

Identification of Network Effects with Spatially Endogenous Covariates: Theory, Simulations and an Empirical Application*

Santiago Pereda-Fernández[†] and Paolo Zacchia[‡]

August 2022

Abstract

Researchers interested in the estimation of peer and network effects, even if these are algebraically identified, still need to address the problem of correlated effects. In this paper we characterize the identification conditions for consistently estimating all the parameters of a spatially autoregressive or linear-in-means model when the structure of social or peer effects is exogenous, but the observed and unobserved characteristics of agents are cross-correlated over some given metric space. We show that identification is possible if the network of social interactions is non-overlapping up to enough degrees of separation, and the spatial matrix that characterizes the co-dependence of individual unobservables and peers' characteristics is known up to a multiplicative constant. We propose a GMM approach for the estimation of the model's parameters, and we evaluate its performance through Monte Carlo simulations. Finally, we show that in a classical empirical application about classmates our approach might estimate statistically non-significant peer effects when conventional approaches register them as significant.

JEL Classification Codes: C21, C31, D85

Keywords: Peer Effects, Networks, Correlated Effects, Spatial Correlation

*We express our heartfelt gratitude to Tiziano Arduini, Manuel Arellano, Yann Bramoullé, Áureo de Paula, Bryan S. Graham, Ida Johnsson and Michele Pellizzari for helpful discussions and advice. Likewise, we extend our thanks to all seminar participants who, at the 2016 North American Summer Meeting of the Econometric Society, the 2016 Conference of the International Association for Applied Econometrics, GREQAM Université de Marseille, the New Economic School, and the 2018 European Winter Meeting of the Econometric Society, provided helpful comments and feedback regarding this paper or its previously circulated versions. We are especially grateful to Giacomo De Giorgi, Lorenzo Peccati, Michele Pellizzari and Silvia Redaelli for sharing the data used in our empirical application. We are the sole responsible for any outstanding mistakes and omissions.

[†]Universidad de Cantabria; e-mail: santiago.pereda@bancaditalia.it.

[‡]Charles University and the Czech Academy of Sciences, e-mail: Paolo.Zacchia@cerge-ei.cz.

1 Introduction

A sizable body of empirical economic research concerns the analysis of peer effects, network effects and more generally “social effects,” i.e. mutual externalities induced by socio-economic interaction. Within this literature, peer effects in education occupy a 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 also numerous (Glaeser et al., 1996; Duflo and Saez, 2003; Mas and Moretti, 2009).¹ In the face of a growing empirical evidence, econometric analysis has struggled for a while to provide a unique structural interpretation to observed group correlations in socio-economic outcomes. Over time, advances have been made: to unambiguously identify the effect of social interactions, the current econometric theory and practice emphasize the use of instrumental variables based upon the observable characteristics of indirectly connected agents in structures of social interactions with a non-trivial topology, such as networks (Bramoullé et al., 2009). However, this kind of approach is largely confined to a restricted set of settings where such characteristics, as well as the structure of socio-economic interactions itself, are both as good as exogenous. This makes these studies liable to the critique, which was put forward most notably by Angrist (2014), according to which the current results in the literature are likely to reflect spurious correlations due to unobserved “correlated effects” that are shared between peers.

By contrast, in this paper we examine a cross-sectional model of social interactions where observed and unobserved individual characteristics are: *i.* both cross-correlated across individuals in some metric space, and *ii.* mutually dependent on one another. Our point of departure is a “Spatially Autoregressive” model (Cliff and Ord, 1981), hereinafter SAR, whose econometrics has been analyzed extensively (Lee, 2007a,b; Lee et al., 2010; Lin and Lee, 2010; Liu and Lee, 2010; Lee and Liu, 2010). Similarly to other papers, we derive our empirical model from an explicit theoretical (strategic) 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.

¹Studies 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 (2020). Other related strands of literature include the one about peer effects in scientific production (Azoulay et al., 2010; Waldinger, 2012) and that about learning externalities (Conley and Udry, 2010).

We explicitly illustrate that in such a framework, the type of endogenous dependence that we allow for not only makes standard estimates of social effects inconsistent, but can also be – under some specifications – observationally equivalent to the so-called “exogenous” or “contextual” effects of peers’ characteristics that are often featured in studies about social interactions. Both observations resonate with the aforementioned critique of the whole empirical literature about social effects.

The main contribution of our paper is to show that within this framework, social effects are identified without resorting to external instruments. We analyze a scenario where the observable characteristics of socio-economic agents depend in a linear fashion on both their own unobservables and on those of other agents, which makes such characteristics both endogenous and cross-correlated. We impose no restriction upon the spatial matrices that characterize this type of endogeneity, except that they are known to the econometrician up to a multiplicative parameter that quantifies the extent of endogeneity. As we elaborate later, knowing the *structure* but not the *intensity* of this type of spatial correlation is arguably realistic in those empirical settings that motivate our work. For example, in peer networks observable characteristics, possibly *all* of them, are likely correlated on the basis of individual previous backgrounds, be they professional, cultural or geographical; in firm-level networks instead, the spatial correlation of key firm-level variables is likely shaped by similarities in technological and product market characteristics. Still, in our analysis we also explore the practical implications of knowing the structure in question imperfectly (misspecification).

The main identifying assumption extends those by Bramoullé et al. (2009), as it requires that the structure of social interactions is non-overlapping up to an additional degree of separation in network space relative to their original results. The intuition is that the type of endogeneity featured in our framework introduces a bias which is observationally equivalent to higher-order network effects; the bias can be explicitly controlled for by accounting for the correlation between an individual’s outcome and the characteristics of higher-order indirect connections in the network. In order to do that, such correlations must be separately identified at different degrees of separation. In the more general version of our model we also introduce a number of covariance restrictions. Their key role is to identify the parameters associated with the primitive components of the error term’s covariance structure (which we allow to be arbitrarily autocorrelated and moving-averaged in network space) and that affect the expression of the aforementioned bias which we control for via first-order moment conditions.

Leveraging upon the moment conditions upon which our identification results are based, we propose a GMM approach for the joint estimation of both social effects and the other parameters of our framework. We derive the asymptotic properties of our estimator and we evaluate its performance in Monte Carlo simulations. Furthermore, we showcase it empirically by applying it to the setting and data from the study by De Giorgi et al. (2010), which is about peer effects in the classroom between students of Bocconi University in Italy. Although peer groups are formed exogenously in that setting, it is arguable that the observable characteristics of students – such as their high school grades – are cross-correlated in a predictable fashion, e.g. as a function of two students’ geographical provenance. Indeed, the estimates of peer effects obtained through an application of our method which accounts for cross-correlation structures of this predictable kind are smaller in magnitude when compared to other customary approaches, and they are not statistically significant. We find that this pattern holds under multiple assumptions about the dependence structure, which echoes an observation we draw from Monte Carlo simulations: our approach can still outcompete the alternatives under misspecification of the cross-correlation between the error term and the observable characteristics. Overall, we interpret these results as a warning against the incautious interpretation of observed cross-correlations in individual outcomes as the result of some structural, behavioral mechanisms such as peer effects.

To better frame our contribution, it is worth to summarize the intellectual history of the workhorse framework in many studies on social effects: the “linear-in-means” model (a special case of an augmented SAR model). In a seminal paper, Manski (1993) highlighted the “reflection problem:” social effects occurring in segregated groups are hard to identify, because group characteristics and group outcomes are simultaneous. Since then, econometricians have striven to characterize conditions under which social effects can be disentangled from confounding factors. The aforementioned, influential contribution by Bramoullé et al. (2009) illustrates how to identify social effects when the latter are shaped via networked structures of interaction where connections are not necessarily transitive; this is especially appealing as networks typically provide more realistic descriptions of real-world social relationships. Blume et al. (2015) incorporate their identification results – as well as one based upon covariance restrictions which builds on Graham (2008) – within a more extended framework. Thanks to these and other efforts, it is now well understood that complex patterns of individual dependence make the identification of social effects, if anything, easier.

Yet most of these analyses either maintain the assumption that the model’s error term is conditionally independent of the observable characteristics and the structure of interactions, or they assume structures of dependence which are not as general and potentially pervasive as ours, and which hence allow for relatively simple solutions.² Obviously, the spatial econometrics literature has examined correlated unobservables at length (Kelejian and Prucha, 1998, 2007, 2010; Kapoor et al., 2007; Drukker et al., 2013); however, individual covariates are typically assumed exogenous in such studies. In a recent survey of the literature about peer effects in networks, Bramoullé et al. (2020) discuss several randomization-based attempts aimed at addressing endogeneity in the composition of peer groups: a problem which is distinct, albeit related, to that of correlated effects. The survey cites an earlier, incomplete version of our paper as the only recent contribution that attempts a structural approach to address the issue of generalized correlated effects, a method potentially amenable to observational studies. Our idea of exploiting the very spatial structure of endogenous cross-correlation for the sake of identification builds upon some previous work by Zacchia (2020).³

As mentioned, the literature has focused at length on a key issue: the possibility that the actual networked structure of interactions is itself endogenous. In an influential contribution, Goldsmith-Pinkham and Imbens (2013) adopt a Bayesian approach in order to estimate an extension of the linear-in-means model where the probability that two peers are linked depends on the degree of similarity between their observable and unobservable characteristics (“homophily”). Following a suggestion originally given by Blume et al. (2015), some scholars (Arduini et al., 2015; Johnsson and Moon, 2021) later developed a control function approach to account for endogeneity of the network. These methods embed, within a SAR-like framework, a network formation model based on Graham (2017).⁴ In other, more empirical contributions, the network or part of it is random (Sacerdote, 2001; De Giorgi et al., 2010; Carrell et al., 2013). We argue that randomizing the peer groups is not sufficient to solve the problem of

²The leading case is again provided by Bramoullé et al. (2009), who allow for fixed effects specific to each of the multiple “networks” that make up their samples. To remove them, they propose local data demeaning procedures that conceptually precede their main two-stages estimation approach.

³Zacchia (2020) 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.

⁴In a recent contribution, Kuersteiner and Prucha (2020) 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.

correlated effects if spatial correlation in the unobservables is pervasive. We maintain that the network is exogenous in most of our discussion: thus, we can isolate our key mechanism of interest. We argue, however, that the approach we propose would also work under some particular, realistic specifications of endogenous network formation.

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 some covariance restrictions, but unlike these papers, their role in identification is to disentangle the autonomous covariance structure of the error term from that of individual covariates, if the two are correlated. Other contributions develop methods for estimating unknown structures of interaction (Rose, 2017b; de Paula et al., 2019) using penalized estimators. While we make no use of such techniques, we argue that they may be adapted for the sake of recovering the structure of spatial correlation that induces endogeneity. We revisit this observation 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 model and the endogeneity specification that we analyze. Section 3 details on the conditions for the identification of social effects. Section 4 introduces our GMM estimator and its asymptotic properties. Section 5 assesses its performance in Monte Carlo simulations. Section 6 discusses our empirical application of the proposed estimator. Lastly, Section 7 concludes the paper. An Appendix provides key mathematical proofs.

2 General Framework

The first part of this section introduces a game-theoretical framework to support our empirical model, and we emphasize its implications for the statistical identification of social effects. The last part of this section characterizes the endogeneity specification that we focus on, and we motivate it with examples inspired by applied research.

2.1 A game of social interactions

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 $i = 1, \dots, N$. 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 $i = 1, \dots, N$. Otherwise, we force no particular structure of the network: we generally allow for asymmetric, directed networks such that for any pair $(i, j) \in \mathcal{I}^2$, the weight g_{ij} implies no restriction upon the weight g_{ji} , and vice versa.

Every player in \mathcal{I} is typified by two variables (x_i, ϵ_i) . We denominate $x_i \in \mathcal{X}$ the *observable characteristic* of player i , and $\epsilon_i \in \mathcal{E}$ his or her *unobservable characteristic*, a terminology that conveys what information regarding either variable is available to econometricians. Both x_i and ϵ_i may be interpreted as the composition of multiple socio-economic factors. Thus, one can easily generalize this framework, as we do later while introducing a more general model with multiple observable characteristics. For simplicity, we set $\mathcal{X} = \mathcal{E} = \mathbb{R}$, although restricted supports could be accommodated easily. We assume that: (i) the random vector of individual observable characteristics $\mathbf{x} = (x_1, \dots, x_N)$, (ii) the random vector of individual unobservables $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_N)$, and (iii) the network \mathcal{G} , are all randomly drawn from a joint probability distribution $\mathcal{F}(\mathbf{x}, \boldsymbol{\epsilon}, \mathcal{G})$, which is known by all agents. This paper deals with particular restrictions on $\mathcal{F}(\cdot)$ such that social effects are identified, despite $(\mathbf{x}, \boldsymbol{\epsilon})$ being mutually dependent (and possibly dependent on \mathcal{G} as well). For the sake of exposition, for the moment we place no *a priori* restriction on $\mathcal{F}(\cdot)$; in the latter part of this section, we introduce and motivate our endogeneity specification of interest.

Players maximize the following “twice exponential” utility function:

$$U_i(e_1, \dots, e_N; x_i, \epsilon_i) = \exp[y_i(e_1, \dots, e_N; x_i, \epsilon_i)] - \exp(e_i), \quad (1)$$

where y_i is the individual-level *outcome* (denoting, say, grades, or production output). The latter is determined through a linear relationship which implies a Cobb-Douglas positive contribution to utility, and that allows for diverse settings such as peer effects in education and R&D spillovers across firms:

$$y_i(e_1, \dots, e_N; x_i, \epsilon_i) = \alpha_0 + \gamma_0 x_i + \mu e_i + \nu \sum_{j=1}^N g_{ij} e_j + \epsilon_i. \quad (2)$$

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 (if the setting of interest is peer effects in education) or R&D investment (for the R&D spillovers setting). 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).⁵ The private and social effects of effort are parametrized as μ and ν , respectively. Note that in this model, all variables (including the weighted sum of peer effort) are complements with one another, unlike in quadratic utility models typical of the peer effects literature (Calvó-Armengol et al., 2009; Blume et al., 2015).

Define the combined parameter $\beta \equiv \nu / (1 - \mu)$. We analyze the model under the following assumptions, the first two of which we maintain throughout the paper.

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

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

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

Assumption 1 makes the positive part of utility non-convex in the strategic variable e_i , and thus the model salient. Assumption 2 limits the impact of social effects on the outcome: first, it ensures uniqueness of the equilibrium while ruling out unrealistically “explosive” solutions with very large values of y_i ; moreover, from a statistical point of view, it bounds the variance-covariance of $\mathbf{y} = (y_1, \dots, y_N)$, which is an increasing function of network connections.⁶ Assumption 3 is typical of the peer effects literature (Manski, 1993; Bramoullé et al., 2009) and provides an interpretation of social effects as 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 maintain Assumption 3 and focus on the conditions for identifying the combined parameter β . Later we relax this hypothesis and we discuss a version of our framework where μ and ν are separately identified through variation in individual in-degree \bar{g}_i . 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 $|\beta| \in [0, 1)$.

⁵For example, the R&D of other firms e_i can both lead to positive knowledge spillovers on output y_i and to negative business stealing effects; see e.g. Bloom et al. (2013), where these two effects are separately identified under the hypothesis that they are mediated by different weights g_{ij} .

⁶We observe that variations of this assumption are typical in the spatial econometrics literature, see e.g. Kelejian and Prucha (2007, 2010); Lee (2007a); Lin and Lee (2010); Liu and Lee (2010).

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

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

By letting the network be generated randomly by nature we abstract from the specifics of the network formation process, as the following result does not depend on it.⁷

Proposition 1. Equilibrium. *For all realizations of $(\mathbf{x}, \boldsymbol{\epsilon}, \mathcal{G})$, under Assumptions 1 and 2 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 $i = 1, \dots, N$ as follows:*

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

where $\alpha \equiv [\alpha_0 + (\mu + \nu) \log \mu] / (1 - \mu)$, $\gamma \equiv \gamma_0 / (1 - \mu)$ and $\varepsilon_i \equiv \epsilon_i / (1 - \mu)$.

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

$$e_j = y_j + \log \mu. \quad (4)$$

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, \dots, e_N) in the $(\mathbb{R}^N, \mathfrak{M})$ metric space, where \mathfrak{M} is the max norm. This implies uniqueness. \square

2.2 Social effects

The reduced form expression (3) that is generated in equilibrium resembles the typical equation of a “linear-in-means” models from the peer effects literature, but comes with additional insights. First, parameter β – corresponding to the *endogenous effect* from the original classification by Manski (1993) – is given a clear behavioral interpretation: it 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

⁷Note that by assuming complete information we make our analysis more general. As discussed by Zacchia (2020), incomplete information provides additional avenues for the identification of social effects, thanks to implicit restrictions on the cross-correlation of strategic variables.

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. In general, however, not even the combined parameter β is identified, as we show next through a constructed example.

Proposition 2. Non-identification of endogenous social effects. *Suppose that in model (3) it is $\gamma = \beta = 0$, hence $y_i = \alpha + \varepsilon_i$. There exist restrictions on $\mathcal{F}(\cdot)$ such that, under Assumptions 1-3, such a simple model is observationally equivalent to:*

$$y_i = \alpha' + \beta' \sum_{j=1}^N g_{ij} y_j + \gamma' x_i + \varepsilon'_i, \quad (5)$$

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

Proof. Let $\mathcal{F}(\cdot)$ be such that $\varepsilon_i = \rho \sum_{j=1}^N g_{ij} \varepsilon_j + v_i$ and $\mathbb{E}[v_i | \mathbf{x}, \mathcal{G}] = \pi_0 + \pi_1 x_i$, where $|\rho| \in (0, 1)$ and $\pi_1 \neq 0$. One can verify that $y_i = \alpha + \varepsilon_i$ is observationally equivalent to model (5) for $\alpha' = \alpha(1 - \rho) + \pi_0$, $\beta' = \rho$, and $\gamma' = \pi_1$. \square

This stylized example highlights the key issue that this paper is concerned with: researchers can mistake behavioral externalities for unobserved confounders that are shared, to some degree, by multiple observations. Although the two mechanisms can deliver indistinguishable statistical patterns in the data, the latter, unlike the former, does not allow to make causal claims or draw policy implications. This was, at heart, the critique of the whole peer effects literature put forward by Angrist (2014).

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*). Although we could easily include an additional term in (2) to include the exogenous effect, such a choice is liable to the following critique.

Proposition 3. Non-identification of exogenous effects. *There exist specific restrictions on $\mathcal{F}(\cdot)$ such that, under Assumptions 1-3, model (3) is observationally equivalent to the following statistical model:*

$$y_i = \alpha'' + \beta'' \sum_{j=1}^N g_{ij} y_j + \gamma'' x_i + \delta'' \sum_{j=1}^N g_{ij} x_j + \varepsilon''_i, \quad (6)$$

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

Proof. Let $\mathcal{F}(\cdot)$ be such that $\varepsilon_i = \rho \sum_{j=1}^N g_{ij} \varepsilon_j + \varepsilon_i''$ and $\mathbb{E}[\varepsilon_i | \mathbf{x}, \mathcal{G}] = \kappa_0 + \kappa_1 x_i$, where $|\rho| \in (0, 1)$ and $\kappa_1 \neq 0$. One can verify that models (3) and (5) are observationally equivalent for $\alpha'' = \alpha + \rho \kappa_0$, $\beta'' = \beta$, $\gamma'' = \gamma$ and $\delta'' = \rho \kappa_1$. \square

This additional example helps make an important point.⁸ If unobservables ε_i are cross-correlated in the network and, in addition, the observables x_i are correlated with the unobservables ε_i , then “contextual effects,” say parametrized by δ' , can emerge as a statistical byproduct of more fundamental structural or stochastic patterns. We see this as a cautionary message to researchers aiming to estimate spillover effects in any setting: the solution of any endogeneity problems due to simultaneous unobservables (and possibly network formation) must precede model specification. For this reason, we restrict our main discussion of identification, our Monte Carlo simulations, as well as our empirical application to models without exogenous effects (still, we allow for them in both our general model and in the construction of our estimator).

2.3 Spatial linear endogeneity

In this paper we focus on restrictions on $\mathcal{F}(\cdot)$ expressed through the following linear statistical relationship between observable characteristics x_i and unobservables ε_i , for $i = 1, \dots, N$:

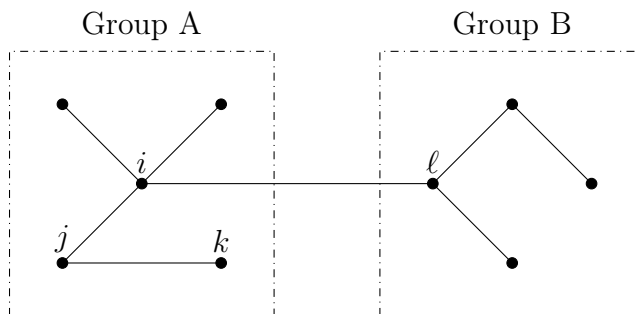
$$x_i = \tilde{x}_i + \xi \sum_{j=1}^N c_{ij} \varepsilon_j. \quad (7)$$

In the above, $\xi \in \mathbb{R}$ is a parameter; \tilde{x}_i is a random variable that we call the *independent component* of the variation of x_i , whose distribution is left unrestricted except for being assumed continuous ($\tilde{x}_i \neq \tilde{x}_j$ almost surely for $i \neq j$) as well as independent of individual unobservables ($\mathbb{E}[\tilde{x}_i \varepsilon_j] = 0$ for any i, j); whereas the weights c_{ij} , that we call *characteristic weights*, introduce the statistical spatial dependence of interest. Like in the case of the adjacency weights g_{ij} , we impose the normalization $c_{ij} \in [0, 1]$; unlike those, however, typically it is $c_{ii} \neq 0$. We collect all the characteristic weights in the N^2 -dimensional set \mathcal{C} , that we call the “characteristics structure.”

⁸In both constructed examples the composite unobservable variable 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 (a spatial AR(1) process can be approximated as a spatial MA(∞) process). Observe that the dependence between the unobservable factor ε_i and the observable x_i is specified in two different ways across the two examples: in Proposition 2, as a linear regression of the sole innovation term of the spatial AR(1) process over x_i , and in Proposition 3, as a linear regression of the entire unobservable ε_i over x_i .

We denominate the relationship given in (7) a case of *spatial linear endogeneity*. This feature of the model can flexibly represent various patterns of interdependence between the observables of one agent and the unobservables of other agents. We find it is useful to discuss a number of settings, that are of interest for applied researchers, where specification (7) can be used to characterize the endogeneity problem of interest.

Segregated groups. Perhaps the simplest cases are those where it is possible to *ex-ante* partition the population of interest into groups that are subject to “common shocks” that affect observables x_i and unobservables ε_i alike. In a schooling context, for example, the quality of teachers and the overall resources made available to a pupil (x_i) may endogenously depend on their preferences and/or the ability (ε_i) of their classmates. This might be induced via an explicit school-level allocation mechanism, if say more motivated students are assigned the best resources or, conversely, more disadvantaged ones are compensated with some extra support. In this case, \mathcal{C} displays a “fully segregated” group structure derived from that of classrooms.⁹ We place no restriction upon the statistical or the topological relationship between \mathcal{C} and \mathcal{G} : the network of interactions can both transcend, and statistically depend upon, the groups defined by \mathcal{C} . This is exemplified in Graph 1, which is inspired by typical schooling environments: friendships are more likely to occur within than between classrooms.



Graph 1: A Cross-Group Friendship Network

Notes. In this graph, nodes (e.g. i, j, k, ℓ) represents observations, edges denote social interactions (e.g. “friendships”) embodied in \mathcal{G} , whereas groups of observations bound within dash-dotted squares depict a fully segregated characteristics structure \mathcal{C} . Thus, it is for example $g_{i\ell} \neq 0$ but $c_{i\ell} = 0$, and at the same time, $g_{ik} = 0$ but $c_{ik} \neq 0$.

⁹By “fully segregated” group structure, we refer to a topological relationship between any triad of observations $(i, j, k) \in \mathcal{T}^3$ such that if 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 $c_{ik} \neq 0 \Leftrightarrow c_{jk} \neq 0$).

Spatial correlation. We do not restrict \mathcal{C} to fully segregated group structures: we allow it to represent any metric space defined over the set of observations \mathcal{I} , with possibly $c_{ij} \neq 0$ for any pair (i, j) . Consider, for example, a population of competing firms, where x_i represents a firm’s feature such as its size, age or product portfolio; ε_i represents unobserved characteristics such as the local environment where the firm operates or other cost factors; while y_i is some outcome of interest (e.g. book value). One can think of settings where, say because of competition in geographical space, *both* y_i and x_i depend on the unobservables of *all* firms of interest. Thus, \mathcal{C} can be specified as a collection of weights that are the inverse of any two firms’ distance in the geographical, technological or product market space. These ideas may also apply to individuals: for example, in our empirical application about university students we examine structures \mathcal{C} consistent with a spatial correlation of observable characteristics that exhibits, for two observations i and j and some $D > 0$, *distance decay*:

$$\text{Cov}(x_i, x_j) \propto \exp(-D \cdot d_{ij}), \quad (8)$$

where d_{ij} is the geographical distance between the districts where students hail from.¹⁰

Limited network dependence. A leading case is that where the characteristics structure and the network topology overlap to some degree: for example, if $c_{ii} = 1$ for $i = 1, \dots, N$ and $c_{ij} = \psi g_{ij}$ for $i \neq j$ and $\psi \in (0, 1]$. Assuming that the elements in $\varepsilon = (\varepsilon_1, \dots, \varepsilon_N)$ are mutually independent, the spatial correlation in the observables increase along the total strength of two observations’ (say i and j) mutual connections:

$$\text{Cov}(x_i, x_j) \propto \psi g_{ji} \text{Var}(\varepsilon_i) + \psi g_{ij} \text{Var}(\varepsilon_j) + \psi^2 \sum_{k=1}^N g_{ik} g_{jk} \text{Var}(\varepsilon_k). \quad (9)$$

How can this partial overlap emerge? Suppose that the characteristics structure \mathcal{C} is exogenous, while the network \mathcal{G} is endogenous and features *homophily*: the conditional distribution of one link, say g_{ij} , is a negative function of $|x_i - x_j|$. Hence, conditional on the observed network topology the characteristics of any two observations that are “close” in network space are expected to be similar, and the two sets \mathcal{C} and \mathcal{G} to be

¹⁰Observe that as per (7), $\text{Cov}(x_i, x_j) \neq 0$ if at least one of the following three conditions is true: (i) $\text{Cov}(\tilde{x}_i, \tilde{x}_j) \neq 0$; (ii) $\xi \neq 0$ and $c_{ij} \neq 0$; (iii) $\xi \neq 0$ and $\text{Cov}(\varepsilon_i, \varepsilon_j) \neq 0$. Hence, a cross-correlation structure such as (8) may arise through a variety of mechanisms; for example, if x_i is some measure of a student’s high-school background while in college (like in our application) student self-selection occurring differentially across districts is consistent with the third condition above.

correlated (in the simple example above, parameter ψ can be interpreted as a measure of the strength of homophily). Note that although we allow for quite general patterns of correlation between \mathbf{x} and \mathcal{G} , we maintain throughout that the vector of unobserved “errors” $\boldsymbol{\varepsilon}$ is mean-independent of the network topology. In the general version of the model developed in Section 3.2, however, we allow second-order moments of the $\boldsymbol{\varepsilon}$ to be general functions of \mathcal{G} . This may accommodate restricted patterns of endogenous network formation featuring homophily on both observables *and* unobservables, but that maintain the mean-independence property of the error term.¹¹

Simple dependence on SARMA errors. A relatively simple case is that where the structural dependence between observables and unobservables does not *directly* involve the errors of other observation (that is, $c_{ij} = 0$ if $i \neq j$) but the errors follow some exogenous pattern of cross-correlation in network space, similarly to the cases exposed in Propositions 2 and 3. Production functions provide an excellent example: according to the First Order Conditions from firm optimization under perfect competition, firm-level shocks ε_i (but not the shocks of other firms ε_j) “transmit” to inputs x_i ; however, shocks can be spatially correlated (due e.g. to technological similarities), something that can be mistaken for spillover effects. If the spatial correlation process is known, e.g. if it follows some Spatial AutoRegressive Moving Average (SARMA) process with known order and parameters, one can specify a characteristics structure \mathcal{C} and a set of “primitive shocks” $\boldsymbol{v} = (v_1, \dots, v_N)$ such that, for $i = 1, \dots, N$:

$$x_i = \tilde{x}_i + \xi \varepsilon_i = \tilde{x}_i + \xi \sum_{j=1}^N c_{ij} v_j, \quad (10)$$

with the resulting model based on \boldsymbol{v} bearing implications for identification similar to those from the baseline case (7). This is best understood within the discussion of the general model in Section 3.2. The latter also allows for unknown SARMA parameters that are identified through covariance restrictions.

¹¹We plan to study models of this sort in future work; here, we develop some intuition. Suppose that ε_i is the composition of multiple “fundamental” shocks or factors, similarly as in [Zacchia \(2020\)](#), Appendix B (where such factors are interpreted as “technologies”). Suppose further that connections in \mathcal{G} are statistically dependent on how many such fundamental factors are shared by any two agents i and j (e.g. two firms are more likely to link up if they share more technologies). Then, connected pairs are likely to have correlated unobservables. However, this does not necessarily bear any implications for the outcome of interest, e.g. productivity: while some technology combinations can be conducive to higher-than-average outcomes (conditionally), others would lead to lower-than-average ones.

3 Identification

In this section we characterize the conditions for the identification of social effects under our endogeneity specification. We first illustrate them in the simpler case with one observable characteristic and no exogenous effect; later we examine a more general model and the implications of further extensions.

3.1 Bivariate SAR model

Our starting point is the structural relationship (3) combined with the spatial linear endogeneity specification (7). It is more convenient to rewrite both expressions using compact notation. Specifically, (3) renders as:

$$\mathbf{y} = \alpha\mathbf{1} + \beta\mathbf{G}\mathbf{y} + \gamma\mathbf{x} + \boldsymbol{\varepsilon}, \quad (11)$$

while (7) renders as:

$$\mathbf{x} = \tilde{\mathbf{x}} + \xi\mathbf{C}\boldsymbol{\varepsilon}, \quad (12)$$

where \mathbf{y} , \mathbf{x} , $\boldsymbol{\varepsilon}$ and $\tilde{\mathbf{x}} = (\tilde{x}_1, \dots, \tilde{x}_N)$ are arrayed as column vectors, while \mathbf{G} and \mathbf{C} are two *adjacency matrices* of dimension $N \times N$ that array the elements of \mathcal{G} and \mathcal{C} , respectively.¹² For illustrative purposes, we omit for now the subindices denoting the sample size N . So long as our assumptions are upheld, we impose no restrictions on the realizations of \mathbf{G} that are allowed by the data generation process. In particular, \mathbf{G} may represent a single large network to which all agents in \mathcal{I} belong, like the body of students from Bocconi University examined by De Giorgi et al. (2010) and in our empirical application, or a sample of smaller networks with no connections between them as in applications based on separate classes, like e.g. in Bramoullé et al. (2009). Recall, though, that under Assumption 3 no agent is allowed to be “isolated.”

Model (11) is a *spatially autoregressive* (SAR) model according to the classification of spatial econometric models by Elhorst (2014).¹³ Hence, our ensuing discussion of identification applies to a particular version of a SAR model where \mathbf{x} and $\boldsymbol{\varepsilon}$ are related through \mathbf{C} as in (12). Note that there is one case where identification is trivial: if \mathbf{C}

¹²In this particular case, we prefer to duplicate notation to highlight the difference between \mathcal{G} as an unstructured set of edges and \mathbf{G} as its arrayed version (and similarly for \mathcal{C} and \mathbf{C}). We feel that this facilitates comparisons with frameworks and models from other contributions in the literature.

¹³Other authors prefer the denomination “mixed regressive-spatially autoregressive” to remark the presence of \mathbf{x} on the right-hand side of (11). Here we opt for a more concise terminology.

has rank less than N ,¹⁴ researchers may find a matrix \mathbf{B} of dimension $N \times N$ such that $\mathbf{B}\mathbf{C} = \mathbf{0}$ and model (11) can be reshaped as:

$$\mathbf{B}\mathbf{y} = \alpha\mathbf{B}\boldsymbol{\iota} + \beta\mathbf{B}\mathbf{G}\mathbf{y} + \gamma\mathbf{B}\mathbf{x} + \mathbf{B}\boldsymbol{\varepsilon}, \quad (13)$$

a transformed SAR model which is identified and estimable via standard approaches since, by construction, $\mathbf{B}\boldsymbol{\varepsilon}$ is mean-independent of $\mathbf{B}\mathbf{x} = \mathbf{B}\tilde{\mathbf{x}}$. A particular example is that where \mathcal{C} describes a set of “fully segregated” groups with characteristics weights that are identical within groups; hence, $\xi\mathbf{C}\boldsymbol{\varepsilon}$ would feature identical values within a group and \mathbf{B} would be a simple group-demeaning matrix.¹⁵ However, we show in both our Monte Carlo simulations and in our empirical application that transformations of this sort can yield estimates that are too imprecise, arguably because they remove much of the relevant statistical variation even if the rank of \mathbf{C} is fairly low. We argue as a consequence that the approach we propose in this paper is useful beyond those only cases where \mathbf{C} has full (or close to full) rank.

To study identification, it is useful to verify that standard moments in the spirit of Lee (2007a), that are based on the spatial lags of \mathbf{x} , are invalid. For the sake of illustration, we assume that the error term $\boldsymbol{\varepsilon}$ is mean independent of the network \mathbf{G} , the characteristics matrix \mathbf{C} , and the independent component of the variation of \mathbf{x} :

$$\mathbb{E}[\boldsymbol{\varepsilon} | \mathbf{G}, \mathbf{C}, \tilde{\mathbf{x}}] = \mathbf{0}. \quad (14)$$

In addition, we also assume that the error term is conditionally homoscedastic:

$$\mathbb{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T | \mathbf{G}, \mathbf{C}, \tilde{\mathbf{x}}] = \sigma_0^2\mathbf{I}. \quad (15)$$

Under this hypothesis, for any nonnegative integer q we obtain:

$$\begin{aligned} \mathbb{E}\left[(\mathbf{G}^q\mathbf{x})^T \boldsymbol{\varepsilon} \mid \mathbf{G}, \mathbf{C}, \tilde{\mathbf{x}}\right] &= \mathbb{E}\left[(\tilde{\mathbf{x}} + \xi\mathbf{C}\boldsymbol{\varepsilon})^T (\mathbf{G}^q)^T \boldsymbol{\varepsilon} \mid \mathbf{G}, \mathbf{C}, \tilde{\mathbf{x}}\right] \\ &= \xi \cdot \mathbb{E}\left[\boldsymbol{\varepsilon}^T (\mathbf{G}^q\mathbf{C})^T \boldsymbol{\varepsilon} \mid \mathbf{G}, \mathbf{C}, \tilde{\mathbf{x}}\right] \\ &= \xi \cdot \text{Tr}(\mathbf{C}\mathbf{G}^q \cdot \mathbb{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T | \mathbf{G}, \mathbf{C}, \tilde{\mathbf{x}}]) \\ &= \xi\sigma_0^2 \cdot \text{Tr}(\mathbf{C}\mathbf{G}^q). \end{aligned} \quad (16)$$

¹⁴We are grateful to an anonymous referee for highlighting this issue and spelling it in these terms.

¹⁵This is analogous to the within transformation for the removal of fixed effects in panel data or to the data transformations by Bramoullé et al. (2009) that remove network-specific common effects.

In this derivation, the second line exploits the fact that $\mathbb{E} \left[(\mathbf{G}^q \tilde{\mathbf{x}})^T \boldsymbol{\varepsilon} \mid \mathbf{G}, \mathbf{C}, \tilde{\mathbf{x}} \right] = 0$ as per (14), an expression that can be recast in terms of an unconditional moment:

$$\mathbb{E} \left[(\mathbf{x} - \xi \mathbf{C} \boldsymbol{\varepsilon})^T \mathbf{G}^q \boldsymbol{\varepsilon} \right] = 0, \quad (17)$$

for $q = 0, 1, 2$ or higher (we will return later to expression (17) because it helps develop intuition for our identification result). The result derived in (16) highlights why standard moments have a non-zero expectation under our endogeneity specification:¹⁶ if the characteristics structure and the network topology overlap to some extent (i.e. $\text{Tr}(\mathbf{C}\mathbf{G}^q) \neq 0$), the individual unobservables are correlated with the observable characteristics of one's "peers-of-peers." Notice that we derived an explicit expression for the bias. This suggests a natural set of moment conditions $\mathbf{m}(\boldsymbol{\vartheta})$ for the identification of the combined parameters $\boldsymbol{\vartheta} \equiv (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \xi^*)$, where $\xi^* \equiv \xi \sigma_0^2$:

$$\mathbb{E}[\mathbf{m}(\boldsymbol{\vartheta})] = \mathbb{E}[\mathbf{K}^T (\mathbf{y} - \boldsymbol{\alpha} \iota - \boldsymbol{\beta} \mathbf{G} \mathbf{y} - \boldsymbol{\gamma} \mathbf{x}) - \xi^* \boldsymbol{\lambda}] = \mathbf{0}; \quad (18)$$

with:

$$\begin{aligned} \mathbf{K} &\equiv \begin{bmatrix} \iota & \mathbf{x} & \mathbf{G}\mathbf{x} & \mathbf{G}^2\mathbf{x} \end{bmatrix}, \\ \boldsymbol{\lambda} &\equiv \begin{bmatrix} 0 & \text{Tr}(\mathbf{C}) & \text{Tr}(\mathbf{C}\mathbf{G}) & \text{Tr}(\mathbf{C}\mathbf{G}^2) \end{bmatrix}^T. \end{aligned}$$

In fact, identification of $\boldsymbol{\vartheta}$ is possible under quite general conditions.

Proposition 4. Identification of the bivariate SAR model. *Consider the statistical model expressed by equations (11), (12), (14) and (15); and suppose that matrices \mathbf{C} and \mathbf{G} are observed. If the three matrices \mathbf{I} , \mathbf{G} and \mathbf{G}^2 are linearly independent of one another and the traces $\text{Tr}(\mathbf{C})$, $\text{Tr}(\mathbf{C}\mathbf{G})$ and $\text{Tr}(\mathbf{C}\mathbf{G}^2)$ are not simultaneously all zeros, the combined parameters $\boldsymbol{\vartheta} \equiv (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \xi^*)$ are globally identified.*

Proof. Express (18) as a function of an arbitrary parameter vector $\tilde{\boldsymbol{\vartheta}} = (\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\gamma}}, \tilde{\xi}^*)$:

$$\begin{aligned} \mathbb{E}[\mathbf{m}(\tilde{\boldsymbol{\vartheta}})] &= \mathbb{E} \left[(\mathbf{K}_{\tilde{\mathbf{x}}} + \mathbf{K}_{\boldsymbol{\varepsilon}})^T \left[(\mathbf{S}_{\tilde{\mathbf{x}}} + \mathbf{S}_{\boldsymbol{\varepsilon}}) (\boldsymbol{\vartheta}_{\setminus \xi^*} - \tilde{\boldsymbol{\vartheta}}_{\setminus \xi^*}) + \boldsymbol{\varepsilon} \right] - \tilde{\xi}^* \boldsymbol{\lambda} \right] \\ &= \left[\mathbb{E}[\mathbf{K}_{\tilde{\mathbf{x}}}^T \mathbf{S}_{\tilde{\mathbf{x}}} + \mathbf{K}_{\boldsymbol{\varepsilon}}^T \mathbf{S}_{\boldsymbol{\varepsilon}}] \quad \boldsymbol{\lambda} \right] (\boldsymbol{\vartheta} - \tilde{\boldsymbol{\vartheta}}), \end{aligned} \quad (19)$$

¹⁶In an "Addendum" to the Appendix we analyze the bias entailed by conventional methods under our assumptions. This helps appreciate how the bias depends on the topology of the problem.

where $\vartheta_{\xi^*} = (\alpha, \beta, \gamma)$, $\tilde{\vartheta}_{\tilde{\xi}^*} = (\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma})$, and:

$$\begin{aligned}\mathbf{K}_{\tilde{x}} &\equiv \begin{bmatrix} \iota & \tilde{\mathbf{x}} & \mathbf{G}\tilde{\mathbf{x}} & \mathbf{G}^2\tilde{\mathbf{x}} \end{bmatrix}, \\ \mathbf{K}_{\varepsilon} &\equiv \begin{bmatrix} \mathbf{0} & \xi\mathbf{C}\varepsilon & \xi\mathbf{G}\mathbf{C}\varepsilon & \xi\mathbf{G}^2\mathbf{C}\varepsilon \end{bmatrix}, \\ \mathbf{S}_{\tilde{x}} &\equiv \begin{bmatrix} \iota & \mathbf{G}(\mathbf{I} - \beta\mathbf{G})^{-1}(\alpha\iota + \gamma\tilde{\mathbf{x}}) & \tilde{\mathbf{x}} \end{bmatrix}, \\ \mathbf{S}_{\varepsilon} &\equiv \begin{bmatrix} \mathbf{0} & \mathbf{G}(\mathbf{I} - \beta\mathbf{G})^{-1}(\mathbf{I} + \gamma\xi\mathbf{C})\varepsilon & \xi\mathbf{C}\varepsilon \end{bmatrix},\end{aligned}$$

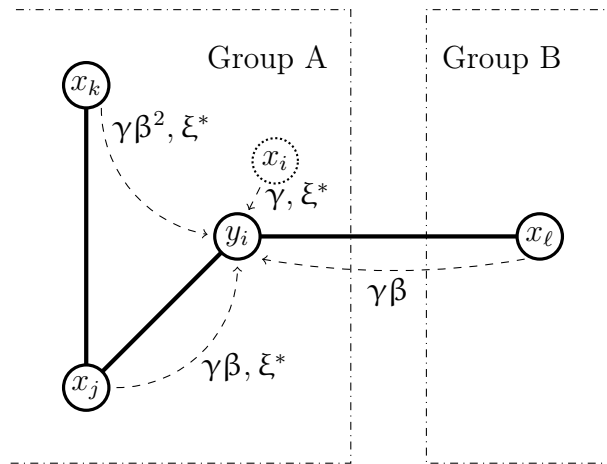
with $\mathbf{K} = \mathbf{K}_{\tilde{x}} + \mathbf{K}_{\varepsilon}$, $\mathbf{y} = (\mathbf{S}_{\tilde{x}} + \mathbf{S}_{\varepsilon})\vartheta_{\xi^*} + \varepsilon$, $\mathbb{E}[\mathbf{K}_{\tilde{x}}^T\mathbf{S}_{\varepsilon}] = \mathbb{E}[\mathbf{K}_{\varepsilon}^T\mathbf{S}_{\tilde{x}}] = \mathbb{E}[\mathbf{K}_{\tilde{x}}^T\varepsilon] = \mathbf{0}$, and $\mathbb{E}[\mathbf{K}_{\varepsilon}^T\varepsilon] = \xi^*\boldsymbol{\lambda}$. The parameter set ϑ is uniquely identified if the only solution that sets (19) at zero is $\tilde{\vartheta} = \vartheta$, which is ensured if matrix $\begin{bmatrix} \mathbb{E}[\mathbf{K}_{\tilde{x}}^T\mathbf{S}_{\tilde{x}} + \mathbf{K}_{\varepsilon}^T\mathbf{S}_{\varepsilon}] & \boldsymbol{\lambda} \end{bmatrix}$ has rank 4. This holds under the maintained conditions. \square

This is a powerful result: it states that if the researcher has some knowledge about the spatial extent of the process which relates the observable characteristics of agents to the unobservables of some others, then the parameters of the SAR model, including the “endogenous” social effect β , can be identified under the same conditions given by Bramoullé et al. (2009): that the network \mathcal{G} is not shaped according to a “fully overlapping” group structure. In addition, it is necessary that the characteristics matrix \mathbf{C} overlaps at least partially with the network, but otherwise it is left unrestricted; it is allowed to assume a group structure or even to coincide with the adjacency matrix \mathbf{G} . The latter condition, however, is generally moot, since its violation would prevent the identification of the combined parameter ξ^* , but not that of the main parameters of interest (α, β, γ) . In fact, if the spatial correlation of individual characteristics is unrelated to the network \mathcal{G} there is no endogeneity, and standard “peers-of-peers” instruments are valid! This point also illustrates why estimates from empirical studies where \mathcal{G} is randomized might still be inconsistent. In fact, if \mathcal{C} is “pervasive,” i.e. for most pairs $(i, j) \in \mathcal{I}^2$ it is $c_{ij} \neq 0$, it is likely that $\text{Tr}(\mathbf{C}\mathbf{G}^q) \neq 0$ even with a random network. We illustrate this point in our empirical application.

We illustrate the intuition behind identification in two ways: algebraic-statistical and graphical. Regarding the former, note that \mathbf{y} can be solved as:

$$\mathbf{y} = (\mathbf{I} - \beta\mathbf{G})^{-1}[\alpha\iota + \gamma(\tilde{\mathbf{x}} + \xi\mathbf{C}\varepsilon) + \varepsilon] = \sum_{s=0}^{\infty} \beta^s \mathbf{G}^s [\alpha\iota + \gamma\tilde{\mathbf{x}} + (\mathbf{I} + \gamma\xi\mathbf{C})\varepsilon]. \quad (20)$$

Thus, by an argument *à la* Bramoullé et al. (2009) the model is identified via a set of instruments of the form $\mathbf{G}^s \tilde{\mathbf{x}}$, which are unfeasible since $\tilde{\mathbf{x}}$ is unobserved. Expression (20) also suggests that if $\tilde{\mathbf{x}}$ and \mathbf{C} are both observed, ξ is separately identified. The appealing nonlinear moments (17), instrumental to our derivation, embed both ideas: they recast the unfeasible moment conditions so that the independent component of \mathbf{x} is backed up from its constituent parts. This is feasible as ξ is identified internally to this approach. One can also gather intuition about the joint identification of all the parameters via a graph. Consider the four observations (i, j, k, ℓ) that are involved in both the network and the group structure represented in Graph 1. According to (20), the variation of y_i is explained by the variation of all the elements in (x_i, x_j, x_k, x_ℓ) , albeit in a different way. This is represented in Graph 2, which “zooms in” the four nodes of interest and in addition, it displays some labeled dashed arrows showing what parameters does each observed characteristic contribute to identify. For example, both nodes j and ℓ are connected to i ; hence, variation in both x_j and x_ℓ helps identify the combined parameter $\gamma\beta$. However, x_j (unlike x_ℓ) also contributes to the identification of ξ^* , because node j (unlike node ℓ) belongs to the same “group” as node i .



Graph 2: Identification: graphical intuition

Notes. This graph elaborates the analysis of nodes (i, j, k, ℓ) from Graph 1, which are related through both a network structure \mathcal{G} (represented by circles and straight lines) and a “grouped” characteristics structure \mathcal{C} (delimited by dash-dotted lines). Directed dashed arrows that connect the variables encapsulated in either node are labeled by the parameter combinations that every observable characteristic on the sending side of the arrow (x_i, x_j, x_k or x_ℓ , with y_i always on the receiving side) contributes to identify per (20). Variable x_i is enclosed in a dotted circle to remark that it does not arise from a node (an observation) different from y_i ’s.

3.2 Multivariate SDM model

The identification result illustrated by Proposition 4 is restricted to a simple model under very restrictive structure of the error terms and assumptions. We thus turn to the discussion of the following, more general model:

$$\mathbf{y} = \alpha\mathbf{1} + \beta\mathbf{G}\mathbf{y} + \mathbf{X}\boldsymbol{\gamma} + \mathbf{G}\mathbf{X}\boldsymbol{\delta} + \boldsymbol{\varepsilon}, \quad (21)$$

where \mathbf{X} is a $N \times K$ matrix of K random observable characteristics, $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_K)$ is the vector of K direct effects associated with each of these, while $\boldsymbol{\delta} = (\delta_1, \dots, \delta_K)$ are the K “contextual effects.” Under Elhorst’s (2014) classification, this is a standard multivariate “Spatial Durbin Model” (SDM); when \mathbf{G} is row-normalized, this model is known as the “linear-in-means” in the peer effects literature. Model (21) can easily follow from an extension of our theoretical framework, where nature initially draws $(\mathbf{X}, \boldsymbol{\varepsilon}, \mathcal{G})$ from some more general distribution $\mathcal{F}(\cdot)$.¹⁷ This model is interesting in two non-alternative cases: (i) all observable characteristics \mathbf{X} are endogenous, structurally dependent on $\boldsymbol{\varepsilon}$, (ii) researchers aim for consistent estimation of all parameters in $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$, in addition to β . In fact, if researchers are only interested in the identification and consistent estimation of β , a single exogenous observable characteristic \mathbf{x} suffices under the conditions characterized by Bramoullé et al. (2009).

In addition, we allow for a more general structure of the error term $\boldsymbol{\varepsilon}$, which we make explicit through the following assumptions.

Assumption 4. Primitive shocks: *there exists a set of N “primitive” i.i.d. shocks $\mathbf{v} \equiv (v_1, \dots, v_N)^\top$ such that $\mathbb{E}[\mathbf{v}] = \mathbf{0}$ and, for some $d > 0$, $\mathbb{E}[|v_i|^{4+d}] < \infty$ for $i = 1, \dots, N$.*

Assumption 5. SARMA Unobservables: *the unobservable characteristics follow a stationary Spatial Autoregressive Moving Average process of order (A, M) :*

$$\boldsymbol{\varepsilon} = (\mathbf{I} - \phi_1\mathbf{F}_1 - \phi_2\mathbf{F}_2 - \dots - \phi_A\mathbf{F}_A)^{-1} (\mathbf{I} + \psi_1\mathbf{E}_1 + \psi_2\mathbf{E}_2 + \dots + \psi_M\mathbf{E}_M) \mathbf{v},$$

where $(\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_A)$ and $(\mathbf{E}_1, \mathbf{E}_2, \dots, \mathbf{E}_M)$ are two possibly identical sets of linearly independent $N \times N$ matrices; $\mathbf{I} - \sum_{a=1}^A \phi_a \mathbf{F}_a$ and $\mathbf{I} + \sum_{m=1}^M \psi_m \mathbf{E}_m$ are both invertible, and the associated parameters lie within the unit circle: $\|\boldsymbol{\phi}\|_2 < 1$ and $\|\boldsymbol{\psi}\|_2 < 1$.

¹⁷The analysis, including Proposition 1, would proceed largely unchanged. As already mentioned, one can think of x_i from Section 2 as the composition of K factors, with $\gamma_0 x_i = (1 - \mu)(\mathbf{X}\boldsymbol{\gamma} + \mathbf{G}\mathbf{X}\boldsymbol{\delta})$.

Together, these two assumptions characterize the stochastic properties of the error term, which is allowed to have a very general spatial correlation structure expressed in terms of a sequence of “primitive” well-behaved shocks. The spatially autoregressive component of the error term is defined by the sequence of matrices $(\mathbf{F}_1, \dots, \mathbf{F}_A)$ and the parameter set $\boldsymbol{\phi} = (\phi_1, \dots, \phi_A)$; the moving average part is encapsulated by matrices $(\mathbf{E}_1, \dots, \mathbf{E}_M)$ and parameters $\boldsymbol{\psi} = (\psi_1, \dots, \psi_M)$. Both matrix sequences are allowed to coincide and depend on the network structure; a leading case that we allow is $\mathbf{F}_a = \mathbf{G}^a$ for $a \leq A$ and $\mathbf{E}_m = \mathbf{G}^m$ for $m \leq M$. While this specification is quite flexible, we are especially interested in the Spatial Moving Average (SMA) component of the process. If the spatially autocorrelated component is missing ($\boldsymbol{\phi} = \mathbf{0}$) the SMA process alone implies zero spatial autocorrelation for observations unrelated through the \mathbf{E}_m matrices. We find this empirical property to be a good approximation of some real-world stylized facts about variables that are diffused in networks.¹⁸ Although our identification results extend to any SARMA process, our estimation framework and Monte Carlo simulations specialize to a simple SMA(1) process, or $(A, M) = (0, 1)$.

The next assumption generalizes expression (7), which characterizes the spatial extent of endogeneity, to the multivariate case. In particular, we associate a different characteristics matrix \mathbf{C}_k to each of the K variables in \mathbf{X} . For the sake of exposition we assume that $\mathbf{C}_k \neq \mathbf{0}$ for $k = 1, \dots, K$. This facilitates the ensuing discussion about the identification of the multiplicative parameters associated with these matrices.

Assumption 6. Multivariate Spatial Linear Endogeneity: *each column of \mathbf{X} is given, for $k = 1, \dots, K$, by:*

$$\mathbf{X}_{*,k} = \tilde{\mathbf{x}}_k + \xi_k \mathbf{C}_k \boldsymbol{\nu}, \quad (22)$$

where $\xi_k \in \mathbb{R}$, $\mathbf{C}_k \neq \mathbf{0}$ is an $N \times N$ characteristics matrix specific to the k -th observable characteristic, while $\tilde{\mathbf{x}}_k$ is a random vector of length K with finite mean. In addition, we assume that for any two $k, k' = 1, \dots, K$ with possibly $k = k'$, the probability limit defined as $\Xi_{kk'} \equiv \text{plim} N^{-1} \sum_{i=1}^N (\tilde{x}_{ki} - \mathbb{E}[\tilde{x}_{ki}])(\tilde{x}_{k'i} - \mathbb{E}[\tilde{x}_{k'i}])$ is finite.

¹⁸In a study about the health outcomes of children, Christakis and Fowler (2013) find that most variables present a spatial autocorrelation in the space of friendship network up to two degrees of distance. Zacchia (2020) observes the same property for the R&D investment of high-tech firms that are connected through research collaborations. In addition, he argues that this property can follow from an underlying SMA(1) process of technological shock, and that it is a good approximation of a model of network formation driven by a homophily dynamic, where two firms link up with some probability only if their unobservables are similar.

Observe a difference relative to the simpler one-characteristic case given in (7): in the latter, matrix \mathbf{C} multiplies the error terms ε_i 's of the model; in (22), each of the K characteristic matrices \mathbf{C}_k multiplies the “primitive” shocks v_i 's. We believe that this specification, together with our SARMA specification of the error term, is more flexible and can capture a number of realistic specifications of endogeneity, as already discussed in Section 2.3.¹⁹ However, our results about identification and estimation would be easy to extend to a setup where the specification of endogeneity in (22) were to involve ε (which would still follow a generalized SARMA process) instead of \mathbf{v} .

Our final identification-related assumption is the following.

Assumption 7. Exogeneity of the spatial structures: *conditional upon the set $\mathcal{S} = \{\mathbf{G}; \mathbf{C}_1, \dots, \mathbf{C}_K; \tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_K\}$ that comprises: (i) the network adjacency matrix, (ii) the K characteristics matrices, as well as (iii) the K independent component of the individual characteristics, the primitive shocks have mean zero and are homoscedastic:*

$$\mathbb{E}[\mathbf{v} | \mathcal{S}] = \mathbf{0}, \quad (23)$$

$$\mathbb{E}[\mathbf{v}\mathbf{v}^T | \mathcal{S}] = \sigma^2 \mathbf{I}. \quad (24)$$

This assumption generalizes (14) and (15) to the general model. As in the previous discussion about the simpler case, this assumption allows to isolate the source of endogeneity introduced via (22) from other confounding factors, such as the endogeneity of the networked structure of interaction or that of the characteristics structure.

Before expressing our main identification result, an important remark is in order. The assumptions that we have introduced so far, especially Assumptions 5, 6 and 7, characterize the *complete* structure of dependence between the individual observable and unobservable characteristics of any two observations, in whatever metric space of interest, for example the pairwise distance in the network $(\mathcal{I}, \mathcal{G})$. Thus, our framework evades the critique by Goldsmith-Pinkham and Imbens (2013), who lament the lack of *general* results on cross-observation dependence that would allow identification and inference the entire data generation process for those settings where all observations are related through a large network (which we allow e.g. in our empirical application). Although our hypotheses about cross-observation dependence are restrictive, they are also very flexible, and the key parameters that characterize them are identified.

¹⁹See e.g. the discussion around (10). There, $\mathbf{C} = \left(\mathbf{I} - \sum_{a=1}^A \phi_a \mathbf{F}_a\right)^{-1} \left(\mathbf{I} + \sum_{m=1}^M \psi_m \mathbf{E}_m\right)$.

We are now ready to characterize our main result. In what follows, we gather the parameters that measure the extent of spatial endogeneity as $\boldsymbol{\xi} = (\xi_1, \dots, \xi_K)$.

Theorem 1. General Identification Result. *Under Assumptions 1-7, the parameters $\boldsymbol{\theta} \equiv (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\xi}, \boldsymbol{\phi}, \boldsymbol{\psi}, \sigma^2)$ are globally identified if $\boldsymbol{\beta}\boldsymbol{\gamma}_k + \boldsymbol{\delta}_k \neq 0$ for at least one $k = 1, \dots, K$; and the following three conditions hold simultaneously:*

- (i) *the matrices \mathbf{I} , \mathbf{G} , \mathbf{G}^2 and \mathbf{G}^3 are linearly independent of one another;*
- (ii) *for $k = 1, \dots, K$, the four traces gathered in the following vector:*

$$\boldsymbol{\lambda}_k \equiv \left[\text{Tr}(\mathbf{C}_k) \quad \text{Tr}(\mathbf{C}_k \mathbf{G}) \quad \text{Tr}(\mathbf{C}_k \mathbf{G}^2) \quad \text{Tr}(\mathbf{C}_k \mathbf{G}^3) \right]^T$$

are not simultaneously all zeros;

- (iii) *the researcher can observe some $P \geq 1 + A + M$ matrices $\{\mathbf{P}_p\}_{p=1}^P$ of size $N \times N$ that are all linearly independent of one another.*

Proof. See the Appendix. The proof extends the identification arguments by Lee and Liu (2010). Proposition 4 is a restricted case of this Theorem. \square

Theorem 1 provides a general identification result for linear-in-means models that feature contextual effects, when the observable characteristics of individuals, the error terms and the interaction structure itself are structurally dependent. In addition, we allow for an error term which is allowed to follow a very general stochastic process, and we show that under specific conditions the associated parameters are identified. The latter is, to the best of our knowledge, a novel result in the spatial econometrics literature, which so far has prevalently examined models whose errors follow simple spatially autoregressive processes.²⁰ It is useful to discuss how the conditions established in the theorem enable identification of the model's parameters. We first focus on the "linear" parameters of (21) and the parameters $\boldsymbol{\xi}$ that measure endogeneity; then we move to the components of the SARMA structure of the error term.

First, observe that the condition that social and contextual effects do not cancel out for at least one observable characteristic, i.e. $\exists k \in \{0, 1, \dots, K\} : \boldsymbol{\beta}\boldsymbol{\gamma}_k + \boldsymbol{\delta}_k \neq 0$, is standard in linear-in-means models (or else $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ cannot be distinguished). Next,

²⁰For example, Kapoor et al. (2007); Kelejian and Prucha (2010); Drukker et al. (2013), among the others, analyze SAR(1) disturbances, while Lee and Liu (2010) consider higher order SAR processes.

to motivate condition (i) return to our simplified analysis in Proposition 4. There, if matrix \mathbf{G}^3 is linearly independent from \mathbf{I} , \mathbf{G} and \mathbf{G}^2 , another moment condition like (16) with $q = 3$ can be exploited for identification. In the general case we exploit QK sets of moments of the following kind, for $q = 1, \dots, Q$, $Q \geq 4$ and $k = 1, \dots, K$:

$$\mathbb{E} [\mathbf{x}_k^T \mathbf{G}^{q-1} \boldsymbol{\varepsilon} - \lambda_{qk}] = 0, \quad (25)$$

where λ_{qk} depends on the assumed SARMA process. Extending the intuition illustrated via Graph 2, the moments based on higher powers of \mathbf{G} allow identification of the contextual effect $\boldsymbol{\delta}$. Condition (ii), which is necessary for the identification of $\boldsymbol{\xi}$, corresponds with the “not all-zero traces” requirement from Proposition 4. Like that one, this condition is not very interesting: were it not to hold, endogeneity would not be a salient problem in the model.²¹ The variance components are identified through standard covariance restrictions of the following kind, for $p = 1, \dots, P$:

$$\mathbb{E} [\boldsymbol{\varepsilon}^T \mathbf{P}_p \boldsymbol{\varepsilon} - \lambda_p] = 0, \quad (26)$$

where again λ_p may vary across cases. Clearly, condition (iii) is necessary to rule out collinearity between the P moments; this also applies to the λ_p elements. A natural choice for the moment matrices is $\mathbf{P}_p = \mathbf{G}^{p-1}$, especially where P is small.

We conclude our treatment of identification with a discussion on the realism and applicability of our assumptions. Our hypotheses about the spatial correlation structure of the error term (Assumption 5), the formulation of endogeneity (Assumption 6) and the network of social interactions (Theorem 1) are quite general and can accommodate a wide variety of settings. Our key assumptions are two: first, that the error term is mean-independent of the key topological structures (Assumption 7); second, that the researchers know the characteristics matrices \mathbf{C}_k up to ξ_k for $k = 1, \dots, K$. Exogeneity of the network is a standard condition in this literature; as argued, while in future work this assumption shall be relaxed, in this paper it allows us to focus on our “spatial” endogeneity mechanism. Conversely, information about the characteristic structures for each covariate may be difficult to obtain in empirical applications; while it is not required that the *intensity* of endogeneity is known (the ξ_k parameters are identified and can be estimated) lack of knowledge about its spatial *structure* can lead to misspecification. In such cases, we advocate using our approach as a testing

²¹In this case, the model is still identified following the elimination of selected elements of $\boldsymbol{\xi}$.

tool (for example, as part of robustness checks) to verify that the results are not driven by the spatial cross-correlation between observed and unobserved characteristics. We illustrate our approach in our empirical application, where conventional estimators deliver statistically significant social effects, but we are concerned whether these are driven by some instance of correlated effects.

3.3 Extensions

Finally, we analyze two simple extensions of our framework. First, we show how it can accommodate multiple networks and the relative fixed effects. Next, we discuss how the primitive parameters μ and ν from our analytical framework, which are combined in β , can be identified under certain conditions.

Network-level fixed effects

The use of the third power of \mathbf{G} bears some analogies with the scenario analyzed by Bramoullé et al. (2009), where the adjacency matrix represents a set of disconnected networks, to each of which is associated a separate fixed effect, and where the use of indirect connections of third degree is necessary once such fixed effects are partialled out. The difference is that here, it is necessary to remove the endogenous component expressed in (22) too. The following corollary is consequent to this observation.

Corollary 1. *If the model of interest is:*

$$\mathbf{y} = \mathbf{D}\boldsymbol{\alpha}^* + \beta\mathbf{G}\mathbf{y} + \mathbf{X}\boldsymbol{\gamma} + \mathbf{G}\mathbf{X}\boldsymbol{\delta} + \boldsymbol{\varepsilon}, \quad (27)$$

where \mathbf{D} represents a set of D dummy variables, each for a separate component of the network \mathcal{G} , and $\boldsymbol{\alpha}^* = (\alpha_1, \dots, \alpha_D)$ is a vector of associated fixed effects, the parameters $\boldsymbol{\theta} \equiv (\boldsymbol{\alpha}^*, \beta, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\xi}, \boldsymbol{\phi}, \boldsymbol{\psi}, \sigma^2)$ are identified if, in addition to the conditions expressed in Theorem 1, also matrix \mathbf{G}^4 is linearly independent of matrices \mathbf{I} , \mathbf{G} , \mathbf{G}^2 and \mathbf{G}^3 .

Proof. This follows straightforwardly from “network differencing” equation (27) by pre-multiplying the data (\mathbf{X}, \mathbf{y}) by $\mathbf{I} - \mathbf{G}$ as in Bramoullé et al. (2009). The identification of the differenced model would follow as per our previous analysis with $\boldsymbol{\alpha} = 0$; the resulting moments are a function of \mathbf{G}^4 which thus must be linearly independent of its lower powers. The fixed effects $\boldsymbol{\alpha}^*$ are residually identified as a subnetwork-specific set of intercepts. \square

Under this approach, also the error term is transformed as $(\mathbf{I} - \mathbf{G}) \boldsymbol{\varepsilon}$. This poses no problems to identification and estimation, as the terms λ_{qk} in (25) and λ_p in (26) can still be calculated, their expressions incorporating the transformation matrix $(\mathbf{I} - \mathbf{G})$.

Identification of μ and ν

In our framework, parameter β represents a composite equilibrium effect: it encloses the direct effect of peers' effort, ν , amplified by the equilibrium response of individual effort, $(1 - \mu)^{-1}$. Because of Assumption 3 (row-normalization of \mathbf{G}) the two parameters μ and ν disappear from the reduced form equilibrium equation. However, note that when this hypothesis is dropped, under our framework (11) becomes:

$$\mathbf{y} = (\alpha - \zeta) \boldsymbol{\iota} + \beta \mathbf{G} \mathbf{y} + \gamma \mathbf{x} + \zeta \bar{\mathbf{g}} + \boldsymbol{\varepsilon}, \quad (28)$$

where $\zeta \equiv (1 - \mu)^{-1} \nu \log \mu$ and $\bar{\mathbf{g}} \equiv \mathbf{G} \boldsymbol{\iota}$ 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 $\exp(\zeta/\beta) = \mu$, if the observable characteristics x_i 's and the network \mathcal{G} are exogenous the primitive parameters μ and ν are separately identified in (28). The 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).²² 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 the latter to be a realistic hypothesis.²³

In our framework, μ and ν are separately identified also under a mildly restrictive instance of endogeneity.

Corollary 2. *Under the conditions expressed by Theorem 1 but Assumption 3, if $\bar{\mathbf{g}}$ is linearly independent of the unit vector $\boldsymbol{\iota}$ or any other covariate $\mathbf{X}_{\cdot,k}$ and, in addition, $\mathbb{E}[\bar{\mathbf{g}}^T \boldsymbol{\varepsilon}] = 0$ holds, then parameters μ and ν are separately identified.*

Proof. Re-define the residual as $\boldsymbol{\varepsilon}(\boldsymbol{\theta}, \zeta) = \mathbf{y} - (\alpha - \zeta) \boldsymbol{\iota} - \beta \mathbf{G} \mathbf{y} - \mathbf{X} \boldsymbol{\gamma} - \mathbf{G} \mathbf{X} \boldsymbol{\delta} - \zeta \bar{\mathbf{g}}$, and add $\mathbb{E}[\bar{\mathbf{g}}^T \boldsymbol{\varepsilon}(\boldsymbol{\theta}, \zeta)] = 0$ to the moments from the proof of Theorem 1. One can verify that this does not affect the rank properties of the matrices examined therein. \square

²²Note that the exact relationship between β , μ and ν depends on functional form assumptions of our model, but the intuition is more general.

²³If 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 the approach taken in studies of R&D spillovers, since researchers can typically observe the R&D expenditures of firms.

Essentially, if the observable characteristics are endogenous as per Assumption 6, *but the intensity of individual connections is independent of individual unobservables*, $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ can be separately identified by adding the additional regressor \bar{g}_i . Note that some form of statistical dependence of the adjacency matrix \mathbf{G} on the characteristics matrices \mathbf{C}_k is still allowed. Scenarios where the identifying assumption is violated are obvious: for example, a very skilled pupil or a very successful firm may find themselves with more (or more intense) connections. In future work, we plan to examine under what conditions can $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ be separately identified even if that hypothesis fails.

4 Estimation

The moment conditions that support our main identification results lend themselves naturally to GMM estimation. In this section we describe how the estimation framework introduced by Lee (2007a) can be adapted to our assumed forms of endogeneity. In doing so, we specialize – as mentioned earlier – to a simple stochastic process that governs our the error term: a spatial moving average of first degree. This facilitates the asymptotic analysis, yet the results can be extended to any SARMA process.

Assumption 8. SMA(1) Unobservables: $\boldsymbol{\phi} = \mathbf{0}$ and $\psi_m = 0$ for $m \geq 2$.

In what follows, we write $\boldsymbol{\psi} = \boldsymbol{\psi}_1$ and $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\xi}, \boldsymbol{\psi}, \sigma^2)$. We also denote the true parameter values as $\boldsymbol{\theta}_0$, we introduce N subscripts, and we define the following matrices for $q = 1, \dots, Q$:

$$\mathbf{Q}_{q,N} \equiv \mathbf{X}_N^T \mathbf{G}_N^{q-1}.$$

Our GMM estimator is based on a set of $1 + QK + P$ moments conditions, with $Q \geq 4$ and $P \geq 2$, that explicitly correct the bias of conventional moments:²⁴

$$\mathbb{E}[\mathbf{m}_N(\boldsymbol{\theta}_0) - \boldsymbol{\lambda}_N(\boldsymbol{\theta}_0)] = \mathbf{0}. \quad (29)$$

To better describe our moment conditions, express the structural residual as:

$$\boldsymbol{\varepsilon}_N(\boldsymbol{\theta}) = \mathbf{y}_N - \boldsymbol{\alpha}\iota_N - \boldsymbol{\beta}\mathbf{G}_N\mathbf{y}_N - \mathbf{X}_N\boldsymbol{\gamma} - \mathbf{G}_N\mathbf{X}_N\boldsymbol{\delta},$$

²⁴We simulated an estimation of our model using moment conditions (17); however, this exercise is outperformed by the main simulation which is based on the bias-adjusted moments (18), that we discuss in Section 5. Both sets of moments follow from the same data generation process and thus should be equivalent, but the linear ones are computationally more convenient.

hence, it is:

$$\mathbf{m}_N(\boldsymbol{\theta}) = \left[\boldsymbol{\iota}^\top \boldsymbol{\varepsilon}_N(\boldsymbol{\theta}) \quad \cdots \quad \boldsymbol{\varepsilon}_N^\top(\boldsymbol{\theta}) \mathbf{Q}_{q,N}^\top \quad \cdots \quad \boldsymbol{\varepsilon}_N^\top(\boldsymbol{\theta}) \mathbf{P}_{p,N} \boldsymbol{\varepsilon}_N(\boldsymbol{\theta}) \quad \cdots \right]^\top,$$

for $q = 1, \dots, Q$ and $p = 1, \dots, P$. As for vector $\boldsymbol{\lambda}_N(\boldsymbol{\theta})$, its first element is given by $\lambda_{1,N}(\boldsymbol{\theta}) = 0$, while the others are:

$$\lambda_{1+qk,N}(\boldsymbol{\theta}) \equiv \sigma^2 \xi_k \text{Tr} \left[\mathbf{C}_{k,N}^\top \mathbf{G}_N^{q-1} (\mathbf{I}_N + \boldsymbol{\psi} \mathbf{E}_N) \right],$$

for $q = 1, \dots, Q$ and $k = 1, \dots, K$; and:

$$\lambda_{1+QK+p,N}(\boldsymbol{\theta}) \equiv \sigma^2 \text{Tr} \left[(\mathbf{I}_N + \boldsymbol{\psi} \mathbf{E}_N)^\top \mathbf{P}_{p,N} (\mathbf{I}_N + \boldsymbol{\psi} \mathbf{E}_N) \right],$$

for $p = 1, \dots, P$. For some $\boldsymbol{\theta}$, the sample moments are, simply:

$$\bar{\mathbf{m}}_N(\boldsymbol{\theta}) \equiv \frac{1}{N} [\mathbf{m}_N(\boldsymbol{\theta}) - \boldsymbol{\lambda}_N(\boldsymbol{\theta})], \quad (30)$$

while our GMM estimator $\hat{\boldsymbol{\theta}}_{GMM}$ is the usual minimizer in the parameter space $\boldsymbol{\Theta}$:

$$\hat{\boldsymbol{\theta}}_{GMM} = \arg \min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \bar{\mathbf{m}}_N^\top(\boldsymbol{\theta}) \mathbf{W}_N \bar{\mathbf{m}}_N(\boldsymbol{\theta}), \quad (31)$$

where \mathbf{W}_N is a weighting matrix. We derive the asymptotic properties of the estimator under the following additional assumptions.

Assumption 9. Bounded Parameter Space: $\boldsymbol{\Theta}$ is bounded.

Assumption 10. Probability Limits of the Covariates: the independent component of x_{ik} are such that $N^{-1} \sum_{i=1}^N (\tilde{x}_{ik} - \mathbb{E}[\tilde{x}_{ik}]) = o_P(1)$ for all $k = 1, \dots, K$.

Assumptions 9 and 10 are typical regularity conditions that are necessary to ensure consistency of the GMM estimator.

Assumption 11. Bounded Characteristics: matrix $\mathbf{C}_{k,N}$ is bounded by $\bar{C}_k < \infty$, that is $\sum_{j=1}^N c_{k,ij} < \bar{C}_k$ for $i = 1, \dots, N$, for all $k = 1, \dots, K$.

Assumption 12. Bounded Adjacencies: the network's adjacency matrix \mathbf{G}_N and its corresponding Leontiev inverse $(\mathbf{I}_N - \beta_0 \mathbf{G}_N)^{-1}$ are uniformly bounded in both row and column sums in absolute value.

Assumption 13. Bounded Moment Matrices: *all the matrices $(\mathbf{Q}_{1,N}, \dots, \mathbf{Q}_{Q,N})$ and $(\mathbf{P}_{1,N}, \dots, \mathbf{P}_{P,N})$ used in the moment conditions are all uniformly bounded in both row and column sums in absolute value.*

Assumptions 11-13 all ensure that the moments in question have finite variance. Note that Assumptions 12-13 have their analogues in Lee (2007a), while Assumption 11 is specific to our framework. Also observe that bounded adjacencies are implied by the row normalization of \mathbf{G}_N . Yet Assumption 12 may be useful when row normalization is dropped, e.g. if interest falls on the separate identification of $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$.

Under the maintained assumptions, one can derive standard asymptotic properties for our GMM estimator. Extending the next result to more general SARMA processes or to heteroscedastic primitive shocks is conceptually straightforward but tedious.

Theorem 2. Asymptotics of the GMM estimator. *Under Assumptions 1-13, and holding the identification conditions detailed in Theorem 1, $\widehat{\boldsymbol{\theta}}_{GMM}$ is a consistent estimator of $\boldsymbol{\theta}_0$ and has the following limiting distribution:*

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

where (i) $\boldsymbol{\Omega}_0 \equiv \text{plim} \frac{1}{N} \text{Var} [\mathbf{m}_N(\boldsymbol{\theta}_0)]$, (ii) $\mathbf{J}_0 \equiv \text{plim} \frac{\partial}{\partial \boldsymbol{\theta}^T} \overline{\mathbf{m}}_N(\boldsymbol{\theta}_0)$, (iii) $\mathbf{W}_0 \equiv \text{plim} \mathbf{W}_N$.

Proof. See the Appendix. The proof is based on the results by Lee (2007a), which in turn rely on White (1996) as well as Kelejian and Prucha (2001). \square

The choice of the optimal weighting matrix \mathbf{W}_N is informed by the same considerations advanced by Lee (2007a), to whom we refer for the details (we use a parallel notation for the moment matrices $\mathbf{Q}_{q,N}$ and $\mathbf{P}_{p,N}$ for ease of comparison). However, in general, this GMM estimator is not efficient, as efficiency is also affected by the choice of moments used for estimation. We explore how this affects the finite sample performance of the estimator in the next section. Still, one can construct an asymptotically efficient estimator, as Lee and Liu (2010) did for a spatial autoregression model with exogenous individual characteristics \mathbf{X} . As in the theory of optimal instruments, the moments must be chosen so that the Jacobian of the moment conditions is equal to the inverse of the asymptotic variance. Because the moment functions depend on $\boldsymbol{\theta}$ itself, the computation of such moments requires an initial \sqrt{N} -consistent estimator; thus, an efficient estimator would proceed in two steps. We leave the analysis of such an estimator for future work.

5 Monte Carlo

We evaluate the performance of our GMM estimator across Monte Carlo simulations. These are all based on a minimal data generation process (d.g.p.): the bivariate SAR model (11) without contextual effects, combined with the spatial linear endogeneity specification expressed in (7). We study a number of “experiments,” that is groups of simulations, that differ by the *type* of the characteristic matrix \mathbf{C} used in the d.g.p. to generate the observable characteristic \mathbf{x} . Examples of different such types include the identity matrix, a matrix with a block structure as in the example from Graph 1, or functions of a network to which all simulated observations belong. To minimize the dependence of our results from specific topologies, in all the simulations or repetitions of an experiment, we generate a new network adjacency matrix \mathbf{G} ; we do the same for characteristic structures \mathbf{C} with possibly irregular types, e.g. the network functions. More specifically, all \mathbf{G} matrices are randomly generated through the “small-world” algorithm by Watts and Strogatz (1998) with constant parameters.²⁵ We also let the error term follow a simple first order spatial moving average process as per Assumption 8, and we set $\mathbf{E} = \mathbf{G}$, i.e. the SMA(1) process is governed by the network.

The following expression for the simulated values of \mathbf{y} summarizes our d.g.p.:

$$\mathbf{y} = (\mathbf{I} - \beta\mathbf{G})^{-1} [\alpha\mathbf{1} + \gamma(\tilde{\mathbf{x}} + \xi\sigma\mathbf{C}\mathbf{v}_y) + \chi\mathbf{w} + \sigma(\mathbf{I} + \psi\mathbf{G})\mathbf{v}_y]$$

where \mathbf{w} is a vector of N independent draws from the continuous uniform distribution with support on $(0, 1)$, which we leverage to compare the performance of our estimator against one based on “external instruments;” χ is a real parameter; \mathbf{v}_y is a vector of N independent draws from a standard normal distribution; while $\tilde{\mathbf{x}}$ is generated as:

$$\tilde{\mathbf{x}} = 0.3 \cdot \mathbf{H}\mathbf{v}_x$$

where \mathbf{v}_x are yet N more independent draws from the standard normal distribution,

²⁵By this algorithm, all observations are first ordered along a line and connected to an even number of B neighbors; this defines an initial set of pairwise binary associations $g_{0,ij} = g_{0,ji} \in \{0, 1\}$, with $g_{0,ii} = 0$ for every node i . Subsequently, all links are subject to random rewiring (the link is deleted, and one of the two involved nodes becomes connected with a random third node) with probability φ . This procedure yields a new topology $g_{1,ij} = g_{1,ji}$ (still without self-links) with associated adjacency matrix \mathbf{G}_1 . The final row-normalized adjacency matrix is obtained as $\mathbf{G} = \text{diag}(\mathbf{G}_1\mathbf{1})\mathbf{G}_1$. In all our simulations we set $B = 2$ and $\varphi = 0.25$. These choices combined ensure a good overlap between the network adjacency matrices \mathbf{G} and the characteristic matrix \mathbf{C} across all our experiments.

and \mathbf{H} is an $N \times N$ matrix that can introduce spatial correlation in the independent component of \mathbf{x} . Like \mathbf{C} , the type of matrix \mathbf{H} is fixed within an experiment but the actual topology can vary across repetitions. In all our simulations we set $N = 500$.

In every experiment, we compare eight estimators against one another. The first four estimators are variations of our proposed GMM estimator, that are summarized as follows; (1) a version based on a smaller set of moments, $Q = 3$ and $P = 2$, where $\mathbf{P}_1 = \mathbf{I}$ and $\mathbf{P}_2 = \mathbf{G}$; (2) a version with more moments, $Q = 4$ and $P = 3$, where \mathbf{P}_1 and \mathbf{P}_2 are as before and in addition, $\mathbf{P}_3 = \mathbf{G}^2$; (3) a version with an even larger set of moments, $Q = 5$ and $P = 4$, with additionally $\mathbf{P}_4 = \mathbf{G}^3$; (4) a version like the latter, but where estimation is based on a “misspecified” characteristic matrix \mathbf{C}_e , as detailed later. All four GMM estimators return estimates for $(\alpha, \beta, \gamma, \chi, \xi, \psi, \sigma)$. The other four estimators are a naïve OLS estimator which takes $\mathbf{G}\mathbf{y}$, \mathbf{x} and \mathbf{w} as independent variables, and three different 2SLS estimators based on the following set of instruments:

$$\mathbf{Z} \equiv \begin{bmatrix} \iota & \mathbf{w} & \mathbf{z} & \mathbf{G}\mathbf{z} & \mathbf{G}^2\mathbf{z} \end{bmatrix}. \quad (32)$$

In (32), it is either (a) $\mathbf{z} = \mathbf{x}$, yielding an 2SLS estimator akin to the one proposed by Bramoullé et al. (2009); (b) $\mathbf{z} = \mathbf{G}\mathbf{w}$, yielding a 2SLS estimator solely based on the exogenous regressor and its spatial lags; or (c) $\mathbf{z} = \mathbf{B}\mathbf{x}$, where \mathbf{B} is a matrix such that $\mathbf{B}\mathbf{C} = \mathbf{0}$ as per the discussion in Section 3.1, yielding a consistent 2SLS estimator based on transformations of \mathbf{x} that are purged of the endogenous component.²⁶ Thus, we compare our GMM estimator to several simpler alternatives that are likely to occur in the empirical practice. These simpler estimators return estimates for $(\alpha, \beta, \gamma, \chi)$.

We summarize the results of our simulated estimates in Tables 1 and 2. For every experiment-estimator combination, we report the median and – in parentheses – the standard deviation of point estimates for the estimated parameters of interest across 1,000 repetitions. We first describe Experiment 1 from Table 1, our “baseline” case, in depth. In all its repetitions, we set $\mathbf{H} = \mathbf{I} + \mathbf{G}$ (hence the independent component of \mathbf{x} displays cross-correlation) and \mathbf{C} is similarly constructed, but instead of \mathbf{G} , we add to the identity matrix a different matrix derived from a “small world algorithm” as per footnote 25. Because of how that algorithm works, \mathbf{H} and \mathbf{C} are correlated, though distinct. The three GMM estimators that are based on a correctly specified matrix \mathbf{C} , as expected, all display a good performance at estimating the real parameters set

²⁶To accommodate those \mathbf{C} matrices that are by construction not of full rank, we specify \mathbf{B} as the annihilator matrix based on the Moore-Penrose pseudoinverse \mathbf{C}^+ : $\mathbf{B} = \mathbf{I} - \mathbf{C}\mathbf{C}^+$.

Table 1: Monte Carlo Simulations (part one)

Target Parameter	Experiment 1: $\mathbf{H} = \mathbf{I} + \mathbf{G}$, $\mathbf{C} = \mathbf{I} + \mathbf{a}$ different small world, $\mathbf{C}_e = \mathbf{I} + \mathbf{G}$					Experiment 2: $\mathbf{H} = \mathbf{I} + \mathbf{G}$, $\mathbf{C} = \mathbf{I} + \mathbf{G}$, $\mathbf{C}_e = \mathbf{I} + \mathbf{G} + \mathbf{G}^2$										
	GMM1	GMM2	GMM3	GMM4	OLS	2SLSa	2SLSb	2SLSa	OLS	GMM4	GMM3	GMM2	GMM1	2SLSa	2SLSb	2SLSc
$\alpha = 0.25$	0.257 (0.059)	0.251 (0.035)	0.252 (0.037)	0.259 (0.032)	0.234 (0.011)	0.249 (0.031)	0.256 (0.163)	0.236 (0.358)	0.231 (0.010)	0.206 (0.056)	0.249 (0.037)	0.248 (0.036)	0.296 (0.674)	0.190 (0.051)	0.251 (0.032)	0.300 (0.649)
$\beta = 0.40$	0.396 (0.050)	0.400 (0.030)	0.399 (0.031)	0.393 (0.026)	0.414 (0.008)	0.400 (0.027)	0.397 (0.069)	0.411 (0.315)	0.417 (0.008)	0.437 (0.047)	0.401 (0.031)	0.402 (0.031)	0.361 (0.579)	0.453 (0.045)	0.399 (0.023)	0.358 (0.485)
$\gamma = 0.50$	0.495 (0.056)	0.485 (0.047)	0.486 (0.046)	0.536 (0.035)	0.580 (0.012)	0.587 (0.017)	0.534 (1.257)	0.479 (0.500)	0.586 (0.011)	0.493 (0.085)	0.492 (0.029)	0.490 (0.031)	0.527 (0.372)	0.564 (0.028)	0.576 (0.368)	0.778 (8.095)
$\chi = 1.00$	0.998 (0.012)	0.999 (0.011)	0.999 (0.011)	1.000 (0.008)	0.997 (0.007)	1.000 (0.009)	1.000 (0.018)	0.999 (0.084)	0.996 (0.007)	0.999 (0.013)	0.999 (0.009)	0.999 (0.009)	1.006 (0.121)	0.989 (0.014)	1.000 (0.014)	0.998 (0.132)
$\xi = 0.20$	0.199 (0.090)	0.215 (0.082)	0.215 (0.079)	0.107 (0.070)	-	-	-	-	0.202 (0.058)	0.072 (0.036)	0.202 (0.036)	0.204 (0.038)	0.184 (0.058)	-	-	-
$\psi = 0.25$	0.267 (0.091)	0.267 (0.076)	0.267 (0.076)	0.226 (0.071)	-	-	-	-	0.266 (0.083)	0.207 (0.137)	0.262 (0.068)	0.266 (0.072)	0.259 (0.083)	-	-	-
$\sigma = 0.05$	0.052 (0.006)	0.052 (0.004)	0.052 (0.004)	0.049 (0.003)	-	-	-	-	0.051 (0.003)	0.051 (0.006)	0.051 (0.003)	0.051 (0.003)	0.058 (0.150)	-	-	-

Target Parameter	Experiment 3: $\mathbf{H} = \mathbf{I} + \mathbf{G}$, $\mathbf{C} = \mathbf{I} + \mathbf{G} + \mathbf{G}^2$, $\mathbf{C}_e = \mathbf{I} + \mathbf{G}$					Experiment 4: $\mathbf{H} = \mathbf{I} + \mathbf{G}$, \mathbf{C} : groups of size 10, \mathbf{C}_e : \mathbf{C} 's groups split in half										
	GMM1	GMM2	GMM3	GMM4	OLS	2SLSa	2SLSb	2SLSa	OLS	GMM4	GMM3	GMM2	GMM1	2SLSa	2SLSb	2SLSc
$\alpha = 0.25$	0.145 (0.987)	0.239 (0.048)	0.242 (0.053)	0.240 (0.208)	0.260 (0.008)	0.354 (0.021)	0.250 (0.021)	0.312 (0.420)	0.228 (0.012)	0.241 (0.038)	0.246 (0.038)	0.246 (0.037)	0.264 (0.132)	0.226 (0.042)	0.249 (0.032)	0.237 (0.047)
$\beta = 0.40$	0.488 (0.810)	0.410 (0.040)	0.407 (0.045)	0.411 (0.218)	0.391 (0.006)	0.309 (0.014)	0.400 (0.367)	0.346 (0.367)	0.420 (0.009)	0.408 (0.032)	0.404 (0.032)	0.403 (0.031)	0.391 (0.099)	0.421 (0.037)	0.401 (0.023)	0.411 (0.041)
$\gamma = 0.50$	0.364 (1.504)	0.496 (0.043)	0.498 (0.048)	0.506 (0.501)	0.631 (0.008)	0.685 (0.013)	0.625 (0.226)	0.683 (0.289)	0.522 (0.015)	0.487 (0.063)	0.501 (0.053)	0.503 (0.048)	0.493 (0.086)	0.520 (0.023)	0.525 (0.488)	0.495 (0.023)
$\chi = 1.00$	0.992 (0.046)	0.997 (0.011)	0.997 (0.012)	0.993 (0.128)	1.002 (0.006)	1.020 (0.009)	1.000 (0.009)	1.012 (0.090)	0.996 (0.008)	0.998 (0.010)	0.999 (0.009)	0.999 (0.008)	0.996 (0.021)	0.995 (0.011)	1.000 (0.015)	0.998 (0.012)
$\xi = 0.20$	0.198 (0.040)	0.189 (0.032)	0.187 (0.037)	0.235 (0.117)	-	-	-	-	0.216 (0.211)	0.230 (0.223)	0.230 (0.223)	0.216 (0.211)	0.240 (0.361)	-	-	-
$\psi = 0.25$	0.217 (0.102)	0.222 (0.094)	0.232 (0.098)	0.232 (0.141)	-	-	-	-	0.249 (0.061)	0.251 (0.061)	0.253 (0.061)	0.279 (0.101)	0.279 (0.101)	-	-	-
$\sigma = 0.05$	0.078 (0.325)	0.051 (0.005)	0.051 (0.005)	0.053 (0.113)	-	-	-	-	0.050 (0.003)	0.051 (0.003)	0.051 (0.003)	0.050 (0.002)	0.055 (0.022)	-	-	-

Target Parameter	Experiment 5: $\mathbf{H} = \mathbf{C}$: groups of size 10, \mathbf{C}_e : \mathbf{C} 's (\mathbf{H} 's) groups split in half					Experiment 6: \mathbf{H} : groups of size 10, $\mathbf{C} = \mathbf{I} + \mathbf{G}$, $\mathbf{C}_e = \mathbf{I} + \mathbf{G} + \mathbf{G}^2$										
	GMM1	GMM2	GMM3	GMM4	OLS	2SLSa	2SLSb	2SLSa	OLS	GMM4	GMM3	GMM2	GMM1	2SLSa	2SLSb	2SLSc
$\alpha = 0.25$	0.248 (0.144)	0.245 (0.045)	0.242 (0.046)	0.236 (0.048)	0.229 (0.011)	0.225 (0.059)	0.250 (0.039)	0.224 (0.068)	0.240 (0.009)	0.216 (0.064)	0.244 (0.043)	0.246 (0.041)	0.255 (0.080)	0.196 (0.134)	0.248 (0.026)	0.269 (0.443)
$\beta = 0.40$	0.402 (0.113)	0.405 (0.037)	0.407 (0.038)	0.412 (0.041)	0.419 (0.009)	0.422 (0.051)	0.401 (0.023)	0.423 (0.060)	0.409 (0.007)	0.428 (0.053)	0.405 (0.036)	0.404 (0.036)	0.397 (0.066)	0.447 (0.118)	0.401 (0.018)	0.384 (0.377)
$\gamma = 0.50$	0.477 (0.161)	0.509 (0.134)	0.508 (0.120)	0.484 (0.141)	0.578 (0.020)	0.575 (0.035)	0.586 (0.951)	0.574 (0.049)	0.685 (0.019)	0.497 (0.147)	0.487 (0.063)	0.491 (0.053)	0.505 (0.068)	0.659 (0.080)	0.650 (0.523)	0.476 (1.318)
$\chi = 1.00$	0.997 (0.017)	0.999 (0.008)	0.999 (0.009)	0.998 (0.010)	0.996 (0.007)	0.995 (0.014)	1.000 (0.015)	0.995 (0.016)	0.998 (0.007)	0.998 (0.018)	0.999 (0.009)	0.999 (0.011)	0.999 (0.011)	0.990 (0.029)	1.000 (0.012)	1.002 (0.080)
$\xi = 0.20$	0.222 (0.210)	0.185 (0.151)	0.189 (0.162)	0.410 (0.341)	-	-	-	-	0.195 (0.033)	0.078 (0.040)	0.195 (0.025)	0.195 (0.022)	0.187 (0.033)	-	-	-
$\psi = 0.25$	0.268 (0.087)	0.249 (0.061)	0.249 (0.065)	0.249 (0.069)	-	-	-	-	0.201 (0.192)	0.201 (0.192)	0.265 (0.114)	0.259 (0.097)	0.255 (0.091)	-	-	-
$\sigma = 0.05$	0.054 (0.027)	0.051 (0.007)	0.051 (0.004)	0.052 (0.007)	-	-	-	-	0.051 (0.004)	0.051 (0.012)	0.051 (0.004)	0.051 (0.004)	0.052 (0.005)	-	-	-

Notes. For every experiment-estimator combination, each column in this table reports the median and the standard deviation (in parentheses) of the relevant point estimates across 1,000 repetitions, all with $N = 500$. Each experiment is characterized by a combination of $(\mathbf{H}, \mathbf{C}, \mathbf{C}_e)$, as summarized by the corresponding headers. The estimators are as follows: GMM1: $Q = 3$, $P = 2$, $\mathbf{C}_e = \mathbf{C}$; GMM2: $Q = 4$, $P = 3$, $\mathbf{C}_e \neq \mathbf{C}$; GMM3: $Q = 5$, $P = 4$, $\mathbf{C}_e = \mathbf{C}$; GMM4: $Q = 5$, $P = 4$, $\mathbf{C}_e \neq \mathbf{C}$; OLS is self-explanatory; 2SLSa: $\mathbf{z} = \mathbf{x}$; 2SLSb: $\mathbf{z} = \mathbf{G}\mathbf{w}$; 2SLSc: $\mathbf{z} = \mathbf{B}\mathbf{x}$. See the text for additional details.

Table 2: Monte Carlo Simulations (part two)

Target Parameter	Experiment 7: as in experiment 1, Table 1, but with higher β					Experiment 8: as in experiment 1, Table 1, but with lower γ										
	GMM1	GMM2	GMM3	GMM4	OLS	2SLSa	2SLSb	2SLSc	GMM1	GMM2	GMM3	GMM4	OLS	2SLSa	2SLSb	2SLSc
$\alpha = 0.25$	0.256 (0.060)	0.252 (0.035)	0.253 (0.036)	0.259 (0.034)	0.232 (0.012)	0.254 (0.032)	0.250 (0.032)	0.244 (0.402)	0.262 (0.108)	0.252 (0.040)	0.253 (0.044)	0.259 (0.040)	0.233 (0.011)	0.253 (0.102)	0.249 (0.073)	0.237 (0.438)
$\beta = 0.50$	0.496 (0.043)	0.499 (0.025)	0.498 (0.025)	0.494 (0.024)	0.513 (0.008)	0.497 (0.024)	0.500 (0.019)	0.504 (0.297)	0.392 (0.087)	0.399 (0.033)	0.398 (0.037)	0.393 (0.033)	0.415 (0.009)	0.397 (0.089)	0.400 (0.037)	0.411 (0.379)
$\gamma = 0.50$	0.498 (0.059)	0.485 (0.046)	0.487 (0.045)	0.535 (0.036)	0.579 (0.012)	0.589 (0.017)	0.548 (0.570)	0.472 (0.651)	0.198 (0.053)	0.187 (0.044)	0.189 (0.043)	0.233 (0.036)	0.283 (0.012)	0.289 (0.023)	0.285 (0.656)	0.184 (0.617)
$\chi = 1.00$	0.999 (0.014)	0.999 (0.010)	0.999 (0.010)	1.000 (0.009)	1.000 (0.007)	1.001 (0.010)	1.000 (0.012)	1.000 (0.099)	0.996 (0.014)	0.999 (0.008)	0.999 (0.011)	1.000 (0.008)	0.997 (0.007)	1.000 (0.021)	1.000 (0.013)	0.999 (0.089)
$\xi = 0.20$	0.196 (0.090)	0.217 (0.081)	0.214 (0.082)	0.107 (0.071)	-	-	-	-	0.184 (0.093)	0.213 (0.079)	0.211 (0.078)	0.108 (0.071)	-	-	-	-
$\psi = 0.25$	0.266 (0.086)	0.270 (0.074)	0.269 (0.073)	0.225 (0.071)	-	-	-	-	0.267 (0.100)	0.267 (0.070)	0.266 (0.073)	0.225 (0.071)	-	-	-	-
$\sigma = 0.05$	0.052 (0.006)	0.052 (0.004)	0.052 (0.004)	0.049 (0.003)	-	-	-	-	0.054 (0.011)	0.052 (0.004)	0.052 (0.004)	0.049 (0.003)	-	-	-	-

Target Parameter	Experiment 9: as in experiment 1, Table 1, but with $\xi = 0$ (no endogeneity)					Experiment 10: as in experiment 1, Table 1, but with $\psi = 0$ (i.i.d. errors)										
	GMM1	GMM2	GMM3	GMM4	OLS	2SLSa	2SLSb	2SLSc	GMM1	GMM2	GMM3	GMM4	OLS	2SLSa	2SLSb	2SLSc
$\alpha = 0.25$	0.245 (0.069)	0.246 (0.037)	0.247 (0.040)	0.252 (0.045)	0.228 (0.012)	0.251 (0.048)	0.249 (0.030)	0.246 (0.398)	0.266 (0.051)	0.255 (0.032)	0.255 (0.032)	0.269 (0.028)	0.247 (0.010)	0.273 (0.032)	0.255 (0.134)	0.253 (0.359)
$\beta = 0.40$	0.405 (0.059)	0.403 (0.030)	0.402 (0.033)	0.399 (0.037)	0.419 (0.009)	0.399 (0.042)	0.401 (0.020)	0.347 (0.347)	0.387 (0.043)	0.396 (0.028)	0.395 (0.027)	0.383 (0.024)	0.403 (0.008)	0.380 (0.028)	0.397 (0.058)	0.314 (0.314)
$\gamma = 0.50$	0.499 (0.079)	0.498 (0.062)	0.497 (0.060)	0.498 (0.030)	0.490 (0.015)	0.500 (0.023)	0.486 (0.514)	0.528 (0.528)	0.490 (0.061)	0.475 (0.062)	0.478 (0.056)	0.524 (0.042)	0.575 (0.011)	0.587 (0.018)	0.533 (1.056)	0.481 (0.677)
$\chi = 1.00$	0.998 (0.014)	0.999 (0.007)	0.999 (0.008)	0.999 (0.008)	0.996 (0.007)	1.000 (0.012)	1.000 (0.014)	1.000 (0.077)	1.001 (0.010)	1.000 (0.010)	1.000 (0.009)	1.003 (0.008)	0.999 (0.007)	1.004 (0.010)	1.000 (0.017)	1.002 (0.085)
$\xi = 0.00$	0.041 (0.091)	0.051 (0.068)	0.049 (0.067)	0.016 (0.033)	-	-	-	-	0.217 (0.101)	0.236 (0.103)	0.234 (0.097)	0.135 (0.080)	-	-	-	-
$\psi = 0.25$	0.262 (0.068)	0.255 (0.049)	0.253 (0.054)	0.250 (0.056)	-	-	-	-	0.082 (0.091)	0.083 (0.078)	0.079 (0.076)	0.055 (0.073)	-	-	-	-
$\sigma = 0.05$	0.052 (0.006)	0.051 (0.002)	0.051 (0.002)	0.051 (0.002)	-	-	-	-	0.053 (0.006)	0.053 (0.006)	0.053 (0.005)	0.050 (0.004)	-	-	-	-

Target Parameter	Experiment 11: as in experiment 1, Table 1, but with higher σ					Experiment 12: $\mathbf{H} = 2(\mathbf{I} + \mathbf{G})$, $\mathbf{C} = \mathbf{I}$ + a different small world, $\mathbf{C}_e = \mathbf{I} + \mathbf{G}$										
	GMM1	GMM2	GMM3	GMM4	OLS	2SLSa	2SLSb	2SLSc	GMM1	GMM2	GMM3	GMM4	OLS	2SLSa	2SLSb	2SLSc
$\alpha = 0.25$	0.250 (0.107)	0.248 (0.066)	0.249 (0.069)	0.256 (0.060)	0.195 (0.021)	0.250 (0.052)	0.250 (0.074)	0.201 (1.066)	0.225 (0.093)	0.234 (0.056)	0.238 (0.056)	0.237 (0.055)	0.241 (0.011)	0.322 (0.030)	0.251 (0.425)	0.246 (0.425)
$\beta = 0.40$	0.402 (0.091)	0.403 (0.056)	0.402 (0.059)	0.395 (0.050)	0.449 (0.016)	0.400 (0.045)	0.399 (0.049)	0.852 (0.852)	0.423 (0.079)	0.415 (0.049)	0.412 (0.048)	0.413 (0.050)	0.407 (0.008)	0.337 (0.026)	0.399 (0.022)	0.403 (0.378)
$\gamma = 0.50$	0.481 (0.114)	0.471 (0.095)	0.471 (0.092)	0.578 (0.060)	0.644 (0.024)	0.674 (0.033)	0.657 (0.842)	7.352 (7.352)	0.470 (0.096)	0.479 (0.062)	0.478 (0.062)	0.538 (0.067)	0.574 (0.012)	0.601 (0.013)	0.564 (0.400)	0.494 (0.300)
$\chi = 1.00$	0.995 (0.022)	0.997 (0.018)	0.997 (0.019)	0.998 (0.015)	0.989 (0.014)	1.000 (0.018)	1.001 (0.024)	1.011 (0.542)	0.992 (0.018)	0.995 (0.016)	0.995 (0.015)	0.993 (0.020)	0.999 (0.007)	1.013 (0.010)	0.999 (0.015)	1.000 (0.076)
$\xi = 0.20$	0.202 (0.089)	0.214 (0.083)	0.214 (0.084)	0.096 (0.063)	-	-	-	-	0.209 (0.095)	0.212 (0.088)	0.216 (0.091)	0.085 (0.081)	-	-	-	-
$\psi = 0.25$	0.265 (0.106)	0.264 (0.086)	0.264 (0.088)	0.216 (0.079)	-	-	-	-	0.217 (0.124)	0.218 (0.106)	0.223 (0.106)	0.252 (0.123)	-	-	-	-
$\sigma = 0.10$	0.104 (0.011)	0.103 (0.008)	0.103 (0.008)	0.097 (0.005)	-	-	-	-	0.056 (0.016)	0.053 (0.008)	0.053 (0.008)	0.050 (0.007)	-	-	-	-

Notes. For every experiment-estimator combination, each column in this table reports the median and the standard deviation (in parentheses) of the relevant point estimates across 1,000 repetitions, all with $N = 500$. Each experiment is characterized by a combination of $(\mathbf{H}, \mathbf{C}, \mathbf{C}_e)$, as summarized by the corresponding headers. The estimators are as follows: GMM1: $Q = 3$, $P = 2$, $\mathbf{C}_e = \mathbf{C}$; GMM2: $Q = 4$, $P = 3$, $\mathbf{C}_e \neq \mathbf{C}$; GMM3: $Q = 5$, $P = 4$, $\mathbf{C}_e = \mathbf{C}$; GMM4: $Q = 5$, $P = 4$, $\mathbf{C}_e \neq \mathbf{C}$; OLS is self-explanatory; 2SLSa: $\mathbf{z} = \mathbf{x}$; 2SLSb: $\mathbf{z} = \mathbf{G}\mathbf{w}$; 2SLSc: $\mathbf{z} = \mathbf{B}\mathbf{x}$. See the text for additional details.

of the d.g.p., which are reported in the table. The number of instruments being used does not appear too consequential. It is interesting to examine the estimates based on a misspecified matrix \mathbf{C}_e (“GMM4”), which in the baseline case we set as equal to $\mathbf{C}_e = \mathbf{I} + \mathbf{G} = \mathbf{H}$: this introduces a bias in our estimates, but not a particularly pronounced one for the main parameters of interests β and γ (unlike the endogeneity parameter ξ , for which the bias is more pronounced). More conventional estimators all deliver biased estimates for either β or γ , or both, that are comparable to those obtained from the misspecified GMM.²⁷ In the case of the third 2SLS estimator (with $\mathbf{z} = \mathbf{B}\mathbf{x}$), the bias seems coupled with an exceedingly large variability of the estimates: as hinted in Section 2.3, the transformation implied by \mathbf{B} is bound to remove much of the independent variable’s variation, which in turn is likely to exacerbate the bias of the GMM estimator in small samples.

Experiments 2 through 6 from Table 1 are analogous to the baseline case. All of them are characterized by different combinations of matrices \mathbf{H} and \mathbf{C} , which may include functions of \mathbf{G}^2 or matrices representing “fully segregated” group structures as per the discussion in Section 2.3 and the representation of Graph 1 (in this case, all connections in a group are equally weighted). When we perform GMM estimates with a misspecified matrix \mathbf{C}_e , the latter is typically chosen so as to be fairly, but not overly similar to the true \mathbf{C} ; for example, if \mathbf{C} has a group structure, \mathbf{C}_e is obtained by splitting the original groups in half. The results are qualitatively very similar to the baseline case. The main difference is that when \mathbf{C} has a group structure, the third 2SLS estimator with $\mathbf{z} = \mathbf{B}\mathbf{x}$ performs much better, as the implied transformation is akin to the conventional within-transformation for panel data.²⁸ Table 2 shows the results from additional simulations, all built around the baseline: Experiments 7 and 8 attempt alternative values for β and γ , respectively; Experiments 9 and 10 shut down the entire endogeneity channel ($\xi = 0$) or the error terms’ SMA(1) process ($\psi = 0$), respectively; lastly, Experiments 11 and 12 increase the variance of two key elements of the d.g.p.: the error term and the independent component of \mathbf{x} , respectively. The interpretation of all those results is the same as in the baseline case.

²⁷Notably, the 2SLS estimates for β when $\mathbf{z} = \mathbf{w}$ are not biased, but the bias associated with γ appears larger. In general, in all these experiments β seldom displays a larger bias; this is likely due to the inclusion of \mathbf{w} in our d.g.p. and of moments based on it across all our estimators. In different Monte Carlo simulations that omit \mathbf{w} , which we experimented with (though we do not report them here for the sake of brevity), the bias associated with β is typically larger.

²⁸Also note that when $\mathbf{C} = \mathbf{I} + \mathbf{G} + \mathbf{G}^2$ and $\mathbf{C}_e = \mathbf{I} + \mathbf{G}$ (Experiment 3) the misspecification bias does not appear sizable, not even in the case of parameter ξ .

In summary, the GMM estimator that we propose appears to perform well, and while it has demanding requirements (mainly the knowledge of the true characteristic matrix \mathbf{C}), departures from the ideal scenario – such as the misspecified estimates that we examine – do not seem to yield worse estimates than those obtained from more conventional estimators; in some cases the estimates for the main parameter of interest β are very similar to those obtained via exogenous instruments, which may also be hard to obtain. With the exception of the quite regular case where \mathbf{C} has a fully segregated group structure, transformations of the data that purge the endogenous component of \mathbf{x} do not seem to be a viable alternative. All these considerations make a case in favor of our proposed estimator in the applied econometric practice.²⁹

6 Empirical Application

To illustrate how our proposed method can help account for correlated effects in an actual empirical study about social effects, we leverage the setting and data from the influential study by De Giorgi et al. (2010). This contribution estimates peer effects between students who started their undergraduate studies at Bocconi University in Italy in 1998.³⁰ A key feature of this paper is that peer groups are shaped according to a non-overlapping, networked structure of social interactions \mathcal{G} that is determined exogenously. Specifically, students from different undergraduate programs at Bocconi University used to take common foundational courses over their first year and a half of curriculum; there were multiple, parallel version of each common course, and students were randomly allocated into them. Students are defined as “peers” over the later part of their curriculum if they attended together a given number of common courses out of seven (this number is set as four in the authors’ preferred specification).³¹ We refer to the original paper for a full-fledged description of the setting and data.

We estimate the simple SAR model (11) on the data provided by De Giorgi et al.

²⁹We also obtained separate results that alter the parameters of the small-world algorithm with respect to the baseline, by either setting $B = 4$ or $\varphi = 0.9$ (see footnote 25). While we do not report these results for brevity, they display the patterns that are qualitatively identical to the baseline’s.

³⁰Bocconi University offers undergraduate and graduate programs in Economics, Finance, Business Administration, and – to a lesser extent – in other Social Sciences.

³¹There were in total nine common courses, of which two were in legal subjects and were excluded by the authors. The two law classes had unusually low attendance rates and thus a lower number of parallel sessions; consequently, including them in the count would inappropriately inflate the number of peers that each student has.

(2010) using the same network matrix \mathbf{G} from their favorite specification of the peer structure (which is row-normalized). However, our model differs from theirs in several respects. First, to keep the analysis simple we abstract from contextual effects, and therefore we abstain from specifying a complete linear-in-means model. Furthermore, we choose different outcome and explanatory variables. In the original paper, y_i is a dummy variable that denotes major choice (Economics vs. Business), whereas in our case it denotes the Bocconi GPA that excludes the initial common courses. We choose not to use a dummy variable as our main dependent variable so as to avoid contradictions between our estimated model and our maintained assumptions about the error term. Furthermore, in order to keep our illustrative application simple we use only one right-hand side variable x_i : the grade that a student has received in the high school final exams.³² In the original paper, the authors use other right-hand side variables, including the score achieved in the Bocconi admission test and a dummy variable encoding whether a student had indicated Economics as the preferred major at the time of applying at Bocconi. We find that these variables have little predictive power over our chosen y_i outcome.

While we believe that our chosen x_i variable is representative of a student's prior educational achievements or background, it is certainly not devoid of problems. First, it is arguably endogenous: it is likely to depend upon the unobserved individual ability or motivation (which are encoded in the error term ε_i) alongside our outcome variable y_i . This would be not so much of a problem for the sake of estimating social effects if such unobserved components were independent across students. It turns out that there are reasons to suspect the existence of cross-correlation between the error terms of different students, and that it occurs along geographical lines. Note that Bocconi is a prestigious university within Italy, certainly not a cheap one to attend by national standards;³³ while located in Milan in Lombardy, about half of its student body hails from outside that region. For such students the cost of attending Bocconi is higher in comparative terms; thus, they are likely to be representative of a more (self-)selected subset of the population of potential students. This may be especially true for those

³²In Italy, completion of high school is conditional upon passing a centrally-managed nationwide exam (which differs by type of high school); marks in this exam are awarded on a scale of 100 with 60 being the passing mark.

³³In making these claims, we would like to remark that neither of us has graduated from or been employed at Bocconi University. One of us briefly attended one of its undergraduate programs before dropping out, and he was later rejected by its flagship Master's program in Economics.

students coming from the regions of central and southern Italy with a markedly lower income per capita (about one fourth of our sample).

In light of these observations, we model endogeneity similarly to (7):

$$x_i = \tilde{x}_i + \xi \varepsilon_i, \quad (33)$$

while introducing cross-correlation in the error term as follows:

$$\varepsilon_i = \sum_{i=1}^N c_{ij} v_i, \quad (34)$$

with $c_{ii} = 1$ for $i = 1, \dots, N$. Together, (33) and (34) are combined as a specification of the high-school mark x_i that is consistent with Assumption 6 and hence, with our econometric model (the weights c_{ij} are collected in matrix \mathbf{C}). Note that (34) treats the error term as a spatial MA(1) process with a fixed parameter;³⁴ although we could estimate this parameter through our procedure, we prefer to assume the entire spatial structure of the errors so as to keep the analysis as simple as possible.

Specifically, we experiment with two alternative structures, defined in terms of the spatial correlation of the error term that is expressed by matrix $\mathbf{C}\mathbf{C}^T$. In both cases, the spatial correlation is normalized by the sample size N so as to comply with our statistical assumptions illustrated in Section 4. The two structures are as follows.

1. The first, denoted as $\mathcal{C}_{(1)}$, is such that $\text{Cov}(\varepsilon_i, \varepsilon_j) = \sigma/N$ if students i and j originate from the same “historical area” of Italy, and $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$ otherwise. We define as “historical areas” those polities with a shared history, subculture and economic structure that existed on the Italian territory before the process of Italian unification was set in motion in 1859.³⁵
2. The second, denoted as $\mathcal{C}_{(2)}$, is such that $\text{Cov}(\varepsilon_i, \varepsilon_j) = \exp(-D \cdot d_{ij}) \cdot \sigma/N$ for every pair (i, j) , where d_{ij} is the distance between the geographical centroids of two students’ *provinces* of origin³⁶ while $D > 0$ is a constant. The resulting

³⁴This model implied by (33) and (34) can be written as $\mathbf{x} = \tilde{\mathbf{x}} + \xi \mathbf{C}\mathbf{v}$, and it is equivalent to a model expressed as $\mathbf{x} = \tilde{\mathbf{x}} + \xi \mathbf{C}'\boldsymbol{\varepsilon}$ where $\mathbf{C}' = \mathbf{I}$ and $\boldsymbol{\varepsilon} = (\mathbf{I} + \psi \mathbf{E})\mathbf{v} = \mathbf{C}\mathbf{v}$.

³⁵We consider the two small historical duchies of Parma-Piacenza and Modena-Reggio as one such polity. We also experimented with alternative subdivisions, including one based on the classification, within the Romance linguistic family, of the regional language that is prevalent in each province (in Italy, traditional regional languages are still widely spoken); in all cases we obtained similar results.

³⁶Provinces are traditional administrative units of Italy, equivalent to the NUTS-3 level. In 1998

spatial correlation of ε_i is consistent with distance decay as in (8).

In both cases, we calculate the characteristics matrix \mathbf{C} by factorizing the target spatial correlation structure via eigendecomposition; hence, weights c_{ij} can be interpreted as the relative importance of student j on the variance of student i 's characteristics. We believe both structures to be potentially capable of capturing the correlated effects that can potentially jeopardize the identification of social effects.³⁷

Before presenting our estimation results, it is worthwhile to discuss raw patterns of spatial correlation in the data, for both the explanatory variable x_i and the outcome y_i . Specifically, we calculate the Moran's I statistic and its associated standard error (Kelejian and Prucha, 2001), for the two variables in question and for alternative characteristics structures \mathcal{C} , using a modified version of $\mathbf{C}\mathbf{C}^T$ (such that all diagonal elements are set at zero) as the weight matrix. For structure $\mathcal{C}_{(2)}$, we experiment with two fairly spaced values of D , that is $D \in \{2, 6\}$. The results are reported in Table 3. Notably, the spatial correlation in final high school grades x_i is calculated as about 0.10 for structure $\mathcal{C}_{(1)}$, and within the range 0.04-0.07 (depending on the value of D) for structure $\mathcal{C}_{(2)}$. This suggests that that our structures of choice explain at least in part the co-dependence of x_i across students. Conversely, the Moran's I statistics for the later Bocconi GPA y_i is calculated at typically smaller values (between 0.01 and 0.02) across all structures \mathcal{C} . According to our econometric framework, these patterns could result from correlated effects that propagate across the observations of both x_i and y_i if the parameter that drives endogeneity, that is ξ , is positive. Note that the standard errors associated with all six estimated Moran's I statistics are very small, which is customary in analyses about spatial correlation.

The analysis of the Moran's I statistics does not allow to draw conclusion about the extent of endogeneity or the existence of some residual spatial correlation in the outcome variable y_i that can be interpreted as peer effects. Instead, estimating our model would allow to tackle both questions at once. The empirical estimates based on our GMM framework presented in Section 4, alongside other estimates that are based on more conventional methods, are displayed in Table 4. We focus our discussion on the key parameters β and ξ . The first column displays the OLS estimates of the SAR model (11), which delivers an estimate of $\beta \simeq 0.12$ that is not statistically significant.

there were 101 provinces. We set $d_{ij} = 0$ if $i = j$ or the two students hail from the same province.

³⁷We treat non-Italian students (who represent less than 2 per cent of the sample) as follows: in $\mathcal{C}_{(1)}$ they are identified as a separate block in the spatial correlation matrix $\mathbf{C}\mathbf{C}^T$, whereas in $\mathcal{C}_{(2)}$ they are treated as hailing from an additional, very distant "province."

Table 3: Empirical Application: Moran’s I Statistics

	Variable: x_i			Variable: y_i		
	$\mathcal{C}_{(1)}$	$\mathcal{C}_{(2)}$	$\mathcal{C}_{(2)}$	$\mathcal{C}_{(1)}$	$\mathcal{C}_{(2)}$	$\mathcal{C}_{(2)}$
Moran’s I	0.099 (0.001)	0.043 (0.001)	0.067 (0.001)	0.016 (0.001)	0.010 (0.001)	0.019 (0.001)
D	–	2	6	–	2	6

Notes. This table reports the Moran’s I statistic, as well as its associated standard error (in parentheses), for the indicated variables and structures \mathcal{C} , using a modified version of $\mathbf{C}\mathbf{C}^T$ (with all diagonal elements set at zero) as the Moran weight matrix. For structures of the $\mathcal{C}_{(2)}$ type, the table reports results for two different values of D , as indicated in the bottom row (if applicable). Across all calculations the sample size equals $N = 1, 141$.

Since OLS is by construction inconsistent in this setting, we attempt 2SLS estimation based on instrument sets *à la* Bramoullé et al. (2009) of varying size. Both attempts register similar values of $\beta \simeq 0.34$: in addition to being sizable, these estimates are also statistically significant. If this result were confirmed, it would provide the basis for policy interventions motivated on a social multiplier of about $1/(1 - \beta) \simeq 1.51$.

The last three columns of Table 4 – those grouped under the label ‘PFZ’ – report the estimates based on our framework for the same characteristics structures used to calculate the Moran’s I statistics. The pattern is similar across the three cases. Most notably, for every characteristic structure β is estimated negative and not statistically significant. In addition, ξ is estimated statistically significant, with point estimates ranging from about 6.5 to about 10.6. These results suggest that our method, however crudely implemented and despite relying on precise assumptions about the structure of spatial correlation, has the potential to turn around empirical estimates of network effects that, while registered by conventional methods, are suspected to be the consequence of correlated effects. This observation could prove useful to researchers intent on verifying whether their estimates of network effects are indeed robust. We also find it interesting to observe that while our estimates of ξ register endogeneity, the estimates for γ do not differ much across the six specifications reported in Table 4.³⁸ This suggests that estimates of network effects are more sensitive, relative to standard regressors, to the kind of endogeneity that we examine in this paper.

³⁸The typical estimate of γ is about 12. The interpretation of this figure is that every 10 additional points that students receive in their final high school exam translate, on average, into a later Bocconi GPA which is higher by 1.2 points on a scale of 30.

Table 4: Empirical Application: Model Estimates

	OLS/IV			PFZ		
	OLS	IV	2SLS	$\mathcal{C}_{(1)}$	$\mathcal{C}_{(2)}$	$\mathcal{C}_{(2)}$
α	14.151*** (2.428)	8.399*** (3.653)	8.164*** (3.680)	16.951*** (0.858)	17.308*** (0.830)	17.052*** (0.851)
β	0.117 (0.089)	0.334** (0.136)	0.343** (0.137)	-0.026 (0.037)	-0.003 (0.031)	-0.019 (0.036)
γ	11.287*** (0.521)	11.258*** (0.523)	11.257*** (0.523)	12.571*** (0.753)	11.374*** (0.397)	12.203*** (0.730)
ξ	–	–	–	9.150** (3.892)	6.541*** (2.435)	10.631** (4.794)
σ	–	–	–	1.596*** (0.020)	1.591*** (0.018)	1.593*** (0.019)
D	–	–	–	–	2	6
N	1,141	1,141	1,141	1,141	1,141	1,141

Notes. Each column in this table reports estimates of model (11) performed on the data by De Giorgi et al. (2010), using the variables y_i and x_i described in the text. The first three columns report either OLS, IV or 2SLS results, where: ‘OLS’ is self-explanatory; the ‘IV’ estimates are obtained by instrumenting $\mathbf{G}\mathbf{y}$ with $\mathbf{G}\mathbf{x}$, while the ‘2SLS’ estimates use both $\mathbf{G}\mathbf{x}$ and $\mathbf{G}^2\mathbf{x}$ as instruments. The remaining three columns display the results obtained by our estimator (‘PFZ’) assuming different structures \mathcal{C} . For structures of the $\mathcal{C}_{(2)}$ type, the table reports the results for two different values of D , which are indicated in the bottom row if applicable. The sample size is denoted by N . Asterisk series: *, ** and *** denote statistical significance of point estimates respectively at the 10 per cent, 5 per cent and 1 per cent level.

7 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. The requirements for identification are quite general: it suffices that the network of social interactions is exogenous, not fully-overlapping in only a slightly stronger sense relative to the identification conditions by Bramoullé et al. (2009), and that the spatial structure of endogeneity (the dependence of individual covariates on peers’ unobservables) is known by the econometrician up to a multiplicative constant. This approach can be applied to studies about peer effects where the the right-hand side individual characteristics used for identification are possibly endogenous and affected by correlated effects. In our empirical application based

on the study by De Giorgi et al. (2010), we show that applying our approach under different specifications of the spatial structure of endogeneity leads to precise zero estimates of the social effects, where conventional methods would instead estimate positive and statistically significant effects.

We envision three areas for future work. First, we plan to extend our approach to more general specifications of the stochastic process driving endogeneity, such as non-linear ones or with conditionally heteroscedastic primitive errors. To this end, we plan to investigate the applicability of semi-parametric estimators or control function approaches that are less reliant upon linear functional forms. Second, we plan to relax the assumption about exogeneity of the network \mathcal{G} , by incorporating either control function methods *à la* Arduini et al. (2015) or Johnsson and Moon (2021), or a GMM approach for panel data in the spirit of Kuersteiner and Prucha (2020).³⁹ Third, and last, we believe it would be worthwhile to integrate the recent literature that exploits penalized estimators in order to recover an unknown network structure (Rose, 2017b; de Paula et al., 2019) within our framework. Specifically, we believe that with partial information about the network structure, this kind of approaches may help identify an unknown characteristics structure \mathcal{C} , or the SARMA structure of the error term, and thus mitigate the main requirement of our approach: that is, the *a priori* knowledge of the structure in question.

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.” EIEF Working Paper 15/09.
- Azoulay, Pierre, Joshua S. Graff Zivin, and Jialan Wang (2010) “Superstar Extinction,” *The Quarterly Journal of Economics*, Vol. 125, No. 2, pp. 549–589.

³⁹Note that in our setup, it is already possible to relax this assumption. Consider for example the simple linear specification of endogeneity discussed at length in subsection 2.3. If (14) were replaced with:

$$\mathbb{E}[\tilde{\mathbf{x}} | \mathbf{G}, \mathbf{C}, \varepsilon] = \mathbf{0}.$$

our derivation and identification result would hold regardless. What is key is that *either* the structural error term ε *or* the independent component of individual characteristics $\tilde{\mathbf{x}}$ is mean independent of the network structure (and the characteristics structure too). The plausibility of either hypothesis depends on the empirical application; in this paper we have maintained the former, rather than the latter, for ease of exposition. A further idea for future work is to examine which models of network formation – intended as restrictions on $\mathcal{F}(\cdot)$ – are consistent with these assumptions.

- 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.
- (2020) “Peer Effects in Networks: A Survey,” *Annual Review of Economics*, Vol. 12, pp. 603–629.
- 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.
- Christakis, Nicholas A. and James H. Fowler (2013) “Social contagion theory: examining dynamic social networks and human behavior,” *Statistics in medicine*, Vol. 32, No. 4, pp. 556–577.
- 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,” *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 (2019) “Identifying network ties from panel data: theory and an application to tax competition.” CeMMAP Working Paper 55/19.
- 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. 111, No. 2, pp. 507–48.
- Goldsmith-Pinkham, Paul and Guido W. Imbens (2013) “Social Networks and the Identification of Peer Effects,” *Journal of Business & Economic Statistics*, Vol. 31, No. 3, pp. 253–264.

- 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*, Vol. 85, No. 4, pp. 1033–1063.
- 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 (2021) “Estimation of Peer Effects in Endogenous Social Networks: Control Function Approach,” *The Review of Economics and Statistics*, Vol. 103, No. 2, pp. 328–345.
- Kapoor, Mudit, Harry H. Kelejian, and Ingmar R. Prucha (2007) “Panel data models with spatially correlated error components,” *Journal of Econometrics*, Vol. 140, No. 1, pp. 97–130.
- 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 (2020) “Dynamic Spatial Panel Models: Networks, Common Shocks, and Sequential Exogeneity,” *Econometrica*, Vol. 88, No. 5, pp. 2109–2146.
- Lee, Lung-fei (2007a) “GMM and 2SLS estimation of mixed regressive, spatial autoregressive models,” *Journal of Econometrics*, Vol. 137, No. 2, pp. 489–514.
- (2007b) “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.
- Lee, Lung-fei and Xiaodong Liu (2010) “Efficient GMM estimation of high order spatial autoregressive models with autoregressive disturbances,” *Econometric Theory*, Vol. 26, No. 1, pp. 187–230.
- 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.
- 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.
- 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.” Unpublished working paper.
- 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.
- Waldinger, Fabian (2012) “Peer Effects in Science: Evidence from the Dismissal of Scientists in Nazi Germany,” *The 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. (1996) *Estimation, Inference and Specification Analysis*: Cambridge University Press.
- Zacchia, Paolo (2020) “Knowledge Spillovers through Networks of Scientists,” *The Review of Economic Studies*, Vol. 84, No. 7, pp. 1989–2018.

Appendix – Mathematical Proofs

Proof of Theorem 1

Preliminaries. Before proceeding to the extended proof of the Theorem, it is useful to establish some notation. Let $\boldsymbol{\theta}_1 \equiv (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta})$, $\boldsymbol{\theta}_2 \equiv (\boldsymbol{\phi}, \boldsymbol{\psi})$,

$$\boldsymbol{\Psi}_{AM}(\boldsymbol{\phi}, \boldsymbol{\psi}) \equiv (\mathbf{I} - \phi_1 \mathbf{F}_1 - \cdots - \phi_A \mathbf{F}_A)^{-1} (\mathbf{I} + \psi_1 \mathbf{E}_1 + \cdots + \psi_M \mathbf{E}_M), \quad (\text{A.1})$$

and:

$$\boldsymbol{\varepsilon}(\boldsymbol{\theta}_1) \equiv \mathbf{y} - \boldsymbol{\alpha} - \boldsymbol{\beta} \mathbf{G} \mathbf{y} - \mathbf{X} \boldsymbol{\gamma} - \mathbf{G} \mathbf{X} \boldsymbol{\delta}.$$

as well as:

$$\mathbf{u}(\boldsymbol{\theta}) = \frac{1}{\sigma} \boldsymbol{\Psi}_{AM}^{-1}(\boldsymbol{\phi}, \boldsymbol{\psi}) \boldsymbol{\varepsilon}(\boldsymbol{\theta}_1),$$

an object that exists only so long as $\boldsymbol{\Psi}_{AM}(\boldsymbol{\phi}, \boldsymbol{\psi})$ is nonsingular. In what follows, we consider some vectors with the same length as our parameters of interest (or subsets thereof), as arguments of functions like the above, and others. We denote such vectors by a small tilde, e.g. $\tilde{\boldsymbol{\phi}}$ or $\tilde{\boldsymbol{\psi}}$. Importantly, we assume without loss of generality that $\tilde{\boldsymbol{\phi}}$ and $\tilde{\boldsymbol{\psi}}$ always lie within the unit circle, and that:

$$\boldsymbol{\Psi}_{AM}(\tilde{\boldsymbol{\phi}}, \tilde{\boldsymbol{\psi}}) \equiv (\mathbf{I} - \tilde{\phi}_1 \mathbf{F}_1 - \cdots - \tilde{\phi}_A \mathbf{F}_A)^{-1} (\mathbf{I} + \tilde{\psi}_1 \mathbf{E}_1 + \cdots + \tilde{\psi}_M \mathbf{E}_M), \quad (\text{A.2})$$

is always nonsingular. If this were not the case, by construction the values of $\tilde{\boldsymbol{\phi}}$ and $\tilde{\boldsymbol{\psi}}$ cannot be mistaken for the true values of $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$, and thus our identification proof would be moot as far as the SARMA process that drives the error term is concerned.

It is also useful to further elaborate on the expressions (A.1) and (A.2):

$$\begin{aligned} \boldsymbol{\Psi}_{AM}(\boldsymbol{\phi}, \boldsymbol{\psi}) &= \left(\mathbf{I} - \sum_{a=1}^A \phi_a \mathbf{F}_a \right)^{-1} \left[\left(\mathbf{I} + \sum_{m=1}^M \tilde{\psi}_m \mathbf{E}_m \right) + \sum_{m=1}^M (\psi_m - \tilde{\psi}_m) \mathbf{E}_m \right], \\ \boldsymbol{\Psi}_{AM}(\tilde{\boldsymbol{\phi}}, \tilde{\boldsymbol{\psi}})^{-1} &= \left(\mathbf{I} + \sum_{m=1}^M \tilde{\psi}_m \mathbf{E}_m \right)^{-1} \left[\left(\mathbf{I} - \sum_{a=1}^A \phi_a \mathbf{F}_a \right) + \sum_{a=1}^A (\phi_a - \tilde{\phi}_a) \mathbf{F}_a \right]. \end{aligned}$$

Hence, the product $\boldsymbol{\Phi}_{AM}(\boldsymbol{\theta}_2, \tilde{\boldsymbol{\theta}}_2) \equiv \boldsymbol{\Psi}_{AM}^{-1}(\tilde{\boldsymbol{\phi}}, \tilde{\boldsymbol{\psi}}) \cdot \boldsymbol{\Psi}_{AM}(\boldsymbol{\phi}, \boldsymbol{\psi})$ can be expressed as:

$$\begin{aligned} \boldsymbol{\Phi}_{AM}(\boldsymbol{\theta}_2, \tilde{\boldsymbol{\theta}}_2) &= \mathbf{I} + \sum_{m=1}^M (\psi_m - \tilde{\psi}_m) \boldsymbol{\Upsilon}_m^\psi + \sum_{a=1}^A (\phi_a - \tilde{\phi}_a) \boldsymbol{\Upsilon}_a^\phi + \\ &\quad + \sum_{m=1}^M \sum_{a=1}^A (\psi_m - \tilde{\psi}_m) (\phi_a - \tilde{\phi}_a) \boldsymbol{\Upsilon}_{am}^{\psi, \phi} \end{aligned}$$

where $(\mathbf{\Upsilon}_1^\psi, \dots, \mathbf{\Upsilon}_M^\psi)$ are M matrices that are functions of $\boldsymbol{\theta}_2, \tilde{\boldsymbol{\theta}}_2$ and of the M \mathbf{E}_m matrices, $(\mathbf{\Upsilon}_1^\phi, \dots, \mathbf{\Upsilon}_A^\phi)$ are A matrices that are functions of $\boldsymbol{\theta}_2, \tilde{\boldsymbol{\theta}}_2$ and of the A \mathbf{F}_a matrices, while $(\mathbf{\Upsilon}_{11}^{\psi,\phi}, \dots, \mathbf{\Upsilon}_{1A}^{\psi,\phi}, \dots, \mathbf{\Upsilon}_{M1}^{\psi,\phi}, \dots, \mathbf{\Upsilon}_{MA}^{\psi,\phi})$ are some MA matrices that are functions of all the aforementioned elements. One can verify that all these matrices are linearly independent of one another so long as \mathbf{E}_m and \mathbf{F}_a are themselves linearly independent, as maintained by Assumption 5.

Proof proper. Define the following set of moment conditions, whose expectation is zero according to our model:

$$\begin{aligned} g_0(\boldsymbol{\theta}) &= \iota^\top \mathbf{u}(\boldsymbol{\theta}) \\ g_{1,qk}(\boldsymbol{\theta}) &= \mathbf{x}_k^\top \mathbf{G}^{q-1} \mathbf{u}(\boldsymbol{\theta}) - \xi_k \lambda_{1,kq}^* \quad \text{for } k = 1, \dots, K \text{ and } q = 1, \dots, Q, \\ g_{2,p}(\boldsymbol{\theta}) &= \mathbf{u}(\boldsymbol{\theta})^\top \mathbf{P}^p \mathbf{u}(\boldsymbol{\theta}) - \lambda_{2,p}^* \quad \text{for } p = 1, \dots, P, \end{aligned}$$

where $Q \geq 4, P \geq 1 + A + M$, and:

$$\begin{aligned} \lambda_{1,kq}^* &\equiv \text{Tr}(\mathbf{C}_k^\top \mathbf{G}^{q-1}), \\ \lambda_{2,p}^* &\equiv \text{Tr}(\mathbf{P}_p). \end{aligned}$$

Observe that these moment conditions differ slightly from those described in Section 4. We stack these moments vertically to make up the vector $\mathbf{g}(\boldsymbol{\theta})$ of length $1 + QK + P$. This proof shows that the the expectation of $\mathbf{g}(\boldsymbol{\theta})$ is set at zero by a unique sequence of parameter values, or structure. It is useful to partition $\mathbf{g}(\boldsymbol{\theta})$ in two blocks: the one includes the first $1 + QK$ rows, and the one that includes the remaining P rows, and analyze the expectations of these two blocks in sequence.

Following some manipulation, the expectation of the first block can be written as a function of $\tilde{\boldsymbol{\theta}}$, an arbitrary vector of parameter values, as:

$$\begin{aligned} \mathbb{E} \left[\mathbf{g}_1(\tilde{\boldsymbol{\theta}}) \right] &= \mathbb{E} \left[(\mathbf{K}_{\tilde{\mathbf{x}}} + \mathbf{K}_{\mathbf{u}})^\top \cdot \Psi_{AM}^{-1}(\tilde{\boldsymbol{\phi}}, \tilde{\boldsymbol{\psi}}) \cdot (\mathbf{S}_{\tilde{\mathbf{x}}} + \mathbf{S}_{\mathbf{u}}) \right] (\boldsymbol{\theta}_1 - \tilde{\boldsymbol{\theta}}_1) + \\ &\quad + \frac{\sigma}{\tilde{\sigma}} \mathbb{E} \left[(\mathbf{K}_{\tilde{\mathbf{x}}} + \mathbf{K}_{\mathbf{u}})^\top \cdot \Psi_{AM}^{-1}(\tilde{\boldsymbol{\phi}}, \tilde{\boldsymbol{\psi}}) \cdot \Psi_{AM}(\boldsymbol{\phi}, \boldsymbol{\psi}) \cdot \mathbf{v} \right] - \Lambda \tilde{\boldsymbol{\xi}} \quad (\text{A.3}) \end{aligned}$$

where $\mathbf{K}_{\tilde{\mathbf{x}}}$ and $\mathbf{K}_{\mathbf{u}}$ are two matrices of dimension $N \times 1 + QK$, while $\mathbf{S}_{\tilde{\mathbf{x}}}$ and $\mathbf{S}_{\mathbf{u}}$ are two matrices of dimension $N \times 2(1 + K)$:

$$\begin{aligned} \mathbf{K}_{\tilde{\mathbf{x}}} &\equiv \begin{bmatrix} \iota & \tilde{\mathbf{X}} & \mathbf{G}\tilde{\mathbf{X}} & \dots & \mathbf{G}^{q-1}\tilde{\mathbf{X}} \end{bmatrix}, \\ \mathbf{K}_{\mathbf{u}} &\equiv \begin{bmatrix} \mathbf{0} & [\mathbf{I} & \mathbf{G} & \dots & \mathbf{G}^{q-1}] [\iota_Q^\top \otimes [[\mathbf{C}_1 \mathbf{v} \dots \mathbf{C}_K \mathbf{v}] \boldsymbol{\xi}]] \end{bmatrix}, \\ \mathbf{S}_{\tilde{\mathbf{x}}} &\equiv \begin{bmatrix} \iota & \mathbf{G}(\mathbf{I} - \beta \mathbf{G})^{-1} (\alpha \iota + \tilde{\mathbf{X}} \boldsymbol{\gamma} + \mathbf{G} \tilde{\mathbf{X}} \boldsymbol{\delta}) & \tilde{\mathbf{X}} & \mathbf{G} \tilde{\mathbf{X}} \end{bmatrix}, \\ \mathbf{S}_{\mathbf{u}} &\equiv \begin{bmatrix} \mathbf{0} & \mathbf{G}(\mathbf{I} - \beta \mathbf{G})^{-1} \left(\sum_{k=1}^K \gamma_k \xi_k \mathbf{C}_k + \Psi_{AM}(\boldsymbol{\phi}, \boldsymbol{\psi}) \right) \mathbf{v} \\ & [\mathbf{C}_1 \mathbf{v} \dots \mathbf{C}_K \mathbf{v}] \boldsymbol{\xi} & \mathbf{G} [\mathbf{C}_1 \mathbf{v} \dots \mathbf{C}_K \mathbf{v}] \boldsymbol{\xi} \end{bmatrix}, \end{aligned}$$

$\tilde{\mathbf{X}} \equiv [\tilde{\mathbf{x}}_1 \ \dots \ \tilde{\mathbf{x}}_K]$ is a $N \times K$ matrix that gathers the independent components of \mathbf{X} , and:

$$\mathbf{\Lambda} \equiv \begin{bmatrix} 0 & 0 & \dots & 0 \\ \text{Tr}(\mathbf{C}_1) & 0 & \dots & 0 \\ 0 & \text{Tr}(\mathbf{C}_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \text{Tr}(\mathbf{C}_K) \\ \text{Tr}(\mathbf{G}\mathbf{C}_1) & 0 & \dots & 0 \\ 0 & \text{Tr}(\mathbf{G}\mathbf{C}_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \text{Tr}(\mathbf{G}\mathbf{C}_K) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Tr}(\mathbf{C}_1\mathbf{G}^Q) & 0 & \dots & 0 \\ 0 & \text{Tr}(\mathbf{C}_2\mathbf{G}^Q) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \text{Tr}(\mathbf{C}_K\mathbf{G}^{Q-1}) \end{bmatrix},$$

is a $1 + QK \times K$ matrix with full column rank if $\text{Tr}(\mathbf{C}_k)$, $\text{Tr}(\mathbf{C}_k\mathbf{G})$, \dots , $\text{Tr}(\mathbf{C}_k\mathbf{G}^{Q-1})$ are not simultaneously all zero for $k = 1, \dots, K$, i.e. condition (ii) of the Theorem.

Suppose that (A.3) equals a vector of zeroes. Following some more manipulation, the block can be written as:

$$\begin{aligned} \mathbb{E} \left[\mathbf{g}_1(\tilde{\boldsymbol{\theta}}) \right] &= \mathbb{E} \left[(\mathbf{K}_{\tilde{\mathbf{x}}} + \mathbf{K}_{\mathbf{u}})^T \cdot \boldsymbol{\Psi}_{AM}^{-1}(\boldsymbol{\phi}, \boldsymbol{\psi}) \cdot (\mathbf{S}_{\tilde{\mathbf{x}}} + \mathbf{S}_{\mathbf{u}}) \right] (\boldsymbol{\theta}_1 - \tilde{\boldsymbol{\theta}}_1) + \\ &+ \frac{\sigma}{\tilde{\sigma}} \mathbf{\Lambda} (\boldsymbol{\xi} - \tilde{\boldsymbol{\xi}}) + \frac{1}{\tilde{\sigma}} \mathbf{\Lambda} \tilde{\boldsymbol{\xi}} (\sigma - \tilde{\sigma}) + \frac{\sigma}{\tilde{\sigma}} \mathbb{E} \left[(\mathbf{K}_{\tilde{\mathbf{x}}} + \mathbf{K}_{\mathbf{u}})^T \left(\boldsymbol{\Phi}_{AM}(\boldsymbol{\theta}_2, \tilde{\boldsymbol{\theta}}_2) - \mathbf{I} \right) \mathbf{v} \right] = \mathbf{0} \end{aligned} \quad (\text{A.4})$$

where the matrix that pre-multiplies $(\boldsymbol{\theta}_1 - \tilde{\boldsymbol{\theta}}_1)$ on the right-hand side of the first line is of full column rank under condition (i) of the Theorem: i.e. that matrices \mathbf{I} , \mathbf{G} , \mathbf{G}^2 and \mathbf{G}^3 are all linearly independent of one another. Observe that if $\boldsymbol{\theta}_2 = \tilde{\boldsymbol{\theta}}_2$ (that is, if the $A + M$ parameters of the SARMA process are identified), the fourth element of the developed expression (the second expectation) equals zero, and thus, being $\mathbf{\Lambda}$ of full column rank, the entire first block can only be zero in expectation if and only if $\boldsymbol{\theta}_1 = \tilde{\boldsymbol{\theta}}_1$, $\boldsymbol{\xi} = \tilde{\boldsymbol{\xi}}$ and $\sigma = \tilde{\sigma}$, which would prove global identification. The remainder of this proof shows that, holding (A.4), the SARMA parameters are identified through the second block of moments. To this end, we exploit the solution for $(\boldsymbol{\theta}_1 - \tilde{\boldsymbol{\theta}}_1)$ that is implied by (A.4), which exists thanks to the already mentioned rank properties of the matrices involved. We write it in terms of two matrices \mathbf{T}_σ and $\mathbf{T}_\boldsymbol{\phi}$.

$$(\boldsymbol{\theta}_1 - \tilde{\boldsymbol{\theta}}_1) = \mathbf{T}_\sigma (\sigma - \tilde{\sigma}) + \mathbf{T}_\boldsymbol{\phi} \mathbb{E} \left[(\mathbf{K}_{\tilde{\mathbf{x}}} + \mathbf{K}_{\mathbf{u}})^T \left(\boldsymbol{\Phi}_{AM}(\boldsymbol{\theta}_2, \tilde{\boldsymbol{\theta}}_2) - \mathbf{I} \right) \mathbf{v} \right] \quad (\text{A.5})$$

We now turn our attention to the second block of moments. Write the expectation of its generic p -th element, for $p = 1, \dots, P$, as the following function of $\tilde{\boldsymbol{\theta}}$:

$$\begin{aligned} \mathbb{E} \left[g_{2,p} \left(\tilde{\boldsymbol{\theta}} \right) \right] &= \frac{1}{\tilde{\sigma}^2} \mathbb{E} \left[\left[\boldsymbol{\Psi}_{AM}^{-1} \left(\tilde{\boldsymbol{\phi}}, \tilde{\boldsymbol{\psi}} \right) \cdot \left(\mathbf{S}_{\tilde{\mathbf{x}}} + \mathbf{S}_{\mathbf{u}} \right) \cdot \left(\boldsymbol{\theta}_1 - \tilde{\boldsymbol{\theta}}_1 \right) + \boldsymbol{\Phi}_{AM} \left(\boldsymbol{\theta}_2, \tilde{\boldsymbol{\theta}}_2 \right) \cdot \boldsymbol{\nu} \right]^T \right. \\ &\quad \left. \cdot \mathbf{P}_p \cdot \left[\boldsymbol{\Psi}_{AM}^{-1} \left(\tilde{\boldsymbol{\phi}}, \tilde{\boldsymbol{\psi}} \right) \cdot \left(\mathbf{S}_{\tilde{\mathbf{x}}} + \mathbf{S}_{\mathbf{u}} \right) \cdot \left(\boldsymbol{\theta}_1 - \tilde{\boldsymbol{\theta}}_1 \right) + \boldsymbol{\Phi}_{AM} \left(\boldsymbol{\theta}_2, \tilde{\boldsymbol{\theta}}_2 \right) \cdot \boldsymbol{\nu} \right] \right] - \text{Tr} \left(\mathbf{P}_p \right). \end{aligned}$$

Substituting (A.5) into the above ultimately leads to a polynomial of the kind:

$$\begin{aligned} \mathbb{E} \left[g_{2,p} \left(\tilde{\boldsymbol{\theta}} \right) \right] &= \sum_{a=1}^A \left(\tilde{\phi}_a - \phi_a \right) \omega_{\phi_a} + \sum_{m=1}^M \left(\tilde{\psi}_m - \psi_m \right) \omega_{\psi_m} \\ &+ \sum_{a_1=1}^A \sum_{a_2=1}^A \left(\tilde{\phi}_{a_1} - \phi_{a_1} \right) \left(\tilde{\phi}_{a_2} - \phi_{a_2} \right) \omega_{\phi_{a_1} \phi_{a_2}} \\ &+ \sum_{m_1=1}^M \sum_{m_2=1}^M \left(\tilde{\psi}_{m_1} - \psi_{m_1} \right) \left(\tilde{\psi}_{m_2} - \psi_{m_2} \right) \omega_{\psi_{m_1} \psi_{m_2}} \\ &+ \sum_{a=1}^A \sum_{m=1}^M \left(\tilde{\phi}_a - \phi_a \right) \left(\tilde{\psi}_m - \psi_m \right) \omega_{\phi_a \psi_m} \\ &+ \sum_{a_1=1}^A \sum_{a_2=1}^A \sum_{m=1}^M \left(\tilde{\phi}_{a_1} - \phi_{a_1} \right) \left(\tilde{\phi}_{a_2} - \phi_{a_2} \right) \left(\tilde{\psi}_m - \psi_m \right) \omega_{\phi_{a_1} \phi_{a_2} \psi_m} \\ &+ \sum_{a=1}^A \sum_{m_1=1}^M \sum_{m_2=1}^M \left(\tilde{\phi}_a - \phi_a \right) \left(\tilde{\psi}_{m_1} - \psi_{m_1} \right) \left(\tilde{\psi}_{m_2} - \psi_{m_2} \right) \omega_{\phi_a \psi_{m_1} \psi_{m_2}} \\ &+ \sum_{a_1=1}^A \sum_{a_2=1}^A \sum_{m_1=1}^M \sum_{m_2=1}^M \left(\tilde{\phi}_{a_1} - \phi_{a_1} \right) \left(\tilde{\phi}_{a_2} - \phi_{a_2} \right) \left(\tilde{\psi}_{m_1} - \psi_{m_1} \right) \left(\tilde{\psi}_{m_2} - \psi_{m_2} \right) \omega_{\phi_{a_1} \phi_{a_2} \psi_{m_1} \psi_{m_2}} \end{aligned}$$

such that a solution for all the ω roots can be derived (although this is overwhelmingly tedious) exists and it can be expressed in terms of the $\boldsymbol{\Upsilon}_m^\psi$ and $\boldsymbol{\Upsilon}_a^\phi$ and $\boldsymbol{\Upsilon}_{ma}^{\psi,\phi}$ matrices. As these are linearly independent, the same argument as in Lee and Liu (2010) shows that all the moments of second order are only set at zero in expectation for $\boldsymbol{\theta}_2 = \tilde{\boldsymbol{\theta}}_2$. For the reasons given in our discussion of the first block, this completes the proof.

Proof of Theorem 2

In this proof, we denote by $\mathbf{x}_{k,N}^*$ the k -th column of \mathbf{X}_N for $k = 1, \dots, K$; its expected value $\mathbb{E} [\mathbf{x}_{k,N}] = \mathbb{E} [\tilde{\mathbf{x}}_{k,N}^*]$, where $\tilde{\mathbf{x}}_{k,N}$ is defined in Assumption 6, corresponds with the k -th column of $\mathbb{E} [\mathbf{X}_N]$. We also write the unconditional expected value of \mathbf{y}_N as:

$$\mathbb{E} [\mathbf{y}_N] = \left(\mathbf{I}_N - \beta_0 \mathbf{G}_N \right)^{-1} \left(\alpha_0 \boldsymbol{\iota}_N + \mathbb{E} [\mathbf{X}_N] \boldsymbol{\gamma}_0 + \mathbf{G}_N \mathbb{E} [\mathbf{X}_N] \boldsymbol{\delta}_0 \right).$$

We also introduce some additional auxiliary notation. For a start, the matrix:

$$\tilde{\mathbf{G}}_N(\boldsymbol{\beta}) \equiv \mathbf{G}_N (\mathbf{I}_N - \boldsymbol{\beta} \mathbf{G}_N)^{-1},$$

helps further define the following vectors:

$$\begin{aligned} \mathbf{d}_N(\boldsymbol{\theta}) &\equiv (\boldsymbol{\alpha}_0 - \boldsymbol{\alpha}) \boldsymbol{\iota}_N + (\boldsymbol{\beta}_0 - \boldsymbol{\beta}) \mathbf{G}_N \mathbb{E}[\mathbf{y}_N] + \mathbb{E}[\mathbf{X}_N] (\boldsymbol{\gamma}_0 - \boldsymbol{\gamma}) + \mathbf{G}_N \mathbb{E}[\mathbf{X}_N] (\boldsymbol{\delta}_0 - \boldsymbol{\delta}), \\ \mathbf{e}_N(\boldsymbol{\theta}) &\equiv \boldsymbol{\varepsilon}_N + (\mathbf{X}_N - \mathbb{E}[\mathbf{X}_N]) (\boldsymbol{\gamma}_0 - \boldsymbol{\gamma}) + \mathbf{G}_N (\mathbf{X}_N - \mathbb{E}[\mathbf{X}_N]) (\boldsymbol{\delta}_0 - \boldsymbol{\delta}) \\ &\quad + (\boldsymbol{\beta}_0 - \boldsymbol{\beta}) \tilde{\mathbf{G}}_N(\boldsymbol{\beta}_0) [\boldsymbol{\varepsilon}_N + (\mathbf{X}_N - \mathbb{E}[\mathbf{X}_N]) \boldsymbol{\gamma}_0 + \mathbf{G}_N (\mathbf{X}_N - \mathbb{E}[\mathbf{X}_N]) \boldsymbol{\delta}_0]. \end{aligned}$$

Note that $\boldsymbol{\varepsilon}_N(\boldsymbol{\theta}) = \mathbf{d}_N(\boldsymbol{\theta}) + \mathbf{e}_N(\boldsymbol{\theta})$. In addition, the following K matrices will be helpful throughout:

$$\boldsymbol{\Gamma}_{k,0}(\boldsymbol{\theta}) \equiv \left[(\gamma_{k,0} - \gamma_k) \mathbf{I}_N + (\delta_{k,0} - \delta_k) \mathbf{G}_N + (\boldsymbol{\beta}_0 - \boldsymbol{\beta}) \tilde{\mathbf{G}}_N(\boldsymbol{\beta}_0) (\gamma_{k,0} \mathbf{I}_N + \delta_{k,0} \mathbf{G}_N) \right],$$

where $k = 1, \dots, K$. Finally, observe that the GMM weighting matrix \mathbf{W}_N can be written as:

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

where \mathbf{A}_N is a square matrix of dimension $1 + QK + P$ and such that $\mathbf{A}_N \xrightarrow{P} \mathbf{A}_0$ and $\text{rank}(\mathbf{A}_N) \geq \dim|\boldsymbol{\theta}|$, where $\mathbf{A}_0^T \mathbf{A}_0 = \mathbf{W}_0$. This implies that the vector $\mathbf{A}_N \mathbf{m}_N(\boldsymbol{\theta})$ can be decomposed as:

$$\begin{aligned} \frac{1}{N} \mathbf{A}_N \mathbf{m}_N(\boldsymbol{\theta}) &= \frac{1}{N} a_{1,N} \boldsymbol{\iota}_N^T + \frac{1}{N} \left[\sum_{q=1}^Q \sum_{k=1}^K a_{1+(q-1)K+k,N} \mathbf{q}_{qk,N} \right. \\ &\quad \left. + \sum_{p=1}^P a_{1+QK+p,N} \boldsymbol{\varepsilon}_N^T(\boldsymbol{\theta}) \mathbf{P}_{p,N} \right] \boldsymbol{\varepsilon}_N(\boldsymbol{\theta}), \end{aligned} \tag{A.6}$$

where the $1 + QK + P$ elements written as $a_{.,N}$ are appropriate combinations of the elements of \mathbf{A}_N . Our main proof of consistency is based on this decomposition; later we refer to the ‘‘first’’ and the ‘‘second’’ element of (A.6) as the two summations laid out within brackets respectively in the first and second line of the above display.

Before we get to the proof proper, one final preparatory step is useful. We later further decompose the elements of (A.6) into smaller bits, through some auxiliary vectors and matrices; it is helpful to introduce these arrays immediately. They are: (i) some $K(1 + K)$ matrices, which are written as $\mathbf{R}_{k,N}^*(\boldsymbol{\theta})$ and $\mathbf{R}_{kk',N}^*(\boldsymbol{\theta})$, and are indexed by $k, k' = 1, \dots, K$:

$$\mathbf{R}_{k,N}^*(\boldsymbol{\theta}) \equiv \sum_{q=1}^Q \mathbf{G}_N^{q-1} a_{1+qk,N} \left[\mathbf{I}_N + (\boldsymbol{\beta}_0 - \boldsymbol{\beta}) \tilde{\mathbf{G}}_N(\boldsymbol{\beta}_0) \right]$$

$$\mathbf{R}_{kk',N}^*(\boldsymbol{\theta}) \equiv \sum_{q=1}^Q \mathbf{G}_N^{q-1} a_{1+qk,N} \boldsymbol{\Gamma}_{k,0}(\boldsymbol{\theta}),$$

(ii) a set of $K + 1$ vectors written as $\mathbf{l}_{0,N}^{**}(\boldsymbol{\theta})$ and as $\mathbf{l}_{k,N}^{**}(\boldsymbol{\theta})$ for $k = 1, \dots, K$;

$$\begin{aligned} \mathbf{l}_{0,N}^{**}(\boldsymbol{\theta}) &\equiv \mathbf{d}_N^T(\boldsymbol{\theta}) \sum_{p=1}^P a_{1+QK+p,N} \mathbf{P}_{p,N} \left[\mathbf{I}_N + (\boldsymbol{\beta}_0 - \boldsymbol{\beta}) \tilde{\mathbf{G}}_N(\boldsymbol{\beta}_0) \right] \\ \mathbf{l}_{k,N}^{**}(\boldsymbol{\theta}) &\equiv \mathbf{d}_N^T(\boldsymbol{\theta}) \sum_{p=1}^P a_{1+QK+p,N} \mathbf{P}_{p,N} \boldsymbol{\Gamma}_{k,0}(\boldsymbol{\theta}), \end{aligned}$$

(iii) another set of $1 + K + K^2$ matrices, written as $\mathbf{R}_{0,N}^{**}(\boldsymbol{\theta})$, $\mathbf{R}_{k,N}^{**}(\boldsymbol{\theta})$ and $\mathbf{R}_{kk',N}^{**}(\boldsymbol{\theta})$, and indexed by $k, k' = 1, \dots, K$:

$$\begin{aligned} \mathbf{R}_{0,N}^{**}(\boldsymbol{\theta}) &\equiv \left(\mathbf{I}_N + (\boldsymbol{\beta}_0 - \boldsymbol{\beta}) \tilde{\mathbf{G}}_N^T(\boldsymbol{\beta}_0) \right) \sum_{p=1}^P a_{1+QK+p,N} \mathbf{P}_{p,N} \left(\mathbf{I}_N + (\boldsymbol{\beta}_0 - \boldsymbol{\beta}) \tilde{\mathbf{G}}_N(\boldsymbol{\beta}_0) \right) \\ \mathbf{R}_{k,N}^{**}(\boldsymbol{\theta}) &\equiv \left(\mathbf{I}_N + (\boldsymbol{\beta}_0 - \boldsymbol{\beta}) \tilde{\mathbf{G}}_N^T(\boldsymbol{\beta}_0) \right) \sum_{p=1}^P a_{1+QK+p,N} (\mathbf{P}_{p,N} + \mathbf{P}_{p,N}^T) \boldsymbol{\Gamma}_{k,0}(\boldsymbol{\theta}) \\ \mathbf{R}_{kk',N}^{**}(\boldsymbol{\theta}) &\equiv \boldsymbol{\Gamma}_{k',0}^T(\boldsymbol{\theta}) \left[\sum_{p=1}^P a_{1+QK+p,N} \mathbf{P}_{p,N} \right] \boldsymbol{\Gamma}_{k,0}(\boldsymbol{\theta}). \end{aligned}$$

We now proceed to our main argument. In order to establish consistency of $\hat{\boldsymbol{\theta}}_{GMM}$, it is necessary to show uniform convergence in probability for all the elements that comprise the vector $\mathbf{A}_N \mathbf{m}_N(\boldsymbol{\theta})$. Consider the first element in brackets in (A.6):

$$\begin{aligned} \sum_{q=1}^Q \sum_{k=1}^K a_{1+(q-1)K+k,N} \mathbf{q}_{qk,N} \boldsymbol{\varepsilon}_N(\boldsymbol{\theta}) &= \underbrace{\sum_{q=1}^Q \sum_{k=1}^K a_{1+(q-1)K+k,N} (\mathbf{G}_N^{q-1} \mathbf{x}_{k,N}^*)^T \mathbf{d}_N(\boldsymbol{\theta})}_{\equiv l_N^*(\boldsymbol{\theta})} \\ &\quad + \underbrace{\sum_{q=1}^Q \sum_{k=1}^K a_{1+(q-1)K+q,N} (\mathbf{G}_N^{q-1} \mathbf{x}_{k,N}^*)^T \mathbf{e}_N(\boldsymbol{\theta})}_{\equiv r_N^*(\boldsymbol{\theta})}, \end{aligned}$$

where $l_N^*(\boldsymbol{\theta})$ is given by:

$$\frac{1}{N} l_N^*(\boldsymbol{\theta}) = \frac{1}{N} \sum_{q=1}^Q \sum_{k=1}^K a_{1+(q-1)K+k,N} (\mathbf{G}_N^{q-1} \mathbb{E}[\mathbf{x}_{k,N}^*])^T \mathbf{d}_N^T(\boldsymbol{\theta}) + o_P(1),$$

while $r_N^*(\boldsymbol{\theta})$ can be expressed as a function of the $\mathbf{R}_{k,N}^*(\boldsymbol{\theta})$ and $\mathbf{R}_{kk',N}^*(\boldsymbol{\theta})$ matrices defined above (note: the second line continues on the next page):

$$\begin{aligned}
\frac{1}{N}r_N^*(\boldsymbol{\theta}) &= \frac{1}{N} \sum_{k=1}^K (\mathbf{x}_{k,N}^* - \mathbb{E}[\mathbf{x}_{k,N}^*])^\top \mathbf{R}_{k,N}^*(\boldsymbol{\theta}) (\mathbf{x}_{k,N}^* - \mathbb{E}[\mathbf{x}_{k,N}^*]) \\
&\quad + \frac{1}{N} \sum_{k=1}^K \sum_{k'=1}^K (\mathbf{x}_{k',N}^* - \mathbb{E}[\mathbf{x}_{k',N}^*])^\top \mathbf{R}_{k,k',N}^*(\boldsymbol{\theta}) \boldsymbol{\varepsilon}_N \\
&= \sigma_0^2 \frac{1}{N} \sum_{k=1}^K \xi_{0,k} \text{Tr}(\mathbf{C}_{k,N}^\top \mathbf{R}_{k,N}^*(\boldsymbol{\theta}) (\mathbf{I}_N + \psi_0 \mathbf{E}_N)) \\
&\quad + \frac{1}{N} \sum_{k=1}^K \sum_{k'=1}^K \xi_{0,k} \xi_{0,k'} \text{Tr}(\mathbf{C}_{k,N}^\top \mathbf{R}_{k,k',N}^*(\boldsymbol{\theta}) \mathbf{C}_{k,N}) \\
&\quad + \frac{1}{N} \sum_{k=1}^K \sum_{k'=1}^K \Xi_{kk'} \cdot \text{Tr}(\mathbf{R}_{k,k',N}^*(\boldsymbol{\theta})) + o_P(1).
\end{aligned}$$

Similarly, the second term in brackets in (A.6) can be decomposed as:

$$\begin{aligned}
\sum_{p=1}^P a_{1+QK+p,N} \boldsymbol{\varepsilon}_N^\top(\boldsymbol{\theta}) \mathbf{P}_{p,N} \boldsymbol{\varepsilon}_N(\boldsymbol{\theta}) &= \sum_{p=1}^P a_{1+QK+p,N} \mathbf{d}_N^\top(\boldsymbol{\theta}) \mathbf{P}_{p,N} \mathbf{d}_N(\boldsymbol{\theta}) \\
&\quad + 2 \underbrace{\sum_{p=1}^P a_{1+QK+p,N} \mathbf{d}_N^\top(\boldsymbol{\theta}) \mathbf{P}_{p,N} \mathbf{e}_N(\boldsymbol{\theta})}_{\equiv l_N^{**}(\boldsymbol{\theta})} + \underbrace{\sum_{p=1}^P a_{1+QK+p,N} \mathbf{e}_N^\top(\boldsymbol{\theta}) \mathbf{P}_{p,N} \mathbf{e}_N(\boldsymbol{\theta})}_{\equiv r_N^{**}(\boldsymbol{\theta})},
\end{aligned}$$

where $l_N^{**}(\boldsymbol{\theta})$ is written in terms of $\mathbf{l}_{0,N}^{**}(\boldsymbol{\theta})$ and $\mathbf{l}_{k,N}^{**}(\boldsymbol{\theta})$:

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

while the term $r_N^{**}(\boldsymbol{\theta})$ can be related to $\mathbf{R}_{0,N}^{**}(\boldsymbol{\theta})$, $\mathbf{R}_{k,N}^{**}(\boldsymbol{\theta})$ and $\mathbf{R}_{kk',N}^{**}(\boldsymbol{\theta})$:

$$\begin{aligned}
\frac{1}{N} r_N^{**}(\boldsymbol{\theta}) &= \frac{1}{N} \boldsymbol{\varepsilon}_N^\top \mathbf{R}_{0,N}^{**}(\boldsymbol{\theta}) \boldsymbol{\varepsilon}_N + \frac{1}{N} \sum_{k=1}^K \boldsymbol{\varepsilon}_N^\top \mathbf{R}_{k,N}^{**}(\boldsymbol{\theta}) (\mathbf{x}_{k,N}^* - \mathbb{E}[\mathbf{x}_{k,N}^*]) \\
&\quad + \frac{1}{N} \sum_{k'=1}^K (\mathbf{x}_{k',N}^* - \mathbb{E}[\mathbf{x}_{k',N}^*])^\top \mathbf{R}_{k,k',N}^{**}(\boldsymbol{\theta}) (\mathbf{x}_{k,N}^* - \mathbb{E}[\mathbf{x}_{k,N}^*]) \\
&= \sigma_0^2 \frac{1}{N} \text{Tr} \left((\mathbf{I}_N + \psi_0 \mathbf{E}_N)^\top \mathbf{R}_{0,N}^{**}(\boldsymbol{\theta}) (\mathbf{I}_N + \psi_0 \mathbf{E}_N) \right)
\end{aligned}$$

$$\begin{aligned}
& + \frac{1}{N} \sum_{k=1}^K \sigma_0^2 \xi_{0,k} \text{Tr} \left((\mathbf{I}_N + \psi_0 \mathbf{E}_N)^T \mathbf{R}_{k,N}^{**}(\boldsymbol{\theta}) \mathbf{C}_{k,N} \right) \\
& + \frac{1}{N} \sum_{k=1}^K \sum_{k'=1}^K \sigma_0^2 \xi_{0,k} \xi_{0,k'} \text{Tr} \left(\mathbf{C}_{k,N}^T \mathbf{R}_{k,k',N}^{**}(\boldsymbol{\theta}) \mathbf{C}_{k,N} \right) \\
& + \frac{1}{N} \sum_{k=1}^K \sum_{k'=1}^K \Xi_{kk'} \cdot \text{Tr} \left(\mathbf{R}_{k,k',N}^{**}(\boldsymbol{\theta}) \right) + o_P(1).
\end{aligned}$$

Note that $N^{-1} \mathbf{v}^T \{ \mathbf{x}_{k,N}^* - \mathbb{E}[\mathbf{x}_{k,N}^*] \} = o_P(1)$ for $k = 1, \dots, K$ uniformly in $\boldsymbol{\theta} \in \Theta$ by Lemmas A.3 and A.4 in Lee (2007a). Since Θ is bounded and all the terms $l_N^*(\boldsymbol{\theta})$, $r_N^*(\boldsymbol{\theta})$, $l_N^{**}(\boldsymbol{\theta})$ and $r_N^{**}(\boldsymbol{\theta})$ can be expressed as appropriate functions of the relevant parameters, uniform convergence follows. Since $\mathbf{m}_N(\boldsymbol{\theta})$ is also quadratic in $\boldsymbol{\theta}$ and Θ is bounded, then $\mathbb{E}[\mathbf{m}_N(\boldsymbol{\theta})]$ is uniformly equicontinuous in Θ . This result, along with the identification conditions, implies that the identification uniqueness condition for $\mathbb{E}[\mathbf{m}_N^T(\boldsymbol{\theta}) \mathbf{A}_N^T \mathbf{A}_N \mathbf{m}_N(\boldsymbol{\theta})]$ is satisfied. Thus, the consistency of the GMM estimator follows from standard arguments (White, 1996).

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

$$\sqrt{N} \left(\widehat{\boldsymbol{\theta}}_{GMM} - \boldsymbol{\theta}_0 \right) = - \left[\mathbf{J}_N^T \left(\widehat{\boldsymbol{\theta}}_{GMM} \right) \mathbf{W}_N \mathbf{J}_N \left(\bar{\boldsymbol{\theta}} \right) \right]^{-1} \mathbf{J}_N^T \left(\widehat{\boldsymbol{\theta}}_{GMM} \right) \mathbf{W}_N \sqrt{N} \mathbf{m}_N \left(\boldsymbol{\theta}_0 \right).$$

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

$$\sqrt{N} \mathbf{A}_N \mathbf{m}_N(\boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{A}_0 \boldsymbol{\Omega}_0 \mathbf{A}_0^T). \quad (\text{A.7})$$

Hence, the main result would follow if $\mathbf{J}_N \left(\widehat{\boldsymbol{\theta}}_{GMM} \right) = \mathbf{J}_0 + o_P(1)$. Note that:

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

where $\mathbf{0}_N$ is shorthand for an N -dimensional vector of zeros. Leaving $\frac{\partial}{\partial \boldsymbol{\theta}^T} \boldsymbol{\lambda}_N(\boldsymbol{\theta})$ aside for the moment, we focus on a submatrix of the first term on the right-hand side, that is the last P rows of the second column. This vector comprises the derivatives of

the P second-order moments with respect to β ; the analysis of the rest of the matrix is just a simpler case. By Lemmas A.3 and A.4 in Lee (2007a), one can write every p -th element of said subvector, for $p = 1, 2, \dots, P$, as:

$$\frac{1}{N} \boldsymbol{\varepsilon}_N^T(\boldsymbol{\theta}) \mathbf{P}_{p,N} \tilde{\mathbf{G}}_N(\beta_0) (\boldsymbol{\alpha}_0 \boldsymbol{\iota} + \mathbf{X}_N \boldsymbol{\gamma}_0 + \mathbf{G}_N \mathbf{X}_N \boldsymbol{\delta}_0 + \boldsymbol{\varepsilon}_N) = b_{p,N} + v_{p,N} + t_{p,N} + f_{p,N},$$

where $\mathbf{y}_N(\boldsymbol{\theta}_0) \equiv (\mathbf{I}_N - \beta_0 \mathbf{G}_N)^{-1} (\boldsymbol{\alpha}_0 \boldsymbol{\iota}_N + \mathbf{X}_N \boldsymbol{\gamma}_0 + \mathbf{G}_N \mathbf{X}_N \boldsymbol{\delta}_0 + \boldsymbol{\varepsilon}_N)$. The terms on the right-hand side are instead given by:

$$b_{p,N} = \frac{1}{N} \mathbf{d}_N^T(\boldsymbol{\theta}) \mathbf{P}_{p,N} \mathbf{G}_N \mathbf{y}_N(\boldsymbol{\theta}_0) = \frac{1}{N} \mathbf{d}_N^T(\boldsymbol{\theta}) \mathbf{P}_{p,N} \mathbf{G}_N \mathbb{E}[\mathbf{y}_N] + o_P(1),$$

and:

$$\begin{aligned} v_{p,N} &= \frac{1}{N} \boldsymbol{\varepsilon}_N^T \mathbf{P}_{p,N} \mathbf{G}_N \mathbf{y}_N(\boldsymbol{\theta}_0) = \sigma_0^2 \frac{1}{N} \text{Tr} \left[\mathbf{P}_{p,N} \tilde{\mathbf{G}}_N(\beta_0) \right] \\ &+ \sum_{k=1}^K \frac{1}{N} \sigma_0^2 \text{Tr} \left[(\mathbf{I}_N + \psi_0 \mathbf{E}_N)^T \mathbf{P}_{p,N} \tilde{\mathbf{G}}_N(\beta_0) (\gamma_{k,0} \mathbf{C}_{k,N} + \delta_{k,0} \mathbf{G}_N \mathbf{C}_{k,N}) \boldsymbol{\xi}_{k,0} \right] + o_P(1), \end{aligned}$$

and:

$$\begin{aligned} t_{p,N} &= \frac{1}{N} (\beta_0 - \beta) \boldsymbol{\varepsilon}_N^T \tilde{\mathbf{G}}_N(\beta_0)^T \mathbf{P}_{p,N} \mathbf{G}_N \mathbf{y}_N(\boldsymbol{\theta}_0) \\ &= \sigma_0^2 (\beta_0 - \beta) \frac{1}{N} \left\{ \text{Tr} \left(\tilde{\mathbf{G}}_N^T(\beta_0) \mathbf{P}_{p,N} \tilde{\mathbf{G}}_N(\beta_0) \right) + \sum_{k=1}^K \text{Tr} \left[(\mathbf{I}_N + \psi_0 \mathbf{E}_N)^T \right. \right. \\ &\quad \left. \left. \cdot \tilde{\mathbf{G}}_N^T(\beta_0) \mathbf{P}_{p,N} \tilde{\mathbf{G}}_N(\beta_0) (\gamma_{k,0} \mathbf{C}_{k,N} + \delta_{k,0} \mathbf{G}_N \mathbf{C}_{k,N}) \boldsymbol{\xi}_{k,0} \right] \right\} + o_P(1), \end{aligned}$$

and:

$$\begin{aligned} f_{p,N} &= \frac{1}{N} \sum_{k=1}^K (\mathbf{x}_{k',N}^* - \mathbb{E}[\mathbf{x}_{k',N}^*])^T \boldsymbol{\Gamma}_{k,0}^T(\boldsymbol{\theta}) \mathbf{P}_{p,N} \mathbf{G}_N \mathbf{y}_N(\boldsymbol{\theta}_0) \\ &= \frac{1}{N} \sum_{k=1}^K \boldsymbol{\xi}_{k,0} \mathbf{C}_{k,N}^T \boldsymbol{\Gamma}_{k,0}^T(\boldsymbol{\theta}) \mathbf{P}_{p,N} \tilde{\mathbf{G}}_N(\beta_0) \left(\mathbf{I}_N \sigma_0^2 + \sum_{k'=1}^K (\gamma_{k',0} \mathbf{I}_N + \delta_{k',0} \mathbf{G}_N) \mathbf{C}_{k,N} \right) \\ &\quad + \frac{1}{N} \sum_{k=1}^K \sum_{k'=1}^K \Xi_{kk'} \cdot \mathbf{P}_{p,N} \tilde{\mathbf{G}}_N(\beta_0) (\gamma_{k',0} \mathbf{I}_N + \delta_{k',0} \mathbf{G}_N) + o_P(1). \end{aligned}$$

All the probability limits above imply uniform convergence for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Evaluating these terms at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, implies $d_{p,N} = v_{p,N} = t_{p,N} = f_{p,N} = o_P(1)$ as $\mathbf{d}_N(\boldsymbol{\theta}_0) = \mathbf{0}$.

Collecting these results together gives:

$$\begin{aligned} \frac{1}{N} \boldsymbol{\varepsilon}^T(\boldsymbol{\theta}) \mathbf{P}_{p,N} \mathbf{G}_N \mathbf{y}_N(\boldsymbol{\theta}_0) &= \sigma_0^2 \text{Tr} \left[\mathbf{P}_{p,N} \tilde{\mathbf{G}}_N(\boldsymbol{\beta}_0) \right] \\ &+ \sigma_0^2 \sum_{k=1}^K \xi_{k,0} \text{Tr} \left[(\mathbf{I}_N + \psi_0 \mathbf{E}_N)^T \mathbf{P}_{p,N} \tilde{\mathbf{G}}_N(\boldsymbol{\beta}_0) \mathbf{G}_N \right] + o_P(1). \end{aligned}$$

Furthermore, some tedious analysis reveals that $\frac{1}{N} \frac{\partial}{\partial \boldsymbol{\theta}^T} \boldsymbol{\lambda}_N(\boldsymbol{\theta}) = \frac{1}{N} \frac{\partial}{\partial \boldsymbol{\theta}^T} \boldsymbol{\lambda}(\boldsymbol{\theta}_0) + o_P(1)$; hence, the $P \times 1$ submatrix of $\mathbf{J}_N(\boldsymbol{\theta})$ under examination has the desired properties. Finally, consistency of $\hat{\boldsymbol{\theta}}_{GMM}$ also straightforwardly implies that $\mathbf{J}_N(\bar{\boldsymbol{\theta}}) \xrightarrow{p} \mathbf{J}_0$. These considerations, together with equation A.7, yield the desired result through the usual application of Slutsky's theorem.

Addendum: Bias of Conventional Methods, Analysis

This section elaborates the analysis of the bias entailed by conventional methods for the estimation of social effects – specifically Bramoullé et al. (2009, henceforth BDF) – as anticipated in footnote 16 of the main text. First, recall that under an exogeneity assumption about the matrix of covariates \mathbf{X} , BDF proposed a consistent estimator which employs the spatial lags of the covariates themselves as instruments. To better understand the source of endogeneity in the model presented in this paper, it is useful to examine the source of endogeneity for OLS under the exogeneity assumption in BDF. Recall the SAR model (11), written without N subscripts:

$$\mathbf{y} = \alpha\mathbf{1} + \beta\mathbf{G}\mathbf{y} + \gamma\mathbf{x} + \boldsymbol{\varepsilon},$$

and note that under homoscedasticity, OLS is based on the following moments:

$$\mathbb{E} [\mathbf{1}^T \boldsymbol{\varepsilon}] = 0 \tag{A.8}$$

$$\mathbb{E} [(\mathbf{G}\mathbf{y})^T \boldsymbol{\varepsilon}] = \sigma_0^2 \text{Tr} ((\mathbf{I} - \beta\mathbf{G})^{-1} \mathbf{G}^T) \tag{A.9}$$

$$\mathbb{E} [\mathbf{x}^T \boldsymbol{\varepsilon}] = 0, \tag{A.10}$$

where (A.9) is better understood by noting that:

$$\mathbb{E} [(\mathbf{G}\mathbf{y})^T \boldsymbol{\varepsilon}] = \mathbb{E} [\boldsymbol{\varepsilon}^T (\mathbf{I} - \beta\mathbf{G})^{-1} \mathbf{G}^T \boldsymbol{\varepsilon}].$$

The bias arising from endogeneity is proportional to the right-hand side of (A.9). Since $\mathbf{G}\mathbf{y}$ linearly depends on $\boldsymbol{\varepsilon}$, this moment is non-zero in expectation, and therefore OLS is inconsistent. This is circumvented by substituting it by the moment:

$$\mathbb{E} [(\mathbf{G}^q \mathbf{y})^T \boldsymbol{\varepsilon}] = 0$$

for some positive integer q . This moment equals zero in expectation and is therefore valid so long as the adjacency matrix \mathbf{G} satisfies the conditions spelled out by BDF (i.e. \mathbf{I} , \mathbf{G} and \mathbf{G}^2 need to be linearly independent).

The model we consider generalizes that by BDF by making \mathbf{x} and $\boldsymbol{\varepsilon}$ correlated. Consider for simplicity the case with only one individual covariate ($K = 1$) as well as a SARMA(0,1) specification. For some positive integer q , our key moments are given by the equations:

$$\begin{aligned} \mathbb{E} [(\mathbf{G}\mathbf{y})^T \boldsymbol{\varepsilon}] &= \mathbb{E} [(\gamma\mathbf{x} + \boldsymbol{\varepsilon})^T (\mathbf{I} - \beta\mathbf{G})^{-1} \mathbf{G}\boldsymbol{\varepsilon}] \\ &= \gamma\xi\sigma^2 \text{Tr} (\mathbf{C}^T (\mathbf{I} - \beta\mathbf{G})^{-1} \mathbf{G}^T (\mathbf{I} + \psi\mathbf{E})) \\ &\quad + \sigma^2 \text{Tr} ((\mathbf{I} - \beta\mathbf{G})^{-1} \mathbf{G}^T), \end{aligned} \tag{A.11}$$

and:

$$\mathbb{E} \left[(\mathbf{G}^q \mathbf{x})^T \boldsymbol{\varepsilon} \right] = \xi \sigma^2 \text{Tr} \left((\mathbf{G}^q \mathbf{C})^T (\mathbf{I} + \psi \mathbf{E}) \right). \quad (\text{A.12})$$

When the model features endogeneity ($\xi \neq 0$), both moments (A.11) and (A.12) are non-zero in expectation. Observe that (A.11) is composed of two terms: one that encodes the endogeneity of $\mathbf{G}\mathbf{y}$ relative to $\boldsymbol{\varepsilon}$, and one that captures the endogeneity between \mathbf{x} and $\boldsymbol{\varepsilon}$. On the other hand, the bias in (A.12) is entirely due to the endogeneity of \mathbf{x} . The bias depends crucially on the interaction between the network adjacency matrix \mathbf{G} and the characteristics matrix \mathbf{C} , which determines the spatial correlation of the different variables at hand. Note also that the spatial MA(1) term of $\boldsymbol{\varepsilon}$, expressed by the term $\psi \mathbf{E}$, amplifies the diffusion across individuals, but it does not cause a bias to the BDF moments so long as the individual characteristics are exogenous ($\xi = 0$).