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

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

June 2024

Abstract

Conventional methods for the estimation of peer, social or network effects are invalid if individual unobservables and covariates correlate across observations. 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 covariates 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 typical 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 Graham, Ida Johnsson and Michele Pellizzari for helpful discussions and their 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@unican.es`.

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

1 Introduction

A sizable body of economic research examines peer effects, network effects and more generally "social effects:" 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] Originally, our understanding of of social effects was hindered by the so-called "reflection problem" (Manski, 1993).[2] However, over time advances have been made. To 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; Blume et al., 2015).[3] Yet, this approach is largely confined to settings where the observable characteristics in question, in addition to the structure of socio-economic interactions, are both as good as exogenous. This makes studies based on such settings liable to a critique that was put forward most notably by Angrist (2014). According to it, 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 the observed and unobserved individual characteristics are: (i) 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).[4] Similarly to related empirical models, ours can be derived from an explicit theoretical framework

[1]Other examples include studies about knowledge spillovers across firms (Jaffe, 1986; Bloom et al., 2013; Zacchia, 2020), peer effects in scientific production (e.g. Azoulay et al., 2010; Waldinger, 2012) and learning externalities (Conley and Udry, 2010).

[2]Social effects occurring within segregated groups with homogeneous relationships between their members are hard to identify, as group characteristics are simultaneous with group outcomes.

[3]Bramoullé et al. (2009) highlighted in particular the identification power of networked structures of interaction in actual empirical settings. The estimation framework that they adopt dates back in spatial econometrics to at least Kelejian and Prucha (1998).

[4]A relevant strand of this literature (Kelejian and Prucha, 2004; Liu, 2014, 2020; Liu and Saraiva, 2015, 2019; Yang and Lee, 2017; Cohen-Cole et al., 2018) examines simultaneous equations models (SEMs) with spatially autoregressive terms for one or more of its endogenous variables. As discussed more elaborately in section 2.2, our model can be seen as a particular kind of SEM.

featuring strategic interaction, which can accommodate contexts ranging from peer effects in the classroom to Research and Development (R&D) spillovers. In our model, the combinations of features (i) and (ii) above not only makes standard estimates of social effects inconsistent, but can also be 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 empirical literature about peer or social effects. This raises the question of whether the latter are testable in such a framework.

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. In peer networks for example, 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. While the moment conditions that motivate our identification results are non-linear in the structural error term, for practical purposes they are best expressed as standard linear moments augmented by a bias-correction term. In our econometric framework we also introduce a number of covariance restrictions, which correspond more closely to the second-order moments introduced by Lee (2007a) and appearing in many other studies, and that lead to efficiency improvements.[5]

We propose a GMM approach for the joint estimation of both social effects and all other parameters of our model. We derive the asymptotic properties of the resulting 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 based on an application of our method which accounts for geography-driven cross-correlation are typically smaller in magnitude compared to customary approaches, and often not statistically significant. This pattern holds under specific assumptions about the dependence structure, but is robust to perturbations of it. This 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.

It is useful to elaborate upon our contribution to the econometrics of social effects. Most studies in this tradition 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.[6] 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, 2023), yet individual covariates are typically assumed exogenous in such studies.

[5]Many contributions to the econometrics of social and peer effects have explored the identification power of covariance restrictions and quadratic moment conditions more generally (e.g. Glaeser et al., 1996; Moffitt, 2001; Graham, 2008; Davezies et al., 2009; Pereda-Fernández, 2017; Rose, 2017a).

[6]The leading case is given by Bramoullé et al. (2009), who allow for fixed effects specific to each of the multiple "networks" that make up their samples. To remove these effects, they propose local data demeaning procedures that precede their main two-stages estimation approach.

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).[7]

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; while also elaborating on other selected aspects of our analysis.

2 Model

2.1 Description

We examine an econometric model that relates $K+1$ observable variables with one another: a vector of *outcomes* $\mathbf{y}$ of dimension N, and a matrix of *explanatory variables* $\mathbf{X}$ of dimension $N \times K$. Here, N is the sample size; the dependence of algebraic objects on N is, for simplicity, for the moment treated as implicit in our notation. Our model is summarized by the following system of equations:

$$\mathbf{y} = \alpha\boldsymbol{\iota} + \beta\mathbf{G}\mathbf{y} + \mathbf{X}\boldsymbol{\gamma} + \mathbf{G}\mathbf{X}\boldsymbol{\delta} + \boldsymbol{\varepsilon} \tag{1}$$

$$\mathbf{X}_{*,k} = \widetilde{\mathbf{x}}_k + \xi_k \mathbf{C}_k \boldsymbol{\varepsilon} \quad \text{for } k = 1, \dots, K \tag{2}$$

$$\boldsymbol{\varepsilon} = (\mathbf{I} + \psi\mathbf{E})\,\boldsymbol{\upsilon}, \tag{3}$$

where $\boldsymbol{\upsilon}$ is a vector of dimension N that collects the *fundamental disturbances* of our model, which we distinguish from the *structural errors* $\boldsymbol{\varepsilon}$; $\boldsymbol{\iota}$ is a vector of dimension

[7] Zacchia (2020) analyzes a model of R&D spillovers in which firms' unobservables are correlated in a network of R&D relationships, and are simultaneous to the R&D of connected firms. To identify spillover effects, he constructs IVs motivated on the finite empirical spatial correlation of R&D.

N whose all elements equal one; $\mathbf{G}$, $\mathbf{E}$ and $\mathbf{C}_k$ (for $k = 1, \ldots, K$) are $N \times N$ matrices whose interpretation is elaborated next;[8] for $k = 1, \ldots, K$, $\mathbf{X}_{*,k}$ represents the k-th column of $\mathbf{X}$, while $\widetilde{\mathbf{x}}_k$ is a random vector of dimension N that we call the *independent component* of $\mathbf{X}_{*,k}$; lastly, the system features $3(1 + K)$ parameters, which we collect as $\boldsymbol{\vartheta} = (\alpha, \beta, \boldsymbol{\gamma}, \boldsymbol{\delta})$ and $\boldsymbol{\theta} = (\boldsymbol{\vartheta}, \boldsymbol{\xi}, \psi)$, and where $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_K)$, $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_K)$ and $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_K)$ are parameter vectors of dimension K.

Equation (1) is our main structural equation: an augmented[9] SAR model. Its key element is the "spatial lag" $\mathbf{Gy}$, which is governed by the *spatial weighting matrix* $\mathbf{G}$, also called *adjacency matrix* in network settings. The elements g_{ij} of this matrix (for $i, j = 1, \ldots, N$) represent the intensity of social interactions directed from observation j to observation i. In a peer effects setting, for example, a higher value of g_{ij} denotes a stronger influence of j on i. The parameter β associated with the spatial lag encodes the magnitude of social effects; this interpretation follows from the derivation of (1), which is standard in the literature (and that we also revisit in the Appendix), as an equilibrium relationship in a game of social interactions. The parameters $\boldsymbol{\delta}$ represent instead the *contextual*, direct effects of an observation's socio-economic connections on its own outcomes. Under the terminology introduced by Manski (1993), parameters β and $\boldsymbol{\delta}$ are called the *endogenous* and *exogenous* effects, respectively. We adopt the following assumption, which is standard in the literature.

Assumption 1. Bounded adjacencies: *matrix* $\mathbf{G}$ *has a zero diagonal,* $\beta \in (\beta_L, \beta_U)$ *is restricted to an interval such that matrix* $(\mathbf{I} - \beta\mathbf{G})$ *is non-singular, and both* $\mathbf{G}$ *and* $(\mathbf{I} - \beta\mathbf{G})^{-1}$ *are uniformly bounded in absolute value for both row and column sums.*

We impose otherwise no restriction on $\mathbf{G}$. In network settings, in particular, (1) can flexibly accommodate interactions that occur either in a large network with a single connected component, or in multiple, smaller, distinct networks.[10]

Expression (2) is the key innovation of our model: it introduces, as we designate it, a "spatial linear endogeneity" (SLE) specification for the explanatory variables in $\mathbf{X}$. In particular, (2) specifies that every column of $\mathbf{X}$ depends on the structural errors $\boldsymbol{\varepsilon}$ in a linear fashion through the multiplicative term $\xi_k \mathbf{C}_k$. We refer to the collection

[8] It is implicitly understood that in (3) and elsewhere, $\mathbf{I}$ is an identity matrix of dimension $N \times N$.

[9] Under Elhorst's (2014) classification of spatial econometric models, (1) is recognized as a "Spatial Durbin Model" due to the inclusion of the $\mathbf{GX}$ term.

[10] In a similar vein, $\mathbf{G}$ can accommodate directed or undirected, weighted or unweighted networked structures of interaction.

$\{\mathbf{C}_k\}_{k=1}^K$ as the model's *characteristic matrices*; unlike the corresponding parameters $\boldsymbol{\xi}$, we treat them as known by the econometrician. Therefore, (2) describes a scenario where the econometrician is aware that the covariates in $\mathbf{X}$ endogenously depend on $\boldsymbol{\varepsilon}$ according to a pre-determined spatial pattern, but does not know the *magnitude* of this dependence. The choice of a characteristic matrix depends on one's application; for example, both $\mathbf{C}_k = \mathbf{I}$ and $\mathbf{C}_k = \mathbf{I} + \mathbf{G}$ are potentially valid choices for any given k. We impose no other restriction not contemplated by the following assumption.

Assumption 2. Bounded characteristics: *all characteristic matrices in* $\{\mathbf{C}_k\}_{k=1}^K$ *are uniformly bounded in absolute value for both row and column sums.*

This assumption disciplines the variance of $\mathbf{X}$. Later in this section, we elaborate on examples and scenarios that are well accommodated by our SLE specification in (2).

The "independent components" $\widetilde{\mathbf{x}}_k$, as specified in (2) for $k = 1, \ldots, K$, capture the part of each explanatory variable that is exogenous to $\boldsymbol{\varepsilon}$. In a schooling context, for example, family background likely correlates with unobserved ability as well as with other factors. Let $\widetilde{\mathbf{X}}$ be the $N \times K$ matrix that isolates all independent components of $\mathbf{X}$, such that $\widetilde{\mathbf{X}}_{*,k} = \widetilde{\mathbf{x}}_k$ for $k = 1, \ldots, K$. We assume the following about $\widetilde{\mathbf{X}}$.

Assumption 3. Exogenous independent components: *matrix* $\widetilde{\mathbf{X}}$ *is exogenous, non-stochastic, bounded and has full column rank* K*; in addition, all its columns are linearly independent of* $\boldsymbol{\iota}$*, and* $\lim_{N\to\infty} N^{-1}\widetilde{\mathbf{X}}^{\mathrm{T}}\widetilde{\mathbf{X}}$ *exists and is nonsingular.*

Assumption 3 adapts standard full rank conditions to our model, while establishing exogeneity of $\widetilde{\mathbf{X}}$. The latter in particular loads (2) with additional interpretation. We think of (2) as a set of structural equations that disentangle the exogenous from the endogenous components of each explanatory variable. This raises questions about the consequences of misspecification in our model: an issue examined later in the paper.

Lastly, (3) specifies a first-order spatial moving average, SMA(1), for the structural error terms $\boldsymbol{\varepsilon}$. While many studies in this literature focus on autoregressive processes for the error term (among the others: Kapoor et al., 2007; Kelejian and Prucha, 2010; Drukker et al., 2023) we entertain the SMA(1) case since it more closely aligns with our network-based applications of interest. In fact, when $\mathbf{E} = \mathbf{G}$ and the elements of $\boldsymbol{\upsilon}$ are mutually independent, (3) implies zero correlation between the structural errors of observation pairs at three or more degrees of separation.[11] Regardless, as for other

[11] In a study about the health outcomes of children, Christakis and Fowler (2013) find that most

models in the literature, it is conceptually easy (albeit tedious) to extend our results to spatially-autoregressive-and-moving-average processes of indeterminate order for the structural errors.[12,13] We make the following assumptions about the fundamental disturbances $\boldsymbol{\upsilon}$, which drive the errors $\boldsymbol{\varepsilon}$ and thus, the endogenous components of $\mathbf{X}$.

Assumption 4. Fundamental disturbances: *it is* $\mathbb{E}[\boldsymbol{\upsilon}] = \mathbf{0}$, *and:*

$$\boldsymbol{\Sigma} \equiv \mathbb{E}\left[\boldsymbol{\upsilon}\boldsymbol{\upsilon}^{\mathrm{T}}\right] = \operatorname{diag}\left(\sigma_1^2, \dots, \sigma_N^2\right) = \begin{bmatrix} \sigma_1^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_N^2 \end{bmatrix},$$

with $\sigma_i^2 \in (0, \infty)$ *for* $i = 1, \dots, N$. *Furthermore, for some* $d > 0$, $\mathbb{E}\left[|\upsilon_i|^{4+d}\right] < \infty$ *for* $i = 1, \dots, N$.

Thus, the elements of $\boldsymbol{\upsilon}$ are mutually independent and display heteroschedasticity of arbitrary form, as in many contributions to the spatial econometrics literature (e.g. Kelejian and Prucha, 2010; Lin and Lee, 2010; Liu and Saraiva, 2019, and more). We also make the following assumption about the primitives of the SMA(1) process.

Assumption 5. Spatial MA errors: *matrix* $\mathbf{E}$ *has a zero diagonal,* $\psi \in (\psi_L, \psi_U)$ *is restricted to an interval such that matrix* $(\mathbf{I} + \psi\mathbf{E})$ *is non-singular, and both* $\mathbf{E}$ *and* $(\mathbf{I} + \psi\mathbf{E})^{-1}$ *are uniformly bounded in absolute value for both row and column sums.*

We finalize the description of the model by making one more assumption.

Assumption 6. Exogenous spatial matrices: *the adjacency matrix* $\mathbf{G}$, *the SMA matrix* $\mathbf{E}$ *and the characteristic matrices* $\{\mathbf{C}_k\}_{k=1}^{K}$ *are all exogenous.*

Assumption 6 is standard in spatial econometrics: we extend it to the characteristic matrices, treating the spatial breadth of SLE that these capture as unrelated to the model's disturbances. As far as $\mathbf{G}$ is concerned, we acknowledge the growing interest

variables of interest display spatial autocorrelation in the friendship network only up to two degrees of separation. Zacchia (2020) observes the same pattern in his study on R&D spillovers.

[12]In a previous version of this paper, we provided identification results for this more general case under a slightly modified setup with homoschedastic $\boldsymbol{\upsilon}$.

[13]Specifying such a general data generation process for $\boldsymbol{\varepsilon}$ can go a long way to approximate the true structure of spatial dependence. In network settings, this would help elude a critique by Goldsmith-Pinkham and Imbens (2013), who lament the lack of general results that enable inference when all observations are related through a large network and whose unobservables are mutually dependent. As pointed out by an anonymous referee, however, identification can be weak under such a strategy.

for the implications of endogenous network formation in models about social effects.[14] However, assuming that $\mathbf{G}$ is exogenous helps isolate and address the SLE mechanism: as we elaborate in the discussion of our empirical application, even randomizing the peer groups is not sufficient to solve the problem induced by SLE if spatial correlation in the unobservables is pervasive and covariates are even mildly endogenous.

2.2 Discussion

It is useful to elaborate on the relationship of this model with the broader literature, as well as on its relevance for actual empirical research.[15] The SLE specification in (2) is fairly general; thus, it can accommodate many instances of endogeneity from actual applications (some examples of which are discussed later in this section). At the same time, it can be seen as a restricted version of a more general simultaneous equations model (SEM). In fact, for $k = 1, \ldots, K$ and $\varsigma_k = \xi_k^{-1}\widetilde{\mathbf{x}}_k$, (2) can be rewritten as:

$$\mathbf{X}_{*,k} = \left(\xi_k^{-1}\mathbf{I} + \gamma_k\mathbf{C}_k + \delta_k\mathbf{C}_k\mathbf{G}\right)^{-1}\mathbf{C}_k\left[\left(\mathbf{I} - \beta\mathbf{G}\right)\mathbf{y} - \alpha\iota - \mathbf{X}_{\setminus k}\boldsymbol{\gamma}_{\setminus k} - \mathbf{G}\mathbf{X}_{\setminus k}\boldsymbol{\delta}_{\setminus k}\right] + \varsigma_k \tag{4}$$

provided that $\xi_k \neq 0$ and $\left(\xi_k^{-1}\mathbf{I} + \gamma_k\mathbf{C}_k + \delta_k\mathbf{C}_k\mathbf{G}\right)$ is nonsingular (here, $\mathbf{X}_{\setminus k}$, $\boldsymbol{\gamma}_{\setminus k}$ and $\boldsymbol{\delta}_{\setminus k}$ denote respectively $\mathbf{X}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$, but deprived of their k-th column or element). Interpreting ς_k as an exogenous and *stochastic*[16] unobserved determinant of $\mathbf{X}_{*,k}$, (4) appears shaped as a structural equation that links the (endogenous) variable $\mathbf{X}_{*,k}$ to the other K endogenous variables of the model, including $\mathbf{y}$. SEMs featuring spatial lags have been extensively studied in spatial econometrics (Kelejian and Prucha, 2004; Liu, 2014, 2020; Liu and Saraiva, 2015, 2019; Yang and Lee, 2017; Cohen-Cole et al., 2018); in these studies, identification of the structural parameters is typically obtained via appropriate (exclusion) restrictions, as in classical SEMs. By contrast, our model establishes restrictions implicitly in (4), and through the characteristic matrices $\mathbf{C}_k$. We show that if the latter are known by the econometrician, $\boldsymbol{\theta}$ is identified without resorting to exogenous (instrumental) variables, untypical in classical SEMs.

[14] Extant proposals to address this issue include: (i) the Bayesian method by Goldsmith-Pinkham and Imbens (2013); (ii) control function approaches, as in Arduini et al. (2015) as well as Johnsson and Moon (2021): both build on Blume et al. (2015) and Graham (2017); (iii) a GMM framework for panel data, as in the more recent contribution by Kuersteiner and Prucha (2020).

[15] We express our gratitude to two anonymous referees who encouraged us to develop many of the ideas exposed in this discussion.

[16] This departure from Assumption 3 helps illustrate the connection between our model and SEMs. Non-stochasticity of $\widetilde{\mathbf{X}}$ can be relaxed at an expositional cost for our model's asymptotic analysis.

An implication of this relationship is that by construction, the parameters of the model cannot be straightforwardly interpreted in a causal sense. Thus, for example, any parameter γ_k (for $k = 1, \ldots, K$), cannot be used to draw conclusions about the "effect" of $\mathbf{X}_{*,k}$ on $\mathbf{y}$ (because the two are simultaneous) *unless* in the real world one can implement policies that manipulate the independent component $\widetilde{\mathbf{x}}_k$. Suppose for example that (1) is a model about firm productivity and R&D spillovers: hence, the column of $\mathbf{X}$ that encodes R&D would include an endogenous component ($\xi_k \mathbf{C}_k \boldsymbol{\varepsilon}$), which depends on firm choices, and an exogenous one ($\widetilde{\mathbf{x}}_k$), which incorporates factors subject to external manipulation, such as governmental grants to perform R&D. It is more difficult to imagine such policies in other settings, like peer effects at school. Nevertheless, our model can be used for two main purposes. First, it allows to *test* for the existence of social, spillover and network effects of various sort that are embodied in $(\beta, \boldsymbol{\delta})$, which is interesting *per se*. Second, knowledge of $\boldsymbol{\theta}$ identifies the "impulse response functions" that describe how a shock in $\boldsymbol{\varepsilon}$ propagates through socio-economic agents as a function of the matrices that describe cross-dependence ($\mathbf{G}$, $\mathbf{E}$ and the $\mathbf{C}_k$ matrices) and how it ultimately affects $\mathbf{y}$. An economically relevant example is that of a technology shock in a setting that features knowledge spillovers between firms.

When interest falls exclusively on testing for the existence of the key "endogenous" social effect β, it is fair to wonder if the model we propose is necessary at all. In fact, β can be identified (as implied by typical econometric models of social effects) via a single exogenous covariate, or an external instrument. There are at least three reasons to employ our model in practice. First, the literature in both economics and sociology emphasizes the need to disentangle endogenous (β) from exogenous ($\boldsymbol{\delta}$) effects. Our model allows to estimate exogenous effects for endogenous variables in $\mathbf{X}$. Second, it may be difficult to observe exogenous covariates or instruments, especially in network settings. For example, in our empirical application discussed in Section 6 we suspect even predetermined variables (like the gender or geographical origin of undergraduate students from Bocconi university) to be endogenous, due to issues of self-selection. In addition, under spatial correlation in the unobservables as per e.g. (3), endogeneity propagates in the network and invalidates standard moment conditions (built around higher order spatial lags of the exogenous variable/instrument) that are typically used to identify β. Third, identification based on a single covariate or instrument may be weak, possibly resulting in imprecise estimates. Our estimation approach can instead yield efficiency improvements, provided that SLE is correctly specified.

2.3 Examples

The applicability of our framework in actual empirical analysis largely hinges on the econometrician's ability to correctly specify the characteristic matrices that determine the SLE specification per (2). Here we offer a number of examples where the choices about $\mathbf{C}_k$ appear natural. To facilitate this discussion, we momentarily impose $K = 1$ and drop the k subscripts where normally due; in addition, we use x_i, $\widetilde{x}_i$, y_i and ε_i to denote individual elements of $\mathbf{X}$, $\widetilde{\mathbf{X}}$, $\mathbf{y}$ and $\boldsymbol{\varepsilon}$, respectively.

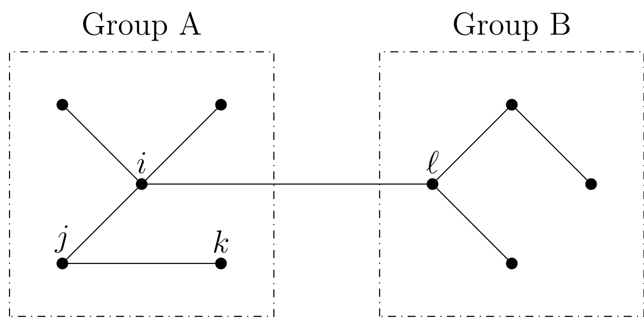
Structural endogeneity. In some settings, economic theory suggests particular spatial structures for SLE. Consider, for example, the classical endogeneity problem in the estimation of production functions (Marschak and Andrews, 1944). If x_i is the only variable log-input in a Cobb-Douglas setting, then $\mathbf{C} = \mathbf{I}$, as profit maximization induces firms to scale, in a predictable fashion, their input usage according to their unobserved shocks ε_i. Suppose now that x_i represents *knowledge capital*, which leads to productivity *spillovers* among firms according to $\mathbf{G}$, as in econometric specifications that follow the tradition by Jaffe (1989).[17] Then, economic theory suggests that firms' choice of x_i would reflect both their own shock ε_i and that of other firms directly or indirectly related through the network $\mathbf{G}$, because investment in knowledge is a public good game. Thus, one can show that in equilibrium it is $\mathbf{C} = (\mathbf{I} - a\mathbf{G})^{-1}$, where $a \geq 0$ depends on the information structure of the game (the closer the game to a complete information benchmark, the higher a).

Segregated groups. In some settings, it is natural to partition the population of interest between groups subject to "common shocks" that affect observables x_i and unobservables ε_i alike. In a schooling environment, 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 can be induced via an explicit school-level allocation mechanism, if more motivated students are assigned the best resources, or conversely, if more disadvantaged ones are compensated with extra support. Hence, $\mathbf{C}$ would display a "segregated" group structure derived from that of classrooms.[18] Matrices $\mathbf{C}$ and $\mathbf{G}$ need not be related: social interactions can

[17] These are essentially variations of (1) with $\beta = 0$ and more exclusion restrictions on $\boldsymbol{\delta}$.

[18] By "segregated" group structure we refer to a regular network topology such that, for any triad (i, j, k), if i and j are connected they are also either both connected or both disconnected to k (hence, transitivity applies), but at least some agent pairs are disconnected. This implies a block structure of the adjacency matrix: using $\mathbf{C}$ as an example, if $c_{ij} \neq 0$ then $c_{ik} \neq 0 \Leftrightarrow c_{jk} \neq 0$.

transcend classrooms, while at the same time, two classmates may not be friends. This is exemplified in Graph 1, which is inspired by typical schooling environments.



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 $\mathbf{G}$, whereas groups of observations bound within dash-dotted squares depict groups or blocks represented by $\mathbf{C}$.

Induced homophily. The issue of *homophily* in networks: the observed tendency of connections to be more likely between pair of nodes that share more characteristics, has attracted the attention of numerous social scientists, including econometricians (e.g. Graham, 2017). Usually, homophily is explained via network formation: *ex ante* similarities facilitate the establishment of links. The reverse causal mechanism (from links to similarities) has attracted less attention, but is no less plausible. Consider a scenario where in a school, connections are externally set by some agent, like a teacher or a trainer. It is likely that as a result of peer effects, students who are thus bound would develop similarities in dimensions x_i such as sport preferences and attitudes. If x_i is then incorporated in a main model of peer effects on other outcomes y_i (such as health-related ones), this instance of *induced* homophily can be accounted for via a SLE specification with, say, $\mathbf{C} = \mathbf{I} + \mathbf{G}$. Note that this simple scenario lends itself naturally to a SEM interpretation of our model, as per (4).

Measurement error. SLE can easily accommodate measurement error.[19] Let the "true" model be:

$$y_i = \alpha + \beta \sum_{j \neq i} g_{ij} y_j + \gamma \widetilde{x}_i + \eta_i$$

where η_i is some "true" error term, $x_i = \widetilde{x}_i + \omega_i$, and ω_i is measurement error in $\widetilde{x}_i$.

[19] We express our thanks to an anonymous referee who prompted us to develop this observation.

The econometrician can only observe x_i, hence the *actual* error term is $\varepsilon_i = \eta_i - \gamma\omega_i$. SLE is isomorphic to this stylized model if $\mathbf{C} = \mathbf{I}$ and $\mathbb{E}\left[\eta_i | \omega_i\right] = \left(\xi^{-1} + \gamma\right)\omega_i$, but the insight applies more generally.

3 Identification

3.1 Preliminaries

This section illustrates results about the identification of our model's parameters, and some extensions. Before proceeding, a preliminary consideration is in order. There is in fact a particular case where identification is trivial. Let $\mathbf{C}_k = \mathbf{C}$ for $k = 1, \dots, K$: if $\mathbf{C}$ has rank less than N, researchers may find a matrix $\mathbf{B}$ of dimension $N \times N$ such that $\mathbf{BC}\boldsymbol{\varepsilon} = \mathbf{0}$ and model (1) can be reshaped as:

$$\mathbf{By} = \alpha\mathbf{B}\iota + \beta\mathbf{BGy} + \boldsymbol{\gamma}\mathbf{BX} + \boldsymbol{\delta}\mathbf{BGX} + \mathbf{B}\boldsymbol{\varepsilon}, \tag{5}$$

a transformed SAR model which is identified and estimable via standard approaches since, by construction, $\mathbf{B}\boldsymbol{\varepsilon}$ is independent of $\mathbf{BX} = \mathbf{B}\widetilde{\mathbf{X}}$. A particular example is that where $\mathbf{C}$ is a block matrix describing "segregated" groups (as in the second example from the previous subsection) and whose elements are identical within each group; as a result, $\mathbf{C}\boldsymbol{\varepsilon}$ would feature identical values within a group and $\mathbf{B}$ would be a simple group-demeaning matrix.[20] Similar solutions might be found even if the characteristic matrices differ across covariates (for example, if one matrix $\mathbf{C}_k$ describes segregated groups that nest those of another matrix $\mathbf{C}_{k'}$, for $k \neq k'$). Our approach is relevant if solutions of this sort are unavailable, or researchers seek efficiency gains. However, our Monte Carlo simulations show 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.

3.2 Moments

Our identification results are based on a set of linear and quadratic moment conditions that build on the tradition initiated by Lee (2007a). A key characteristic of our setup

[20] This 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.

is that the linear moments feature an explicit bias correction term derived from more primitive zero-covariance conditions; these are actually *nonlinear* in the parameters. To illustrate, let $\mathbf{Q}_q \equiv \mathbf{G}^{q-1}\mathbf{X}$ for $q \in \mathbb{N}$. Because $\mathbf{X}$ is endogenous, for any integer q it is $\mathbb{E}\left[\mathbf{Q}_q^{\mathrm{T}}\boldsymbol{\varepsilon}\right] \neq \mathbf{0}$.[21] However, our setup suggests a natural set of appropriate moments built around the independent components $\widetilde{\mathbf{X}}$, that are by construction independent of $\boldsymbol{\varepsilon}$ (Assumption 3). Consider the following set of K moment conditions, for $q \in \mathbb{N}$:

$$\mathbb{E}\left[\left(\mathbf{G}^{q-1}\widetilde{\mathbf{X}}\right)^{\mathrm{T}}\boldsymbol{\varepsilon}\left(\boldsymbol{\theta}\right)\right] = \mathbf{0}, \tag{6}$$

where $\boldsymbol{\varepsilon}\left(\boldsymbol{\theta}\right) = \left(\mathbf{I} - \beta\mathbf{G}\right)\mathbf{y} - \alpha\boldsymbol{\iota} - \mathbf{X}\boldsymbol{\gamma} - \mathbf{G}\mathbf{X}\boldsymbol{\delta}$. Note that, by the SLE specification in (2), for $k = 1, \ldots, K$ the k-th row of (6) can be recast as:

$$\mathbb{E}\left[\left(\mathbf{X}_{*,k} - \xi_k\mathbf{C}_k\boldsymbol{\varepsilon}\left(\boldsymbol{\theta}\right)\right)^{\mathrm{T}}\left(\mathbf{G}^{q-1}\right)^{\mathrm{T}}\boldsymbol{\varepsilon}\left(\boldsymbol{\theta}\right)\right] = 0. \tag{7}$$

The quadratic form inside the expectation above is nonlinear in $\boldsymbol{\theta}$. The expression on the left-hand side of (7), however, can be developed as follows:

$$\begin{aligned}
\mathbb{E}\left[\left(\mathbf{G}^{q-1}\mathbf{X}_{*,k}\right)^{\mathrm{T}}\boldsymbol{\varepsilon}\left(\boldsymbol{\theta}\right)\right] &= \xi_k\mathbb{E}\left[\boldsymbol{\varepsilon}^{\mathrm{T}}\left(\boldsymbol{\theta}\right)\left(\mathbf{G}^{q-1}\mathbf{C}_k\right)^{\mathrm{T}}\boldsymbol{\varepsilon}\left(\boldsymbol{\theta}\right)\right] \\
&= \xi_k\mathrm{Tr}\left(\left(\mathbf{G}^{q-1}\mathbf{C}_k\right)^{\mathrm{T}}\mathbb{E}\left[\boldsymbol{\varepsilon}\left(\boldsymbol{\theta}\right)\boldsymbol{\varepsilon}^{\mathrm{T}}\left(\boldsymbol{\theta}\right)\right]\right) \\
&= \xi_k\mathrm{Tr}\left(\boldsymbol{\Upsilon}\left(\boldsymbol{\theta}\right)\mathbf{G}^{q-1}\mathbf{C}_k\right)
\end{aligned}$$

where:

$$\boldsymbol{\Upsilon}\left(\boldsymbol{\theta}\right) \equiv \left(\mathbf{I} + \psi\mathbf{E}\right)\mathbb{E}\left[\mathrm{diag}\left(\upsilon_1^2\left(\boldsymbol{\theta}\right), \ldots, \upsilon_N^2\left(\boldsymbol{\theta}\right)\right)\right]\left(\mathbf{I} + \psi\mathbf{E}\right)^{\mathrm{T}}$$

with $\boldsymbol{\upsilon}\left(\boldsymbol{\theta}\right) = \left(\upsilon_1\left(\boldsymbol{\theta}\right), \ldots, \upsilon_N\left(\boldsymbol{\theta}\right)\right) = \left(\mathbf{I} + \psi\mathbf{E}\right)^{-1}\boldsymbol{\varepsilon}\left(\boldsymbol{\theta}\right)$. It is thus straightforward to specify the *bias-correction* terms associated with moments of the $\mathbb{E}\left[\mathbf{Q}_q^{\mathrm{T}}\boldsymbol{\varepsilon}\right]$ form: these are Q vectors of dimension K collected as:

$$\boldsymbol{\lambda}_{1,q}^{\mathrm{T}}\left(\boldsymbol{\theta}\right) = \begin{bmatrix}\lambda_{1,q,1}\left(\boldsymbol{\theta}\right) & \ldots & \lambda_{1,q,K}\left(\boldsymbol{\theta}\right)\end{bmatrix}$$

for $q = 1, \ldots, Q$ and where, for $k = 1, \ldots, K$, it is:

$$\lambda_{1,q,k}\left(\boldsymbol{\theta}\right) = \xi_k\mathrm{Tr}\left(\boldsymbol{\Upsilon}\left(\boldsymbol{\theta}\right)\mathbf{G}^{q-1}\mathbf{C}_k\right). \tag{8}$$

[21] In Addendum A (Appendix) we analyze in more detail the bias entailed by conventional methods under our assumptions. This helps appreciate how the bias depends on the topology of the problem.

Accordingly, we formulate $1+QK$ "linear" moments expressed as:

$$\mathbb{E}\left[\mathbf{m}_1(\boldsymbol{\theta})\right] = \mathbb{E}\left[\begin{bmatrix} m_{1,0}(\boldsymbol{\theta}) & \mathbf{m}_{1,1}^{\mathrm{T}}(\boldsymbol{\theta}) & \dots & \mathbf{m}_{1,Q}^{\mathrm{T}}(\boldsymbol{\theta}) \end{bmatrix}^{\mathrm{T}}\right] = \mathbf{0} \tag{9}$$

where $m_{1,0}(\boldsymbol{\theta}) = \boldsymbol{\iota}^{\mathrm{T}}\boldsymbol{\varepsilon}(\boldsymbol{\theta})$ and, for $q = 1, \dots, Q$, $\mathbf{m}_{1,q}(\boldsymbol{\theta}) = \mathbf{Q}_q^{\mathrm{T}}\boldsymbol{\varepsilon}(\boldsymbol{\theta}) - \boldsymbol{\lambda}_{1,q}(\boldsymbol{\theta})$. While these moments are linear in $\boldsymbol{\varepsilon}(\boldsymbol{\theta})$, they are actually quadratic in the parameters $\boldsymbol{\theta}$, because of the bias-correction terms $\boldsymbol{\lambda}_{1,q}(\boldsymbol{\theta})$.

In addition, we establish P quadratic moments expressed as follows:

$$\mathbb{E}\left[\mathbf{m}_2(\boldsymbol{\theta})\right] = \mathbb{E}\left[\begin{bmatrix} m_{2,1}(\boldsymbol{\theta}) & \dots & m_{2,P}(\boldsymbol{\theta}) \end{bmatrix}^{\mathrm{T}}\right] = \mathbf{0} \tag{10}$$

where, for $p = 1, \dots, P$, $m_{2,p}(\boldsymbol{\theta}) = \boldsymbol{\varepsilon}^{\mathrm{T}}(\boldsymbol{\theta})\mathbf{P}_p\boldsymbol{\varepsilon}(\boldsymbol{\theta}) - \lambda_{2,p}(\boldsymbol{\theta})$, $\mathbf{P}_p$ is some appropriate $N \times N$ matrix, while

$$\lambda_{2,p}(\boldsymbol{\theta}) = \mathrm{Tr}\left(\boldsymbol{\Upsilon}(\boldsymbol{\theta})\mathbf{P}_p\right). \tag{11}$$

We make the following standard assumption about the $\mathbf{P}_p$ matrices.[22]

Assumption 7. Bounded quadratic moments: *The $\{\mathbf{P}_p\}_{p=1}^{P}$ matrices are linearly independent of one another and are all uniformly bounded in absolute value in both row and column sums.*

Since at least Lee (2007a), quadratic moments like (10) are motivated on the efficiency improvements that they deliver in a GMM estimation framework, as they leverage the observed correlations between observations. We elaborate on the appropriate choice of the $\mathbf{P}_p$ matrices in Section 4. Typically, these matrices are required to have a zero trace (under homoschedasticity) or, more generally, a zero diagonal for $\boldsymbol{\varepsilon}^{\mathrm{T}}(\boldsymbol{\theta})\mathbf{P}_p\boldsymbol{\varepsilon}(\boldsymbol{\theta})$ to be zero in expectation. By introducing the bias-correction terms (11) into (10) we can dispense with this requirement, thus keeping the framework more general.

3.3 Result

We are now ready to express some *sufficient* conditions about the linear moments (9) that ensure identification of $\boldsymbol{\theta}$.

[22] Boundedness of the $\mathbf{Q}_q$ matrices, another standard requirement, is here implied by Assumptions 1, 2 and 3.

Theorem 1. General Identification Result. *Under the maintained assumptions,* θ *is globally identified if the following conditions hold simultaneously:*

(i) $\beta\gamma_k + \delta_k \neq 0$ *for at least one* $k = 1, \dots, K$*;*

(ii) *the matrices* $\mathbf{I}$*,* $\mathbf{G}$*,* $\mathbf{G}^2$ *and* $\mathbf{G}^3$ *are linearly independent of one another;*

(iii) *for all* $k = 1, \dots, K$*, the four traces gathered in the following vector:*

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

are all simultaneously nonzero; moreover, $\xi_k \neq 0$.

Proof. See the Appendix; the proof strategy is adapted from Lee and Liu (2010). □

We find it useful to illustrate the logic of the proof in a simplified homoschedastic setting. Suppose that $K = 1$ (all k subscripts are dropped), $\boldsymbol{\Sigma} = \mathbf{I}$, $\delta = \psi = 0$, and $Q = 3$. Thus, the model reads as $\mathbf{y} = \alpha\boldsymbol{\iota} + \beta\mathbf{G}\mathbf{y} + \gamma\tilde{\mathbf{x}} + (\mathbf{I} + \gamma\xi\mathbf{C})\,\boldsymbol{\upsilon}$. Because $\tilde{\mathbf{x}}$ is unobserved, identification rests on "bias-corrected" linear moments as in (9); here we abstract from quadratic moments. Let $\boldsymbol{\theta}_0 = (\alpha_0, \beta_0, \gamma_0, \xi_0)$ be the "true" parameter vector. In this simplified setting, one can show that an attempt to evaluate (9) at an "impostor" structure $\tilde{\boldsymbol{\theta}} = \left(\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma}, \tilde{\xi}\right)$ returns:

$$\mathbb{E}\left[\mathbf{m}_1\left(\tilde{\boldsymbol{\theta}}\right)\right] = \left[\boldsymbol{\Pi}_0^* + \boldsymbol{\Pi}_1^*\right]\left(\boldsymbol{\theta}_0 - \tilde{\boldsymbol{\theta}}\right) \tag{12}$$

where, for $\widetilde{\mathbf{G}}_0 \equiv \mathbf{G}\left(\mathbf{I} - \beta_0\mathbf{G}\right)^{-1}$:

$$\boldsymbol{\Pi}_0^* = \mathbb{E}\begin{bmatrix} N & \boldsymbol{\iota}^{\mathrm{T}}\widetilde{\mathbf{G}}_0\left(\alpha\boldsymbol{\iota} + \gamma\tilde{\mathbf{x}}\right) & \boldsymbol{\iota}^{\mathrm{T}}\tilde{\mathbf{x}} & 0 \\ \tilde{\mathbf{x}}^{\mathrm{T}}\boldsymbol{\iota} & \tilde{\mathbf{x}}^{\mathrm{T}}\widetilde{\mathbf{G}}_0\left(\alpha\boldsymbol{\iota} + \gamma\tilde{\mathbf{x}}\right) & \tilde{\mathbf{x}}^{\mathrm{T}}\tilde{\mathbf{x}} & 0 \\ \left(\mathbf{G}\tilde{\mathbf{x}}\right)^{\mathrm{T}}\boldsymbol{\iota} & \left(\mathbf{G}\tilde{\mathbf{x}}\right)^{\mathrm{T}}\widetilde{\mathbf{G}}_0\left(\alpha\boldsymbol{\iota} + \gamma\tilde{\mathbf{x}}\right) & \left(\mathbf{G}\tilde{\mathbf{x}}\right)^{\mathrm{T}}\tilde{\mathbf{x}} & 0 \\ \left(\mathbf{G}^2\tilde{\mathbf{x}}\right)^{\mathrm{T}}\boldsymbol{\iota} & \left(\mathbf{G}^2\tilde{\mathbf{x}}\right)^{\mathrm{T}}\widetilde{\mathbf{G}}_0\left(\alpha\boldsymbol{\iota} + \gamma\tilde{\mathbf{x}}\right) & \left(\mathbf{G}^2\tilde{\mathbf{x}}\right)^{\mathrm{T}}\tilde{\mathbf{x}} & 0 \end{bmatrix},$$

and:

$$\boldsymbol{\Pi}_1^* = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \xi_0\mathrm{Tr}\left(\mathbf{C}\left(\mathbf{I} + \gamma_0\xi_0\mathbf{C}^{\mathrm{T}}\right)\widetilde{\mathbf{G}}_0^{\mathrm{T}}\right) & \xi_0\mathrm{Tr}\left(\mathbf{C}\mathbf{C}^{\mathrm{T}}\right) & \mathrm{Tr}\left(\mathbf{C}\right) \\ 0 & \xi_0\mathrm{Tr}\left(\mathbf{G}\mathbf{C}\left(\mathbf{I} + \gamma_0\xi_0\mathbf{C}^{\mathrm{T}}\right)\widetilde{\mathbf{G}}_0^{\mathrm{T}}\right) & \xi_0\mathrm{Tr}\left(\mathbf{G}\mathbf{C}\mathbf{C}^{\mathrm{T}}\right) & \mathrm{Tr}\left(\mathbf{G}\mathbf{C}\right) \\ 0 & \xi_0\mathrm{Tr}\left(\mathbf{G}^2\mathbf{C}\left(\mathbf{I} + \gamma_0\xi_0\mathbf{C}^{\mathrm{T}}\right)\widetilde{\mathbf{G}}_0^{\mathrm{T}}\right) & \xi_0\mathrm{Tr}\left(\mathbf{G}^2\mathbf{C}\mathbf{C}^{\mathrm{T}}\right) & \mathrm{Tr}\left(\mathbf{G}^2\mathbf{C}\right) \end{bmatrix}.$$

In (12), matrix $\mathbf{\Pi}_0^*$ results from the exogenous component of the model ($\tilde{\mathbf{x}}$) while $\mathbf{\Pi}_1^*$ follows from the endogenous part (e.g. $\xi\mathbf{C}\boldsymbol{\varepsilon}$). While neither matrix has full rank, one can verify that their sum does so long as $\xi_0 \neq 0$ and the fourth column of $\mathbf{\Pi}_1^*$ is not zero, i.e. condition (iii) of the Theorem holds. Hence, for (12) to equal zero $\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta}_0$ must hold. The proof of Theorem 1 develops this argument for the general case.

This analysis clarifies the role of condition (iii) of the Theorem: it simply requires that the SLE specification per (2) is meaningful, i.e. the K covariates are actually all endogenous. Ultimately, this is a moot requirement: if some covariates are exogenous, but researchers still want to estimate $\boldsymbol{\vartheta}$ in its entirety (for example, because interest falls on elements of $\boldsymbol{\delta}$ for specific endogenous covariates), they can proceed by simply placing appropriate restrictions on $\boldsymbol{\xi}$ in (9). Conditions (i) and (ii) are more standard in models about peer and network effects, and spatial econometrics more generally. The former requires that social and contextual effects do not cancel out for at least one observable characteristic, as otherwise β and $\boldsymbol{\delta}$ cannot be disentangled. The latter requires increasingly higher-order powers of $\mathbf{G}$ to vary in their identification power: in networks this occurs for example when connections are not transitive (Bramoullé et al., 2009). With respect to standard models, however, condition (ii) extends to one additional degree distance (observe how it involves $\mathbf{G}^3$). This is necessary to establish enough linearly independent moment conditions that also identify $\boldsymbol{\xi}$.[23]

More intuition about identification can be obtained in two ways: algebraic-statistic and graphical. We develop the former first. In the same simplified setting we exploited to derive (12), the reduced form for $\mathbf{y}$ can be expressed as:

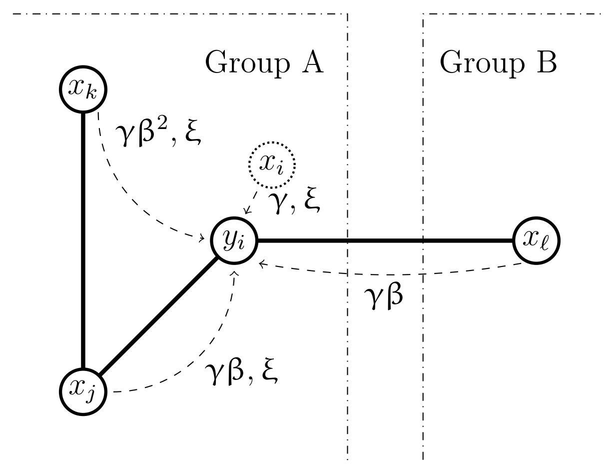
$$\mathbf{y} = \sum_{s=0}^{\infty} \beta^s \mathbf{G}^s \left[\alpha\boldsymbol{\iota} + \gamma\widetilde{\mathbf{x}} + (\mathbf{I} + \gamma\xi\mathbf{C})\,\boldsymbol{\upsilon}\right]. \tag{13}$$

Thus, by an argument *à la* Kelejian and Prucha (1998) the model is identified via a set of instruments of the form $\mathbf{G}^s\widetilde{\mathbf{x}}$, which are unfeasible since $\widetilde{\mathbf{x}}$ is unobserved. Expression (13) also suggests that if $\widetilde{\mathbf{x}}$ and $\mathbf{C}$ are both observed, ξ is identified separately from γ. Expression (7), instrumental to the construction of our linear moments, embed both ideas: it reshapes the "unfeasible" moments (6) so that the independent component of $\mathbf{x}$ is backed up from its constituent parts. This is possible as ξ is internally identified thanks to the knowledge of the characteristics matrix $\mathbf{C}$; again, (13) suggests why this

[23] In light of the discussion from Section 2.2, we observe an analogy between our condition (ii) and identification conditions developed for SEMs with spatial lags, like in Liu and Saraiva (2019).

is the case. In fact, the endogenous component of $\mathbf{x}$ propagates through the structure of social or spatial interactions $\mathbf{G}$, and is therefore reflected at higher-order distances. This further clarifies why $\mathbf{I}$, $\mathbf{G}$, $\mathbf{G}^2$ must be linearly independent of $\mathbf{G}^3$ too.

To appreciate the graphical intuition, consider the four observations (i, j, k, ℓ) involved in both the network and the group structure from Graph 1. According to (13), the variation of y_i is explained by the variation of all the elements in (x_i, x_j, x_k, x_ℓ), albeit through different mechanisms and "effects" (parameters). This is represented in Graph 2, which "zooms in" the four nodes in question and in addition, displays some labeled dashed arrows that indicate 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. Intuitively, this identifies ξ. In addition, the direct effect of x_i on y_i contributes to the identification of both γ and ξ. As a result, a comparison of the "effect" of x_i on y_i against that of x_j on y_i allows, via knowledge of ξ, to disentangle γ from β, as in models with exogenous covariates.



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 $\mathbf{G}$ (represented by circles and straight lines) and a "grouped" characteristics structure $\mathbf{C}$ (delimited by dash-dotted lines). Directed dashed arrows that connect the variables encapsulated in either node are labeled by to 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 (13). 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.

3.4 Extensions

Our framework, and its associated identification results, can be extended in several directions. Here we briefly outline two such extensions, which we find to be especially relevant for some of the network-based applications that inspired our model; for each of them, we express the identification conditions in specific corollaries to Theorem 1. We leave the analysis of other extensions to future work.

Network-level fixed effects. As mentioned in Section 2, in network settings our model can accommodate both the case where adjacency matrix $\mathbf{G}$ represents a large network with a unique connected component (for example, an input-output network of firms) and that where it describes multiple disconnected networks (say, friends from different classes or schools). In the latter case, $\mathbf{G}$ has a block-diagonal structure. With several separate networks, researchers may want to estimate an extended model like:

$$\mathbf{y} = \mathbf{D}\boldsymbol{\alpha}^* + \beta\mathbf{G}\mathbf{y} + \mathbf{X}\boldsymbol{\gamma} + \mathbf{G}\mathbf{X}\boldsymbol{\delta} + \boldsymbol{\varepsilon}, \tag{14}$$

where the $N \times D$ matrix $\mathbf{D}$ collects D dummy variables, each encoding an observation's association with a particular block (network) of $\mathbf{G}$, while $\boldsymbol{\alpha}^* = (\alpha_1, \ldots, \alpha_D)$ are the corresponding "network-level" fixed effects.

Corollary 1. *If the model of interest is (14), the parameters* $\theta^* \equiv (\boldsymbol{\alpha}^*, \beta, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\xi}, \psi)$ *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 (14) 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 $\alpha = 0$; 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. □

With this approach, the error term is transformed as $(\mathbf{I} - \mathbf{G})\,\boldsymbol{\varepsilon}$. Hence, for the sake of identification and estimation the bias-correction terms $\lambda_{1,qk}$ in (9) and $\lambda_{2,p}$ in (10) are transformed accordingly (the calculations are straightforward).

Overall connection strength. Occasionally, researchers may want to estimate the direct effect on $\mathbf{y}$ of a measure that represents the total intensity or "strength" of the network connections directed to each observation, which we denote by $\bar{\mathbf{g}} = \mathbf{G}\iota$ (the *indegree* vector). Thus, the structural equation of interest becomes:

$$\mathbf{y} = \alpha\iota + \beta\mathbf{G}\mathbf{y} + \mathbf{X}\gamma + \mathbf{G}\mathbf{X}\delta + \phi\bar{\mathbf{g}} + \varepsilon, \tag{15}$$

where ϕ is a new parameter of interest.[24] Since, under Assumption 6, $\mathbf{G}$ is exogenous, the extended identification conditions are straightforward in this case.

Corollary 2. *In (15), ϕ is identified separately from θ if $\bar{\mathbf{g}}$ is linearly independent of ι and all columns of $\widetilde{\mathbf{X}}$.*

Proof. Extend (9) with additional moments of the form $\mathbb{E}\left[(\mathbf{G}^{q-1}\bar{\mathbf{g}})^{\mathrm{T}}\varepsilon(\theta,\phi)\right] = \mathbf{0}$, for $q = 1, \ldots, Q$. An revised proof of Theorem 1 holds under this extended setup. □

Hence, identification requires that the network indegree $\bar{\mathbf{g}}$ is neither constant (which rules out "row-normalized" specifications of $\mathbf{G}$ that are especially popular in the peer effects literature) nor perfectly predicted by the exogenous, independent components of $\mathbf{X}$. The evaluation of the latter requirement depends on one's application.

4 Estimation

The moment conditions that support our main identification results lend themselves naturally to GMM estimation. In this section we show how the estimation framework introduced by Lee (2007a) can be adapted to our model with SLE. In what follows, we denote the "true" parameter values as θ_0 and we introduce N subscripts to denote the dependence of a particular algebraic object (random variable, vector or matrix) on the sample size. We make also make an additional assumption.

Assumption 8. Parameter space: $\theta_0 \in \Theta \subset \mathbb{R}^{3(1+K)}$*, that is, θ_0 belongs to the interior of a parameter space denoted as Θ which is compact and convex.*

Assumption 8 is standard. With regard in particular to the two parameters β and ψ, it complements Assumptions 1 and 5.[25]

[24] In Addendum B, we specify a model of strategic interaction that microfounds (15). In this game, players take choices about a costly public good ("effort"). According to the model, the identification of ϕ backs up primitive parameters interpreted as the private and social returns of effort, respectively.

[25] These implicitly constrain β and ψ to lie within an interval that is symmetric around zero and whose length is at most twice the inverse of the spectral radius of $\mathbf{G}$ and $\mathbf{E}$, respectively. As a result, $(\mathbf{I} - \beta\mathbf{G})$ and $(\mathbf{I} + \psi\mathbf{E})$ are both non-singular (see the discussion by Kelejian and Prucha, 2010).

We collect all the moment conditions of the model as:

$$\mathbb{E}\left[\mathbf{m}_N\left(\boldsymbol{\theta}_0\right)\right]=\mathbb{E}\begin{bmatrix}\mathbf{m}_{1,N}\left(\boldsymbol{\theta}_0\right)\\ \mathbf{m}_{2,N}\left(\boldsymbol{\theta}_0\right)\end{bmatrix}=\mathbf{0}. \tag{16}$$

where $\mathbf{m}_{1,N}\left(\boldsymbol{\theta}\right)$ and $\mathbf{m}_{2,N}\left(\boldsymbol{\theta}\right)$ denote respectively the vectors of "linear" and quadratic moments (inclusive of the bias-correction terms) as in (9) and (10).[26] The construction of appropriate sample analogs of (16) is hindered by the fact that the bias-correction terms are functions of the unknown matrix $\boldsymbol{\Sigma}_N$. We circumvent this issue by replacing those terms with appropriate consistent estimators of them. Thus, for a given $\boldsymbol{\theta}\in\boldsymbol{\Theta}$, we construct the following sample moments:

$$\overline{\mathbf{m}}_N\left(\boldsymbol{\theta}\right)\equiv\frac{1}{N}\begin{bmatrix}\overline{\mathbf{m}}_{1,N}\left(\boldsymbol{\theta}\right)\\ \overline{\mathbf{m}}_{2,N}\left(\boldsymbol{\theta}\right)\end{bmatrix}. \tag{17}$$

Here, the linear moments are $\overline{\mathbf{m}}^{\mathrm{T}}_{1,N}\left(\boldsymbol{\theta}\right)=\left[m_{1,0,N}\left(\boldsymbol{\theta}\right)\;\;\overline{\mathbf{m}}^{\mathrm{T}}_{1,1,N}\left(\boldsymbol{\theta}\right)\;\;\dots\;\;\overline{\mathbf{m}}^{\mathrm{T}}_{1,Q,N}\left(\boldsymbol{\theta}\right)\right]$ with $\overline{\mathbf{m}}^{\mathrm{T}}_{1,q,N}\left(\boldsymbol{\theta}\right)=\mathbf{Q}_{q,N}\boldsymbol{\varepsilon}_N\left(\boldsymbol{\theta}\right)-\overline{\boldsymbol{\lambda}}_{1,q,N}\left(\boldsymbol{\theta}\right)$ for $q=1,\dots,Q$, and where $\overline{\boldsymbol{\lambda}}_{1,q,N}\left(\boldsymbol{\theta}\right)$ is a vector of dimension K whose elements are expressed, for $k=1,\dots,K$, as:

$$\overline{\lambda}_{1,q,k,N}\left(\boldsymbol{\theta}\right)=\xi_k\mathrm{Tr}\left(\overline{\boldsymbol{\Upsilon}}_N\left(\boldsymbol{\theta}\right)\mathbf{G}_N^{q-1}\mathbf{C}_{k,N}\right).$$

The above is unlike (8) because $\overline{\boldsymbol{\Upsilon}}_N\left(\boldsymbol{\theta}\right)$ here features no expectation:

$$\overline{\boldsymbol{\Upsilon}}_N\left(\boldsymbol{\theta}\right)=\left(\mathbf{I}_N+\psi\mathbf{E}_N\right)\mathrm{diag}\left(\upsilon_1^2\left(\boldsymbol{\theta}\right),\dots,\upsilon_N^2\left(\boldsymbol{\theta}\right)\right)\left(\mathbf{I}_N+\psi\mathbf{E}_N\right)^{\mathrm{T}}. \tag{18}$$

Similarly, the quadratic moments are $\overline{\mathbf{m}}^{\mathrm{T}}_{2,N}\left(\boldsymbol{\theta}\right)=\left[\overline{m}_{2,1,N}\left(\boldsymbol{\theta}\right)\;\;\dots\;\;\overline{m}_{2,P,N}\left(\boldsymbol{\theta}\right)\right]$ where, for $p=1,\dots,P$, $\overline{m}_{2,p,N}\left(\boldsymbol{\theta}\right)=\boldsymbol{\varepsilon}^{\mathrm{T}}_N\left(\boldsymbol{\theta}\right)\mathbf{P}_{p,N}\boldsymbol{\varepsilon}_N\left(\boldsymbol{\theta}\right)-\overline{\lambda}_{2,p,N}\left(\boldsymbol{\theta}\right)$ and

$$\overline{\lambda}_{2,p,N}\left(\boldsymbol{\theta}\right)=\mathrm{Tr}\left(\overline{\boldsymbol{\Upsilon}}_N\left(\boldsymbol{\theta}\right)\mathbf{P}_{p,N}\right),$$

which replaces (11). Our GMM estimator is thus:

$$\widehat{\boldsymbol{\theta}}_{GMM}=\underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\arg\min}\,\overline{\mathbf{m}}^{\mathrm{T}}_N\left(\boldsymbol{\theta}\right)\mathbf{W}_N\overline{\mathbf{m}}_N\left(\boldsymbol{\theta}\right), \tag{19}$$

[26] In practical implementations of our GMM estimator (e.g. in the Monte Carlo) we also entertained alternative mathematical formulations of the "linear" moments based on (7); as expected, the results are similar. We noted, however, that the baseline formulation that features the bias-correction terms $\boldsymbol{\lambda}_{1,q}\left(\boldsymbol{\theta}\right)$ is computationally faster, more convenient; we thus maintain it as our favorite.

where $\mathbf{W}_N$ is a suitable weighting matrix that complies with the following assumption, which is key to ensure finite moments of (19).

Assumption 9. Weighting matrix: *$\mathbf{W}_N$ has a probability limit: $\mathbf{W}_N \xrightarrow{p} \mathbf{W}_0$; it can be decomposed as:*

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

where $\mathbf{A}_N$ is a conformable matrix such that $\mathbf{A}_N \xrightarrow{p} \mathbf{A}_0$ with $\mathbf{A}_0^{\mathrm{T}} \mathbf{A}_0 = \mathbf{W}_0$, and with $\operatorname{rank}(\mathbf{A}_N) \geq \dim |\boldsymbol{\theta}|$*. In addition, $\mathbf{A}_N$ is uniformly bounded in absolute value in both row and column sums, and all its elements are also bounded.*

The asymptotic properties of $\widehat{\boldsymbol{\theta}}_{GMM}$ are established via the following result.

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

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

where $\boldsymbol{\Omega}_0 \equiv \operatorname{plim} \mathbb{V}\mathrm{ar}\left[\overline{\mathbf{m}}_N(\boldsymbol{\theta}_0)\right]$ *and* $\mathbf{J}_0 \equiv \operatorname{plim} \mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\theta}^{\mathrm{T}}} \overline{\mathbf{m}}_N(\boldsymbol{\theta}_0)\right]$.

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). □

To perform statistical inference in actual empirical applications, it is necessary to use a consistent estimator of $\boldsymbol{\Omega}_0$: its sample analogue is the most natural choice.[27]

As with analogous models in spatial econometrics, the efficiency of our estimator depends on the choice of moments, more specifically the $\mathbf{Q}_{q,N}$ and $\mathbf{P}_{p,N}$ matrices. In the next section we provide limited evidence, via Monte Carlo simulations, that the number and type of moments can affect the estimator's finite sample performance. Efficiency can be further improved by choosing optimal moment-weighting matrices $\mathbf{Q}_{q,N}$ and $\mathbf{P}_{p,N}$. As in the analysis by Lin and Lee (2010) about the SAR model with unrestricted heteroschedasticity, the optimal instrument matrices in our model would also depend on the unknown matrix $\boldsymbol{\Sigma}_N$. Additional efficiency improvements may be obtained with a two-step procedure (a first step delivers a consistent, but less efficient estimator of $\boldsymbol{\theta}$; a second step updates the moments); however, we reserve the analysis of this and other refinements of our proposed estimator to future work.

[27] Because $\boldsymbol{\Omega}_0$ is a function of fourth-order moments of $\boldsymbol{\upsilon}_N$, its expression and that of its sample analogues are convoluted; for the sake of exposition, they are provided in Addendum C.

5 Monte Carlo

We evaluate the performance of our GMM estimator across Monte Carlo simulations based on a simplified version of our model. Specifically, we allow for two covariates: an exogenous one that we denote by $\mathbf{w}$, and an endogenous one, written as $\mathbf{x}$; we drop the exogenous effects $\boldsymbol{\delta}$, and we enforce homoschedasticity. While in this section we focus on one particular experiment (the "baseline"), we also examine other experiments that differ in some details of the data generation process. To minimize the dependence of our results from a specific topology of the adjacency, or spatial weighting matrix $\mathbf{G}$, in all the simulations or repetitions of an experiment, we generate a new matrix $\mathbf{G}$; in most experiments we do the same for characteristic matrices $\mathbf{C}$. More specifically, these matrices are randomly generated via the "small-world" algorithm by Watts and Strogatz (1998) with constant parameters. In particular, we set the number of links for each simulated observation at $B = 2$, and the link rewiring probability at $b = 0.25$.[28]

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

$$\mathbf{y} = (\mathbf{I} - \beta\mathbf{G})^{-1}\left[\alpha\iota + \gamma\left(\widetilde{\mathbf{x}} + \xi\sigma\mathbf{C}\left(\mathbf{I} + \psi\mathbf{G}\right)\boldsymbol{\upsilon_y}\right) + \chi\mathbf{w} + \sigma\left(\mathbf{I} + \psi\mathbf{G}\right)\boldsymbol{\upsilon_y}\right],$$

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; $\boldsymbol{\upsilon_y}$ is a vector of N independent draws from a standard normal distribution; while $\widetilde{\mathbf{x}}$ is generated as:

$$\widetilde{\mathbf{x}} = 0.3 \cdot \mathbf{H}\boldsymbol{\upsilon_x},$$

where $\boldsymbol{\upsilon_x}$ are yet N more independent draws from the standard normal distribution, and $\mathbf{H}$ is an $N \times N$ matrix. When $\mathbf{H} \neq \mathbf{I}$ the independent component of $\mathbf{x}$ features spatial correlation, which is arguably realistic. While in most simulations reported here we set $\mathbf{H} = \mathbf{I} + \mathbf{G}$ (therefore, $\mathbf{H}$ varies across repetitions), we also experimented

[28]The small-world algorithm is initialized by ordering all observations are first ordered along a line and connecting them 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 b. This procedure yields an updated 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 $\mathbf{G} = \text{diag}\left(\mathbf{G}_1\iota\right)\mathbf{G}_1$. If derived through a distinct random realization of this algorithm, matrix $\mathbf{C}$ is obtained likewise. Our combined choices for B and b ensure a good overlap between the adjacency matrices $\mathbf{G}$ and the characteristic matrix $\mathbf{C}$ across our experiments.

with different choices of $\mathbf{H}$, and observed that these have no substantive bearing on the results. In all our simulations we set $N = 500$ and $\sigma = 0.05$.

In all experiments, we compare nine estimators with one another. Four estimators are variations of our proposed GMM estimator, where moments (17) are constructed using a characteristic matrix, denoted by $\mathbf{C}_e$ which may or may not coincide with that used in the d.g.p.: $\mathbf{C}$. These four variations are summarized as follows; (1) one based on a smaller set of moments, $Q = 3$ and $P = 2$, where $\mathbf{P}_1 = \mathbf{I}$ and $\mathbf{P}_2 = \mathbf{G}$, and $\mathbf{C}_e^* = \mathbf{C}$; (2) as the previous one but with more moments: $Q = 4$ and $P = 3$, where $\mathbf{P}_3 = \mathbf{G}^2$; (3) one with even more moments: $Q = 5$ and $P = 4$, and where $\mathbf{P}_4 = \mathbf{G}^3$; (4) one like the latter (more moments), but where the estimation algorithm employs a "misspecified" characteristic matrix $\mathbf{C}_e^* = \mathbf{C}_e \neq \mathbf{C}$. In particular, in the baseline $\mathbf{C}$ is randomly generated via the small-world algorithm (hence it most likely differs from $\mathbf{I}+\mathbf{G}$), but $\mathbf{C}_e^* = \mathbf{I}+\mathbf{G}$. Hence, through this exercise we assess the implications of an incorrect choice of the characteristic matrix, and in particular, the arguably realistic scenario where the econometrician believes that $\mathbf{C}$ and $\mathbf{G}$ capture the same patterns of spatial correlation, but this is only partly or approximately correct. All four GMM estimators return estimates for $(\alpha, \beta, \gamma, \chi, \xi, \psi)$.

The next four estimators are a naïve OLS estimator which takes $\mathbf{Gy}$, $\mathbf{x}$ and $\mathbf{w}$ as independent variables, and three different 2SLS estimators based on the following set of instruments:

$$\mathbf{Z} \equiv \begin{bmatrix} \iota & \mathbf{w} & \mathbf{z} & \mathbf{Gz} & \mathbf{G}^2\mathbf{z} \end{bmatrix},$$

where $\mathbf{z}$ is either: (a) $\mathbf{z} = \mathbf{x}$, yielding an 2SLS estimator akin to the one proposed by Bramoullé et al. (2009); (b) $\mathbf{z} = \mathbf{Gw}$, yielding a 2SLS estimator solely based on the exogenous regressor and its spatial lags; or (c) $\mathbf{z} = \mathbf{Bx}$, where $\mathbf{B}$ is a matrix such that $\mathbf{BC} = \mathbf{0}$ as per the discussion in Section 3.2, yielding a consistent 2SLS estimator based on transformations of $\mathbf{x}$ that are purged of the endogenous component.[29] With some abuse of terminology, we call the ninth estimator a 3SLS. Inspired by Kelejian and Prucha (2004), we construct another 2SLS estimator based on a Cochrane-Orcutt transformation of our model informed by 2SLS estimates with $\mathbf{z} = \mathbf{Gw}$.[30] Hence, we

[29]In those experiments where the $\mathbf{C}$ matrices are by construction of full rank, we specify $\mathbf{B}$ as the annihilator matrix based on the Moore-Penrose pseudoinverse $\mathbf{C}^+$: $\mathbf{B} = \mathbf{I} - \mathbf{CC}^+$.

[30]We are grateful to an anonymous referee for encouraging us to develop this specific comparison. The model by Kelejian and Prucha (2004) features simultaneous equations, exogenous instruments, and (unlike our model) spatially autoregressive errors. Our Cochrane-Orcutt transformation reflects in particular our SMA errors.

compare our GMM estimator to several simpler alternatives that are likely to occur in the empirical practice. These simpler estimators return estimates for $(\alpha, \beta, \gamma, \chi)$.

Table 1: Monte Carlo Simulations: baseline

Target Parameter	Experiment 1, baseline: $\mathbf{H} = \mathbf{I} + \mathbf{G}$; $\mathbf{C} = \mathbf{I} +$ a different small world; $\mathbf{C}_e^* = \mathbf{I} + \mathbf{G}$								
	GMM1	GMM2	GMM3	GMM4	OLS	2SLSa	2SLSb	2SLSc	3SLS
$\alpha = 0.25$	0.248	0.246	0.246	0.287	0.239	0.180	0.251	0.191	0.249
	(0.027)	(0.029)	(0.031)	(0.026)	(0.008)	(0.034)	(0.028)	(0.527)	(0.160)
	[0.063]	[0.053]	[0.051]	[0.027]	[0.009]	[0.029]	[0.055]	[1.961]	[0.533]
	{0.943}	{0.937}	{0.913}	{0.629}	{0.000}	{0.015}	{0.055}	{0.848}	{0.248}
$\beta = 0.40$	0.402	0.403	0.404	0.369	0.410	0.461	0.399	0.453	0.399
	(0.022)	(0.024)	(0.025)	(0.021)	(0.007)	(0.030)	(0.019)	(0.469)	(0.117)
	[0.053]	[0.044]	[0.042]	[0.023]	[0.007]	[0.025]	[0.036]	[1.835]	[0.402]
	{0.944}	{0.943}	{0.918}	{0.616}	{0.000}	{0.000}	{0.023}	{0.571}	{0.155}
$\gamma = 0.50$	0.200	0.200	0.200	0.244	0.280	0.268	0.279	0.227	0.349
	(0.014)	(0.014)	(0.014)	(0.012)	(0.005)	(0.008)	(0.255)	(0.446)	(1.812)
	[0.011]	[0.010]	[0.010]	[0.007]	[0.005]	[0.008]	[0.618]	[2.934]	[5.851]
	{0.804}	{0.799}	{0.796}	{0.014}	{0.000}	{0.000}	{0.517}	{0.877}	{0.694}
$\chi = 1.00$	1.000	0.999	0.999	1.004	0.998	0.987	1.000	0.988	1.000
	(0.008)	(0.008)	(0.008)	(0.007)	(0.007)	(0.010)	(0.014)	(0.120)	(0.043)
	[0.007]	[0.007]	[0.007]	[0.005]	[0.006]	[0.009]	[0.029]	[0.422]	[0.243]
	{0.826}	{0.818}	{0.801}	{0.720}	{0.000}	{0.000}	{0.002}	{0.061}	{0.072}
$\xi = 10.0$	9.839	9.816	9.813	6.619	–	–	–	–	–
	(0.841)	(0.782)	(0.868)	(0.822)					
	[0.672]	[0.633]	[0.629]	[0.519]					
	{0.799}	{0.834}	{0.776}	{0.008}					
$\psi = 0.25$	0.240	0.238	0.233	0.158	–	–	–	–	–
	(0.066)	(0.067)	(0.074)	(0.092)					
	[0.124]	[0.105]	[0.100]	[0.063]					
	{0.950}	{0.951}	{0.927}	{0.586}					

Notes: This table summarizes the results of the "baseline" Monte Carlo experiment. For each estimator-parameter combination, this table reports: (i) the median estimate, (ii) the standard deviation of the estimates (in parentheses), (iii) the average standard error (in square brackets), and (iv) the proportion of non-rejected "true" null hypotheses for that parameter (in curly brackets), across 1,000 repetitions with pseudo-sample size $N = 500$. The data generation process and estimators are based on the $(\mathbf{H}, \mathbf{C}, \mathbf{C}_e^*)$ matrices as summarized in the table's header. The estimators are as follows; GMM1: $Q = 3$, $P = 2$, $\mathbf{C}_e = \mathbf{C}$; GMM2: $Q = 4$, $P = 3$, $\mathbf{C}_e = \mathbf{C}$; GMM3: $Q = 5$, $P = 4$, $\mathbf{C}_e = \mathbf{C}$; GMM4: $Q = 5$, $P = 4$, $\mathbf{C}_e = \mathbf{C}_e^*$; OLS is self-explanatory; 2SLSa: $\mathbf{z} = \mathbf{x}$; 2SLSb: $\mathbf{z} = \mathbf{Gw}$; 2SLSc: $\mathbf{z} = \mathbf{Bx}$; 3SLS: $\mathbf{z} = \mathbf{Gw}$ with ensuing Cochrane-Orcutt transformation and re-estimation. See the text for further details. All other experiments are summarized in tables, also reported in Addendum D, that follow the same structure as this one.

We summarize the results of our baseline simulations in Table 1. For each combination of estimator and parameter, we report the median and the standard deviation of point estimates for the estimated parameters of interest across 1,000 repetitions. In addition, we report the average standard errors and the proportion of non-rejected "true" null hypotheses for that particular parameter. All GMM estimators based on a correctly specified matrix $\mathbf{C}$ (that is, "GMM1," "GMM2" and "GMM3") display a good performance, as expected, at estimating the true parameters of the d.g.p. (which are reported in the table). The number of moments does not appear too consequential. It is interesting to examine the estimates based on a misspecified matrix $\mathbf{C}_e$ ("GMM4"): this introduces a bias, but not a particularly pronounced one for the key parameters of interests β and γ (unlike the endogeneity parameter ξ, for which the bias is more pronounced). More conventional estimators typically deliver biased estimates for β, γ, or for both that are comparable to those from misspecified GMM. The 2SLS and 3SLS estimators based on the exogenous regressor $\mathbf{w}$ perform somewhat better, but still not perfectly, which we interpret as an instance of weak identification.[31] For the third 2SLS estimator (where $\mathbf{z} = \mathbf{B}\mathbf{x}$), the bias appears coupled with an exceedingly large variability of the estimates: as hinted in Section 3.2, 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.

We also performed other experiments. We are especially interested in cases where the characteristic matrices are based on segregated groups, as in one of our examples: the reason is that, as discussed in section 3.1, identification is attainable via a simpler approach. Table 2 summarizes the outcome of two experiment variations where $\mathbf{C}$ is not generated via small-world draws separate from $\mathbf{G}$'s, but has a pre-defined block-diagonal structure. The results are qualitatively unchanged from the baseline of Table 1; it is especially worth noting that not even in this case do the "2SLSc" estimates, which set $\mathbf{z} = \mathbf{B}\mathbf{x}$, appear to outperform our GMM approach with misspecified $\mathbf{C}_e$. Other results, not reported here for brevity, are also analogous.[32] To summarize, the GMM estimator that we propose appears to perform well, and despite its demanding requirements (knowledge of the true characteristic matrix $\mathbf{C}$), departures from the ideal scenario, like the instances of misspecification we examined, do not appear worse than using more conventional estimators. Furthermore, transformations of the data that purge the endogenous component of $\mathbf{x}$ do not seem to be a viable alternative. All these considerations, combined, make a case in favor of using our proposed estimator in the applied econometric practice.

[31] In general β seldom displays much of a bias; this is likely due to the inclusion of $\mathbf{w}$ in our d.g.p. and moments. In unreported experiments where $\mathbf{w}$ is omitted, the bias of β is typically larger.

[32] Some of these results, which are informed by perturbations of the parameter values in the d.g.p., the small-world algorithm, or by other $\mathbf{H}$, $\mathbf{C}$ and $\mathbf{C}_e^*$ matrices, are reviewed in Addendum D.

Table 2: Monte Carlo Simulations: group-based characteristic matrices

Target Parameter	Experiment 2: $\mathbf{H} = \mathbf{I} + \mathbf{G}$; $\mathbf{C}$: groups of size 10; $\mathbf{C}_e^*$: $\mathbf{C}$'s groups, evenly split								
	GMM1	GMM2	GMM3	GMM4	OLS	2SLSa	2SLSb	2SLSc	3SLS
$\alpha = 0.25$	0.250 (0.024) [0.030] {0.881}	0.251 (0.025) [0.023] {0.845}	0.250 (0.026) [0.024] {0.837}	0.228 (0.028) [0.030] {0.808}	0.229 (0.010) [0.010] {0.000}	0.214 (0.025) [0.026] {0.000}	0.251 (0.045) [0.085] {0.065}	0.221 (0.029) [0.030] {0.001}	0.341 (2.932) [4.557] {0.299}
$\beta = 0.40$	0.400 (0.020) [0.025] {0.884}	0.400 (0.020) [0.019] {0.855}	0.400 (0.021) [0.020] {0.846}	0.418 (0.023) [0.026] {0.824}	0.418 (0.008) [0.008] {0.000}	0.432 (0.022) [0.022] {0.000}	0.399 (0.028) [0.055] {0.026}	0.425 (0.026) [0.026] {0.000}	0.355 (1.450) [1.968] {0.185}
$\gamma = 0.50$	0.503 (0.022) [0.020] {0.838}	0.503 (0.022) [0.015] {0.784}	0.504 (0.021) [0.015] {0.797}	0.495 (0.027) [0.038] {0.914}	0.541 (0.010) [0.010] {0.000}	0.533 (0.016) [0.016] {0.000}	0.543 (0.424) [0.951] {0.465}	0.488 (0.020) [0.019] {0.000}	–0.799 (39.491) [96.26] {0.711}
$\chi = 1.00$	1.000 (0.008) [0.005] {0.751}	1.000 (0.008) [0.005] {0.731}	1.000 (0.008) [0.005] {0.732}	0.998 (0.008) [0.006] {0.753}	0.996 (0.008) [0.008] {0.000}	0.993 (0.009) [0.009] {0.000}	1.000 (0.014) [0.025] {0.004}	0.994 (0.010) [0.010] {0.000}	1.016 (0.443) [1.552] {0.079}
$\xi = 10.0$	9.063 (3.504) [4.510] {0.921}	9.047 (3.379) [2.912] {0.885}	8.947 (3.334) [3.087] {0.890}	18.141 (9.412) [15.48] {0.782}	–	–	–	–	–
$\psi = 0.25$	0.248 (0.047) [0.060] {0.914}	0.249 (0.047) [0.047] {0.888}	0.248 (0.049) [0.048] {0.882}	0.223 (0.051) [0.074] {0.927}	–	–	–	–	–

Target Parameter	Experiment 3: $\mathbf{H} = \mathbf{C}$: groups of size 10; $\mathbf{C}_e^*$: $\mathbf{C}$'s ($\mathbf{H}$'s) groups, evenly split								
	GMM1	GMM2	GMM3	GMM4	OLS	2SLSa	2SLSb	2SLSc	3SLS
$\alpha = 0.25$	0.249 (0.026) [0.029] {0.878}	0.249 (0.027) [0.026] {0.860}	0.249 (0.026) [0.026] {0.878}	0.224 (0.030) [0.029] {0.773}	0.231 (0.010) [0.010] {0.000}	0.216 (0.030) [0.030] {0.003}	0.251 (0.028) [0.057] {0.061}	0.212 (0.033) [0.032] {0.003}	0.248 (0.088) [0.490] {0.355}
$\beta = 0.40$	0.401 (0.021) [0.024] {0.886}	0.401 (0.022) [0.022] {0.866}	0.401 (0.021) [0.022] {0.880}	0.422 (0.025) [0.025] {0.787}	0.416 (0.008) [0.008] {0.000}	0.430 (0.026) [0.026] {0.000}	0.399 (0.019) [0.043] {0.018}	0.434 (0.029) [0.028] {0.000}	0.402 (0.058) [0.314] {0.213}
$\gamma = 0.50$	0.503 (0.029) [0.024] {0.818}	0.504 (0.030) [0.021] {0.799}	0.505 (0.029) [0.021] {0.805}	0.491 (0.032) [0.035] {0.895}	0.568 (0.008) [0.012] {0.000}	0.559 (0.018) [0.020] {0.000}	0.582 (0.484) [1.197] {0.521}	0.557 (0.024) [0.027] {0.000}	0.680 (1.659) [10.73] {0.787}
$\chi = 1.00$	1.000 (0.008) [0.005] {0.754}	1.000 (0.008) [0.005] {0.742}	1.000 (0.008) [0.005] {0.741}	0.998 (0.008) [0.006] {0.774}	0.996 (0.007) [0.008] {0.000}	0.993 (0.009) [0.010] {0.000}	0.999 (0.014) [0.033] {0.003}	0.993 (0.010) [0.010] {0.000}	1.000 (0.034) [0.242] {0.085}
$\xi = 10.0$	9.116 (2.717) [2.830] {0.870}	9.071 (2.866) [2.288] {0.850}	8.914 (2.855) [2.333] {0.850}	18.894 (6.145) [7.580] {0.619}	–	–	–	–	–
$\psi = 0.25$	0.246 (0.048) [0.057] {0.903}	0.247 (0.048) [0.050] {0.904}	0.244 (0.050) [0.050] {0.898}	0.223 (0.051) [0.063] {0.908}	–	–	–	–	–

See the notes accompanying Table 1 for a description of this table's structure.

6 Empirical Application

6.1 Motivation

To illustrate how our proposed method can help account for correlated effects in an actual empirical study about social effects, we leverage both the setting and data from the influential paper by De Giorgi et al. (2010). The original aim of their study is the search for peer effects in major choice in higher education settings. While we revisit this research question, we also address a more conventional one, which is about peer effects on academic performance, as measured in terms of final grades or GPA. The setting studied by De Giorgi et al. (2010) is, in fact, suitable to tackle both questions. More specifically, De Giorgi et al. (2010) examine data about students who started their undergraduate studies at Bocconi University, a relevant Italian business school which also offers university degrees in Economics and other social sciences, in 1998. Because Bocconi University attracts highly skilled students from Italy and elsewhere, uncovering peer effects on academic performance would reveal a "social multiplier" that further enhances the value of degrees like Bocconi's.

An attractive feature of the setting originally examined by De Giorgi et al. (2010) is that peer groups are shaped according to a non-overlapping, networked structure of social interactions $\mathbf{G}$ that is determined exogenously. Specifically, students from different undergraduate programs at Bocconi University take common foundational courses over their first year and a half of studies; to reduce class size, the university organizes multiple, parallel versions of each common course; freshmen are randomly allocated into them. In the original paper, the authors defined two students as "peers" if they had been classmates in a given number of common courses out of seven, with the motivation that the bonds established by students over their first three semesters of undergraduate studies would affect their later choices about majors.[33] We refer to the original paper for a full-fledged description of the setting and data.

6.2 Specification

We estimate an augmented version of model (1) on the data provided by De Giorgi et al. (2010), using the same (row-normalized) adjacency matrix $\mathbf{G}$ from their favorite

[33] There were in total nine common courses, of which two were in legal subjects and were excluded by De Giorgi et al. (2010). The two law classes had low attendance rates and thus, a lower number of parallel sessions; consequently, including them would inappropriately inflate a student's peer count.

specification of the network structure, where two students are defined as "peers" if they attended together at least four common courses. However, our revisited analysis differs from the original in two main respects. First, we examine two, rather than one outcomes of interest. In the original paper, the dependent variable is a dummy that denotes major choice (Economics vs. Business Administration): hence, it contradicts the assumptions about the error term maintained in our model. Here, we largely focus on a different, yet *per se* interesting outcome variable that we write as $y_i^{(1)}$, measured on a more continuous scale: the later Bocconi GPA[34] that excludes the initial common courses. For the sake of comparison, we also report results that use the original binary outcome, that we write as $y_i^{(2)}$. Second, we leverage a specific right-hand side variable x_i to construct identifying moment conditions for different estimators: i.e. the grade received by students in high school final exams.[35] This variable has strong predictive power towards both outcomes y_i, but we suspect it to be endogenous. Third, in most specifications we omit contextual or "exogenous" effects, as we find that they typically lead to noisier estimates that complicate comparisons across methods.

Our econometric specification is summarized as follows, for $o = 1, 2$:

$$y_i^{(o)} = \beta \sum_{j=1}^{N} g_{ij} y_i^{(o)} + \gamma x_i + \delta \sum_{j=1}^{N} g_{ij} x_j + \sum_{k=1}^{K'} \chi_k w_{ki} + \varepsilon_i, \tag{20}$$

though in most cases we impose the restriction $\delta = 0$. The K' right-hand side variables w_{ki} in (20) are additional controls that largely overlap with those in the original study: dummies about gender, residence status in Milan, a student's region of origin, type of high school degree (technical school versus an academic-oriented *liceo*) and a student's household income being classified in the top bracket. Among all these controls, we pay special attention to the female dummy; we denote the associated parameter by χ_{fe}. The original study included some additional variables: more specifically, the logarithm of household income and the Bocconi admission test score. We treated the latter both as candidates for our x_i predictor; just like our chosen x_i (the high-school grade) they

[34]In Italian universities like Bocconi, grades are awarded over a scale of 30 points, with 18 being the passing grade. A GPA in Italy is a weighted average of all exam grades, with weights measuring the relative hours load of each course.

[35]In Italy, completion of high school is conditional upon passing a centrally-managed nationwide exam (which differs by type of high school, e.g. technical schools or academic-oriented *licei*) grades in this exam are awarded over a scale of 100 points, with 60 being the passing grade. In the data x_i is rescaled on a zero-to-one measure.

are likely to be endogenous. While experimenting with our proposed GMM approach, however, we found that both candidates typically lead to noisier estimates across all estimators. Since we focused on approaches to address the endogeneity of our favorite predictor x_i, we chose for the sake of consistency to omit other potentially endogenous regressors from the right-hand side of (20) across all specifications we discuss next.[36] We provide some key summary statistics in Table 3; we refer to the original study for more extensive data description and additional statistics.

Table 3: Main variables of interest: summary statistics

	$\mathbf{y}^{(1)}$	$\mathbf{G}\mathbf{y}^{(1)}$	$\mathbf{y}^{(2)}$	$\mathbf{G}\mathbf{y}^{(2)}$	$\mathbf{x}$	$\mathbf{G}\mathbf{x}$	$\mathbf{w}_{fe}$
Mean	26.752	26.755	0.127	0.129	0.863	0.864	0.396
(St. dev.)	(2.049)	(0.522)	(0.333)	(0.088)	(0.112)	(0.027)	(0.489)

Notes. This table reports the mean and the standard deviation of key variables, denoted in the column headers by their corresponding compact notation (e.g., $\mathbf{x}$ is the vector of x_i observations; $\mathbf{w}_{fe}$ is the vector of female dummies). Across all calculations the sample size equals $N = 1,141$. St. dev.: standard deviation.

While we believe that our chosen x_i variable is representative of a student's prior educational achievements or background, as hinted we suspect it to be endogenous. In fact, it is likely to depend upon the unobserved individual ability or motivation, as encoded in the error term ε_i, that also affect the outcomes y_i. This would not affect the identification of social effects if such unobserved components were independent across students. However, there are reasons to suspect the existence of a spatial correlation between the error terms of different students which occurs along geographical lines. Note that Bocconi is a prestigious university within Italy, certainly not a cheap one to attend by national standards;[37] 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 relatively more (self-)selected subset of the population of potential students. This may be especially salient for those students coming from those central and southern

[36] The original study also included a significant predictor of major choice: a dummy variable that indicates whether a student declared Economics (instead of Business) as their favorite major before taking the final decision at the end of the initial common courses. This is an obvious instrument for the identification of social effects in our secondary outcome of interest: major choice, and we have no reason to suspect it endogenous. As the objective of our analysis is to showcase our proposed method in a real setting where endogeneity is salient, we chose to omit this variable from the analysis.

[37] We would like to remark that neither of us has ever graduated from or been employed at Bocconi University. One of us briefly attended one of its undergraduate programs before dropping out.

regions of Italy (about one fourth of our sample) with a markedly lower income per capita and higher overall costs for attending Bocconi.

In light of these observations, we model endogeneity as follows:

$$x_i = \widetilde{x}_i + \xi \varepsilon_i \tag{21}$$

$$\varepsilon_i = \sum_{i=1}^{N} \tilde{c}_{ij} \upsilon_i, \tag{22}$$

where: $\tilde{c}_{ii} \in (0, 1]$ for $i = 1, \ldots, N$; $\tilde{c}_{ij} = \tilde{c}_{ji} \in [0, 1)$ for any (i, j) pair with $i \neq j$; and υ_i, for $i = 1, \ldots, N$, is a random shock with finite but otherwise unrestricted variance. This specification departs slightly from our workhorse model described in (1)-(2)-(3) because it imposes the restriction that $\psi = 0$. In fact, it corresponds with a variation of our model where $\mathbf{C} = \mathbf{I}$ and the structure of spatial correlation for both x_i and ε_i is completely characterized by the $\tilde{c}_{ij}$ weights. We took this approach for two reasons: on the one hand, we observed that in this particular setting, estimates of ψ based on our model with $\mathbf{C} = \mathbf{I}$ turn out to be somewhat noisy; on the other hand, we aimed to experiment with structures of spatial correlation that are non-linear in geographical distance. We note that this approach additionally showcases our model's flexibility. Estimation of this modified (or restricted) model proceeds straightforwardly.

We experiment with two types of spatial correlation structures that are motivated upon our concerns about geography-induced self-selection. They are characterized by different choices for the $\widetilde{\mathbf{C}}$ "characteristic matrices" that collect the $\tilde{c}_{ij}$ weights, and such that the variance-covariance matrix of both x_i and ε_i is proportional to $\widetilde{\mathbf{C}}\widetilde{\mathbf{C}}^{\mathrm{T}}$.

1. The first type returns a structure of spatial correlation featuring *distance decay*:

$$\mathbb{C}\mathrm{ov}\left(x_i, x_j\right) \propto \exp\left(-D \cdot d_{ij}\right).$$

   for every pair (i, j), where d_{ij} is the distance between the geographical centroids of two students' *provinces* of origin.[38] For the sake of this paper's application, we illustrate results based on the (simplifying) choice $D = 1$. We thus construct a characteristic matrix consistent with the resulting pattern, and denoted by $\widetilde{\mathbf{C}}_d$, by eigendecomposing the target variance-covariance matrix $\widetilde{\mathbf{C}}_d\widetilde{\mathbf{C}}_d^{\mathrm{T}}$.

[38]Provinces are traditional administrative units of Italy. In 1998 there were 101 *provinces*, grouped in 20 larger *regions*. We set $d_{ij} = 0$ if $i = j$ or the two students hail from the same province.

2. The second type features geographically-informed "segregated groups" as in one of the examples from Section 2.3. In particular, we specify two partitions of the Italian territory between non-overlapping areas, and for every pair (i, j), we set

$$\tilde{c}_{o,ij} = \left|\mathcal{H}_o(i)\right|^{-1} \mathbb{1}\left[j \in \mathcal{H}_o(i)\right]$$

where for $o = 1, 2$, $\mathcal{H}_o(i)$ denotes the *geographical area* that student i belongs to according to one of the two partitions.[39] This yields two characteristic matrices denoted as $\widetilde{\mathbf{C}}_{h1}$ and $\widetilde{\mathbf{C}}_{h2}$, respectively. The first partition is based on the polities that historically existed on the Italian territory before the process of political unification of the Italian peninsula was started in 1859.[40] The second partition instead classifies Italian provinces according to the regional language which is traditionally most widely spoken in the local area.[41] Both definitions correspond to different groupings of Italian provinces, which often transcend the borders of modern regions, that capture similarities in history, subculture and economic structure of different areas.[42] In both cases, we set $\tilde{c}_{o,ii} = 1$ for $i = 1, \ldots, N$.

Some considerations are common across all characteristics matrix that we employ. First, they comply with identification condition (ii) spelled out by Theorem 1. Second, they also comply with Assumption 2, as required for Theorem 2. Third, their entries are fairly comparable in magnitude. Table 4 qualifies these statements quantitatively: it reports, for our three characteristic matrices, the means and the standard deviations of the elements of the diagonal of $\widetilde{\mathbf{C}}\mathbf{G}$, as well as of the entries of either triangle of $\widetilde{\mathbf{C}}\widetilde{\mathbf{C}}^{\mathrm{T}}$. The former, in particular, verifies that condition (ii) of Theorem 1 holds in our setting. While none of the characteristic matrices that we use is likely to capture the true spatial correlation, we expect them all to approximate it to some degree.

[39] In this case, both covariances $\mathbb{C}\mathrm{ov}(\varepsilon_i, \varepsilon_j)$ and $\mathbb{C}\mathrm{ov}(x_i, x_j)$ are decreasing in $\left|\mathcal{H}_o(i)\right|$, as we expect larger groups or areas to be more heterogeneous.

[40] There are a few differences between the $\mathcal{H}_1(i)$ groups we use to construct $\widetilde{\mathbf{C}}_{h1}$ and the 1859 political map of Italy. First, we detach both Sardinia and Sicily from their parent kingdoms ("Sardinia-Piedmont" and "Two Sicilies"). Second, we split Lombardy-Venetia into its constituent parts. Third, we treat the two small historical duchies of Parma-Piacenza and Modena-Reggio as one polity.

[41] Italian traditional regional languages, such as Lombard, Friulian, Neapolitan or Sardinian, are still widely spoken nowadays. Although most of them belong to the Romance linguistic family, they often lack mutual intelligibility, hence their colloquial denomination as "dialects" may be erroneous.

[42] It is important to comment on how we treat the non-Italian students (that amount to less than 2 per cent of the dataset). In constructing $\widetilde{\mathbf{C}}_d$ they are treated as hailing from an additional, very distant "province." In the matrices of the $\widetilde{\mathbf{C}}_{ho}$ kind instead, they are identified as a separate group.

Table 4: Characteristic matrices: summary statistics

	Diagonal of $\widetilde{\mathbf{C}}\mathbf{G}$			Upper/lower triangle of $\widetilde{\mathbf{C}}\widetilde{\mathbf{C}}^{\mathrm{T}}$		
	$\widetilde{\mathbf{C}}_d$	$\widetilde{\mathbf{C}}_{h1}$	$\widetilde{\mathbf{C}}_{h2}$	$\widetilde{\mathbf{C}}_d$	$\widetilde{\mathbf{C}}_{h1}$	$\widetilde{\mathbf{C}}_{h2}$
Mean	0.0001	0.0009	0.0009	0.0006	0.0026	0.0026
(St. dev.)	(0.0078)	(0.0006)	(0.0008)	(0.0002)	(0.0066)	(0.0072)
Obs.	$N = 1,141$			$N(N-1)/2 = 650,370$		

Notes. This table reports, for the three definitions of $\widetilde{\mathbf{C}}$ used in the analysis that are indicated in each column header, the mean and the standard deviation of the N elements of the diagonal of $\widetilde{\mathbf{C}}\mathbf{G}$ (left panel), or of the $N(N-1)/2$ elements of either the lower or the upper triangle, diagonal excluded, of $\widetilde{\mathbf{C}}\widetilde{\mathbf{C}}^{\mathrm{T}}$ (right panel). St. dev.: standard deviation; Obs.: observations.

6.3 Estimates

We thus turn our attention to the empirical results. We begin by reviewing estimates based on conventional IV/2SLS estimators, that are collected in Table 5. Columns (1) through (4) of this table display estimates for our main outcome of interest $y_i^{(1)}$: students' final GPA on graduation. In this case, typical estimates for the social effects β range between 0.13 and 0.57. However, these estimates are statistically significant only in models from columns (1) and (3), obtained via typical instruments based on spatial lags of $\mathbf{x}$, and by dropping the exogenous effect δ. In both cases, the effect is estimated at about $\beta \simeq 0.35$. By contrast, the two estimates of β in columns (2) and (4) are not statistically significant. The former features the exogenous effect δ, which appears to introduce noise. The latter employs (arguably weak) instruments based on the female dummy $\mathbf{w}_{fe}$. Note that the estimates of γ and χ_{fe} are similar across all models. In light of these results, we argue that the most promising route to estimate social effects is through instruments based on the high-school final grades $\mathbf{x}$, provided that any endogeneity issues about this variable are properly addressed. In addition, these preliminary results prompt us to drop δ from our GMM estimates. Note that most parameter estimates about models for the original outcome of interst $y_i^{(2)}$ in De Giorgi et al. (2010): students' major choice, are seldom statistically significant. It is important to remark that our specification differs from the original; we estimate it for comparison's sake. Once again, the specifications with instruments based on spatial lags of $\mathbf{x}$ and without δ, that is (5) and (7), appear more statistically precise (they return statistically significant estimates for γ) while those from columns (6) and (8) feature no statistically significant estimate at all.

Table 5: Empirical estimates: 2SLS

	Outcome variable: $y_i^{(1)}$ (later career GPA)				Outcome variable: $y_i^{(2)}$ (economics major)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
β	0.319**	0.571	0.353**	0.131	0.359	–3.586	0.115	0.634
	(0.136)	(0.878)	(0.139)	(0.360)	(0.448)	(26.90)	(0.410)	(0.496)
γ	11.39***	11.33***	11.37***	10.76**	0.589***	0.648	0.643***	0.114
	(0.525)	(0.551)	(0.522)	(4.151)	(0.096)	(0.401)	(0.095)	(0.861)
δ	–	–2.660	–	–	–	2.595	–	–
		(11.55)				(18.78)		
χ_{fe}	0.234**	0.227**	0.273***	0.262	-0.018	0.003	-0.023	0.005
	(0.101)	(0.101)	(0.101)	(0.235)	(0.020)	(0.140)	(0.020)	(0.049)
$\mathbf{z}_1$	$\mathbf{x}$	$\mathbf{x}$	$\mathbf{x}$	$\mathbf{Gw}_{fe}$	$\mathbf{x}$	$\mathbf{x}$	$\mathbf{x}$	$\mathbf{Gw}_{fe}$
$\mathbf{z}_2$	$\mathbf{Gx}$	$\mathbf{Gx}$	$\mathbf{Gx}$	$\mathbf{G}^2\mathbf{w}_{fe}$	$\mathbf{Gx}$	$\mathbf{Gx}$	$\mathbf{Gx}$	$\mathbf{G}^2\mathbf{w}_{fe}$
$\mathbf{z}_3$	–	$\mathbf{G}^2\mathbf{x}$	–	–	–	$\mathbf{G}^2\mathbf{x}$	–	–
PFE	NO	NO	YES	NO	NO	NO	YES	NO
Obs.	1,141	1,141	1,132	1,141	1,141	1,141	1,132	1,141

Notes. This table reports IV/2SLS estimates of model (20) for either outcome variables of interest, as indicated in the header. Most estimates incorporate the restriction $\delta = 0$ (no exogenous effects) unless they report an estimate for δ. All estimates are based upon orthogonality conditions between the error term and: (i) a constant vector; (ii) the w_{ki} controls; (iii) two or three "instruments" (IVs) $\mathbf{z}_1$, $\mathbf{z}_2$ or $\mathbf{z}_3$ as specified in each column; $\mathbf{z}_3$ only appears in models featuring the exogenous effect. The vector $\mathbf{w}_{fe}$ stacks the "female" dummies. "PFE" denotes "Province Fixed Effects," handled via a preliminary within-transformation. Estimates for parameters other than β, γ, δ, χ_{fe} are omitted. Heteroschedasticity-consistent standard errors are in parentheses. Asterisk series: *, **, and ***; denote statistical significance at the 10, 5 and 1 per cent level, respectively. Obs.: Observations.

Next, we discuss estimates obtained via our proposed GMM approach: they are collected in Table 6. All specifications in this table incorporate the restriction $\delta = 0$: as in estimates based on conventional models, we noted that any attempt to estimate exogenous effects results in overall noisier estimates.[43] We focus on outcome $y_i^{(1)}$ first. Column (1) reports results for the baseline specification of the later career GPA that models spatial endogeneity using a characteristic matrix $\widetilde{\mathbf{C}}_d$ based on geographical spatial decay. The estimate of social effects β is notably smaller than those reported in Table 5 and it is not statistically significant; analogous considerations extend to the estimates of γ and χ_{fe}. The "endogeneity" parameter is estimated $\hat{\xi} \simeq 0.024$ but is not statistically significant either. Columns (2) and (3) report estimates informed by the characteristic matrices $\widetilde{\mathbf{C}}_{h1}$ and $\widetilde{\mathbf{C}}_{h2}$ that are based, respectively, on the historical

[43] In addition, none of these estimates features "province fixed effects." Adding them considerably complicates GMM (numerical) estimation; at the same time, the estimates of Table 5 suggest that they are unlikely to substantially affect the results.

Table 6: Empirical estimates: proposed GMM approach

	Outcome variable: $y_i^{(1)}$ (later career GPA)				Outcome variable: $y_i^{(2)}$ (economics major)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
β	0.096	0.000	0.000	0.076*	0.058	0.058	0.058	0.057
	(0.102)	(0.044)	(0.044)	(0.044)	(0.042)	(0.041)	(0.041)	(0.041)
γ	3.719	11.21***	11.21***	5.544***	0.587	0.587***	0.588***	0.258
	(8.175)	(0.308)	(0.315)	(1.667)	(1.066)	(0.053)	(0.054)	(0.296)
χ_{fe}	0.618	0.237***	0.237***	0.527***	–0.016	–0.016	–0.016	0.000
	(0.428)	(0.051)	(0.051)	(0.111)	(0.062)	(0.010)	(0.010)	(0.027)
ξ	0.024	0.282	0.319	0.020***	0.000	0.835	0.869	0.031
	(0.016)	(0.293)	(0.302)	(0.004)	(0.098)	(0.966)	(1.006)	(0.021)
$\widetilde{\mathbf{C}}$	$\widetilde{\mathbf{C}}_d$	$\widetilde{\mathbf{C}}_{h1}$	$\widetilde{\mathbf{C}}_{h2}$	$\mathbf{I}+\mathbf{G}$	$\widetilde{\mathbf{C}}_d$	$\widetilde{\mathbf{C}}_{h1}$	$\widetilde{\mathbf{C}}_{h2}$	$\mathbf{I}+\mathbf{G}$
Obs.	1,141	1,141	1,141	1,141	1,141	1,141	1,141	1,141

Notes. Each column in this table reports estimates of the model described in (20)-(21)-(22), using the GMM estimator illustrated in Section 4, for both outcome variables as indicated in the header. All estimates incorporate the restriction $\delta = 0$. Each column reports estimates based on a different choice for the characteristic matrix $\widetilde{\mathbf{C}}$ (which collects the $\tilde{c}_{ij}$ weights), as indicated. Point estimates for parameters other than β, γ, χ_{fe} and ξ are omitted. Standard errors are calculated as described in Addendum C; they are reported in parentheses. Asterisk series: *, **, and ***; denote statistical significance at the 10, 5 and 1 per cent level, respectively. Obs.: Observations.

polity of linguistic group to which a student's home province belongs. In these cases, the estimates of γ and χ_{fe} are statistically significant and are in line with those from Table 5. However, β is estimated as a virtual zero and ξ is not statistically significant either. Furthermore, we experiment with a characteristic matrix defined as a function of the peer network: $\widetilde{\mathbf{C}} = \mathbf{I}+\mathbf{G}$: the results are reported in column (4). The resulting estimates for β, γ and χ_{fe} are attenuated relative to those from conventional models, but are statistically significant (for β, only at the 10 per cent level). The endogeneity parameter is instead estimated $\hat{\xi} \simeq 0.02$ and it is significant at the 1 per cent level. In columns (5) through (8) we report model estimates about outcome $y_i^{(2)}$, one for each of the four characteristic matrices used for $y_i^{(1)}$. In general, none of the key parameter estimates are stastically significant, the exception being γ when using characteristic matrices $\mathbf{C}_{h1}$ and $\mathbf{C}_{h2}$: these estimates align with those from Table 5.

We offer the following interpretation of the results for our key outcome of interest: students' final GPA $y_i^{(1)}$. Overall, none of the $\widetilde{\mathbf{C}}$ matrices we construct to model our concern about self-selection of undergraduate Bocconi students along geographical or cultural lines appear to work perfectly: none of them is associated with a statistically significant estimate of ξ. We suspect, however, that the one based on the exponential decay specification: $\widetilde{\mathbf{C}}_d$, is a better approximation.[44] We find it interesting, moreover, that the specification with $\widetilde{\mathbf{C}} = \mathbf{I} + \mathbf{G}$ leads to a statistically significant estimate for ξ. Because peer networks in college cannot predict high-school grades, this is evidently a random occurrence. All characteristic matrices we experimented with are most likely misspecified; yet, as shown via our Monte Carlo simulations, the resulting bias needs not be larger than that of conventional methods. The estimates of β, in particular, appear less economically and stastistically significant than those from Table 5. Similar considerations extend to the binary outcome $y_i^{(2)}$: students' choice of major. We note that, although applied to a regression model that displays low predictive power in the first place, and unlike conventional methods, our approach delivers point estimates of social effects in a neighborhood of $\widehat{\beta} \simeq 0.06$, a figure very close to the main results by De Giorgi et al. (2010).[45]

We draw some more general implications from these results. First, our approach can yield statistically insignificant estimates of social or peer effects when conventional approaches register these as significant. Second, the choice of the characteristic matrix matters. We suggest that when testing whether results about social effects hold under our method, researchers experiment with multiple plausible characteristic matrices (say, as part of robustness checks). Third, observe that no characteristic matrix that we experimented with admits a straightforward transformation of the kind $\mathbf{BC} = \mathbf{0}$, as per the discussion from Section 3.1. Thus, any attempt to differentiate endogeneity out (to later proceed with more conventional methods) must rely on $\mathbf{B}$ transformation matrices based off the Moore-Penrose pseudoinverse of $\mathbf{C}$. In practice, this is likely to result in unrealistic point estimates and large standard errors,[46] making this approach unviable and calling for a full-fledged implementation of the method we propose.

7 Conclusion

This paper shows that one can identify and estimate spillover or social effects within a standard spatial econometric framework, even if the right-hand side characteristics

[44]In a previous version of this paper we reported estimates based on a more restricted version of the model, which enforced homoschedasticity. There, estimates based on the $\widetilde{\mathbf{C}}_d$ matrix delivered statistically significant estimates for ξ.

[45]In this case, enforcing homoschedasticity would yield statistically significant estimates of β and, for some specifications of SLE, of ξ too.

[46]This issue is showcased in Addendum E, which reports estimates based on this alternative "data transformation" approach for this empirical application.

are endogenous, and without resorting to external instruments. The requirements for identification are fairly general: it suffices that the structure of social interactions is exogenous, not fully-overlapping in only a slightly stronger sense relative to standard identification conditions (e.g. Bramoullé et al., 2009), and that the spatial structure of endogeneity (the dependence of individual covariates on other agents' unobservables) is known by the econometrician up to a multiplicative constant. This approach can be applied to studies about spillover effects where the the right-hand side covariates are suspected endogenous and affected by correlated effects. In our empirical application that revisits the study by De Giorgi et al. (2010), we show that under different specifications of the spatial structure of endogeneity, our approach can lead to precise zero estimates of the social effects, while conventional methods would 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. To this end, we envision the use 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 structure $\mathbf{G}$, 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). Third, and last, we believe it would be worthwhile to integrate our framework with recent contributions that exploit penalized estimators to recover an unknown network structure (Rose, 2017b; de Paula et al., 2023). We conjecture that approaches of this sort can also serve another aim: to recover (partially) unknown characteristic matrices $\mathbf{C}$, and thus mitigate the main requirement of our method (knowledge of $\mathbf{C}$), so long as the structure of social interactions $\mathbf{G}$ is at least partially known.

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.

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.

Cohen-Cole, Ethan, Xiaodong Liu, and Yves Zenou (2018) "Multivariate choices and identification of social interactions," *Journal of Applied Econometrics*, Vol. 33, No. 2, pp. 165–178.

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 (2023) "Identifying network ties from panel data: theory and an application to tax competition," *Review of Economic Studies*. Forthcoming.

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.

——— (2023) "Simultaneous equations models with higher-order spatial or social network interactions," *Econometric Theory*, Vol. 39, No. 6, pp. 1154–1201.

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

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

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

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.

——— (2004) "Estimation of simultaneous systems of spatially interrelated cross sectional equations," *Journal of Econometrics*, Vol. 118, No. 1-2, pp. 27–50.

——— (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 (2014) "Identification and Efficient Estimation of Simultaneous Equations Network Models," *Journal of Business Economics and Statistics*, Vol. 32, No. 4, pp. 516–536.

——— (2020) "GMM identification and estimation of peer effects in a system of simultaneous equations," *Journal of Spatial Econometrics*, Vol. 1, No. 1. Article 1.

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.

Liu, Xiaodong and Paulo Saraiva (2015) "GMM estimation of SAR models with endogenous regressors," *Regional Science and Urban Economics*, Vol. 55, pp. 68–79.

——— (2019) "GMM estimation of spatial autoregressive models in a system of simultaneous equations with heteroskedasticity," *Econometric Reviews*, Vol. 38, No. 4, pp. 359–385.

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

Marschak, Jacob and William H. Andrews (1944) "Random simultaneous equations and the theory of production," *Econometrica*, Vol. 12, No. 3/4, pp. 143–205.

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.

Yang, Kai and Lung-fei Lee (2017) "Identification and QML estimation of multivariate and simultaneous equations spatialautoregressive models.," *Journal of Econometrics*, Vol. 196, No. 1, pp. 196–214.

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

This proof's strategy follows Lee and Liu (2010): it evaluates the moment conditions at an "impostor" parameter vector $\tilde{\boldsymbol{\theta}} \in \boldsymbol{\Theta}$, showing that, for $Q \geq 4$:

$$\mathbb{E}\left[\mathbf{m}\left(\tilde{\boldsymbol{\theta}}\right)\right] = \mathbf{0}$$

if and only if $\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta}_0$, where $\boldsymbol{\theta}_0 = (\boldsymbol{\vartheta}_0, \boldsymbol{\xi}_0, \psi_0)$ denotes the "true" parameter vector. It is convenient to develop the analysis separately for the linear and quadratic moment. Define some ancillary objects: *i.* $\boldsymbol{\Lambda}(\boldsymbol{\theta})$, a $1 + QK \times K$ matrix (for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$):

$$\boldsymbol{\Lambda}(\boldsymbol{\theta}) \equiv \begin{bmatrix} 0 & 0 & \dots & 0 \\ \mathrm{Tr}\left(\boldsymbol{\Upsilon}(\boldsymbol{\theta})\mathbf{C}_1\right) & 0 & \dots & 0 \\ 0 & \mathrm{Tr}\left(\boldsymbol{\Upsilon}(\boldsymbol{\theta})\mathbf{C}_2\right) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathrm{Tr}\left(\boldsymbol{\Upsilon}(\boldsymbol{\theta})\mathbf{C}_K\right) \\ \mathrm{Tr}\left(\boldsymbol{\Upsilon}(\boldsymbol{\theta})\mathbf{G}\mathbf{C}_1\right) & 0 & \dots & 0 \\ 0 & \mathrm{Tr}\left(\boldsymbol{\Upsilon}(\boldsymbol{\theta})\mathbf{G}\mathbf{C}_2\right) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathrm{Tr}\left(\boldsymbol{\Upsilon}(\boldsymbol{\theta})\mathbf{G}\mathbf{C}_K\right) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Tr}\left(\boldsymbol{\Upsilon}(\boldsymbol{\theta})\mathbf{G}^{Q-1}\mathbf{C}_1\right) & 0 & \dots & 0 \\ 0 & \mathrm{Tr}\left(\boldsymbol{\Upsilon}(\boldsymbol{\theta})\mathbf{G}^{Q-1}\mathbf{C}_2\right) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathrm{Tr}\left(\boldsymbol{\Upsilon}(\boldsymbol{\theta})\mathbf{G}^{Q-1}\mathbf{C}_K\right) \end{bmatrix};$$

ii. $\mathbf{F}_0 \equiv \mathbf{I} + \psi_0\mathbf{E}$ and $\widetilde{\mathbf{F}} \equiv \mathbf{I} + \tilde{\psi}\mathbf{E}$; *iii.* $\boldsymbol{\Xi}_0$, an $N \times N$ matrix defined as follows:

$$\boldsymbol{\Xi}_0 = \left[\xi_{01}\mathbf{C}_1\mathbf{F}_0\boldsymbol{\upsilon} \quad \dots \quad \xi_{0K}\mathbf{C}_K\mathbf{F}_0\boldsymbol{\upsilon}\right];$$

iv. $\mathbf{K}_{\tilde{\boldsymbol{x}}}$ and $\mathbf{K}_{\boldsymbol{\upsilon}}$, two matrices of dimension $N \times 1 + QK$:

$$\mathbf{K}_{\tilde{\boldsymbol{x}}} \equiv \left[\boldsymbol{\iota} \quad \widetilde{\mathbf{X}} \quad \mathbf{G}\widetilde{\mathbf{X}} \quad \dots \quad \mathbf{G}^{q-1}\widetilde{\mathbf{X}}\right],$$
$$\mathbf{K}_{\boldsymbol{\upsilon}} \equiv \left[\mathbf{0} \quad \boldsymbol{\Xi}_0 \quad \mathbf{G}\boldsymbol{\Xi}_0 \quad \dots \quad \mathbf{G}^{q-1}\boldsymbol{\Xi}_0\right];$$

v. $\mathbf{S}_\beta$, an $N \times N$ matrix defined as follows:

$$\mathbf{S}_\beta \equiv \mathbf{G}\left(\mathbf{I} - \beta_0\mathbf{G}\right)^{-1}\left[\mathbf{I} + \sum_{k=1}^{K}\xi_{0k}\left(\gamma_{0k}\mathbf{I} + \delta_{0k}\mathbf{G}\right)\mathbf{C}_k\right],$$

vi. lastly, $\mathbf{S}_{\widetilde{\boldsymbol{x}}}$ and $\mathbf{S}_{\boldsymbol{\upsilon}}$, two matrices of dimension $N \times 2(1+K)$:

$$\begin{aligned}
\mathbf{S}_{\widetilde{\boldsymbol{x}}} &\equiv \begin{bmatrix} \boldsymbol{\iota} & \mathbf{G}\left(\mathbf{I}-\beta_0\mathbf{G}\right)^{-1}\left(\alpha_0\boldsymbol{\iota}+\widetilde{\mathbf{X}}\boldsymbol{\gamma}_0+\mathbf{G}\widetilde{\mathbf{X}}\boldsymbol{\delta}_0\right) & \widetilde{\mathbf{X}} & \mathbf{G}\widetilde{\mathbf{X}} \end{bmatrix}, \\
\mathbf{S}_{\boldsymbol{\upsilon}} &\equiv \begin{bmatrix} \mathbf{0} & \mathbf{S}_{\beta}\mathbf{F}_0\boldsymbol{\upsilon} & \boldsymbol{\Xi}_0 & \mathbf{G}\boldsymbol{\Xi}_0 \end{bmatrix}.
\end{aligned}$$

We begin by evaluating the expectation of the linear moments at $\tilde{\boldsymbol{\theta}}$:

$$\begin{aligned}
\mathbb{E}\left[\mathbf{m}_1\left(\tilde{\boldsymbol{\theta}}\right)\right] &= \mathbb{E}\left[\left(\mathbf{K}_{\widetilde{\boldsymbol{x}}}+\mathbf{K}_{\boldsymbol{\upsilon}}\right)^{\mathrm{T}}\left(\mathbf{S}_{\widetilde{\boldsymbol{x}}}+\mathbf{S}_{\boldsymbol{\upsilon}}\right)\right]\left(\boldsymbol{\vartheta}_0-\tilde{\boldsymbol{\vartheta}}\right)+ \\
&\quad + \mathbb{E}\left[\left(\mathbf{K}_{\widetilde{\boldsymbol{x}}}+\mathbf{K}_{\boldsymbol{\upsilon}}\right)^{\mathrm{T}}\mathbf{F}_0\boldsymbol{\upsilon}\right]-\boldsymbol{\Lambda}\left(\tilde{\boldsymbol{\theta}}\right)\tilde{\boldsymbol{\xi}} \\
&= \mathbb{E}\left[\mathbf{K}_{\widetilde{\boldsymbol{x}}}^{\mathrm{T}}\mathbf{S}_{\widetilde{\boldsymbol{x}}}+\mathbf{K}_{\boldsymbol{\upsilon}}^{\mathrm{T}}\mathbf{S}_{\boldsymbol{\upsilon}}\right]\left(\boldsymbol{\vartheta}_0-\tilde{\boldsymbol{\vartheta}}\right)+\boldsymbol{\Lambda}\left(\boldsymbol{\theta}_0\right)\boldsymbol{\xi}_0-\boldsymbol{\Lambda}\left(\tilde{\boldsymbol{\theta}}\right)\tilde{\boldsymbol{\xi}} \\
&= \mathbb{E}\left[\mathbf{K}_{\widetilde{\boldsymbol{x}}}^{\mathrm{T}}\mathbf{S}_{\widetilde{\boldsymbol{x}}}+\mathbf{K}_{\boldsymbol{\upsilon}}^{\mathrm{T}}\mathbf{S}_{\boldsymbol{\upsilon}}\right]\left(\boldsymbol{\vartheta}_0-\tilde{\boldsymbol{\vartheta}}\right)+\boldsymbol{\Lambda}\left(\tilde{\boldsymbol{\theta}}\right)\left(\boldsymbol{\xi}_0-\tilde{\boldsymbol{\xi}}\right)- \\
&\quad -\left(\boldsymbol{\Lambda}\left(\tilde{\boldsymbol{\theta}}\right)-\boldsymbol{\Lambda}\left(\boldsymbol{\theta}_0\right)\right)\boldsymbol{\xi}_0. \qquad (A.1)
\end{aligned}$$

This follows since:

$$\mathbf{K}_{\widetilde{\boldsymbol{x}}}+\mathbf{K}_{\boldsymbol{\upsilon}} = \begin{bmatrix} \boldsymbol{\iota} & \mathbf{Q}_1 & \ldots & \mathbf{Q}_Q \end{bmatrix}$$

and:

$$\boldsymbol{\varepsilon}\left(\tilde{\boldsymbol{\theta}}\right) = \mathbf{y}-\left(\mathbf{S}_{\widetilde{\boldsymbol{x}}}+\mathbf{S}_{\boldsymbol{\upsilon}}\right)\tilde{\boldsymbol{\vartheta}} = \left(\mathbf{S}_{\widetilde{\boldsymbol{x}}}+\mathbf{S}_{\boldsymbol{\upsilon}}\right)\left(\boldsymbol{\vartheta}_0-\tilde{\boldsymbol{\vartheta}}\right)+\mathbf{F}_0\boldsymbol{\upsilon},$$

$\mathbb{E}\left[\mathbf{K}_{\widetilde{\boldsymbol{x}}}^{\mathrm{T}}\mathbf{F}_0\boldsymbol{\upsilon}\right] = \mathbb{E}\left[\mathbf{K}_{\widetilde{\boldsymbol{x}}}^{\mathrm{T}}\mathbf{S}_{\boldsymbol{\upsilon}}\right] = \mathbb{E}\left[\mathbf{K}_{\boldsymbol{\upsilon}}^{\mathrm{T}}\mathbf{S}_{\widetilde{\boldsymbol{x}}}\right] = \mathbf{0}$, and $\mathbb{E}\left[\mathbf{K}_{\boldsymbol{\upsilon}}^{\mathrm{T}}\mathbf{F}_0\boldsymbol{\upsilon}\right] = \boldsymbol{\Lambda}\left(\boldsymbol{\theta}_0\right)\boldsymbol{\xi}_0$. To dissect the last two terms in the last line of (A.1), note that, for some $N \times N$ matrix $\mathbf{M}$:

$$\mathrm{Tr}\left(\boldsymbol{\Upsilon}\left(\tilde{\boldsymbol{\theta}}\right)\mathbf{M}\right) = \mathrm{Tr}\left(\mathbf{M}\widetilde{\mathbf{F}}\widetilde{\boldsymbol{\Sigma}}\widetilde{\mathbf{F}}^{\mathrm{T}}\right)$$

where, denoting the Hadamard (elementwise) matrix product by '$\circ$':

$$\widetilde{\boldsymbol{\Sigma}} \equiv \mathbb{E}\left[\mathrm{diag}\left(\upsilon_1^2\left(\tilde{\boldsymbol{\theta}}\right),\ldots,\upsilon_N^2\left(\tilde{\boldsymbol{\theta}}\right)\right)\right] = \mathbf{I}\circ\mathbb{E}\left[\boldsymbol{\upsilon}\left(\tilde{\boldsymbol{\theta}}\right)\boldsymbol{\upsilon}^{\mathrm{T}}\left(\tilde{\boldsymbol{\theta}}\right)\right];$$

while:

$$\begin{aligned}
\widetilde{\mathbf{F}}\widetilde{\boldsymbol{\Sigma}}\widetilde{\mathbf{F}}^{\mathrm{T}} &= \left[\mathbf{F}_0-\mathbf{E}\left(\psi_0-\tilde{\psi}\right)\right]\widetilde{\boldsymbol{\Sigma}}\left[\mathbf{F}_0-\mathbf{E}\left(\psi_0-\tilde{\psi}\right)\right]^{\mathrm{T}} \\
&= \mathbf{F}_0\widetilde{\boldsymbol{\Sigma}}\mathbf{F}_0^{\mathrm{T}}-\left[\mathbf{E}\widetilde{\boldsymbol{\Sigma}}\mathbf{F}_0^{\mathrm{T}}+\mathbf{F}_0\widetilde{\boldsymbol{\Sigma}}\mathbf{E}^{\mathrm{T}}\right]\left(\psi_0-\tilde{\psi}\right)+\left[\mathbf{E}\widetilde{\boldsymbol{\Sigma}}\mathbf{E}^{\mathrm{T}}\right]\left(\psi_0-\tilde{\psi}\right)^2.
\end{aligned}$$

Also note that, by Assumption 5, the following derivation yields a convergent matrix series:

$$\widetilde{\mathbf{F}}^{-1}\mathbf{F}_0 = \left(\mathbf{F}_0^{-1}\widetilde{\mathbf{F}}\right)^{-1} = \left[\mathbf{I}-\mathbf{F}_0^{-1}\mathbf{E}\left(\psi_0-\tilde{\psi}\right)\right]^{-1} = \sum_{r=0}^{\infty}\mathbf{F}_0^{-r}\mathbf{E}^r\left(\psi_0-\tilde{\psi}\right)^r. \qquad (A.2)$$

The previous observations are instrumental to further analysis of $\widetilde{\boldsymbol{\Sigma}}$:

$$\begin{aligned}
\widetilde{\boldsymbol{\Sigma}} &= \mathbf{I} \circ \mathbb{E}\left[\widetilde{\mathbf{F}}^{-1} \boldsymbol{\varepsilon}\left(\tilde{\boldsymbol{\theta}}\right) \boldsymbol{\varepsilon}^{\mathrm{T}}\left(\tilde{\boldsymbol{\theta}}\right)\left(\widetilde{\mathbf{F}}^{-1}\right)^{\mathrm{T}}\right] \\
&= \mathbf{I} \circ \mathbb{E}\left[\widetilde{\mathbf{F}}^{-1}\left[\left(\mathbf{S}_{\widetilde{\boldsymbol{x}}}+\mathbf{S}_{\boldsymbol{\upsilon}}\right)\left(\boldsymbol{\vartheta}_0-\tilde{\boldsymbol{\vartheta}}\right)+\mathbf{F}_0 \boldsymbol{\upsilon}\right]\left[\left(\mathbf{S}_{\widetilde{\boldsymbol{x}}}+\mathbf{S}_{\boldsymbol{\upsilon}}\right)\left(\boldsymbol{\vartheta}_0-\tilde{\boldsymbol{\vartheta}}\right)+\mathbf{F}_0 \boldsymbol{\upsilon}\right]^{\mathrm{T}}\left(\widetilde{\mathbf{F}}^{-1}\right)^{\mathrm{T}}\right] \\
&= \mathbf{I} \circ\left[\widetilde{\mathbf{F}}^{-1} \mathbf{F}_0 \mathbf{F}_0^{-1} \mathbf{S}_{\widetilde{\boldsymbol{x}}}\left(\boldsymbol{\vartheta}_0-\tilde{\boldsymbol{\vartheta}}\right)\left(\boldsymbol{\vartheta}_0-\tilde{\boldsymbol{\vartheta}}\right)^{\mathrm{T}} \mathbf{S}_{\widetilde{\boldsymbol{x}}}^{\mathrm{T}}\left(\widetilde{\mathbf{F}}^{-1} \mathbf{F}_0 \mathbf{F}_0^{-1}\right)^{\mathrm{T}}\right] \\
&\quad+\mathbf{I} \circ \mathbb{E}\left[\widetilde{\mathbf{F}}^{-1} \mathbf{F}_0 \mathbf{F}_0^{-1} \mathbf{S}_{\boldsymbol{\upsilon}}\left(\boldsymbol{\vartheta}_0-\tilde{\boldsymbol{\vartheta}}\right)\left(\boldsymbol{\vartheta}_0-\tilde{\boldsymbol{\vartheta}}\right)^{\mathrm{T}} \mathbf{S}_{\boldsymbol{\upsilon}}^{\mathrm{T}}\left(\widetilde{\mathbf{F}}^{-1} \mathbf{F}_0 \mathbf{F}_0^{-1}\right)^{\mathrm{T}}\right] \\
&\quad+\mathbf{I} \circ \mathbb{E}\left[\widetilde{\mathbf{F}}^{-1} \mathbf{F}_0 \mathbf{F}_0^{-1} \mathbf{S}_{\boldsymbol{\upsilon}}\left(\boldsymbol{\vartheta}_0-\tilde{\boldsymbol{\vartheta}}\right) \boldsymbol{\upsilon}^{\mathrm{T}}\left(\widetilde{\mathbf{F}}^{-1} \mathbf{F}_0\right)^{\mathrm{T}}\right] \\
&\quad+\mathbf{I} \circ \mathbb{E}\left[\widetilde{\mathbf{F}}^{-1} \mathbf{F}_0 \boldsymbol{\upsilon}\left(\boldsymbol{\vartheta}_0-\tilde{\boldsymbol{\vartheta}}\right)^{\mathrm{T}} \mathbf{S}_{\boldsymbol{\upsilon}}^{\mathrm{T}}\left(\widetilde{\mathbf{F}}^{-1} \mathbf{F}_0 \mathbf{F}_0^{-1}\right)^{\mathrm{T}}\right] \\
&\quad+\mathbf{I} \circ \mathbb{E}\left[\widetilde{\mathbf{F}}^{-1} \mathbf{F}_0 \boldsymbol{\upsilon} \boldsymbol{\upsilon}^{\mathrm{T}}\left(\widetilde{\mathbf{F}}^{-1} \mathbf{F}_0\right)^{\mathrm{T}}\right].
\end{aligned}$$

By replacing (A.2) in the above, isolating all expectations $\mathbb{E}\left[\boldsymbol{\upsilon}\boldsymbol{\upsilon}^{\mathrm{T}}\right]=\boldsymbol{\Sigma}$, and with additional substitutions and manipulations, one can develop (A.1) into:

$$\begin{aligned}
\mathbb{E}\left[\mathbf{m}_1\left(\tilde{\boldsymbol{\theta}}\right)\right] &= \left[\boldsymbol{\Pi}_0+\boldsymbol{\Pi}_1\right]\left(\boldsymbol{\vartheta}_0-\tilde{\boldsymbol{\vartheta}}\right) \\
&\quad+\sum_{r=0}^{\infty} \sum_{s=0}^{\infty} \sum_{k=1}^{K} \sum_{u=0}^{2}\left(\psi_0-\tilde{\psi}\right)^{r+s+u}\left(\left(\xi_{0k}-\tilde{\xi}_k\right) \boldsymbol{\varphi}_{u, krs}-\xi_{0k} \boldsymbol{\varphi}_{u, krs}^{*}\right) \\
&\quad+\sum_{r=0}^{\infty} \sum_{s=0}^{\infty} \sum_{k=1}^{K} \sum_{u=0}^{2}\left(\psi_0-\tilde{\psi}\right)^{r+s+u}\left(\left(\xi_{0k}-\tilde{\xi}_k\right)-\xi_{0k}\right) \boldsymbol{\Phi}_{u, krs}\left(\boldsymbol{\vartheta}_0-\tilde{\boldsymbol{\vartheta}}\right) \\
&\quad+\sum_{r=0}^{\infty} \sum_{s=0}^{\infty} \sum_{k=1}^{K} \sum_{u=0}^{2}\left(\psi_0-\tilde{\psi}\right)^{r+s+u}\left(\left(\xi_{0k}-\tilde{\xi}_k\right)-\xi_{0k}\right) \cdot \\
&\qquad \cdot \boldsymbol{\Psi}_{u, krs} \operatorname{vec}\left(\left(\boldsymbol{\vartheta}_0-\tilde{\boldsymbol{\vartheta}}\right)\left(\boldsymbol{\vartheta}_0-\tilde{\boldsymbol{\vartheta}}\right)^{\mathrm{T}}\right), \qquad \text{(A.3)}
\end{aligned}$$

where $\boldsymbol{\Pi}_0 \equiv \mathbb{E}\left[\mathbf{K}_{\widetilde{\boldsymbol{x}}}^{\mathrm{T}} \mathbf{S}_{\widetilde{\boldsymbol{x}}}\right]$ and $\boldsymbol{\Pi}_1 \equiv \mathbb{E}\left[\mathbf{K}_{\boldsymbol{\upsilon}}^{\mathrm{T}} \mathbf{S}_{\boldsymbol{\upsilon}}\right]$ are matrices of size $1+QK \times 2(1+K)$, the $\boldsymbol{\varphi}$ vectors have dimension $1+QK$, the $\boldsymbol{\Phi}$ matrices have size $1+QK \times 2(1+K)$, while the $\boldsymbol{\Psi}$ matrices have size $1+QK \times 4(1+K)^2$. The structure of $\boldsymbol{\Pi}_0$ in particular is standard in the analysis of GMM estimators for models with spatially autoregressive terms (e.g. Lee, 2007a; Lee and Liu, 2010; Lin and Lee, 2010), with the difference that here, $\boldsymbol{\Pi}_0$ only features the independent components of $\mathbf{X}$. Instead:

$$\boldsymbol{\Pi}_1=\left[\begin{array}{llllllll}\mathbf{0} & \boldsymbol{\pi}_0 & \boldsymbol{\pi}_1 & \ldots & \boldsymbol{\pi}_K & \boldsymbol{\pi}_{K+1} & \ldots & \boldsymbol{\pi}_{2K}\end{array}\right]$$

where all the constituent vectors have a zero first entry (hence the first row of $\mathbf{\Pi}_1$ is also zero): for $k = 1, \ldots, K$, $\pi_{0,0} = \pi_{0,k} = \pi_{K+k,0} = 0$. The other entries of $\boldsymbol{\pi}_0$ are:

$$\pi_{0,1+(q-1)K+k} = \xi_{0k} \left[\mathrm{vec}\left(\mathbf{C}_k^{\mathrm{T}} \mathbf{G}^{q-1} \mathbf{S}_{\beta}\right)\right]^{\mathrm{T}} \mathrm{vec}\left(\mathbf{F}_0 \mathbf{\Sigma} \mathbf{F}_0^{\mathrm{T}}\right)$$

for $q = 1, \ldots, Q$ and $k = 1, \ldots, K$; as for the other non-zero vectors of $\mathbf{\Pi}_1$:

$$\pi_{oK+h,1+(q-1)K+k} = \xi_{0k}\xi_{0h} \left[\mathrm{vec}\left(\mathbf{C}_k^{\mathrm{T}} \mathbf{G}^{q-1+o} \mathbf{C}_h\right)\right]^{\mathrm{T}} \mathrm{vec}\left(\mathbf{F}_0 \mathbf{\Sigma} \mathbf{F}_0^{\mathrm{T}}\right)$$

for $o = 0, 1$, $h = 1, \ldots, K$, $q = 1, \ldots, Q$ and $k = 1, \ldots, K$. The $\boldsymbol{\varphi}$ vectors as well as all matrices of the $\mathbf{\Phi}$ and $\mathbf{\Psi}$ kind are instead derived from the analysis of $\mathbf{\Lambda}(\boldsymbol{\theta})$; they are examined next.

To proceed, define the following row vectors of dimension N^2:

$$\begin{aligned}
\mathbf{f}_{0,kq} &\equiv \left[\mathrm{vec}\left(\mathbf{F}_0^{\mathrm{T}} \mathbf{G}^{q-1} \mathbf{C}_k \mathbf{F}_0\right)\right]^{\mathrm{T}} \\
\mathbf{f}_{1,kq} &\equiv \left[\mathrm{vec}\left(\mathbf{F}_0^{\mathrm{T}} \mathbf{G}^{q-1} \mathbf{C}_k \mathbf{E} + \mathbf{E}^{\mathrm{T}} \mathbf{G}^{q-1} \mathbf{C}_k \mathbf{F}_0\right)\right]^{\mathrm{T}} \\
\mathbf{f}_{2,kq} &\equiv \left[\mathrm{vec}\left(\mathbf{E}^{\mathrm{T}} \mathbf{G}^{q-1} \mathbf{C}_k \mathbf{E}\right)\right]^{\mathrm{T}}
\end{aligned}$$

for $k = 1, \ldots, K$ and $q = 1, \ldots, Q$. Also define the following vectors, for $k = 1, \ldots, K$:

$$\mathbf{i}_k = \left[\mathbb{1}\left[k=1\right] \quad \mathbb{1}\left[k=2\right] \quad \ldots \quad \mathbb{1}\left[k=K\right]\right]^{\mathrm{T}},$$

whose entries are equal to zero except in the k-th positions, where they are equal to one. Thus, for $u = 0, 1, 2$, $k = 1, \ldots, K$, and $r, s = 0, 1, \ldots$:

$$\boldsymbol{\varphi}_{u,krs} = \begin{bmatrix} \mathbf{0}^{\mathrm{T}} \\ \mathbf{i}_k \otimes \mathbf{f}_{u,k1} \\ \vdots \\ \mathbf{i}_k \otimes \mathbf{f}_{u,kQ} \end{bmatrix} \mathrm{vec}\left(\mathbf{I} \circ \left[\mathbf{F}_0^{-r} \mathbf{E}^r \mathbf{\Sigma} \left(\mathbf{F}_0^{-s} \mathbf{E}^s\right)^{\mathrm{T}}\right]\right),$$

while

$$\boldsymbol{\varphi}^*_{u,krs} = \boldsymbol{\varphi}_{u,krs} - \mathbb{1}\left[u = r = s = 0\right] \begin{bmatrix} \mathbf{0}^{\mathrm{T}} \\ \mathbf{i}_k \otimes \mathbf{f}_{0,k1} \\ \vdots \\ \mathbf{i}_k \otimes \mathbf{f}_{0,kQ} \end{bmatrix} \mathrm{vec}\left(\mathbf{\Sigma}\right).$$

Here, '$\otimes$' denotes the Kronecker product. Hence, the vectors $\boldsymbol{\varphi}^*_{u,krs}$ differ from their $\boldsymbol{\varphi}_{u,krs}$ counterparts only when $u = r = s = 0$, and indeed, $\boldsymbol{\varphi}^*_{0,k00} = \mathbf{0}$ for $k = 1, \ldots, K$. Consequently, all elements on the right-hand side of (A.3) multiply at least once one element of vector $\boldsymbol{\theta}_0 - \tilde{\boldsymbol{\theta}}$ (the system of equations features no constant). Observe that, for $k = 1, \ldots, K$, the difference between $\boldsymbol{\varphi}_{0,k00}$ and $\boldsymbol{\varphi}^*_{0,k00}$ equals the k-th column of matrix $\mathbf{\Lambda}(\boldsymbol{\theta}_0)$; it results from decomposing vector $\mathbf{\Lambda}(\boldsymbol{\theta}_0)\boldsymbol{\xi}_0$ in (A.1).

For $u = 0, 1, 2$, $k = 1, \ldots, K$, and $r, s = 0, 1, \ldots$:

$$\boldsymbol{\Phi}_{u,krs} = \begin{bmatrix} \mathbf{0}^{\mathrm{T}} \\ \mathbf{i}_k \otimes \mathbf{f}_{u,k1} \\ \vdots \\ \mathbf{i}_k \otimes \mathbf{f}_{u,kQ} \end{bmatrix} \begin{bmatrix} \mathbf{0} & \mathbf{s}_{\beta} & \mathbf{s}_{\gamma_1} & \ldots & \mathbf{s}_{\gamma_K} & \mathbf{s}_{\delta_1} & \ldots & \mathbf{s}_{\delta_K} \end{bmatrix}_{rs}$$

where:

$$\mathbf{s}_{\beta,rs} = \mathrm{vec}\left(\mathbf{I} \circ \left[\mathbf{F}_0^{-r}\mathbf{E}^r\mathbf{F}_0^{-1}\mathbf{S}_{\beta}\boldsymbol{\Sigma}\left(\mathbf{F}_0^{-s}\mathbf{E}^s\right)^{\mathrm{T}} + \mathbf{F}_0^{-r}\mathbf{E}^r\boldsymbol{\Sigma}\mathbf{S}_{\beta}^{\mathrm{T}}\left(\mathbf{F}_0^{-s}\mathbf{E}^s\mathbf{F}_0^{-1}\right)^{\mathrm{T}}\right]\right)$$

and:

$$\mathbf{s}_{\gamma_k,rs} = \xi_{0k}\mathrm{vec}\left(\mathbf{I} \circ \left[\mathbf{F}_0^{-r}\mathbf{E}^r\mathbf{F}_0^{-1}\mathbf{C}_k\boldsymbol{\Sigma}\left(\mathbf{F}_0^{-s}\mathbf{E}^s\right)^{\mathrm{T}} + \mathbf{F}_0^{-r}\mathbf{E}^r\boldsymbol{\Sigma}\mathbf{C}_k^{\mathrm{T}}\left(\mathbf{F}_0^{-s}\mathbf{E}^s\mathbf{F}_0^{-1}\right)^{\mathrm{T}}\right]\right)$$
$$\mathbf{s}_{\delta_k,rs} = \xi_{0k}\mathrm{vec}\left(\mathbf{I} \circ \left[\mathbf{F}_0^{-r}\mathbf{E}^r\mathbf{F}_0^{-1}\mathbf{G}\mathbf{C}_k\boldsymbol{\Sigma}\left(\mathbf{F}_0^{-s}\mathbf{E}^s\right)^{\mathrm{T}} + \mathbf{F}_0^{-r}\mathbf{E}^r\boldsymbol{\Sigma}\left(\mathbf{G}\mathbf{C}_k\right)^{\mathrm{T}}\left(\mathbf{F}_0^{-s}\mathbf{E}^s\mathbf{F}_0^{-1}\right)^{\mathrm{T}}\right]\right)$$

for $k = 1, \ldots, K$. The $\boldsymbol{\Psi}$ matrices are more elaborate; it is useful to partition each of them between $2(1+K)$ vertical blocks of size $1 + QK \times 2(1+K)$:

$$\boldsymbol{\Psi}_{u,krs} = \begin{bmatrix} \boldsymbol{\Psi}_{u,krs,1} & \boldsymbol{\Psi}_{u,krs,2} & \boldsymbol{\Psi}_{u,krs,3} & \ldots & \boldsymbol{\Psi}_{u,krs,2(1+K)} \end{bmatrix},$$

and analyze the K "central" blocks $\boldsymbol{\Psi}_{u,krs,2+h}$, for $h = 1, \ldots, K$, first. For $u = 0, 1, 2$, $k = 1, \ldots, K$, and $r, s = 0, 1, \ldots$:

$$\boldsymbol{\Psi}_{u,krs,2+h} = \begin{bmatrix} \mathbf{0}^{\mathrm{T}} \\ \mathbf{i}_k \otimes \mathbf{f}_{u,k1} \\ \vdots \\ \mathbf{i}_k \otimes \mathbf{f}_{u,kQ} \end{bmatrix} \begin{bmatrix} \mathbf{v}_{\alpha,h} & \mathbf{v}_{\beta,h} & \mathbf{v}_{\gamma_1,h} & \ldots & \mathbf{v}_{\gamma_K,h} & \mathbf{v}_{\delta_1,h} & \ldots & \mathbf{v}_{\delta_K,h} \end{bmatrix}_{rs}$$

where:

$$\mathbf{v}_{\alpha,h,rs} = \mathrm{vec}\left(\mathbf{I} \circ \left[\mathbf{F}_0^{-r}\mathbf{E}^r\mathbf{F}_0^{-1}\tilde{\mathbf{x}}_h\boldsymbol{\iota}^{\mathrm{T}}\left(\mathbf{F}_0^{-s}\mathbf{E}^s\mathbf{F}_0^{-1}\right)^{\mathrm{T}}\right]\right)$$

and:

$$\mathbf{v}_{\beta,h,rs} = \mathrm{vec}\left(\mathbf{I} \circ \left[\mathbf{F}_0^{-r}\mathbf{E}^r\mathbf{F}_0^{-1}\left(\tilde{\mathbf{x}}_h\left(\alpha_0\boldsymbol{\iota} + \widetilde{\mathbf{X}}\boldsymbol{\gamma}_0 + \mathbf{G}\widetilde{\mathbf{X}}\boldsymbol{\delta}_0\right)^{\mathrm{T}}\left(\mathbf{I} - \beta_0\mathbf{G}^{\mathrm{T}}\right)^{-1}\mathbf{G}^{\mathrm{T}} + \right.\right.\right.$$
$$\left.\left.\left. + \ \xi_{0,h}\mathbf{C}_h\boldsymbol{\Upsilon}_0\mathbf{S}_{\beta}^{\mathrm{T}}\right)\left(\mathbf{F}_0^{-s}\mathbf{E}^s\mathbf{F}_0^{-1}\right)^{\mathrm{T}}\right]\right),$$

where, using shorthand notation, $\boldsymbol{\Upsilon}_0 = \mathbf{F}_0\boldsymbol{\Sigma}\mathbf{F}_0^{\mathrm{T}}$; and, for $k = 1, \ldots, K$:

$$\mathbf{v}_{\gamma_k,h,rs} = \mathrm{vec}\left(\mathbf{I} \circ \left[\mathbf{F}_0^{-r}\mathbf{E}^r\mathbf{F}_0^{-1}\left[\tilde{\mathbf{x}}_h\tilde{\mathbf{x}}_k^{\mathrm{T}} + \xi_{0,h}\xi_{0,k}\mathbf{C}_h\boldsymbol{\Upsilon}_0\mathbf{C}_k^{\mathrm{T}}\right]\left(\mathbf{F}_0^{-s}\mathbf{E}^s\mathbf{F}_0^{-1}\right)^{\mathrm{T}}\right]\right)$$

$$\mathbf{v}_{\delta_k,h,rs} = \operatorname{vec}\left(\mathbf{I} \circ \left[\mathbf{F}_0^{-r}\mathbf{E}^r\mathbf{F}_0^{-1}\left[\tilde{\mathbf{x}}_h\tilde{\mathbf{x}}_k^{\mathrm{T}}\mathbf{G}^{\mathrm{T}} + \xi_{0,h}\xi_{0,k}\mathbf{C}_h\boldsymbol{\Upsilon}_0\left(\mathbf{G}\mathbf{C}_k^{\mathrm{T}}\right)^{\mathrm{T}}\right]\left(\mathbf{F}_0^{-s}\mathbf{E}^s\mathbf{F}_0^{-1}\right)^{\mathrm{T}}\right]\right).$$

The expressions for the blocks $\boldsymbol{\Psi}_{u,krs,2+K+h}$, for $h = 1, \ldots, K$, are derived similarly: with some abuse of notation (which makes for a more succinct exposition) it is enough to replace $\tilde{\mathbf{x}}_h$ with $\mathbf{G}\tilde{\mathbf{x}}_h$ and $\mathbf{C}_h$ with $\mathbf{G}\mathbf{C}_h$ in the vectors $\mathbf{v}_{\alpha,h,rs}$, $\mathbf{v}_{\beta,h,rs}$, $\mathbf{v}_{\gamma_k,h,rs}$ and $\mathbf{v}_{\delta_k,h,rs}$ above. Blocks $\boldsymbol{\Psi}_{u,krs,1}$ and $\boldsymbol{\Psi}_{u,krs,2}$ are also obtained similarly. In the former case, one replaces $\tilde{\mathbf{x}}_h$ with ι and $\xi_{0h}\mathbf{C}_h$ with an $N \times N$ matrix of zeroes. In the latter case, one replaces $\tilde{\mathbf{x}}_h$ with the second column of $\mathbf{S}_{\tilde{\boldsymbol{x}}}$ and $\xi_{0h}\mathbf{C}_h$ with $\mathbf{S}_\beta$.

Under the maintained assumptions expression (A.3) equals zero only when $\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta}_0$. First, it is easy to verify that since the system features infinitely many powers of $\psi_0 - \tilde{\psi}$ that multiply linearly independent vectors, it must hold that $\psi_0 = \tilde{\psi}$. This simplifies the system to:

$$\begin{aligned}
\mathbb{E}\left[\mathbf{m}_1\left(\tilde{\boldsymbol{\theta}}\right)\right] = &\left[\boldsymbol{\Pi}_0 + \boldsymbol{\Pi}_1 + \sum_{k=1}^{K}\left(\left(\xi_{0k} - \tilde{\xi}_k\right) - \xi_{0k}\right)\boldsymbol{\Phi}_{0,k00}\right]\left(\boldsymbol{\vartheta}_0 - \tilde{\boldsymbol{\vartheta}}\right) \\
&+ \left[\sum_{k=1}^{K}\left(\left(\xi_{0k} - \tilde{\xi}_k\right) - \xi_{0k}\right)\boldsymbol{\Psi}_{0,k00}\right]\operatorname{vec}\left(\left(\boldsymbol{\vartheta}_0 - \tilde{\boldsymbol{\vartheta}}\right)\left(\boldsymbol{\vartheta}_0 - \tilde{\boldsymbol{\vartheta}}\right)^{\mathrm{T}}\right) \\
&+ \left[\boldsymbol{\Lambda}\left(\boldsymbol{\theta}_0\right)\right]\left(\boldsymbol{\xi}_0 - \tilde{\boldsymbol{\xi}}\right) = \mathbf{0}.
\end{aligned}$$

One can verify that for $Q \geq 4$, the row rank of the system of equations is at least as large as the number of parameters $\boldsymbol{\theta}$: $3(1 + K)$, so long as the three conditions from the theorem's statement: (i), (ii) and (iii), hold. Their role is as follows.

(i) Prevents social effects from canceling out in the reduced form of $\mathbf{y}$ as in models with exogenous $\mathbf{X}$; this can deliver a deficient rank for $\mathbf{S}_{\tilde{\boldsymbol{x}}}$, and hence $\boldsymbol{\Pi}_0 + \boldsymbol{\Pi}_1$.

(ii) Guarantees, together with Assumption 3, linear independence of at least $2+3K$ rows of $\boldsymbol{\Pi}_0$ (this requires $Q \geq 4$, hence the condition extending to $\mathbf{G}^3$).

(iii) Ensures full column rank of the $\boldsymbol{\Lambda}(\boldsymbol{\theta})$ matrices, and hence linear independence of all $\boldsymbol{\varphi}_{u,krs}$ vectors, all $\boldsymbol{\Phi}_{u,krs}$ matrices, as well as all $\boldsymbol{\Psi}_{u,krs}$ matrices for different values of $k = 1, \ldots, K$ and given u, r and s.

Note that these are sufficient, not necessary conditions. As in Lee and Liu (2010) and analogous models, quadratic moments can supplement deficient identification.

We thus turn our attention to the quadratic moments: we analyze them summarily, without developing a full solution like (A.3). Let:

$$\begin{aligned}
\mathbb{E}\left[m_{2,p}\left(\tilde{\boldsymbol{\theta}}\right)\right] = \mathbb{E}\Bigg[&\left[\left(\mathbf{S}_{\tilde{\boldsymbol{x}}} + \mathbf{S}_{\boldsymbol{u}}\right)\left(\boldsymbol{\vartheta}_0 - \tilde{\boldsymbol{\vartheta}}\right) + \mathbf{F}_0\boldsymbol{\upsilon}\right]^{\mathrm{T}} \cdot \mathbf{P}_p \cdot \\
&\cdot \left[\left(\mathbf{S}_{\tilde{\boldsymbol{x}}} + \mathbf{S}_{\boldsymbol{u}}\right)\left(\boldsymbol{\vartheta}_0 - \tilde{\boldsymbol{\vartheta}}\right) + \mathbf{F}_0\boldsymbol{\upsilon}\right]\Bigg] - \operatorname{Tr}\left(\mathbf{P}_p\boldsymbol{\Upsilon}\left(\tilde{\boldsymbol{\theta}}\right)\right)
\end{aligned}$$

i.e. the expectation of some generic p-th element of the second block, for $p = 1, \dots, P$, expressed as a function of an impostor structure $\tilde{\boldsymbol{\theta}}$. By developing and manipulating the quadratic form inside the expectation, the above can be reformulated as:

$$\begin{aligned}\mathbb{E}\left[m_{2,p}\left(\tilde{\boldsymbol{\theta}}\right)\right] &= \operatorname{vec}\left[\left(\mathbf{S}_{\tilde{\boldsymbol{x}}}^{\mathrm{T}}\mathbf{P}_p\mathbf{S}_{\tilde{\boldsymbol{x}}} + \mathbb{E}\left[\mathbf{S}_{\boldsymbol{\upsilon}}^{\mathrm{T}}\mathbf{P}_p\mathbf{S}_{\boldsymbol{\upsilon}}\right]\right)\right]^{\mathrm{T}} \operatorname{vec}\left(\left(\boldsymbol{\vartheta}_0 - \tilde{\boldsymbol{\vartheta}}\right)\left(\boldsymbol{\vartheta}_0 - \tilde{\boldsymbol{\vartheta}}\right)^{\mathrm{T}}\right) + \\ &+ \left[\operatorname{vec}\left(\mathbf{P}_p + \mathbf{P}_p^{\mathrm{T}}\right)\right]^{\mathrm{T}} \mathbf{S}_{\boldsymbol{\vartheta}}^{\mathrm{v}}\left(\boldsymbol{\vartheta}_0 - \tilde{\boldsymbol{\vartheta}}\right) + \operatorname{Tr}\left(\mathbf{P}_p\left(\boldsymbol{\Upsilon}\left(\boldsymbol{\theta}_0\right) - \boldsymbol{\Upsilon}\left(\tilde{\boldsymbol{\theta}}\right)\right)\right) \end{aligned} \quad \text{(A.4)}$$

where $\mathbf{S}_{\boldsymbol{\vartheta}}^{\mathrm{v}}$ is a matrix of size $N^2 \times 2(1+K)$ expressed as follows:

$$\mathbf{S}_{\boldsymbol{\vartheta}}^{\mathrm{v}} = \begin{bmatrix}\mathbf{0} & \operatorname{vec}\left(\mathbf{S}_{\beta}\boldsymbol{\Upsilon}_0\right) & \dots & \xi_{0k}\operatorname{vec}\left(\mathbf{C}_k\boldsymbol{\Upsilon}_0\right) & \dots & \dots & \xi_{0k}\operatorname{vec}\left(\mathbf{G}\mathbf{C}_k\boldsymbol{\Upsilon}_0\right) & \dots\end{bmatrix},$$

where the final $2K$ columns are understood to run over $k = 1, \dots, K$ twice. The last term in (A.4):

$$\operatorname{Tr}\left(\mathbf{P}_p\left(\boldsymbol{\Upsilon}\left(\boldsymbol{\theta}_0\right) - \boldsymbol{\Upsilon}\left(\tilde{\boldsymbol{\theta}}\right)\right)\right) = \operatorname{Tr}\left(\mathbf{P}_p\left(\mathbf{F}_0\boldsymbol{\Sigma}\mathbf{F}_0^{\mathrm{T}} - \widetilde{\mathbf{F}}\widetilde{\boldsymbol{\Sigma}}\widetilde{\mathbf{F}}^{\mathrm{T}}\right)\right)$$

can be analyzed similarly as in the case of the first block of linear moments, yielding a system of P higher-order polynomials of $\boldsymbol{\theta}_0 - \tilde{\boldsymbol{\theta}}$. A further inspection of (A.4) reveals that the coefficients of these equations are linearly independent, because the $\{\mathbf{P}_p\}_{p=1}^{P}$ matrices are themselves linearly independent under Assumption 7; consequently, the quadratic moments are zero in expectation if and only if $\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta}_0$.

Proof of Theorem 2

Before proceeding we establish some auxiliary notation. For $k = 1, \dots, K$, let $\mathbf{x}_{k,N}$ be k-th column of $\mathbf{X}_N$ (which is given as $\mathbf{X}_{*,k}$ in the text) and let $\mathbb{E}\left[\mathbf{x}_{k,N}\right] = \mathbb{E}\left[\widetilde{\mathbf{x}}_{k,N}\right]$ be its expected value. Thus, $\mathbb{E}\left[\mathbf{x}_{k,N}\right]$ is the k-th column of $\mathbb{E}\left[\mathbf{X}_N\right]$. Write the unconditional expected value of $\mathbf{y}_N$ as follows:

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

Let $\widetilde{\mathbf{G}}_N(\beta) \equiv \mathbf{G}_N\left(\mathbf{I}_N - \beta\mathbf{G}_N\right)^{-1}$, and define the following vectors:

$$\begin{aligned}\mathbf{d}_N(\boldsymbol{\theta}) &\equiv (\alpha_0 - \alpha)\boldsymbol{\iota}_N + (\beta_0 - \beta)\mathbf{G}_N\mathbb{E}\left[\mathbf{y}_N\right] + \mathbb{E}\left[\mathbf{X}_N\right](\boldsymbol{\gamma}_0 - \boldsymbol{\gamma}) + \mathbf{G}_N\mathbb{E}\left[\mathbf{X}_N\right](\boldsymbol{\delta}_0 - \boldsymbol{\delta}), \\ \mathbf{e}_N(\boldsymbol{\theta}) &\equiv \boldsymbol{\varepsilon}_N + \left(\mathbf{X}_N - \mathbb{E}\left[\mathbf{X}_N\right]\right)(\boldsymbol{\gamma}_0 - \boldsymbol{\gamma}) + \mathbf{G}_N\left(\mathbf{X}_N - \mathbb{E}\left[\mathbf{X}_N\right]\right)(\boldsymbol{\delta}_0 - \boldsymbol{\delta}) + \\ &+ (\beta_0 - \beta)\widetilde{\mathbf{G}}_N(\beta_0)\left[\boldsymbol{\varepsilon}_N + \left(\mathbf{X}_N - \mathbb{E}\left[\mathbf{X}_N\right]\right)\boldsymbol{\gamma}_0 + \mathbf{G}_N\left(\mathbf{X}_N - \mathbb{E}\left[\mathbf{X}_N\right]\right)\boldsymbol{\delta}_0\right].\end{aligned}$$

Observe that $\boldsymbol{\varepsilon}_N(\boldsymbol{\theta}) = \mathbf{d}_N(\boldsymbol{\theta}) + \mathbf{e}_N(\boldsymbol{\theta})$. For $k = 1, \dots, K$, define the following $N \times N$ auxiliary matrices:

$$\boldsymbol{\Gamma}_{k,N}(\boldsymbol{\theta}) \equiv \left[\left(\gamma_{0,k} - \gamma_k\right)\mathbf{I}_N + \left(\delta_{0,k} - \delta_k\right)\mathbf{G}_N + (\beta_0 - \beta)\widetilde{\mathbf{G}}_N(\beta_0)\left(\gamma_{0,k}\mathbf{I}_N + \delta_{0,k}\mathbf{G}_N\right)\right],$$

where $\gamma_{0,k}$ and $\delta_{0,k}$ are the "true" values of γ_k and δ_k, respectively. Define yet another $N \times N$ matrix:

$$\mathbf{\Delta}_N(\theta) \equiv \left[\mathbf{I}_N + (\beta_0 - \beta)\,\widetilde{\mathbf{G}}_N(\beta_0)\right],$$

and note that: $\mathbf{e}_N(\theta) = \sum_{k=1}^{K} \mathbf{\Gamma}_{k,N}(\theta)\left[\mathbf{x}_{k,N} - \mathbb{E}\left[\mathbf{x}_{k,N}\right]\right] + \mathbf{\Delta}_N(\theta)\,\boldsymbol{\varepsilon}_N$. Furthermore, let $\mathbf{\Upsilon}_{0,N} = \mathbf{\Upsilon}_N(\theta_0)$, $\mathbf{F}_{0,N} = \mathbf{I}_N + \psi_0 \mathbf{E}_N$, and $\mathbf{F}_N = \mathbf{I}_N + \psi \mathbf{E}_N$ (the dependence on ψ is implicit in $\mathbf{F}_N$); let $a_{i,j,N}$ be the (i,j)-th element of $\mathbf{A}_N$; let $\mathbf{q}_{q,k,N}$ be the k-th *row* of $\mathbf{Q}_{q,N}$, for $q = 1, \dots, Q$ and $k = 1, \dots, K$; and lastly, write

$$\overline{\boldsymbol{\lambda}}_N(\theta) = \left[0 \quad \overline{\boldsymbol{\lambda}}_{1,1,N}^{\mathrm{T}}(\theta) \quad \dots \quad \overline{\boldsymbol{\lambda}}_{1,Q,N}^{\mathrm{T}}(\theta) \quad \overline{\lambda}_{2,1,N}(\theta) \quad \dots \quad \overline{\lambda}_{2,P,N}(\theta)\right]^{\mathrm{T}},$$

a vector that collects the bias-correction terms of the moments, and whose individual elements are denoted, with shorthand notation, by $\overline{\lambda}_{\ell',N}(\theta)$, for $\ell' = 1, \dots, 1+QK+P$.

This proof focuses on the analysis of the influence function

$$\mathbf{n}_N(\theta) \equiv \frac{1}{N}\mathbf{A}_N \mathbf{m}_N(\theta),$$

whose individual elements, denoted by $n_{\ell,N}(\theta)$ for $\ell = 1, \dots, 1+QK+P$, are:

$$n_{\ell,N}(\theta) = \frac{1}{N}\left[a_{\ell,1,N}\boldsymbol{\iota}_N^{\mathrm{T}}\boldsymbol{\varepsilon}_N(\theta) + \sum_{q=1}^{Q}\sum_{k=1}^{K} a_{\ell,1+(q-1)K+k,N}\mathbf{q}_{q,k,N}\boldsymbol{\varepsilon}_N(\theta) + \right.$$
$$\left. + \sum_{p=1}^{P} a_{\ell,1+QK+p,N}\boldsymbol{\varepsilon}_N^{\mathrm{T}}(\theta)\,\mathbf{P}_{p,N}\boldsymbol{\varepsilon}_N(\theta) - \sum_{\ell'=2}^{1+QK+P} a_{\ell,\ell',N}\overline{\lambda}_{\ell',N}(\theta)\right]. \quad \text{(A.5)}$$

In what follows, we examine the three summations in brackets above. In the process, we repeatedly apply Lemmas A.1, A.2 and A.3 by Lin and Lee (2010). In particular,

$$\frac{1}{N}\boldsymbol{\upsilon}_N^{\mathrm{T}}\mathbf{M}_N\boldsymbol{\upsilon}_N = \frac{1}{N}\mathbb{E}\left[\boldsymbol{\upsilon}_N^{\mathrm{T}}\mathbf{M}_N\boldsymbol{\upsilon}_N\right] + o_P(1) = \frac{1}{N}\mathrm{Tr}\left[\mathbf{M}_N\mathbf{\Sigma}_N\right] + o_P(1)$$

where $\mathbf{M}_N$ is some matrix that depends on the context. The conditions underpinning those lemmas are supported here by Assumptions 1, 2, 3, 4, 5, 7, and 9.

Start from the first summation and note that:

$$\sum_{q=1}^{Q}\sum_{k=1}^{K} a_{\ell,1+(q-1)K+k,N}\mathbf{q}_{qk,N}\boldsymbol{\varepsilon}_N(\theta) =$$
$$= \underbrace{\sum_{q=1}^{Q}\sum_{k=1}^{K} a_{\ell,1+(q-1)K+k,N}\mathbf{q}_{q,k,N}\mathbf{d}_N(\theta)}_{\equiv\, \varpi_{\ell,N}^{*}(\theta)} + \underbrace{\sum_{q=1}^{Q}\sum_{k=1}^{K} a_{\ell,1+(q-1)K+k,N}\mathbf{q}_{q,k,N}\mathbf{e}_N(\theta)}_{\equiv\, \rho_{\ell,N}^{*}(\theta)},$$

where, uniformly in $\boldsymbol{\theta} \in \boldsymbol{\Theta}$:

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

while, given some $K(2+K)$ auxiliary (for the given ℓ and for $k,h = 1,\dots,K$):

$$\begin{aligned}
\widetilde{\boldsymbol{\Delta}}^{*}_{\ell,k,N}(\boldsymbol{\theta}) &\equiv \sum_{q=1}^{Q} a_{\ell,1+(q-1)K+k,N}\left(\mathbf{G}_N^{q-1}\right)^{\mathrm{T}}\boldsymbol{\Delta}_N(\boldsymbol{\theta})\\
\widetilde{\boldsymbol{\Gamma}}^{*}_{\ell,k,N}(\boldsymbol{\theta}) &\equiv \sum_{q=1}^{Q} a_{\ell,1+(q-1)K+k,N}\left(\mathbf{G}_N^{q-1}\right)^{\mathrm{T}}\left[\boldsymbol{\Delta}_N(\boldsymbol{\theta}) - \mathbf{I}_N\right]\\
\widetilde{\boldsymbol{\Gamma}}^{*}_{\ell,k,h,N}(\boldsymbol{\theta}) &\equiv \sum_{q=1}^{Q} a_{\ell,1+(q-1)K+k,N}\left(\mathbf{G}_N^{q-1}\right)^{\mathrm{T}}\boldsymbol{\Gamma}_{h,N}(\boldsymbol{\theta}),
\end{aligned}$$

it is, again uniformly in $\boldsymbol{\theta} \in \boldsymbol{\Theta}$:

$$\begin{aligned}
\frac{1}{N}\rho^{*}_{\ell,N}(\boldsymbol{\theta}) = {}& \frac{1}{N}\sum_{k=1}^{K}\left(\mathbf{x}_{h,N} - \mathbb{E}\left[\mathbf{x}_{k,N}\right]\right)^{\mathrm{T}}\widetilde{\boldsymbol{\Delta}}^{*}_{\ell,k,N}(\boldsymbol{\theta})\,\boldsymbol{\varepsilon}_N\\
&+ \frac{1}{N}\sum_{k=1}^{K}\sum_{h=1}^{K}\left(\mathbf{x}_{k,N} - \mathbb{E}\left[\mathbf{x}_{k,N}\right]\right)^{\mathrm{T}}\widetilde{\boldsymbol{\Gamma}}^{*}_{\ell,k,h,N}(\boldsymbol{\theta})\left(\mathbf{x}_{h,N} - \mathbb{E}\left[\mathbf{x}_{h,N}\right]\right)\\
= {}& \frac{1}{N}\sum_{\ell'=2}^{1+QK} a_{\ell,\ell',N}\lambda_{\ell',N}(\boldsymbol{\theta}_0) + \frac{1}{N}\sum_{k=1}^{K}\xi_{0,k}\mathrm{Tr}\left(\mathbf{C}_{k,N}^{\mathrm{T}}\widetilde{\boldsymbol{\Gamma}}^{*}_{\ell,k,N}(\boldsymbol{\theta})\,\boldsymbol{\Upsilon}_{0,N}\right)\\
&+ \frac{1}{N}\sum_{k=1}^{K}\sum_{h=1}^{K}\xi_{0,k}\xi_{0,h}\mathrm{Tr}\left(\mathbf{C}_{k,N}^{\mathrm{T}}\widetilde{\boldsymbol{\Gamma}}^{*}_{\ell,k,h,N}(\boldsymbol{\theta})\,\mathbf{C}_{h,N}\boldsymbol{\Upsilon}_{0,N}\right) + o_P(1). \quad \text{(A.6)}
\end{aligned}$$

The second summation in (A.5) can instead be decomposed as:

$$\begin{aligned}
\sum_{p=1}^{P} a_{\ell,1+QK+p,N}\boldsymbol{\varepsilon}_N^{\mathrm{T}}(\boldsymbol{\theta})\,\mathbf{P}_{p,N}\boldsymbol{\varepsilon}_N(\boldsymbol{\theta}) = {}& \sum_{p=1}^{P} a_{\ell,1+QK+p,N}\mathbf{d}_N^{\mathrm{T}}(\boldsymbol{\theta})\,\mathbf{P}_{p,N}\mathbf{d}_N(\boldsymbol{\theta}) + \\
+ \underbrace{\sum_{p=1}^{P} a_{\ell,1+QK+p,N}\mathbf{d}_N^{\mathrm{T}}(\boldsymbol{\theta})\left(\mathbf{P}_{p,N} + \mathbf{P}_{p,N}^{\mathrm{T}}\right)\mathbf{e}_N(\boldsymbol{\theta})}_{\equiv\, \boldsymbol{\varpi}^{**}_{\ell,N}(\boldsymbol{\theta})} &+ \underbrace{\sum_{p=1}^{P} a_{\ell,1+QK+p,N}\mathbf{e}_N^{\mathrm{T}}(\boldsymbol{\theta})\,\mathbf{P}_{p,N}\mathbf{e}_N(\boldsymbol{\theta})}_{\equiv\, \rho^{**}_{\ell,N}(\boldsymbol{\theta})},
\end{aligned}$$

where two asterisks differentiate any function of $\boldsymbol{\theta}$ from the corresponding one in the analysis of the first summation. Thus, uniformly in $\boldsymbol{\theta} \in \boldsymbol{\Theta}$:

$$\frac{1}{N}\boldsymbol{\varpi}_{\ell,N}^{**}(\boldsymbol{\theta}) = \frac{1}{N}\sum_{p=1}^{P} a_{\ell,1+QK+p,N}\mathbf{d}_N^{\mathrm{T}}(\boldsymbol{\theta})\left(\mathbf{P}_{p,N} + \mathbf{P}_{p,N}^{\mathrm{T}}\right)\Bigg[\boldsymbol{\Delta}_N(\boldsymbol{\theta})\boldsymbol{\varepsilon}_N + \\ + \sum_{k=1}^{K}\boldsymbol{\Gamma}_{k,N}(\boldsymbol{\theta})\left[(\mathbf{x}_{k,N} - \mathbb{E}[\mathbf{x}_{k,N}]) + \xi_{0,k}\mathbf{C}_{k,N}\boldsymbol{\varepsilon}_N\right]\Bigg] = o_P(1),$$

while, given $2 + K + K^2$ auxiliary matrices (for the given ℓ and for $k, h = 1, \ldots, K$):

$$\widetilde{\boldsymbol{\Delta}}_{\ell,N}^{**}(\boldsymbol{\theta}) \equiv \sum_{p=1}^{P} a_{\ell,1+QK+p,N}\boldsymbol{\Delta}_N(\boldsymbol{\theta})\mathbf{P}_{p,N}\boldsymbol{\Delta}_N(\boldsymbol{\theta})$$

$$\widetilde{\boldsymbol{\Gamma}}_{\ell,N}^{**}(\boldsymbol{\theta}) \equiv \sum_{p=1}^{P} a_{\ell,1+QK+p,N}\left[\boldsymbol{\Delta}_N(\boldsymbol{\theta})\mathbf{P}_{p,N}\boldsymbol{\Delta}_N(\boldsymbol{\theta}) - \mathbf{P}_{p,N}\right]$$

$$\widetilde{\boldsymbol{\Gamma}}_{\ell,k,N}^{**}(\boldsymbol{\theta}) \equiv \sum_{p=1}^{P} a_{\ell,1+QK+p,N}\boldsymbol{\Delta}_N(\boldsymbol{\theta})\left(\mathbf{P}_{p,N} + \mathbf{P}_{p,N}^{\mathrm{T}}\right)\boldsymbol{\Gamma}_{k,N}(\boldsymbol{\theta})$$

$$\widetilde{\boldsymbol{\Gamma}}_{\ell,k,h,N}^{**}(\boldsymbol{\theta}) \equiv \sum_{p=1}^{P} a_{\ell,1+QK+p,N}\boldsymbol{\Gamma}_{k,N}^{\mathrm{T}}(\boldsymbol{\theta})\mathbf{P}_{p,N}\boldsymbol{\Gamma}_{h,N}(\boldsymbol{\theta}),$$

one obtains, uniformly in $\boldsymbol{\theta} \in \boldsymbol{\Theta}$:

$$\begin{aligned}\frac{1}{N}\rho_{\ell,N}^{**}(\boldsymbol{\theta}) &= \frac{1}{N}\boldsymbol{\varepsilon}_N^{\mathrm{T}}\widetilde{\boldsymbol{\Delta}}_{\ell,N}^{**}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_N + \frac{1}{N}\sum_{k=1}^{K}\boldsymbol{\varepsilon}_N^{\mathrm{T}}\widetilde{\boldsymbol{\Gamma}}_{\ell,k,N}^{**}(\boldsymbol{\theta})(\mathbf{x}_{k,N} - \mathbb{E}[\mathbf{x}_{k,N}]) \\ &\quad + \frac{1}{N}\sum_{k=1}^{K}\sum_{h=1}^{K}(\mathbf{x}_{k,N} - \mathbb{E}[\mathbf{x}_{k,N}])^{\mathrm{T}}\widetilde{\boldsymbol{\Gamma}}_{\ell,k,h,N}^{**}(\boldsymbol{\theta})(\mathbf{x}_{h,N} - \mathbb{E}[\mathbf{x}_{h,N}]) \\ &= \frac{1}{N}\sum_{\ell'=2+QK}^{1+QK+P} a_{\ell,\ell',N}\lambda_{\ell',N}(\boldsymbol{\theta}_0) + \frac{1}{N}\mathrm{Tr}\left(\widetilde{\boldsymbol{\Gamma}}_{\ell,N}^{**}(\boldsymbol{\theta})\boldsymbol{\Upsilon}_{0,N}\right) \\ &\quad + \frac{1}{N}\sum_{k=1}^{K}\xi_{0,k}\mathrm{Tr}\left(\widetilde{\boldsymbol{\Gamma}}_{\ell,k,N}^{**}(\boldsymbol{\theta})\mathbf{C}_{k,N}\boldsymbol{\Upsilon}_{0,N}\right) \\ &\quad + \frac{1}{N}\sum_{k=1}^{K}\sum_{h=1}^{K}\xi_{0,k}\xi_{0,h}\mathrm{Tr}\left(\mathbf{C}_{k,N}^{\mathrm{T}}\widetilde{\boldsymbol{\Gamma}}_{\ell,k,h,N}^{**}(\boldsymbol{\theta})\mathbf{C}_{h,N}\boldsymbol{\Upsilon}_{0,N}\right) + o_P(1). \quad (A.7)\end{aligned}$$

Lastly, consider the third summation within (A.5), and express any of its elements as $\overline{\lambda}_{\ell',N}(\boldsymbol{\theta}) = \mathrm{Tr}\left(\mathbf{M}_{\ell',N}\overline{\boldsymbol{\Upsilon}}_N(\boldsymbol{\theta})\right)$, for $\ell' = 2, \ldots, 1 + QK + P$, where $\mathbf{M}_{\ell',N}$ is some matrix that depends on the position of the index ℓ' (for example, if $\ell' = 1 + QK + P$, it is $\mathbf{M}_{\ell',N}=\mathbf{P}_{P,N}$). One can decompose this term further as:

$$\begin{aligned}\overline{\lambda}_{\ell',N}(\boldsymbol{\theta}) &= \operatorname{Tr}\left(\mathbf{M}_{\ell',N}\mathbf{F}_N\left[\operatorname{diag}\left(\upsilon_1^2(\boldsymbol{\theta}),\dots,\upsilon_N^2(\boldsymbol{\theta})\right)\right]\mathbf{F}_N^{\mathrm{T}}\right)\\ &= \left[\operatorname{vec}\left(\mathbf{F}_N^{\mathrm{T}}\mathbf{M}_{\ell',N}\mathbf{F}_N\right)\right]^{\mathrm{T}}\operatorname{vec}\left(\mathbf{I}_N\circ\left[\mathbf{F}_N^{-1}\boldsymbol{\varepsilon}_N(\boldsymbol{\theta})\boldsymbol{\varepsilon}_N^{\mathrm{T}}(\boldsymbol{\theta})\left(\mathbf{F}_N^{-1}\right)^{\mathrm{T}}\right]\right)\end{aligned}$$

where the second vector in the second line only has N non-zero entries, and each of these is a second-degree polynomial of the elements of $\boldsymbol{\varepsilon}_N(\boldsymbol{\theta})$. Hence, one can write:

$$\begin{aligned}\overline{\lambda}_{\ell',N}(\boldsymbol{\theta}) &= \sum_{i=1}^{N}\sum_{j=1}^{N} l_{\ell',i,j,N}(\psi)\,\varepsilon_{i,N}(\boldsymbol{\theta})\,\varepsilon_{j,N}(\boldsymbol{\theta})\\ &= \boldsymbol{\varepsilon}_N^{\mathrm{T}}(\boldsymbol{\theta})\,\mathbf{L}_{\ell',N}(\psi)\,\boldsymbol{\varepsilon}_N(\boldsymbol{\theta})\end{aligned}$$

where $\mathbf{L}_{\ell',N}(\psi)$ is a matrix whose (i,j)-th entries, denoted here by $l_{\ell',i,j,N}(\psi)$, can be expressed as functions of $\mathbf{M}_{\ell',N}$, $\mathbf{F}_N$ and $\mathbf{F}_N^{-1}$. Since $\mathbf{F}_N=\mathbf{F}_{0,N}-\mathbf{E}_N(\psi_0-\psi)$ and

$$\mathbf{F}_N^{-1}=\mathbf{F}_N^{-1}\mathbf{F}_{0,N}\mathbf{F}_{0,N}^{-1}=\left[\mathbf{I}_N+\mathbf{F}_N^{-1}\mathbf{E}_N(\psi_0-\psi)\right]\mathbf{F}_{0,N}^{-1},$$

such a matrix is a function of ψ. Under the model's assumptions, $\mathbf{L}_{\ell',N}(\psi)$ is bounded in absolute value in both row and column sums. Because $\overline{\lambda}_{\ell',N}(\boldsymbol{\theta})$ is a quadratic form of $\boldsymbol{\varepsilon}_N(\boldsymbol{\theta})$, the third summation in (A.5) can be developed similarly to the second one. Thus, uniformly in $\boldsymbol{\theta}\in\boldsymbol{\Theta}$:

$$\begin{aligned}\sum_{\ell=2}^{1+QK+P} a_{\ell,\ell',N}\overline{\lambda}_{\ell',N}(\boldsymbol{\theta}) = &\sum_{\ell=2}^{1+QK+P} a_{\ell,\ell',N}\mathbf{d}_N^{\mathrm{T}}(\boldsymbol{\theta})\,\mathbf{L}_{\ell',N}(\psi)\,\mathbf{d}_N(\boldsymbol{\theta})+\\ &+\frac{1}{N}\sum_{\ell'=2}^{1+QK+P} a_{\ell,\ell',N}\lambda^{*}_{\ell',N}(\psi)+\frac{1}{N}\operatorname{Tr}\left(\widetilde{\boldsymbol{\Gamma}}^{***}_{\ell,N}(\boldsymbol{\theta})\,\boldsymbol{\Upsilon}_{0,N}\right)\\ &+\frac{1}{N}\sum_{k=1}^{K}\xi_{0,k}\operatorname{Tr}\left(\widetilde{\boldsymbol{\Gamma}}^{***}_{\ell,k,N}(\boldsymbol{\theta})\,\mathbf{C}_{k,N}\boldsymbol{\Upsilon}_{0,N}\right)\\ &+\frac{1}{N}\sum_{k=1}^{K}\sum_{h=1}^{K}\xi_{0,k}\xi_{0,h}\operatorname{Tr}\left(\mathbf{C}_{k,N}^{\mathrm{T}}\widetilde{\boldsymbol{\Gamma}}^{***}_{\ell,k,h,N}(\boldsymbol{\theta})\,\mathbf{C}_{h,N}\boldsymbol{\Upsilon}_{0,N}\right)\\ &+o_P(1),\end{aligned}\tag{A.8}$$

where, for $\ell'=2,\dots,1+QK+P$, by Lemma A.3 in Lin and Lee (2010):

$$\frac{1}{N}\boldsymbol{\varepsilon}_N^{\mathrm{T}}\mathbf{L}_{\ell',N}(\psi)\,\boldsymbol{\varepsilon}_N\equiv\frac{1}{N}\underbrace{\operatorname{Tr}\left[\mathbf{L}_{\ell',N}(\psi)\,\boldsymbol{\Upsilon}_{0,N}\right]}_{\equiv\lambda^{*}_{\ell',N}(\psi)}+o_P(1),$$

which implicitly provides the definition for the $\lambda^{*}_{\ell',N}(\psi)$ terms that appear in (A.8);

and where, again, some additional $1+K+K^2$ auxiliary matrices are being used:

$$\widetilde{\mathbf{\Gamma}}^{***}_{\ell,N}(\boldsymbol{\theta}) \equiv \sum_{\ell'=2}^{1+QK+P} a_{\ell,\ell',N}\left[\mathbf{\Delta}_N(\boldsymbol{\theta})\mathbf{L}_{\ell',N}(\boldsymbol{\psi})\mathbf{\Delta}_N(\boldsymbol{\theta}) - \mathbf{L}_{\ell',N}(\boldsymbol{\psi})\right]$$
$$\widetilde{\mathbf{\Gamma}}^{***}_{\ell,k,N}(\boldsymbol{\theta}) \equiv \sum_{\ell'=2}^{1+QK+P} a_{\ell,\ell',N}\mathbf{\Delta}_N(\boldsymbol{\theta})\left(\mathbf{L}_{\ell',N}(\boldsymbol{\psi}) + \mathbf{L}^{\mathrm{T}}_{\ell',N}(\boldsymbol{\psi})\right)\mathbf{\Gamma}_{k,N}(\boldsymbol{\theta})$$
$$\widetilde{\mathbf{\Gamma}}^{***}_{\ell,k,h,N}(\boldsymbol{\theta}) \equiv \sum_{\ell'=2}^{1+QK+P} a_{\ell,\ell',N}\mathbf{\Gamma}^{\mathrm{T}}_{k,N}(\boldsymbol{\theta})\mathbf{L}_{\ell',N}(\boldsymbol{\psi})\mathbf{\Gamma}_{h,N}(\boldsymbol{\theta}),$$

for the given ℓ and for $k,h = 1,\ldots,K$ (this time using three asterisks).

It is useful to study the $\lambda^*_{\ell',N}(\boldsymbol{\psi})$ terms more elaborately. As $\mathbf{\Upsilon}_{0,N} = \mathbf{F}_{0,N}\mathbf{\Sigma}_N\mathbf{F}^{\mathrm{T}}_{0,N}$, for $\ell' = 2,\ldots,1+QK+P$ it is:

$$\begin{aligned}
\lambda^*_{\ell',N}(\boldsymbol{\psi}) &= \mathrm{Tr}\left[\mathbf{\Sigma}_N^{\frac{1}{2}}\mathbf{F}^{\mathrm{T}}_{0,N}\mathbf{L}_{\ell',N}(\boldsymbol{\psi})\mathbf{F}_{0,N}\mathbf{\Sigma}_N^{\frac{1}{2}}\right] \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N} l_{\ell',i,j,N}(\boldsymbol{\psi})\,\mathbf{f}_{0,i,N}\mathbf{f}^{\mathrm{T}}_{0,j,N} \\
&= \left[\mathrm{vec}\left(\mathbf{F}^{\mathrm{T}}_N\mathbf{M}_{\ell',N}\mathbf{F}_N\right)\right]^{\mathrm{T}}\mathrm{vec}\left(\mathbf{I}_N \circ \left[\mathbf{F}^{-1}_N\mathbf{F}_{0,N}\mathbf{\Sigma}_N\mathbf{F}^{\mathrm{T}}_{0,N}\left(\mathbf{F}^{-1}_N\right)^{\mathrm{T}}\right]\right),
\end{aligned}$$

where $\mathbf{f}_{0,i,N}$ here denotes the i-th *row* of $\mathbf{F}_{0,N}\mathbf{\Sigma}_N^{\frac{1}{2}}$, for $i = 1,\ldots,N$. This derivation is obtained by reverse-engineering the transformation leading to $\mathbf{L}_{\ell',N}(\boldsymbol{\psi})$. For $\boldsymbol{\psi} = \boldsymbol{\psi}_0$, because $\mathbf{\Sigma}_N$ is a diagonal matrix this expression simplifies to:

$$\lambda^*_{\ell',N}(\boldsymbol{\psi}_0) = \left[\mathrm{vec}\left(\mathbf{F}^{\mathrm{T}}_{0,N}\mathbf{M}_{\ell',N}\mathbf{F}_{0,N}\right)\right]^{\mathrm{T}}\mathrm{vec}\left(\mathbf{\Sigma}_N\right) = \lambda_{\ell',N}(\boldsymbol{\theta}_0),$$

and therefore in the influence function, the $\lambda^*_{\ell',N}(\boldsymbol{\psi}_0)$ terms in (A.8) cancel out with the corresponding $\lambda_{\ell',N}(\boldsymbol{\theta}_0)$ terms in (A.6) and (A.7). Because $\mathbf{d}_N(\boldsymbol{\theta})$ and all matrices of the $\mathbf{\Gamma}^*_{\ell,\cdot,N}(\boldsymbol{\theta})$, $\mathbf{\Gamma}^{**}_{\ell,\cdot,N}(\boldsymbol{\theta})$ and $\mathbf{\Gamma}^{***}_{\ell,\cdot,N}(\boldsymbol{\theta})$ kind are also equal to zero(es) when evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, one concludes that $\mathbf{n}_N(\boldsymbol{\theta}_0) = o_P(1)$, as expected. To complete this part of the proof, note that $\mathbf{n}_N(\boldsymbol{\theta})$ is quadratic in $\boldsymbol{\theta}$ and $\mathbf{\Theta}$ is bounded, hence, $\mathbb{E}[\mathbf{n}_N(\boldsymbol{\theta})]$ is uniformly equicontinuous in $\mathbf{\Theta}$. This fact, along with the identification conditions for $\boldsymbol{\theta}$ (as specified in Theorem 1), implies that the identification uniqueness conditions for $\mathbb{E}\left[\mathbf{n}^{\mathrm{T}}_N(\boldsymbol{\theta})\mathbf{n}_N(\boldsymbol{\theta})\right]$ are satisfied. Hence, consistency of the GMM estimator follows from standard arguments (White, 1996).

Next, we move to the proof of asymptotic normality. The typical manipulation of the GMM First Order Conditions via the Mean Value Theorem gives:

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

where $\overline{\boldsymbol{\theta}}_N$ is a convex combination of $\widehat{\boldsymbol{\theta}}_{GMM}$ and $\boldsymbol{\theta}_0$, and:

$$\mathbf{J}_N(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}^{\mathrm{T}}} \overline{\mathbf{m}}_N(\boldsymbol{\theta}) = \underbrace{= -\frac{1}{N} \begin{bmatrix} \boldsymbol{\iota}_N^{\mathrm{T}} \\ \mathbf{Q}_{1,N} \\ \vdots \\ \mathbf{Q}_{Q,N} \\ 2\boldsymbol{\varepsilon}_{\boldsymbol{N}}^{\mathrm{T}}(\boldsymbol{\theta})\,\mathbf{P}_{1,N} \\ \vdots \\ 2\boldsymbol{\varepsilon}_{\boldsymbol{N}}^{\mathrm{T}}(\boldsymbol{\theta})\,\mathbf{P}_{P,N} \end{bmatrix} \begin{bmatrix} \boldsymbol{\iota}_N & \mathbf{G}_N \mathbf{y}_N & \mathbf{X}_N & \mathbf{G}_N \mathbf{X}_N & \mathbf{0}_N & \dots & \mathbf{0}_N \end{bmatrix} - \frac{1}{N} \frac{\partial \overline{\boldsymbol{\lambda}}_N(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\mathrm{T}}}}_{\equiv \mathbf{J}_N^*(\boldsymbol{\theta})}.$$

An inspection of (A.5) and of its decomposition reveals that $\mathbf{A}_N \mathbf{m}_N(\boldsymbol{\theta}_0)$ is a a vector of linear-quadratic forms of $\boldsymbol{\varepsilon}_N$, and hence of $\boldsymbol{\upsilon}_N$. For $\ell = 1, \dots, 1 + QK + P$:

$$N n_{\ell,N}(\boldsymbol{\theta}_{\mathbf{0}}) = \sum_{q=1}^{Q} \sum_{k=1}^{K} a_{\ell,1+(q-1)K+k,N} \left[\left(\mathbf{q}_{q,k,N} - \boldsymbol{\upsilon}_N^{\mathrm{T}} \mathbf{F}_{0,N}^{\mathrm{T}} \mathbf{C}_{k,N} \mathbf{G}_N^{q-1} \xi_{0,k} \right) \mathbf{F}_{0,N} \boldsymbol{\upsilon}_N \right] + \\ + \sum_{p=1}^{P} a_{\ell,1+QK+p,N} \operatorname{Tr}\left(\mathbf{P}_{p,N} \mathbf{F}_{0,N} \left(\boldsymbol{\upsilon}_N \boldsymbol{\upsilon}_N^{\mathrm{T}} - \boldsymbol{\Sigma}_N \right) \mathbf{F}_{0,N}^{\mathrm{T}} \right).$$

By Assumptions 1, 2, 3, 4, 5, 7, and 9, the stochastic portion of the expression above complies with the conditions of Theorem 1 by Kelejian and Prucha (2001), hence:

$$\sqrt{N} \mathbf{A}_N \mathbf{m}_N(\boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \mathbf{A}_0 \boldsymbol{\Omega}_0 \mathbf{A}_0^{\mathrm{T}}\right). \tag{A.9}$$

We proceed to show that $\mathbf{J}_N\left(\widehat{\boldsymbol{\theta}}_{GMM}\right) = \mathbf{J}_0 + o_P(1)$. We thus study the quantity:

$$\frac{d}{d\beta}\left(\frac{1}{N} \boldsymbol{\varepsilon}_N^{\mathrm{T}}(\boldsymbol{\theta})\,\mathbf{M}_N \boldsymbol{\varepsilon}_N(\boldsymbol{\theta})\right) = -\frac{2}{N} \boldsymbol{\varepsilon}_N^{\mathrm{T}}(\boldsymbol{\theta})\,\mathbf{M}_N \mathbf{G}_N \mathbf{y}_N \tag{A.10}$$

for some given absolutely bounded matrix $\mathbf{M}_N$, showing uniform convergence over $\boldsymbol{\Theta}$. If $\mathbf{M}_N = \mathbf{P}_{p,N}$ for any $p = 1, \dots, P$, for example, this quantity lies the second column of $\mathbf{J}_N^*(\boldsymbol{\theta})$; other elements of $\mathbf{J}_N^*(\boldsymbol{\theta})$ are simpler cases of it. Expand (A.10) as:

$$-\frac{2}{N} \boldsymbol{\varepsilon}_N^{\mathrm{T}}(\boldsymbol{\theta})\,\mathbf{M}_N \widetilde{\mathbf{G}}_N(\beta_0)\left(\alpha_0 \boldsymbol{\iota} + \mathbf{X}_N \boldsymbol{\gamma}_0 + \mathbf{G}_N \mathbf{X}_N \boldsymbol{\delta}_0 + \boldsymbol{\varepsilon}_N\right) = \\ = -2\left(\epsilon_N^*(\boldsymbol{\theta}) + \eta_N^*(\boldsymbol{\theta}) + \sum_{k=1}^{K} \upsilon_{k,N}^*(\boldsymbol{\theta})\right),$$

where, uniformly in $\boldsymbol{\theta} \in \boldsymbol{\Theta}$:

$$\epsilon_N^* (\boldsymbol{\theta}) = \frac{1}{N} \mathbf{d}_N^{\mathrm{T}} (\boldsymbol{\theta}) \mathbf{M}_N \mathbf{G}_N \mathbf{y}_N = \frac{1}{N} \mathbf{d}_N^{\mathrm{T}} (\boldsymbol{\theta}) \mathbf{M}_N \mathbf{G}_N \mathbb{E} [\mathbf{y}_N] + o_P (1) ,$$

and, uniformly in $\boldsymbol{\theta} \in \boldsymbol{\Theta}$:

$$\begin{aligned} \eta_N^* (\boldsymbol{\theta}) &= \frac{1}{N} \boldsymbol{\varepsilon}_N^{\mathrm{T}} \boldsymbol{\Delta}_N^{\mathrm{T}} (\boldsymbol{\theta}) \mathbf{M}_N \mathbf{G}_N \mathbf{y}_N \\ &= \frac{1}{N} \sum_{k=1}^{K} \xi_{0,k} \mathrm{Tr} \left(\boldsymbol{\Delta}_N^{\mathrm{T}} (\boldsymbol{\theta}) \mathbf{M}_N \widetilde{\mathbf{G}}_N (\beta_0) (\gamma_{0,k} \mathbf{I}_N + \delta_{0,k} \mathbf{G}_N) \mathbf{C}_{k,N} \boldsymbol{\Upsilon}_{0,N} \right) \\ &\quad \frac{1}{N} \mathrm{Tr} \left(\boldsymbol{\Delta}_N^{\mathrm{T}} (\boldsymbol{\theta}) \mathbf{M}_N \widetilde{\mathbf{G}}_N (\beta_0) \boldsymbol{\Upsilon}_{0,N} \right) + o_P (1) , \end{aligned}$$

and, for $k = 1, \ldots, K$, uniformly in $\boldsymbol{\theta} \in \boldsymbol{\Theta}$:

$$\begin{aligned} \upsilon_{k,N}^* (\boldsymbol{\theta}) &= \frac{1}{N} (\mathbf{x}_{k,N} - \mathbb{E} [\mathbf{x}_{k,N}])^{\mathrm{T}} \boldsymbol{\Gamma}_{0,k}^{\mathrm{T}} (\boldsymbol{\theta}) \mathbf{M}_N \mathbf{G}_N \mathbf{y}_N \\ &= \frac{1}{N} \xi_{0,k} \sum_{h=1}^{K} \xi_{0,h} \mathrm{Tr} \left(\mathbf{C}_{k,N}^{\mathrm{T}} \boldsymbol{\Gamma}_{0,k}^{\mathrm{T}} (\boldsymbol{\theta}) \mathbf{M}_N \widetilde{\mathbf{G}}_N (\beta_0) (\gamma_{0,h} \mathbf{I}_N + \delta_{0,h} \mathbf{G}_N) \mathbf{C}_{h,N} \boldsymbol{\Upsilon}_{0,N} \right) \\ &\quad \frac{1}{N} \xi_{0,k} \mathrm{Tr} \left(\mathbf{C}_{k,N}^{\mathrm{T}} \boldsymbol{\Gamma}_{0,k}^{\mathrm{T}} (\boldsymbol{\theta}) \mathbf{M}_N \widetilde{\mathbf{G}}_N (\beta_0) \boldsymbol{\Upsilon}_{0,N} \right) + o_P (1) . \end{aligned}$$

Evaluating these terms at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ and collecting them gives:

$$\begin{aligned} - \frac{2}{N} \boldsymbol{\upsilon}^{\mathrm{T}} (\boldsymbol{\theta}_0) \mathbf{M}_N \mathbf{G}_N \mathbf{y}_N &= - \frac{2}{N} \mathrm{Tr} \left[\mathbf{M}_N \widetilde{\mathbf{G}}_N (\beta_0) \boldsymbol{\Upsilon}_{0,N} \right] - \\ &\quad - \frac{2}{N} \sum_{k=1}^{K} \xi_{0,k} \mathrm{Tr} \left[\mathbf{M}_N \widetilde{\mathbf{G}}_N (\beta_0) (\gamma_{0,k} \mathbf{I}_N + \delta_{0,k} \mathbf{G}_N) \mathbf{C}_{k,N} \boldsymbol{\Upsilon}_{0,N} \right] + o_P (1) , \end{aligned}$$

yielding the uniform convergence result that is sought after, which as argued can be generalized to all components of $\mathbf{J}_N (\boldsymbol{\theta})$. Since the GMM estimator is consistent, this straightforwadly implies $\mathbf{J}_N \left(\overline{\boldsymbol{\theta}}_N \right) = \mathbf{J}_0 + o_P (1)$ as well.

We conclude the analysis by examining the Jacobian of $\overline{\boldsymbol{\lambda}}_N (\boldsymbol{\theta})$. For $q = 1, \ldots, Q$ and $k = 1, \ldots, K$:

$$\begin{aligned} \frac{\partial \overline{\lambda}_{1,q,k,N} (\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\mathrm{T}}} &= \xi_k \boldsymbol{\upsilon}_N (\boldsymbol{\theta})^{\mathrm{T}} \left(\mathbf{I}_N \circ \left[\mathbf{F}_N^{\mathrm{T}} \left[\mathbf{G}_N^{q-1} \mathbf{C}_{k,N} + \left(\mathbf{G}_N^{q-1} \mathbf{C}_{k,N} \right)^{\mathrm{T}} \right] \mathbf{F}_N \right] \right) \frac{\partial \boldsymbol{\upsilon}_N (\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\mathrm{T}}} + \\ &\quad + \sum_{k=1}^{K} \left[\mathbf{0}_{2+2K}^{\mathrm{T}} \quad \mathbf{i}_k^{\mathrm{T}} \mathrm{Tr} \left(\mathbf{F}_N \mathbf{U}_N (\boldsymbol{\theta}) \mathbf{F}_N^{\mathrm{T}} \mathbf{G}_N^{q-1} \mathbf{C}_{k,N} \right) \quad 0 \right] + \\ &\quad + \left[\mathbf{0}_{2+3K}^{\mathrm{T}} \quad \xi_k \mathrm{Tr} \left(\mathbf{G}_N^{q-1} \mathbf{C}_{k,N} \left(\mathbf{E}_N \mathbf{U}_N (\boldsymbol{\theta}) \mathbf{F}_N^{\mathrm{T}} + \mathbf{F}_N \mathbf{U}_N (\boldsymbol{\theta}) \mathbf{E}_N^{\mathrm{T}} \right) \right) \right] \end{aligned}$$

where $\mathbf{0}_{2+2K}$ $(\mathbf{0}_{2+3K})$ is a vector of $2+2K$ $(2+3K)$ zeroes; vector $\mathbf{i}_k$, for $k = 1, \ldots, K$, is as defined in the proof of Theorem 1;

$$\mathbf{U}_N(\boldsymbol{\theta}) \equiv \operatorname{diag}\left(\upsilon_1^2(\boldsymbol{\theta}), \ldots, \upsilon_N^2(\boldsymbol{\theta})\right)$$

and:

$$\frac{\partial \boldsymbol{\upsilon}_N(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\mathrm{T}}} = -\mathbf{F}_N^{-1}\left[\boldsymbol{\iota}_N \quad \mathbf{G}_N\mathbf{y}_N \quad \mathbf{X}_N \quad \mathbf{G}_N\mathbf{X}_N \quad \mathbf{0}_N \quad \ldots \quad \mathbf{0}_N\right].$$

Similarly, for $p = 1, \ldots, P$:

$$\begin{aligned}\frac{\partial \overline{\lambda}_{1,p,N}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\mathrm{T}}} = \boldsymbol{\upsilon}_N(\boldsymbol{\theta})^{\mathrm{T}}\left(\mathbf{I}_N \circ \left[\mathbf{F}_N^{\mathrm{T}}\left(\mathbf{P}_{p,N} + \mathbf{P}_{p,N}^{\mathrm{T}}\right)\mathbf{F}_N\right]\right)\frac{\partial \boldsymbol{\upsilon}_N(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\mathrm{T}}} + \\ + \left[\mathbf{0}_{2+3K}^{\mathrm{T}} \quad \operatorname{Tr}\left(\mathbf{P}_{p,N}\left(\mathbf{E}_N\mathbf{U}_N(\boldsymbol{\theta})\mathbf{F}_N^{\mathrm{T}} + \mathbf{F}_N\mathbf{U}_N(\boldsymbol{\theta})\mathbf{E}_N^{\mathrm{T}}\right)\right)\right].\end{aligned}$$

All these derivatives retain the mathematical structure of (A.10); this follows because, as shown previously, $\overline{\boldsymbol{\lambda}}_N(\boldsymbol{\theta})$ can be expressed as a vector of quadratic forms of $\boldsymbol{\varepsilon}_N(\boldsymbol{\theta})$. Consequently, they can be decomposed similarly to (A.10) so as to show that:

$$\frac{1}{N}\frac{\partial \overline{\boldsymbol{\lambda}}_N(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\mathrm{T}}} = \frac{1}{N}\frac{\partial \boldsymbol{\lambda}_N(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^{\mathrm{T}}} + o_P(1).$$

These results are combined with (A.9) via Slutsky's Theorem to establish asymptotic normality of the proposed GMM estimator. This concludes the proof.

Addendum A Bias of conventional methods

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 21 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 bias of both OLS and the BDF moments. Consider a simplified version of (1), with $K = 1$, $\delta = \psi = 0$, and homoschedastic errors ($\boldsymbol{\Sigma} = \sigma_0^2 \mathbf{I}$, $\sigma_0^2 > 0$):

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

With exogenous $\mathbf{x}$ ($\xi = 0$ or $\mathbf{C} = \mathbf{0}$), OLS would be based on the following moments:

$$\begin{aligned} \mathbb{E}\left[\boldsymbol{\iota}^{\mathrm{T}}\boldsymbol{\varepsilon}\right] &= 0 \\ \mathbb{E}\left[(\mathbf{G}\mathbf{y})^{\mathrm{T}}\boldsymbol{\varepsilon}\right] &= \sigma_0^2 \mathrm{Tr}\left((\mathbf{I} - \beta\mathbf{G})^{-1}\mathbf{G}^{\mathrm{T}}\right) \\ \mathbb{E}\left[\mathbf{x}^{\mathrm{T}}\boldsymbol{\varepsilon}\right] &= 0. \end{aligned} \tag{A.11}$$

The bias arising from endogneity is proportional to the right-hand side of (A.11). Since $\mathbf{G}\mathbf{y}$ linearly depends on $\boldsymbol{\varepsilon}$, this moment is non-zero in expectation, and therefore OLS is inconsistent. BDF circumvent this problems by replacing (A.11) with:

$$\mathbb{E}\left[(\mathbf{G}^q\mathbf{x})^{\mathrm{T}}\boldsymbol{\varepsilon}\right] = 0$$

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

The model we consider introduces correlation between $\mathbf{x}$ and $\boldsymbol{\varepsilon}$. The key OLS and BDF moments are, under endogeneity ($\xi \neq 0$, $\mathbf{C} \neq \mathbf{0}$):

$$\begin{aligned} \mathbb{E}\left[(\mathbf{G}\mathbf{y})^{\mathrm{T}}\boldsymbol{\varepsilon}\right] &= \mathbb{E}\left[(\gamma\mathbf{x} + \boldsymbol{\varepsilon})^{\mathrm{T}}(\mathbf{I} - \beta\mathbf{G})^{-1}\mathbf{G}\boldsymbol{\varepsilon}\right] \\ &= \sigma^2 \mathrm{Tr}\left(\left[\gamma\xi\mathbf{C}^{\mathrm{T}} + \mathbf{I}\right](\mathbf{I} - \beta\mathbf{G})^{-1}\mathbf{G}^{\mathrm{T}}\right) \end{aligned} \tag{A.12}$$

and, for $q \in \mathbb{N}_0$:

$$\mathbb{E}\left[(\mathbf{G}^q\mathbf{x})^{\mathrm{T}}\boldsymbol{\varepsilon}\right] = \sigma^2\xi\mathrm{Tr}\left((\mathbf{G}^q\mathbf{C})^{\mathrm{T}}\right). \tag{A.13}$$

Hence, both (A.12) and (A.13) are non-zero. Expression (A.12) comprises two terms, that encode the endogeneity of $\mathbf{G}\mathbf{y}$ and $\mathbf{x}$, respectively. Instead, the bias in (A.13) is entirely due to the endogeneity of $\mathbf{x}$. Both biases depend crucially on the interaction between the spatial weighting (or adjacency) matrix $\mathbf{G}$ and the characteristics matrix $\mathbf{C}$, which jointly determine the spatial correlation of the variables at hand.

Addendum B A game of social interactions

The literature in spatial econometrics and peer or social effects has emphasized the microfoundation of econometric models with spatial lags of the dependent variable as equilibrium outcomes of games of social interactions (e.g. public good games) where agents have quadratic utilities and cost functions. Here we propose a complementary microfoundation based on exponential functions instead. Its advantages are twofold. First, it generalizes to familiar settings in economics such as games between firms with Cobb-Douglas production functions. Second, it loads model (15), which features the network indegree as a right-hand side variable of interest, with interpretation.

We consider a public good game where N players indexed as $i = 1, \dots, N$ interact in a network, which is exogenous and whose topology is summarized by an adjacency matrix $\mathbf{G}$ with zero diagonal and entries g_{ij}. This game applies to settings like peer effects at school and firms' R&D investment decisions with knowledge spillovers: our running examples. Players are heterogeneous and typified by a variable that we denote as χ_i (e.g. the prior background of students, or the established production capacity of firms). Players maximize the following "twice exponential" utility function:

$$U_i\left(e_1, \dots, e_N; \chi_i\right) = \exp\left[y_i\left(e_1, \dots, e_N; \chi_i\right)\right] - \exp\left(e_i\right), \tag{A.14}$$

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 contribution to the "benefit" component of utility (A.14):

$$y_i\left(e_1, \dots, e_N; \chi_i\right) = \pi + \mu e_i + \nu \sum_{i=1}^{N} g_{ij} e_j + \chi_i. \tag{A.15}$$

Individual outcomes depend on χ_i and on a costly strategic variable e_i that we name *effort*: it represents, for instance, time dedicated to homeworks in a peer effects setting or R&D investment. Because of social interactions and externalities, y_i also depends on the effort of all the other players that an agent is connected to. The private and social effects of effort are parametrized as $\mu > 0$ and $0 < \nu < 1$, respectively (we could allow $-1 < \nu < 0$ to introduce negative externalities, as in models about the "business stealing" effect of R&D, but we prefer to retain the public good interpretation of e_i). 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 (e.g. Blume et al., 2015). We further define the combined parameter $\beta \equiv \nu / (1 - \mu)$ and assume that matrix $\mathbf{I} - \beta \mathbf{G}$ is nonsingular.

We discuss a game of complete information: that is, $\boldsymbol{\chi} \equiv (\chi_1, \dots, \chi_N)$ is common knowledge. The first order conditions are readily summarized by the expression:

$$e_j = y_j + \log \mu$$

for $i = 1, \ldots, N$, and it is easy to verify that this model's (unique) Nash equilibrium is expressed as:

$$y_i = \alpha + \beta \sum_{j=1}^{N} g_{ij} y_j + \phi \sum_{j=1}^{N} g_{ij} + \frac{\chi_i}{1-\nu} \tag{A.16}$$

for $i = 1, \ldots, N$, with $\alpha \equiv [\pi + \mu \log \mu] / (1-\mu)$ and $\phi \equiv \nu \log \mu / (1-\mu)$. Now let:

$$(1-\nu)^{-1} \boldsymbol{\chi} = \mathbf{X}\boldsymbol{\gamma} + \mathbf{G}\mathbf{X}\boldsymbol{\delta} + \boldsymbol{\varepsilon}$$

where $\boldsymbol{\varepsilon}$ is a vector of unobserved shocks and $\mathbf{X}$ represents observed variables possibly dependent on $\boldsymbol{\varepsilon}$ as per the SLE specification (2). Stacking (A.16) over all observations would then yield model (15) in the text. Note that the separate identification of β and ϕ as per Corollary 2 allows to solve for the more fundamental, primitive parameters μ and ν, *even if effort is unobserved*. The main structural equation (1) in the text is obtained from this model when the adjacency matrix is "row-normalized" ($\bar{\mathbf{g}} = \boldsymbol{\iota}$) as typical in studies about peer effects. In this case, the intercept of the model embodies both the intercept of (A.16) and ϕ; specifically, it is $\alpha = [\pi + (\mu + \nu) \log \mu] / (1-\mu)$.

Addendum C Estimation of the asymptotic variance

What follows elaborates upon variance-covariance matrix of the model's moments $\boldsymbol{\Omega}_0$ that is introduced in Section 4. The analysis of this matrix and of its sample analogue clarifies how standard errors for our proposed estimator are to be calculated. Write:

$$\boldsymbol{\Omega}_0 = \frac{1}{N} \begin{bmatrix} \omega_{N,1,1} & \cdots & \omega_{N,1,1+QK+P} \\ \vdots & \ddots & \vdots \\ \omega_{N,1+QK+P,1} & \cdots & \omega_{N,1+QK+P,1+QK+P} \end{bmatrix} + o_P(1),$$

and note that this matrix is symmetric: $\omega_{N,i,j} = \omega_{N,j,i}$ for all $i, j = 1, \ldots, 1+QK+P$. Let $\boldsymbol{\Sigma}_N^* = \operatorname{diag}(\mathbb{E}[\upsilon_1^3], \mathbb{E}[\upsilon_2^3], \ldots, \mathbb{E}[\upsilon_N^3])$: an $N \times N$ matrix. In addition:

$$\boldsymbol{\Sigma}_N^{\ddagger} = \mathbf{I}_N \otimes \boldsymbol{\Sigma}_N^{\dagger,diag} + \left(\boldsymbol{\iota}_N \boldsymbol{\iota}_N^{\mathrm{T}} - \mathbf{I}_N\right) \otimes \boldsymbol{\Sigma}_N^{**,out}$$

is an $N^2 \times N^2$ matrix whose constituent blocks, written as $\boldsymbol{\Sigma}_N^{\dagger,diag}$ and $\boldsymbol{\Sigma}_N^{\dagger,out}$, are the following symmetric $N \times N$ matrices:

$$\boldsymbol{\Sigma}_N^{\dagger,out} = \begin{bmatrix} \mathbb{E}[\upsilon_1^2]\,\mathbb{E}[\upsilon_1^2] & \mathbb{E}[\upsilon_1^2]\,\mathbb{E}[\upsilon_2^2] & \cdots & \mathbb{E}[\upsilon_1^2]\,\mathbb{E}[\upsilon_N^2] \\ \mathbb{E}[\upsilon_2^2]\,\mathbb{E}[\upsilon_1^2] & \mathbb{E}[\upsilon_2^2]\,\mathbb{E}[\upsilon_2^2] & \cdots & \mathbb{E}[\upsilon_2^2]\,\mathbb{E}[\upsilon_N^2] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{E}[\upsilon_N^2]\,\mathbb{E}[\upsilon_1^4] & \mathbb{E}[\upsilon_N^2]\,\mathbb{E}[\upsilon_2^2] & \cdots & \mathbb{E}[\upsilon_N^2]\,\mathbb{E}[\upsilon_N^2] \end{bmatrix},$$

and $\boldsymbol{\Sigma}_N^{\dagger,diag} = \boldsymbol{\Sigma}_N^{\dagger,out} - \mathbf{I}_N \circ \boldsymbol{\Sigma}_N^{\dagger,out} + \operatorname{diag}(\mathbb{E}[\upsilon_1^4], \mathbb{E}[\upsilon_2^4], \ldots, \mathbb{E}[\upsilon_N^4])$.

Using these matrices as well as $\mathbf{F}_{0,N} = \mathbf{I}_N + \psi_0 \mathbf{E}_N$, $\mathbf{\Omega}_0$ is completely characterized by the following expressions, for $k, k' = 1, \ldots, N$, $q, q' = 1, \ldots, Q$, and $p, p' = 1, \ldots, P$:

$$\begin{aligned}
\omega_{N,1,1} &= \iota_N^{\mathrm{T}} \mathbf{F}_{0,N}^{\mathrm{T}} \mathbf{\Sigma}_N \mathbf{F}_{0,N} \iota_N \\
\omega_{N,1+K(q-1)+k,1} &= \xi_{0,k} \mathrm{Tr}\left(\mathbf{F}_{0,N} \circ \mathbf{F}_{0,N}^{\mathrm{T}} \left(\mathbf{G}_N^{q-1}\right)^{\mathrm{T}} \mathbf{F}_{0,N} \circ \mathbf{\Sigma}_N^{*}\right) \\
&\quad + \iota_N^{\mathrm{T}} \mathbf{F}_{0,N}^{\mathrm{T}} \mathbf{\Sigma}_N \mathbf{F}_{0,N} \mathbf{G}_N^{q-1} \widetilde{\mathbf{x}}_{k,N} \\
\omega_{N,1+QK+p,1} &= \mathrm{Tr}\left(\mathbf{F}_{0,N} \circ \mathbf{F}_{0,N}^{\mathrm{T}} \mathbf{P}_N \mathbf{F}_{0,N} \circ \mathbf{\Sigma}_N^{*}\right) \\
\omega_{N,1+K(q-1)+k,1+K(q'-1)+k'} &= \xi_{0,k} \xi_{0,k'} \mathrm{vec}\left(\mathbf{F}_{0,N}^{\mathrm{T}} \mathbf{G}_N^{q-1} \mathbf{C}_{k,N} \mathbf{F}_{0,N}\right)^{\mathrm{T}} \cdot \\
&\quad \cdot \mathbf{\Sigma}_N^{\ddagger} \mathrm{vec}\left(\mathbf{F}_{0,N}^{\mathrm{T}} \mathbf{G}_N^{q'-1} \mathbf{C}_{k',N} \mathbf{F}_{0,N}\right) \\
&\quad + \xi_{0,k} \iota_N^{\mathrm{T}} \left(\mathbf{F}_{0,N}^{\mathrm{T}} \mathbf{G}_N^{q-1} \mathbf{C}_{k,N} \mathbf{F}_{0,N} \circ \mathbf{\Sigma}_N^{*}\right) \mathbf{F}_{0,N}^{\mathrm{T}} \mathbf{G}_N^{q'-1} \widetilde{\mathbf{x}}_{k',N} \\
&\quad + \xi_{0,k'} \iota_N^{\mathrm{T}} \left(\mathbf{F}_{0,N}^{\mathrm{T}} \mathbf{G}_N^{q'-1} \mathbf{C}_{k',N} \mathbf{F}_{0,N} \circ \mathbf{\Sigma}_N^{*}\right) \mathbf{F}_{0,N}^{\mathrm{T}} \mathbf{G}_N^{q-1} \widetilde{\mathbf{x}}_{k,N} \\
&\quad - \xi_{0,k} \mathrm{Tr}\left(\mathbf{F}_{0,N}^{\mathrm{T}} \left(\mathbf{G}_N^{q-1} \mathbf{C}_{k,N}\right)^{\mathrm{T}} \mathbf{F}_{0,N} \mathbf{\Sigma}_N\right) \cdot \\
&\quad \cdot \xi_{0,k'} \mathrm{Tr}\left(\mathbf{F}_{0,N}^{\mathrm{T}} \left(\mathbf{G}_N^{q'-1} \mathbf{C}_{k',N}\right)^{\mathrm{T}} \mathbf{F}_{0,N} \mathbf{\Sigma}_N\right) \\
&\quad + \left(\mathbf{G}_N^{q-1} \widetilde{\mathbf{x}}_{k,N}\right)^{\mathrm{T}} \mathbf{F}_{0,N}^{\mathrm{T}} \mathbf{\Sigma}_N \mathbf{F}_{0,N} \mathbf{G}_N^{q'-1} \widetilde{\mathbf{x}}_{k',N} \\
\omega_{N,1+K(q-1)+k,1+QK+p} &= \xi_{0,k} \mathrm{vec}\left(\mathbf{F}_{0,N}^{\mathrm{T}} \mathbf{G}_N^{q-1} \mathbf{C}_{k,N} \mathbf{F}_{0,N}\right)^{\mathrm{T}} \mathbf{\Sigma}_N^{\ddagger} \mathrm{vec}\left(\mathbf{F}_{0,N}^{\mathrm{T}} \mathbf{P}_{p,N}^{\mathrm{T}} \mathbf{F}_{0,N}\right) \\
&\quad + \iota_N^{\mathrm{T}} \left(\mathbf{F}_{0,N}^{\mathrm{T}} \mathbf{P}_{p,N} \mathbf{F}_{0,N} \circ \mathbf{\Sigma}_N^{*}\right) \mathbf{F}_{0,N}^{\mathrm{T}} \mathbf{G}_N^{q-1} \widetilde{\mathbf{x}}_{k,N} \\
&\quad - \xi_{0,k} \mathrm{Tr}\left(\mathbf{F}_{0,N}^{\mathrm{T}} \left(\mathbf{G}_N^{q-1} \mathbf{C}_{k,N}\right)^{\mathrm{T}} \mathbf{F}_{0,N} \mathbf{\Sigma}_N\right) \cdot \\
&\quad \cdot \mathrm{Tr}\left(\mathbf{F}_{0,N}^{\mathrm{T}} \mathbf{P}_{p,N} \mathbf{F}_{0,N} \mathbf{\Sigma}_N\right) \\
\omega_{N,1+QK+p,1+QK+p'} &= \mathrm{vec}\left(\mathbf{F}_{0,N}^{\mathrm{T}} \mathbf{P}_{p,N} \mathbf{F}_{0,N}\right)^{\mathrm{T}} \mathbf{\Sigma}_N^{\ddagger} \mathrm{vec}\left(\mathbf{F}_{0,N}^{\mathrm{T}} \mathbf{P}_{p',N} \mathbf{F}_{0,N}\right) \\
&\quad - \mathrm{Tr}\left(\mathbf{F}_{0,N}^{\mathrm{T}} \mathbf{P}_{p,N} \mathbf{F}_{0,N} \mathbf{\Sigma}_N\right) \mathrm{Tr}\left(\mathbf{F}_{0,N}^{\mathrm{T}} \mathbf{P}_{p',N} \mathbf{F}_{0,N} \mathbf{\Sigma}_N\right).
\end{aligned}$$

A consistent estimator of $\mathbf{\Omega}_0$ is obtained as:

$$\widehat{\mathbf{\Omega}}_N = \frac{1}{N} \begin{bmatrix} \widehat{\omega}_{N,1,1} & \cdots & \widehat{\omega}_{N,1,1+QK+P} \\ \vdots & \ddots & \vdots \\ \widehat{\omega}_{N,1+QK+P,1} & \cdots & \widehat{\omega}_{N,1+QK+P,1+QK+P} \end{bmatrix}$$

where for $i, j = 1, \ldots, 1 + QK + P$, each element $\widehat{\omega}_{N,i,j}$ is obtained as an approppriate counterpart of $\omega_{N,i,j}$. In particular, after obtaining the GMM estimate one calculates:

$$\begin{aligned}
\widehat{\mathbf{F}}_N &= \mathbf{I}_N + \widehat{\psi}_{GMM} \mathbf{E}_N \\
\widehat{\boldsymbol{\upsilon}}_N &= \widehat{\mathbf{F}}_N^{-1} \left(\mathbf{y}_N - \widehat{\alpha}_{GMM} \iota - \widehat{\beta}_{GMM} \mathbf{G}_N \mathbf{y}_N - \mathbf{X}_N \widehat{\boldsymbol{\gamma}}_{GMM} - \mathbf{G}_N \mathbf{X}_N \widehat{\boldsymbol{\delta}}_{GMM}\right)
\end{aligned}$$

and thus, $\mathbf{\Sigma}_N$ is replaced with $\operatorname{diag}\left(\widehat{\upsilon}_1^2,\ldots,\widehat{\upsilon}_N^2\right)$; $\mathbf{\Sigma}_N^*$ with $\operatorname{diag}\left(\widehat{\upsilon}_1^3,\ldots,\widehat{\upsilon}_N^3\right)$; $\mathbf{\Sigma}_N^\ddagger$ with an analogous $N^2 \times N^2$ matrix where all moments of the $\mathbb{E}\left[\upsilon_i^o\right]$ kind, for $i = 1, \ldots, N$ and $o = 2, 4$, are replaced simply with $\widehat{\upsilon}_i^o$; $\mathbf{F}_{0,N}$ is replaced with $\widehat{\mathbf{F}}_N$; and lastly, $\widetilde{\mathbf{x}}_{k,N}$, for $k = 1, \ldots, K$, is replaced with:

$$\widehat{\widetilde{\mathbf{x}}}_{k,N} = \mathbf{X}_{*,k,N} - \widehat{\xi}_{k,GMM}\mathbf{C}_{k,N}\widehat{\mathbf{F}}_N\widehat{\boldsymbol{\upsilon}}_N.$$

If the model also featured moments of the $\mathbb{E}\left[\mathbf{z}_N^{\mathrm{T}}\boldsymbol{\varepsilon}\left(\boldsymbol{\theta}\right)\right] = 0$ kind, where $\mathbf{z}_N$ is some exogenous variable, the analysis of $\mathbf{\Omega}_0$ and its consistent estimator is straightforward to extend. For each such variable, one would add one row and one column to $\mathbf{\Omega}_0$; for simplicity, let them both indexed by 0. Thus, the elements $\omega_{N,0,*}$ from the extra row are akin to those of the $\omega_{N,1,*}$ kind, but with $\mathbf{z}_N$ instead of $\boldsymbol{\iota}_N$; similarly, $\mathbf{z}_N$ replaces $\boldsymbol{\iota}_N$ in the extra column. The element where the extra row and column cross is simply:

$$\omega_{N,0,0} = \mathbb{E}\left[\left(\mathbf{z}_N^{\mathrm{T}}\boldsymbol{\varepsilon}\left(\boldsymbol{\theta}\right)\right)^2\right] = \mathbf{z}_N^{\mathrm{T}}\mathbf{F}_{0,N}^{\mathrm{T}}\mathbf{\Sigma}_N\mathbf{F}_{0,N}\mathbf{z}_N.$$

The construction of $\widehat{\mathbf{\Omega}}_N$ in this case requires the calculation of the residuals $\widehat{\boldsymbol{\upsilon}}_N$ to be appropriately revised to accommodate the additional variable(s) $\mathbf{z}_N$.

Addendum D Additional Monte Carlo experiments

The tables of this section summarize additional Monte Carlo experiments built around variations of the baseline from Section 5. What follows is a brief summary of each.

- Table A.1 provides an example about how our baseline simulation from Table 1 responds to slight perturbations of key regression parameter such as β or γ. The results are virtually unchanged. Analogous results are obtained for different values of β and/or γ.

- Table A.2 removes some key parameters from the d.g.p.: either ξ or ψ is zero. In the former case, endogeneity is removed from the model; in the latter case, the structural errors are independent. The simulation results change as expected; in particular, when $\xi = 0$ conventional methods based on spatial lags of $\mathbf{x}$ ('2SLSb' and '3SLS') appear to estimate β accurately.

- Table A.3 is obtained by modifying the small-world algorithm, so that it either yields denser ($B = 4$) or more irregular ($b = 0.95$) networks with respect to the baseline. The results are qualitatively the same as in Table 1.

- Table A.4 reports on an exercise about changing features of the model's spatial structure; in particular, $\mathbf{C} = \mathbf{I} + \mathbf{G}$. Interestingly, if $\mathbf{C}$ is misspecified by setting $\mathbf{C}_e = \mathbf{I} + \mathbf{G} + \mathbf{G}^2$, the estimate of ξ worsens relative to the baseline, but those of β and γ do not appear to deteriorate to the same extent.

Table A.1: Monte Carlo Simulations, perturbed parameters (part one)

Target Parameter	Experiment A1: as in experiment 1, but with stronger social effects ($\beta = 0.5$)								
	GMM1	GMM2	GMM3	GMM4	OLS	2SLSa	2SLSb	2SLSc	3SLS
$\alpha = 0.25$	0.249	0.250	0.249	0.285	0.234	0.218	0.251	0.212	0.250
	(0.022)	(0.023)	(0.024)	(0.020)	(0.008)	(0.014)	(0.056)	(0.341)	(0.155)
	[0.025]	[0.025]	[0.024]	[0.016]	[0.008]	[0.013]	[0.156]	[2.142]	[0.686]
	{0.902}	{0.901}	{0.872}	{0.448}	{0.000}	{0.000}	{0.060}	{0.787}	{0.253}
$\beta = 0.50$	0.501	0.500	0.501	0.475	0.511	0.524	0.499	0.527	0.501
	(0.015)	(0.015)	(0.016)	(0.013)	(0.005)	(0.010)	(0.025)	(0.261)	(0.084)
	[0.018]	[0.017]	[0.017]	[0.011]	[0.005]	[0.010]	[0.068]	[2.220]	[0.383]
	{0.910}	{0.905}	{0.874}	{0.434}	{0.000}	{0.000}	{0.017}	{0.310}	{0.121}
$\gamma = 0.50$	0.500	0.500	0.500	0.538	0.576	0.568	0.569	0.473	0.554
	(0.013)	(0.013)	(0.014)	(0.011)	(0.005)	(0.008)	(0.425)	(0.758)	(1.223)
	[0.010]	[0.010]	[0.010]	[0.007]	[0.006]	[0.007]	[1.360]	[12.787]	[6.530]
	{0.828}	{0.830}	{0.808}	{0.018}	{0.000}	{0.000}	{0.272}	{0.524}	{0.546}
$\chi = 1.00$	0.999	0.999	0.999	1.003	0.997	0.993	1.000	0.990	1.000
	(0.008)	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)	(0.016)	(0.070)	(0.040)
	[0.005]	[0.005]	[0.005]	[0.004]	[0.006]	[0.007]	[0.046]	[0.306]	[0.208]
	{0.764}	{0.766}	{0.761}	{0.697}	{0.000}	{0.000}	{0.005}	{0.039}	{0.057}
$\xi = 10.0$	9.853	9.813	9.824	6.968	–	–	–	–	–
	(0.819)	(0.774)	(0.855)	(0.720)					
	[0.547]	[0.550]	[0.552]	[0.405]					
	{0.771}	{0.794}	{0.769}	{0.001}					
$\psi = 0.25$	0.241	0.244	0.236	0.233	–	–	–	–	–
	(0.068)	(0.071)	(0.077)	(0.097)					
	[0.087]	[0.086]	[0.083]	[0.052]					
	{0.931}	{0.924}	{0.899}	{0.665}					

Target Parameter	Experiment A2: as in experiment 1, but with weaker covariate effects ($\gamma = 0.2$)								
	GMM1	GMM2	GMM3	GMM4	OLS	2SLSa	2SLSb	2SLSc	3SLS
$\alpha = 0.25$	0.248	0.246	0.246	0.287	0.239	0.180	0.251	0.191	0.249
	(0.027)	(0.029)	(0.031)	(0.026)	(0.008)	(0.034)	(0.028)	(0.527)	(0.160)
	[0.063]	[0.053]	[0.051]	[0.027]	[0.009]	[0.029]	[0.055]	[1.961]	[0.533]
	{0.943}	{0.937}	{0.913}	{0.629}	{0.000}	{0.015}	{0.055}	{0.848}	{0.248}
$\beta = 0.40$	0.402	0.403	0.404	0.369	0.410	0.461	0.399	0.453	0.399
	(0.022)	(0.024)	(0.025)	(0.021)	(0.007)	(0.030)	(0.019)	(0.469)	(0.117)
	[0.053]	[0.044]	[0.042]	[0.023]	[0.007]	[0.025]	[0.036]	[1.835]	[0.402]
	{0.944}	{0.943}	{0.918}	{0.616}	{0.000}	{0.000}	{0.023}	{0.571}	{0.155}
$\gamma = 0.20$	0.200	0.200	0.200	0.244	0.280	0.268	0.279	0.227	0.349
	(0.014)	(0.014)	(0.014)	(0.012)	(0.005)	(0.008)	(0.255)	(0.446)	(1.812)
	[0.011]	[0.010]	[0.010]	[0.007]	[0.005]	[0.008]	[0.618]	[2.934]	[5.851]
	{0.804}	{0.799}	{0.796}	{0.014}	{0.000}	{0.000}	{0.517}	{0.877}	{0.694}
$\chi = 1.00$	1.000	0.999	0.999	1.004	0.998	0.987	1.000	0.988	1.000
	(0.008)	(0.008)	(0.008)	(0.007)	(0.007)	(0.010)	(0.014)	(0.120)	(0.043)
	[0.007]	[0.007]	[0.007]	[0.005]	[0.006]	[0.009]	[0.029]	[0.422]	[0.243]
	{0.826}	{0.818}	{0.801}	{0.720}	{0.000}	{0.000}	{0.002}	{0.061}	{0.072}
$\xi = 10.0$	9.839	9.816	9.813	6.619	–	–	–	–	–
	(0.841)	(0.782)	(0.868)	(0.822)					
	[0.672]	[0.633]	[0.629]	[0.519]					
	{0.799}	{0.834}	{0.776}	{0.008}					
$\psi = 0.25$	0.240	0.238	0.233	0.158	–	–	–	–	–
	(0.066)	(0.067)	(0.074)	(0.092)					
	[0.124]	[0.105]	[0.100]	[0.063]					
	{0.950}	{0.951}	{0.927}	{0.586}					

See the notes accompanying Table 1 for a description of this table's structure.

Table A.2: Monte Carlo Simulations, perturbed parameters (part two)

Target Parameter	Experiment A3: as in experiment 1, but with no endogeneity ($\xi = 0$)								
	GMM1	GMM2	GMM3	GMM4	OLS	2SLSa	2SLSb	2SLSc	3SLS
$\alpha = 0.25$	0.250	0.253	0.251	0.252	0.228	0.252	0.250	0.251	0.247
	(0.038)	(0.039)	(0.039)	(0.038)	(0.012)	(0.048)	(0.032)	(0.334)	(0.079)
	[0.025]	[0.025]	[0.025]	[0.027]	[0.011]	[0.048]	[0.076]	[0.687]	[0.610]
	{0.771}	{0.746}	{0.723}	{0.767}	{0.000}	{0.015}	{0.067}	{0.824}	{0.233}
$\beta = 0.40$	0.400	0.398	0.400	0.398	0.419	0.398	0.400	0.399	0.403
	(0.031)	(0.032)	(0.032)	(0.031)	(0.009)	(0.042)	(0.021)	(0.293)	(0.070)
	[0.021]	[0.021]	[0.020]	[0.022]	[0.009]	[0.042]	[0.046]	[0.604]	[0.423]
	{0.775}	{0.751}	{0.724}	{0.763}	{0.000}	{0.005}	{0.029}	{0.553}	{0.142}
$\gamma = 0.50$	0.496	0.496	0.498	0.499	0.490	0.500	0.474	0.487	0.538
	(0.040)	(0.036)	(0.038)	(0.023)	(0.016)	(0.023)	(0.842)	(0.336)	(2.280)
	[0.023]	[0.020]	[0.021]	[0.012]	[0.013]	[0.025]	[4.236]	[0.650]	[11.64]
	{0.739}	{0.718}	{0.721}	{0.692}	{0.000}	{0.000}	{0.624}	{0.542}	{0.738}
$\chi = 1.00$	1.000	1.000	1.000	1.000	0.996	1.001	1.000	1.001	1.000
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.012)	(0.017)	(0.069)	(0.034)
	[0.005]	[0.005]	[0.005]	[0.005]	[0.008]	[0.012]	[0.072]	[0.136]	[0.206]
	{0.717}	{0.723}	{0.716}	{0.713}	{0.000}	{0.000}	{0.009}	{0.044}	{0.059}
$\xi = 0.00$	0.073	0.105	0.042	0.019	–	–	–	–	–
	(0.903)	(0.850)	(0.896)	(0.507)					
	[0.510]	[0.451]	[0.488]	[0.267]					
	{0.708}	{0.686}	{0.703}	{0.691}					
$\psi = 0.25$	0.254	0.256	0.255	0.254	–	–	–	–	–
	(0.050)	(0.050)	(0.050)	(0.048)					
	[0.032]	[0.032]	[0.032]	[0.033]					
	{0.778}	{0.775}	{0.761}	{0.768}					

Target Parameter	Experiment A4: as in experiment 1, but with i.n.i.d. structural errors ($\psi = 0$)								
	GMM1	GMM2	GMM3	GMM4	OLS	2SLSa	2SLSb	2SLSc	3SLS
$\alpha = 0.25$	0.250	0.250	0.250	0.299	0.245	0.231	0.251	0.254	0.249
	(0.021)	(0.022)	(0.022)	(0.021)	(0.007)	(0.016)	(0.033)	(0.326)	(0.480)
	[0.026]	[0.024]	[0.023]	[0.014]	[0.008]	[0.015]	[0.066]	[1.268]	[0.592]
	{0.905}	{0.886}	{0.863}	{0.170}	{0.000}	{0.000}	{0.051}	{0.735}	{0.200}
$\beta = 0.40$	0.400	0.400	0.400	0.358	0.404	0.416	0.399	0.399	0.394
	(0.018)	(0.018)	(0.019)	(0.017)	(0.006)	(0.013)	(0.021)	(0.269)	(0.443)
	[0.023]	[0.021]	[0.020]	[0.011]	[0.006]	[0.012]	[0.037]	[0.996]	[0.465]
	{0.918}	{0.894}	{0.877}	{0.154}	{0.000}	{0.000}	{0.024}	{0.476}	{0.124}
$\gamma = 0.50$	0.500	0.500	0.500	0.547	0.571	0.566	0.577	0.536	0.359
	(0.017)	(0.017)	(0.018)	(0.012)	(0.006)	(0.009)	(0.414)	(1.034)	(5.357)
	[0.013]	[0.013]	[0.013]	[0.007]	[0.006]	[0.008]	[0.977]	[4.445]	[6.684]
	{0.836}	{0.832}	{0.814}	{0.005}	{0.000}	{0.000}	{0.268}	{0.544}	{0.493}
$\chi = 1.00$	1.000	0.999	0.999	1.007	0.999	0.996	0.999	0.999	1.011
	(0.009)	(0.009)	(0.009)	(0.008)	(0.007)	(0.007)	(0.019)	(0.057)	(0.245)
	[0.007]	[0.006]	[0.006]	[0.004]	[0.007]	[0.007]	[0.042]	[0.201]	[0.279]
	{0.836}	{0.817}	{0.814}	{0.562}	{0.000}	{0.000}	{0.006}	{0.035}	{0.046}
$\xi = 10.0$	9.705	9.679	9.653	6.026	–	–	–	–	–
	(0.868)	(0.833)	(0.935)	(0.795)					
	[0.584]	[0.582]	[0.601]	[0.453]					
	{0.760}	{0.768}	{0.753}	{0.000}					
$\psi = 0.00$	0.000	–0.002	–0.004	0.028	–	–	–	–	–
	(0.068)	(0.069)	(0.070)	(0.104)					
	[0.107]	[0.102]	[0.098]	[0.050]					
	{0.941}	{0.936}	{0.928}	{0.559}					

See the notes accompanying Table 1 for a description of this table's structure.

Table A.3: Monte Carlo Simulations: perturbed parameters (part three)

Target Parameter	Experiment A5: as in experiment 1, but with more network connections ($B = 4$)								
	GMM1	GMM2	GMM3	GMM4	OLS	2SLSa	2SLSb	2SLSc	3SLS
$\alpha = 0.25$	0.248	0.249	0.249	0.294	0.227	0.176	0.249	0.299	0.250
	(0.028)	(0.027)	(0.028)	(0.030)	(0.011)	(0.028)	(0.033)	(1.399)	(0.140)
	[0.032]	[0.033]	[0.034]	[0.023]	[0.012]	[0.023]	[0.058]	[6.954]	[0.772]
	{0.897}	{0.909}	{0.895}	{0.518}	{0.000}	{0.000}	{0.066}	{0.982}	{0.370}
$\beta = 0.40$	0.401	0.401	0.401	0.364	0.419	0.462	0.400	0.358	0.400
	(0.022)	(0.022)	(0.023)	(0.024)	(0.009)	(0.023)	(0.024)	(1.182)	(0.095)
	[0.026]	[0.027]	[0.028]	[0.018]	[0.010]	[0.019]	[0.043]	[5.883]	[0.545]
	{0.894}	{0.905}	{0.897}	{0.502}	{0.000}	{0.000}	{0.023}	{0.868}	{0.226}
$\gamma = 0.50$	0.500	0.500	0.501	0.539	0.576	0.559	0.579	0.519	0.554
	(0.016)	(0.016)	(0.016)	(0.016)	(0.008)	(0.012)	(0.271)	(0.981)	(1.237)
	[0.010]	[0.010]	[0.011]	[0.007]	[0.008]	[0.011]	[0.547]	[6.104]	[7.776]
	{0.768}	{0.764}	{0.779}	{0.064}	{0.000}	{0.000}	{0.325}	{0.906}	{0.664}
$\chi = 1.00$	1.000	1.000	1.000	1.001	0.998	0.993	1.000	1.004	0.999
	(0.008)	(0.008)	(0.008)	(0.008)	(0.007)	(0.008)	(0.011)	(0.107)	(0.043)
	[0.004]	[0.004]	[0.005]	[0.004]	[0.007]	[0.007]	[0.020]	[0.594]	[0.251]
	{0.733}	{0.743}	{0.740}	{0.676}	{0.000}	{0.000}	{0.003}	{0.105}	{0.071}
$\xi = 10.0$	9.888	9.870	9.884	5.764	–	–	–	–	–
	(0.993)	(0.953)	(0.999)	(0.702)					
	[0.590]	[0.597]	[0.632]	[0.383]					
	{0.731}	{0.738}	{0.750}	{0.000}					
$\psi = 0.25$	0.238	0.238	0.233	0.303	–	–	–	–	–
	(0.088)	(0.092)	(0.097)	(0.122)					
	[0.108]	[0.113]	[0.116]	[0.064]					
	{0.933}	{0.932}	{0.926}	{0.639}					

Target Parameter	Experiment A6: as in experiment 1, but with a higher rewiring chance ($b = 0.9$)								
	GMM1	GMM2	GMM3	GMM4	OLS	2SLSa	2SLSb	2SLSc	3SLS
$\alpha = 0.25$	0.252	0.252	0.253	0.291	0.238	0.245	0.249	0.275	–0.244
	(0.022)	(0.021)	(0.023)	(0.037)	(0.009)	(0.014)	(0.023)	(2.166)	(15.70)
	[0.033]	[0.026]	[0.026]	[0.022]	[0.008]	[0.013]	[0.046]	[27.70]	[10.87]
	{0.939}	{0.923}	{0.915}	{0.404}	{0.000}	{0.000}	{0.029}	{0.589}	{0.174}
$\beta = 0.40$	0.398	0.399	0.398	0.367	0.410	0.405	0.401	0.379	0.678
	(0.017)	(0.017)	(0.019)	(0.030)	(0.007)	(0.012)	(0.015)	(1.898)	(8.878)
	[0.027]	[0.021]	[0.021]	[0.018]	[0.006]	[0.011]	[0.033]	[24.29]	[6.173]
	{0.942}	{0.930}	{0.914}	{0.396}	{0.000}	{0.000}	{0.014}	{0.194}	{0.112}
$\gamma = 0.50$	0.496	0.498	0.495	0.555	0.571	0.573	0.563	0.336	3.942
	(0.024)	(0.022)	(0.031)	(0.013)	(0.005)	(0.005)	(0.284)	(4.571)	(107.1)
	[0.046]	[0.032]	[0.031]	[0.008]	[0.006]	[0.006]	[0.637]	[58.45]	[75.73]
	{0.970}	{0.960}	{0.926}	{0.005}	{0.000}	{0.000}	{0.198}	{0.257}	{0.491}
$\chi = 1.00$	1.000	1.000	1.000	1.001	0.998	0.999	1.000	1.007	1.066
	(0.008)	(0.008)	(0.008)	(0.007)	(0.006)	(0.007)	(0.015)	(0.509)	(2.109)
	[0.006]	[0.005]	[0.005]	[0.004]	[0.007]	[0.007]	[0.034]	[6.515]	[1.630]
	{0.796}	{0.777}	{0.769}	{0.713}	{0.000}	{0.000}	{0.007}	{0.012}	{0.044}
$\xi = 10.0$	9.570	9.532	9.364	2.843	–	–	–	–	–
	(1.402)	(1.407)	(1.720)	(1.936)					
	[2.305]	[1.675]	[1.650]	[0.873]					
	{0.968}	{0.967}	{0.945}	{0.009}					
$\psi = 0.25$	0.252	0.250	0.252	0.301	–	–	–	–	–
	(0.069)	(0.066)	(0.072)	(0.072)					
	[0.094]	[0.072]	[0.070]	[0.045]					
	{0.878}	{0.883}	{0.830}	{0.612}					

See the notes accompanying Table 1 for a description of this table's structure.

Table A.4: Monte Carlo Simulations: alternative spatial matrices

Target Parameter	Experiment A7: $\mathbf{H} = \mathbf{C} = \mathbf{I} + \mathbf{G}$; $\mathbf{C}_e^* = \mathbf{I} + \mathbf{G} + \mathbf{G}^2$								
	GMM1	GMM2	GMM3	GMM4	OLS	2SLSa	2SLSb	2SLSc	3SLS
$\alpha = 0.25$	0.249	0.249	0.248	0.119	0.246	0.232	0.250	0.340	0.256
	(0.027)	(0.028)	(0.029)	(0.074)	(0.006)	(0.026)	(0.023)	(2.791)	(0.148)
	[0.018]	[0.018]	[0.018]	[0.037]	[0.008]	[0.017]	[0.047]	[84.82]	[0.589]
	{0.812}	{0.796}	{0.794}	{0.232}	{0.000}	{0.000}	{0.045}	{0.952}	{0.258}
$\beta = 0.40$	0.400	0.400	0.401	0.504	0.404	0.416	0.400	0.307	0.395
	(0.022)	(0.022)	(0.024)	(0.059)	(0.006)	(0.023)	(0.016)	(2.588)	(0.135)
	[0.015]	[0.015]	[0.015]	[0.029]	[0.006]	[0.014]	[0.031]	[74.19]	[0.441]
	{0.817}	{0.796}	{0.788}	{0.234}	{0.000}	{0.000}	{0.014}	{0.850}	{0.168}
$\gamma = 0.50$	0.504	0.506	0.506	0.444	0.585	0.577	0.570	0.197	0.542
	(0.025)	(0.026)	(0.028)	(0.103)	(0.006)	(0.016)	(0.197)	(8.082)	(1.821)
	[0.022]	[0.020]	[0.022]	[0.039]	[0.006]	[0.010]	[0.463]	[716.1]	[6.484]
	{0.890}	{0.855}	{0.842}	{0.435}	{0.000}	{0.000}	{0.218}	{0.870}	{0.519}
$\chi = 1.00$	1.000	1.000	1.001	1.001	0.999	0.997	1.000	1.021	1.000
	(0.008)	(0.008)	(0.008)	(0.014)	(0.006)	(0.009)	(0.011)	(0.876)	(0.030)
	[0.004]	[0.004]	[0.004]	[0.011]	[0.006]	[0.007]	[0.023]	[25.76]	[0.181]
	{0.717}	{0.709}	{0.713}	{0.705}	{0.000}	{0.000}	{0.004}	{0.263}	{0.066}
$\xi = 10.0$	9.961	9.962	9.975	1.020	–	–	–	–	–
	(0.953)	(0.962)	(1.010)	(1.854)					
	[0.700]	[0.621]	[0.678]	[0.501]					
	{0.808}	{0.775}	{0.792}	{0.002}					
$\psi = 0.25$	0.224	0.216	0.212	0.278	–	–	–	–	–
	(0.108)	(0.112)	(0.117)	(0.301)					
	[0.092]	[0.084]	[0.091]	[0.119]					
	{0.884}	{0.841}	{0.836}	{0.523}					

Target Parameter	Experiment A8: $\mathbf{H}$: 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	2SLSc	3SLS
$\alpha = 0.25$	0.248	0.250	0.249	0.151	0.251	0.247	0.250	0.272	0.230
	(0.029)	(0.031)	(0.033)	(0.112)	(0.005)	(0.033)	(0.024)	(0.971)	(0.698)
	[0.020]	[0.021]	[0.021]	[0.024]	[0.007]	[0.017]	[0.052]	[50.24]	[1.591]
	{0.832}	{0.812}	{0.784}	{0.369}	{0.000}	{0.000}	{0.042}	{0.952}	{0.237}
$\beta = 0.40$	0.402	0.400	0.401	0.478	0.399	0.402	0.400	0.381	0.414
	(0.024)	(0.026)	(0.027)	(0.090)	(0.006)	(0.029)	(0.016)	(0.767)	(0.459)
	[0.017]	[0.017]	[0.017]	[0.019]	[0.006]	[0.015]	[0.030]	[35.26]	[1.048]
	{0.830}	{0.815}	{0.791}	{0.367}	{0.000}	{0.000}	{0.016}	{0.865}	{0.153}
$\gamma = 0.50$	0.502	0.504	0.504	0.504	0.603	0.601	0.589	0.418	0.437
	(0.024)	(0.024)	(0.025)	(0.126)	(0.006)	(0.021)	(0.221)	(2.590)	(4.711)
	[0.014]	[0.014]	[0.015]	[0.031]	[0.006]	[0.011]	[0.518]	[242.0]	[13.16]
	{0.780}	{0.740}	{0.739}	{0.241}	{0.000}	{0.000}	{0.207}	{0.963}	{0.532}
$\chi = 1.00$	1.000	1.000	1.000	1.003	1.000	1.000	1.000	1.003	1.004
	(0.008)	(0.008)	(0.008)	(0.018)	(0.006)	(0.010)	(0.010)	(0.130)	(0.141)
	[0.004]	[0.004]	[0.004]	[0.011]	[0.006]	[0.006]	[0.021]	[5.501]	[0.440]
	{0.707}	{0.713}	{0.707}	{0.708}	{0.000}	{0.000}	{0.005}	{0.198}	{0.074}
$\xi = 10.0$	10.010	9.986	10.013	–0.106	–	–	–	–	–
	(0.937)	(0.959)	(0.937)	(1.778)					
	[0.498]	[0.449]	[0.495]	[0.281]					
	{0.732}	{0.662}	{0.722}	{0.000}					
$\psi = 0.25$	0.231	0.230	0.225	0.035	–	–	–	–	–
	(0.084)	(0.081)	(0.086)	(0.388)					
	[0.044]	[0.039]	[0.043]	[0.103]					
	{0.700}	{0.685}	{0.688}	{0.331}					

See the notes accompanying Table 1 for a description of this table's structure.

Addendum E Data transformations: results

Table A.5 below reports results from estimates of a variation of model (5), adapted to the specification (20) from our empirical application (with $\delta = 0$), where matrix $\mathbf{B}$ is constructed in such a way that, for some given choice of the characteristic matrix $\mathbf{C}$, it is $\mathbf{BC} = \mathbf{0}$. In particular, we set $\mathbf{B} = \mathbf{I} - \mathbf{CC}^{+}$, where $\mathbf{C}^{+}$ is the Moore-Penrose pseudoinverse of $\mathbf{C}$. The results are characterized by large standard errors and point estimates that are at times implausible, especially for the social effects parameter β.

Table A.5: Empirical estimates: data transformations

	Outcome variable: $y_i^{(1)}$ (later career GPA)				Outcome variable: $y_i^{(2)}$ (economics major)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
β	0.348***	–2.240	2.426***	0.219	0.649	–1.388	10.94***	–0.078
	(0.134)	(4.939)	(0.229)	(0.174)	(0.402)	(3.003)	(1.656)	(0.439)
γ	11.38***	–6.629	10.49***	10.43***	0.461***	–0.123	2.948***	0.681***
	(0.521)	(13.90)	(0.927)	(0.606)	(0.101)	(2.995)	(0.738)	(0.120)
χ_{fe}	0.190*	1.814*	2.187***	0.477***	0.010	–0.235	0.659***	–0.055***
	(0.100)	(1.024)	(0.195)	(0.123)	(0.020)	(0.177)	(0.207)	(0.020)
$\mathbf{C}$	$\widetilde{\mathbf{C}}_d$	$\widetilde{\mathbf{C}}_{h1}$	$\widetilde{\mathbf{C}}_{h2}$	$\mathbf{I} + \mathbf{G}$	$\widetilde{\mathbf{C}}_d$	$\widetilde{\mathbf{C}}_{h1}$	$\widetilde{\mathbf{C}}_{h2}$	$\mathbf{I} + \mathbf{G}$
RFE	NO	NO	YES	NO	NO	NO	YES	NO
Obs.	1,141	1,141	1,141	1,141	1,141	1,141	1,141	1,141

Notes. Each column in this table reports IV/2SLS estimates of a *transformed* version of model (20), for both outcome variables as indicated in the header. Both sides of the model equation (in vectoral form), are pre-multiplied by a matrix $\mathbf{B}$ such that, for a given choice of matrix $\mathbf{C}$ as specified in each column, $\mathbf{BC} = \mathbf{0}$, resulting in a variation of model (5). Specifically, $\mathbf{B} = \mathbf{I} - \mathbf{CC}^{+}$, where $\mathbf{C}^{+}$ is the Moore-Penrose pseudoinverse of $\mathbf{C}$. All estimates incorporate the restriction $\delta = 0$ (no exogenous effects). All estimates are based upon orthogonality conditions between the transformed error term and: (i) a constant vector; (ii) the w_{ki} controls; (iii) two "instruments" (IVs) $\mathbf{Bx}$ and $\mathbf{GBx}$, where $\mathbf{x}$ stacks all high-school final grades. Point estimates for parameters other than β, γ, χ_{fe} are omitted. Heteroschedasticity-consistent standard errors are in parentheses. Asterisk series: *, **, and ***; denote statistical significance at the 10, 5 and 1 per cent level, respectively. Obs.: Observations.