The international platform of Ludwig-Maximilians University's Center for Economic Studies and the Ifo Institute

Working Papers

www.cesifo.org/wp

# Dynamic Spatial Panel Models: Networks, Common Shocks, and Sequential Exogeneity

Guido M. Kuersteiner
Ingmar R. Prucha

CESIFO WORKING PAPER NO. 5445
CATEGORY 12: EMPIRICAL AND THEORETICAL MODELS
JULY 2015



***ISSN 2364-1428***

CESifo
Center for Economic Studies & Ifo Institute

CESifo *Working Paper No. 5445*

# Dynamic Spatial Panel Models: Networks, Common Shocks, and Sequential Exogeneity


## Abstract

This paper considers a class of GMM estimators for general dynamic panel models, allowing for cross sectional dependence due to spatial lags and due to unspecified common shocks. We significantly expand the scope of the existing literature by allowing for endogenous spatial weight matrices, time-varying interactive effects, as well as weakly exogenous covariates. The model is expected to be useful for empirical work in both macro and microeconomics. An important area of application is in social interaction and network models where our specification can accommodate data dependent network formation. We discuss explicit examples from the recent social interaction literature. Identification of spatial interaction parameters is achieved through a combination of linear and quadratic moment conditions. We develop an orthogonal forward differencing transformation to aid in the estimation of factor components while maintaining orthogonality of moment conditions. This is an important ingredient to a tractable asymptotic distribution of our estimators. In the social interactions example, orthogonal forward differencing amounts to controlling for unobserved correlated effects by combining multiple outcome measures.


JEL-Code: C010, C210, C230, C310, C330.


*Guido M. Kuersteiner*
*Department of Economics*
*University of Maryland*
*USA - College Park, MD 20742*
*kuersteiner@econ.umd.edu*

*Ingmar R. Prucha*
*Department of Economics*
*University of Maryland*
*USA - College Park, MD 20742*
*prucha@econ.umd.edu*


July 17, 2015

# 1 Introduction[1]

In this paper we develop an estimation theory for a dynamic panel data model with interactive fixed effects and sequentially (rather than only strictly) exogenous regressors.[2] The model allows for cross sectional dependence, which may stem from two potential sources: (i) Cross sectional interactions in the form of "spatial" lags in the endogenous variables, exogenous variables, and/or disturbances. The notion of "space" should be interpreted liberally, and is not confined to geographic space. Importantly, proximity between cross sectional units is not defined by physical location, but in essence only in terms of some measure of distance, which could be a measure of social distance. (ii) Cross sectional dependence may also stem from common shocks, which are accounted for by some underlying $\sigma$-field, but are otherwise left unspecified in line with Andrews (2005) and Ahn et al. (2013). However, in contrast to those papers, and as in Kuersteiner and Prucha (2013), we do not maintain that the data are conditionally i.i.d. The common shocks may effect all variables, including the common factors appearing in the interactive fixed effects. Another important feature is that the weights used in forming the spatial lags are allowed to be stochastic. Our analysis assumes the availability of data indexed by $i = 1, \ldots, n$ in the cross sectional dimension and $t = 1, \ldots, T$ in what is typically referred to as the time dimension, although we allow for a broader interpretation of this index. Our focus is on short panels with $T$ fixed.

The classical dynamic panel data literature has generally assumed that the data are distributed independently and identically in the cross sectional dimension. This included the data on the exogenous variables, which were predominantly treated as sequentially exogenous. The assumption of cross sectional independence is satisfied in many settings where the randomly sampled cross sectional units correspond to individuals, firms, etc., and/or decisions are not interdependent. However in many other settings the assumption may be violated. This includes situations where there are spillovers between units, common

[1] We gratefully acknowledge financial support from the National Institute of Health through the SBIR grant R43 AG027622 and R44 AG027622. We thank David M. Drukker, Stata, for his very helpful collaboration on computations issues. Earlier versions of the paper were presented at the International Panel Data Conference 2013, London, the Econometric Workshop 2104, Shanghai, Joint Statistical Meetings 2014, Boston, Labor Workshop 2014, Laax, VII World Conference of the Spatial Econometrics Association, 2014, Zurich, 14th International Workshop of Spatial Econometrics and Statistics 2015, Paris, as well as at seminars at Michigan State University, Penn State University, Columbia University, University of Rochester, Chicago Booth, University of Michigan, Colorado University and Harvard/MIT. We would like to thank the participants of those conferences and seminars for their helpful comments.

[2] Endogenous regressors in addition to spatial lags of the l.h.s. variable can in principle be accommodated as well, at the cost of additional notation to separate covariates that can be used as instruments from those that cannot. We do not explicitly account for this possibility to save on notation.

shocks, or where the cross sectional units refer to counties, states, countries or industries or generally, in situations where cross-sections are not sampled at random.

A widely used approach to model cross sectional dependence is through common factors; see, e.g., Phillips and Sul (2003, 2007), Bai and Ng (2006a,b), Pesaran (2006), and Andrews (2005). Recent contributions to panel data models with interactive fixed effects include Ahn et al. (2013), Bai (2009, 2013), and Moon and Weidner (2013a,b). This represents an important class of models, however they are not geared towards modeling cross sectional interactions.[3] In addition to general unmodelled cross-sectional dependence stemming from common shocks, our approach also allows for interactive fixed effects as considered in Holtz-Eakin et al. (1988) and Ahn et al. (2001, 2013).

In the spatial literature a widely used approach to model spatial interactions is through spatial lags dating back to Whittle (1954). Those models are often referred to as Cliff-Ord (1973, 1981) type models. Dynamic panel data models that allow for spatial interactions in terms of spatial lags have recently been considered by Mutl (2006), and Yu, de Jong and Lee (2008, 2012), Elhorst (2010), Lee and Yu (2014) and Su and Yang (2014). None of these papers consider interactive fixed effects and common shocks. In contrast to the theory developed in this paper, all of those papers assume that the exogenous variables are strictly exogenous, treat the spatial weights as fixed, do not consider higher order spatial lags, and postulate homoskedasticity for the basic innovations. The literature on panel data models, which allows for both cross sectional interactions in terms of spatial lags and for common shocks, is limited. It includes Chudik and Pesaran (2013), Bai and Li (2013), and Pesaran and Torsetti (2011). All of these papers assume that both $n$ and $T$ tend to infinity, and the latter two papers only consider a static setup.

This paper extends the literature on dynamic panel data models in several important directions by allowing for - possibly higher order - Cliff-Ord type spatial lags in the endogenous and exogenous variables and in the disturbance process, and furthermore by allowing for common shocks. The latter are left unspecified, in contrast to, e.g. Chudik and Pesaran (2013). Our specification is fully dynamic by allowing for both lagged dependent variables and other weakly exogenous and cross-sectionally dependent covariates. An important limitation of the typical specification of Cliff-Ord models is that the spatial interaction matrices are assumed to be known constants. We relax this constraint by allowing for stochastic data

[3] All of these papers assume that both $n$ and $T$ tend to infinity, except for Ahn et al. (2013) and Bai (2013). Apart from not being geared towards modeling cross sectional interactions, a difference between the setups considered in the latter two papers to the setup considered in this paper is that in those papers the observations are modeled as independent in the cross sectional dimension conditional on the common shocks.

dependent weights. The weights can be endogenous in the sense that they can be correlated with the disturbances. Data dependent weights are important tools in modelling network formation and other aspects of social interactions, such as group composition and group heterogeneity. Added flexibility of our model is achieved by allowing for interactive effects that can be used to capture outcome and individual specific effects as well as unmodelled cross-sectional dependence. We exemplify the importance of these extensions to the Cliff-Ord model with applications to the recent literature on social interaction models of Graham (2008) and Carrell, Sacerdote and West (2013). We expect this extension to be important in other applications. An example are models of growth convergence with productivity spillovers depending on trade shares.

We propose a Generalized Method of Moments (GMM) estimator for our extended class of dynamic spatial panel models with unobserved common shocks, and establish its asymptotic properties. With the data and multiplicative factors allowed to depend on common shocks, our asymptotic theory needs to accommodate objective functions that are stochastic in the limit. We present a set of general results that establish the properties of M-estimators in situations with random limiting objective functions and stochastic estimands. Our analysis builds on the classical M-estimation theory summarized in, e.g., Newey and McFadden (1994) for the case of i.i.d. data and Gallant and White (1988), White (1994), and Pötscher and Prucha (1997) for the case of non-i.i.d. data. The CLT developed in this paper extends our earlier results in Kuersteiner and Prucha (2013) to the case of linear-quadratic moment conditions. Quadratic moments play a key role in identifying cross-sectional interaction parameters but pose major challenges in terms of tractability of the weight matrix which in general depends on hard to estimate cross-sectional sums of moments. We achieve significant simplifications and tractability by developing a quasi-forward differencing transformation to eliminate interactive effects while ensuring orthogonality of the transformed moments. This transformation contains the Helmert transformation as a special case. We also provide general results regarding the variances and covariances of linear quadratic forms of forward differences.

The paper is organized as follows. Section 2 introduces the model and considers examples. Section 3 introduces a generalized forward differencing operator to remove interactive effects and provides general results on the variances and covariances of linear quadratic forms. Section 4 defines the GMM estimator and discusses identification. Section 5 contains formal assumptions while Section 6 contains the theoretical results establishing asymptotic properties of the estimators. Concluding remarks are given in Section 7. Basic results regarding stable convergence in distribution as well as all proofs are relegated to the appendices. A supplementary appendix available separately provides additional details for the proofs.

## 2 Model

We consider panel data $\{y_t, x_t, z_t\}_{t=1}^T$ defined on a common probability space $(\Omega, \mathcal{F}, P)$, where $y_t = [y_{1t}, ...., y_{nt}]'$, $x_t = [x'_{1t}, ..., x'_{nt}]'$, and $z_t = [z'_{1t}, ..., z'_{nt}]'$ denote the vector of the endogenous variables, and the matrices of $k_x$ weakly exogenous and $k_z$ strictly exogenous variables. All variables are allowed to vary with the cross sectional sample size $n$, although we suppress this dependence for notational convenience. The specification allows for temporal dynamics in that $x_{it}$ may include a finite number of time lags of the endogenous variables.

Our setup allows for fairly general forms of cross-sectional dependence. As in Andrews (2005) and Kuersteiner and Prucha (2013) we allow in each period $t$ for the regressors and disturbances (and thus for the dependent variable) to be affected by common shocks, which are captured by a sigma field, say, $\mathcal{C}_t \subset \mathcal{F}$ but which is otherwise left unspecified. In the following let $\mathcal{C} = \mathcal{C}_1 \vee \ldots \vee \mathcal{C}_T$ where $\vee$ denotes the sigma field generated by the union of two sigma fields. An important special case where common shocks are not present arises when $\mathcal{C}_t = \mathcal{C} = \{\emptyset, \Omega\}$.

Alternatively or concurrently with common shocks we allow for cross sectional dependence from "spatial lags" in the endogenous variables, the exogenous variables and in the disturbance process. Our specification accommodates higher order spatial lags, as well as time lags thereof which may be included in the $x_{it}$. Spatial lags represent weighted cross sectional averages, where the weights will typically be reflective of some measure of distance between units.[4] We emphasize that distance does not have to be geographic distance, and could, for example and as illustrated below, be a measure of social distance. The spatial weights will be summarized by $n \times n$ spatial weight matrices denoted as $M_{p,t} = (m_{p,ijt})$ and $\underline{M}_{q,t} = (\underline{m}_{q,ijt})$. We do assume that the weights $M_{p,t}$ and $\underline{M}_{q,t}$ are known or observed in the data.

In the following $\varepsilon_t = [\varepsilon_{1t}, ..., \varepsilon_{nt}]'$ denotes the vector of regression disturbances, $u_t = [u_{1t}, ..., u_{nt}]'$ denotes the vector of unobserved idiosyncratic disturbances, and $\mu$ is an $n \times 1$ vector of unobserved factor loadings or individual specific fixed effects, which may be time varying through a common unobserved factor $f_t$. The factor $f_t$ is assumed to be measurable with respect to a sigma field $\mathcal{C}_t$. Furthermore, let $\lambda$ and $\rho$ be $P$ respectively $Q$ dimensional vectors of parameters with typical elements $\lambda_p$ and $\rho_q$ and define $R_t(\lambda) = \sum_{p=1}^P \lambda_p M_{p,t}$

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

and $\underline{R}_t(\rho) = I - \sum_{q=1}^{Q} \rho_q \underline{M}_{q,t}$ for a spatial autoregressive error term or $\underline{R}_t(\rho) = (I + \sum_{q=1}^{Q} \rho_q \underline{M}_{q,t})^{-1}$ for a spatial moving average error term.

The dynamic and cross sectionally dependent panel data model we consider can be written as

$$\begin{aligned} y_t &= R_t(\lambda) y_t + x_t \beta_x + z_t \beta_z + \varepsilon_t = W_t \delta + \varepsilon_t, \\ \underline{R}_t(\rho) \varepsilon_t &= \mu f_t + u_t, \end{aligned} \tag{1}$$

where $W_t = [M_{1,t} y_t, \ldots, M_{P,t} y_t, x_t, z_t]$ and $\delta = [\lambda', \beta']'$ with $\beta = [\beta_x', \beta_z']'$, which are the parameters of interest. As a normalization we take $m_{p,iit} = \underline{m}_{q,iit} = 0$, and $f_T = 1$.

Note that (1) is a system of $n$ equations describing simultaneous interactions between the individual units. The weighted averages, say, $\overline{y}_{p,it} = \sum_{j=1}^{n} m_{p,ijt} y_{jt}$ and $\overline{\varepsilon}_{q,it} = \sum_{j=1}^{n} \underline{m}_{q,ijt} \varepsilon_{jt}$ model contemporaneous direct cross-sectional interactions in the dependent variables and the disturbances. Those weighted averages are typically referred to as spatial lags, and the corresponding parameters are typically referred to as spatial autoregressive parameters. The specification of spatial lags dates back to Cliff and Ord (1973,1981).[5] In contrast to standard specifications of Cliff-Ord type spatial models we do not assume that the weights are given constants, but allow them to be stochastic. The weights are allowed to be endogenous in that they can depend on $\mu_1, \ldots, \mu_n$ and $u_{it}$, apart from predetermined variables and common shocks, and thus can be correlated with the disturbances $\varepsilon_t$.[6] In fact, we do not impose any particular restrictions on how the weights are generated. This extension is important for example to model sequential group formation as in Carrell et.al. (2013) or endogenous network formation as in Goldsmith-Pinkham and Imbens (2013).

For $i = 1, ..., n$ let $z_i^o = (z_{i1}, \ldots, z_{iT})$, $x_{it}^o = [x_{i1}, \ldots, x_{it}]$, $u_{it}^o = [u_{i1}, \ldots, u_{it}]$, $u_{-i,t} = [u_{i1}, \ldots, u_{i-1,t}, u_{i+1,t}, \ldots u_{nt}]$. We next formulate our main moment conditions for the idiosyncratic disturbances.

**Assumption 1** *Let $K_u$ be some finite constant (which is taken, w.o.l.o.g., to be greater then one), and define the sigma fields*

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

[5] An alternative specification, analogous to specifications considered in Baltagi et al (2008), would be to model the disturbance process in (1) as $\varepsilon_t = \phi f_t + v_t$, where $\phi$ and $v_t$ follow possibly different spatial autoregressive processes. Since we are not making any assumptions on the unobserved components $\mu$ it is readily seen that the above specification includes this case, provided that the spatial weights do not depend on $t$.

[6] In addition to the spatial lags in $y_t$ other endogenous variables could be readily included on the r.h.s. of (1). We do not explicitly list those variables for notational simplicity. In cases where the weights depend on $u_{it}$ contemporaneous spatial lags in the exogenous variables would be separated out of $x_t$ and $z_t$, and included as additional endogenous variables.

*and*

$$\mathcal{Z}_n = \sigma(\{z_j^o, \mu_j\}_{j=1}^n).$$

*For some $\delta > 0$ and all $t = 1, \ldots, T$, $i = 1, \ldots, n$, $n \geq 1$:*
*(i) The $2 + \delta$ absolute moments of the random variables $x_{it}, z_{it}$, $u_{it}$, and $\mu_i$ exist, and the moments are uniformly bounded by $K$.*
*(ii) Then the following conditional moment restrictions hold for some constant $c_u > 0$:*

$$E\left[u_{it}|\mathcal{B}_{n,i,t} \vee \mathcal{C}\right] = 0, \tag{2}$$

$$E\left[u_{it}^2|\mathcal{B}_{n,i,t} \vee \mathcal{C}\right] = \sigma_t^2 \varrho_i^2 \quad \text{with} \quad \sigma_t^2, \varrho_i^2 \geq c_u, \tag{3}$$

$$E\left[|u_{it}|^{2+\delta} |\mathcal{B}_{n,i,t} \vee \mathcal{C}\right] \leq K_u. \tag{4}$$

*The variance components $\gamma_\sigma = (\sigma_1^2, \ldots, \sigma_T^2)'$ are assumed to be measurable w.r.t. $\mathcal{C}$. The variance components $\varrho_i^2 = \varrho_i^2(\gamma_\varrho)$ are taken to depend on a finite dimensional parameter vector $\gamma_\varrho$ and are assumed to be measurable w.r.t. $\mathcal{Z}_n \vee \mathcal{C}$.*

Condition (2) clarifies the distinction between weakly exogenous covariates $x_{it}$ and strictly exogenous covariates $z_{it}$. The later enter the conditioning set at all leads and lags. We shall also use the notation $\Sigma_\sigma = \operatorname{diag}(\sigma_t^2)$ and $\Sigma_\varrho = \operatorname{diag}(\varrho_i^2)$. As a normalization we may, e.g., take $\sigma_T^2 = 1$ or $n^{-1}\operatorname{tr}(\Sigma_\varrho) = 1$. Specifications where $\sigma_t^2$ and $\varrho_i^2$ are non-stochastic, and specifications where the $u_{it}$ are conditionally homoskedastic are covered as special cases.

The reduced form of the model is given by

$$\begin{aligned} y_t &= (I_n - R_t(\lambda))^{-1} W_t\delta + (I_n - R_t(\lambda))^{-1} \varepsilon_t, \\ \varepsilon_t &= \underline{R}_t(\rho)^{-1} (\mu f_t + u_t). \end{aligned} \tag{5}$$

Applying a Cochrane-Orcutt type transformation by premultiplying the first equation in (1) with $\underline{R}_t(\rho)$ yields

$$\underline{R}_t(\rho) y_t = \underline{R}_t(\rho) W_t \delta + \mu f_t + u_t. \tag{6}$$

The first example illustrates the use of both spatial interaction terms and interactive effects in a social interaction model.

**Example 1 (Social Interactions)** *Graham (2008) considers a linear social interactions model. Here we analyze a modification of Graham (2008) assuming that we have, as is the case for the data Graham (2008) analyzes, two measures of academic achievement $y_{crt}$ where $c$ is the classroom index, $r$ is the individual student in classroom $c$, and $t = 1, 2$ indicates two distinct performance measures. These could be the quantitative and verbal* *scores on a standardized test. We use two outcome measures to control for unobserved and possibly correlated student characteristics. Identification rests on the assumption that all unobserved, cross-sectionally correlated characteristics affect both outcomes in the same way, up to an unobserved scale factor which can very with the performance measure. To the best of our knowledge this is a new identification strategy in the social interactions literature where disentangling unobserved correlated effects from social interactions is a major challenge. More specifically, student specific characteristics are modelled using a common factor structure of the form* $u_{crt} + \kappa_{cr} f_t$ *where* $\kappa_{cr}$ *is a student fixed effect for student* $r$ *in classroom* $c$, $f_t$ *is a common factor, and the* $u_{crt}$ *are idiosyncratic components. Student fixed effects* $\kappa_{cr}$ *are allowed to be arbitrarily correlated between individuals.*

*Assuming a sample of* $C$ *classrooms of size* $n_c$, *and a total number of students* $n = n_1 + ... + n_C$, *we consider the following version of Graham's model:*[7]

$$y_{crt} = \alpha_c f_t + (u_{crt} + \kappa_{cr} f_t) + \rho \left[ (n_c - 1)^{-1} \sum_{l=1, l \neq r}^{n_c} (u_{clt} + \kappa_{cl} f_t) \right]$$

*where* $\alpha_c$ *captures class room level heterogeneity. Organizing the data by class room for each test* $t$, *let* $y_{ct} = [y_{c1t}, ...., y_{cn_ct}]'$, $u_{ct} = [u_{c1t}, ...., u_{cn_ct}]'$*and* $\kappa_c = [\kappa_{c1}, \ldots, \kappa_{cn_c}]'$. *Defining* $\mathbf{1}_c = (1, ...., 1)'$ *as a* $n_c \times 1$ *vector,* $M_{(c)} = (\mathbf{1}_c \mathbf{1}_c' - I_c) / (n_c - 1)$ *where* $I_c$ *is the* $n_c \times n_c$ *identity matrix, Graham's model can be written as* $y_{ct} = \mu_c^* f_t + \left(I_c + \rho M_{(c)}\right) u_{ct}$ *with* $\mu_c^* = \alpha_c \mathbf{1}_c + (I_c + \rho M_{(c)}) \kappa_c$, *or in stacked notation*

$$y_t = \mu^* f_t + (I_n + \rho M) u_t \tag{7}$$

*where* $y_t = [y_{1t}', ..., y_{Ct}']'$, $M = \text{diag}_{c=1}^{C} \left(M_{(c)}\right)$, $u_t = [u_{1t}', ..., u_{Ct}']'$ *and* $\mu^* = [\mu_1^{*\prime}, \ldots, \mu_C^{*\prime}]'$. *It is easily seen that model (7) is a special case of what we called the reduced form (5) if we impose the restrictions* $\delta = 0$, $R_t(\lambda) = 0$, $\mu = (I_n + \rho M)^{-1} \mu^*$ *and* $\underline{R}_t(\rho) = (I_n + \rho M)^{-1}$. *The inverse of* $I_n + \rho M$ *exists (in closed form) as long as* $\rho \neq -1$ *and* $\rho \neq n_c - 1$. *for all c. Let* $s_c$ *be an indicator variable for small classes, and define* $z_{rc} = s_c$. *Furthermore, let* $i = i(r, c)$ *be a one-to-one mapping and consider the simplified information set* $\mathcal{B}_{n,i(r,c),t} = \sigma(\{z_j^o, u_{t-1j}^o, \mu_j\}_{j=1}^n, u_{t,-i(r,c)})$. *Then, consistent with Graham (2008), we assume*

$$E\left[u_{rct} | \mathcal{B}_{n,i(r,c),t} \vee \mathcal{C}\right] = 0 \tag{8}$$

*and model the conditional variances as*

$$E\left[u_{rct}^2 | \mathcal{B}_{n,i(r,c),t} \vee \mathcal{C}\right] = \sigma_t^2 \varrho_{i(r,c)}^2 \tag{9}$$

[7] Unlike Graham we use averages that do not include an individual's own characteristics to account for peer interaction effects. This modification is a simple parameter normalization that is convenient for our purposes. Our identification strategy does not rely on this normalization.

*where* $\varrho^2_{i(r,c)} = \varrho^2_L s_c + \varrho^2_S(1 - s_c)$*. We note that both moment restrictions can be relaxed for purposes of identifying the model. As a normalization we take* $\varrho^2_S = 1$ *and denote* $\gamma_\varrho = \varrho^2_L$*. (Class room heterogeneity can be captured through the fixed effects.) Since* $E[y_t|\mathcal{C}] = E[\mu^*|\mathcal{C}]\, f_t$ *the common factor* $f_t$ *accounts for different means in scores of the two tests. These could be due to different levels of difficulty, or different standards in grading. The assumption that* $\kappa_{cr}$ *is constant across the two tests implies that on average (but not necessarily individually) higher ability students tend to do better on both tests.*
*Clearly the above moment conditions are covered as a special case of (2) and (3). We note that, while remaining within our general framework, the above example can be extended to accommodate some randomness in the assignment of students to classrooms, assuming that the selection process is defined by a set of strictly exogenous variables, and that the elements of* $M$ *are measurable w.r.t. the corresponding* $\sigma$*-field. The matrix* $M$ *then provides a convenient description of the realized network.*[8]

The next example is based on Carrell et. al. (2013) with certain simplifications to aid exposition. It illustrates the use of higher order, and data-dependent spatial lags to model within-group heterogeneity. By allowing $R_t(\lambda)$ to depend on predetermined outcomes we can accommodate the fact that, as in Carrell et.al. (2013), group membership is not exogenous.

**Example 2 (Group Level Heterogeneity)** *With the same notation as in Example 1 let* $D_c$ *be a diagonal matrix of size* $n_c$ *with diagonal elements* $d_{jj} = 1$ *if student* $j$ *is a high type (for example as measured by SAT scores) and* $d_{jj} = 0$ *if student* $j$ *is a low type. The test score for test* $t$ *and class* $c$ *then is, following Carrell et.al., given by* $y_{ct} = 1_c\alpha_c f_t + \left(\lambda_1 (I - D_c) M_{(c)} + \lambda_2 D_c M_{(c)}\right) y_{ct} + u_{ct}$*, or in stacked notation*

$$y_t = (\lambda_1 M_1 + \lambda_2 M_2)\, y_t + \mu f_t + u_t \tag{10}$$

*with* $M_1 = \mathrm{diag}_{c=1}^{C}\left[(I - D_c) M_{(c)}\right]$, $M_2 = \mathrm{diag}_{c=1}^{C}\left[D_c M_{(c)}\right]$*, and* $\mu = [1_1'\alpha_1, \ldots, 1_C'\alpha_C]'$*. The implication of the model is that students with high type react differently to average performance in the class than students with low type. In this example* $R_t(\lambda) = \lambda_1 M_1 + \lambda_2 M_2$*, and* $\underline{R}_t(\rho) = I$*.*

Apart from applications to social interactions the model in (1) has a wide range of potential applications in empirical micro, macro, regional science and urban economics, geography and health. Our setup contains the standard Cliff-Ord type spatial models as

[8]For further discussions of the use of Cliff-Ord type spatial models in modeling social interactions see, e.g., Liu and Lee (2010), Lee et al. (2010), Blume et al. (2011), and Patacchini et al. (2013).

special cases. In the past such models have, for example, been applied in empirical studies of spatial price competition, growth convergence, real estate prices, financial contagions, and technology adoption. The next example is in the area of health, and considers the spread of an infectious disease.

**Example 3 (Infectious Disease)** *Let $y_{it}$ denote the rate of infection with genital herpes in county $i$ in period $t$, and consider the following illustrative model:*[9]

$$y_{it} = \lambda \overline{y}_{it} + y_{it-1}\phi + \overline{y}_{it-1}\overline{\phi} + r_{it}\gamma + \overline{r}_{it}\overline{\gamma} + g_{it}\delta + \overline{g}_{it}\overline{\delta} + \mu f_t + \varepsilon_{it}, \tag{11}$$

*where $r_{it}$ denotes a vector of sequentially exogenous variables, which may include the rate of vaccination, $g_{it}$ denotes a vector of strictly exogenous variables, which may include demographic variables, and $\overline{y}_{it}$, $\overline{r}_{it}$, and $\overline{g}_{it}$ denote spatial lags to account for interactions of the population across counties. Then clearly the above dynamic model is a special case of the scalar representation of model (1) with $x_{it} = [y_{it-1}, \overline{y}_{it-1}, r_{it}, \overline{r}_{it}]$, $z_{it} = [g_{it}, \overline{g}_{it}]$ and parameter vectors $\beta_x = [\phi, \overline{\phi}, \gamma', \overline{\gamma}']'$ and $\beta_z = [\delta', \overline{\delta}']'$. Now suppose that the spatial weights of the spatial lag $\overline{y}_{it}$ are modelled as $m_{ij} = \sum_{p=1}^{P} \alpha_p (1/d_{ij})^p$ where $d_{ij}$ is a distance measure and the $\alpha_r$ are unknown parameters. Then*

$$\lambda \overline{y}_{it} = \lambda \sum_{j=1}^{n} m_{ij} y_{jt} = \sum_{p=1}^{P} \lambda_p \left[ \sum_{j=1}^{n} m_{p,ij} y_{jt} \right] \tag{12}$$

*with $\lambda_p = \lambda \alpha_p$ and $m_{p,ij} = (1/d_{ij})^p$, and substitution of the expression in (12) into (11) yields a p-order spatial autoregressive model as considered in (1).*

In the above examples the spatial weights do not vary with $t$. We emphasize that in our general model we allow for the spatial weights to vary with $t$, and to depend on sequentially and strictly exogenous variables. As a result, the model can also be applied to situations where the location decision of a unit is a function of sequentially and strictly exogenous variables, in that we can allow for the distance between units to vary with $t$ and to depend on those variables.

# 3 Forward Differencing and Orthogonality of Linear Quadratic Forms

In the classical panel literature the Helmert transformation was proposed by Arellano and Bover (1995) as an alternative forward filter that, unlike differencing, eliminates fixed ef-

[9] The specification is in line with specifications considered in, e.g., Keeling and Rohani (2008) and Chen et al. (2014).

fects without introducing serial correlation.[10] In this section we introduce an orthogonal quasi-forward differencing transformation for the more general case where factors $f_t$ appear in the model. Our moment conditions involve both linear and quadratic forms of forward differenced disturbances. Thus we also present a general result on the variances and covariances of linear quadratic forms, which establishes sufficient conditions for the orthogonality of both the linear and the quadratic forms.

For $\eta_{ti} = \mu_i f_t + u_{it}$ and $t = 1, \ldots, T-1$ consider the forward differences

$$\eta_{it}^{+} = \textstyle\sum_{s=t}^{T} \pi_{ts}\eta_{is}, \qquad u_{it}^{+} = \textstyle\sum_{s=t}^{T} \pi_{ts}u_{is}$$

with $\pi_t = [0, \ldots, 0, \pi_{tt}, \ldots, \pi_{tT}]$. Now define the upper triangular $T-1 \times T$ matrix $\Pi = [\pi_1', .., \pi_T']'$ and let $f = [f_1, \ldots, f_T]$. Then $\Pi f = 0$ is a sufficient condition for the transformation to eliminate the unit specific components such that $u_{it}^{+} = \eta_{it}^{+}$. If in addition $\Pi\Sigma_\sigma\Pi' = I$ then under our assumptions the transformed errors $u_{it}^{+}$ will be uncorrelated across $i$ and $t$. The following proposition introduces such a transformation. To emphasize that the elements of $\Pi$ are functions of the $f_t$'s and $\sigma_t$'s we sometimes write $\pi_{ts}(f, \gamma_\sigma)$.

**Proposition 1** *(Generalized Helmert Transformation) Let $F = (f_{ts})$ be a $T-1 \times T$ quasi differencing matrix with diagonal elements $f_{tt} = 1$, $f_{t,t+1} = -f_t/f_{t+1}$, and all other elements zero. Let $U$ be an upper triagonal $T-1 \times T-1$ matrix such that $F\Sigma_\sigma F' = UU'$. Then, the $T-1 \times T$ matrix $\Pi = U^{-1}F$ is upper triagonal and satisfies $\Pi f = 0$ and $\Pi\Sigma_\sigma\Pi' = I$. Explicit formulas for the elements of $\Pi = \Pi(f, \gamma_\sigma)$ are given as*

$$\begin{aligned}
&\pi_{tt}(f, \gamma_\sigma) = \left(\sqrt{\phi_{t+1}/\phi_t}\right)/\sigma_t, \\
&\pi_{ts}(f, \gamma_\sigma) = -f_t f_s \left(\sqrt{\phi_{t+1}/\phi_t}\right) / \left(\phi_{t+1}\sigma_t\sigma_s^2\right) \text{ for } s > t, \\
&\pi_{ts} = 0 \text{ for } s < t.
\end{aligned}$$

*with $\phi_t = \sum_{\tau=t}^{T}(f_\tau/\sigma_\tau)^2$ For computational purposes observe that $\phi_t = (f_t/\sigma_t)^2 + \phi_{t+1}$. Also note that if $\sigma_T^2 = 1$ as a normalizations, then $f_T/\sigma_T = 1$.*

Proposition 1 is an important result because it gives explicit expressions for the elements of $\Pi$. Such expression are crucial from a computational point of view, especially if $f_t$ is estimated as an unobserved parameter of the model. Although we do not adopt this in the following, for computational purposes it may furthermore be convenient to re-parameterize the model in terms $\underline{f}_t = f_t/\sigma_t$ and $\sigma_t$ in place of $f_t$ and $\sigma_t$. We note that for $f_t = 1$ and $\sigma_t = 1$ we obtain as a special case the Helmert transformation with $\pi_{tt} = \sqrt{(T-t)/(T-t+1)}$ and $\pi_{ts} = -\sqrt{(T-t)/(T-t+1)}/(T-t)$ for $s > t$.

[10] Hayakawa (2006) extends the Helmert transformation to systems estimators of panel models by using arguments based on GLS transformations similar to Hayashi and Sims (1983) and Arellano and Bover (1995).

We also note that because $Ff = 0$ any transformation of the form $\Pi(f, \bar{\gamma}_\sigma) = \bar{U}^{-1}F$ with $F\bar{\Sigma}_\sigma F' = \bar{U}\bar{U}'$ and $\bar{\Sigma}_\sigma = \mathrm{diag}(\bar{\gamma}_\sigma)$ some positive diagonal matrix removes the interactive effect. An important special case is the transformation with weights $\pi_{ts}(f, 1_T)$ corresponding to $\bar{\Sigma}_\sigma = I_T$.

In (1) the disturbance process was specified to depend only on a single factor for simplicity. Now suppose that the disturbance process is generalized to $\underline{R}_t(\rho)\varepsilon_t = \mu^1 f_t^1 + \ldots + \mu^P f_t^P + u_t$ where $f_t^p$ denotes the $p$-th factor and $\mu^p$ the corresponding vector of factor loadings. We note that multiple factors can be handled by recursively applying the above generalized Helmert transformation, yielding a $T - P \times T$ transformation matrix $\Pi = \Pi_P \ldots \Pi_2 \Pi_1$ where the matrices $\Pi_p$ are of dimension $(T-p) \times (T-p+1)$, and $\Pi_1 \Sigma_\sigma \Pi_1' = I_{T-1}$, $\Pi_p \Pi_p' = I_{T-p}$ for $p > 1$, and $\Pi_p(\Pi_{p-1} ... \Pi_1 f^p) = 0$ with $f^p = [f_1^p, \ldots, f_T^p]'$. Of course, this in turn implies that $\Pi \Sigma_\sigma \Pi' = I_{T-P}$ and $\Pi[f^1, \ldots, f^P] = 0$. The elements of each of the $\Pi_p$ matrices have the same structure as those given in Proposition 1. A more detailed discussion, including a discussion of a convenient normalization for the factors, is given in an supplementary appendix.

We next give a general result on the variance covariances of linear quadratic forms based on forward differenced, but not necessarily orthogonally forward differences, disturbances. The optimal weight matrix of a GMM estimator based on both linear and quadratic moment conditions depends on these covariances. Simplifying them as much as possible is critical to the implementation of the estimator. Our result establishes the conditions under which such simplifications can be achieved. We also give sufficient conditions for the validity of linear and quadratic moment conditions.

**Proposition 2** [11] *Let the information sets $\mathcal{B}_{n,i,t}$, $\mathcal{B}_{n,t}$, $\mathcal{Z}_n$ be as defined in Section 2. Furthermore assume that for all $t = 1, \ldots, T$, $i = 1, \ldots, n$, $n \geq 1$, $E[u_{it}|\mathcal{B}_{n,i,t} \vee \mathcal{C}] = 0$, $E[u_{it}^2|\mathcal{B}_{n,i,t} \vee \mathcal{C}] = \varrho_i^2 \sigma_t^2 > 0$, $E[u_{it}^3|\mathcal{B}_{n,i,t} \vee \mathcal{C}] = \mu_{3,it}$, $E[u_{it}^4|\mathcal{B}_{n,i,t} \vee \mathcal{C}] = \mu_{4,it}$, where $\sigma_t$ is finite and measurable w.r.t. $\mathcal{C}$, and $\varrho_i$, $\mu_{3,it}$ and $\mu_{4,it}$ are finite and measurable w.r.t. $\mathcal{Z}_n \vee \mathcal{C}$. Define $\Sigma_\varrho = \mathrm{diag}(\varrho_1^2, ..., \varrho_n^2)$ and $\Sigma_\sigma = \mathrm{diag}(\sigma_1^2, ..., \sigma_T^2)$. Let $A_t = (a_{ijt})$ and $B_t = (b_{ijt})$ be $n \times n$ matrices, and let $a_t = (a_{it})$ and $b_t = (b_{it})$ be $n \times 1$ vectors, where $a_{ijt}$, $b_{ijt}$, $a_{it}$, $b_{it}$ are measurable w.r.t. $\mathcal{B}_{n,t} \vee \mathcal{C}$. Let $\pi_t = [0, \ldots, 0, \pi_{tt}, \ldots, \pi_{tT}]$ and $\gamma_t = [0, \ldots, 0, \gamma_{tt}, \ldots, \gamma_{tT}]$ be $1 \times T$ vectors where $\pi_{t\tau}$ and $\gamma_{t\tau}$ are measurable w.r.t. $\mathcal{C}$, and consider the forward differences $u_t^+ = [u_{1t}^+, \ldots, u_{nt}^+]'$ and $u_t^\times = [u_{1t}^\times, \ldots, u_{nt}^\times]'$ with*

$$u_{it}^+ = \textstyle\sum_{s=t}^T \pi_{ts} u_{is} = \pi_t u_i', \qquad \textit{and} \qquad u_{it}^\times = \textstyle\sum_{s=t}^T \gamma_{ts} u_{is} = \gamma_t u_i'.$$

[11] Further details and an explicit proof are given in the Supplementary Appendix B.

*Then*

$$E\left[u_t^{+\prime}A_t u_t^{\times} + u_t^{+\prime}a_t|\mathcal{C}\right] = \pi_t\Sigma_\sigma\gamma_t \operatorname{tr}\left[E\left(A_t\Sigma_\varrho|\mathcal{C}\right)\right], \tag{13}$$

$$\begin{aligned}
&\operatorname{Cov}(u_t^{+\prime}A_t u_t^{\times} + a_t' u_t^{+}, u_t^{+\prime}B_t u_t^{\times} + b_t' u_t^{+}|\mathcal{C}) \qquad (14)\\
&\quad = (\pi_t\Sigma_\sigma\pi_t')(\gamma_t\Sigma_\sigma\gamma_t')E\left[\operatorname{tr}(A_t\Sigma_\varrho B_t'\Sigma_\varrho)|\mathcal{C}\right] + (\pi_t\Sigma_\sigma\gamma_t')^2 E\left[\operatorname{tr}(A_t\Sigma_\varrho B_t\Sigma_\varrho)|\mathcal{C}\right]\\
&\quad +(\pi_t\Sigma_\sigma\pi_t')E\left[a_t'\Sigma_\varrho b_t|\mathcal{C}\right] + \mathcal{K}_1,\\
&\operatorname{Cov}(u_t^{+\prime}A_t u_t^{\times} + a_t' u_t^{+}, u_s^{+\prime}B_s u_s^{\times} + b_s' u_s^{+}|\mathcal{C}) = \mathcal{K}_2 \qquad \textit{for all } t > s, \qquad (15)
\end{aligned}$$

*where $\mathcal{K}_1$ and $\mathcal{K}_2$ are random functionals that depend on $a_t$, $b_t$, $A_t$ and $B_t$. Explicit expressions for $\mathcal{K}_1$ and $\mathcal{K}_2$ are given in the supplementary appendix. Sufficient conditions that ensure that $E\left[u_t^{+\prime}A_t u_t^{\times} + u_t^{+\prime}a_t|\mathcal{C}\right] = 0$ and that $\mathcal{K}_1 = \mathcal{K}_2 = 0$ are that $\operatorname{vec}_D(A_t) = \operatorname{vec}_D(B_t) = 0$, $\Pi = \Gamma$ with $\Pi f = 0$ and $\Pi\Sigma_\sigma\Pi' = I$. Specialized expressions for $\mathcal{K}_1$ and $\mathcal{K}_2$ when one or several of these conditions fail are again given in the supplementary appendix.*

First consider the special case where $\Sigma_\varrho = \varrho^2 I$. In this case a sufficient condition for the validity of the moment conditions $E\left[u_t^{+\prime}A_t u_t^{\times} + u_t^{+\prime}a_t|\mathcal{C}\right] = 0$ is that $\operatorname{tr}(A_t) = 0$. Consistent with this observation and under cross sectional homoskedasticity, quadratic moment conditions where only the trace of the weight matrices is assumed to be zero, have been considered frequently in the spatial literature[12]However, $\operatorname{tr}(A_t) = 0$ does not insure that the linear quadratic forms are uncorrelated across time because the terms $\mathcal{K}_1$ and $\mathcal{K}_2$ are not necessarily zero, even in the case of orthogonally transformed disturbances, i.e., $\Pi = \Gamma$ and $\Pi\Sigma_\sigma\Pi' = I$. This is in contrast to the case of pure linear forms (where $A_t = B_t = 0$). The terms $\mathcal{K}_1$ and $\mathcal{K}_2$ depend on cross-sectional sums that are potentially difficult to estimate.

Next consider the case where (possibly) $\Sigma_\varrho \neq \varrho^2 I$. In this case a sufficient condition for $E\left[u_t^{+\prime}A_t u_t^{\times} + u_t^{+\prime}a_t|\mathcal{C}\right] = 0$ is that $\operatorname{vec}_D(A_t) = 0$. We note that with $\operatorname{vec}_D(A_t) = 0$ no restrictions on $E\left[u_{it}^2|\mathcal{B}_{n,i,t} \vee \mathcal{C}\right]$ are necessary to ensure $E\left[u_t^{+\prime}A_t u_t^{\times} + u_t^{+\prime}a_t|\mathcal{C}\right] = 0$. An inspection of $\mathcal{K}_1$ and $\mathcal{K}_2$ shows that strengthening the assumptions to $\operatorname{vec}_D(A_t) = \operatorname{vec}_D(B_t) = 0$ for all $t$ and using orthogonally transformed disturbances ensures that $\mathcal{K}_1 = \mathcal{K}_2 = 0$, and thus simplifies the optimal GMM weight matrix. Also note that under this setup the expressions on the r.h.s. of (14) simplify to $E\left[\operatorname{tr}(A_t\Sigma_\varrho(B_t + B_t')\Sigma_\varrho)|\mathcal{C}\right] + E\left[a_t'\Sigma_\varrho b_t|\mathcal{C}\right]$, while (15) implies that the linear quadratic forms are uncorrelated over time. Another important implication of Proposition 2 is that under this setup the covariances between linear sample moments and quadratic sample moments are zero. Expressions for the variance of linear quadratic forms are obtained as a special case where $A_t = B_t$ and $a_t = b_t$. The results

[12] See, e.g., Kelejian and Prucha (1998,1999), Lee and Liu (2010) and Lee and Yu (2014).

of Proposition 2 are consistent with some specialized results given in Kelejian and Prucha (2001, 2010) under the assumption that the coefficients $a_t$ and $A_t$ in the linear quadratic forms are non-stochastic.

# 4 Moment Conditions and GMM Estimator

For clarity we denote the true parameters of interest and the true auxiliary parameters as $\theta_0 = (\delta_0', \rho_0', f_0')'$ and $\gamma_0 = \left(\gamma_{0,\varrho}', \gamma_{0,\sigma}'\right)'$. Using (6) we define

$$u_t^+(\theta_0, \gamma_\sigma) = \sum_{s=t}^T \pi_{ts}\left(f_0, \gamma_\sigma\right) u_s = \sum_{s=t}^T \pi_{ts}\left(f_0, \gamma_\sigma\right) \underline{R}_s(\rho_0)\left[y_s - W_s\delta_0\right], \tag{16}$$

with the weights $\pi_{ts}(.,.)$ of the forward differencing operation defined by Proposition 1. Note that this operation removes the unobserved individual effects even if $\gamma_\sigma \neq \gamma_{0,\sigma}$. Our estimators utilize both linear and quadratic moment conditions based on

$$u_{*t}^+(\theta_0, \gamma) = \Sigma_\varrho(\gamma_\varrho)^{-1/2} u_t^+(\theta_0, \gamma_\sigma). \tag{17}$$

Considering moment conditions based on $u_{*t}^+(\theta_0, \gamma)$ is sufficiently general to cover initial estimators with $\Sigma_\sigma = I_T$ and $\Sigma_\varrho = I_n$. Quadratic moment conditions are often required to identify parameters associated with spatial lags in the disturbance process and may further increase the efficiency of estimators due to spatial correlation in the data generating process. Quadratic moment conditions have been exploited routinely in the spatial literature. They can be motivated by inspecting the score of the log-likelihood function of spatial models; see, e.g., Anselin (1988, p. 64) for the score of a spatial ARAR(1,1) model. Quadratic moment conditions were introduced by Kelejian and Prucha (1998,1999) for GMM estimation of a cross sectional spatial ARAR(1,1) model, and have more recently been used in the context of panel data; see, e.g., Kapoor, Kelejian and Prucha (2007), Lee and Yu (2014).

Let $h_{it} = (h_{it}^r)$ be some $1 \times p_t$ vector of instruments, where the instruments are measurable w.r.t. $\mathcal{B}_{n,t} \vee \mathcal{C}$. Also, consider the $n \times 1$ vectors $h_t^r = (h_{it}^r)_{i=1,\ldots,n}$,, then by Assumption 1 and Proposition 2 we have the following linear moment conditions for $t = 1, \ldots, T-1$,

$$E\begin{bmatrix} h_t^{1\prime} u_{*t}^+(\theta_0, \gamma) \\ \vdots \\ h_t^{p_t\prime} u_{*t}^+(\theta_0, \gamma) \end{bmatrix} = E\left[\sum_{i=1}^n h_{it}' u_{*it}^+(\theta_0, \gamma)\right] = 0 \tag{18}$$

with $u_{*it}^+(\theta_0, \gamma) = u_{it}^+(\theta_0, \gamma_\sigma)/\varrho_i(\gamma_\varrho)$. For the quadratic moment conditions, let $a_{ij,t} = (a_{ij,t}^r)$ be a $1 \times q_t$ vector of weights, where the weights are measurable w.r.t. $\mathcal{B}_{n,t} \vee \mathcal{C}$. Also consider the $n \times n$ matrices $A_t^r = (a_{ij,t}^r)_{i,j=1,\ldots,n}$ such that by Assumption 1 and Proposition 2, and imposing the constraint that $a_{ii,t} = 0$ one obtains the following quadratic moment conditions for $t = 1, \ldots, T-1$,

$$E\begin{bmatrix} u_{*t}^{+}(\theta_0,\gamma)' A_t^1 u_{*t}^{+}(\theta_0,\gamma) \\ \vdots \\ u_{*t}^{+}(\theta_0,\gamma)' A_t^{q_t} u_{*t}^{+}(\theta_0,\gamma) \end{bmatrix} = E\left[\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij,t}' u_{*it}^{+}(\theta_0,\gamma) u_{*jt}^{+}(\theta_0,\gamma)\right] = 0. \qquad (19)$$

The requirement that $a_{ii,t} = 0$ is generally needed for (19) to hold, unless $\Sigma_{0,\varrho} = I_n$. W.o.l.o.g. we also maintain that $a_{ij,t} = a_{ji,t}$.

By allowing for subvectors of $h_{it}$ and $a_{ij,t}$ to be zero and by setting $q_t = p_t$, the above moment conditions can be stacked and written more compactly as

$$\begin{aligned} E\overline{m}_t(\theta_0,\gamma) &= 0, \qquad \text{with} \qquad (20) \\ \overline{m}_t(\theta_0,\gamma) &= n^{-1/2}\sum_{i=1}^{n} h_{it}' u_{*it}^{+}(\theta_0,\gamma) + n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij,t}' u_{*it}^{+}(\theta_0,\gamma) u_{*jt}^{+}(\theta_0,\gamma). \end{aligned}$$

It is worth noting that the formulation in (20) allows for nontrivial linear combinations of (18) and (19) in addition to simply stacking both sets of moments. The particular form of (20) is motivated by a need to minimize cross-sectional and temporal correlations between empirical moments. Proposition 2 shows that only a very judicious choice of moment conditions, moment weights $A_t$ and forward differences $\Pi$ leads to a moment vector covariance matrix that can be estimated reasonably easily.

Let $\theta = (\delta', \rho', f')'$ and $\gamma = \left(\gamma_\varrho', \gamma_\sigma'\right)'$ denote some vector of parameters, let $p = \sum_{t=1}^{T-1} p_t$, and define the $p \times 1$ normalized stacked sample moment vector corresponding to (20) as

$$\overline{m}_n(\theta,\gamma) = \left[\overline{m}_1(\theta,\gamma)', \ldots, \overline{m}_{T-1}(\theta,\gamma)'\right]. \qquad (21)$$

For some estimator $\bar{\gamma}_n$ of the auxiliary parameters $\gamma$ and a $p \times p$ moment weights matrix $\tilde{\Xi}_n$ the GMM estimator for $\theta_0$ is defined as

$$\tilde{\theta}_n\left(\bar{\gamma}_n\right) = \arg\min_{\theta \in \underline{\Theta}_\theta} n^{-1}\overline{m}_n(\theta,\bar{\gamma}_n)' \tilde{\Xi}_n \overline{m}_n(\theta,\bar{\gamma}_n) \qquad (22)$$

where the parameter space $\underline{\Theta}_\theta$ is defined in more detail in Section 5. The parameter $\gamma$ is a nuisance parameter that can either be fixed at an a priori value or estimated in a first step.

For the practical implementation of $\tilde{\theta}_n$ choices of the instruments $h_{it}$ and weights $a_{ijt}$ need to be made. Clearly $x_{it}^{o}$ and $z_i$ are available as possible instruments. However, when the spatial weights are measurable w.r.t. $\mathcal{B}_{n,t} \vee \mathcal{C}$, then taking guidance from the spatial literature the instrument vector $h_{it}$ may not only contain $x_{it}^{o}$ and $z_i$, but also spatial lags thereof. One motivation for this is that for classical spatial autoregressive models the conditional mean of the explanatory variables can be expressed as a linear combination of the exogenous regressors and spatial lags thereof, including higher order spatial lags. Again, when the spatial weights are measurable w.r.t. $\mathcal{B}_{n,t} \vee \mathcal{C}$, then taking guidance from the spatial literature possible choices for the matrices $A_t^r = (a_{ijt}^r)$ include the spatial weight matrices up to period $t$ and powers thereof (with the diagonal elements set to zero). With endogenous weights, in the sense that the weights also depend on contemporaneous idiosyncratic disturbances, possible candidates for $A_t^r$ can be based on projections of the weights onto $\mathcal{B}_{n,t} \vee \mathcal{C}$, or can be constructed from spatial weight matrices up to period $t-1$. We note that the case where the spatial weights are measurable w.r.t. $\mathcal{B}_{n,t} \vee \mathcal{C}$ already covers situations where endogeneity only stems from the spatial weights being dependent on the unit specific effects.

For specific models the construction of a linear quadratic criterion function that identifies the parameters of interest requires a careful analysis. This is illustrated with the following example.

**Example 1 (Social Interactions, Cont.)** *We illustrate the identification of the parameters* $\theta_0 = (\rho_0, f_{0,1})$ *in the social interaction model (setting* $f_{0,2} = 1$ *as a normalization). For identification we can exploit the availability of at least two outcome measures in the form of scores from different tests. We also exploit the fact that scores are typically measured as positive integers. Utilizing Proposition 1 with* $T = 2$ *set* $\pi_{11} = \pi_{11}(f_1, \gamma_\sigma) = v$, $\pi_{12} = \pi_{12}(f_1, \gamma_\sigma) = -f_1 v$, *with* $v = v(f_1, \gamma_\sigma) = 1/\sqrt{f_1^2\sigma_2^2 + \sigma_1^2}$ *and* $\gamma_\sigma = (\sigma_1^2, \sigma_2^2)'$, *and let* $\Sigma_\varrho(\gamma_\varrho) = diag_{c=1}^C \left[\phi_c^2 I_c\right]$ *with* $\phi_c^2 = \varrho_L^2 s_c + (1 - s_c)$ *and* $\gamma_\varrho = \varrho_L^2$. *Consistent with (17) define*

$$u_{*1}^+(\theta, \gamma) = \pi_{11}(f_1, \gamma_\sigma)\Sigma_\varrho(\gamma_\varrho)^{-1/2} y_1^*(\rho) + \pi_{12}(f_1, \gamma_\sigma)\Sigma_\varrho(\gamma_\varrho)^{-1/2} y_2^*(\rho).$$

*with* $y_t^*(\rho) = (I + \rho M)^{-1} y_t$, *and consider the moment vector*

$$m(\theta, \gamma) = n^{-1/2} \begin{bmatrix} u_{*1}^+(\theta_0, \gamma)'\mathbf{1}_n \\ u_{*1}^+(\theta_0, \gamma)' A u_{*1}^+(\theta_0, \gamma) \end{bmatrix} \tag{23}$$

*with* $A = M$. *Let the parameter space for* $\theta$ *and* $\gamma = (\sigma_1^2, \sigma_2^2, \varrho_L^2)$ *be a subset of* $(-1, K_\rho) \times [-K, K]$ *and* $[K^{-1}, K]^3$, *respectively, with* $K_\rho = \min_c n_c/2 - 1 > 0$ *and for some* $0 < K < \infty$ *sufficiently large. Given the conditions on the first and second moments of the* $u_{cjt}$ *and the maintained assumptions it can be shown that* $E[m(\theta, \gamma)] = 0$ *if and only if* $\theta = \theta_0$ *for all admissible* $\gamma$.[13] *We note that* $E[m(\theta_0, \gamma)] = 0$ *follows immediately from Proposition 2. Identification of* $f_{0,1}$ *can be established from an inspection of the linear moment condition. Identification of the social interaction parameter* $\rho_0$ *can be established via an analysis of the quadratic moment condition.*

[13] A proof is given in the auxiliary appendix in Proposition D.1.

*A formal analysis of the identifiability of $\theta_0$ is given in the auxiliary appendix; see Proposition D.1. In fact, this analysis shows that $\theta_0$ can be identified even if we relax our restrictions on the idiosyncratic errors and allow for those errors to be correlated between tests (but not between individuals). We emphasize further that no restrictions on $\mu$ (other than the fact that $\mu$ does not depend on $t$) are needed to identify $\rho_0$. Also, the analysis in the appendix shows that neither knowledge of nor restrictions on heteroskedasticity in $t$ and $i$ are necessary for the identification of $\theta_0$. Thus initial estimates of $\theta_0$ can be obtained with $\Sigma_\sigma = I_T$ and $\Sigma_\varrho = I_n$.*

# 5 Formal Assumptions

In the following we state the set of assumptions which we employ, in addition to Assumption 1, in establishing the consistency and limiting distribution of our GMM estimator. We first postulate a set of assumptions regarding the instruments $h_{it}$ and weights $a_{ijt}$. Let $\xi$ denote some some random variable, then $\|\xi\|_s \equiv (E[|\xi|^s])^{1/s}$ denotes the $s$-norm of $\xi$ for $s \geq 1$.

**Assumption 2** *Let $\delta > 0$, and let $K_h$, $K_a$ and $K_f$ denote finite constants (which are taken, w.o.l.o.g., to be greater then one and do not vary with any of the indices and $n$), then the following conditions hold for $t = 1, \ldots, T$ and $i = 1, \ldots, n$:*
*(i) The elements of the $1 \times p_t$ vector of instruments $h_{it} = [h_{ir,t}]_{r=1,\ldots,p_t}$ are measurable w.r.t. $\mathcal{B}_{n,t} \vee \mathcal{C}$. Furthermore, $\|h_{irt}\|_{2+\delta} \leq K_h < \infty$ for some $\delta > 0$.*
*(ii) The elements of the $1 \times p_t$ vector of weights $a_{ij,t} = [a_{ijr,t}]_{r=1,\ldots,p_t}$ are measurable w.r.t. $\mathcal{B}_{n,t} \vee \mathcal{C}$. Furthermore, $a_{ii,t} = 0$ and $a_{ij,t} = a_{ji,t}$, and $\sum_{j=1}^n |a_{ijr,t}| \leq K_a < \infty$, and $\sum_{j=1}^n \|a_{ijr,t}\|_{2+\delta} \leq K_a < \infty$.*
*(iii) The factors $f_t$, with $f_T = 1$ as a normalization, are measurable w.r.t. $\mathcal{C}$ and satisfy $|f_t| \leq K_f$.*

In the case where the $a_{ijr,t}$ are non-stochastic $\|a_{ijr,t}\|_{2+\delta} = |a_{ijr,t}|$. The next assumption summarizes the assumed convergence behavior of sample moments of $h_{it}$ and $a_{ijt}$. The assumption allows for the observations to be cross sectionally normalized by $\varrho_i$, where $\varrho_i$ may differ from $\varrho_{0,i}$.

**Assumption 3** *Let the elements of $\Sigma_\varrho = \operatorname{diag}_{i=1}^n(\varrho_i^2)$ be measurable w.r.t. $\mathcal{Z}_n \vee \mathcal{C}$ with $0 < c_u^\varrho < \varrho_i^2 < C_u^\varrho < \infty$. The following holds for $t = 1, \ldots, T-1$:*

$$n^{-1} \sum_{i=1}^n E\left[ \left( \frac{\varrho_{0,i}}{\varrho_i} \right)^2 h_{it}' h_{it} \middle| \mathcal{C} \right] \xrightarrow{p} V_{t,\varrho}^h, \quad n^{-1} \sum_{i=1}^n \sum_{j=1}^n E\left[ \left( \frac{\varrho_{0,i}}{\varrho_i} \right)^2 \left( \frac{\varrho_{0,j}}{\varrho_j} \right)^2 a_{ij,t}' a_{ij,t} \middle| \mathcal{C} \right] \xrightarrow{p} V_{t,\varrho}^a,$$

*where the elements of $V^h_{t,\varrho}$ and $V^a_{t,\varrho}$ are finite a.s. and measurable w.r.t. $\mathcal{C}$, and*

$$V^h_{t,n,\varrho} = n^{-1}\sum_{i=1}^{n}\left(\frac{\varrho_{0,i}}{\varrho_i}\right)^2 h'_{it}h_{it} \xrightarrow{p} V^h_{t,\varrho}, \quad V^a_{t,n,\varrho} = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{\varrho_{0,i}}{\varrho_i}\right)^2\left(\frac{\varrho_{0,j}}{\varrho_j}\right)^2 a'_{ij,t}a_{ij,t} \xrightarrow{p} V^a_{t,\varrho}.$$

*The matrix $V_\varrho = \operatorname{diag}_{t=1}^{T-1}(V_{t,\varrho})$ with $V_{t,\varrho} = V^h_{t,\varrho} + 2V^a_{t,\varrho}$ is a.s. positive definite.*

For the case where $\varrho_i = \varrho_{0,i}$ we will use in the following the simplified notation $V^h_t$, $V^q_t$, $V_t$ and $V$ for the matrices defined in the above assumption. The spatial weights matrices, the spatial lag matrices $R_t(\lambda)$ and $\underline{R}_t(\rho)$, and the parameters are assumed to satisfy the following assumption.

**Assumption 4** *(i) The elements of the spatial weights matrices $M_{p,t}$ and $\underline{M}_{q,t}$ are observed. (ii) All diagonal elements of $M_{p,t}$ and $\underline{M}_{q,t}$ are zero. (iii) $\lambda_{n,0} \in \Theta_\lambda$, $\rho_{n,0} \in \Theta_\rho$, $\beta_{n,0} \in \Theta_\beta$, $f_{n,0} \in \Theta_f$ and $\gamma_{n,0} \in \Theta_\gamma$ where $\Theta_\lambda \subseteq \mathbb{R}^P$, $\Theta_\rho \subseteq \mathbb{R}^Q$, $\Theta_\beta \subseteq \mathbb{R}^{k_x+k_z}$, $\Theta_f \subseteq \mathbb{R}^{T-1}$ and $\Theta_\gamma \subseteq \mathbb{R}^{p_\gamma}$ are open and bounded. Furthermore, $\lambda_{n,0} \to \lambda_*$, $\rho_{n,0} \to \rho_*$, $\beta_{n,0} \to \beta_*$, $f_{n,0} \to f_*$, $\gamma_{n,0} \to \gamma_*$ as $n \to \infty$ with $\lambda_* \in \Theta_\lambda$, $\rho_* \in \Theta_\rho$, $\beta_* \in \Theta_\beta$, $f_* \in \Theta_f$, $\gamma_* \in \Theta_\gamma$ and where $f_*$ and $\gamma_*$ are $\mathcal{C}$-measurable. (iii) For some compact sets $\underline{\Theta}_\lambda$, $\underline{\Theta}_\beta$, $\underline{\Theta}_\rho$ and $\underline{\Theta}_f = [-K, K]$ we have $\Theta_\lambda \subseteq \underline{\Theta}_\lambda$, $\Theta_\beta \subseteq \underline{\Theta}_\beta$, $\Theta_\rho \subseteq \underline{\Theta}_\rho$ and $\Theta_f \subseteq \underline{\Theta}_f$. (iv) The matrices $R_t(\lambda)$ and $\underline{R}_t(\rho)$ are defined for $\lambda \in \underline{\Theta}_\lambda$, $\rho \in \underline{\Theta}_\rho$ and nonsingular for $\lambda \in \Theta_\lambda$, $\rho \in \Theta_\rho$.*

The GMM estimator is optimized over the set $\underline{\Theta}_\theta = \underline{\Theta}_\lambda \times \underline{\Theta}_\beta \times \underline{\Theta}_\rho \times \underline{\Theta}_f$. We observe, as will be discussed in more detail below, that under the above assumptions the sample moment vector $\overline{m}_n(\theta, \gamma)$ given in (21), and thus the objective function of the GMM estimator, are well defined for all $\theta \in \underline{\Theta}_\theta$.

The next assumption postulates a basic smoothness condition for the cross sectional variance components and states basic assumptions regarding the convergence behavior of the sample moments. (The first part of the assumption also ensures that the measurability conditions and boundedness conditions of Assumption 3 are maintained over the entire parameter space.)

**Assumption 5** [14]*(i) The cross sectional variances components $\varrho_i^2(\gamma_\varrho)$ are differentiable and satisfy the measurability conditions and boundedness conditions of Assumption 3 for $\gamma_\varrho \in \Theta_{\gamma_\varrho}$.*
*(ii) For $t \le \tau \le s$ let $C_s$ be a $n \times n$ matrix of the form $\Upsilon$, $\Upsilon\underline{R}_s(\rho)$, $\Upsilon(\partial\underline{R}_s(\rho)/\partial\rho_q$, $\Upsilon A^r_t\Upsilon$, $\underline{R}_\tau(\rho)'\Upsilon A^r_t\Upsilon\underline{R}_s(\rho)$, or $\underline{R}_\tau(\rho)'\Upsilon A^r_t\Upsilon(\partial\underline{R}_s(\rho)/\partial\rho_q)$, where $\Upsilon$ is an $n \times n$ positive diagonal*

[14]We implicitly assume that all derivatives are well defined on an open set containing the optimization space.

*matrix with elements which are uniformly bounded and measurable w.r.t.* $\mathcal{Z}_n \vee \mathcal{C}$*. Then for* $\rho \in \underline{\Theta}_\rho$ *the probability limits* ($n \to \infty$) *of*

$$\begin{array}{lll} n^{-1}h'_{r,t}C_s y_s, & n^{-1}h'_{r,t}C_s W_s, & n^{-1}y'_\tau C_s W_s, \\ n^{-1}W'_\tau C_s y_s, & n^{-1}y'_\tau C_s y_s, & n^{-1}W'_\tau C_s W_s, \end{array} \tag{24}$$

*exist for* $r = 1, \ldots, p_t$*, and the probability limits are measurable w.r.t.* $\mathcal{C}$*, continuous in* $\rho$*, and bounded in absolute value, where the bound does not depend on* $\rho$*.*

We note that typically those probability limits will coincide with the probability limits of the corresponding expectations w.r.t. to $\mathcal{C}$, e.g.,

$$\operatorname*{plim}_{n\to\infty} n^{-1}h'_{r,t}C_s y_s = \operatorname*{plim}_{n\to\infty} E\left[n^{-1}h'_{r,t}C_s y_s|\mathcal{C}\right].$$

For autoregressive disturbance processes $\underline{R}_\tau(\rho) = I - \sum_{q=1}^{Q}\rho_q \underline{M}_{q,t}$. In this case we have $\underline{R}_\tau(\rho)'\Upsilon A_t^r \Upsilon \underline{R}_s(\rho) = \sum_{q=0}^{Q}\sum_{p=0}^{Q}\rho_q\rho_p \underline{M}_{q,\tau}\Upsilon A_t^r \Upsilon \underline{M}_{p,s}$ with $\rho_0 = -1$ and $\underline{M}_{0,\tau} = \underline{M}_{0,s} = I$, and furthermore $\partial \underline{R}_s(\rho)/\partial\rho_q = \underline{M}_{q,s}$. Observing that the parameter space $\underline{\Theta}_\rho$ is compact it is readily seen that for Assumption 5 to hold it suffices to consider matrices $C_s$ of the form $\Upsilon$, $\Upsilon A_t^r \Upsilon$, $\Upsilon A_t^r \Upsilon \underline{M}_{p,s}$, or $\underline{M}'_{q,\tau}\Upsilon A_t^r \Upsilon \underline{M}_{p,s}$, and we can suppress $\rho$ from the assumption.

The following assumption guarantees that the moment conditions identify the parameter $\theta_0$. To cover initial estimators for $\theta_0$ our setup allows both for situations where the estimator for $\theta_0$ is based on a consistent or an inconsistent estimator of the auxiliary parameters $\gamma_0$. In the following let $\bar{\gamma}_n \xrightarrow{p} \bar{\gamma}_*$ with $\bar{\gamma}_n \in \Theta_\gamma$ and $\bar{\gamma}_* \in \Theta_\gamma$ denote a particular estimator and its limit. For consistent estimators of the auxiliary parameters $\bar{\gamma}_* = \gamma_*$, and for inconsistent estimators $\bar{\gamma}_* \neq \gamma_*$. The latter covers the case where in the computation of the first stage estimator for $\theta_0$ all auxiliary parameters are set equal to some fixed values, i.e., the case where $\bar{\gamma}_n = \gamma_* = \bar{\gamma}$.

**Assumption 6** *Let* $\delta_*, \rho_*, f_*, \gamma_*$ *be as defined in Assumption 4, let* $\theta_* = (\delta'_*, \rho'_*, f'_*)'$*, and let* $\bar{\gamma}_n \xrightarrow{p} \bar{\gamma}_*$ *with* $\bar{\gamma}_n \in \Theta_\gamma$ *and* $\bar{\gamma}_* \in \Theta_\gamma$*, where* $\bar{\gamma}_*$ *is* $\mathcal{C}$*-measurable. Furthermore, for* $\theta \in \underline{\Theta}_\theta$ *let* $\mathfrak{m}(\theta) = \operatorname{plim}_{n\to\infty} n^{-1/2}\overline{m}_n(\theta, \bar{\gamma}_*)$ *and* $G(\theta) = \operatorname{plim}_{n\to\infty} \partial n^{-1/2}\overline{m}_n(\theta, \gamma_*)/\partial\theta$.[15] *Then the following is assumed to hold:*

*(i)* $\theta_*$ *is identifiable unique in the sense that* $\mathfrak{m}(\theta_*) = 0$ *a.s. and for every* $\varepsilon > 0$*,*

$$\inf_{\{\theta\in\underline{\Theta}_\theta : |\theta-\theta_*|>\varepsilon\}} \|\mathfrak{m}(\theta)\| > 0 \ \text{a.s.} \tag{25}$$

[15] Lemma A.5 establishes the existence of the limit of the moment vector $\mathfrak{m}(\theta)$ and the limit of the derivatives of the moment vector $G(\theta)$. To keep our notation simple, we have suppressed the dependence of $\mathfrak{m}(\theta)$ on $\bar{\gamma}_*$. The limiting matrix $G(\theta)$ is only considered at $\bar{\gamma}_* = \gamma_*$.

*(ii)* $\sup_{\theta\in\underline{\Theta}_\theta}\left\|n^{-1/2}\overline{m}_n(\theta,\bar{\gamma}_n)-\mathfrak{m}(\theta)\right\|=o_p(1)$ *for* $\bar{\gamma}_n\xrightarrow{p}\bar{\gamma}_*$.
*(iii)* $\sup_{\theta\in\underline{\Theta}_\theta}\left\|\partial n^{-1/2}\overline{m}_n(\theta,\bar{\gamma}_n)/\partial\theta-G(\theta)\right\|=o_p(1)$ *for* $\bar{\gamma}_n\xrightarrow{p}\gamma_*$, *and*

$$\operatorname*{plim}_{n\to\infty}\partial n^{-1/2}\overline{m}_n(\bar{\theta}_n,\bar{\gamma}_n)/\partial\gamma=0$$

*for* $\bar{\theta}_n\xrightarrow{p}\theta_*$ *and* $\bar{\gamma}_n\xrightarrow{p}\gamma_*$.

We furthermore maintain the following assumptions regarding the moment weighting matrix of our GMM estimator.

**Assumption 7** *Suppose* $\tilde{\Xi}_n\xrightarrow{p}\Xi$ *, where* $\Xi$ *is* $\mathcal{C}$*-measurable with a.s. finite elements, and* $\Xi$ *is positive definite a.s.*

For autoregressive disturbance processes where $\underline{R}_t(\rho)=I-\sum_{q=1}^{Q}\rho_q\underline{M}_{q,t}$ our specification allows for the true autoregressive parameters to be arbitrarily close to a singular point.[16] Technically we distinguish between the parameter space and the optimization space, which defines the estimator. Since our specification of the moment vector does not rely on $R_t(\lambda)^{-1}$ or $\underline{R}_t(\rho)^{-1}$ it remains well defined even for parameter values where $R_t(\lambda)$ and $\underline{R}_t(\rho)$ are singular. Thus for autoregressive processes we can specify the optimization space to be a compact set $\underline{\Theta}_\theta=\underline{\Theta}_\lambda\times\underline{\Theta}_\beta\times\underline{\Theta}_\rho\times\underline{\Theta}_f$ containing the parameter space, without restricting the class of admissible models. We note that given that $f_T=1$ the weights $\pi_{ts}=\pi_{ts}(f,\gamma_\sigma)$ of the Generalized Helmert transformation defined in Proposition 1 are well defined on $\underline{\Theta}_f\times\underline{\Theta}_\gamma$. At the same time we note that when $\underline{R}_t(\rho)$ parametrizes a spatial moving average process or is a of a more general form, it may not be well defined on the boundaries of an open parameter space $\Theta_\rho$. In such cases it is necessary to shrink the set $\Theta_\rho$ sufficiently such that $\underline{R}_t(\rho)$ exists on $\underline{\Theta}_\rho$.[17]

# 6 Asymptotic Properties of the GMM Estimator

## 6.1 Consistency

Consider a sequence of estimators of the auxiliary parameters $\bar{\gamma}_n\xrightarrow{p}\bar{\gamma}_*$. The objective function of the GMM estimator $\tilde{\theta}_n(\bar{\gamma}_n)$ defined in (22) is then given by $\mathcal{R}_n(\theta)=n^{-1}\overline{m}_n(\theta,\bar{\gamma}_n)'\tilde{\Xi}_n\overline{m}_n(\theta,\bar{\gamma}_n)$. Correspondingly consider the "limiting" objective function

[16] This is in contrast to some of the recent panel data literature; see, e.g., Lee and Yu (2014).

[17] An example is $\underline{R}(\rho)=(I_n-\rho 1_n 1_n/n)^{-1}$ and $\Theta_\rho=(-1,1)$. Then, $\underline{R}(\rho)$ is not defined for $\rho=1$. In this case, we choose $\varepsilon>0$, $\Theta_\rho=(-1,1-\varepsilon)$ and $\underline{\Theta}_\rho=[-1,1-\varepsilon]$.

$\mathcal{R}(\theta) = \mathfrak{m}(\theta)'\Xi\mathfrak{m}(\theta)$ with $\mathfrak{m}(\theta) = \text{plim}_{n\to\infty} n^{-1/2}\overline{m}_n(\theta, \bar{\gamma}_*)$. Because $\mathfrak{m}(\theta)$ and $\Xi$ are generally stochastic in the presence of common factors it follows that $\mathcal{R}(\theta)$ and the minimizer $\theta_*$ are also generally stochastic. Consistency proofs need to account for the randomness in $\mathcal{R}(\theta)$ and $\theta_*$. The results given in this section build, in particular, on Gallant and White (1988), White (1984), Newey and McFadden (1994), Pötscher and Prucha (1997, ch 3).[18]

The following proposition holds for general criterion functions $\mathcal{R}_n : \Omega \times \underline{\Theta}_\theta \to \mathbb{R}$ and $\mathcal{R} : \Omega \times \underline{\Theta}_\theta \to \mathbb{R}$, the finite sample objective function and the corresponding "limiting" objective function, respectively. They include, but are not limited to the particular specification of $\mathcal{R}_n$ and $\mathcal{R}$ for our GMM estimator given above. The notation emphasizes that $\mathcal{R}$ is a random function. Furthermore $\widehat{\theta}_n = \widehat{\theta}_n(\omega)$ and $\theta_* = \theta_*(\omega)$ are the "minimizers" of $\mathcal{R}_n(\omega, \theta)$ and $\mathcal{R}(\omega, \theta)$, where both $\widehat{\theta}_n$ and $\theta_*$ are implicitly assumed to be well defined random variables. For the following we also adopt the convention that the variables in any sequence, that is claimed to converge in probability, are measurable. We now have the following general module for proving consistency.

**Proposition 3** *(i) Suppose that the minimizer $\theta_* = \theta_*(\omega)$ of $\mathcal{R}(\omega, \theta)$ is identifiably unique in the sense that for every $\epsilon > 0$, $\inf_{\{\theta\in\underline{\Theta}_\theta:|\theta-\theta_*|\geq\varepsilon\}} \mathcal{R}(\omega, \theta) - \mathcal{R}(\omega, \theta_*(\omega)) > 0$ a.s. (ii) Suppose furthermore that $\sup_{\theta\in\underline{\Theta}_\theta} |\mathcal{R}_n(\omega, \theta) - \mathcal{R}(\omega, \theta)| \to 0$ a.s. [i.p.] as $n \to \infty$. Then for any sequence $\widehat{\theta}_n$ such that eventually $\mathcal{R}_n(\omega, \widehat{\theta}_n(\omega)) = \inf_{\theta\in\underline{\Theta}_\theta} \mathcal{R}_n(\omega, \theta)$ holds, we have $\widehat{\theta}_n \to \theta_*$ a.s. [i.p.] as $n \to \infty$.*

We note that for the above proposition compactness of $\underline{\Theta}_\theta$ is not needed. The definition of identifiable uniqueness adopted in the above proposition extends the notion of identifiable uniqueness to stochastic limiting functions and stochastic minimizers. In case the limiting objective function is non-stochastic it reduces to the usual definition of identification.

Assumptions 6(i) and 7 are crucial in establishing that $\theta_*$ is identifiable unique in the sense of Proposition 3. We therefore have the following consistency result.

**Theorem 1** *(Consistency) Suppose Assumptions 1-7 hold for some estimator of the auxiliary parameters $\bar{\gamma}_n \xrightarrow{p} \bar{\gamma}_*$. Then $\tilde{\theta}_n(\bar{\gamma}_n) - \theta_{n,0} \xrightarrow{p} 0$ as $n \to \infty$.*

**Remark 1** *Assumptions 6(iii) is not required by the above theorem. We note that the theorem covers the case where $\bar{\gamma}_n = \tilde{\gamma}_n$ and $\tilde{\gamma}_n$ is a consistent estimator of the auxiliary parameters, as well as the case where $\bar{\gamma}_n = \bar{\gamma}_* = \bar{\gamma}$ for all $n$. The latter case is relevant*

[18] The latter reference also provides citations to the earlier fundamental contributions to the consistency proof of M-estimators in the statistics literature. We would like to thank Benedikt Pötscher for very helpful discussions on extending the notion of identifiable uniqueness to stochastic analogue functions, and the propositions presented in this section.

*for first stage estimators that are based on arbitrarily fixed variance parameters. For $\gamma_\sigma$ an obvious choice is $\bar{\gamma}_\sigma = \mathbf{1}_T$. For $\gamma_\varrho$ convenient choices depend on the specifics of the model. In many situations the first stage estimator will be based on the choice $\varrho_i^2(\bar{\gamma}_\varrho) = 1$.*

## 6.2 Central Limit Theorem

In the following we establish the limiting distribution of the sample moment vector $\overline{m}_n = \overline{m}_n\left(\theta_0, \gamma_{0,\sigma}, \gamma_\varrho\right)$ defined by (21), evaluated at the true parameters, except possible for the specification of the cross sectional variance components $\varrho_i^2$. In particular, the results allow for the leading cases $\varrho_i^2 = \varrho_{0,i}^2$ and $\varrho_i^2 = 1$. Observe from (20) that the subvectors of $\overline{m}_n$ are given by

$$\begin{aligned} &\overline{m}_t(\theta_0, \gamma_{0,\sigma}, \gamma_\varrho) = n^{-1/2} \textstyle\sum_{i=1}^n h_{it}' u_{*it}^+ + n^{-1/2} \sum_{i=1}^n \sum_{j=1}^n a_{ij,t}' u_{*it}^+ u_{*jt}^+, \\ &u_{*it}^+ = u_{*it}^+(\theta_0, \gamma_{0,\sigma}, \gamma_\varrho) = \textstyle\sum_{s=t}^T \pi_{ts}\left(f_0, \gamma_{0,\sigma}\right) u_{is}/\varrho_i, \end{aligned} \tag{26}$$

We derive the limiting distribution of $\overline{m}_n$ by showing that it can be represented as a sum of martingale differences to which the CLT of Kuersteiner and Prucha (2013, Theorem 1) can be applied. In fact, we not only establish convergence in distribution of $\overline{m}_n$, but we establish $\mathcal{C}$-stable convergence in distribution of $\overline{m}_n$.[19] This allows us to accommodate random norming which is needed because our criterion function may converge to mixed normal limiting distributions. The result allows us to establish the joint limiting distribution for $(\overline{m}_n, A)$ for any matrix valued random variable $A$ that is $\mathcal{C}$ measurable. Establishing joint limits is a requirement for the continuous mapping theorem to apply and thus critical for the asymptotic analysis of estimators and test statistics.

The CLT for the sample moment vector $\overline{m}_n$ given below establishes $V_\varrho$, defined in Assumption 3, as the limiting variance covariance matrix. The form of $V_\varrho$ is consistent with the results on the variance covariances of linear quadratic forms given in Proposition 2, after specializing those results to the case of orthogonally transformed disturbances, and symmetric weight matrices with zero diagonal elements. We emphasize that due to (i) employing an orthogonal transformation of the disturbances to eliminate the unit specific effects and (ii) considering matrices with zero diagonal elements in forming the quadratic moment conditions, all correlations across time are zero. An inspection of Proposition 2 also shows that the expressions for the variances and covariances are much more complex for non-orthogonal transformations, and that the use of matrices with non-zero diagonal

[19] See Renyi (1963), Aldous and Eagleson (1978), Hall and Heyde (1980) and Kuersteiner and Prucha (2013) for a definitions and discussion of $\mathcal{C}$-stable convergence.

elements in forming the quadratic moment conditions can introduce components which may be difficult to estimate.

**Theorem 2** *Let the transformation matrix* $\Pi = \Pi(f_0, \gamma_{0,\sigma})$ *be as defined in Proposition 1, and suppose Assumptions 1-3 hold with* $\varrho_i^2 = \varrho_i^2(\gamma_\varrho)$ *and* $V_\varrho = \operatorname{diag}_{t=1}^{T-1}(V_{t,\varrho})$ *and* $V_{t,\varrho} = V_{t,\varrho}^h + 2V_{t,\varrho}^a$.
*(i) Then*

$$\overline{m}_n\left(\theta_0, \gamma_{0,\sigma}, \gamma_\varrho\right) \xrightarrow{d} V_\varrho^{1/2}\xi \quad (\mathcal{C}\text{-stably}), \tag{27}$$

*where* $\xi \sim N(0, I_p)$*, and* $\xi$ *and* $\mathcal{C}$ *(and thus* $\xi$ *and* $V_\varrho$*) are independent.*
*(ii) Let* $A$ *be some* $p_* \times p$ *matrix that is* $\mathcal{C}$ *measurable with finite elements and rank* $p_*$ *a.s., then*

$$A\overline{m}_n \xrightarrow{d} (AV_\varrho A')^{1/2}\xi_*, \tag{28}$$

*where* $\xi_* \sim N(0, I_{p_*})$*, and* $\xi_*$ *and* $\mathcal{C}$ *(and thus* $\xi_*$ *and* $AV_\varrho A'$*) are independent.*

Theorem 2 can be of interest in itself as a CLT for vectors of linear quadratic forms of transformed innovations. As a special case the theorem also covers linear quadratic forms in the original innovations: for $f_T = \sigma_T = 1$, $f_t = 0$ for $t < T$ and $\varrho_i^2 = \varrho_{0,i}^2$ we have $u_{*it}^+ = u_{it}/(\sigma_{0,t}\varrho_{0,i})$. The result generalizes Theorem 2 in Kuersteiner and Prucha (2013). We emphasize that our result differs from existing results on CLTs for quadratic forms in various respects:[20] First it considers linear quadratic forms in a panel framework. To the best of our knowledge, other results only consider single indexed variables. As stressed in Kuersteiner and Prucha (2013) the widely used CLT for martingale differences by Hall and Heyde (1980) is not generally compatible with a panel data situation. Second, Theorem 2 allows for the presence of common factors which can be handled, because Theorem 2 establishes convergence in distribution $\mathcal{C}$-stably. Third, the theorem covers orthogonally transformed variables, and demonstrates how these transformations very significantly simplify the correlation structure between the linear quadratic forms.

## 6.3 Limiting Distribution

The next theorem establishes basic properties for the limiting distribution of the GMM estimator $\tilde{\theta}_n(\tilde{\gamma}_n)$ when $\tilde{\gamma}_n$ is a consistent estimator of the auxiliary parameters so that $\tilde{\gamma}_n - \gamma_{n,0} \xrightarrow{p} 0$ and $\gamma_{n,0} \xrightarrow{p} \gamma_*$. Let $G_n(\theta, \gamma) = \partial n^{-1/2}\overline{m}_n(\theta, \gamma)/\partial\theta$ and $G(\theta) = \operatorname{plim}_{n\to\infty} G_n(\theta, \gamma_*)$

[20] See, e.g., Atchad and Cattaneo (2012), Doukhan et al. (1996), Gao and Yongmiao (2007), Giraitis and Taqqu (1998), and Kelejian and Prucha (2001) for recent contributions. To the best of our knowledge the result is also not covered in the literature on $U$-statistics; see, e.g., Korolyuk and Borovskich (1994) for a review.

as defined in Assumption 6. In deriving the limiting distribution of $\tilde{\theta}_n(\tilde{\gamma}_n)$ we establish that $G(\theta)$ exists, and that $G(\theta)$ is $\mathcal{C}$-measurable for all $\theta \in \underline{\Theta}_\theta$, and continuous in $\theta$. Let $G = G(\theta_*)$ and observe that $G$ is $\mathcal{C}$-measurable, since $\theta_*$ is $\mathcal{C}$-measurable in light of Assumption 4.

**Theorem 3** *(Asymptotic Distribution). Suppose Assumptions 1-7 holds for $\bar{\gamma} = \tilde{\gamma}_n$ with $\tilde{\gamma}_n - \gamma_{n,0} = O_p(n^{-1/2})$ and $\varrho_i^2 = \varrho_{0,i}^2 = \varrho_i^2(\gamma_{0,\varrho})$, and that $G$ has full column rank a.s. Then, (i)*

$$n^{1/2}(\tilde{\theta}_n\left(\tilde{\gamma}_n\right) - \theta_{n,0}) \xrightarrow{d} \Psi^{1/2}\xi_*, \quad \text{as } n \to \infty,$$

*where $\xi_*$ is independent of $\mathcal{C}$ (and hence of $\Psi$), $\xi_* \sim N(0, I_{p_\theta})$ and*

$$\Psi = (G'\Xi G)^{-1}G'\Xi V \Xi G(G'\Xi G)^{-1}. \tag{29}$$

*(ii) Suppose $B$ is some $q \times p_\theta$ matrix that is $\mathcal{C}$ measurable with finite elements and rank $q$ a.s., then*

$$Bn^{1/2}(\tilde{\theta}_n\left(\tilde{\gamma}_n\right) - \theta_{n,0}) \xrightarrow{d} (B\Psi B')^{1/2}\xi_{**},$$

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

The matrix $V$ is defined in Assumption 3. Since $\varrho_i^2 = \varrho_{0,i}^2$ the expression simplifies to $V = \operatorname{diag}_{t=1}^{T-1}\left(V_t^h + 2V_t^a\right)$ with $V_t = V_t^h + 2V_t^a$, where $n^{-1}\sum_{i=1}^n E\left[h_{it}'h_{it}\middle|\mathcal{C}\right] \xrightarrow{p} V_t^h$ and $n^{-1}\sum_{i=1}^n\sum_{j=1}^n E\left[a_{ij,t}'a_{ij,t}\middle|\mathcal{C}\right] \xrightarrow{p} V_t^a$. In light of Assumption 3 a consistent estimator of $V$ is

$$\widetilde{V}_n = \operatorname{diag}_{t=1}^{T-1}\left(V_{t,n}^h + 2V_{t,n}^a\right), \tag{30}$$

where $V_{t,n}^h = n^{-1}\sum_{i=1}^n h_{it}'h_{it}$, $V_{t,n}^a = n^{-1}\sum_{i=1}^n\sum_{j=1}^n a_{ij,t}'a_{ij,t}$.

For efficiency, conditional on $\mathcal{C}$, we select $\Xi = V^{-1}$, in which case $\Psi = \left[G'V^{-1}G\right]^{-1}$. The corresponding feasible efficient GMM estimator is then obtained by choosing $\tilde{\Xi}_n = \widetilde{V}_n^{-1}$yielding

$$\hat{\theta}_n = \arg\min_{\theta\in\underline{\Theta}_\theta} \overline{m}_n(\theta, \tilde{\gamma}_n)'\widetilde{V}_n^{-1}\overline{m}_n(\theta, \tilde{\gamma}_n). \tag{31}$$

Clearly $\widetilde{V}_{(n)}^{-1} \xrightarrow{p} V^{-1}$ in light of Assumption 3, with $V^{-1}$ being $\mathcal{C}$-measurable with $a.s.$ finite elements, and with $V^{-1}$ positive definite $a.s.$ Furthermore, from the proof of Theorem 3, $G_n(\hat{\theta}_n, \tilde{\gamma}_n) \xrightarrow{p} G$ where $G$ is $\mathcal{C}$-measurable with $a.s.$ finite elements, and with full column rank $a.s.$, we have that $\hat{\Psi}_n = \left[G_n'(\hat{\theta}_n, \tilde{\gamma}_n)\widetilde{V}_n^{-1}G_n(\hat{\theta}_n, \tilde{\gamma}_n)\right]^{-1}$ is a consistent estimator for $\Psi$.

Let $R$ be a $q \times p_\theta$ full row rank matrix and $r$ a $q \times 1$ vector, and consider the Wald statistic

$$T_n = \left\|\left(R\hat{\Psi}_nR'\right)^{-1/2}\sqrt{n}(R\hat{\theta}_n - r)\right\|^2 \tag{32}$$

to test the null hypothesis $H_0 : R\theta_{n,0} = r$ against the alternative $H_1 : R\theta_{n,0} \neq r$. The next theorem shows that $T_n$ is distributed asymptotically chi-square, even if $\Psi$ is allowed to be random due to the presence of common factors represented by $\mathcal{C}$.

**Theorem 4** *Suppose the assumptions of Theorem 3 hold. Then*

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

*Furthermore*

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

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

As remarked above, an initial consistent GMM estimator $\bar{\theta}_n$ can be obtained by choosing $\tilde{\Xi}_n = I$ and $\bar{\gamma} = 1$, or equivalently by using the identity matrices as estimators for $\Sigma_\sigma$ and $\Sigma_\varrho$.

# 7 Conclusion

In this paper we develop an estimation theory for a panel data model with interactive effects that permits for the data generating process to be (time) dynamic, and cross sectionally dependent. Cross sectional dependence may stem from "Cliff-Ord type spatial" interactions as well as from common shocks. The model allows for spatial interactions in the endogenous variables, exogenous variables, and/or disturbances. The spatial interaction matrices are themselves allowed to be endogenously determined. The model also allows for heteroskedasticity in the time and cross sectional dimensions. In addition, our theory also accommodates regressors that are only sequentially (rather than strictly) exogenous.

The paper considers a class of GMM estimators based on linear and quadratic moment conditions and forward differenced data. It provides results on the consistency and the limiting distribution of the estimators. The paper first develops a quasi-forward differencing transformation, which eliminates the interactive effects while ensuring orthogonality of the transformed moments. This transformation contains the Helmert transformation as a special case. The paper also gives general results regarding the variances and covariances of linear quadratic forms of forward differences. Due to the presence of common factors the limiting distribution of the GMM estimator is nonstandard as a multivariate mixture normal, which leads to the need for random norming. Despite of this it is shown that corresponding Wald test statistics have the usual $\chi^2$-distribution.

The estimation theory developed here is expected to be useful for analyzing a wide range of data in micro economics, including social interactions, as well as in macro economics.

Our theory is general in nature. Future work will examine specific models and estimators in more detail. The exact specification of instruments and the estimation of nuisance parameters are best handled on a case by case basis.

# A Appendix: Proofs

## A.1 Martingale Difference Representation

To establish a martingale difference representation of the sample moment vector $\overline{m}_n = \overline{m}_n(\theta_0, \gamma_{0,\sigma}, \gamma_\varrho)$ defined by (21) and (26) we employ the following sub-$\sigma$-fields of $\mathcal{F}$ ($i = 1, \ldots, n$):

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

with $\mathcal{F}_{n,0} = \mathcal{C}$. Let $\lambda = (\lambda_1', \ldots, \lambda_{T-1}')' \in \mathbb{R}^p$ be a fixed vector with $\lambda'\lambda = 1$. Using the Cramer-Wold device and utilizing (26) consider $\lambda'\overline{m}_n = S_1 + S_2$ with $S_1 = n^{-1/2}\sum_{i=1}^n \sum_{t=1}^{T-1} \lambda_t' h_{it}' u^+_{*it}$ and $S_2 = n^{-1/2}\sum_{i=1}^n \sum_{t=1}^{T-1} \lambda_t' \sum_{j=1}^n a'_{ij,t} u^+_{*it} u^+_{*jt}$ where $u^+_{*it} = u^+_{it}/\varrho_i = (\varrho_{0,i}/\varrho_i)\left[u^+_{it}/\varrho_{0,i}\right]$ with $u^+_{it}/\varrho_{0,i} = u^+_{it}(\theta_0, \gamma_{0,\sigma})/\varrho_{0,i} = \sum_{s=t}^T \pi_{ts}\left(f_0, \gamma_{0,\sigma}\right)\left[u_{is}/\varrho_{0,i}\right]$. Since $\varrho_{0,i}$ and $\varrho_i$ satisfies the same measurability properties as $h_{it}$ and $a_{ij,t}$, and since $0 < c_u^\varrho < \varrho_{0,i}^2, \varrho_i^2 < C_u^\varrho < \infty$, we can w.o.l.o.g. set $\varrho_{0,i} = \varrho_i = 1$ and implicitly absorb these terms into $h_{it}$ and $a_{ij,t}$. Then

$$S_1 = n^{-1/2}\sum_{i=1}^n \sum_{t=1}^{T-1} \lambda_t' h_{it}' \sum_{u=t}^T \pi_{tu} u_{iu} = \sum_{t=1}^T \sum_{i=1}^n c_{it} u_{it}, \tag{A.2}$$

with

$$c_{it} = \sum_{s=1}^t \lambda_s' h_{is}' \pi_{st} \tag{A.3}$$

and where we set $\lambda_T = 0$. Note that $c_{it}$ only depends on $h_{is}$ with $s \le t$ and $\pi_{st}$, and thus is measurable w.r.t. $\mathcal{B}_{n,t} \vee \mathcal{C}$. This implies that $c_{it}$ is measurable w.r.t. $\mathcal{F}_{n,(t-1)n+i}$ and $\mathcal{B}_{n,i,t} \vee \mathcal{C}$. Next, observe that

$$S_2 = \sum_{t=1}^T \sum_{i=1}^n 2\left(\sum_{j=1}^{i-1} u_{it} u_{jt} c_{ij,tt} + \sum_{s=1}^{t-1}\sum_{j=1}^n u_{it} u_{js} c_{ij,ts}\right) \tag{A.4}$$

with

$$c_{ij,ts} = \sum_{\tau=1}^s \lambda_\tau' a'_{ij,\tau} \pi_{\tau s} \pi_{\tau t} \tag{A.5}$$

for $s \le t$. Observe that $c_{ij,ts} = c_{ji,ts}$ and $c_{ij,10} = 0$ per our convention on summation, and that $c_{ij,ts}$ only depends on $a_{ij,\tau}$ for $\tau \le s \le t$. Thus $c_{ij,ts}$ is measurable w.r.t. $\mathcal{B}_{n,s} \vee \mathcal{C}$. This implies that $c_{ij,ts}$ is measurable w.r.t. $\mathcal{F}_{n,(s-1)n+i}$ and $\mathcal{B}_{n,i,s} \vee \mathcal{C}$. By Equations (A.2) and (A.4) it follows that $\lambda'\overline{m}_n = \sum_{v=1}^{Tn+1} X_{n,v}$ with $X_{n,1} = 0$ and, for $t = 1, \ldots, T, i = 1, \ldots, n$,

$$X_{n,(t-1)n+i+1} = n^{-1/2} u_{it}\left(c_{it} + 2\left(\sum_{j=1}^{i-1} c_{ij,tt} u_{jt} + \sum_{j=1}^n \sum_{s=1}^{t-1} c_{ij,ts} u_{js}\right)\right) \tag{A.6}$$

where $\lambda_T = 0$. Given the judicious construction of the random variables $X_{n,v}$ and the information sets $\mathcal{F}_{n,v}$ with $v = (t-1)n+i+1$ we see that $\mathcal{F}_{n,v-1} \subseteq \mathcal{F}_{n,v}$, $X_{n,v}$ is $\mathcal{F}_{n,v}$-measurable, and that $E\left[X_{n,v}|\mathcal{F}_{n,v-1}\right] = E\left[X_{n,(t-1)n+i+1}|\mathcal{F}_{n,(t-1)n+i}\right] = 0$ in light of Assumption 1 and observing that $\mathcal{F}_{n,(t-1)n+i} \subseteq \mathcal{B}_{n,i,t} \vee \mathcal{C}$. This establishes that $\{X_{n,v}, \mathcal{F}_{n,v}, 1 \leq v \leq Tn+1, n \geq 1\}$ is a martingale difference array.[21]

## A.2 Lemmas

**Lemma A.1** *Suppose Assumptions 1 - 3 hold with $\varrho_{0,i}^2 = \varrho_i^2 = 1$, and let $c_{it}$ and $c_{ij,ts}$ be as defined in (A.3) and (A.5) with $\pi_{ts} = \pi_{ts}\left(f_0, \gamma_{0,\sigma}\right)$. Then the following bounds hold for some constant $K$ with $1 < K < \infty$*
*(i)* $E\left[\left|c_{it}\right|^{2+\delta}\right] \leq K$,
*(ii)* $\sum_{i=1}^n \left|c_{ij,ts}\right| \leq K$,
*(iii) for* $q \geq 1$, $\sum_{i=1}^n \left|c_{ij,ts}\right|^q \leq K$,
*(iv) for* $1 \leq q \leq 2+\delta$, $\sum_{j=1}^n \left\|c_{ij,ts}\right\|_q \leq K$,
*(v) for* $1 \leq q \leq 2+\delta$, $E\left[\left|u_{it}\right|^q|\mathcal{F}_{n,(t-1)n+i}\right] \leq K$,
*(vi) for* $s \leq t$, $1 \leq q \leq 2+\delta$, $E\left[\sum_{i=1}^n \left|u_{is}\right|^q \left|c_{ij,ts}\right| |\mathcal{B}_{n,s} \vee \mathcal{C}\right] \leq K$,
*(vii) for* $s \leq t$, $1 \leq q \leq 2+\delta$, $E\left[\left(\sum_{i=1}^n \left|u_{is}\right| \left|c_{ij,ts}\right|\right)^q |\mathcal{B}_{n,s} \vee \mathcal{C}\right] \leq K$.

**Proof.** See supplementary appendix "Proofs of Lemmas". ■

**Lemma A.2** *Suppose Assumptions 1 - 3 hold with $\varrho_{0,i}^2 = \varrho_i^2 = 1$, and let $c_{it}$ and $c_{ij,ts}$ be as defined in (A.3) and (A.5) with $\pi_{ts} = \pi_{ts}\left(f_0, \gamma_{0,\sigma}\right)$. Let $\varsigma_{it}^{(1)} = c_{it}^2$, $\varsigma_{it}^{(2)} = 4\left(\sum_{j=1}^{i-1} c_{ij,tt}u_{jt}\right)^2$, $\varsigma_{it}^{(3)} = 4\left(\sum_{s=1}^{t-1}\sum_{j=1}^n c_{ij,ts}u_{js}\right)^2$, $\varsigma_{it}^{(4)} = 4c_{it}\sum_{j=1}^{i-1} c_{ij,tt}u_{jt}$, $\varsigma_{it}^{(5)} = 4c_{it}\sum_{s=1}^{t-1}\sum_{j=1}^n c_{ij,ts}u_{js}$ and $\varsigma_{it}^{(6)} = 8\sum_{j=1}^{i-1} c_{ij,tt}u_{jt}\sum_{s=1}^{t-1}\sum_{l=1}^n c_{il,ts}u_{ls}$.*
*Define the limits*

$$\varsigma_t^{(1)} = \operatorname*{plim}_{n\to\infty} n^{-1}\sum_{i=1}^n E\left[c_{it}^2|\mathcal{C}\right], \ \varsigma_t^{(2)} = \operatorname*{plim}_{n\to\infty} 2\sigma_{0,t}^2 n^{-1}\sum_{i=1}^n\sum_{j=1}^n E\left[c_{ij,tt}^2|\mathcal{C}\right],$$

$$\varsigma_t^{(3)} = \operatorname*{plim}_{n\to\infty}\sum_{s=1}^{t-1} 4\sigma_{0,s}^2 n^{-1}\sum_{i=1}^n\sum_{j=1}^n E\left[c_{ji,ts}^2|\mathcal{C}\right].$$

*Then for* $m = 1, 2, 3$,

$$n^{-1}\sum_{i=1}^n \varsigma_{it}^{(m)} \xrightarrow{p} \varsigma_t^{(m)} \quad \text{as } n \to \infty.$$

*Furthermore,* $n^{-1}\sum_{i=1}^n \varsigma_{it}^{(4)} \xrightarrow{p} 0$, $n^{-1}\sum_{t=1}^T \sigma_{0,t}^2\sum_{i=1}^n \varsigma_{it}^{(5)} \to 0$ *and* $n^{-1}\sum_{i=1}^n \varsigma_{it}^{(6)} \xrightarrow{p} 0$ *as* $n \to \infty$.

[21] As to potential alternative selections of the information sets, we note that defining $\mathcal{F}_{n,(t-1)n+i} = \mathcal{B}_{n,i,t} \vee \mathcal{C}$ yields information sets that are not adaptive, and defining $\mathcal{F}_{n,(t-1)n+i} = \sigma\left\{\left(x_{j1}^o, z_j, \mu_j\right)_{j=1}^n\right\} \vee \mathcal{C}$ would violate the condition that $X_{n,v}$ is $\mathcal{F}_{n,v}$-measurable.

**Proof.** See supplementary appendix. ■

The next lemma will be useful for, e.g., establishing the consistency of variance co-variance matrix estimators. We consider general (not necessarily criterion) functions $\mathcal{R}_n : \Omega \times \underline{\Theta}_\theta \to \mathbb{R}$ and $\mathcal{R} : \Omega \times \underline{\Theta}_\theta \to \mathbb{R}$.

**Lemma A.3** *Suppose $\mathcal{R}(\omega, \theta)$ is a.s. uniformly continuous on $\underline{\Theta}_\theta$, where $\underline{\Theta}_\theta$ is a subset of $\mathbb{R}^{p_\theta}$, suppose $\widehat{\theta}_n$ and $\theta_*$ are random vectors with $\widehat{\theta}_n \to \theta_*$ a.s. [i.p.], and*

$$\sup_{\theta \in \underline{\Theta}_\theta} |\mathcal{R}_n(\omega, \theta) - \mathcal{R}(\omega, \theta)| \to 0 \text{ a.s.[i.p.] as } n \to \infty, \tag{A.7}$$

*then*

$$\mathcal{R}_n(\omega, \widehat{\theta}_n) - \mathcal{R}(\omega, \theta_*) \to 0 \text{ a.s.[i.p.] as } n \to \infty. \tag{A.8}$$

**Proof.** See Supplementary Appendix. ■

The next lemma is useful in establishing uniform convergence of the objective function of the GMM estimator from uniform convergence of the sample moments. In the following proposition $\mathfrak{m}_n : \Omega \times \underline{\Theta}_\theta \to \mathbb{R}^m$ and $\mathfrak{m} : \Omega \times \underline{\Theta}_\theta \to \mathbb{R}^m$ should be viewed as the sample moment vector and the corresponding "limiting" moment vector.

**Lemma A.4** *Suppose $\underline{\Theta}_\theta$ is compact, $\mathfrak{m}(\omega, \theta) \subseteq K \subseteq \mathbb{R}^{p_m}$ for all $\theta \in \underline{\Theta}_\theta$ a.s. with $K$ compact, and*

$$\sup_{\theta \in \underline{\Theta}_\theta} \|\mathfrak{m}_n(\omega, \theta) - \mathfrak{m}(\omega, \theta)\| \to 0 \text{ a.s.[i.p.] as } n \to \infty. \tag{A.9}$$

*Furthermore, let $\Xi_n$ and $\Xi$ be $p_m \times p_m$ real valued random matrices, and suppose that $\Xi_n - \Xi \to 0$ a.s. [i.p.] where $\Xi$ is finite a.s.. Then*

$$\sup_{\theta \in \underline{\Theta}_\theta} \left| \mathfrak{m}_n(\omega, \theta)' \Xi_n \mathfrak{m}_n(\omega, \theta) - \mathfrak{m}(\omega, \theta)' \Xi \mathfrak{m}(\omega, \theta) \right| \to 0 \text{ a.s.[i.p.] as } n \to \infty. \tag{A.10}$$

**Proof.** See Supplementary Appendix. ■

**Lemma A.5** *Suppose Assumptions 1- 5 hold, and let $\bar{\gamma}_n \xrightarrow{p} \bar{\gamma}_*$ with $\bar{\gamma}_n \in \Theta_\gamma$ and $\bar{\gamma}_* \in \Theta_\gamma$, where $\bar{\gamma}_*$ is $\mathcal{C}$-measurable. Then*
*(i) $\mathfrak{m}(\theta) = \operatorname{plim}_{n \to \infty} n^{-1/2} \overline{m}_n(\theta, \bar{\gamma}_*)$ exists for each $\theta \in \underline{\Theta}_\theta$,with $\mathfrak{m} : \Omega \times \underline{\Theta}_\theta \to K$ where $K$ a compact subset of $\mathbb{R}^p$, $\mathfrak{m}(\theta)$ is $\mathcal{C}$-measurable for each $\theta \in \underline{\Theta}$.*
*(ii) $G(\theta) = \operatorname{plim}_{n \to \infty} \partial n^{-1/2} \overline{m}_n(\theta, \gamma_*)/\partial \theta$ exists and is finite for each $\theta \in \underline{\Theta}_\theta$, $G(\theta)$ is $\mathcal{C}$-measurable for each $\theta \in \underline{\Theta}$, and $G(\theta)$ is uniformly continuous on $\underline{\Theta}_\theta$.*

**Proof.** See Supplementary Appendix. ■

## A.3 Main Results

**Proof of Proposition 1.** Given the explicit expressions for the elements of $\Pi$ the claims of the proposition can be readily verified by straight forward calculations.[22] ■

**Proof of Proposition 2.** The proof of the proposition uses methodology similar to that used in establishing (A.12) below in the proof of Theorem 2. Explicit derivations are available in the supplementary appendix. ■

**Proof of Proposition 3.** An inspection of the proof of, e.g., Lemma 3.1 in Pötscher and Prucha (1997) shows that the proof of the a.s. version of their Lemma 3.1 goes through even if the "limiting" objective functions $\overline{R}_n$ and the minimizers $\overline{\beta}_n$ are allowed to be random, and the identifiable uniqueness assumption (3.1) is only assumed to holds a.s.. The convergence i.p. version of the proposition follows again from a standard subsequence argument. Consequently Proposition 3 is seen to hold as a special case of the generalized Lemma 3.1 in Pötscher and Prucha (1997). ■

**Proof of Theorem 1.** $\mathcal{R}_n(\theta) = n^{-1}\overline{m}_n(\theta,\bar{\gamma}_n)'\tilde{\Xi}_n\overline{m}_n(\theta,\bar{\gamma}_n)$ and $\mathcal{R}(\theta) = \mathfrak{m}(\theta)'\Xi\mathfrak{m}(\theta)$. We use Proposition 3 to prove the theorem. Under the maintained assumptions, $\theta_*$ is identifiable unique in the sense of Condition (i) of Proposition 3. This is seen to hold in light of Condition (25) of Assumption 6, and by observing that $\mathcal{R}(\theta_*) = \mathfrak{m}(\theta_*)'\Xi\mathfrak{m}(\theta_*) = 0$ and

$$\mathcal{R}(\theta) = \mathfrak{m}(\theta)'\Xi\mathfrak{m}(\theta) \geq \lambda_{\min}(\Xi)\left\|\mathfrak{m}(\theta)\right\|^2,$$

with $\lambda_{\min}(\Xi) > 0$ a.s. by Assumption 7. To verify Condition (ii) of Proposition 3 we employ Lemma A.4. By Lemma A.5 we have $\mathfrak{m}(\theta) \in K$, where $K$ is compact, and $\mathfrak{m}(\theta)$ is $\mathcal{C}$-measurable. By Assumption 6 we have

$$\sup_{\theta\in\underline{\Theta}_\theta}\left\|n^{-1/2}m_n(\theta,\bar{\gamma}_n) - \mathfrak{m}(\theta)\right\| = o_p(1).$$

Furthermore, observe that by Assumptions 7 we have $\tilde{\Xi}_n - \Xi = o_p(1)$ where $\Xi$ is $\mathcal{C}$-measurable and finite a.s. Having verified all assumptions of Lemma A.4 it follows from that Lemma that also Condition (ii) of Proposition 3, i.e.,

$$\sup_{\theta\in\underline{\Theta}_\theta}\left\|\mathcal{R}_n(\theta) - \mathcal{R}(\theta)\right\| = o_p(1),$$

holds. Having verified both conditions of Proposition 3 it follows from that proposition that $\tilde{\theta}_n(\bar{\gamma}_n) - \theta_* \xrightarrow{p} 0$ and consequently $\tilde{\theta}_n(\bar{\gamma}_n) - \theta_{n,0} \xrightarrow{p} 0$ as $n \to \infty$. ■

**Proof of Theorem 2.** To derive the limiting distribution we apply the martingale difference central limit theorem (MD-CLT) developed in Kuersteiner and Prucha (2013),

[22] A constructive proof, which allowed us to find the explicit expressions for the elements of $\Pi$, is significantly more involved and available on request.

which is given as Theorem 1 in that paper. To apply the MD-CLT we verify that the assumptions maintained by the theorem hold here. Observe that $\mathcal{F}_0 = \bigcap_{n=1}^{\infty} \mathcal{F}_{n,0} = \mathcal{C}$ and $\mathcal{F}_{n,0} \subseteq \mathcal{F}_{n,1}$ for each $n$ and $E\left[X_{n,1}|\mathcal{F}_{n,0}\right] = 0$ where $X_{n,v}$ is defined in (A.6). In the proof of Theorem 2 of Kuersteiner and Prucha (2013) it is shown that the following conditions are sufficient for conditions (14)-(16) there, postulated by the MD-CLT, to hold:

$$\sum_{v=1}^{k_n} E\left[|X_{n,v}|^{2+\delta}\right] \to 0, \tag{A.11}$$

$$V_{nk_n}^2 = \sum_{v=1}^{k_n} E\left[X_{n,v}^2|\mathcal{F}_{n,v-1}\right] \xrightarrow{p} \eta^2, \tag{A.12}$$

$$\sup_n E\left[V_{nk_n}^{2+\delta}\right] = \sup_n E\left[\left(\sum_{v=1}^{k_n} E\left[X_{n,v}^2|\mathcal{F}_{n,v-1}\right]\right)^{1+\delta/2}\right] < \infty. \tag{A.13}$$

with $k_n = Tn + 1$. In the following we verify (A.11)-(A.13) with $\eta^2 = v_\lambda = \lambda' V \lambda$, for any $\lambda \in \mathbb{R}^p$ such that $\lambda'\lambda = 1$.

For the verification of Condition (A.11) let $q = 2+\delta$, $1/q+1/p = 1$ and $v = (t-1)n+i+1$. Observe that using inequality (1.4.4) in Bierens (1994) we have

$$\begin{aligned} |X_{n,v}|^q \quad &\leq \quad \frac{2^q(T+1)^q}{n^{1+\delta/2}} |u_{it}|^q \left\{ |c_{it}|^q + \left( \sum_{j=1}^{i-1} |c_{ij,tt}|^{1/p} |c_{ij,tt}|^{1/q} |u_{jt}| \right)^q \right. \\ &\qquad \left. + \sum_{s=1}^{t-1} \left( \sum_{j=1}^{n} |c_{ij,ts}|^{1/p} |c_{ij,ts}|^{1/q} |u_{js}| \right)^q \right\} \end{aligned}$$

such that by Hölder's inequality

$$\begin{aligned} |X_{n,v}|^q \quad &\leq \quad \frac{2^q(T+1)^q}{n^{1+\delta/2}} |u_{it}|^q \left\{ |c_{it}|^q + \left( \sum_{j=1}^{i-1} |c_{ij,tt}| \right)^{q/p} \sum_{j=1}^{i-1} |c_{ij,tt}| |u_{jt}|^q \right. \\ &\qquad \left. + \sum_{s=1}^{t-1} \left( \sum_{j=1}^{n} |c_{ij,ts}| \right)^{q/p} \left( \sum_{j=1}^{n} |c_{ij,ts}| |u_{js}|^q \right) \right\}. \end{aligned}$$

Consequently, recalling from Section A.1 that $c_{it}$ and $c_{ij,ts}$ are measurable w.r.t. $\mathcal{F}_{n,(t-1)n+i}$

it follows that

$$\begin{aligned} E\left[\left|X_{n,v}\right|^q \left|\mathcal{F}_{n,v-1}\right.\right] &\leq \frac{2^q(T+1)^q}{n^{1+\delta/2}} E\left[\left|u_{it}\right|^q \left|\mathcal{F}_{n,(t-1)n+i}\right.\right] \left\{ \left|c_{it}\right|^q + \left(\sum_{j=1}^{i-1} \left|c_{ij,tt}\right|\right)^{q/p} \sum_{j=1}^{i-1} \left|c_{ij,tt}\right| \left|u_{jt}\right|^q \right. \\ &\quad \left. + \sum_{s=1}^{t-1} \left(\sum_{j=1}^{n} \left|c_{ij,ts}\right|\right)^{q/p} \left(\sum_{j=1}^{n} \left|c_{ij,ts}\right| \left|u_{js}\right|^q\right) \right\} \\ &\leq \frac{2^q(T+1)^q}{n^{1+\delta/2}} K \left\{ \left|c_{it}\right|^q + K^{q/p} \sum_{s=1}^{t} \left(\sum_{j=1}^{n} \left|c_{ij,ts}\right| \left|u_{js}\right|^q\right) \right\} \end{aligned}$$

where we have used bounds in Lemma A.1(ii),(v) to establish the last inequality. Employing Lemma A.1(i) and (vi) we have

$$\begin{aligned} E\left[\left|X_{n,v}\right|^q\right] &= E\left[E\left[\left|X_{n,v}\right|^q \left|\mathcal{F}_{n,v-1}\right.\right]\right] \\ &\leq \frac{2^q(T+1)^q}{n^{1+\delta/2}} K \left\{ E\left[\left|c_{it}\right|^q\right] + K^{q/p} \sum_{s=1}^{t} \left(\sum_{j=1}^{n} E\left[\left|c_{ij,ts}\right| \left|u_{js}\right|^q\right]\right) \right\} \\ &\leq \frac{2^q(T+1)^q}{n^{1+\delta/2}} K \left(K + TK^{q/p+1}\right). \end{aligned}$$

Consequently, recalling that $k_n = Tn+1$,

$$\sum_{v=1}^{k_n} E\left[\left|X_{n,v}\right|^{2+\delta}\right] \leq \sum_{v=1}^{k_n} E\left[E\left[\left|X_{n,v}\right|^{2+\delta} \left|\mathcal{F}_{n,v-1}\right.\right]\right] \leq \frac{2^{2+\delta}(T+1)^{3+\delta} K^2}{n^{\delta/2}} \left(1 + TK^{1+\delta}\right) \to 0,$$

which verifies condition (A.11).

To verify (A.12) with $\eta^2 = v_\lambda = \lambda' V \lambda$ we first calculate

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

Recall from Section A.1 that the $\varrho_{0,i}^2$ and $\varrho_i$ are absorbed into $h_{it}$ **and** $a_{ij,t}$, and thus by Assumption 1 we have $E\left[u_{it}^2 | \mathcal{F}_{n,(t-1)n+i}\right] = \sigma_{0,t}^2$. Furthermore, recalling that $c_{it}$ and $c_{ij,ts}$ are measurable w.r.t. $\mathcal{F}_{n,(t-1)n+i}$.we have

$$\begin{aligned} E\left[X_{n,v}^2 | \mathcal{F}_{n,v-1}\right] &= E\left[X_{n,(t-1)n+i+1}^2 | \mathcal{F}_{n,(t-1)n+i}\right] \\ &= n^{-1} \sigma_{0,t}^2 \left(c_{it} + 2\sum_{j=1}^{i-1} c_{ij,tt} u_{jt} + 2\sum_{s=1}^{t-1} \sum_{j=1}^{n} c_{ij,ts} u_{js}\right)^2 \\ &= \sigma_{0,t}^2 n^{-1} \sum_{m=1}^{6} \varsigma_{it}^{(m)} \end{aligned}$$

where the $\varsigma_{it}^{(m)}$ are defined in Lemma A.2. Thus

$$V_{nk_n}^2 = \sum_{v=1}^{k_n} E\left[X_{n,v}^2|\mathcal{F}_{n,v-1}\right] = \sum_{m=1}^{6}\sum_{t=1}^{T}\sigma_{0,t}^2 n^{-1}\sum_{i=1}^{n}\varsigma_{it}^{(m)}. \tag{A.14}$$

Given the probability limits of $n^{-1}\sum_{i=1}^{n}\varsigma_{it}^{(m)}$, for $m = 1,\ldots,6$ derived in Lemma A.2 we have

$$V_{nk_n}^2 = \sum_{v=1}^{k_n} E\left[X_{n,v}^2|\mathcal{F}_{n,v-1}\right] = \sum_{m=1}^{6}\sum_{t=1}^{T}\sigma_{0,t}^2 n^{-1}\sum_{i=1}^{n}\varsigma_{it}^{(m)} \xrightarrow{p} \eta_*^2$$

with

$$\begin{aligned}
\eta_*^2 &= \sum_{t=1}^{T}\sigma_{0,t}^2\left(\varsigma_t^{(1)}+\varsigma_t^{(2)}+\varsigma_t^{(3)}\right) = \operatorname*{plim}_{n\to\infty}\left(\sum_{t=1}^{T}\sigma_{0,t}^2 n^{-1}\sum_{i=1}^{n}E\left[c_{it}^2|\mathcal{C}\right]\right)\\
&\quad+\operatorname*{plim}_{n\to\infty}\left(2\sum_{t=1}^{T}\sigma_{0,t}^4 n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}E\left[c_{ij,tt}^2|\mathcal{C}\right]+4\sum_{t=1}^{T}\sigma_{0,t}^2\sum_{s=1}^{t-1}\sigma_{0,s}^2 n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}E\left[c_{ji,ts}^2|\mathcal{C}\right]\right).
\end{aligned}$$

Recall that for $t = 1,\ldots,T$ we have $c_{it} = \sum_{\tau=1}^{t}\lambda_\tau' h_{i\tau}'\pi_{\tau t} = \sum_{\tau=1}^{T-1}\lambda_\tau' h_{i\tau}'\pi_{\tau t}$ where the last equality holds since $\pi_{\tau t} = 0$ for $\tau > t$. Thus

$$\begin{aligned}
\sum_{u=1}^{T}\sigma_{0,u}^2\sum_{i=1}^{n}c_{iu}^2 &= \sum_{u=1}^{T}\sigma_{0,u}^2\sum_{i=1}^{n}\sum_{t=1}^{T-1}\lambda_t' h_{it}'\pi_{tu}\sum_{\tau=1}^{T-1}\lambda_\tau' h_{i\tau}'\pi_{\tau u}\\
&= \sum_{i=1}^{n}\sum_{t=1}^{T-1}\sum_{\tau=1}^{T-1}\lambda_t' h_{it}'\lambda_\tau' h_{i\tau}'\left(\pi_t\Sigma_{0,\sigma}\pi_\tau'\right) = \sum_{i=1}^{n}\sum_{t=1}^{T-1}\lambda_t' h_{it}'\lambda_\tau' h_{it}\lambda_t
\end{aligned}$$

observing that $\pi_t\Sigma_{0,\sigma}\pi_\tau' = \sum_{u=1}^{T}\sigma_{0,u}^2\pi_{tu}\pi_{\tau u}$ and $\Pi\Sigma_{0,\sigma}\Pi' = I_{T-1}$.

Recall further that for $t = 1,\ldots,T$, $s \le t$, we have $c_{ij,ts} = \sum_{\tau=1}^{s}\lambda_\tau' a_{ij,\tau}'\pi_{\tau s}\pi_{\tau t} = \sum_{\tau=1}^{T-1}\lambda_\tau' a_{ij,\tau}'\pi_{\tau s}\pi_{\tau t}$ where the last equality holds since $\pi_{\tau s} = 0$ for $\tau > s$. Thus, by straight forward algebra,

$$\begin{aligned}
&2\sum_{t=1}^{T}\sigma_{0,t}^4\sum_{i,j=1}^{n}c_{ij,tt}^2+4\sum_{t=1}^{T}\sigma_{0,t}^2\sum_{s=1}^{t-1}\sigma_{0,s}^2\sum_{i,j=1}^{n}c_{ji,ts}^2 = 2\sum_{t,s=1}^{T}\sigma_{0,t}^2\sigma_{0,s}^2\sum_{i,j=1}^{n}c_{ji,ts}^2\\
&= 2\sum_{t,s=1}^{T-1}\sum_{i,j=1}^{n}\lambda_t' a_{ij,t}'\lambda_s' a_{ij,s}'\left(\pi_t\Sigma_{0,\sigma}\pi_s'\right)^2 = 2\sum_{t=1}^{T-1}\sum_{i,j=1}^{n}\lambda_t' a_{ij,t}' a_{ij,t}\lambda_t,
\end{aligned}$$

observing again that $\Pi\Sigma_{0,\sigma}\Pi' = I_{T-1}$. From this we see that

$$\begin{aligned}
\eta_*^2 &= \operatorname*{plim}_{n\to\infty}\sum_{t=1}^{T-1}\lambda_t'\left\{n^{-1}\sum_{i=1}^{n}E\left[h_{it}'h_{it}|\mathcal{C}\right]+2n^{-1}\sum_{i,j=1}^{n}E\left[a_{ij,t}'a_{ij,t}|\mathcal{C}\right]\right\}\lambda_t\\
&= \sum_{t=1}^{T-1}\lambda_t'\left[V_t^h+2V_t^a\right]\lambda_t = \lambda' V\lambda,
\end{aligned}$$

which establishes that indeed $V_{nk_n}^2 \xrightarrow{p} \eta^2 = \lambda' V \lambda$.

Finally, we verify Condition (A.13). Analogously as in the verification of Condition (A.11) observe that using the triangle inequality

$$\begin{aligned} |X_{n,v}|^2 \quad &\leq \quad \frac{4(T+1)^2}{n} |u_{it}|^2 \left\{ |c_{it}|^2 + \left( \sum_{j=1}^{i-1} |c_{ij,tt}|^{1/2} |c_{ij,tt}|^{1/2} |u_{jt}| \right)^2 \right. \\ &\left. + \sum_{s=1}^{t-1} \left( \sum_{j=1}^{n} |c_{ij,ts}|^{1/2} |c_{ij,ts}|^{1/2} |u_{js}| \right)^2 \right\} \end{aligned}$$

and by subsequently applying Hölder's inequality we have

$$\begin{aligned} |X_{n,v}|^2 \quad &\leq \quad \frac{4(T+1)^2}{n} |u_{it}|^2 \left\{ |c_{it}|^2 + \left( \sum_{j=1}^{i-1} |c_{ij,tt}| \right) \sum_{j=1}^{i-1} |c_{ij,tt}| |u_{jt}|^2 \right. \\ &\left. + \sum_{s=1}^{t-1} \left( \sum_{j=1}^{n} |c_{ij,ts}| \right) \left( \sum_{j=1}^{n} |c_{ij,ts}| |u_{js}|^2 \right) \right\}. \end{aligned}$$

Consequently in light of Lemma A.1 (ii) and (v)

$$\begin{aligned} & E\left[ |X_{n,v}|^2 | \mathcal{F}_{n,v-1} \right] \\ \leq \quad & \frac{4(T+1)^2}{n} E\left[ |u_{it}|^2 | \mathcal{F}_{n,(t-1)n+i} \right] \left\{ |c_{it}|^2 + K \sum_{j=1}^{i-1} |c_{ij,tt}| |u_{jt}|^2 \right. \\ & \left. + K \sum_{s=1}^{t-1} \sum_{j=1}^{n} |c_{ij,ts}| |u_{js}|^2 \right\} \\ \leq \quad & \frac{4(T+1)^2 K^2}{n} \left\{ |c_{it}|^2 + \sum_{j=1}^{i-1} |c_{ij,tt}| |u_{jt}|^2 + \sum_{s=1}^{t-1} \sum_{j=1}^{n} |c_{ij,ts}| |u_{js}|^2 \right\}. \end{aligned}$$

In light of the above inequality

$$
\begin{aligned}
& E\left[V_{nk_n}^{2+\delta}\right] \\
= \; & E\left[\left(\sum_{v=1}^{k_n} E\left[X_{n,v}^2|\mathcal{F}_{n,v-1}\right]\right)^{1+\delta/2}\right] \\
\leq \; & \frac{2^{2+\delta}(T+1)^{2+\delta}K^{2+\delta}}{n^{1+\delta/2}} E\left[\left\{\sum_{v=1}^{k_n}\left(|c_{it}|^2+\sum_{j=1}^{i-1}|c_{ij,tt}|\,|u_{jt}|^2+\sum_{s=1}^{t-1}\sum_{j=1}^{n}|c_{ij,ts}|\,|u_{js}|^2\right)\right\}^{1+\delta/2}\right] \\
\leq \; & \frac{2^{2+\delta}(T+1)^{2+\delta}K^{2+\delta}k_n^{\delta/2}}{n^{1+\delta/2}} \sum_{v=1}^{k_n} E\left[\left(|c_{it}|^2+\sum_{j=1}^{i-1}|c_{ij,tt}|\,|u_{jt}|^2+\sum_{s=1}^{t-1}\sum_{j=1}^{n}|c_{ij,ts}|\,|u_{js}|^2\right)^{1+\delta/2}\right] \\
\leq \; & \frac{3^{\delta/2}2^{2+\delta}(T+1)^{2+\delta}K^{2+\delta}k_n^{\delta/2}}{n^{1+\delta/2}} \sum_{v=1}^{k_n}\left\{E\left[|c_{it}|^{2+\delta}\right]+E\left[\left(\sum_{j=1}^{i-1}|c_{ij,tt}|\,|u_{jt}|^2\right)^{1+\delta/2}\right]\right. \\
& \left. +T^{\delta/2}\sum_{s=1}^{t-1}E\left[\left(\sum_{j=1}^{n}|c_{ij,ts}|\,|u_{js}|^2\right)^{1+\delta/2}\right]\right\}
\end{aligned}
$$

where we have used repeatedly inequality (1.4.3) in Bierens(1994). By Lemma A.1 (i) we have $E\left[|c_{it}|^{2+\delta}\right] \leq K$. Applying Hölder's inequality with $q = 1+\delta/2$ and $1/p+1/q = 1$, and utilizing Lemma A.1 (ii)-(vi) we have:

$$
\begin{aligned}
& E\left[\left(\sum_{j=1}^{n}|c_{ij,ts}|\,|u_{js}|^2\right)^{1+\delta/2}\right] = E\left[\left(\sum_{j=1}^{n}|c_{ij,ts}|^{1/p}\,|c_{ij,ts}|^{1/q}\,|u_{js}|^2\right)^{1+\delta/2}\right] \\
\leq \; & E\left[\left(\sum_{j=1}^{n}|c_{ij,ts}|\right)^{q/p}\left(\sum_{j=1}^{n}|c_{ij,ts}|\,|u_{js}|^{2+\delta}\right)\right] \leq K^{q/p}\sum_{j=1}^{n}E\left[|c_{ij,ts}|\,|u_{js}|^{2+\delta}\right] \leq K^{1+q/p}
\end{aligned}
$$

and by the same arguments $E\left[\left(\sum_{j=1}^{i-1}|c_{ij,tt}|\,|u_{jt}|^2\right)^{1+\delta/2}\right] \leq K^{1+q/p}$. Consequently, observing that $q/p = \delta/2$ and $k_n/n \leq T+1$,

$$
\begin{aligned}
E\left[V_{nk_n}^{2+\delta}\right] & \leq \frac{3^{\delta/2}2^{2+\delta}(T+1)^{2+\delta}K^{2+\delta}k_n^{\delta/2}3T^{1+\delta/2}k_nK^{1+\delta/2}}{n^{1+\delta/2}} \\
& \leq 3^{1+\delta/2}2^{2+\delta}(T+1)^{4+2\delta}K^{3+3\delta/2} < \infty
\end{aligned}
$$

which verifies condition (A.13). Consequently it follows from Kuersteiner and Prucha (2013, Theorem 1) that $\lambda'\overline{m}_n = \sum_{v=1}^{Tn+1} X_{n,v} \xrightarrow{d} \eta\xi_0$ ($\mathcal{C}$-stably), where $\xi_0$ and $\mathcal{C}$ are independent. Applying the Cramer-Wold device - see, e.g., Kuersteiner and Prucha (2013, Proposition

A.2) it follows further that $\overline{m}_n \xrightarrow{d} V^{1/2}\xi$ ($\mathcal{C}$-stably) where $\xi \sim N(0, I_p)$ and $\xi$ and $\mathcal{C}$ are independent.

Recall that in establishing the martingale difference representation of $\lambda'\overline{m}_n$ we have absorbed $\varrho_{0,i}/\varrho_i$ into $h_{it}$ and $a_{ijt}$. The expression for $V_\varrho$ given in Assumption 3 is obtained upon reversing this absorption. ■

**Proof of Theorem 3.** The proof follows from standard arguments. Details are given in the supplementary appendix. ■

**Proof of Theorem 4.** As remarked in the text, $\widetilde{V}_n^{-1} \xrightarrow{p} V^{-1}$ with $V^{-1}$ being $\mathcal{C}$-measurable with *a.s.* finite elements, and with $V^{-1}$ positive definite *a.s.* Furthermore, as established in the proof of Theorem 3, $G_n(\hat{\theta}_n, \tilde{\gamma}_n) \xrightarrow{p} G$ where $G$ is $\mathcal{C}$-measurable with *a.s.* finite elements, and with full column rank *a.s.* Thus $\hat{\Psi}_n = \left(G_n(\hat{\theta}_n, \tilde{\gamma}_n)'\widetilde{V}_n^{-1}G_n(\hat{\theta}_n, \tilde{\gamma}_n)\right)^{-1} \xrightarrow{p} \Psi = (G'V^{-1}G)^{-1}$. It now follows from part (i) of Theorem 3 that

$$n^{1/2}(\hat{\theta}_n - \theta_{n,0}) \xrightarrow{d} \Psi^{1/2}\xi_*, \tag{A.15}$$

where $\xi_*$ is independent of $\mathcal{C}$ (and hence of $\Psi$), $\xi \sim N(0, I_{p_\theta})$. In light of (A.15), the consistency of $\hat{\Psi}_n$, and given that $R$ has full row rank $q$ it follows furthermore that under $H_0$

$$\begin{aligned}
\left(R\hat{\Psi}R'\right)^{-1/2} n^{1/2}(R\hat{\theta}_n - r) &= \left(R\hat{\Psi}R'\right)^{-1/2} R\left(n^{1/2}(\hat{\theta}_n - \theta_{n,0})\right) \\
&= \left(R\Psi R'\right)^{-1/2} R\left(n^{1/2}(\hat{\theta}_n - \theta_{n,0})\right) + o_p(1).
\end{aligned}$$

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

$$\left(R\hat{\Psi}R'\right)^{-1/2} n^{1/2}(R\hat{\theta}_n - r) \xrightarrow{d} \xi_{**} \tag{A.16}$$

where $\xi_{**} \sim N(0, I_q)$. Hence, in light of the continuous mapping theorem, $T_n$ converges in distribution to a chi-square random variable with $q$ degrees of freedom. The claim that $\hat{\Psi}_n^{-1/2}\sqrt{n}(\hat{\theta}_n - \theta_{n,0}) \xrightarrow{d} \xi_*$ is seen to hold as a special case of (A.16) with $R = I$ and $r = \theta_0$. ■

## References


[1] Ahn, S. G.,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.

[2] _______(2013): "Panel data models with multiple time-varying individual effects," *Journal of Econometrics* 174, 1-14.

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

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

[5] Anselin, L., 1988, *Spatial Econometrics: Methods and Models*, Kluwer, Boston.

[6] Arellano, M and O. Bover, 1995, Another look at the instrumental variable estimation of error-components models, *Journal of Econometrics*, 68, 29-51.

[7] Atchad, Y.F., and M.D. Cattaneo, 2012, Limit theorems for quadratic forms of Markov chains, Department of Economics, University of Michigan, Working Paper.

[8] Bai, J., 2009, Panel data models with interactive fixed effects, *Econometrica* 77, 1229-1279.

[9] Bai, J., 2013, Likelihood approach to dynamic panel models with interactive effects, MRPA working paper.

[10] Bai, J., and K. Li, 2013, Spatial panel data models with common shocks, MRPA working paper.

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

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

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

[14] Blume, L.E., W.A. Brock, S.N. Durlauf, and Y.M. Ioannides, 2011, Identification of social interactions. In J. Benhabib, M.O. Jackson and A. Bisin, eds., *Handbook of Social Economics*, Vol. 1B, North-Holland, Amsterdam, 853-964.

[15] Carrell, S. E., B. I. Sacerdote, and J. E. West, 2013, From natural variation to optimal policy? The importance of endogenous peer group formation, *Econometrica,* 81, vol. 3, 855-882.

[16] Chen, D., B. Moulin, and J. Wu, eds., 2014, *Analyzing and Modeling Spatial and Temporal Dynamics of Infectious Diseases*, Wiley, New York.

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

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

[19] Chudik, A., and M.H. Pesaran, 2013, Common Correlated Effects Estimation of Heterogeneous dynamic panel data models with weakly exogenous regressors, CES Ifo working paper.

[20] Doukhan, P., J.R. Leon, and P. Soulier, 1996, Central and non central limit theorems for quadratic forms of strongly dependent Gaussian fields. *Brazilian Journal of Probability and Statistics* 10, 205-223.

[21] Elhorst, P., 2010, Dynamic panels with endogenous interaction effects when $T$ is small, *Regional Science and Urban Economics* 20, 272-282.

[22] Gallant, A.R., and H. White, 1988, *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*, Basil Blackwell, New York.

[23] Gao, J., and H. Yongmiao, 2007, Central limit theorems for weighted quadratic forms of dependent processes with applications in specification testing, MPRA Working Paper.

[24] Giraitis, L., and M.S. Taqqu, 1998, Central limit theorems for quadratic forms with time-domain conditions. *Annals of Probability* 26, 377–398.

[25] Goldsmith-Pinkham, P. and G. W. Imbens, 2013, Social Networks and the Identification of Peer Effects, *Journal of Business & Economic Statistics,* 31, pp. 253-264.

[26] Graham, B. S., Identifying social interactions through conditional variance restrictions, *Econometrica,* 76, vol. 3, 643-660.

[27] Hayakawa, K, 2006, The asymptotic properties of the system GMM estimator in dynamic panel data models when both N and T are large. mimeo.

[28] Hayashi, F., and C Sims, 1983, Nearly Efficient Estimation of Time Series Models with Predetermined, but not Exogenous, Instruments, *Econometrica* 51, 783-798.

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

[30] Holtz-Eakin, D.,W. Newey, and H. Rosen (1988), Estimating Vector Autoregressions With Panel Data, *Econometrica*, 56, 1371–1395.

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

[32] Keeling, M.J., and P. Rohani, 2008, *Modeling Infectious Diseases in Humans and Animals*, Princeton University Press, Princeton.

[33] Kelejian, H.H., and I.R. Prucha, 1998, A generalized spatial two stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances, *Journal of Real Estate Finance and Economics* 17, 99-121.

[34] Kelejian, H.H., and I.R. Prucha, 1999, A generalized moments estimator for the autoregressive parameter in a spatial model, *International Economic Review* 40, 509-533.

[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] Korolyuk, V.S., and Y.V. Borovskich, 1994, *Theory of U-Statistics*, Springer Verlag, Dordrecht.

[37] Kuersteiner, G.M., and I.R. Prucha, 2013, Limit theory for panel data models with cross sectional dependence and sequential exogeneity, *Journal of Econometrics* 174, 107-126.

[38] Lee, L.-F., and X. Liu, 2010, Efficient GMM estimation of higher order spatial autoregressive models with autoregressive disturbances, *Econometric Theory* 26, 187-230.

[39] Lee, L.-F., X. Liu and X. Lin, 2010, Specification and estimation of social interaction models with network structures, *Econometrics Journal* 13, 145-176.

[40] Lee, L.-F., and J. Yu, 2014, Efficient GMM estimation of spatial dynamic panel data models with fixed effects, *Journal of Econometrics* 180, 174–197.

[41] Liu, X., and L.-F. Lee, 2010, GMM estimation of social interaction models with centrality, *Journal of Econometrics* 159, 99-115.

[42] Moon, H.R., and M. Weidner, 2013a, Dynamic linear panel regression models with interactive fixed effects, CeMMaP working paper.

[43] Moon, H.R., and M. Weidner, 2013b, Linear regression for panel with unknown number of factors as interactive fixed effects, CeMMaP working paper.

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

[45] Newey, W.K. and D.L. McFadden. Large sample estimation and hypothesis testing. In R.F. Engle and D.L. McFadden (eds.), *Handbook of Econometrics*, Vol. IV, pp. 2113-2245. Amsterdam: Elsevier Science B.V., 1994.

[46] Patacchini, E., E. Rainone and Y. Zenou, 2013, Heterogeneous peer effects in education, Syracuse University, Department of Economics working paper.

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

[48] Pesaran, M.H., and E. Tosetti, 2011, Large panels with common factors and spatial correlation, *Journal of Econometrics* 161, 182-202.

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

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

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

[52] Renyi, A, 1963, On stable sequences of events. *Sankya Ser. A*, 25, 293-302.

[53] Su, L. J. and Z.L.Yang, 2014, QML estimation of dynamic panel models with spatial errors. *Journal of Econometrics*, forthcoming

[54] White, H., 1994, *Estimation, Inference and Specification Analysis*, Cambridge University Press, New York.

[55] 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.

[56] 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.