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

Guido M. Kuersteiner*and Ingmar R. Prucha†

First Version: July 17, 2015; Last Revised: March 1, 2020

Abstract

This paper considers a class of generalized methods of moments (GMM) estimators for general dynamic panel models, allowing for weakly exogenous covariates and cross sectional dependence due to spatial lags, unspecified common shocks and time-varying interactive effects. We significantly expand the scope of the existing literature by allowing for endogenous time varying spatial weight matrices without imposing explicit structural assumptions on how the weights are formed. An important area of application is in social interaction and network models where our specification can accommodate data dependent network formation. We consider an exemplary social interaction model and show how identification of the interaction parameters is achieved through a combination of linear and quadratic moment conditions. For the general setup 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 general, the asymptotic distribution of our estimators is found to be mixed normal due to random norming. However, the asymptotic distribution of our test statistics is still chi-square.

*Department of Economics, University of Maryland, College Park, MD 20742, e-mail: kuersteiner@econ.umd.edu

†Department of Economics, University of Maryland, College Park, MD 20742, e-mail: prucha@econ.umd.edu

1 Introduction[1]

Network and social interaction models have recently attracted attention both in empirical work as well as in econometric theory. In this paper we develop Generalized Methods of Moments (GMM) estimators for panel data with network structure. Our focus is on estimating linear models for outcome variables that may depend on outcomes and covariates of others in the network. We assume that the network structure is observed but do not impose any explicit structural restrictions on the process that generates the network. We allow for the network to change dynamically and being formed endogenously. Implicit restrictions we impose are in the form of high level assumptions about the convergence of sample moments. These assumptions imply restrictions on the amount of cross-sectional dependence one can allow for in covariates and on how dense the network can be. The assumptions are similar to high level assumptions imposed in Kuersteiner and Prucha (2013). In addition, when networks are formed endogenously we do assume that some sequentially exogenous covariates predict network formation. Recent work on the estimation of models with endogenous weights includes Goldsmith-Pinkham and Imbens (2013), Han and Lee (2016), Hsieh and Lee (2016) who propose Bayesian methods, Qu and Lee (2015), Qu, Lee and Yu (2017), Shi and Lee (2018) proposing quasi maximum likelihood estimators, Kelejian and Piras (2014) proposing GMM, Auerbach (2016) who develops a local matching estimator, and Arduini, Patacchini and Rainone (2015) and Johnson and Moon (2017) using a control function approach. All these papers assume specific generating mechanisms for the network formation process.

Because we do not estimate parameters of the network formation model and because our GMM estimators are identified from moment restrictions imposed on the idiosyncratic errors, our approach can be completely agnostic about the way the network is formed, at least as long as the network formation is sequentially exogenous. When network matrices are endogenous, in the sense of being correlated with the idiosyncratic model errors, instruments for network matrices are required for identification. These instruments are constructed from

[1]We gratefully acknowledge financial support from the National Institute of Health through the SBIR grant R43 AG027622, 1 R43 AG056199-01 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 thank seminar and conference participants, as well as two co-editors, Ulrich Müller and Elie Tamer, and referees for their helpful comments.

sequentially exogenous covariates that predict network formation. While not required for our estimators, a network formation model may be helpful in thinking about such predictors.

Our work also extends the estimation theory for dynamic panel data models with higher order spatial lags to allow for interactive fixed effects, unobserved common factors affecting covariates and error terms and sequentially (rather than only strictly) exogenous regressors.[2] Our treatment of common shocks is inspired by Andrews (2005). Unlike Andrews (2005) 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. Our analysis is for panel data with the cross-sectional sample size n tending to infinity while the number time periods T is fixed. Our treatment of interactive effects is related to the large literature on panel models including Phillips and Sul (2003, 2007), Bai and Ng (2006a,b), Pesaran (2006), Bai (2009, 2013), Moon and Weidner (2015,2017). We propose a new quasi differencing transformation, given in Proposition 1, that we call the generalized Helmert transform to eliminate individual factor loadings and treat factors as estimands. Our transformation combines and extends quasi differencing proposed by Holtz-Eakin, Newey and Rosen (1988) and the Helmert transform of Arellano and Bover (1995) into an orthonormal forward filtering procedure with estimated filter weights. Our estimators are most closely related to the fixed T GMM estimators of Ahn et al. (2013).

The moment conditions of our GMM estimator depend on a general result, Theorem 1, for the mean, variances and covariances of linear-quadratic forms of transformed disturbances. The limiting properties of our GMM estimator and associated test statistics, given in Theorems 2-4, are based on Proposition 2, which establishes the consistency of stochastic minimizers and on Proposition 3 which is a new stable central limit theorem (CLT) for linear and quadratic forms. The CLT is suitable to handle the type of unmodeled cross-sectional dependence in covariates and heteroskedastic errors we allow for and builds on the CLT for linear forms of Kuersteiner and Prucha (2013). The CLT, as well as simplifications of the asymptotic variance of our estimators that are possible because of the way the generalized Helmert transform is designed, are critical inputs to the asymptotic methods for inference that we propose.

Our work also relates to the spatial literature dating back to Whittle (1954), Anselin (1988) and Cliff and Ord (1973, 1981), and the GMM framework based on linear and quadratic moment conditions introduced in Kelejian and Prucha (1998,1999) and Kapoor et al. (2007) for cross sectional and panel data. Dynamic panel data models with spatial

[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 notation to separate covariates that can be used as instruments from those that cannot. For ease of exposition we do not explicitly account for this possibility.

interactions 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 (2015). Papers combining spatial lags and common shocks include Chudik and Pesaran (2015), Bai and Li (2013), and Pesaran and Torsetti (2011). All of these papers assume that both n and T tend to infinity, they do not consider endogenous spatial weight matrices, and the latter two papers only consider a static setup. We significantly expand the scope of these models by allowing for dynamic and endogenous network formation in combination with interactive effects and common shocks affecting the covariates in non-parametric ways while still being able to provide tractable inference procedures.

Section 2 contains a worked example that illustrates the main features of our theoretical results, which are presented in Section 3. Appendix A contains formal assumptions, Appendix B develops the generalized Helmert transformation, and Appendix C contains proofs.[3]

2 Example and Motivation

We consider a stylized social interactions model to illustrate the main ideas behind our estimators and to illustrate identification and estimation for the general cross sectional interaction model considered in Section 3. Assume that we observe outcomes, collected in a vector $y_t = [y_{1t}, \ldots, y_{nt}]'$, for n individuals with exogenous characteristics collected in a matrix z_t^1. Interactivity between individuals is captured by an observed and possibly time varying $n \times n$ network interaction matrix M_t. Our setup allows for M_t to be determined endogenously, and allows for endogenous and exogenous peer or network effects captured, respectively, by $M_t y_t$ and $M_t z_t^1$ (Manski, 1993). Consider the following simple linear social interactions model with time-varying fixed effects,

$$y_t = \lambda M_t y_t + Z_t \beta + \varepsilon_t = W_t \delta + \varepsilon_t, \qquad \varepsilon_t = \mu f_t + u_t, \qquad t = 1, \ldots, T, \tag{1}$$

where $Z_t = [z_t^1, M_t z_t^1]$ is a $n \times p_z$ matrix, $\varepsilon_t = [\varepsilon_{1t}, ..., \varepsilon_{nt}]'$ denotes the vector of regression disturbances, $\mu = [\mu_1, \ldots, \mu_n]'$ denotes the vector of fixed effects, f_t is an unobserved scalar factor, $u_t = [u_{1t}, ..., u_{nt}]'$ denotes the vector of unobserved idiosyncratic disturbances, $W_t = [M_t y_t, Z_t]$, and $\delta = [\lambda, \beta']'$ is the vector of unknown parameters. For expositional purposes we assume that u_{it} is i.i.d. in both indices and we set $T = 2$ in this section. We relax both assumptions in Section 3 where u_{it} is allowed to be heteroskedastic and where independence is replaced by conditional mean independence.

[3] A supplementary appendix available separately provides additional details for the proofs.

The model in (1) illustrates the following main contributions of our paper: (i) We show how to handle endogenous and time varying spatial weight matrices and interactive fixed effects using linear and quadratic moment conditions. (ii) We show how a novel generalized Helmert transformation of the model can be used to eliminate the fixed effects μ, and orthogonalize both the linear and quadratic moments. We use the orthogonalization to simplify the criterion function and demonstrate how the simplification can be used to prove identification, facilitate inference and construct estimation algorithms. (iii) We illustrate how projections can be used to instrument for endogeneity in M_t. (iv) We develop a new CLT for linear quadratic moments, capable of handling the unmodeled cross-sectional dependence we allow for.

Define $z_t = [z_t^1, \zeta_t]$, where the matrix ζ_t collects additional exogenous variables which may be only partially observed, and where the number of columns of ζ_t can depend on n. For our example we assume further that z_t is strictly exogenous, and consider GMM estimators for the parameter δ based on the moment condition

$$E\left[u_{it}|z_1, z_2, \mu\right] = 0, \;\; t = 1, 2. \tag{2}$$

To keep the example simple, we also assume that conditionally on z_1, z_2 and μ the elements of $u = (u_1', u_2')'$ are mutually independent and identically distributed $(0, \sigma^2)$. It is well known that the parameter δ may not be identified by linear moment conditions alone, see for example Manski (1993), Kelejian and Prucha (2002), Kelejian et al. (2006), Lee (2007), Bramoulle, Djebbari and Fortin (2009) and de Paula (2017). Consistent with the spatial literature, to overcome the limitations of linear IV, our GMM estimator augments linear with quadratic moment conditions.

Model (1) accounts for cross-sectional correlation stemming both from individual interaction as well as from common factors. To make progress on our inference problem, we develop a novel generalized quasi-differencing transformation that efficiently eliminates the fixed effects μ.[4] We refer to this transformation as the generalized Helmert transform. When $T = 2$, we can, without loss of generality, normalize $f_2 = 1$. The transform for y_1 is then defined as $y_1^+ = (y_1 - f_1 y_2)/\sqrt{\left(f_1^2 + 1\right)\sigma^2}$. Using anlogous notation for the other variables we note that $u_1^+ = \varepsilon_1^+$. Since f_1 is unobserved in general, we treat it as a parameter to be estimated in Section 3. For expositional purposes we assume for now that f_1 is known. An important special case where f_1 is known, and equal to 1, is the pure fixed effects model. In this case our generalized quasi-differencing transformation is the same as

[4]See Section 3 and Proposition 1.

the Helmert transform. The transformed version of (1) can be written as

$$y_1^+ = \lambda(M_1 y_1)^+ + Z_1^+\beta + u_1^+ = W_1^+\delta + u_1^+. \tag{3}$$

It is convenient to adopt the following notation for the transformed spatial lag,

$$\overline{y}_1^+ = (M_1 y_1)^+ = (M_1 y_1 - f_1 M_2 y_2) / \sqrt{\left(f_1^2 + 1\right)\sigma^2}. \tag{4}$$

We formulate GMM estimators which exploit restrictions implied by (2) and the assumption, maintained for this example, that the elements of u are i.i.d. Let $h^r = (h_i^r)$, $r = 1, ..., p$, be a set of $n \times 1$ instrument vectors, and let $A^r = (a_{ij}^r)$, $r = 1, ..., q$, be a set of $n \times n$ symmetric matrices of instruments with zero diagonal elements $a_{ii}^r = 0$, where the elements of h^r and A^r are observed and measurable w.r.t. z_1, z_2, μ. It then follows from (2) that

$$E\left[h^{r\prime} u_1^+\right] = 0, \qquad E\left[u_1^{+\prime} A^r u_1^+\right] = 0. \tag{5}$$

The spatial and peer effects literatures have suggested to construct h^r and A^r from functions of M_t and z_t^1. When M_t is exogenous, similar ideas, explored in more detail below, can be applied in our setting. When M_t is potentially endogenous, these ideas need to be modified. For more detail, assume that M_t is generated as

$$M_t = M_t(\tau_t^o, \upsilon_t^o, \mu, \nu) \tag{6}$$

where $M_t(.)$ is an unknown function, τ_t is a matrix of strictly exogenous variables which may partially overlap with those in z_t^1, and $\tau_t^o = [\tau_1, \ldots, \tau_t]$. Unobserved innovations are collected in a matrix υ_t, and $\upsilon_t^o = [\upsilon_1, \ldots, \upsilon_t]$. Finally, ν is a vector $\nu = [\nu_1, \ldots, \nu_n]'$ of further unobserved unit specific effects for the network formation process. We assume for our example that (u_t, υ_t) are i.i.d. in t. When υ_t and/or ν are dependent with u_t we refer to M_t as endogenous. In this case we may think of ζ_t to contain the exogenous variables τ_t of the network formation process (or the subset of strictly exogenous variables not already included in z_t^1). When υ_t and ν are independent of u_t we refer to M_t as exogenous. In this case we may think of the matrix ζ_t to contain τ_t (or the subset of strictly exogenous variables not already included in z_t^1) as well as υ_t and ν, or more conveniently M_t.[5] The case where $M_t = M$ is time invariant corresponds to $\tau_t = \tau$ and $\upsilon_t = \upsilon$. All variables are allowed to vary with the cross-sectional sample size n, although we suppress this dependence for notational convenience. When M_t is endogenous we propose to predict M_t with $M_t^* = M_t^*(\tau_t^o)$ in the

[5]Suppose the mean of u_t conditional on $\{z_s^1, \tau_s, \upsilon_s\}_{s=1}^T, \mu, \nu$ is zero, then by iterated expectations so is the mean of u_t conditional on $\{z_s^1, M_s\}_{s=1}^T, \mu$. Consequently, when M_t is exogenous, M_t is measurable w.r.t. $z_1, \ldots, z_t$ and μ under either interpretation of ζ_t. Also note that exogeneity is defined w.r.t. u_t, while M_t may be endogenous w.r.t. ε_t, because it could depend on μ.

construction of instruments. The choice of the function $M_t^*(.)$ may be motivated from a specific network formation model as discussed below, or be more empirically oriented as is typical for reduced form IV approaches.

2.1 Estimator

We now discuss in more detail how to construct the estimator and how to select instruments h^r and A^r. To keep the presentation of the example simple, we take $\sigma^2 = 1$, and defer the discussion of the general case to Section 3. Let $u_1^+(\delta) = y_1^+ - W_1^+\delta$ denote the vector of transformed model errors, and let $\overline{m}_{n,\mathfrak{l}}(\delta) = n^{-1/2}\left[h^{1\prime}u_1^+(\delta), ..., h^{p\prime}u_1^+(\delta)\right]'$, leading to the linear moment conditions $E\left[\overline{m}_{n,\mathfrak{l}}(\delta_0)\right] = 0$. Similarly, let $\overline{m}_{n,\mathfrak{q}}(\delta) = n^{-1/2}\left[u_1^+(\delta)'A^1u_1^+(\delta), ..., u_1^+(\delta)'A^qu_1^+(\delta)\right]'$, with the corresponding quadratic moment conditions $E\left[\overline{m}_{n,\mathfrak{q}}(\delta_0)\right] = 0$. The linear and quadratic moment functions can be stacked as $\overline{m}_n(\delta) = [\overline{m}_{n,\mathfrak{l}}(\delta)', \overline{m}_{n,\mathfrak{q}}(\delta)']'$ and the moment conditions written more compactly as

$$E\left[\overline{m}_n(\delta_0)\right] = 0. \tag{7}$$

The generalized Helmert transformation greatly simplifies the correlation structure between linear and quadratic moments, as compared to other linear transformations used in the literature to eliminate the fixed effects, e.g., by subtracting unit sample averages over time. We exploit these simplifications to set up the GMM criterion function, which as a result conveniently decomposes into a sum of two components, one based on linear moments and one based on quadratic moments.

Let $V_n^h = n^{-1}\sum_{i=1}^n h_i'h_i$ with $h_i = [h_i^1, \ldots, h_i^p]$ and $V_n^a = n^{-1}\sum_{i=1}^n\sum_{j=1}^n a_{ij}a_{ij}'$ with $a_{ij} = [a_{ij}^1, \ldots, a_{ij}^q]$ where a_{ij}^r is the ij-th element of the instrument matrix A^r. Exploiting the orthogonality of the elements of u_1^+ and that $a_{ii}^r = 0$, it can be shown that $E\left[\overline{m}_n(\delta_0)\overline{m}_n(\delta_0)' \mid z_1, z_2, \mu\right] = \tilde{\Xi}_n^{-1}$ where $\tilde{\Xi}_n = \operatorname{diag}[(V_n^h)^{-1}, (2V_n^a)^{-1}]$. The GMM estimator for δ_0 is defined as the minimand of $n^{-1}\overline{m}_n(\delta)'\tilde{\Xi}_n\overline{m}_n(\delta)$ and can be represented as

$$\tilde{\delta}_n = \arg\min_{\delta\in\underline{\Theta}_\delta} n^{-1}\left[\overline{m}_{n,\mathfrak{l}}(\delta)'\left(V_n^h\right)^{-1}\overline{m}_{n,\mathfrak{l}}(\delta) + \overline{m}_{n,\mathfrak{q}}(\delta)'\left(2V_n^a\right)^{-1}\overline{m}_{n,\mathfrak{q}}(\delta)\right], \tag{8}$$

where $\underline{\Theta}_\delta$ is a compact set.

We next explore explicit choices for h^r and A^r, and discuss how, in line with the spatial literature, the structure of Model 1 can be exploited towards finding additional instruments. We first consider the case where $M_t = M$ is time invariant and exogenous. As discussed, when M_t is exogenous it is convenient to think of M_t as being part of ζ_t and thus of $z_t = [z_t^1, \zeta_t]$. Consequently $E\left[M^s z_t^1 u_1^+\right] = 0$ for $s = 0, 1, \ldots$. Observe that the reduced form of y_t is $y_t = (I - \lambda M_t)^{-1}[Z_t\beta + \varepsilon_t]$. For exogenous time invariant $M_t = M$ the reduced form of the quasi-differenced model (3) is given by

$$y_1^+ = (I - \lambda M)^{-1}[Z_1^+\beta + u_1^+], \tag{9}$$

because in this case $\overline{y}_1^+ = My_1^+$. Using (9) and assuming $\|\lambda M\| < 1$ we have

$$E\left[My_1^+ \mid z_1, z_2, \mu\right] = M(I - \lambda M)^{-1}Z_1^+\beta = \sum_{s=0}^{\infty} \lambda^s M^{s+1} Z_1^+\beta.$$

From this we see that the optimal instrument for My_1^+ is a nonlinear function of unknown parameters and $M^s z_t^1$, $s = 0, 1, \ldots$. This suggests that the set of instruments h^r, $r = 1, \ldots, p$ can be taken to correspond to the the linearly independent columns of $\{M^s z_t^1,\ s = 0, 1, \ldots\}$ with $t = 1, 2$. This set can be viewed as providing an approximation to the optimal instruments. Kelejian and Prucha (1998,1999) make a corresponding observation within the context of a spatial cross sectional model and suggested the use of higher order spatial lags of the exogenous variables as additional instruments.

From the reduced form it further follows that

$$VC\left[y_1^+ \mid z_1, z_2, \mu\right] = (I - \lambda M)^{-1}(I - \lambda M')^{-1} = \sum_{s=0}^{\infty}\sum_{\tau=0}^{\infty} \lambda^{s+\tau} M^s M'^{\tau}.$$

As in the spatial literature, and also motivated by an inspection of the score of the Gaussian log-likelihood function, this suggests that the A^r, $r = 1, \ldots, q$ can be chosen from the set $\{M^s M^{\tau\prime} - diag(M^s M^{\tau\prime}),\ s, \tau = 0, 1...\}$. Without loss of generality we can work with symmetrized versions of these matrices, with $(M + M')/2$ and $MM' - diag(MM')$ as leading selections.

In the case where M_t is time varying arguments analogous to those above suggest that the instruments h^r and the matrices A^r can be chosen from $\{M_t^s z_t^1,\ s = 0, 1, \ldots; t = 1, 2\}$ and $\{M_t^s M_t^{\tau\prime} - diag(M_t^s M_t^{\tau\prime}), s, \tau = 0, 1...; t = 1, 2\}$. In the case where M_t is endogenous it is convenient to think of the exogenous variables τ_t, which affect network formation, to be included in ζ_t and thus in $z_t = [z_t^1, \zeta_t]$. In this case we can replace M_t in the above expressions with projections on z_1, z_2. We discuss possible practical choices in the next section where the context of an explicit network formation model makes it easier to give specific recommendations. Our general setup allows for situations where $E\left[u_{it}|z_1, \ldots, z_T, \upsilon_1, \ldots, \upsilon_{t-1}\mu, \nu\right] = 0$. In this situation and $T > 2$ a further simple alternative would be to replace M_t by M_{t-1} in the above expressions.

2.2 Network Formation

Explicit assumptions about the network formation process in (6) are not needed for our GMM estimators, especially when M_t is exogenous. Nevertheless, a specific model for (6) may be useful to check the plausibility of high level assumptions or, in case of endogenous M_t, aid in the construction of valid instruments. We illustrate these points by considering the network formation model analyzed by Graham (2016). A growing literature on estimation of network formation models includes Chandrasekhar (2016), de Paula (2017), Graham (2016), Leung (2016), Ridder and Sheng (2016) and Sheng (2016). However, our focus is on developing a GMM estimator for the parameters δ that is robust to the network formation process, rather than on the estimation of the network formation process itself.

We start our discussion of a specific model for (6) by assuming that we observe relationships between individuals through the indicator variable $d_{ij,t}$, where $d_{ij,t} = 1$ if individuals i and j are related in period t, and $d_{ij,t} = 0$ otherwise. Let $\sum_{j=1}^{n} d_{ij,t} = n_{i,t}$ be the number of relationships of i in period t and define the $n \times n$ matrix $M_t = (m_{ij,t})$ with $m_{ij,t} = d_{ij,t}/n_{i,t}$. Assume that the adjacency matrix $D_t = (d_{ij,t})$ is formed by a dynamic network formation model as in Graham (2016). Let $\psi_{ij} = \psi_{ji} = \nu_i + \nu_j + \alpha_\mu |\mu_i - \mu_j|$ be the utility from matching on unobserved characteristics ν_i and μ_i and define a link at time $t = 1$ as

$$d_{ij,1} = 1\{\alpha_0 + \alpha_\tau |\tau_i - \tau_j| + \psi_{ij} + \upsilon_{ij,1} > 0\} 1\{s_{ij} \leq c\} \tag{10}$$

where $s_{ij} = s_{ji}$ is a measure of "distance" between i and j, and c is a finite constant. For simplicity the covariates τ_i are taken to be time invariant scalars. A simple example for the above model arises in situations where τ_i refers to physical location, $s_{ij} = |\tau_i - \tau_j|$ and individuals only form links if they are in sufficiently close proximity. Another simple example arises in situations where interactions are formed within groups. In this case we define $s_{ij} = |\tau_i - \tau_j|$, where $\tau_i \in \{1, 2, 3...\}$ represents a group index, and $c = 0$. Another example where, say, τ_i in (10) refers to gender, race, income, etc., and interactions are formed within groups can readily be accommodated if τ_i is taken to be multivariate (combined with a trivial relabeling of the variables). Further generalizations to multivariate and time varying τ_{it} are straightforward. More generally, we can model s_{ij} as a function of τ such that $s_{ij} = s_{ij}(\tau)$. To illustrate dynamic network formation we assume that at time $t = 2$ links are formed based both on characteristics and on whether a direct or indirect link existed at time $t = 1$. For this purpose define $\ell_{ij,1} = \sum_{k=1}^{n} d_{ik,1} d_{jk,1}$ as the number of common links between i and j in period 1. Links at time $t = 2$ are then formed according to

$$d_{ij,2} = 1\{\alpha_0 + \alpha_1 d_{ij,1} + \alpha_2 \ell_{ij,1} + \alpha_\tau |\tau_i - \tau_j| + \psi_{ij} + \upsilon_{ij,2} > 0\} 1\{s_{ij} \leq c\}. \tag{11}$$

Endogeneity of $d_{ij,t}$ is now modeled as follows. Let $\upsilon_{ij,t} = \tilde{\upsilon}_{ij,t} + \epsilon_{ijt}$ where $\tilde{\upsilon}_{ij,t} = \tilde{\upsilon}_{ji,t}$ is correlated with u_{it} and u_{jt}, and the ϵ_{ijt} are time varying link specific shocks. Assume that $\tilde{\upsilon}_t = (\tilde{\upsilon}_{ij,t})$ and $\epsilon_t = (\epsilon_{ijt})$ are i.i.d. over time, independent of each other, and independent of τ,μ, ν. Furthermore, the elements of ϵ_t are i.i.d., independent of u_t and follow a logistic distribution. Given this setup, $\upsilon_{ij,t}$ contemporaneously depends on u_t through $\tilde{\upsilon}_{ij,t}$.

We deal with the endogeneity of $d_{ij,t}$ by replacing them with predictors that are based on functions $d^*_{ij,t}(\tau)$ of the exogenous variables τ. A search for predictive functions may be motivated by considering the non-parametric reduced forms $E[d_{ij,1}|\tau_i, \tau_j]$ and $E[d_{ij,2}|\tau_i, \tau_j, d_{ij,1}, \ell_{ij,1}]$. Let $\Lambda(a) = \exp(a) / (1 + \exp(a))$ denote the cumulative distribution function of the Logistic distribution and let $c_{ij,1} = \alpha_0 + \alpha_\tau |\tau_i - \tau_j|$, $\alpha_\tau < 0$. It follows that

$$E[d_{ij,1}|\tau_i, \tau_j] = E_{\upsilon_1}[\Lambda(c_{ij,1} + \psi_{ij} + \tilde{\upsilon}_{ij,1})] 1\{s_{ij} \le c\} \tag{12}$$

where for given τ the expectation E_{υ_1} is with respect to the joint distribution of ψ_{ij} and $\tilde{\upsilon}_{ij,1}$.[6] Similarly, one obtains, for $c_{ij,2} = \alpha_0 + \alpha_1 d_{ij,1} + \alpha_2 \ell_{ij,1} + \alpha_\tau |\tau_i - \tau_j|$, that

$$E[d_{ij,2}|\tau_i, \tau_j, d_{ij,1}, \ell_{ij,1}] = E_{\upsilon_2}[\Lambda(c_{ij,2} + \psi_{ij} + \tilde{\upsilon}_{ij,2})] 1\{s_{ij} \le c\} \tag{13}$$

where for given τ the (conditional) expectation E_{υ_2} is with respect to the joint distribution of ψ_{ij} and $\tilde{\upsilon}_{ij,2}$, conditional on $d_{ij,1}, \ell_{ij,1}$. A series expansion of $\Lambda(a)$ around $c_{ij,1}$ under the integral in (12) can be used to obtain candidate predictors for $d_{ij,1}$. A simple approach consists in using only the leading term $\Lambda(c_{ij,1}) 1\{s_{ij} \le c\}$ and by setting $d^*_{ij,1} = \Lambda(c_{ij,1}) 1\{s_{ij} \le c\}$. The case for $d_{ij,2}$ is slightly more complicated. While $d_{ij,1}$ and $\ell_{ij,1}$ are sequentially exogenous for $d_{ij,2}$, they are not exogenous relative to u_{i1}, which enters the moment condition through the transformed error u_1^+. We therefore replace $c_{ij,2}$ with $c^*_{ij,2} = \alpha_0 + \alpha_1 d^*_{ij,1} + \alpha_2 \ell^*_{ij,1} + \alpha_\tau |\tau_i - \tau_j|$ where $\ell^*_{ij,1} = \sum_{k=1}^n d^*_{ik,1} d^*_{jk,1}$. We then use the predictor $d^*_{ij,2} = \Lambda(c^*_{ij,2}) 1\{s_{ij} \le c\}$. Using the notation $c_{ij,1} = c^*_{ij,1}$, the tail behavior of $d^*_{ij,t}$ is proportional to $\exp\left(-2c^*_{ij,t}\right)$ as $c^*_{ij,t}$ becomes large. This motivates an alternative specification $d^*_{ij,t} = \exp\left(-2(c^*_{ij,t} - \alpha_0)\right) 1\{s_{ij} \le c\}$. To accommodate that the "α parameters" are unknown, we can simply use

$$d^*_{ij,1} = \exp(-\varkappa |\tau_i - \tau_j|) 1\{s_{ij} \le c\}, \tag{14}$$

$$d^*_{ij,2} = \exp(-\varkappa |\tau_i - \tau_j|) \Lambda\left(\varkappa_d d^*_{ij,1}\right) \Lambda\left(\varkappa_\ell \ell^*_{ij,1}\right) 1\{s_{ij} \le c\}. \tag{15}$$

with some non-negative "$\varkappa$ parameters" chosen by the econometrician. Another possibility is to define $d^*_{ij,2} = d^*_{ij,1}$, which may be attractive in situations where M_t varies slowly over time. In this case we could, instead, specify $d^*_{ij,1} = d^*_{ij,2} = 1\{|\tau_i - \tau_j| \le c_\xi\} 1\{s_{ij} \le c\}$. The tuning parameters $\varkappa$ and c_ξ can be obtained, for example, by splitting the sample into two

[6] Since $d_{ij,1}$ does not directly depend on u_1 it is enough to integrate over the marginal distribution of $\psi_{ij} + \tilde{\upsilon}_{ij,1}$. The LHS of (12) is, for a given marginal distribution, invariant to the joint distribution of $\psi_{ij} + \tilde{\upsilon}_{ij,1}$ and u_1.

parts and fitting a parametric model for $d_{ij,t}$ on the first part. If data at $t = 0$ are available, then using that time period to estimate the tuning parameters is a natural choice.[7]

Using either of these predictors, we set M_t^* with typical element $m_{ij,t}^* = d_{ij,t}^*/n_{i,t}^*$ where $n_{i,t}^* = \sum_{j=1}^n d_{ij,t}^*$. Instrument vectors h^r and matrices A^r can now be constructed as discussed above, but with M_t replaced by M_t^*. In panel models with $T > 2$, M_{t-1} is sequentially exogenous for u_t^+ and correlated with M_t and M_{t+1}. In this scenario M_t and M_{t+1} can be replaced with M_{t-1} in the formulations for h^r and A^r. Alternatively, if $\tilde{v}_{ij,t}$ only depends on lagged u_{is} for $s < t$, then M_t is sequentially exogenous and can be used to form instruments.

The above discussion is intended to illustrate how a parametric model for $d_{ij,t}$ may be useful in the construction of possible instruments. However, it is important to stress that such a model is by no means required. A more empirically oriented approach of finding exogenous variables with good predictive power for $d_{ij,t}$ may work just as well.

2.3 Identification and Regularity Conditions

We now discuss high level conditions for identification and give an empirical criterion that can be used to assess identification based on linear and quadratic moments. We then show that the network formation example given in Section 2.2 satisfies regularity conditions required for our theoretical results in Section 3. It proves helpful to collect the instruments in the $n \times p$ matrix $H = [h^1, ..., h^p]$ and to observe that $V_n^h = n^{-1}H'H$.

Assumption EX *Let y be generated according to (1), and assume that the instruments h^r and matrices A^r satisfy the conditions stated above. Let $\delta_0 = (\lambda_0, \beta_0')'$ where $\lambda_0 \in \Theta_\lambda$ with $\Theta_\lambda = (-1, 1)$ and $\beta_0 \in \Theta_\beta$ where Θ_β is an open and bounded subset of $\mathbb{R}^{p_z}$. Furthermore assume that*
(i) $n^{-1}H'u_1^+ = o_p(1)$, $n^{-1}u_1^{+\prime}A^r u_1^+ = o_p(1)$,
(ii) $\operatorname{plim} n^{-1}H'\bar{y}_1^+ = \Gamma_{HMy}$, $\operatorname{plim} n^{-1}H'Z_1^+ = \Gamma_{HZ}$, $\operatorname{plim} n^{-1}W_1^{+\prime}A^r u_1^+ = \Gamma_{WA_r u}$, *and* $\operatorname{plim} n^{-1}W_1^{+\prime}A^r W_1^+ = \Gamma_{WA_r W}$ *are finite for all* $r = 1, .., q$,
(iii) $\operatorname{plim} V_n^h = V^h$ *and* $\operatorname{plim} V_n^a = V^a$ *are finite with* V^h *and* V^a *nonsingular.*

The postulated convergence assumptions are at the level typically assumed in a general analysis of M-estimators; see e.g., Amemiya (1985, pp. 110). The assumptions $n^{-1}H'u_1^+ = o_p(1)$, $n^{-1}u_1^{+\prime}A^r u_1^+ = o_p(1)$ are the asymptotic analogue of the orthogonality conditions (5). Let $\Gamma_{HW} = \operatorname{plim} n^{-1}H'W_1^+ \equiv [\Gamma_{HMy}, \Gamma_{HZ}]$, and consider the $q \times 2$ matrices $S = \operatorname{plim}$

[7] Other possibilities for selecting nuisance parameters include fitting of a model for $d_{ij,t}$ on the estimation sample or cross-validation. A theoretical analysis of these methods is beyond the scope of the paper.

S_n with

$$S_{r,n} = n^{-1}\left[y_1^{+\prime}Q_H^{\prime}A^rQ_H\bar{y}_1^{+}, \bar{y}_1^{+\prime}Q_H^{\prime}A^rQ_H\bar{y}_1^{+}\right]$$

and $S_n = \left[S_{1,n}^{\prime}, ..., S_{q,n}^{\prime}\right]^{\prime}$ where $Q_H = I - Z_1^{+}(Z_1^{+\prime}P_HZ_1^{+})^{-1}Z_1^{+\prime}P_H$ with $P_H = H\left(H^{\prime}H\right)^{-1}H^{\prime}$. The following lemma establishes conditions for identification irrespective of whether M_t is endogenous or exogenous.

Lemma EX1 *Let Assumption EX hold. Then,*
i) if Γ_{HW} *has full column rank, then* $\text{plim}\, n^{-1/2}m_{n,\mathfrak{l}}(\delta) = 0$ *has a unique solution at* $\delta = \delta_0$*, and the parameters are identifiable from the linear moment condition alone.*
ii) if Γ_{HW} *does not have full column rank, but* Γ_{HZ} *and* S *have full column rank, then* $\text{plim}\, n^{-1/2}m_n(\delta) = 0$ *has a unique solution at* $\delta = \delta_0$ *and the parameters are identifiable from the linear and quadratic moment conditions.*

Part (i) of the lemma assumes that Γ_{HW} has full column rank. This condition is maintained in Kelejian and Prucha (1998), and subsequent papers on instrumental variable estimators for spatial network models. If Γ_{HZ} has full column rank, this condition is equivalent to postulating that Γ_{HMy} is not collinear with Γ_{HZ}.

Part (ii) shows that by utilizing the quadratic moment conditions identification is still possible even if Γ_{HW} does not have full column rank. We maintain that Γ_{HZ} has full column rank, which is a standard instrument relevance condition typically imposed in IV settings. Given that Γ_{HZ} has full column rank we have $\Gamma_{HMy} = \Gamma_{HZ}c$ for some vector c. This arises for example if $\bar{y}_1^{+}$ is collinear with Z_1^{+}.

Our adopted data transformation has the advantage that the objective function of the GMM estimator given by (8) is additive in the parts involving the linear and quadratic moment conditions. Given this structure we show in the proof of the lemma that asymptotically all solutions of the linear moment conditions are described by the relation $\beta(\lambda) - \beta_0 = -c(\lambda - \lambda_0)$. Substitution of this expression for $\beta(\lambda)$ into the quadratic moment conditions yields

$$\text{plim}\, n^{-1/2}\overline{m}_{n,\mathfrak{q}}(\lambda, \beta(\lambda)) = S\begin{bmatrix} 1/2 & 0 \\ \lambda_0 & 1 \end{bmatrix}^{-1}[\lambda - \lambda_0, (\lambda - \lambda_0)^2]^{\prime}. \tag{16}$$

Equations (16) have a unique solution at $\lambda = \lambda_0$ if S has full column rank. This in turn implies that linear and quadratic moment conditions have a unique solution at $\delta = \delta_0$; see Lee (2007, pp. 493) for a corresponding discussion of a cross sectional spatial model. In an application it may be convenient to check this condition by checking on the non-singularity of $S_n^{\prime}S_n$. A necessary condition for S_n to have full column rank is that y_1^{+} and $\bar{y}_1^{+}$ do not lie in the space spanned by Z_1^{+}. This condition is likely satisfied since the reduced form (9) depends on both Z_1^{+} and u_1^{+}.

Assumption EX postulates that $n^{-1}h^{r\prime}u_1^+ = o_p(1)$, $n^{-1}u_1^{+\prime}A^r u_1^+ = o_p(1)$. The next lemma implies these assumptions from lower level conditions that can be imposed on the model in Section 2.2. The lemma also provides specific choices of h^r and A^r for which these conditions are satisfied.

Lemma EX2 *Suppose the network is generated by (10) and (11), and suppose Assumption EX holds, except for postulating that $n^{-1}h^{r\prime}u_1^+ = o_p(1)$ and $n^{-1}u_1^{+\prime}A^r u_1^+ = o_p(1)$ holds. Then the following statements are true for all $i = 1, ..., n$ and $n \geq 1$, with bounding constants K, K_h, K_a, K_z that do not depend on i, j, n or t:*
(a) A sufficient condition for $n^{-1}h^{r\prime}u_1^+ = o_p(1)$ and $n^{-1}u_1^{+\prime}A^r u_1^+ = o_p(1)$ to hold is that $\|h_{ir}\|_{2+\delta} \leq K_h < \infty$ for some $\delta > 0$, and $\sum_{j=1}^n \left|a_{ij}^r\right| \leq K_a < \infty$.
(b) Suppose that $\sum_{l=1}^n d_{il,t} \geq 1$, $s_{ij} = s_{ji}$ and
(i) $\sum_{j=1}^n 1\{s_{ij} \leq c\} \leq K < \infty$,
(ii) $\sum_{j=1}^n (\Pr(s_{ij} \leq c))^{1/[s(2+\delta)]} \leq K < \infty$, $\left\|z_t^1\right\|_{4+\delta} \leq K_z < \infty$ for some $\delta > 0$ and some $s = 1, 2, ...$, and the instruments h^r are taken from $\{M_t^\tau z_t^1,\ \tau = 0, \ldots, s\}$ and the matrices A^r are of the form $A^r = (\underline{A}^r + \underline{A}^{r\prime})/2$ with $\underline{A}^r$ taken from $\{M_t^{\tau-\sigma}M_t^{\sigma\prime} - diag(M_t^{\tau-\sigma}M_t^{\sigma\prime})$, $0 \leq \sigma \leq \tau, \tau = 1, \ldots, s\}$ with $t = 1, 2$. Then the sufficient conditions in (a) are satisfied. Furthermore, for some finite K_a we have $\sum_{j=1}^n \left\|a_{ij}^r\right\|_{2+\delta} \leq K_a$.

Part (b) of the lemma shows that for our exemplary network model the specific selection for h^r and A^r satisfy the properties postulated for our general model; cp. Assumption 2(i),(ii) in Appendix A. As shown in the proof of the lemma in the supplemental appendix, the condition in (b)(ii) that $\sum_{j=1}^n (\Pr(s_{ij} \leq c))^{1/[s(2+\delta)]} \leq K$ is implied by the stronger condition $\sum_{j=1}^n 1\{\Pr(s_{ij} \leq c) > 0\} \leq K$. If $\Pr(s_{ij} \leq c) = 0$ implies $1\{s_{ij} \leq c\} = 0$ then (b)(i) and (b)(ii) can be replaced with $\sum_{j=1}^n 1\{\Pr(s_{ij} \leq c) > 0\} \leq K$. The summability condition in (b) allows for all individuals in the network to potentially be connected, albeit with small probability for most connections, while the stronger condition rules out most connections with probability one.

A computational algorithm to obtain consistent starting values using both linear and quadratic moment conditions is based on partialling out the term $Z_t\beta$ using the linear moment conditions only. This is possible because β is identified by the linear moment conditions for any fixed value of λ. Let $\hat{\beta}(\lambda) = \left(Z_1^{+\prime}P_H Z_1^+\right)^{-1} Z_1^{+\prime}P_H\left(y_1^+ - \lambda\bar{y}_1^+\right)$ be the 2SLS estimator of a linear IV regression of $\left(y_1^+ - \lambda\bar{y}_1^+\right)$ on Z_1^+ using instruments H and set $\tilde{\delta}_n(\lambda) = \left(\lambda, \hat{\beta}(\lambda)'\right)$. The second step consists in substituting $\tilde{\delta}_n(\lambda)$ into the quadratic moment conditions and in minimizing the quadratic part of the moment function. The algorithm can be summarized as follows:

Algorithm EX *Let* $\overline{m}_n(\delta), \hat{\beta}_z(\lambda)$ *and* $\tilde{\delta}_n(\lambda)$ *be as defined before. Let* $m_{n,q,r}\left(\tilde{\delta}_n(\lambda)\right) = u_1^+(\delta)' A^r u_1^+(\delta)$
(1) Solve the problem $\tilde{\lambda} = \arg\min_\lambda n^{-1} \overline{m}_{n,\mathfrak{q}}\left(\tilde{\delta}_n(\lambda)\right)' (V_n^a)^{-1} \overline{m}_{n,\mathfrak{q}}\left(\tilde{\delta}_n(\lambda)\right)$.
(2) Set the starting values to $\hat{\lambda} = \tilde{\lambda}$, $\hat{\beta}_z = \hat{\beta}_z\left(\hat{\lambda}\right)$.

When Assumption EX holds it follows from (16) that $m_{n,\mathfrak{q}}\left(\tilde{\delta}_n(\lambda)\right) = 2(\lambda_0 - \lambda)\gamma_b + (\lambda_0 - \lambda)^2 \gamma_c + o_p(1)$ where γ_b and γ_c are constant vectors. In large samples $m_{n,\mathfrak{q}}\left(\tilde{\delta}_n(\lambda)\right) = 0$ has only one solution if S has full column rank. As a result, Algorithm EX provides starting values that are consistent estimates asymptotically. Using starting values obtained from Algorithm EX in a subsequent full optimization step as in (8) leads to parameter estimates that have the limiting distributions derived in Section 3.

2.4 Monte Carlo

We conduct a Monte Carlo experiment with data generated from (1) with $Z_t = \left[z_t^1, M_t z_t^1\right]$, $T = 2$, $f_1 = f_2 = 1$ and networks formed according to (10) and (11). In our first design M_t is exogenous w.r.t. $\varepsilon_t^+ = u_t^+$. We set $p_z = 2$ and draw μ_i, u_{it}, ν_i and z_{it}^1 mutually independently from standard Gaussian distributions, while $\upsilon_{ij,t} = \upsilon_{ji,t}$ is drawn independently from a logistic distribution. The location characteristics τ_i are drawn independently from uniform distributions with heterogeneous means, $\tau_i \sim U[i, i+2]$, and $s_{ij} = 1\{|\tau_i - \tau_j| < 10\}$. We set $\alpha_0 = 1, \alpha_\tau = -1, \beta_1 = 1$ and $\alpha_\mu = -.1$. We vary λ in $\{.1, .5, .7\}$ and set $\beta_2 = -(\lambda + \Delta)\beta_1$ where Δ takes values in $\{.1, .5, 1\}$. Linear instruments are $H = \left[z_1^1, z_2^1, M_1 z_1^1, M_2 z_2^1, M_1^2 z_1^1, M_2^2 z_2^1, M_1^3 z_1^1, M_2^3 z_2^1\right]$, and quadratic moment conditions are formed with $A^1 = (M_1 + M_1')/2$, $A^2 = (M_2 + M_2')/2$, $A^3 = M_1' M_1 - \mathrm{diag}(M_1' M_1)$ and $A^4 = M_2' M_2 - \mathrm{diag}(M_2' M_2)$. As shown in Bramoulle, Djebbari and Fortin (2009) and de Paula (2017) the model is not identified by linear moment conditions if $\beta_2 = -\lambda\beta_1$, which is consistent with a failure of a general condition for identification by linear moment restrictions given in Kelejian and Prucha (1998). Our Monte Carlo design thus approaches the point of non-identification for linear IV as Δ shrinks towards zero. We consider sample sizes of $n = 250$ and $n = 500$ for all designs. Table 1 reports results for the estimator of λ using conventional OLS of y_1^+ on W_1^+, two stage least squares (2SLS) of y_1^+ on W_1^+ using H as instruments and our linear-quadratic GMM (GMM) estimator defined in (8). We use Algorithm EX to find starting values, followed by a full optimization step over the entire criterion function. The computational complexity of minimizing the linear-quadratic criterion of GMM is essentially independent of the cross-sectional sample size n.[8] For $\lambda = .1$

[8]Matlab replication code for the simulations is available from the authors.

endogeneity is relatively mild leading to OLS being reasonably unbiased, at least in absolute terms. As λ increases to .5 and .7, OLS becomes seriously biased. The 2SLS estimator performs well when $\Delta = 1$, although biases exist in the small sample case where $n = 250$. As the sample size increases to $n = 500$ the bias considerably drops and the Mean Absolute Error (MAE) significantly improves. However, as Δ moves towards .1 the performance of linear IV starts to rapidly deteriorate even in the large sample design with $n = 500$. This first manifests itself in elevated MAE's and, as $\Delta = .1$, in severely biased estimates and large MAE values. GMM on the other hand shows very robust performance across all designs and clearly dominates all estimators in both sample sizes and for all parameter values. It is essentially unbiased even when $n = 250$, with a percentage median bias of 5% or less when $\lambda > .1$ and around 10% median bias for $\lambda = .1$. For the larger sample size the bias further drops and is substantially smaller than the bias of the other two estimators. The MAE is significantly smaller for GMM than either for OLS or 2SLS in all designs and for both sample sizes.

We also consider a design where M_t is endogenous w.r.t. $\varepsilon_t^+ = u_t^+$. We generate $\upsilon_{ij} = \upsilon_{ji}$ by setting $\upsilon_{ij,t} = (u_{it} + u_{jt})/2 + \epsilon_{ij,t}$ where $\epsilon_{ij,t}$ is independent logistic. All other parameters are the same as in the case where M_t is exogenous. We predict the endogenous M_t with M_t^* using the functional forms in (14) and (15). The parameters for the prediction are set at $\chi = .75, \chi_d = 1$, $\chi_\ell = 1$ and $c = 5$. Linear instruments and quadratic instruments are formed as in the case with exogenous M_t, except that in H and A^j the matrix M_t is replaced with M_t^*. Simulation results are reported in Table 2. The OLS estimator is somewhat more biased than in the case of exogenous M_t with a corresponding increase in the MAE. The 2SLS estimator now is significantly more biased than in the exogenous M_t case. The MAE of 2SLS is accordingly significantly inflated. GMM is somewhat more biased than in the case with exogenous M_t. It is much less biased than OLS in all designs and also much less biased than 2SLS. The MAE of the GMM estimator rises somewhat as λ increases but overall is very insensitive to Δ. At $n = 250$ it clearly dominates OLS both in terms of bias and MAE except when $\lambda = .1$ and $\Delta = 1$. It also dominates 2SLS in terms of MAE across all parametrization of the model. When $n = 500$ the GMM estimator dominates OLS clearly across the entire parameter space. The 2SLS estimator continues to do poorly except when $\Delta = 1$ and $\lambda = .7$. GMM on the other hand does well across the entire parameter space with low bias and MAE that is not very sensitive to the DGP. Overall GMM clearly dominates 2SLS when M_t is endogenous and approximated by M_t^*.

3 The General Model

3.1 Specification

We consider a fairly general panel data model, which covers the example in Section 2 as a special case. In addition, it allows for higher order and time dependent spatial lags, weakly exogenous covariates and unobserved common factors that we treat as unknown parameters. Let $\{y_t, x_t, z_t\}_{t=1}^T$ be a panel data set defined on a common probability space $(\Omega, \mathcal{F}, P)$, where $y_t = [y_{1t}, ..., y_{nt}]'$, $x_t = [x'_{1t}, ..., x'_{nt}]'$, and $z_t = [z'_{1t}, ..., z'_{nt}]'$ are of dimension $n \times 1$, $n \times k_x$ and $n \times k_z$. The dynamic and cross sectionally dependent panel data model we consider can then be written as

$$\begin{aligned} y_t &= \textstyle\sum_{p=1}^P \lambda_p M_{p,t} y_t + Z_t \beta + \varepsilon_t = W_t \delta + \varepsilon_t, \\ \varepsilon_t &= \textstyle\sum_{q=1}^Q \rho_q \underline{M}_{q,t} \varepsilon_t + \mu f_t + u_t, \end{aligned} \tag{17}$$

where Z_t is a $n \times k$ matrix composed of columns of $x_t^1, z_t^1, M_{1,t} x_t^1, M_{1,t} z_t^1, \ldots, M_{P,t} x_t^1, M_{P,t} z_t^1$ and a finite number of time lags thereof, $W_t = [M_{1,t} y_t, \ldots, M_{P,t} y_t, Z_t]$ and $\delta = [\lambda', \beta']'$ are the parameters of interest. As for the exemplary model discussed in the previous section $z_t = [z_t^1, \zeta_t]$ is a matrix of k_z strictly exogenous variables, where z_t^1 denotes the strictly exogenous variables in the regression, and ζ_t denotes additional strictly exogenous variables which may affect network formation, and where the dimension of ζ_t may depend on n. In addition we now also include k_x weakly exogenous covariates $x_t = [x_t^1, \xi_t]$, which we partition in an analogous manner. 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. Consistent with the exemplary social interaction models discussed in the previous section, we allow for network interdependencies in the form of "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, where spatial lags of predetermined variables should be viewed as being included in x_{it}. The $n \times n$ spatial weight matrices are denoted as $M_{p,t} = (m_{p,ijt})$ and $\underline{M}_{q,t} = (\underline{m}_{q,ijt})$. We do assume that the matrices $M_{p,t}$ and $\underline{M}_{q,t}$ are known or observed in the data. As a normalization we take $m_{p,iit} = \underline{m}_{q,iit} = 0$.

Alternatively or concurrently, we allow in each period t for the regressors and disturbances to be affected by common shocks. As in Andrews (2005) and Kuersteiner and Prucha (2013), those common shocks are captured by a sigma field, say, $\mathcal{C}_t \subset \mathcal{F}$, but are otherwise left unspecified. 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 $\mathcal{C}_t = \mathcal{C} = \{\emptyset, \Omega\}$.

We also allow for interactive effects in the error term where μ 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 λ and ρ be, respectively, P and Q dimensional vectors of parameters with typical elements λ_p and ρ_q.

Note that (17) 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. In line with the literature on spatial networks we refer to those weighted averages as spatial lags, and to the corresponding parameters as spatial autoregressive parameters.[9] 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 ε_t.[10] In fact, and in contrast to most of the recent literature discussed in the introduction on models with endogenous spatial weights, we do not impose any particular restrictions on how the weights are generated.

For $i = 1, ..., n$ let $z_i^o = (z_{i1}, \ldots, z_{iT})$, $x_{it}^o = [x_{i1}, \ldots, x_{it}]$, $u_{it}^o = [u_{i1}, \ldots, u_{it}]$, $u_{-i,t} = [u_{i1}, \ldots, u_{i-1,t}, u_{i+1,t}, ...u_{nt}]$. 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 than one), and define the sigma fields*

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

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 μ_i exist, and the

[9] An alternative specification, analogous to specifications considered in Baltagi et al (2008), would be to model the disturbance process in (17) as $\varepsilon_t = \phi f_t + v_t$, where ϕ and v_t follow possibly different spatial autoregressive processes. Since we are not making any assumptions on the unobserved components μ it is readily seen that the above specification includes this case, provided that the spatial weights do not depend on t.

[10] It is for this reason that we list spatial lags of x_t and z_t separately in defining the regressors in Z_t. If the $M_{p,t}$ are strictly exogenous we can incorporate those spatial lags w.o.l.o.g. into x_t and z_t. The matrix Z_t may also contain additional endogenous variables, apart from the spatial lags in y_t. We do not explicitly list those variables for notational simplicity.

moments are uniformly bounded by a generic constant K.

(ii) The following conditional moment restrictions hold for some constant $c_u > 0$:

$$E\left[u_{it}|\mathcal{B}_{n,i,t} \vee \mathcal{C}\right] = 0, \tag{18}$$

$$E\left[u_{it}^2|\mathcal{B}_{n,i,t} \vee \mathcal{C}\right] = \sigma_t^2 \varrho_i^2 \quad \textit{with} \quad \sigma_t^2, \varrho_i^2 \geq c_u, \tag{19}$$

$$E\left[\left|u_{it}\right|^{2+\delta}|\mathcal{B}_{n,i,t} \vee \mathcal{C}\right] \leq K_u. \tag{20}$$

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 γ_ϱ and are assumed to be measurable w.r.t. $\mathcal{Z}_n \vee \mathcal{C}$.

Condition (18) clarifies the distinction between weakly exogenous covariates x_{it} and strictly exogenous covariates z_{it}. The latter enter the conditioning set at all leads and lags. The conditioning sets $\mathcal{B}_{n,i,t}$ and $\mathcal{B}_{n,t}$ can be expanded to include additional conditioning variables without affecting the analysis. In the following we use the notation $\Sigma_\sigma = \operatorname{diag}(\sigma_t^2)$ and $\Sigma_\varrho = \operatorname{diag}(\varrho_i^2)$. As a normalization we may take $\sigma_T^2 = 1$ or $n^{-1}\operatorname{tr}(\Sigma_\varrho) = 1$. Specifications where σ_t^2 and ϱ_i^2 are non-stochastic, and specifications where the u_{it} are conditionally homoskedastic are covered as special cases.

In addition to Assumption 1 we maintain Assumptions 2-7, which are collected in Appendix A for ease of presentation. We note that those assumptions do not maintain that the f_t are non-stochastic, but only maintain that the f_t are measurable w.r.t. $\mathcal{C}$. As a normalization we maintain $f_T = 1$. The unit specific effects μ are left unspecified and are allowed to be correlated with the covariates.

Define $R_t(\lambda) = I_n - \sum_{p=1}^P \lambda_p M_{p,t}$ and $\underline{R}_t(\rho) = I_n - \sum_{q=1}^Q \rho_q \underline{M}_{q,t}$, then the reduced form of the model is given by

$$\begin{aligned} y_t &= R_t(\lambda)^{-1}\left(x_t\beta_x + z_t\beta_z + \varepsilon_t\right), \\ \varepsilon_t &= \underline{R}_t(\rho)^{-1}\left(\mu f_t + u_t\right). \end{aligned} \tag{21}$$

Applying a Cochrane-Orcutt type transformation by premultiplying the first equation in (17) with $\underline{R}_t(\rho)$ yields

$$\underline{R}_t(\rho) y_t = \underline{R}_t(\rho) W_t \delta + \mu f_t + u_t. \tag{22}$$

A further transformation of the spatially Cochrane-Orcutt transformed model (22) is needed to eliminate the unit specific effects μ. In the classical panel literature with $f_t = 1$ the Helmert transformation was proposed by Arellano and Bover (1995) as an alternative forward filter that, unlike differencing, eliminates fixed effects without introducing serial correlation in the linear moment conditions underlying their GMM estimator.[11] Building on this idea we first develop an orthogonal quasi-forward differencing transformation for the more general case where factors f_t appear in the model. More specifically, consider the $T \times 1$ vectors $f = [f_1, \ldots, f_T]$ and $u_i = [u_{i1}, ..., u_{iT}]'$ such that $\eta_i = \mu_i f + u_i$ with typical element $\eta_{it} = \mu_i f_t + u_{it}$. A quasi-forward differencing filter has a representation as an upper triangular $T-1 \times T$ matrix Π with the property $\Pi f = 0$. Let $\pi_t = [0, \ldots, 0, \pi_{tt}, \ldots, \pi_{tT}]$ denote the rows of Π, let $\eta_i^+ = \Pi\eta_i$ and $u_i^+ = \Pi u_i$, then $\eta_i^+ = u_i^+$, and the elements of η_i^+ and u_i^+ can be written as

$$\eta_{it}^+ = \sum\nolimits_{s=t}^{T} \pi_{ts}\eta_{is}, \qquad u_{it}^+ = \sum\nolimits_{s=t}^{T} \pi_{ts}u_{is}. \tag{23}$$

If in addition $\Pi\Sigma_\sigma\Pi' = I$ then under our assumptions the transformed errors u_{it}^+ are uncorrelated across i and t. In Proposition 1 in Appendix B we present a generalization of the Helmert transformation that satisfies these two conditions, and give explicit expressions for the elements $\pi_{ts} = \pi_{ts}(f, \gamma_\sigma)$. Such expressions are crucial from a computational point of view, especially if f_t is estimated as an unobserved parameter. A more detailed discussion, including a discussion of a convenient normalization for the factors and how to handle multiple factors, is given in that appendix and a supplementary appendix. Our moment conditions involve both linear and quadratic forms of the forward differenced disturbances.

3.2 Estimator

For clarity we denote the true parameters of interest θ and the true auxiliary variance parameters γ defined in Assumption 1 as $\theta_0 = (\delta_0', \rho_0', f_0')'$ and $\gamma_0 = \left(\gamma_{0,\varrho}', \gamma_{0,\sigma}'\right)'$. Using (22) we define

$$u_t^+(\theta_0, \gamma_\sigma) = \sum\nolimits_{s=t}^{T} \pi_{ts}\left(f_0, \gamma_\sigma\right) u_s = \sum\nolimits_{s=t}^{T} \pi_{ts}\left(f_0, \gamma_\sigma\right) \underline{R}_s(\rho_0) \left[y_s - W_s\delta_0\right], \tag{24}$$

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

$$u_{*t}^+(\theta_0, \gamma) = \Sigma_\varrho(\gamma_\varrho)^{-1/2} u_t^+(\theta_0, \gamma_\sigma). \tag{25}$$

with $\gamma = \left(\gamma_\varrho', \gamma_\sigma'\right)'$. Considering moment conditions based on $u_{*t}^+(\theta_0, \gamma)$ is sufficiently general to cover initial estimators with $\Sigma_\sigma = I_T$ and $\Sigma_\varrho = I_n$. As illustrated in Section 2 quadratic moment conditions are often required to identify parameters associated with spatial lags

[11] 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).

and may further increase the efficiency of estimators by exploiting spatial correlation in the data generating process.

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 Theorem 1 we have the following linear moment conditions for $t = 1, \ldots, T-1$,

$$E \begin{bmatrix} h_t^{1\prime} u_{*t}^+(\theta_0, \gamma) \\ \vdots \\ h_t^{p_t\prime} u_{*t}^+(\theta_0, \gamma) \end{bmatrix} = E \left[\sum_{i=1}^n h_{it}' u_{*it}^+(\theta_0, \gamma) \right] = 0 \tag{26}$$

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 Theorem 1, and imposing the constraint that $a_{ii,t} = 0$ one obtains the following quadratic moment conditions for $t = 1, \ldots, T-1$,

$$E \begin{bmatrix} u_{*t}^+(\theta_0, \gamma)' A_t^1 u_{*t}^+(\theta_0, \gamma) \\ \vdots \\ u_{*t}^+(\theta_0, \gamma)' A_t^{q_t} u_{*t}^+(\theta_0, \gamma) \end{bmatrix} = E \left[\sum_{i=1}^n \sum_{j=1}^n a_{ij,t}' u_{*it}^+(\theta_0, \gamma) u_{*jt}^+(\theta_0, \gamma) \right] = 0. \tag{27}$$

The requirement that $a_{ii,t} = 0$ is generally needed for (27) 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 redefining both p_t and q_t as $p_t + q_t$, the above moment conditions can be stacked and written more compactly as

$$\begin{aligned} &E\left[\overline{m}_{n,t}(\theta_0, \gamma)\right] = 0, \qquad \text{with} \\ &\overline{m}_{n,t}(\theta, \gamma) = n^{-1/2} \sum_{i=1}^n h_{it}' u_{*it}^+(\theta, \gamma) + n^{-1/2} \sum_{i=1}^n \sum_{j=1}^n a_{ij,t}' u_{*it}^+(\theta, \gamma) u_{*jt}^+(\theta, \gamma). \end{aligned} \tag{28}$$

The example in Section 2 is a special case of $\overline{m}_{n,t}(\theta, \gamma)$ with $t = 1$ where $\overline{m}_{n,1}(\theta, \gamma) = \overline{m}_n(\delta) = [\overline{m}_{n,\mathfrak{l}}(\delta)', \overline{m}_{n,\mathfrak{q}}(\delta)']'$, $h_{i1} = \left[h_i^1, ..., h_i^p, \mathbf{0}_q'\right]'$, $a_{ij,1} = \left[\mathbf{0}_p', a_{ij}^1, ..., a_{ij}^q\right]'$ and $\mathbf{0}_k$ is a $k \times 1$ vector of zeros. The formulation in (28) allows for more general forms of the empirical moment function by allowing for nontrivial linear combinations of (26) and (27) in addition to simply stacking both sets of moments. The particular form of (28) is motivated by a need to minimize cross-sectional and temporal correlations between empirical moments. Theorem 1 below provides for sufficient conditions for the choice of moments, moment weights A_t and forward differences Π that lead to a covariance matrix of the moment vector, which 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 (28) as

$$\overline{m}_n(\theta, \gamma) = \left[\overline{m}_1(\theta, \gamma)', \ldots, \overline{m}_{T-1}(\theta, \gamma)'\right]. \tag{29}$$

For some estimator $\bar{\gamma}_n$ of the auxiliary parameters γ and a $p \times p$ moment weights matrix $\tilde{\Xi}_n$ the GMM estimator for θ_0 is defined as

$$\tilde{\theta}_n\left(\bar{\gamma}_n\right) = \arg\min_{\theta \in \underline{\Theta}_\theta} n^{-1} \overline{m}_n(\theta, \bar{\gamma}_n)' \tilde{\Xi}_n \overline{m}_n(\theta, \bar{\gamma}_n) \tag{30}$$

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

The optimal weight matrix of a GMM estimator based on both linear and quadratic moment conditions depends on the variance covariances of linear quadratic forms based on forward differenced disturbances. Simplifying them as much as possible is critical to the implementation of the estimator. The following proposition establishes sufficient conditions under which significant simplifications can be achieved.

Theorem 1 [12] *Let the information sets $\mathcal{B}_{n,i,t}$, $\mathcal{B}_{n,t}$, $\mathcal{Z}_n$ be as defined in Section 3. Furthermore assume that for all $t = 1, \ldots, T$, $i = 1, \ldots, n$, $n \geq 1$, $E\left[u_{it} | \mathcal{B}_{n,i,t} \vee \mathcal{C}\right] = 0$, $E\left[u_{it}^2 | \mathcal{B}_{n,i,t} \vee \mathcal{C}\right] = \varrho_i^2 \sigma_t^2 > 0$, $E\left[u_{it}^3 | \mathcal{B}_{n,i,t} \vee \mathcal{C}\right] = \mu_{3,it}$, $E\left[u_{it}^4 | \mathcal{B}_{n,i,t} \vee \mathcal{C}\right] = \mu_{4,it}$, where σ_t is finite and measurable w.r.t. $\mathcal{C}$, and ϱ_i, $\mu_{3,it}$ and $\mu_{4,it}$ are finite and measurable w.r.t. $\mathcal{Z}_n \vee \mathcal{C}$. Define $\Sigma_\varrho = \operatorname{diag}\left(\varrho_1^2, ..., \varrho_n^2\right)$ and $\Sigma_\sigma = \operatorname{diag}\left(\sigma_1^2, ..., \sigma_T^2\right)$. 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 $\pi_t = [0, \ldots, 0, \pi_{tt}, \ldots, \pi_{tT}]$ be a $1 \times T$ vector where $\pi_{t\tau}$ is measurable w.r.t. $\mathcal{C}$, and consider the forward differences $u_t^+ = \left[u_{1t}^+, \ldots, u_{nt}^+\right]'$ with $u_{it}^+ = \pi_t u_i'$. Assume that $\operatorname{vec}_D(A_t) = \operatorname{vec}_D(B_t) = 0$, $\Pi f = 0$ and $\Pi \Sigma_\sigma \Pi' = I$. Then*

$$E\left[u_t^{+\prime} A_t u_t^+ + u_t^{+\prime} a_t | \mathcal{C}\right] = 0, \tag{31}$$

$$\begin{aligned} &\operatorname{Cov}(u_t^{+\prime} A_t u_t^+ + a_t' u_t^+, u_t^{+\prime} B_t u_t^+ + b_t' u_t^+ | \mathcal{C}) \\ &\quad = E\left[\operatorname{tr}(A_t \Sigma_\varrho (B_t + B_t') \Sigma_\varrho) | \mathcal{C}\right] + E\left[a_t' \Sigma_\varrho b_t | \mathcal{C}\right], \end{aligned} \tag{32}$$

$$\operatorname{Cov}(u_t^{+\prime} A_t u_t^+ + a_t' u_t^+, u_s^{+\prime} B_s u_s^+ + b_s' u_s^+ | \mathcal{C}) = 0 \qquad \text{for all } t > s. \tag{33}$$

[12] A more general version of the proposition, further details and an explicit proof are given in the supplemental appendix in Section D.2. There we consider generalized forms $u_t^{+\prime} A_t u_t^\times + u_t^{+\prime} a_t$, where u_t^+ and $u_t^\times$ contain, respectively, forward differences corresponding to upper triangular matrices Π and Γ. We show that, in general, linear and quadratic moment conditions are correlated when the condition of the theorem that $\operatorname{vec}_D(A_t) = \operatorname{vec}_D(B_t) = 0$, $\Pi = \Gamma$ with $\Pi f = 0$ and $\Pi \Sigma_\sigma \Pi' = I$ fails. We also show that in that case, variances and covariances depend on additional higher order terms that are difficult to estimate.

The proof shows that a sufficient condition for $E\left[u_t^{+\prime}A_t u_t^{+}+u_t^{+\prime}a_t|\mathcal{C}\right]=0$ is $\mathrm{vec}_D(A_t)=0$ where $\mathrm{vec}_D(A_t)$ is the vector of diagonal elements of A_t. We note that with $\mathrm{vec}_D(A_t)=0$ no restrictions on $E\left[u_{it}^2|\mathcal{B}_{n,i,t}\vee\mathcal{C}\right]$ are necessary to ensure $E\left[u_t^{+\prime}A_t u_t^{+}+u_t^{+\prime}a_t|\mathcal{C}\right]=0$. Since A_t is a quantity chosen by the econometrician, the constraint $\mathrm{vec}_D(A_t)=0$ can easily be imposed and is satisfied for the example discussed in Section 2. Setting $\mathrm{vec}_D(A_t)=\mathrm{vec}_D(B_t)=0$ for all t, and using orthogonally transformed disturbances, ensures that variances and covariances in (32) and (33) do not depend on higher order moments and thus simplifies the optimal GMM weight matrix. In particular, (32) implies that linear and quadratic moments are uncorrelated, while (33) implies that the linear quadratic forms are uncorrelated over time. 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 of Theorem 1 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.

The homoskedastic case where $\Sigma_\varrho=\varrho^2 I$ leads to some further simplifications. In that case a sufficient condition for the validity of moment conditions of the from $E\left[u_t^{+\prime}A_t u_t^{+}+u_t^{+\prime}a_t|\mathcal{C}\right]=0$ is that $\mathrm{tr}(A_t)=0$. 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[13]. However, $\mathrm{tr}(A_t)=0$ does not insure that the linear quadratic forms are uncorrelated across time even in the case of orthogonally transformed disturbances, i.e., $\Pi f=0$ and $\Pi\Sigma_\sigma\Pi'=I$.

3.3 Consistency

Consistent with the assumptions in Appendix A let $\theta_*=\lim_{n\to\infty}\theta_{n,0}$ and $\gamma_*=\lim_{n\to\infty}\gamma_{n,0}$. Furthermore, consider a sequence of estimators of the auxiliary parameters $\bar{\gamma}_n\xrightarrow{p}\bar{\gamma}_*$. The objective function of the GMM estimator $\tilde{\theta}_n\left(\bar{\gamma}_n\right)$ defined in (30) 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 $\mathcal{R}(\theta)=\mathfrak{m}(\theta)'\Xi\mathfrak{m}(\theta)$ with $\mathfrak{m}(\theta)=\mathrm{plim}_{n\to\infty}n^{-1/2}\overline{m}_n(\theta,\bar{\gamma}_*)$. Because $\mathfrak{m}(\theta)$ and Ξ are generally stochastic in the presence of common factors it follows that $\mathcal{R}(\theta)$ and the minimizer θ_* are also generally stochastic. The consistency proof needs to account for the randomness in $\mathcal{R}(\theta)$ and θ_*. The consistency results given below build, in particular, on Gallant and White (1988), White (1984), Newey and McFadden (1994), Pötscher and Prucha (1997, ch 3).[14] We first establish a general result for the consistency of estima-

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

[14]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

tors for situations where the limiting objective function and the minimizers are stochastic, which is given as Proposition 2 in Appendix C. This proposition also extends the notion of identifiable uniqueness to stochastic limit functions and minimizers. We then use this result to proof the following theorem establishing consistency.

Theorem 2 *(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* $n \to \infty$.

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

3.4 Limit Theory

The limiting distribution of our GMM estimators depends on the limiting distribution of the sample moment vector $\overline{m}_n = \overline{m}_n(\theta_0, \gamma_{0,\sigma}, \gamma_\varrho)$ defined by (29), evaluated at the true parameters, except possibly for the specification of the cross sectional variance components ϱ_i^2. The reason for this is to accommodate both leading cases $\varrho_i^2 = \varrho_{0,i}^2$ and $\varrho_i^2 = 1$. Our derivation of the limiting distribution of $\overline{m}_n$ is based on Proposition 3 in Appendix C.

Proposition 3 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:[15] 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, Proposition 3 allows for the presence of common factors which can be handled, because Proposition 3 establishes convergence in distribution $\mathcal{C}$-stably.[16] Third, the theorem covers orthogo-

propositions presented in this section.

[15]See, e.g., Atchad and Cattaneo (2012), Doukhan et al. (1996), Gao and Hong (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., Koroljuk and Borovskich (1994) for a review.

[16]We refer the reader to the working paper version for a detailed discussion of $\mathcal{C}$-stable convergence.

nally transformed variables, and demonstrates how these transformations very significantly simplify the correlation structure between the linear quadratic forms.

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 $G(\theta) = \operatorname{plim}_{n\to\infty} G_n(\theta,\gamma_*)$ as defined in Assumption 6. To establish our results we show that $G(\theta)$ exists, and that $G(\theta)$ is $\mathcal{C}$-measurable for all $\theta \in \underline{\Theta}_\theta$, and continuous in θ. Let $G = G(\theta_*)$ and observe that G is $\mathcal{C}$-measurable, since θ_* 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)*

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

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

$$\Psi = (G'\Xi G)^{-1}G'\Xi V \Xi G(G'\Xi G)^{-1}. \tag{34}$$

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

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

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

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

$$\widetilde{V}_n = \operatorname{diag}_{t=1}^{T-1}\left(V_{t,n}^h + 2V_{t,n}^a\right), \tag{35}$$

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

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

$$\hat{\theta}_n = \arg\min_{\theta\in\underline{\Theta}_\theta} \overline{m}_n(\theta,\tilde{\gamma}_n)'\widetilde{V}_n^{-1}\overline{m}_n(\theta,\tilde{\gamma}_n). \tag{36}$$

Clearly $\widetilde{V}_{(n)}^{-1} \xrightarrow{p} V^{-1}$ by 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 Ψ.

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

$$T_n = \left\| \left(R\hat{\Psi}_n R'\right)^{-1/2} \sqrt{n}(R\hat{\theta}_n - r) \right\|^2 \tag{37}$$

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 Ψ is allowed to be random due to the presence of common factors represented by $\mathcal{C}$. A similar result is shown by Andrews (2005).

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 Σ_σ and Σ_ϱ.

4 Conclusion

The paper considers a class of GMM estimators for panel data models that include possibly endogenous and dynamically evolving network or peer effect terms. Identification of these models may require both linear and quadratic moment conditions. We show that a judicious choice of quadratic moments combined with efficient forward differencing of the data leads to tractable limiting approximations of the sampling distribution. Due to the presence of common factors the limiting distribution of the GMM estimator is nonstandard, a multivariate mixture normal. This leads to the need for random norming. Despite of this it is shown that corresponding Wald test statistics have the usual χ^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 some macro economic settings where short panels are used. 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.

Monte Carlo Results for λ : Exogenous M_t

λ	Δ	OLS		2SLS		GMM	
		Bias	MAE	Bias	MAE	Bias	MAE
		(1)	(2)	(3)	(4)	(5)	(6)
				Sample Size $n = 250$			
0.1	1	0.043	0.069	0.015	0.124	0.011	0.059
0.1	0.5	0.052	0.076	0.028	0.216	0.010	0.064
0.1	0.1	0.055	0.078	0.070	0.365	0.012	0.067
0.3	1	0.110	0.111	0.020	0.111	0.014	0.054
0.3	0.5	0.126	0.126	0.054	0.196	0.015	0.059
0.3	0.1	0.132	0.131	0.118	0.335	0.017	0.062
0.5	1	0.145	0.141	0.024	0.090	0.015	0.045
0.5	0.5	0.167	0.163	0.063	0.162	0.016	0.054
0.5	0.1	0.174	0.171	0.150	0.285	0.019	0.061
0.7	1	0.128	0.126	0.020	0.059	0.012	0.034
0.7	0.5	0.151	0.148	0.050	0.110	0.018	0.078
0.7	0.1	0.160	0.157	0.136	0.207	0.024	0.106
				Sample Size $n = 500$			
0.1	1	0.038	0.053	-0.000	0.092	-0.001	0.043
0.1	0.5	0.042	0.059	0.002	0.169	0.001	0.046
0.1	0.1	0.042	0.060	0.036	0.343	0.001	0.048
0.3	1	0.104	0.102	0.004	0.083	0.002	0.039
0.3	0.5	0.118	0.117	0.020	0.153	0.002	0.042
0.3	0.1	0.123	0.122	0.113	0.326	0.003	0.043
0.5	1	0.138	0.137	0.008	0.067	0.004	0.032
0.5	0.5	0.160	0.158	0.028	0.124	0.003	0.035
0.5	0.1	0.167	0.166	0.145	0.280	0.005	0.036
0.7	1	0.124	0.123	0.008	0.044	0.004	0.023
0.7	0.5	0.146	0.145	0.026	0.084	0.007	0.061
0.7	0.1	0.154	0.154	0.123	0.201	0.008	0.070

Table 1. Monte Carlo results are based on 1,000 replications. Results are reported only for estimates of the parameter λ. 'Bias' is the median bias, MAE is the mean absolute error. OLS is the ordinary least squares estimator, 2SLS is the two stage least squares estimator, and GMM is the GMM estimator based on both linear and quadratic moment conditions.

Monte Carlo Results for λ : Endogenous M_t

λ	Δ	OLS		2SLS		GMM	
		Bias	MAE	Bias	MAE	Bias	MAE
		(1)	(2)	(3)	(4)	(5)	(6)
				Sample Size $n = 250$			
0.1	1	0.061	0.093	0.192	0.372	0.025	0.100
0.1	0.5	0.075	0.108	0.280	0.482	0.022	0.102
0.1	0.1	0.080	0.114	0.322	0.533	0.025	0.102
0.3	1	0.150	0.151	0.199	0.329	0.030	0.109
0.3	0.5	0.184	0.182	0.305	0.432	0.031	0.114
0.3	0.1	0.197	0.195	0.363	0.489	0.030	0.108
0.5	1	0.192	0.189	0.173	0.261	0.038	0.127
0.5	0.5	0.235	0.232	0.272	0.351	0.045	0.160
0.5	0.1	0.255	0.250	0.337	0.410	0.042	0.145
0.7	1	0.165	0.164	0.114	0.166	0.038	0.124
0.7	0.5	0.205	0.205	0.196	0.237	0.047	0.158
0.7	0.1	0.224	0.223	0.245	0.284	0.044	0.156
				Sample Size $n = 500$			
0.1	1	0.054	0.071	0.140	0.308	0.010	0.066
0.1	0.5	0.067	0.083	0.248	0.434	0.011	0.067
0.1	0.1	0.072	0.088	0.337	0.518	0.011	0.067
0.3	1	0.146	0.144	0.147	0.264	0.013	0.068
0.3	0.5	0.179	0.175	0.269	0.390	0.015	0.074
0.3	0.1	0.193	0.189	0.370	0.480	0.014	0.070
0.5	1	0.189	0.187	0.117	0.201	0.019	0.099
0.5	0.5	0.233	0.230	0.235	0.312	0.022	0.127
0.5	0.1	0.250	0.248	0.343	0.406	0.021	0.121
0.7	1	0.164	0.163	0.074	0.123	0.025	0.121
0.7	0.5	0.205	0.204	0.160	0.204	0.033	0.161
0.7	0.1	0.223	0.222	0.248	0.285	0.030	0.154

Table 2. Monte Carlo results are based on 1,000 replications. Results are reported only for estimates of the parameter λ. 'Bias' is the median bias, MAE is the mean absolute error. OLS is the ordinary least squares estimator, 2SLS is the two stage least squares estimator, and GMM is the GMM estimator based on both linear and quadratic moment conditions.

A Appendix: 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 $a_{ij,t}$. Let ξ denote some random variable, then $\|\xi\|_s \equiv \left(E\left[|\xi|^s\right]\right)^{1/s}$ denotes the s-norm of ξ 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 $a_{ij,t} = \left[a^r_{ij,t}\right]_{r=1,\ldots,p_t}$ 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 $\sum_{j=1}^n \left|a^r_{ij,t}\right| \leq K_a < \infty$, and $\sum_{j=1}^n \left\|a^r_{ij,t}\right\|_{2+\delta} \leq K_a < \infty$.
(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 $a^r_{ij,t}$ are non-stochastic $\left\|a^r_{ij,t}\right\|_{2+\delta} = \left|a^r_{ij,t}\right|$. The next assumption summarizes the assumed convergence behavior of sample moments of h_{it} and $a_{ij,t}$. The assumption allows for the observations to be cross sectionally normalized by ϱ_i, where ϱ_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 $t = 1, \ldots, T-1$:*

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

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 use the simplified notation V_t^h, V_t^q, 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 $\Theta_\lambda \subseteq \mathbb{R}^P$, $\Theta_\rho \subseteq \mathbb{R}^Q$, $\Theta_\beta \subseteq \mathbb{R}^k$, $\Theta_f \subseteq \mathbb{R}^{T-1}$ 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 γ_* 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 (29), 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 *(i) The cross sectional variance 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 $t \leq \tau \leq s$ let C_s be a $n \times n$ matrix of the form Υ, $\Upsilon \underline{M}_{p,s}$, $\Upsilon A_t^r \Upsilon$, $\Upsilon A_t^r \Upsilon \underline{M}_{p,s}$, or $\underline{M}'_{q,\tau} \Upsilon A_t^r \Upsilon \underline{M}_{p,s}$, where Υ is an $n \times n$ positive diagonal matrix with elements which are uniformly bounded and measurable w.r.t. $\mathcal{Z}_n \vee \mathcal{C}$. Then the probability limits ($n \to \infty$) of

$$\begin{array}{lll} n^{-1} h'_{r,t} C_s y_s, & n^{-1} h'_{r,t} C_s W_s, & n^{-1} y'_\tau C_s W_s, \\ n^{-1} W'_\tau C_s y_s, & n^{-1} y'_\tau C_s y_s, & n^{-1} W'_\tau C_s W_s, \end{array} \tag{38}$$

exist for $r = 1, \ldots, p_t$, and the probability limits are measurable w.r.t. $\mathcal{C}$, and bounded in absolute value.

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.,

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

The following assumption guarantees that the moment conditions identify the parameter θ_0. To cover initial estimators for θ_0 our setup allows both for situations where the estimator for θ_0 is based on a consistent or an inconsistent estimator of the auxiliary parameters γ_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 θ_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 $G(\theta) = \operatorname{plim}_{n\to\infty} \partial n^{-1/2}\overline{m}_n(\theta, \gamma_*)/\partial\theta$.[17] Then the following is assumed to hold:*
(i) θ_* is identifiable unique in the sense that $\mathfrak{m}(\theta_*) = 0$ a.s. and for every $\varepsilon > 0$,

$$\inf_{\{\theta\in\underline{\Theta}_\theta:|\theta-\theta_*|>\varepsilon\}} \|\mathfrak{m}(\theta)\| > 0 \text{ a.s.} \tag{39}$$

(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 Ξ is $\mathcal{C}$-measurable with a.s. finite elements, and Ξ is positive definite a.s.*

[17] Lemma 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 $\bar{\gamma}_* = \gamma_*$.

Assumptions 6(i) and 7 are crucial in establishing that θ_* is identifiable unique in the sense of Proposition 2. Assumptions 6(iii) is not required by Theorem 2.

Our specification allows for the true autoregressive parameters to be arbitrarily close to a singular point of $R_t(\lambda)$ and $\underline{R}_t(\rho)$.[18] 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 $\underline{\Theta}_f \times \underline{\Theta}_\gamma$.

B Appendix: Forward Differencing

Let $u_i^+ = \Pi u_i$ with elements $u_{it}^+ = \sum_{s=t}^{T} \pi_{ts} u_{is}$ denote the vector of forward differenced disturbances, where Π satisfies $\Pi f = 0$ and $\Pi \Sigma_\sigma \Pi' = I$. To emphasize that the elements of Π are functions of the f_t's and σ_t's we sometimes write $\pi_{ts}(f, \gamma_\sigma)$. The next proposition provides explicit expressions for $\pi_{ts}(f, \gamma_\sigma)$.

Proposition 1 [19] *(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 diagonal $T-1 \times T-1$ matrix such that $F \Sigma_\sigma F' = UU'$. Then, the $T-1 \times T$ matrix $\Pi = U^{-1}F$ is upper diagonal 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 Π. Such expression are crucial from a computational point of view, especially if f_t is

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

[19] Further details and an explicit proof are given in the Supplementary Appendix D. While the claims of the proposition are now easy to verify, the original derivation of explicit expressions for the elements of Π posed a substantial challenge.

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 σ_t in place of f_t and σ_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 $s > t$.

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 $\bar{\Sigma}_\sigma = \text{diag}(\bar{\gamma}_\sigma)$ 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 (17) 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 μ^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 Π_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 Π_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 the supplementary appendix.

C Appendix: Proofs

C.1 Martingale Difference Representation

Consider the sample moment vector $\overline{m}_n = \overline{m}_n(\theta_0, \gamma_{0,\sigma}, \gamma_\varrho)$ defined by (29), evaluated at $\theta_0, \gamma_{0,\sigma}$, but allowing for $\gamma_\varrho \neq \gamma_{0,\varrho}$. As discussed in the text, the reason for this is to accommodate both leading cases $\varrho_i^2 = \varrho_{0,i}^2$ and $\varrho_i^2 = 1$. Observe from (28) that the subvectors of $\overline{m}_n$ are given by

$$\begin{aligned} &\overline{m}_{n,t}(\theta_0, \gamma_{0,\sigma}, \gamma_\varrho) = n^{-1/2}\textstyle\sum_{i=1}^n h_{it}' u_{*it}^+ + n^{-1/2}\sum_{i=1}^n\sum_{j=1}^n a_{ij,t}' u_{*it}^+ u_{*jt}^+, \\ &u_{*it}^+ = u_{*it}^+(\theta_0, \gamma_{0,\sigma}, \gamma_\varrho) = \textstyle\sum_{s=t}^T \pi_{ts}(f_0, \gamma_{0,\sigma})\, u_{is}/\varrho_i. \end{aligned} \tag{40}$$

To establish a martingale difference representation of $\overline{m}_n = \overline{m}_n(\theta_0, \gamma_{0,\sigma}, \gamma_\varrho)$ we define the following sub-σ-fields of $\mathcal{F}$ $(i = 1, \ldots, n)$:

$$\begin{aligned}
&\mathcal{F}_{n,i} = \sigma\left(\left\{x_{j1}^o, z_j, \mu_j\right\}_{j=1}^n, \{u_{j1}\}_{j=1}^{i-1}\right) \vee \mathcal{C}, \\
&\mathcal{F}_{n,n+i} = \sigma\left(\left\{x_{j2}^o, z_j, u_{j1}^o, \mu_j\right\}_{j=1}^n, \{u_{j2}\}_{j=1}^{i-1}\right) \vee \mathcal{C}, \\
&\vdots \\
&\mathcal{F}_{n,(T-1)n+i} = \sigma\left(\left\{x_{jT}^o, z_j, u_{j,T-1}^o, \mu_j\right\}_{j=1}^n, \{u_{jT}\}_{j=1}^{i-1}\right) \vee \mathcal{C},
\end{aligned} \tag{41}$$

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 (40) 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 $u_{it}^+/\varrho_{0,i} = u_{it}^+(\theta_0, \gamma_{0,\sigma})/\varrho_{0,i} = \sum_{s=t}^T \pi_{ts}(f_0, \gamma_{0,\sigma})\left[u_{is}/\varrho_{0,i}\right]$. Since $\varrho_{0,i}$ and ϱ_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

$$S_1 = n^{-1/2}\sum_{i=1}^n \sum_{t=1}^{T-1} \lambda_t' h_{it}' \sum_{u=t}^T \pi_{tu} u_{iu} = \sum_{t=1}^T \sum_{i=1}^n c_{it} u_{it}, \tag{42}$$

with

$$c_{it} = \sum_{s=1}^t \lambda_s' h_{is}' \pi_{st} \tag{43}$$

and where we set $\lambda_T = 0$. Note that c_{it} only depends on h_{is} with $s \leq t$ and π_{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

$$S_2 = \sum_{t=1}^T \sum_{i=1}^n 2\left(\sum_{j=1}^{i-1} u_{it} u_{jt} c_{ij,tt} + \sum_{s=1}^{t-1} \sum_{j=1}^n u_{it} u_{js} c_{ij,ts}\right) \tag{44}$$

with

$$c_{ij,ts} = \sum_{\tau=1}^s \lambda_\tau' a_{ij,\tau}' \pi_{\tau s} \pi_{\tau t} \tag{45}$$

for $s \leq t$. 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 $\tau \leq s \leq t$. 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 (42) and (44) it follows that $\lambda'\overline{m}_n = \sum_{v=1}^{Tn+1} X_{n,v}$ with $X_{n,1} = 0$ and, for $t = 1, \ldots, T, i = 1, \ldots, n$,

$$X_{n,(t-1)n+i+1} = n^{-1/2} u_{it}\left(c_{it} + 2\left(\sum_{j=1}^{i-1} c_{ij,tt} u_{jt} + \sum_{j=1}^n \sum_{s=1}^{t-1} c_{ij,ts} u_{js}\right)\right) \tag{46}$$

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.[20]

C.2 Lemmas and Modules for Consistency

Lemma 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 (43) and (45) with $\pi_{ts} = \pi_{ts}(f_0, \gamma_{0,\sigma})$. Then the following bounds hold for some constant K with $1 < K < \infty$*
(i) $E\left[|c_{it}|^{2+\delta}\right] \leq K$,
(ii) $\sum_{i=1}^n |c_{ij,ts}| \leq K$,
(iii) for $q \geq 1$, $\sum_{i=1}^n |c_{ij,ts}|^q \leq K$,
(iv) for $1 \leq q \leq 2+\delta$, $\sum_{j=1}^n \|c_{ij,ts}\|_q \leq K$,
(v) for $1 \leq q \leq 2+\delta$, $E\left[|u_{it}|^q |\mathcal{F}_{n,(t-1)n+i}\right] \leq K$,
(vi) for $s \leq t$, $1 \leq q \leq 2+\delta$, $E\left[\sum_{i=1}^n |u_{is}|^q |c_{ij,ts}| |\mathcal{B}_{n,s} \vee \mathcal{C}\right] \leq K$,
(vii) for $s \leq t$, $1 \leq q \leq 2+\delta$, $E\left[\left(\sum_{i=1}^n |u_{is}| |c_{ij,ts}|\right)^q |\mathcal{B}_{n,s} \vee \mathcal{C}\right] \leq K$.

Proof. See Supplementary Appendix. ■

Lemma 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 (43) and (45) with $\pi_{ts} = \pi_{ts}(f_0, \gamma_{0,\sigma})$. 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* $n\to\infty$.

[20] 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 $\mathcal{F}_{n,(t-1)n+i} = \sigma\left\{(x_{j1}^o, z_j, \mu_j)_{j=1}^n\right\} \vee \mathcal{C}$ would violate the condition that $X_{n,v}$ is $\mathcal{F}_{n,v}$-measurable.

Proof. See Supplementary Appendix. ■

The following proposition regarding the consistency of extremum estimators 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 θ_* 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 2 *(Consistency of Stochastic Minimizers) (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 $n \to \infty$.*

Proof of Proposition 2. 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 2 is seen to hold as a special case of the generalized Lemma 3.1 in Pötscher and Prucha (1997). ■

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.

The next lemma will be useful for, e.g., establishing the consistency of variance covariance 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 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 θ_* are random vectors with $\widehat{\theta}_n \to \theta_*$ a.s. [i.p.], and*

$$\sup_{\theta \in \underline{\Theta}_\theta} |\mathcal{R}_n(\omega, \theta) - \mathcal{R}(\omega, \theta)| \to 0 \text{ a.s.[i.p.] as } n \to \infty, \tag{47}$$

then

$$\mathcal{R}_n(\omega, \widehat{\theta}_n) - \mathcal{R}(\omega, \theta_*) \to 0 \ \textit{a.s.[i.p.] as } n \to \infty. \tag{48}$$

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 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*

$$\sup_{\theta \in \underline{\Theta}_\theta} \|\mathfrak{m}_n(\omega, \theta) - \mathfrak{m}(\omega, \theta)\| \to 0 \ \textit{a.s.[i.p.] as } n \to \infty. \tag{49}$$

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

$$\sup_{\theta \in \underline{\Theta}_\theta} \left| \mathfrak{m}_n(\omega, \theta)' \Xi_n \mathfrak{m}_n(\omega, \theta) - \mathfrak{m}(\omega, \theta)' \Xi \mathfrak{m}(\omega, \theta) \right| \to 0 \ \textit{a.s.[i.p.] as } n \to \infty. \tag{50}$$

Proof. See Supplementary Appendix. ■

Lemma 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 is 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. ■

C.3 Main Results

Proof of Proposition 1. Given the explicit expressions for the elements of Π the claims of the proposition can be readily verified by straight forward calculations.[21] ■

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

Proof of Theorem 1. The proof of the proposition uses methodology similar to that used in establishing (54) below in the proof of Theorem 3. Explicit derivations are available in the Supplementary Appendix. ■

Proof of Theorem 2. $\mathcal{R}_n(\theta) = n^{-1}\overline{m}_n(\theta, \bar{\gamma}_n)'\tilde{\Xi}_n\overline{m}_n(\theta, \bar{\gamma}_n)$ and $\mathcal{R}(\theta) = \mathfrak{m}(\theta)'\Xi\mathfrak{m}(\theta)$. We use Proposition 2 to prove the theorem. Under the maintained assumptions, θ_* is identifiable unique in the sense of Condition (i) of Proposition 2. This is seen to hold in light of Condition (39) 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 2 we employ Lemma 4. By Lemma 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 Ξ is $\mathcal{C}$-measurable and finite a.s. Having verified all assumptions of Lemma 4 it follows from that Lemma that also Condition (ii) of Proposition 2, 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 2 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 $n \to \infty$. ■

In the following we establish the limiting distribution of the sample moment vector $\overline{m}_n = \overline{m}_n(\theta_0, \gamma_{0,\sigma}, \gamma_\varrho)$ defined by (29), evaluated at $\theta_0, \gamma_{0,\sigma}$, but allowing for $\gamma_\varrho \neq \gamma_{0,\varrho}$. We derive the limiting distribution of $\overline{m}_n$ by utilizing the martingale difference representation developed in Appendix C.1, and by applying the CLT of Kuersteiner and Prucha (2013, Theorem 1).

Proposition 3 *(CLT for Linear Quadratic Forms) 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

$$\overline{m}_n(\theta_0, \gamma_{0,\sigma}, \gamma_\varrho) \xrightarrow{d} V_\varrho^{1/2}\xi \quad \textit{(}\mathcal{C}\textit{-stably)}, \tag{51}$$

where $\xi \sim N(0, I_p)$*, and* ξ *and* $\mathcal{C}$ *(and thus* ξ *and* V_ϱ*) 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*

$$A\overline{m}_n \xrightarrow{d} (AV_\varrho A')^{1/2}\xi_*, \tag{52}$$

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

Proof of Proposition 3. To derive the limiting distribution we apply the martingale difference central limit theorem (MD-CLT) developed in Kuersteiner and Prucha (2013), 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 $E[X_{n,1}|\mathcal{F}_{n,0}] = 0$ where $X_{n,v}$ is defined in (46). 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:

$$\sum_{v=1}^{k_n} E\left[|X_{n,v}|^{2+\delta}\right] \to 0, \tag{53}$$

$$V_{nk_n}^2 = \sum_{v=1}^{k_n} E\left[X_{n,v}^2|\mathcal{F}_{n,v-1}\right] \xrightarrow{p} \eta^2, \tag{54}$$

$$\sup_n E\left[V_{nk_n}^{2+\delta}\right] = \sup_n E\left[\left(\sum_{v=1}^{k_n} E\left[X_{n,v}^2|\mathcal{F}_{n,v-1}\right]\right)^{1+\delta/2}\right] < \infty. \tag{55}$$

with $k_n = Tn+1$. In the following we verify (53)-(55) 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 (53) 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

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

such that by Hölder's inequality

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

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

it follows that

$$E\left[|X_{n,v}|^q\,|\mathcal{F}_{n,v-1}\right] \le \frac{2^q(T+1)^q}{n^{1+\delta/2}} E\left[|u_{it}|^q\,|\mathcal{F}_{n,(t-1)n+i}\right]\left\{|c_{it}|^q + \left(\sum_{j=1}^{i-1}|c_{ij,tt}|\right)^{q/p}\sum_{j=1}^{i-1}|c_{ij,tt}|\,|u_{jt}|^q\right.$$

$$\left. + \sum_{s=1}^{t-1}\left(\sum_{j=1}^{n}|c_{ij,ts}|\right)^{q/p}\left(\sum_{j=1}^{n}|c_{ij,ts}|\,|u_{js}|^q\right)\right\}$$

$$\le \frac{2^q(T+1)^q}{n^{1+\delta/2}} K\left\{|c_{it}|^q + K^{q/p}\sum_{s=1}^{t}\left(\sum_{j=1}^{n}|c_{ij,ts}|\,|u_{js}|^q\right)\right\}$$

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

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

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

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

which verifies condition (53).

To verify (54) 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 C.1 that the $\varrho_{0,i}^2$ and ϱ_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 2. Thus

$$V_{nk_n}^2 = \sum_{v=1}^{k_n} E\left[X_{n,v}^2 | \mathcal{F}_{n,v-1}\right] = \sum_{m=1}^{6} \sum_{t=1}^{T} \sigma_{0,t}^2 n^{-1} \sum_{i=1}^{n} \varsigma_{it}^{(m)}. \tag{56}$$

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

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

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 $t = 1, \ldots, T$, $s \leq t$, 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 (55). Analogously as in the verification of Condition (53) observe that using the triangle inequality

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

and by subsequently applying Hölder's inequality we have

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

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

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

In light of the above inequality

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

such that

$$E\left[V_{nk_n}^{2+\delta}\right] \leq \frac{2^{2+\delta}(T+1)^{2+\delta}K^{2+\delta}k_n^{\delta/2}}{n^{1+\delta/2}}\sum_{v=1}^{k_n} E\left[\left(|c_{it}|^2 + \sum_{j=1}^{i-1}|c_{ij,tt}|\,|u_{jt}|^2 + \sum_{s=1}^{t-1}\sum_{j=1}^{n}|c_{ij,ts}|\,|u_{js}|^2\right)^{1+\delta/2}\right]$$

$$\leq \frac{3^{\delta/2}2^{2+\delta}(T+1)^{2+\delta}K^{2+\delta}k_n^{\delta/2}}{n^{1+\delta/2}}\sum_{v=1}^{k_n}\left\{E\left[|c_{it}|^{2+\delta}\right] + E\left[\left(\sum_{j=1}^{i-1}|c_{ij,tt}|\,|u_{jt}|^2\right)^{1+\delta/2}\right]\right.$$

$$\left. +T^{\delta/2}\sum_{s=1}^{t-1}E\left[\left(\sum_{j=1}^{n}|c_{ij,ts}|\,|u_{js}|^2\right)^{1+\delta/2}\right]\right\}$$

where we have used repeatedly inequality (1.4.3) in Bierens(1994). By Lemma 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 1 (ii)-(vi) we have:

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

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

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 $k_n/n \leq T+1$,

$$E\left[V_{nk_n}^{2+\delta}\right] \leq \frac{3^{\delta/2}2^{2+\delta}(T+1)^{2+\delta}K^{2+\delta}k_n^{\delta/2}3T^{1+\delta/2}k_nK^{1+\delta/2}}{n^{1+\delta/2}}$$

$$\leq 3^{1+\delta/2}2^{2+\delta}(T+1)^{4+2\delta}K^{3+3\delta/2} < \infty$$

which verifies condition (55). 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 ξ_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 ξ 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_ϱ 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

$$n^{1/2}(\hat{\theta}_n - \theta_{n,0}) \xrightarrow{d} \Psi^{1/2}\xi_*, \tag{57}$$

where ξ_* is independent of $\mathcal{C}$ (and hence of Ψ), $\xi \sim N(0, I_{p_\theta})$. In light of (57), 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

$$\left(R\hat{\Psi}R'\right)^{-1/2} n^{1/2}(R\hat{\theta}_n - r) \xrightarrow{d} \xi_{**} \tag{58}$$

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 (58) 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] Arduini, T., E. Patacchini and E. Rainone, 2015, Parametric and Semiparametric IV Estimation of Network Models with Selectivity, EIEF Working Paper 15/09.

[7] Arellano, M. and S. Bond, 1991, Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations, *Review of Economic Studies*, 58, 277–297.

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

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

[10] Auerbach, E., 2016, Identification and Estimation of Models with Endogenous Network Formation, working paper.

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

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

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

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

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

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

[17] 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.

[18] Blundell, R. and S. Bond, 1998, Initial conditions and moment restrictions in dynamic panel data models, *Journal of Econometrics,* 87, 115-143.

[19] Bramoullé, Y., H. Djebbari, and B. Fortin, 2009, Identification of peer effects through social networks, *Journal of Econometrics* 150, 41-55

[20] Chandrasekhar, A. (2016), Econometrics of Network Formation, The Oxford Handbook of the Economics of Networks, Edited by Yann Bramoullé, Andrea Galeotti, and Brian Rogers.

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

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

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

[24] Chudik, A., and M.H. Pesaran, 2015, Common Correlated Effects Estimation of Heterogeneous dynamic panel data models with weakly exogenous regressors, , *Journal of Econometrics* 188, 393-420.

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

[26] de Paula, A. (2017), Econometrics of Network Models, Advances in Economics and Econometrics, Eleventh World Congress, Edited by Bo Honoré, Ariel Pakes, Monika Piazzesi and Larry Samuelson, Cambridge University Press.

[27] 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.

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

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

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

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

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

[33] Graham, B., 2016, Homophily and transitivity in dynamic network formation, working paper.

[34] Han, X., and L.-F. Lee, 2016, Bayesian Analysis of Spatial Panel Autoregressive Models With Time-Varying Endogenous Spatial Weight Matrices, Common Factors, and Random Coefficients, *Journal of Business & Economic Statistics* 34, 642-660.

[35] Hayakawa, K, 2015, The asymptotic properties of the system GMM estimator in dynamic panel data models when both N and T are large, *Econometric Theory* 31, 647-667.

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

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

[38] Hsieh, C., and L.F. Lee, 2016, A Social Interactions Model with Endogenous Friendship Formation and Selectivity, *Journal of Applied Econometrics,* 31, 301-319.

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

[40] Johnson, I., and H.R. Moon, 2017, Estimation of Peer Effects in Endogenous Social Networks: Control Function Approach, USC-INET Research paper 17-25.

[41] Jackson, M. O. (2008), Social and Economic Networks. Princeton University Press.

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

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

[44] Kelejian, H.H., and G. Piras, 2014, Estimation of Spatial Models with Endogenous Weighting Matrices, and an Application to a Demand Model for Cigarettes, Regional Science and Urban Economics 46, 140-149.

[45] 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.

[46] 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.

[47] 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.

[48] Kelejian, H.H., and I.R. Prucha, 2002, 2SLS and OLS in a Spatial Autoregressive Model with Equal Spatial Weights, *Regional Science and Urban Economics*, 32, 2002, 691-707.

[49] Kelejian, H.H., I.R. Prucha, and Y. Yuzefovich, Estimation Problems in Models with Spatial Weighting Matrices which have Blocks of Equal Elements, *Journal of Regional Science*, 46, 2006, 507-517.

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

[51] 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.

[52] Lee, L.-F., 2007, GMM and 2SLS estimation of mixed regressive, spatial autoregressive models, *Journal of Econometrics* 137, 489-514.

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

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

[55] 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.

[56] Leung, M.P. (2016), Treatment and Spillover Effects under Network Interference, manuscript.

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

[58] Manski, C.F., 1993, Identification of Endogenous Social Effects: The Reflection Problem, *Review of Economic Studies* 60, 531-542

[59] Moon, H.R., and M. Weidner, 2015, Linear Regression for Panel with Unknown Number of Factors as Interactive Fixed Effects, *Econometrica* 83, 1543-1579.

[60] Moon, H.R., and M. Weidner, 2017, Dynamic Linear Panel Regression Models with Interactive Fixed Effects, *Econometric Theory* 33, 158-195.

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

[62] 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.

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

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

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

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

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

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

[69] Qu, X., and L.-F. Lee, 2015, Estimating a Spatial Autoregressive Model with an Endogenous Spatial Weight Matrix, *Journal of Econometrics* 184, 209-232.

[70] Qu, X., L.-F. Lee and J. Yu, 2017, QML Estimation of Spatial Dynamic Panel Data Models with Endogenous Time Varying Spatial Weights Matrices, *Journal of Econometrics* 197, 173-201.

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

[72] Ridder, G. and S. Sheng (2016), Estimation of Large Network Games, manuscript.

[73] Shi, W., and L.-F. Lee, 2018, A Spatial Panel Data Model with Time Varying Endogenous Weights Matrices and Common Factors, *Regional Science and Urban Economics* 72, 6-34.

[74] Sheng, S. (2016), A Structural Econometric Analysis of Network Formation Games, manuscript.

[75] Su, L. J. and Z.L. Yang, 2015, QML estimation of dynamic panel models with spatial errors, *Journal of Econometrics* 185, 230-258.

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

[77] 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.

[78] 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.