Identification of Social Effects with Endogenous Networks and Covariates: Theory and Simulations

Santiago Pereda-Fernández[1] and Paolo Zacchia[2]

[1]Banca d'Italia

[2]IMT School for Advanced Studies Lucca

March 2018

Abstract

The estimation of spillover and peer effects presents challenges that are still unsolved. In fact, even if separate algebraic identification of the endogenous and exogenous effects is possible, these might be contaminated by the simultaneous dependence of outcomes, covariates and the network structure upon spatially correlated unobservables. In this paper we characterize the identification conditions for consistently estimating all the parameters of a spatially autoregressive or linear-in-means model in presence of linear forms of endogeneity. We show that identification is possible if the spatial correlation of individual covariates and that of unobservables do not overlap, and we relate this idea to a schooling context in which the factors that determine friendships and socio-economic characteristics are different. We propose a GMM estimator to estimate the relevant parameters and we evaluate its performance through Monte Carlo simulations.

JEL Classification Codes: C21, C31, D85

Keywords: Peer Effects, Networks, Network Endogeneity, Common Shocks

1 Introduction

A sizable body of empirical economic research deals with the analysis of peer effects, network effects, social interactions and – more generally – externalities. Among these studies, the literature about peer effects in education occupies perhaps a more prominent position (Sacerdote, 2001; Calvó-Armengol et al., 2009; De Giorgi et al., 2010; Carrell et al., 2013), but applications in more diverse fields are numerous (Glaeser et al., 1996; Duflo and Saez, 2003; Mas and Moretti, 2009).[1] In the face of growing empirical evidence, econometric analysis has struggled for a while to characterize conditions for providing a more structural interpretation to observed group correlations in outcomes. While important advances have been realized, their relevance is limited to a restricted set of empirical settings, in which the characteristics of individuals as well as their structure of socio-economic interactions are both as good as exogenous.

In order to better characterize the position of the present article in this literature, it is worth to summarize the evolution of economists' understanding of the workhorse framework for many studies about social effects: the "linear-in-means" model. Following Manski (1993), who highlighted the so-called *reflection* problem of simultaneity between group characteristics and group outcomes, econometricians have attempted to individuate settings in which endogenous social responses can be identified separately from the influence of common external factors. In an influential contribution, Bramoullé et al. (2009) express the conditions for identification when social effects are shaped by networked structures of interaction, which is particularly appealing as networks typically provide more realistic descriptions of actual social relationships. Blume et al. (2015) incorporate their identification results – as well as one based on covariance restrictions which builds on Graham (2008) – within a larger framework. Thanks to these and other efforts, it is now well understood that complex patterns of individual dependence if anything make the identification of social effects *easier*. Yet all these analyses maintain the assumption that the model's error term is conditionally independent of the individual covariates and the structure of interactions.

[1] Studies of R&D and knowledge spillovers more generally, which follow the tradition initiated by Jaffe (1986, 1989), are seldom counted among these studies. This is quite a notable omission, since the workhorse econometric frameworks employed in this literature are easily seen as variations of the standard spatial models utilized for the estimation of peer effects. More recent contributions about R&D spillovers include Bloom et al. (2013), Lychagin et al. (2016) and Zacchia (2017). Other related strands of literature include the one about peer effects in scientific production (Azoulay et al., 2010; Waldinger, 2011; Oettl, 2012) and that about learning externalities (Conley and Udry, 2010).

By contrast, in this paper we examine a model of social interactions where both individual characteristics and the network that defines paired relationships are simultaneously dependent on individual unobservables. Our departure point is a Spatially Autoregressive (SAR) model (Cliff and Ord, 1981), of which the linear-in-means model is a special case, and whose econometrics has been rigorously analyzed in a number of theoretical papers (Lee, 2007; Lee et al., 2010; Lin and Lee, 2010; Liu and Lee, 2010). In conformity with other articles from this literature, we derive our empirical model from an explicit theoretical framework; unlike most, ours is based on a Cobb-Douglas utility function, and it can accommodate contexts ranging from peer effects in the classroom to R&D spillovers. We discuss how standard estimates of social effects are inconsistent if unobservables correlate with covariates *and* with peers' unobservables, and we illustrate the related identification problem by showing that the errors' cross-correlation is observationally equivalent to typical exogenous or "contextual" effects often featured in linear-in-means models. This resonates with the notorious critique of peer effect studies put forward by Angrist (2014), according to whom the current results in the literature may reflect spurious correlations due to "correlated effects."

The main contribution of our paper is to show that for certain structures of underlying endogeneity, the social effects of interest can be internally identified without resorting to external instruments. Specifically, it is necessary that the cross-correlation of the observable and unobservable characteristics is not collinear in social or network space. In such a case, it is possible to identify some residual variation in peers characteristics that works as an exogenous predictor of peers outcomes, by which the identification of social effects would proceed as in the standard case. We relate this intuition to a schooling setting in which peers establish bonds on the basis of their similarity in unobservable characteristics (like ability), while the observable covariates (like family background) correlate with the unobservables even between non-friends. Once again, the increased complexity of social relationships helps bring about easier identification. Although we initially illustrate our result under some simplifying assumptions, we further show that the basic idea can be extended to more general cases. For our simpler setup, we characterize a GMM approach for the joint estimation of both the social effects and the other parameters of the data generation process. We evaluate the performance of our estimator through some Monte Carlo simulations.

The one we propose is a novel approach in the literature. In line with a recommendation given by Blume et al. (2015), some scholars (Arduini et al., 2015; Johnsson and

Moon, 2017) develop a control function approach to account for network endogeneity. These contributions embed, within a SAR model, the network formation stage and its estimation, which are based on Graham (2017). However, they do not consider the possible simultaneity between the network and individual characteristics, which can be due to correlated unobserved errors. By contrast, our strategy does; in addition, it does not require an explicit model of network formation, so long as the spatial scope of the relevant cross-correlations is known. In this respect, our method extends some previous work by Zacchia (2017).[2] Obviously, the spatial econometrics literature has examined correlated unobservables at length (Kelejian and Prucha, 1998, 2007, 2010; Drukker et al., 2013); however, individual covariates are usually assumed exogenous.[3]

It is useful to relate our article to other papers from the literature about peer and network effects. In addition to the cited contribution by Graham (2008), other papers make use of conditional covariance restrictions to achieve the identification of social effects (Glaeser et al., 1996; Moffitt, 2001; Davezies et al., 2009; Pereda Fernández, 2017; Rose, 2017a). Our method also exploits covariance restrictions, but unlike these papers, their role in identification is to estimate the covariance structure of the model so to isolate the variation of peers' characteristics that is exogenous to individual unobservables. Finally, it is important to observe that in our setup, the matrices that characterize both the social effects and dependence of peers characteristics on individual errors are assumed to be known by the econometrician. In addition, methods developed for estimating the structure of interactions (Manresa, 2017; Rose, 2017b; De Paula et al., 2018) which make use of penalized estimators such as the LASSO (Tibshirani, 1996) cannot be adapted to this setup. We revisit these observations in the conclusion of the paper while suggesting future lines of work.

The remainder of this paper is organized as follows. Section 2 presents our general analytical framework. Section 3 discusses the conditions for the identification of social effects. Section 4 characterizes the GMM estimator associated with our simpler setup. Section 5 demonstrates its qualities in Monte Carlo simulation. Finally, 6 concludes the paper. A separate Appendix provides the proofs for the main results.

[2]Zacchia (2017) analyzes a model of R&D spillovers in which firms' unobservables are correlated in the network of R&D relationships, and are simultaneous to the R&D of connected firms. In order to identify spillover effects, he constructs IVs motivated on the finite empirical spatial correlation of R&D. The framework presented here instead does not restrict the spatial correlation of covariates.

[3]In a recent contribution, Kuersteiner and Prucha (2018) examine a SAR model for panel data in which the interaction matrix is possibly endogenous and covariates are weakly exogenous, and propose an appropriate GMM estimator. In our cross-sectional framework covariates are endogenous.

2 Analytical Framework

In this section we introduce the social interactions game that constitutes the theoretical framework of this paper. We subdivide this section between the description of the model's setup and the discussion of the equilibrium predictions.

2.1 Model's Setup

We consider an abstract setting of social and economic interactions between heterogeneous agents (players) in a network. In order to allow for interdependence between the characteristics of agents and the structure of their connections, we allow nature to randomly draw the *weighted network* $(\mathcal{I}, \mathcal{G})$ that characterizes the social interactions. Here, $\mathcal{I}$ is the set that comprises the N players, who are indexed as $i = 1, \ldots, N$. The N^2-dimensional set $\mathcal{G}$, instead, represents the interaction structure: thus, $g_{ij} \in \mathbb{R}$ denotes the relative strength of the influence exerted by player j on player i (and vice versa). We impose two standard normalizations: $g_{ij} \in [0, 1]$ and $g_{ii} = 0$ for all players $i = 1, \ldots, N$. Furthermore, we assume for simplicity that the network is *undirected* (symmetric), that is $g_{ij} = g_{ji}$ for all pairs $(i, j) \in \mathcal{I}^2$. Our results are easily extended to the case of *directed* (asymmetric) networks, where $g_{ij} \neq g_{ji}$.

Every player in $\mathcal{I}$ is typified by two variables (x_i, ε_i). We denominate $x_i \in \mathcal{X}$ the *observable characteristics* of player i, and $\varepsilon_i \in \mathcal{E}$ his or her *unobservables*: this abstract terminology clearly relates to the information which is available to the econometricians who are in search of social externalities. For simplicity we set $\mathcal{X} = \mathcal{E} = \mathbb{R}$, although many, possibly discrete characteristics could easily be accommodated. We assume that the random vector of individual observable characteristics $\boldsymbol{x} = (x_1, \ldots, x_N)$, the random vector of individual unobservables, $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_N)$ and the network $\mathcal{G}$ are all randomly drawn from a joint probability distribution $\mathcal{F}(\boldsymbol{x}, \boldsymbol{\varepsilon}, \mathcal{G})$, which is known by all agents. We place no a priori restrictions on the distribution $\mathcal{F}(\cdot)$.

The economic content of the description outlined thus far deserves some further discussion. In social networks, the probability of a connection occurring between any two agents is documented to be correlated with their characteristics. For example, friends usually sort on social background and demographics, while R&D spillovers naturally occur between firms belonging to the same technological class. This result is predicted by many models of random and strategic network formation, and the social mechanism by which similar agents are paired to one another bears the name of homophily. However, it is apparent that many of the characteristics that predict the occurrence (or the relative strength) of connections are unobserved by researchers: for example, student friendships may be sorted on ability; likewise, R&D connections may appear more frequently between firms with shared technologies. In either case, the fact that connected agents share some of their unobservables poses identification problems to the econometrician. Zacchia (2017) discusses in more detail how "common unobservables" and "network endogeneity" are two intertwined issues.[4]

Players maximize the following "twice exponential" utility function:

$$U_i\left(e_1, \dots, e_N; x_i, \varepsilon_i\right) = \exp\left[y_i\left(e_1, \dots, e_N; x_i, \varepsilon_i\right)\right] - \exp\left(e_i\right) \tag{1}$$

where y_i is the individual-level *outcome* (denoting, say, grades, or production output) which is determined through the following linear relationship.

$$y_i\left(e_1, \dots, e_N; x_i, \varepsilon_i\right) = \alpha_0 + \gamma_0 x_i + \mu e_i + \nu \sum_{i=1}^{N} g_{ij} e_j + \varepsilon_i \tag{2}$$

The outcomes of individuals depend upon their characteristics (x_i, ε_i) as well as on a costly strategic variable $e_i \in \mathbb{R}$ that we call *effort*: this may represent, for instance, time dedicated to homework or R&D investment. Because of social interactions and externalities, y_i also depends on the effort of all the other players an agent is connected to (possibly negatively). Private and social effects of effort are parametrized as μ and ν, respectively. To make the model realistic, we impose the following restriction.

Assumption 1. Concavity: $\mu \in (0, 1)$

This assumption makes *i.* individual output positively dependent on effort, and *ii.* the utility function concave in $\exp(e_i)$, so that choice trade-offs are cogent. As we discuss later, additional restrictions on ν are necessary to ensure equilibrium uniqueness.

Note two differences between this framework and the quadratic utility model that is typical of the peer effects literature (Calvó-Armengol et al., 2009; Blume et al., 2015). First, the model proposed here outlines a clear distinction between individual choice variables and ultimate outcomes, which undoubtedly gives it more generality.

[4]In addition, in an extension of his theoretical framework (which is tailored to the setting of R&D spillovers) he observes that common models of network formation imply little or no cross-correlation of individual variables at three degrees of separation or more, which is in remarkable agreement with the empirical evidence.

Second, while both utility functions are globally concave in their respective strategic variables, in the case of (1) individual characteristics x_i, ability ε_i, individual effort e_i and the effort e_j of connections are complements. In addition to accommodating functional forms such as those that are typical of production functions, this increases the degree of realism of the model even in other social contexts. For example, more skilled or better supported students may benefit relatively more from devoting more time to homework and independent study, either alone or with their friends. Finally, observe that one could easily introduce heterogeneous weights to the benefit and cost components of (1), but this is beyond the point of the present analysis.

We analyze a game of complete information characterized by the following timing.

1. Nature draws $(\boldsymbol{x}, \boldsymbol{\varepsilon}, \mathcal{G})$ from $\mathcal{F}(\cdot)$. Every player observes the result of this draw.
2. Players simultaneously make their effort choices, and utilities are determined accordingly.

By letting the network be generated randomly by nature we abstract from the specifics of the network formation process, as our results do not depend on it. Also note that by assuming complete information we make our analysis more general. As discussed by Zacchia (2017), in fact, incomplete information provides more avenues for the identification of social effects, in the form of implicit restrictions on the cross-correlation of strategic variables.

2.2 Analysis

We analyze the properties of the equilibrium conditional upon a restriction, that we maintain throughout our discussion, about the combined parameter $\beta \equiv \nu (1 - \mu)^{-1}$.

Assumption 2. Non-explosiveness: $|\beta| \cdot \max_{i \in \mathcal{I}} \sum_{j=1}^{N} g_{ij} \in [0, 1)$

This assumption imposes that social effects do not "dominate" the process of outcome generation. In the game, it ensures uniqueness of the equilibrium by ruling out unrealistically "explosive" scenarios. In statistical terms, this assumption makes it possible that the variation of y_i is not predominantly explained by the cross-correlation of outcomes in the network: we find that otherwise, the identification problems discussed in this article are largely moot, since standard estimators would capture the social effects with little bias relatively to the overall variance of the dependent variable. We observe that variations of this hypothesis are often assumed in the literature.

In standard models of peer effects, it is also routinely assumed that the in-degree of agents is constant and normalized to one, as follows.

Assumption 3. Row Normalization: $\bar{g}_i \equiv \sum_{j=1}^{N} g_{ij} = 1$ *for all* $i = 1, \ldots, N$

Under Assumption 3 social effects represent the individual response to the (weighted) average behavior or characteristics of peers. This contrasts with models where social effects are a function of the total intensity of connections. Throughout most of this paper we will maintain Assumption 3, while concentrating on the identification of the combined parameter β. Later we relax this hypothesis and, among the possible extensions of our approach, we discuss the possibility to separately identify μ and ν by exploiting variation in individual in-degree. Incidentally, observe that Assumption 3 implies that no agent is allowed to be "isolated" (disconnected from the network) and that under row normalization, Assumption 2 reduces to $|\beta| \in [0, 1)$.

Under all the hypotheses outlined thus far, the following result is easily obtained.

Proposition 1. Equilibrium. *For all realizations of* $(\boldsymbol{x}, \boldsymbol{\varepsilon}, \mathcal{G})$*, under Assumptions 1-3 there exists a unique equilibrium of the game, which gives rise to an equation for the outcome* y_i *that can be expressed for each player* $i = 1, \ldots, N$ *as follows:*

$$y_i = \alpha + \beta \sum_{j=1}^{N} g_{ij} y_j + \gamma x_i + \varepsilon_i \tag{3}$$

where $\alpha \equiv (1 - \mu)^{-1} [\alpha_0 + (\mu + \nu) \log \mu]$ *and* $\gamma \equiv (1 - \mu)^{-1} \gamma_0$.

Proof. The First Order Condition from utility maximization can be written, for each player $j = 1, \ldots, N$, as:

$$e_j = y_j + \log \mu \tag{4}$$

substituting this expression into (2) results in (3). Moreover, by substituting (2) into (4) and solving for e_j it is easily seen that – under the non-explosiveness condition – the N First Order Conditions together represent a contraction of $(e_1, \ldots, e_N)$ in the $\left(\mathbb{R}^N, \mathcal{M}\right)$ metric space, where $\mathcal{M}$ is the max norm. This implies uniqueness. □

Let us examine the reduced form expression (3) that is generated in equilibrium. While it resembles the standard equation of linear in means models from the peer effects literature, it provides some additional insights in relationship with the model.

First, parameter β – corresponding to the *endogenous effect* from the original classification by Manski (1993) – is given here a clear behavioral interpretation. In fact, β is equal to the direct effect of connections' effort ν amplified by a factor representing the equilibrium response of individual effort caused by complementarities: intuitively, students put additional effort while firms increase their R&D investment as they are aware of the interdependencies and expect their connections to behave similarly. This interpretation of β is important, since in many empirical studies of social externalities individual "effort" is not observable by researchers.

The second difference with typical linear-in-means models is that in our model we do not include Manski's *exogenous effect*, that is a structural dependence of individual outcomes on the characteristics x_j of peers (also called "contextual" effects). While we could easily include an additional term in (2) to allow for the exogenous effect, we believe that our choice makes it easier to illustrate the following fact.

Proposition 2. Non-identification of contextual effects: *There exist specific restrictions on $\mathcal{F}(\cdot)$ such that the model is observationally equivalent to the following alternative structure:*

$$y_i = \alpha' + \beta' \sum_{j=1}^{N} g_{ij} y_j + \gamma' x_i + \delta' \sum_{j=1}^{N} g_{ij} x_j + \varepsilon'_i \tag{5}$$

where $\delta' \neq 0$ and the random vector $\boldsymbol{\varepsilon}' = (\varepsilon'_1, \ldots, \varepsilon'_N)$ is such that $\mathbb{E}[\boldsymbol{\varepsilon}' | \boldsymbol{x}, \mathcal{G}] = 0$.

Proof. Suppose that $\mathcal{F}(\cdot)$ implies that $\varepsilon_i = \rho \sum_{j=1}^{N} g_{ij} \varepsilon_j + \varepsilon'_i$ and $\mathbb{E}[\varepsilon_j | x_j] = \kappa + \chi x_j$ where $\chi \neq 0$ and $\rho \neq 0$. It is easy to see that under those conditions, models (3) and (5) are observationally equivalent under $\alpha' = \alpha + \rho\kappa$, $\beta' = \beta$, $\gamma' = \gamma$ and $\delta' = \rho\chi$. $\square$

While the particular example that we chose to straightforwardly prove our statement is abstract,[5] it serves to make an important point. If individual unobservables ε_i are correlated in the network – say, because agents form connections by sorting on ability – and, in addition, individual characteristics x_i are also correlated with the unobservables, then "contextual effects" δ' are just a statistical byproduct of these more fundamental structural behavioral patterns. We see this as a cautionary message to researchers aiming to estimate spillover effects in any given economic context: the

[5] In that example ability ε_i follows a first order spatially autoregressive process, which implies that individual unobservables are increasingly dissimilar the farther apart are any two agents in the network (intuitively, a spatial AR(1) process can be approximated as a spatial MA(∞) process).

solution of potential endogeneity problems due to simultaneous unobservables and network formation must precede model specification. Clearly, a similar problem may also affect the main behavioral parameter β of endogenous spillover effects. The rest of this article discusses strategies aimed at disentangling genuine externalities from shared confounders. Throughout most of the exposition we maintain the assumption that individual "effort" is not directly observable by researchers.

3 Identification

In this section we discuss under what conditions it is possible to identify the parameters of model (3) even if individual characteristics and the network are endogenous with respect to the unobservables. Following a description the problem, we illustrate our approach first under simple linear assumptions about the underlying data generation process, and then under more general conditions. At the end of the section we comment on some possible extensions of the proposed methodology.

3.1 SAR models

We find it useful to briefly discuss how the endogeneity problem we are concerned with differs from those of previous analyses. To this end, we re-write model (3) using matrix notation:

$$\mathbf{y} = \alpha\boldsymbol{\iota} + \beta\mathbf{G}\mathbf{y} + \gamma\mathbf{x} + \boldsymbol{\varepsilon} \tag{6}$$

where $\mathbf{y} = (y_1, \ldots, y_N)^{\mathrm{T}}$, $\mathbf{x} = (x_1, \ldots, x_N)^{\mathrm{T}}$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_N)^{\mathrm{T}}$ are the realizations of y_i, x_i and ε_i – respectively – stacked over all the agents; $\mathbf{G}$ instead is the *adjacency matrix* with g_{ij} entries. Following the classification of spatial econometric models by Elhorst (2014), we call this a *spatially autoregressive* (SAR) model.[6] Note that under row-normalization of $\mathbf{G}$ (Assumption 3) any SAR model corresponds to the linear-in-means model typical of peer effects studies, but deprived of contextual effects.

The most apparent econometric problem of model (6) is one of simultaneity: since the y_i's of different agents are structurally dependent on one another, the spatially autoregressive component $\mathbf{Gy}$ of (6) is correlated with the error term – the so-called *reflection* problem – and thus OLS estimation is inconsistent. There is a vast literature

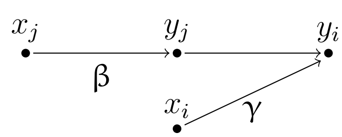
[6] Other authors prefer the denomination "mixed regressive-spatially autoregressive" in order to remark the presence of $\mathbf{x}$ on the right-hand side of (6). Here we opt for a more concise terminology.

in spatial econometrics, which is not our objective to review here, that concerns the ML estimation of (6) under normality assumptions. Semi-parametric approaches to the estimation of models akin to (6) include IV-2SLS (Kelejian and Prucha, 1998) as well as GMM (Lin and Lee, 2010). The former appears of particular relevance here, as it has been extended to models featuring contextual effects and network fixed effects through the influential contribution by Bramoullé et al. (2009).

To understand why internal identification of (6) is possible, observe that if $\mathbf{G}$ is linearly independent from the identity matrix $\mathbf{I}$, as $(\mathbf{I} - \beta\mathbf{G})$ is then invertible the model can be rewritten in a "reduced form" fashion as:

$$\mathbf{y} = (\mathbf{I} - \beta\mathbf{G})^{-1} (\alpha\boldsymbol{\iota} + \gamma\mathbf{x} + \boldsymbol{\varepsilon}) \simeq \sum_{r=1}^{\infty} \beta^r \mathbf{G}^r (\alpha\boldsymbol{\iota} + \gamma\mathbf{x} + \boldsymbol{\varepsilon}) \tag{7}$$

implying, under $\mathbb{E}[\boldsymbol{\varepsilon}|\, \mathbf{x}] = 0$, the existence of an *infinite* set of instrumental variables of the form $(\mathbf{Gx}, \mathbf{G}^2\mathbf{x}, \dots)$ of which only $\mathbf{Gx}$, though, is generally expected to be also relevant. Yet this enough for identification, the intuition being that it is possible to predict the outcomes of connected agents through their characteristics. This idea is exemplified Graph 1, which involves variables $(x_i, y_i; x_j, y_j)$ pertaining to any two connected observations (i, j). In the graph, arrows represent the structural relationships between variables that allow to identify the indicated parameter of interest.



Graph 1: Identification of SAR models

Models featuring contextual effects like a term $\delta\mathbf{Gx}$ on the right-hand side of (6) entail the additional complication that clearly $\mathbf{Gx}$ is not excluded from the structural form. However, $\mathbf{G}^2\mathbf{x}$ would then be a relevant instrument for the identification of β: if contextual effects exist, the characteristics of "friends of friends" affect the outcomes of direct peers – easily extending the intuition above – so that β and δ are separately identified. These ideas are best framed as a system of simultaneous equations, which are generally known to be identified so long as enough instrument exist to satisfy both the order and rank conditions. Here it is the structure of networks that naturally gives rise to the appropriate exclusion restrictions, in the form of the characteristic of others agents that have no direct effect on individual outcomes (Rose, 2017b).

These ideas and the related results are all based on the assumptions of exogenous covariates $\mathbf{x}$. In the systematic analysis of the literature by Blume et al. (2015), an equivalent of assumption $\mathbb{E}[\boldsymbol{\varepsilon}|\,\mathbf{x}] = 0$ is central to all results about identification. In SAR and linear-in-means models, the endogeneity of individual characteristics not only prevents the identification of their specific effect on the outcome of interest, but also of social effects themselves, as the x_i's of peers can no longer be employed as instruments. This fact suggests that the problem is particularly subtle in this case, as its gravity depends on the breadth of endogeneity in network space – that is, to what extent individual unobservables are correlated with peers' characteristics. In what follows we illustrate, starting for the sake of exposition from a linear characterization the problem, under what conditions internal identification of SAR models is possible even if covariates are endogenous.

3.2 Simple Linear Endogeneity

Suppose that both the network $\mathcal{G}$ and agents' characteristics $\boldsymbol{x}$ are statistically related to the unobservables through some auxiliary random variables $\boldsymbol{\upsilon} = (\upsilon_1, \ldots, \upsilon_N) \in \mathbb{R}^N$, drawn from some unspecified distribution, that we call *diffused shocks*. Specifically, assume that $\mathcal{F}(\cdot)$ is characterized by the following relationships.

Assumption 4. Diffused Shocks: $\mathbb{E}[\upsilon_i] = 0$ *and, for some* $d > 0$, $\mathbb{E}\left[|\upsilon_i|^{4+d}\right] < \infty$ *for* $i = 1, \ldots, N$; $\mathbb{E}[\upsilon_i \upsilon_j] = 0$ *for any pair* (i, j) *such that* $i \neq j$.

Assumption 5. SMA(1) Unobservables: $\boldsymbol{\varepsilon} = (\mathbf{I} + \psi\mathbf{G})\,\boldsymbol{\upsilon}$ *where* $|\psi| < 1$.

Assumption 6. Linear Endogeneity: $\boldsymbol{x} = \boldsymbol{\omega} + \xi\mathbf{C}\boldsymbol{\upsilon}$ *where* $\xi \in \mathbb{R}$, $\mathbf{C}$ *is a* $N \times N$ *invertible, nonstochastic* "characteristics matrix" *such that for any* $i \in \mathcal{I}$, $c_{ii} \neq 0$ *is allowed, and* $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_N) \in \mathbb{R}^N$ *is a random vector with the following properties:*

a. *unrestricted mean* $\mathbb{E}[\omega_i]$ *and, for some* $d > 0$, $\mathbb{E}\left[|\omega_i|^{4+d}\right] < \infty$ *for* $i = 1, \ldots, N$;

b. $\boldsymbol{\omega}$ *independent of the diffused shocks* $\boldsymbol{\upsilon}$: $\mathbb{E}\left[\boldsymbol{\omega}\boldsymbol{\upsilon}^{\mathrm{T}}\right] = 0$;

c. *the conditional expectation vector* $\boldsymbol{\varpi}(\alpha, \beta, \gamma, \xi, \psi) \equiv \mathbb{E}[\boldsymbol{\omega}|\,\mathbf{x}] \neq \mathbf{x}$ *is uniquely identified given a vector of parameters* $(\alpha, \beta, \gamma, \xi, \psi)$.

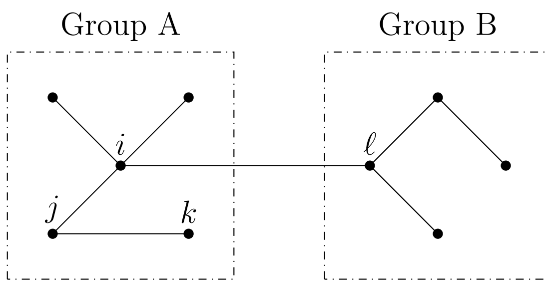
These assumptions together re-define the endogeneity problem by decomposing individual characteristics, both observable (x_i) and unobservable (ε_i), in terms of a sequence of N independent "diffused" shocks υ_i. The latter are closely related to the "common shocks" from other studies of spillover effects or, in Manski's classification, the "correlated effects" that express the correlation between the unobservables of all individuals from the same group. We introduce a different terminology to highlight the fact that these unobservables are not necessarily homogeneous across observations within given groups, but are somehow "shared" by agents as a function of their relative proximity in the structure of social interactions. In fact, it is not even necessary that diffused shocks bear any economic interpretation: they mainly serve to characterize the structure of cross-correlation of (x_i, ε_i) in a tractable way.

These points are made clearer by illustrating in more detail the assumptions above. In particular, Assumption 4 is a standard characterization of primitive shocks: they have zero mean, finite higher moments, and are mutually independent across observations. Assumption 5 expresses the stochastic process that defines the cross-correlation of unobservables in the network. Specifically, here $\boldsymbol{\varepsilon}$ follows a Spatial Moving Average (SMA) process of first degree: ε_i equals υ_i plus a weighted average of the υ_j's of an agent's links, times a multiplicative factor ψ of small enough magnitude so to ensure stationarity. This process could be given a structural interpretation, if say the ability of peers or the technology of connected firms directly affects individual outcomes y_i. Alternatively, the SMA(1) process may be given a mere statistical interpretation, as the representation of the network formation process enclosed in $\mathcal{F}(\cdot)$.[7]

Finally, Assumption 6 describes the dependence of observables x_i's from diffused shocks. Specifically, the random vector $\boldsymbol{x}$ of individual characteristics is decomposed into two additive, stochastically independent terms. The first is a random vector $\boldsymbol{\omega}$ with unrestricted mean and finite higher moments, which represents the independent component of the variation of $\boldsymbol{x}$. The second term is a linear function of the diffused shocks $\boldsymbol{\upsilon}$, as expressed by what we denominate the *"characteristics matrix"* $\mathbf{C}$. This matrix, which we assume nonstochastic but leave unrestricted and not necessarily related to the network $\mathcal{G}$, characterizes the cross-correlation of individual characteristics x_i's. Note that the spatial correlation of $\boldsymbol{x}$ and $\boldsymbol{\varepsilon}$ are related to one another through the diffused shocks: under linearity assumptions this makes for a relatively simple representation of endogeneity in the spirit of Angrist's (2014) mentioned critique.

[7]Note that, in fact, any SMA(1) process implies a cross-correlation of unobservables that extends by two degrees of separation in the network. As discussed by Zacchia (2017), this property is a good approximation of a model of network formation driven by a homophily dynamic: two agents link up with some probability only if they are similar.

Since **C** can generally be any given matrix, Assumption 6 allows to represent a wide category of economic structures of interdependence, including patterns of cross-correlation of the x_i's that are distinct from those of the unobservables. For example, **C** may represent a fully overlapping group structure, whereby if any two agents i and j are connected, they are also either both connected or both disconnected to any third agent k (if $c_{ij} \neq 0$ then $c_{ik} \neq 0 \Leftrightarrow c_{jk} \neq 0$ for all $(i, j, k) \in \mathcal{I}^3$). In terms of our motivating settings, this allows scenarios where family background is correlated across all students within the same classroom because of some social sorting mechanism – say, parental selection of the best teachers – while unobserved ability only correlates across "friends" because of homophily in network formation. A specular interpretation is one where observable firm characteristics are more similar within industries, while R&D spillovers transcend sectors because of other kinds of technological similarities. Both cases can be graphically illustrated as follows.



Graph 2: A Cross-Group Network

In this graph, nodes represent observations while edges denote network connections (as usual). In addition, all nodes belong to different "groups" within which observable characteristics are correlated. Thus $g_{ij} \neq 0$ and $c_{i\ell} \neq 0$, but $g_{ik} = 0$ and $c_{j\ell} = 0$.

Despite being flexible, Assumptions 5-6 may still appear quite restrictive. In particular, point 6.c imposes some additional conditions on the conditional expectation $\mathbb{E}[\boldsymbol{\omega}|\, \mathbf{x}]$: it must be uniquely determined by the parameters and distinct from $\mathbf{x}$ (this holds trivially if matrices **G** and **C** are nonzero, that is if endogeneity is actually a problem). Yet their implications proves useful in order to appreciate the intuition behind our main result that is presented later. To this end, note that these hypotheses imply the following regression function.

$$\mathbb{E}[\mathbf{y}|\, \mathbf{x}] = (\mathbf{I} - \beta\mathbf{G})^{-1} \left[\alpha\iota + \gamma\mathbf{x} + \xi^{-1} (\mathbf{I} + \psi\mathbf{G})\, \mathbf{C}^{-1} (\mathbf{x} - \boldsymbol{\varpi})\right] \tag{8}$$

Observe how, under our restrictive but simple assumptions, expression (8) subsumes all sources of interdependence and endogeneity in the model: the reflection problem, endogenous network formation, as well as endogenous sorting of characteristics. This relationship lets us demonstrate our next statement.

Proposition 3. Identification under simple Linear Endogeneity. *Suppose that Assumptions 1-6 hold with* $\xi \neq 0$. *If all matrices* $\mathbf{I}$, $\mathbf{G}$, $\mathbf{C}^{-1}$, $\mathbf{G}\mathbf{C}^{-1}$ *and* $\mathbf{G}^2\mathbf{C}^{-1}$ *are linearly independent, the parameters* $(\alpha, \beta, \gamma, \xi, \psi)$ *are all identified from the data.*

Proof. This proof is analogous to those given by Bramoullé et al. (2009). Specifically, observe from (8) and from Assumption 6.c that any two structures $(\alpha, \beta, \gamma, \xi, \psi)$ and $(\alpha', \beta', \gamma', \xi', \psi')$ are observationally equivalent only if:

$$(\mathbf{I} - \beta'\mathbf{G})\,\alpha = (\mathbf{I} - \beta\mathbf{G})\,\alpha'$$

$$(\mathbf{I} - \beta'\mathbf{G})\,\gamma = (\mathbf{I} - \beta\mathbf{G})\,\gamma'$$

$$\xi^{-1}\,(\mathbf{I} - \beta'\mathbf{G})\,(\mathbf{I} + \psi\mathbf{G})\,\mathbf{C}^{-1} = \xi'^{-1}\,(\mathbf{I} - \beta\mathbf{G})\,(\mathbf{I} + \psi'\mathbf{G})\,\mathbf{C}^{-1}$$

by summing equations and tediously rearranging terms it appears evident that the three equations can hold simultaneously only if $(\alpha, \beta, \gamma, \xi, \psi) = (\alpha', \beta', \gamma', \xi', \psi')$. □

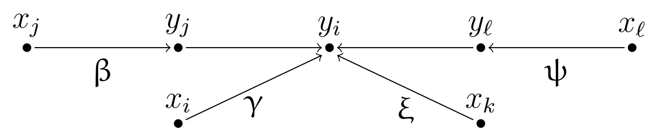
This result states that social effects are identified in a SAR model even if individual characteristics linearly reflect individual unobservables, under some broad conditions. In particular, it is required that $\xi \neq 0$ (clearly if $\xi = 0$ there is no endogeneity and the identification result would still hold) and the network $\mathbf{G}$ relates to the characteristics matrix $\mathbf{C}$ in the way expressed by the statement of the proposition. These conditions are very general: they require that the process of sorting in the network, which affects the cross-correlation of the unobservables, somehow "differs enough" from sorting in the x_i's, as it is discussed above. It is illustrative to describe a particular case where this condition does not hold. Suppose that $\mathbf{C}^{-1} \propto (\mathbf{I} + \psi\mathbf{G})$ or, equivalently, $\boldsymbol{x} \propto \boldsymbol{\varepsilon}$, that is the individual characteristics proportionately reflect the unobservables, as in a textbook model of omitted variable bias. Then, (8) would collapse to:

$$\mathbb{E}\left[\mathbf{y}\middle|\,\mathbf{x}\right] \propto \xi^{-1}\,(\mathbf{I} - \beta\mathbf{G})^{-1}\left[-\alpha\boldsymbol{\varpi} + \gamma\mathbf{x}\right]$$

and identification fails. This makes it clear that what makes identification possible in this context is the partially-overlapping nature of the two kinds of social correlations.

A theoretical case in which $\boldsymbol{x}_k \propto \boldsymbol{\varepsilon}$ (for a higher number of observable characteristics $\boldsymbol{x}_k$, $k = 1, \ldots, K$) is provided by basic models of firm optimization, which predict that the unobserved productivity shock is reflected proportionally by all inputs, giving rise to the so-called "transmission bias."

We again find it useful to express this intuition through a graph.



Graph 3: Identification of SAR models under linear endogeneity

Graph 3 represents the structural relationships between the x and y variables of four observations (i, j, k, ℓ) like those depicted in Graph 2. Consider that if $\mathbf{C}$ has a group structure and is also invertible, $\mathbf{C}^{-1}$ would retain the same group structure. Hence, while $g_{ik} = 0$ because nodes i and k are not connected, they belong to the same group, thus $c'_{ik} \neq 0$ (where c'_{ik} can be defined as the i-k-th entry of matrix $\mathbf{C}^{-1}$). Because x_i and x_k are correlated by assumption there exists an indirect structural correlation between x_k and y_i, which is what allows to identify ξ. Now consider nodes i and ℓ: they do *not* belong to the same group, but they are directly connected. Hence, the correlation between their unobservables is reflected through the correlation between their outcomes, which is what permits the identification of ψ. Once these correlations are accounted for, the identification of γ and β would logically proceed as in a simple SAR model: the former through i's characteristics x_i, while the latter through the x_j of another connected observation j. Note that to facilitate interpretation, in Graph 2 node ℓ is made representing an agent who is connected to i but belongs to a different group. However, this is not strictly necessary for identification: for the linear independence conditions to be fulfilled, $\mathbf{G}$ and $\mathbf{C}$ may well have the same non-zero entries (represent the same network) but with different internal weights.[8]

To complete the discussion about identification under simple linear endogeneity, it is useful again to draw a parallel between a SAR model and simultaneous equations

[8] This is so if, for example, $\mathbf{C} = \mathbf{I} + \psi_1 \mathbf{G}$ where $\psi_1 \neq \psi$. The economic interpretation is that the process of network formation and sorting relates to the observable and unobservable characteristics with different intensities: for example, friendships may be more likely to be determined by individual preferences and ability rather than from socio-economic background. Note that in such a scenario, to obtain separate identification of the other parameters it is necessary that ψ_1 be known a-priori (Proposition 3 does not extend to the identification of ψ_1).

models (SEMs). It is well known from the traditional analysis of SEMs that covariance restrictions can aid identification if linear restrictions are not enough to satisfy both the order and rank conditions. In an extension of the "exogenous x's" case, where the network allows to generate enough restrictions to identify social effects, the partial network overlap between the covariance structure of the observable characteristics x and that of the unobservables ε implicitly generates a set of covariance restrictions, involving different pairs of observations depending on their position in the network and in the characteristic matrix, which allow to identify the data generation process.

3.3 Linear Endogeneity: General Result

Having discussed the intuition for identification under relatively simple hypotheses, we are now ready to illustrate our main result. We are interested in the identification of the parameters of the following model, which extends (6):

$$\mathbf{y} = \alpha \boldsymbol{\iota} + \beta \mathbf{G}\mathbf{y} + \mathbf{X}\boldsymbol{\gamma} + \mathbf{G}\mathbf{X}\boldsymbol{\delta} + \boldsymbol{\varepsilon} \tag{9}$$

where $\mathbf{X}$ is a $N \times K$ data matrix of K observable characteristics (with $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ having full rank), $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_K)$ is the vector of K direct effects associated with each of these characteristics, and $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_K)$ are the K related "contextual effects." In Elhorst's taxonomy, this is a standard multivariate "Spatial Durbin Model" (SDM), otherwise known – under row-normalization of $\mathbf{G}$ – as a linear-in-means model. Note that such a model could easily follow from an extension of our theoretical framework, in which nature initially draws $(\mathbf{X}, \boldsymbol{\varepsilon})$ from $\mathcal{F}(\cdot)$. To keep the problem interesting we presume that all observable characteristics $\mathbf{X}$ are potentially endogenous and structurally dependent on $\boldsymbol{\varepsilon}$, or else single exogenous factors could be enough to identify β under the logic of Graph 1. In addition, we introduce the following assumptions.

Assumption 7. SARMA Unobservables: *the unobservable characteristics follow a stationary Spatial Autoregressive Moving Average process where the autoregressive component has order S:*

$$\boldsymbol{\varepsilon} = (\mathbf{I} - \psi_1 \mathbf{F}_1 - \psi_2 \mathbf{F}_2 - \cdots - \psi_p \mathbf{F}_S)^{-1} \mathbf{E}\boldsymbol{\upsilon}$$

where $\mathbf{E}$ is some unrestricted $N \times N$ matrix, $(\mathbf{F}_1, \mathbf{F}_2, \ldots, \mathbf{F}_S)$ are $N \times N$ matrices such that $\mathbf{I} + \sum_{s=1}^{s'} \mathbf{F}_s$ is invertible for all $s' \leq S$, and $\boldsymbol{\upsilon}$ conforms to Assumption 4.

Assumption 8. Diffused Shocks Separable in $\boldsymbol{x}$: *for $k = 1, \ldots, K$:*

$$\boldsymbol{x}_k = \boldsymbol{\omega}_k + \xi_k \mathbf{C}_k \boldsymbol{\upsilon}, \quad \xi_k \in \mathbb{R}$$

where $\boldsymbol{\omega}_k$ is a random vector that possesses the same properties as $\boldsymbol{\omega}$ from Assumption 6 and $\mathbf{C}_k$ is the characteristics matrix for the k-th observable covariate $\boldsymbol{x}_k$.

Assumption 7 allows for very general linear patterns of cross-correlation in the unobservables ε. On the one hand, the autoregressive component is allowed to be of any order S; on the other hand the moving average component is also quite flexible, since matrix $\mathbf{E}$ is left unrestricted. Assumption 8 instead is a simple extension of Assumption 6 to the multivariate case. Note that to each k-th observable characteristic is associated a different invertible matrix $\mathbf{C}_k$, which may be a complex function of the network or of some group structure. The parameters featured in these assumptions can be collected as the vectors $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_K)$ and $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_K)$. We are now ready to state our main result.

Theorem 1. General Identification Result. *Under Assumptions 1-4 and 7-8, if it holds that $\beta\gamma_k + \delta_k \neq 0$, $\xi_k \neq 0$ for every $k = 1, \ldots, K$ and the matrices in the set*

$$\mathbb{G}_k \equiv \left\{\mathbf{I}, \mathbf{G}, \mathbf{G}^2, \mathbf{E}\mathbf{C}_k^{-1}\right\} \cup \{\mathbf{F}_1, \mathbf{F}_2, \ldots, \mathbf{F}_S\}$$

are linearly independent for every $k = 1, \ldots, K$, then $(\alpha, \beta, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\xi}, \boldsymbol{\psi})$ are identified.

Proof. See the Appendix. The proof makes use of a linear algebra result by Henderson and Searle (1981), but is otherwise a generalization of Proposition 3. □

Theorem 1 provides a general identification result for SAR models (possibly featuring contextual effects) such that the observable and unobservable characteristics are structurally dependent. The conditions under which identification is obtained are quite general, since a large class of cross-correlation structures in the unobservables are allowed by Assumption 7. Identification is obtained if for each covariate, the set $\mathbb{G}_k$ is composed of linearly independent matrices. In economic terms this means, along the lines of the intuition illustrated in the previous discussion, that the network must have a sufficiently small degree of overlap with the social structure that determines the correlation of the characteristics $\boldsymbol{x}_k$. For example, if the characteristic matrix $\mathbf{C}$ is the same for all covariates, this condition is usually satisfied if $\mathbf{C}$ follows a group structure while the $\mathbf{F}_p$ and $\mathbf{E}$ matrices are some function of the network $\mathbf{G}$.

3.4 Identification of μ and ν

In our framework, parameter β – which in studies of peer effects is usually associated with the concept of social multiplier – represents a composite equilibrium effect: it equals the direct effect of peers' effort on the individual outcome ν multiplied by a term corresponding to the equilibrium response of individual effort $(1-\mu)^{-1}$. Under the hypotheses maintained so far, particularly Assumption 3 (row-normalization of $\mathbf{G}$) the two parameters μ and ν disappear from the reduced form equilibrium equation. However, note that dropping the row-normalization hypothesis, (6) would read as:

$$\mathbf{y} = (\alpha - \vartheta)\,\iota + \vartheta\bar{\mathbf{g}} + \beta\mathbf{G}\mathbf{y} + \gamma\mathbf{x} + \boldsymbol{\varepsilon} \tag{10}$$

where $\vartheta \equiv (1-\mu)^{-1}\nu\log\mu$ and $\bar{\mathbf{g}} \equiv \mathbf{G}\iota$ is the vector of individual in-degrees (the overall strength of all one individual's connections, such that $\bar{g}_i = \sum_{j=1}^{N} g_{ij}$).

Since $\exp(\theta/\beta) = \mu$, if the observable characteristics x_i's are exogenous the primitive parameters μ and ν are separately identified in (10). While the exact relationship between structural and reduced form parameters depends upon the functional form assumptions of our model, the associated economic intuition is straightforward: the variation in individual in-degree $\bar{\mathbf{g}}$ conveys additional information about the overall strength of direct spillovers (expressed by the parameter ν). An individual with more friends or a firm with more connections is likely to enjoy more beneficial externalities. While row-normalization is routinely assumed in studies of peer effects, we find this to be a realistic hypothesis worth of being empirically tested.

The obvious potential empirical issue is that individual in-degree $\bar{g}_i$ may be itself endogenous and dependent on the (spatially correlated) unobservables. Intuitively, a very skilled pupil or a very successful firm may find themselves with more or more intense connections. Yet our identification results are easily extended to a framework with a non row-normalized adjacency matrix.

Corollary to Theorem 1: *Under the conditions expressed by Theorem 1, except for Assumption 3 (row-normalization), if $\bar{\mathbf{g}}$ is linearly independent from ι then parameters μ and ν are separately identified.*

The intuition is that if the spatial processes that drive the correlations of $\boldsymbol{x}$ and $\boldsymbol{\varepsilon}$ do not overlap, it is possible to effectively control for the sources of endogeneity; the reduced form "effect" of $\bar{\mathbf{g}}$ can be retrieved through the covariance restrictions implied by the model. In the next section we illustrate how this idea can be subsumed within the set of moment conditions employed in our proposed estimation procedure.[9]

4 Estimation

We are now ready to illustrate the GMM estimator that we propose for the estimation of the SDM – linear-in-means – model (9), and under linear endogeneity. To simplify the problem, we restrict our attention to unobservables following a SMA(1) process, although we allow the matrix that characterizes the process to be different from $\mathbf{G}_N$.

Assumption 9. SMA(1) Unobservables (general): $\boldsymbol{\varepsilon} = (\mathbf{I} + \psi \mathbf{E}_N) \boldsymbol{\upsilon}$ *where* $\mathbf{E}_N$ *is some invertible, row-normalized* $N \times N$ *matrix such that* $\mathbf{E}_N = \mathbf{E}_N \boldsymbol{\iota}$*, and* $|\psi| < 1$.

We also assume that $\alpha = 0$, that $\mathbf{X}$ possibly features a constant vector, and that the problem is homoscedastic, once again with the objective of simplifying the analysis.

Assumption 10. Homoscedasticity: *the conditional variance of the diffused shocks is homoscedastic:* $\mathbb{E}\left[\upsilon_i^2 \middle| \mathbf{x}_i\right] = \sigma^2$ *for all* $i = 1, \dots, N$.

Define $\boldsymbol{\theta}_0 \equiv (\beta_0, \boldsymbol{\gamma}_0, \boldsymbol{\delta}_0, \boldsymbol{\xi}_0, \psi_0, \sigma_0^2)$ as the vector of true parameter values. We construct a GMM estimator based on a set of moment conditions, which we subdivide in two blocks. The first block is constituted by a sequence of $(1+Q)\,K$ bias-adjusted standard linear moments:

$$\boldsymbol{m}_{1,N}\left(\boldsymbol{\theta}_0\right) = \mathbb{E}\left[\boldsymbol{m}_{1,i}^*\left(\boldsymbol{\theta}_0\right)\right] - \boldsymbol{\mu}_{1,N}\left(\boldsymbol{\theta}_0\right) = \mathbf{0} \tag{11}$$

where:

$$\overline{\boldsymbol{m}}_{1,N}^* \equiv \sum_{i=1}^{N} \boldsymbol{m}_{1,i}^*\left(\boldsymbol{\theta}_0\right) \equiv \begin{bmatrix} \mathbf{Q}_{0,N} \\ \mathbf{Q}_{1,N} \\ \mathbf{Q}_{2,N} \\ \vdots \\ \mathbf{Q}_{Q,N} \end{bmatrix} \boldsymbol{\varepsilon}\left(\boldsymbol{\theta}_0\right) \equiv \begin{bmatrix} \mathbf{X}^{\mathrm{T}} \\ \mathbf{X}^{\mathrm{T}}\mathbf{G}_N \\ \mathbf{X}^{\mathrm{T}}\mathbf{G}_N^2 \\ \vdots \\ \mathbf{X}^{\mathrm{T}}\mathbf{G}_N^Q \end{bmatrix} \boldsymbol{\varepsilon}\left(\boldsymbol{\theta}_0\right)$$

[9] If individual "effort" e_i is observable, an alternative route for the separate identification of μ and ν would be based on the structural "production function" (2): this is indeed the approach taken in studies of R&D spillovers, since researchers can typically observe the R&D expenditures of firms. However, our procedure cannot be directly extended to firm production functions, because the nature of endogeneity is likely to violate our identification conditions, at least if firm optimization proceeds according to standard "static" First Order Conditions ($\boldsymbol{x}_k \propto \boldsymbol{\varepsilon}$ for every input k).

which implicitely defines a set of Q matrices $\mathbf{Q}_{q,N}$ for $q = 0, \ldots, Q$, and:

$$\boldsymbol{\mu}_{1,N}(\theta_0) \equiv \sigma\xi \begin{bmatrix} \mathbf{t}_{0,N} \\ \mathbf{t}_{1,N} \\ \mathbf{t}_{2,N} \\ \vdots \\ \mathbf{t}_{Q,N} \end{bmatrix}$$

and where $\mathbf{t}_{q,N} = \left[\mathrm{Tr}\left[\mathbf{C}_{1,N}^{\mathrm{T}}\mathbf{G}_N^q(\mathbf{I}+\psi\mathbf{E}_N)\right], \ldots, \mathrm{Tr}\left[\mathbf{C}_{K,N}^{\mathrm{T}}\mathbf{G}_N^q(\mathbf{I}+\psi\mathbf{E}_N)\right]\right]^{\mathrm{T}}$ is a $K \times 1$ vector for $q = 0, \ldots, Q$. The second block is a set of $P + 1$ covariance restrictions:

$$\boldsymbol{m}_{2,N}(\theta_0) = \mathbb{E}\left[\boldsymbol{m}_{2,i}^*(\theta_0)\right] - \boldsymbol{\mu}_{2,N}(\theta_0) = \mathbf{0} \tag{12}$$

where:

$$\overline{\boldsymbol{m}}_{2,N}^* \equiv \sum_{i=1}^{N} \boldsymbol{m}_{2,i}^*(\theta_0) \equiv \begin{bmatrix} \boldsymbol{\varepsilon}(\theta_0)^{\mathrm{T}}\mathbf{P}_{0,N} \\ \boldsymbol{\varepsilon}(\theta_0)^{\mathrm{T}}\mathbf{P}_{1,N} \\ \boldsymbol{\varepsilon}(\theta_0)^{\mathrm{T}}\mathbf{P}_{2,N} \\ \vdots \\ \boldsymbol{\varepsilon}(\theta_0)^{\mathrm{T}}\mathbf{P}_{P,N} \end{bmatrix} \boldsymbol{\varepsilon}(\theta_0) \equiv \begin{bmatrix} \boldsymbol{\varepsilon}(\theta_0)^{\mathrm{T}}\mathbf{I} \\ \boldsymbol{\varepsilon}(\theta_0)^{\mathrm{T}}\mathbf{G}_N \\ \boldsymbol{\varepsilon}(\theta_0)^{\mathrm{T}}\mathbf{G}_N^2 \\ \vdots \\ \boldsymbol{\varepsilon}(\theta_0)^{\mathrm{T}}\mathbf{G}_N^P \end{bmatrix} \boldsymbol{\varepsilon}(\theta_0)$$

which here implicitly defines a set of P matrices $\mathbf{P}_{p,N}$ for $p = 0, \ldots, P$, and:

$$\boldsymbol{\mu}_{2,N}(\theta_0) \equiv \sigma^2 \begin{bmatrix} \mathrm{Tr}\left[(\mathbf{I}+\psi\mathbf{E}_N)^{\mathrm{T}}(\mathbf{I}+\psi\mathbf{E}_N)\right] \\ \mathrm{Tr}\left[(\mathbf{I}+\psi\mathbf{E}_N)^{\mathrm{T}}\mathbf{G}_N(\mathbf{I}+\psi\mathbf{E}_N)\right] \\ \mathrm{Tr}\left[(\mathbf{I}+\psi\mathbf{E}_N)^{\mathrm{T}}\mathbf{G}_N^2(\mathbf{I}+\psi\mathbf{E}_N)\right] \\ \vdots \\ \mathrm{Tr}\left[(\mathbf{I}+\psi\mathbf{E}_N)^{\mathrm{T}}\mathbf{G}_N^P(\mathbf{I}+\psi\mathbf{E}_N)\right] \end{bmatrix}$$

so that the two blocks can be compactly written as follows.

$$\boldsymbol{m}_N(\theta_0) = \mathbb{E}\left[\boldsymbol{m}_i^*(\theta_0)\right] - \boldsymbol{\mu}_{1,N}(\theta_0) = \mathbf{0} \tag{13}$$

where $\boldsymbol{m}_i^*(\theta_0)$ and $\boldsymbol{\mu}_N(\theta_0)$ result from vertically stacking the two blocks of moments.

To give intuition, consider the case of a simple endogenous SAR with one covariate such as (6). Suppose that the econometrician were able to observe the realizations of $\boldsymbol{\omega}$, and write them as $\boldsymbol{\omega} = (\omega_1, \dots, \omega_N)^{\mathrm{T}}$. Thus, a way to consistently estimate α, β and γ would be to estimate a model analogous to (7) such as the one that follows.

$$\mathbf{y} = (\mathbf{I} - \beta \mathbf{G}_N)^{-1} (\alpha \iota + \gamma \boldsymbol{\omega} + \varepsilon) \simeq \sum_{r=1}^{\infty} \beta^r \mathbf{G}_N^r (\alpha \iota + \gamma \boldsymbol{\omega} + \varepsilon)$$

The problem is that the econometrician does not usually observe $\boldsymbol{\omega}$. However, under the conditions expressed by Theorem 1 it is possible to recover the covariance structure of the model; once that is obtained, $\boldsymbol{\omega}$ can be disentangled from the variation of $\mathbf{x}$ that is due to the diffused shocks; this logic easily extends to a multivariate contexts. Our GMM estimator simultaneously executes both tasks: while the covariance restrictions serve the purpose of estimating of the model's variance structure, the first block of moments estimates the linear parameters of the SDM model.

We define our GMM estimator $\widehat{\theta}_{GMM}$ as the usual minimizer in the parameter space $\boldsymbol{\Theta}$:

$$\widehat{\theta}_{GMM} = \arg\min_{\theta \in \boldsymbol{\Theta}} \overline{\mathbf{m}}_N^{\mathrm{T}} (\theta) \mathbf{W}_N \overline{\mathbf{m}}_N (\theta) \tag{14}$$

where $\overline{\mathbf{m}}_N (\theta) = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{m}_i^* (\theta) - \boldsymbol{\mu}_N (\theta)$ and $\mathbf{W}_N$ is a weighting matrix. We derive the asymptotic properties of the estimator under the following additional assumptions.

Assumption 11. Bounded Parameter Space: $\boldsymbol{\Theta}$ *is bounded.*

Assumption 12. Probability Limits of the Covariates: *the independent component of $\boldsymbol{x}_k$ are such that $N^{-1} \iota^{\mathrm{T}} (\boldsymbol{\omega}_k - \mathbb{E} [\boldsymbol{\omega}_k]) = o_{\mathcal{P}} (1)$ for all $k = 1, \dots, K$.*

Assumptions 11 and 12 are regularity conditions that are necessary to ensure consistency of the GMM estimator.

Assumption 13. Bounded Adjacencies: *the network's "adjacency" matrix $\mathbf{G}_N$ and its corresponding Leontiev inverse $(\mathbf{I} - \beta_0 \mathbf{G}_N)^{-1}$ are uniformly bounded in both row and column sums in absolute value.*

Assumption 14. Bounded Characteristics: *the characteristics matrix $\mathbf{C}_k$ is bounded by $\overline{C}_k < \infty$ for every $k = 1, \dots, K$, that is $\sum_{j=1}^{N} \mathrm{c}_{ij} < \overline{C}_k$ for $i = 1, \dots, N$.*

Assumption 15. Bounded Moments: *the matrices $(\mathbf{Q}_{0,N}, \mathbf{Q}_{1,N}, \dots, \mathbf{Q}_{Q,N})$ and $(\mathbf{P}_{0,N}, \mathbf{P}_{1,N}, \dots, \mathbf{P}_{P,N})$ used in the moment conditions are all uniformly bounded in both row and column sums in absolute value.*

Assumptions 13-15 all ensure that the relevant moments have finite variance. Thus, the asymptotic results are expressed as follows.

Theorem 2. Asymptotics of the GMM estimator. *Under Assumptions 1-4 and 8-15,* $\widehat{\theta}_{GMM}$ *is a consistent estimator of* θ_0 *and has the following limiting distribution:*

$$\sqrt{N}\left(\widehat{\theta}_{GMM}-\theta_0\right)\overset{d}{\to}\mathcal{N}\left(\mathbf{0},\left[\mathbf{J}_0^{\mathrm{T}}\mathbf{W}_0\mathbf{J}_0\right]^{-1}\mathbf{J}_0^{\mathrm{T}}\mathbf{W}_0\boldsymbol{\Omega}_0\mathbf{W}_0\mathbf{J}_0\left[\mathbf{J}_0^{\mathrm{T}}\mathbf{A}_0\mathbf{J}_0\right]^{-1}\right)$$

where here: (i) $\boldsymbol{\Omega}_0\equiv\operatorname{plim}\frac{1}{N}\mathbb{V}\mathrm{ar}\left[\boldsymbol{m}_N\left(\theta_0\right)\right]$*;* (ii) $\mathbf{J}_0\equiv\operatorname{plim}\frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial\theta^{\mathrm{T}}}\mathbb{E}\left[\boldsymbol{m}_N\left(\theta_0\right)\right]$*; and* (iii) $\mathbf{W}_0\equiv\operatorname{plim}\mathbf{W}_N$.

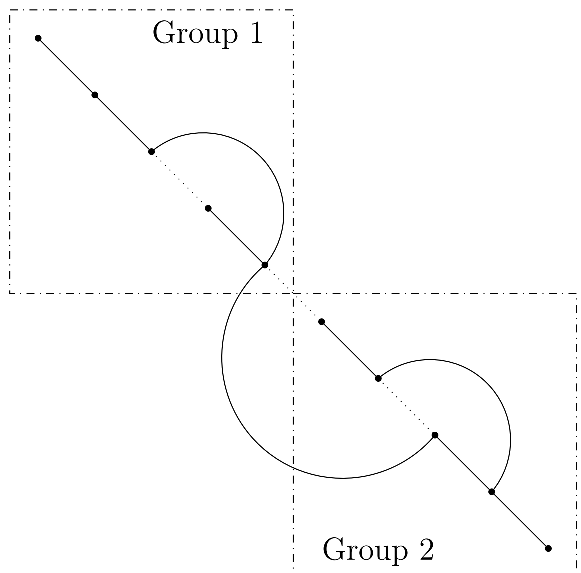
Proof. See the Appendix. The proof makes use of results by White (1994), Kelejian and Prucha (2001) and Lee (2007). □

Observe that if the adjacency matrix $\mathbf{G}_N$ is not normalized (that is, Assumption 3 fails), the GMM estimator can easily be adapted for the separate estimation of the primitive parameters μ and ν. Since under the maintained assumptions $\mathbb{E}\left[\bar{\mathbf{g}}^{\mathrm{T}}\boldsymbol{\omega}_k\right]=0$ for all k, an appropriate moment condition can be employed for the estimation of ϑ, and both μ and ν are recovered later as Minimum Distance estimates; alternatively, the entire GMM problem can be accordingly rephrased in terms of these parameters. Further extensions of this setup appear evident: for example, heteroscedasticity can be introduced by parametrizing σ^2 as a function of the covariates, while the covariance restrictions may be adapted to allow for more general SARMA processes of the error term. All these extensions deserve some dedicated analysis in future work.

5 Monte Carlo

We find it useful to evaluate the performance of our GMM estimator through Monte Carlo simulations. In particular, we simulate a minimal d.g.p.: the SAR model (6) with one covariate and no contextual effects ($\delta=0$), combined with the simple setup of linear endogeneity expressed by Assumptions 4-6 (thus, $\mathbf{G}=\mathbf{E}$). In all simulations we set $N=500$; moreover we construct a homogeneous, block-diagonal characteristics matrix $\mathbf{C}$ which – in our baseline case – is composed by 50 "groups" of size 10. We generate a new matrix $\mathbf{G}$ in each repetition of every simulation, in order to minimize the dependence of our results from a specific network matrix. Specifically, each matrix $\mathbf{G}$ is randomly generated through the 'small-world' algorithm by Watts and Strogatz

(1998); by this procedure, all observations are first ordered along a line and connected to an even number of κ neighbors; next, links are reshuffled with some probability π (connections are unweighted, that is $g_{ij} \in \{0, 1\}$). Given that the initial ordering of observations corresponds with the one used for defining the characteristics matrix, **G** and **C** are guaranteed to have some degree of overlap, although not a complete one. We represent this through the following graphical example.



Graph 4: Partial overlap of **C** and **G**: Example

In Graph 4, 10 nodes are ordered along a line, and split in two symmetrical groups – each of size 5 – which characterize **C**. Through a small-world algorithm with $\kappa = 2$, all nodes are connected in the network with their immediate neighbors on the line, but three links are eventually reshuffled so that the resulting matrix **G** is irregular.

In our baseline simulation, we set the following parameters:[10]

$$(\alpha_0, \beta_0, \gamma_0, \xi_0, \psi_0, \sigma_0) = (.25, .4, .5, .1, .25, .05)$$

note that ψ_0 amounts to five times the standard deviation of the diffused shocks, which results in the introduction of substantial endogeneity into the model. In addition, we set $\kappa = 2$ and $\pi = 0.25$ in the network-generation algorithm. Over 1,000 repetitions, we estimate our model with equally-weighted moment conditions of order $Q = 2$ and $P = 1$; we also compare our estimates of (α, β, γ) with those obtained via OLS as well

[10]Furthermore, we set $\mathbb{V}\text{ar}[\omega_i] = 0.09$, but we are not interested in estimating this parameter.

as with an IV estimator such that $\mathbf{Gx}$ is used as an instrument for $\mathbf{Gy}$. Finally, we repeat the simulation by tuning certain parameters differently relative to the baseline. The results are reported in Tables 1 and 2.

Table 1: Monte Carlo Simulations (part one)

	Baseline			$\beta = 0.50$		
	PFZ	IV	OLS	PFZ	IV	OLS
α	0.256	0.082	0.044	0.254	0.081	0.041
	(0.040)	(0.018)	(0.015)	(0.038)	(0.020)	(0.015)
β	0.385	0.802	0.894	0.492	0.839	0.918
	(0.095)	(0.044)	(0.035)	(0.075)	(0.039)	(0.029)
γ	0.496	0.288	0.229	0.497	0.292	0.232
	(0.050)	(0.034)	(0.029)	(0.050)	(0.035)	(0.029)
ξ	0.100	–	–	0.100	–	–
	(0.013)			(0.013)		
ψ	0.270	–	–	0.260	–	–
	(0.010)			(0.090)		
σ	0.050	–	–	0.050	–	–
	(0.005)			(0.004)		

	$\gamma = 0.2$			$\psi = 0$		
	PFZ	IV	OLS	PFZ	IV	OLS
α	0.252	-0.144	-0.048	0.253	0.046	0.036
	(0.065)	(0.056)	(0.017)	(0.036)	(0.024)	(0.019)
β	0.395	1.347	1.115	0.392	0.891	0.913
	(0.156)	(0.133)	(0.042)	(0.086)	(0.057)	(0.046)
γ	0.188	-0.133	-0.040	0.499	0.190	0.176
	(0.042)	(0.064)	(0.027)	(0.053)	(0.046)	(0.038)
ξ	0.100	–	–	0.996	–	–
	(0.013)			(0.107)		
ψ	0.267	–	–	0.037	–	–
	(0.170)			(0.048)		
σ	0.050	–	–	0.050	–	–
	(0.005)			(0.004)		

Note. Every column reports the median and the standard deviation (in parentheses) of the relevant parameter estimates across 1000 repetitions. 'PFZ' indicates our proposed procedure, 'IV' the estimator obtained by instrumenting $\mathbf{Gy}$ with $\mathbf{Gx}$, while 'OLS' is self-explanatory.

Table 2: Monte Carlo Simulations (part two)

	$\xi = 0$			Group Size: 5		
	PFZ	IV	OLS	PFZ	IV	OLS
α	0.254	0.251	0.104	0.251	0.211	0.179
	(0.044)	(0.028)	(0.018)	(0.016)	(0.009)	(0.011)
β	0.389	0.398	0.749	0.397	0.494	0.571
	(0.106)	(0.067)	(0.044)	(0.038)	(0.022)	(0.026)
γ	0.499	0.498	0.303	0.497	0.292	0.232
	(0.055)	(0.043)	(0.037)	(0.050)	(0.035)	(0.029)
ξ	0.004	–	–	0.101	–	–
	(0.005)			(0.026)		
ψ	0.259	–	–	0.245	–	–
	(0.134)			(0.053)		
σ^2	0.050	–	–	0.050	–	–
	(0.002)			(0.002)		

	$\kappa = 4$			$\pi = 0.9$		
	PFZ	IV	OLS	PFZ	IV	OLS
α	0.263	0.027	-0.012	0.242	0.116	0.099
	(0.056)	(0.037)	(0.026)	(0.038)	(0.016)	(0.013)
β	0.368	0.935	1.029	0.420	0.723	0.762
	(0.135)	(0.088)	(0.061)	(0.092)	(0.037)	(0.032)
γ	0.506	0.226	0.173	0.491	0.418	0.402
	(0.060)	(0.058)	(0.044)	(0.034)	(0.025)	(0.023)
ξ	0.102	–	–	0.093	–	–
	(0.016)			(0.021)		
ψ	0.273	–	–	0.228	–	–
	(0.140)			(0.144)		
σ^2	0.050	–	–	0.049	–	–
	(0.003)			(0.004)		

Note. See the notes for Table 1.

In our baseline simulations, our proposed estimator appears to be quite accurate. While it slightly underestimates β (on average) it contrasts with both OLS and IV estimators, which estimate β about twice as large. We obtain similar results when we set different values of β or γ, or if diffused shocks are made coincident with the error term ($\psi = 0$). If we silence the characteristics matrix channel ($\xi = 0$) IV becomes consistent; however, it behaves similarly as our proposed GMM estimator. The more interesting implications are obtained by altering the parameters that define matrices **C** and **G**. By halving the size of groups in the characteristics matrix, endogeneity is reduced; however, our GMM method still provides accurate estimates, unlike IV or OLS. Increasing the density of **G** (by setting $\kappa = 4$) does not seem to significantly affect the simulated estimates; however, increasing the randomness of links ($\pi = 0.9$) results in β to be slightly overestimated (instead of underestimated) on average. To summarize, it appears that our GMM method – while consistent and preferable to the standard IV estimator – is biased in small samples, in a way that depends on the characteristics of the underlying networks.

6 Conclusion

In this paper we have shown that, under certain configurations of the underlying socio-economic relationships that determine the characteristics and relevant outcomes of economic agents, it is possible to identify and estimate peer or social effects within a standard spatial econometric framework, even if the right-hand side characteristics are themselves endogenous. In fact, the requirements for identification are quite general: it suffices that the spatial correlation of unobservables and individual characteristics does not overlap in the relevant metric space. Our running example of such a setting is a schooling context in which students establish friendships on the basis of certain unobservables; however, their socio-economic characteristics (such as socio-economic status) correlate even among non-friends within certain observable groups (like school classes). In future work, we plan on implementing our GMM methodology to such an empirical application; in addition, we aim at extending it to more general functional forms of the underlying cross-correlations that give rise to endogeneity.

Our contribution is relevant for the developing literature about the econometrics of social effects, which is currently focused on ways to address the problem of network endogeneity. Our approach is more general than that, since it allows for both network and covariates' endogeneity – which, in fact, are two sides of the same coin. Paradoxically, our work also speaks to the recent literature about the identification of interaction networks when these are unknown, under the assumption that individual covariates are exogenous with respect to the error term. Quite the contrary, the application of our proposed estimator requires the researcher's knowledge of the underlying patterns of cross-correlation, for both the unobservables and the covariates. In future work, it would be interesting to analyze how these two approaches can be interacted, say if some form of endogeneity is present in the relevant application but the researcher has access to only partial information about its actual shape.

References

Angrist, Joshua D. (2014) "The perils of peer effects," *Labour Economics*, Vol. 30, pp. 98–108.

Arduini, Tiziano, Eleonora Patacchini, and Edoardo Rainone (2015) "Parametric and Semiparametric IV Estimation of Network Models with Selectivity." Mimeo.

Azoulay, Pierre, Joshua S. Graff Zivin, and Jialan Wang (2010) "Superstar Extinction," *The Quarterly Journal of Economics*, Vol. 125, No. 2, pp. 549–589.

Bloom, Nicholas, Mark Schankerman, and John Van Reenen (2013) "Identifying Technology Spillovers and Product Market Rivalry," *Econometrica*, Vol. 81, No. 4, pp. 1347–1393.

Blume, Lawrence E., William A. Brock, Steven N Durlauf, and Rajshri Jayaraman (2015) "Linear Social Interactions Models," *Journal of Political Economy*, Vol. 123, No. 2, pp. 444–496.

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin (2009) "Identification of peer effects through social networks," *Journal of Econometrics*, Vol. 150, No. 1, pp. 41–55.

Calvó-Armengol, Antoni, Eleonora Patacchini, and Yves Zenou (2009) "Peer Effects and Social Networks in Education," *The Review of Economic Studies*, Vol. 76, No. 4, pp. 1239–1267.

Carrell, Scott E., Bruce I. Sacerdote, and James E. West (2013) "From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation," *Econometrica*, Vol. 81, No. 3, pp. 855–882.

Cliff, Andrew D. and John Keith Ord (1981) *Spatial Processes: Models and Applications*: Pion London.

Conley, Timothy G. and Christopher R. Udry (2010) "Learning about a new technology: Pineapple in Ghana," *The American Economic Review*, Vol. 100, No. 1, pp. 35–69.

Davezies, Laurent, Xavier D'Haultfoeuille, and Denis Fougère (2009) "Identification of peer effects using group size variation," *The Econometrics Journal*, Vol. 12, No. 3, pp. 397–413.

De Giorgi, Giacomo, Michele Pellizzari, and Silvia Redaelli (2010) "Identification of Social Interactions through Partially Overlapping Peer Groups," *American Economic Journal: Applied Economics*, Vol. 2, No. 2, pp. 241–275.

De Paula, Áureo, Imran Rasul, and Pedro Sousa (2018) "Recovering Social Networks from Outcome Data: Identification, Simulations and an Application." Mimeo.

Drukker, David M., Peter Egger, and Ingmar R. Prucha (2013) "On Two-Step Estimation of a Spatial Autoregressive Model with Autoregressive Disturbances and Endogenous Regressors," *Econometric Reviews*, Vol. 32, No. 5-6, pp. 686–733.

Duflo, Esther and Emmanuel Saez (2003) "The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment," *The Quarterly Journal of Economics*, Vol. 118, No. 3, pp. 815–842.

Elhorst, Paul J. (2014) *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*: Springer.

Glaeser, Edward L., Bruce I. Sacerdote, and José A. Scheinkman (1996) "Crime and Social Interactions," *The Quarterly Journal of Economics*, Vol. 11, No. 2, pp. 507–48.

Graham, Bryan S. (2008) "Identifying Social Interactions Through Conditional Variance Restrictions," *Econometrica*, Vol. 76, No. 3, pp. 643–660.

——— (2017) "An Econometric Model of Network Formation with Degree Heterogeneity," *Econometrica*. Forthcoming.

Henderson, Harold V. and Shayle R. Searle (1981) "On Deriving the Inverse of a Sum of Matrices," *SIAM Review*, Vol. 23, No. 1, pp. 53–60.

Jaffe, Adam B. (1986) "Technological opportunity and spillovers of R&D: evidence from firms' patents, profits and market value," *American Economic Review*, Vol. 76, No. 5, pp. 984–1001.

——— (1989) "Characterizing the technological position of firms, with application to quantifying technological opportunity and research spillovers," *Research Policy*, Vol. 18, No. 2, pp. 87–97.

Johnsson, Ida and Hyungsik Roger M. Moon (2017) "Estimation of Peer Effects in Endogenous Social Networks: Control Function Approach." Mimeo.

Kelejian, Harry H. and Ingmar R. Prucha (1998) "A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances," *The Journal of Real Estate Finance and Economics*, Vol. 17, No. 1, pp. 99–121.

——— (2001) "On the asymptotic distribution of the Moran I test statistic with applications," *Journal of Econometrics*, Vol. 104, No. 2, pp. 219–257.

——— (2007) "HAC estimation in a spatial framework," *Journal of Econometrics*, Vol. 140, No. 1, pp. 131–154.

——— (2010) "Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances," *Journal of Econometrics*, Vol. 157, No. 1, pp. 53–67.

Kuersteiner, Guido M. and Ingmar Prucha (2018) "Dynamic Spatial Panel Models: Networks, Common Shocks, and Sequential Exogeneity." Mimeo.

Lee, Lung-fei (2007) "Identification and estimation of econometric models with group interactions, contextual factors and fixed effects," *Journal of Econometrics*, Vol. 140, No. 2, pp. 333–374.

Lee, Lung-fei, Xiaodong Liu, and Xu Lin (2010) "Specification and estimation of social interaction models with network structures," *The Econometrics Journal*, Vol. 13, No. 2, pp. 145–176.

Lin, Xu and Lung-fei Lee (2010) "GMM estimation of spatial autoregressive models with unknown heteroskedasticity," *Journal of Econometrics*, Vol. 157, No. 1, pp. 34–52.

Liu, Xiaodong and Lung-fei Lee (2010) "GMM estimation of social interaction models with centrality," *Journal of Econometrics*, Vol. 157, No. 1, pp. 34–52.

Lychagin, Sergey, Joris Pinkse, Margaret E. Slade, and John Van Reenen (2016) "Spillovers in Space: Does Geography Matter?" *The Journal of Industrial Economics*, Vol. 64, No. 2, pp. 295–335.

Manresa, Elena (2017) "Estimating the Structure of Social Interactions Using Panel Data." Mimeo.

Manski, Charles F. (1993) "Identification of Endogenous Social Effects: The Reflection Problem," *The Review of Economic Studies*, Vol. 60, No. 3, pp. 531–542.

Mas, Alexandre and Enrico Moretti (2009) "Peers at Work," *American Economic Review*, Vol. 99, No. 1, pp. 112–145.

Moffitt, Robert A. (2001) "Policy Interventions, Low-Level Equilibria And Social Interactions," in Steven Durlauf and Peyton Young eds. *Social Dynamics*: MIT Press.

Oettl, Alexander (2012) "Reconceptualizing Stars: Scientist Helpfulness and Peer Performance," *Management Science*, Vol. 56, No. 6, pp. 1122–1140.

Pereda Fernández, Santiago (2017) "Social Spillovers in the Classroom: Identification, Estimation and Policy Analysis," *Economica*, Vol. 84, No. 336, pp. 712–747.

Rose, Christiern D. (2017a) "Identification of peer effects through social networks using variance restrictions," *The Econometrics Journal*, Vol. 20, No. 3, pp. S47–S60.

——— (2017b) "Identification of Spillover Effects using Panel Data." Mimeo.

Sacerdote, Bruce I. (2001) "Peer Effects with Random Assignment: Results for Dartmouth Roommates," *The Quarterly Journal of Economics*, Vol. 116, No. 2, pp. 681–704.

Tibshirani, Robert (1996) "Regression Shrinkage and Selection via the LASSO," *Journal of the Royal Statistical Society*, Vol. 58, No. 1, pp. 267–88.

Waldinger, Fabian (2011) "Peer Effects in Science: Evidence from the Dismissal of Scientists in Nazi Germany," *Review of Economics Studies*, Vol. 72, No. 2, pp. 838–861.

Watts, Duncan J. and Steven H. Strogatz (1998) "Collective dynamics of 'small-world' networks," *Nature*, Vol. 393, No. 6684, pp. 440–442.

White, Halbert L. (1994) *Estimation, Inference and Specification Analysis*: Cambridge University Press.

Zacchia, Paolo (2017) "Knowledge Spillovers through Networks of Scientists." Mimeo.

Appendix – Mathematical Proofs

Proof of Theorem 1

Denote the k-th column of $\mathbf{X}$ as $\mathbf{x}_k^*$, and write the conditional expectation of $\mathbf{y}$ given $\mathbf{X}$ as:

$$\begin{aligned}\mathbb{E}\left[\mathbf{y}\right|\mathbf{X}] = (\mathbf{I}-\beta\mathbf{G})^{-1}\sum_{k=1}^{K}[&\alpha\iota+\gamma_k\mathbf{x}_k^*+\delta_k\mathbf{G}\mathbf{x}_k^*+\\&+\xi_k^{-1}\left(\mathbf{I}-\psi_1\mathbf{F}_1-\psi_2\mathbf{F}_2-\cdots-\psi_p\mathbf{F}_S\right)^{-1}\mathbf{E}\mathbf{C}_k^{-1}\left(\mathbf{x}_k^*-\boldsymbol{\varpi}_k\right)]\end{aligned}$$

where $\boldsymbol{\varpi}_k = \mathbb{E}\left[\boldsymbol{\omega}_k | \mathbf{X}\right]$ has the same properties as $\boldsymbol{\varpi}$ in the one-covariate case (Assum. 6). Thus, given two alternative structures $(\alpha, \beta, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\xi}, \boldsymbol{\psi})$ and $(\alpha', \beta', \boldsymbol{\gamma}', \boldsymbol{\delta}', \boldsymbol{\xi}', \boldsymbol{\psi}')$, observational equivalence requires that $(\mathbf{I} - \beta'\mathbf{G})\,\alpha = (\mathbf{I} - \beta\mathbf{G})\,\alpha'$ and that, for each $k = 1, \dots, K$:

$$(\mathbf{I} - \beta'\mathbf{G})\,(\gamma_k \boldsymbol{\iota} + \delta_k \mathbf{G}) = (\mathbf{I} - \beta\mathbf{G})\,(\gamma'_k \boldsymbol{\iota} + \delta'_k \mathbf{G})$$

which – see Proposition 1 in Bramoullé et al. (2009) – is only possible if $\mathbf{I}$, $\mathbf{G}$ and $\mathbf{G}^2$ are linearly dependent, if $\beta\gamma_k + \delta_k = 0$, or if $(\beta, \gamma_k, \delta_k) = (\beta', \gamma'_k, \delta'_k)$; and that:

$$\begin{aligned}(\mathbf{I} - \beta'\mathbf{G})\,\xi_k^{-1}\,(\mathbf{I} - \psi_1\mathbf{F}_1 - \psi_2\mathbf{F}_2 - \dots - \psi_p\mathbf{F}_S)^{-1}\,\mathbf{E}\mathbf{C}_k^{-1} = \\ = (\mathbf{I} - \beta\mathbf{G})\,\xi_k'^{-1}\left(\mathbf{I} - \psi'_1\mathbf{F}_1 - \psi'_2\mathbf{F}_2 - \dots - \psi'_p\mathbf{F}_S\right)^{-1}\mathbf{E}\mathbf{C}_k^{-1}\end{aligned}$$

which under the hypotheses of the Theorem, is only possible if $(\xi, \boldsymbol{\psi}) = (\xi', \boldsymbol{\psi}')$. To see why, define $\boldsymbol{\Phi}_s \equiv (\mathbf{I} - \psi_1\mathbf{F}_1 - \psi_2\mathbf{F}_2 - \dots - \psi_p\mathbf{F}_s)$ for $s = 1, \dots, S$, and observe that by a matrix algebra result by Henderson and Searle (1981), for $s = 1, \dots, S-1$:

$$\boldsymbol{\Phi}_{s+1}^{-1} = \boldsymbol{\Phi}_s^{-1} - \psi_{s+1}\boldsymbol{\Phi}_s^{-1}\mathbf{F}_{s+1}\left(\mathbf{I} - \psi_{s+1}\mathbf{F}_{s+1}\boldsymbol{\Phi}_s^{-1}\right)\boldsymbol{\Phi}_s^{-1}$$

thus, applying this result recursively shows that under the hypotheses of the theorem $\boldsymbol{\Phi}_S^{-1}$ is uniquely determined by a set of linearly independent matrices $(\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_S)$ and a parameter vector $\boldsymbol{\psi}$. Combining all these considerations reveals that under the maintained hypotheses the two alternative structures above must be identical in order to deliver the same observations, which proves the theorem. Note that the associated corollary follows straightforwardly by including $\vartheta\overline{\mathbf{g}}$ in the regression function.

Proof of Theorem 2

First, let us index all the relevant quantities by N and introduce some notation:

$$\begin{aligned}\mathbb{E}\left[\mathbf{X}_N\right] &\equiv \left[\mathbb{E}\left[\boldsymbol{\omega}_{1,N}\right]\mathbf{C}_{1,N}\boldsymbol{\iota} \quad \mathbb{E}\left[\boldsymbol{\omega}_{2,N}\right]\mathbf{C}_{2,N}\boldsymbol{\iota} \quad \dots \quad \mathbb{E}\left[\boldsymbol{\omega}_{K,N}\right]\mathbf{C}_{K,N}\boldsymbol{\iota}\right] \\ \mathbb{E}\left[\mathbf{y}_N\right] &\equiv (\mathbf{I}_N - \beta_0\mathbf{G}_N)^{-1}\left(\mathbb{E}\left[\mathbf{X}_N\right]\boldsymbol{\gamma}_0 + \mathbf{G}_N\mathbb{E}\left[\mathbf{X}_N\right]\boldsymbol{\delta}_0\right) \\ \mathbf{d}_N(\boldsymbol{\theta}) &\equiv (\beta_0 - \beta)\,\mathbf{G}_N\mathbb{E}\left[\mathbf{y}_N\right] + \mathbb{E}\left[\mathbf{X}_N\right](\boldsymbol{\gamma}_0 - \boldsymbol{\gamma}) + \mathbf{G}_N\mathbb{E}\left[\mathbf{X}_N\right](\boldsymbol{\delta}_0 - \boldsymbol{\delta})\end{aligned}$$

and:

$$\mathbf{W}_N = \mathbf{A}_N^{\mathrm{T}}\mathbf{A}_N$$

where $\mathbf{A}_N$ is a matrix of dimension $\overline{O} \times \overline{O}$ where $\overline{O} \equiv (1+Q)\,K + P + 1$ and such that $\operatorname{rank}(\mathbf{A}_N) \geq \dim|\boldsymbol{\theta}|$ and $\mathbf{A}_n \overset{p}{\to} \mathbf{A}_0$, where $\mathbf{A}_0^{\mathrm{T}}\mathbf{A}_0 = \mathbf{W}_0$. Observe that, by denoting the j-ℓ entries of $\mathbf{A}_N$ as $a_{j\ell,N}$ and by $\mathbf{q}^*_{qk,N}$ the k-th row of $\mathbf{Q}_{q,N}$ (for $k = 1, \dots, K$), it is:

$$\mathbf{A}_N\overline{\mathbf{m}}_N(\boldsymbol{\theta}) = \frac{1}{N}\sum_{j=1}^{\overline{O}}\left[\sum_{q=0}^{Q}\sum_{k=1}^{K} a_{j(qK+k),N}\mathbf{q}^*_{qk,N} + \sum_{p=0}^{P} a_{j\left(\overline{Q}+1+p\right),N}\boldsymbol{\varepsilon}_N^{\mathrm{T}}(\boldsymbol{\theta})\,\mathbf{P}_{p,N}\right]\boldsymbol{\varepsilon}_N(\boldsymbol{\theta})$$

where $\overline{Q} \equiv (1+Q)K$; to establish consistency of $\widehat{\theta}_{GMM}$, it is necessary to show uniform convergence in probability for all the elements in the quadratic form above. The case of the linear moments is standard, hence the focus here is on the covariance restrictions. To this end, note that:

$$\sum_{p=0}^{P} a_{j(\overline{Q}+1+p),N} \boldsymbol{\varepsilon}_N^{\mathrm{T}}(\boldsymbol{\theta}) \mathbf{P}_{p,N} \boldsymbol{\varepsilon}_N(\boldsymbol{\theta}) = \sum_{p=0}^{P} a_{j(\overline{Q}+1+p),N} \mathbf{d}_N^{\mathrm{T}}(\boldsymbol{\theta}) \mathbf{P}_{p,N} \mathbf{d}_N(\boldsymbol{\theta}) + \\ + \underbrace{\sum_{p=0}^{P} a_{j(\overline{Q}+1+p),N} \mathbf{d}_N^{\mathrm{T}}(\boldsymbol{\theta}) \mathbf{P}_{p,N} \mathbf{e}_N(\boldsymbol{\theta})}_{\equiv l_N(\boldsymbol{\theta})} + \underbrace{\sum_{p=0}^{P} a_{j(\overline{Q}+1+p),N} \mathbf{e}_N^{\mathrm{T}}(\boldsymbol{\theta}) \mathbf{P}_{p,N} \mathbf{e}_N(\boldsymbol{\theta})}_{\equiv r_N(\boldsymbol{\theta})}$$

where:

$$\mathbf{e}_N(\boldsymbol{\theta}) \equiv \boldsymbol{\varepsilon}_N + (\mathbf{X}_N - \mathbb{E}[\mathbf{X}_N])(\boldsymbol{\gamma}_0 - \boldsymbol{\gamma}) + \mathbf{G}_N(\mathbf{X}_N - \mathbb{E}[\mathbf{X}_N])(\boldsymbol{\delta}_0 - \boldsymbol{\delta}) + \\ + (\beta_0 - \beta)\mathbf{G}(\mathbf{I}_N - \beta_0 \mathbf{G}_N)^{-1}[\boldsymbol{\varepsilon}_N + (\mathbf{X}_N - \mathbb{E}[\mathbf{X}_N])\boldsymbol{\gamma}_0 + \mathbf{G}_N(\mathbf{X}_N - \mathbb{E}[\mathbf{X}_N])\boldsymbol{\delta}_0]$$

therefore, there are $K+1$ appropriate $1 \times N$ vectors $\mathbf{l}_{1,N}(\boldsymbol{\theta})$ and $\mathbf{l}_{k,N}(\boldsymbol{\theta})$ and $2K+1$ appropriate $N \times N$ matrices $\mathbf{R}_{0,N}(\boldsymbol{\theta})$, $\mathbf{R}_{k,N}(\boldsymbol{\theta})$ and $\mathbf{R}_{k',N}(\boldsymbol{\theta})$ for $k, k' = 1, \ldots, K$ such that:

$$\frac{1}{N} l_N(\boldsymbol{\theta}) = \frac{1}{N} \mathbf{l}_{0,N}(\boldsymbol{\theta}) \boldsymbol{\varepsilon}_N(\boldsymbol{\theta}) + \frac{1}{N} \sum_{k=1}^{K} \mathbf{l}_{k,N}(\boldsymbol{\theta}) \left(\mathbf{x}_{k,N}^* - \mathbb{E}\left[\mathbf{x}_{k,N}^*\right]\right) = o_{\mathcal{P}}(1)$$

$$\frac{1}{N} r_N(\boldsymbol{\theta}) = \frac{1}{N} \boldsymbol{\varepsilon}_N^{\mathrm{T}}(\boldsymbol{\theta}) \mathbf{R}_{0,N}(\boldsymbol{\theta}) \boldsymbol{\varepsilon}_N(\boldsymbol{\theta}) + \frac{1}{N} \sum_{k=1}^{K} \boldsymbol{\varepsilon}_N^{\mathrm{T}}(\boldsymbol{\theta}) \mathbf{R}_{k,N}(\boldsymbol{\theta}) \left(\mathbf{x}_{k,N}^* - \mathbb{E}\left[\mathbf{x}_{k,N}^*\right]\right) + \\ + \frac{1}{N} \sum_{k'=1}^{K} \left(\mathbf{x}_{k',N}^* - \mathbb{E}\left[\mathbf{x}_{k',N}^*\right]\right)^{\mathrm{T}} \mathbf{R}_{k',N}(\boldsymbol{\theta}) \left(\mathbf{x}_{k',N}^* - \mathbb{E}\left[\mathbf{x}_{k',N}^*\right]\right) = o_{\mathcal{P}}(1)$$

where $\mathbf{x}_{k,N}^*$ is the k-th column of $\mathbf{X}_N$, such that for each $k = 1, \ldots, K$:

$$\frac{1}{N} \boldsymbol{\iota}^{\mathrm{T}} \left\{\mathbf{x}_{k,N}^* - \mathbb{E}\left[\mathbf{x}_{k,N}^*\right]\right\} = \frac{1}{N} \boldsymbol{\iota}^{\mathrm{T}} \left\{\boldsymbol{\omega}_{k,N} - \mathbb{E}[\boldsymbol{\omega}_{k,N}] + \xi_k \mathbf{C}_{k,N} \mathbb{E}[\boldsymbol{\upsilon}_N]\right\} = o_{\mathcal{P}}(1)$$

and similarly for $\mathbf{x}_{k',N}^*$. The results above entail uniform convergence in probability in $\boldsymbol{\Theta}$, since the latter space is bounded, $l_N(\boldsymbol{\theta})$ and $r_N(\boldsymbol{\theta})$ are quadratic functions in the relevant parameters, and because of Lemmas A.3 and A.4 in Lee (2007). By extending the argument to the linear moments, it follows that since $\overline{\mathbf{m}}_N(\boldsymbol{\theta})$ is also quadratic in $\boldsymbol{\theta}$ and $\boldsymbol{\Theta}$ is bounded, then $N^{-1}\mathbf{A}_N \mathbb{E}[\overline{\mathbf{m}}_N(\boldsymbol{\theta})]$ is uniformly equicontinuous in $\boldsymbol{\Theta}$, therefore the identification uniqueness condition for $N^{-2}\mathbf{A}_N \mathbb{E}\left[\overline{\mathbf{m}}_N^{\mathrm{T}}(\boldsymbol{\theta}) \mathbf{A}_N^{\mathrm{T}} \mathbf{A}_N \overline{\mathbf{m}}_N(\boldsymbol{\theta})\right]$ is satisfied and the GMM estimator is consistent (White, 1994).

It remains to show that $\widehat{\theta}_{GMM}$ is also asymptotically normal. The usual application of the Mean Value Theorem to the First Order Conditions gives:

$$\sqrt{N}\left(\widehat{\theta}_{GMM}-\theta_0\right)=-\left[\mathbf{J}_N^{\mathrm{T}}\left(\widehat{\theta}_{GMM}\right)\mathbf{W}_N\mathbf{J}_N\left(\overline{\theta}\right)\right]^{-1}\mathbf{J}_N^{\mathrm{T}}\left(\widehat{\theta}_{GMM}\right)\mathbf{W}_N\sqrt{N}\overline{\mathbf{m}}_N\left(\theta_0\right)$$

where $\mathbf{J}_N(\theta)=\frac{\partial}{\partial\theta}\overline{\mathbf{m}}_N(\theta)$. By Theorem 1 in Kelejian and Prucha (2001):

$$\sqrt{N}\mathbf{A}_N\overline{\mathbf{m}}_N\left(\theta_0\right)\xrightarrow{d}\mathcal{N}\left(\mathbf{0},\mathbf{A}_0\mathbf{\Omega}_0\mathbf{A}_0^{\mathrm{T}}\right)$$

hence the main result would follow if $\mathbf{J}_N\left(\widehat{\theta}_{GMM}\right)=\mathbf{J}_0+o_{\mathcal{P}}(1)$. Note that:

$$\mathbf{J}_N(\theta)=-\frac{1}{N}\begin{bmatrix}\mathbf{Q}_{0,N}\\ \mathbf{Q}_{1,N}\\ \vdots\\ \mathbf{Q}_{Q,N}\\ 2\boldsymbol{\varepsilon}^{\mathrm{T}}(\theta)\mathbf{P}_{0,N}\\ 2\boldsymbol{\varepsilon}^{\mathrm{T}}(\theta)\mathbf{P}_{1,N}\\ \vdots\\ 2\boldsymbol{\varepsilon}^{\mathrm{T}}(\theta)\mathbf{P}_{P,N}\end{bmatrix}\begin{bmatrix}\mathbf{G}_N\mathbf{y}_N & \mathbf{X}_N & \mathbf{G}_N\mathbf{X}_N & \mathbf{0}_N & \mathbf{0}_N & \mathbf{0}_N\end{bmatrix}+\frac{\partial\boldsymbol{\mu}_N(\theta)}{\partial\theta^{\mathrm{T}}}$$

where $\mathbf{0}_N$ is shorthand for an N-dimensional vector of zeros. Leaving $\frac{\partial}{\partial\theta^{\mathrm{T}}}\boldsymbol{\mu}_N(\theta)$ aside for the moment, we focus on the first term on the right-hand side, and in particular on the submatrix formed by the last $P+1$ rows – the covariance restrictions – and the first column – corresponding with the derivative with respect to β – since the analysis of the rest of the matrix is just a simpler case. By Lemmas A.3 and A.4 in Lee (2007), one can write each p-th element of said submatrix-vector, for $p=0,1,2,\dots,P$, as:

$$\frac{1}{N}\boldsymbol{\varepsilon}^{\mathrm{T}}(\theta)\mathbf{P}_{p,N}\mathbf{G}_N\left(\mathbf{I}_N-\beta_0\mathbf{G}_N\right)^{-1}\left(\mathbf{X}_N\boldsymbol{\gamma}_0+\mathbf{G}_N\mathbf{X}_N\boldsymbol{\delta}_0+\boldsymbol{\varepsilon}_N\right)=d_{p,N}+v_{p,N}+t_{p,N}+f_{p,N}$$

where, writing $\mathbf{y}_N(\theta_0)\equiv\left(\mathbf{I}_N-\beta_0\mathbf{G}_N\right)^{-1}\left(\mathbf{X}_N\boldsymbol{\gamma}_0+\mathbf{G}_N\mathbf{X}_N\boldsymbol{\delta}_0+\boldsymbol{\varepsilon}_N\right)$, the terms on the right-hand side are given by:

$$d_{p,N}=\frac{1}{N}\mathbf{d}_N^{\mathrm{T}}(\theta)\mathbf{P}_{p,N}\mathbf{G}_N\mathbf{y}_N(\theta_0)=\frac{1}{N}\mathbf{d}^{\mathrm{T}}(\theta)\mathbf{P}_{p,N}\mathbf{G}_N\mathbb{E}\left[\mathbf{y}_N\right]+o_{\mathcal{P}}(1)$$

and:

$$\begin{aligned}v_{p,N}=\frac{1}{N}\boldsymbol{\varepsilon}_N^{\mathrm{T}}\mathbf{P}_{p,N}\mathbf{G}_N\mathbf{y}_N(\theta_0)&=\sigma_0^2\mathrm{Tr}\left[\mathbf{P}_{p,N}\mathbf{G}_N\left(\mathbf{I}_N-\beta_0\mathbf{G}_N\right)^{-1}\right]+\\&+\sigma_0\xi_0\mathrm{Tr}\left[\left(\mathbf{I}_N-\psi_0\mathbf{E}_N\right)^{\mathrm{T}}\mathbf{P}_{p,N}\mathbf{G}_N\left(\mathbf{I}_N-\beta_0\mathbf{G}_N\right)^{-1}\mathbf{G}_N\right]+o_{\mathcal{P}}(1)\end{aligned}$$

and:

$$t_{p,N} = \frac{1}{N}\boldsymbol{\varepsilon}_N^{\mathrm{T}} \left(\mathbf{I}_N - \beta_0 \mathbf{G}_N\right)^{-1} \mathbf{G}_N \mathbf{P}_{p,N} \mathbf{G}_N \mathbf{y}_N \left(\theta_0\right) = \sigma_0^2 \mathrm{Tr}\left[\mathbf{P}_{p,N} \mathbf{G}_N \left(\mathbf{I}_N - \beta_0 \mathbf{G}_N\right)^{-1}\right] + \\ + \sigma_0 \xi_0 \mathrm{Tr}\left[\left(\mathbf{I}_N - \psi_0 \mathbf{E}_N\right)^{\mathrm{T}} \mathbf{P}_{p,N} \mathbf{G}_N \left(\mathbf{I}_N - \beta_0 \mathbf{G}_N\right)^{-1} \mathbf{G}_N\right] + o_{\mathcal{P}}\left(1\right)$$

and:

$$f_{p,N} = \frac{1}{N} \sum_{k=1}^{K} \left(\mathbf{x}_{k',N}^{*} - \mathbb{E}\left[\mathbf{x}_{k',N}^{*}\right]\right)^{\mathrm{T}} \left[\left(\beta_0 - \beta\right)\left(\gamma_{k,0} \mathbf{I}_N + \delta_{k,0} \mathbf{G}_N\right) \mathbf{G}_N \left(\mathbf{I}_N - \beta_0 \mathbf{G}_N\right)^{-1} + \\ + \left(\gamma_{k,0} + \gamma_k\right) \mathbf{I}_N + \left(\delta_{k,0} + \delta_k\right) \mathbf{G}_N\right]^{\mathrm{T}} \mathbf{P}_{p,N} \mathbf{G}_N \mathbf{y}_N \left(\theta_0\right) = \sum_{k=1}^{K} f_{p,k,N}$$

to complete the analysis of which, denote by $\boldsymbol{\Xi}_{k,N}$ the K matrices corresponding with the expression in brackets above, and define $\phi_{k,0} \equiv \mathbb{V}\mathrm{ar}\left[\omega_{i,k}\right]$; thus, the K elements of the summation can be expressed as:

$$f_{p,k,N} = \left(\gamma_{k,0} + \delta_{k,0}\right)\left\{\xi_{k,0}^2 \mathrm{Tr}\left[\mathbf{G}_N \boldsymbol{\Xi}_{k,N}\right] \mathbf{P}_{p,N} \mathbf{G}_N \left(\mathbf{I}_N - \beta_0 \mathbf{G}_N\right)^{-1} \mathbf{G}_N\right\} + \\ + \left(\gamma_{k,0} + \delta_{k,0}\right)\left\{\phi_{k,0}^2 \mathrm{Tr}\left[\mathbf{C}_{k,N} \boldsymbol{\Xi}_{k,N}\right] \mathbf{P}_{p,N} \mathbf{G}_N \left(\mathbf{I}_N - \beta_0 \mathbf{G}_N\right)^{-1} \mathbf{C}_{k,N}\right\} + \\ + \left(\gamma_{k,0} + \delta_{k,0}\right)\left\{\xi_{k,0} \phi_{k,0} \mathrm{Tr}\left[\mathbf{G}_N \boldsymbol{\Xi}_{k,N}\right] \mathbf{P}_{p,N} \mathbf{G}_N \left(\mathbf{I}_N - \beta_0 \mathbf{G}_N\right)^{-1} \left(\mathbf{I}_N + \psi_0 \mathbf{E}_N\right)\right\} + o_{\mathcal{P}}\left(1\right)$$

for $k = 1, \ldots, K$. All the probability limits above imply uniform convergence for any $\theta \in \boldsymbol{\Theta}$; collecting these results together gives:

$$\frac{1}{N}\boldsymbol{\varepsilon}^{\mathrm{T}}\left(\theta\right) \mathbf{P}_{p,N} \mathbf{G}_N \mathbf{y}_N \left(\theta_0\right) = \sigma_0^2 \mathrm{Tr}\left[\mathbf{P}_{p,N} \mathbf{G}_N \left(\mathbf{I}_N - \beta_0 \mathbf{G}_N\right)^{-1}\right] + \\ + \sigma_0 \sum_{k=1}^{K} \xi_{k,0} \mathrm{Tr}\left[\left(\mathbf{I}_N + \psi_0 \mathbf{E}_N\right)^{\mathrm{T}} \mathbf{P}_{p,N} \mathbf{G}_N \left(\mathbf{I}_N - \beta_0 \mathbf{G}_N\right)^{-1} \mathbf{G}_N\right] + o_{\mathcal{P}}\left(1\right)$$

moreover, since some tedious analysis reveals that $\frac{\partial}{\partial \theta^{\mathrm{T}}} \boldsymbol{\mu}_N\left(\theta\right) = \frac{\partial}{\partial \theta^{\mathrm{T}}} \boldsymbol{\mu}\left(\theta\right) + o_{\mathcal{P}}\left(1\right)$, it follows that the $\left(P+1\right) \times 1$ submatrix of $\mathbf{J}_N\left(\theta\right)$ under examination has the desired properties. Extending these considerations to all the elements of $\mathbf{J}_N\left(\theta\right)$ proves the asymptotic normality result as stated by Theorem.