Estimation of Endogenous Network Externalities

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

[1]Banca d'Italia
[2]IMT School for Advanced Studies Lucca

February 2016

Abstract

The estimation of spillover and peer effects presents challenges that are still unsolved. In particular, even if separate algebraic identification of the endogenous and exogenous effects is possible, these might be contaminated by the simultaneous dependence of outcomes and covariates upon unobserved factors. In this paper we propose a methodology to consistently estimate all the parameters of a linear-in-means model in presence of endogeneity, with non perfectly overlapping groups and when individual "effort" is observable. We show how the number of necessary "external" instruments can be reduced to one by exploiting theoretical restrictions. Specifically, spillover effects are identified by appropriate restrictions on the equilibrium excess covariance between nodes that are sufficiently distant in a network. All but one of the other parameters, which correspond to the elasticities of inputs in production functions, are identified via the use of "input ratios" as instruments.

JEL Classification Codes: C31, C57, D24, D85

Keywords: Spillovers, Networks, Production Functions, Covariance Restriction

Simple economic analysis predicts that students with better abilities have, all else equal, stronger incentives to put effort towards school grades. Likewise, firms embedding more original ideas and knowledge will face higher returns from R&D. However, both students and firms operate in social contexts characterized by spillover effects. Since pupils reciprocally influence each other, the innate skills of a student may also alter the incentives of his or her peers. Similarly, the specificities of one company affect the return of R&D of its neighboring firms in the technological space. In both cases individual characteristics, individual behavior and group behavior are intertwined in a continuous causal chain. Within this framework the sharp distinction between *exogenous* and *endogenous* spillover effects (Manski, 1993), originating in the sociological literature, appears less relevant. The existence of unobserved factors, such as ability or embedded knowledge, that influence group behavior is arguably a more severe threat to the econometric identification of spillover effects than how simultaneity or *reflection* is typically thought to be.

This paper makes two related contributions, one theoretical and one econometric. On the theoretical side, we analyze a game of spillover effects, played in an unrestricted network, within a production function framework. In this game, a single variable that we generally call *effort* induces externalities over players' connections. We discuss how a production function approach is equally suitable to the analysis of circumstances like peer effects in the classroom. We show that, because of complementarities between effort and other individual characteristics, this model is observationally equivalent to a linear in means model with nonzero contextual effects when effort is not observed. Hence, we argue that the main identification problem in the search of spillover effects is not the algebraic separation of the exogenous and endogenous types of effect, but the inability of the econometrician to observe actual individual behavior (effort) and its consequences on other agents' outcomes. In such a case, by estimating linear in means models the econometrician may run the risk of mistaking interactions between individual characteristics and effort for contextual effects.

Econometric identification of production functions is notoriously difficult. In the context of our model, even if effort were actually observed, it would be threatened by the presence of unobserved individual characteristics. Our econometric contribution is a methodology for the estimation of production functions with spillover effects, by resorting to only one external exclusion restriction. To this end, we exploit specific theoretical predictions of our model. Spillovers effects, in particular, are identified via conditional covariance restrictions relative to the equilibrium outcomes of sufficiently distant nodes in the network. Importantly, this approach can be adapted to allow for general correlation patterns in the unobservables across the network, also known as "common shocks." All other parameters but one are instead identified via natural instruments, the *input ratios*, that are based upon the model's first order conditions. This allows to reduce the need for external instruments to one, or even zero in presence of panel data and with proper timing assumptions. To the best of our knowledge, this intuition is new in the literature on the estimation of production functions.

This paper relates to, and combines, two different strands in the literature. We extend the studies on the estimation of spillover and peer effects by introducing a model that allows, relative to others (Calvó-Armengol et al., 2009; Blume et al., 2015) for more general patterns of interactions between individual variables. Furthermore, we show how higher order moment restrictions, first employed by Graham (2008) for the identification of spillover effects in a way eventually extended by Pereda Fernández (2015), can be effective in addressing the problem of "common shocks."[1] In addition, we contribute to the literature on the estimation of production functions. Most methods that have been proposed to account for "unobserved shocks" require panel data, and are motivated on some specific timing assumptions about the adjustment of state variables. Some methods are semiparametric (Olley and Pakes, 1996; Levinsohn and Petrin, 2003), while others are more structural (Doraszelski and Jaumandreu, 2013). We propose a simple approach using instruments motivated by the model's first order conditions, which are combined together with other exclusion restrictions – as well as with covariance restrictions for the identification of spillover effects – in a single GMM problem. This method also works with cross-sectional data.

This paper is organized as follows. In Section 1 we introduce the analytical model. We discuss the model's predictions about conditional covariance restrictions, as well as its observational equivalence with linear in means models in which the effort variable is missing. In Section 2 we develop the econometrics. We first discuss identification and estimation with cross sectional data and no common shocks, to eventually accomodate both the latter and panel data. In Section 3 we show the results of a Monte Carlo simulation of our estimation strategy, while comparing different procedures. Finally, in Section 4 we spell out plans for the completion and extensions of this work.

[1]In a well known critique of the whole peer effects literature, Angrist (2014) considers common shocks to be the prime threat for the identification of social externalities.

1 Analytical Framework

In this section we characterize the general game-theoretic model of social interactions on which this paper is based. We provide results that are relevant for the model's empirical counterpart, which we are interested in estimating. This section is divided in three parts. In the first one we outline the setup of the model. In the second part we discuss the equilibrium properties of the model, in particular covariance restrictions that are useful for the identification of the model's parameters. In the third part we compare the model and its empirical implications to those from previous studies.

1.1 Model's Setup

We consider a general game of complete information played on a network $\langle N, \mathbf{G} \rangle$. This network is composed by N players whose interactions are summarized by an adjacency matrix $\mathbf{G}$.[2] Each entry $g_{ij} \in [0,1]$ of this matrix denotes the strength of the relationship directed from player j to player i, with $g_{ii} = 0$. Throughout the paper we assume the network to be *undirected*: that is, relationships are symmetric between all players ($g_{ij} = g_{ji}$ for all i, j). For any given economic context, players are heterogeneous agents interested in achieving the best possible outcome conditional on taking costly actions. Their objective function features network externalities: a choice variable that we call *effort* reciprocally influences the outcomes of players who are linked in the network. This introduces strategic interdependence in players' choices.

Specifically, players get positive utility from some combination of: *i.* the individual *effort* choice $S_i \in \mathbb{R}_+$; *ii.* all other players' effort $\mathbf{s}_{-i} = (S_i, \dots, S_{i-1}, S_{i+1}, \dots, S_N)'$, with $\mathbf{s} = \left(S_i, \mathbf{s}'_{-i}\right)' \in \mathbb{R}^N_+$; *iii.* some "ordinary" *choice variables*, which are distinct from effort and are denoted with the vector $\mathbf{x}_i = (X_{i1}, \dots, X_{iK})' \in \mathbb{R}^K_+$; *iv.* idiosyncratic *invariant characteristics*, that are represented by the vector $\mathbf{z}_i = (Z_{i1}, \dots, Z_{iH})' \in \mathbb{R}^H_+$; *v.* a *luck* component, denoted by an i.i.d. random variable $\varepsilon_i \in \mathbb{R}$ with mean zero ($\mathbb{E}[\varepsilon_i] = 0$). All these factors contribute to the "revenue" (positive) component Y_i of utility, and their interaction is expressed by a Cobb-Douglas function.

$$Y_i(\mathbf{s}, \mathbf{x}_i, \mathbf{z}_i) = \exp(\alpha + \varepsilon_i) \left(\prod_{k=1}^{K} X_{ik}^{\beta_k} \right) S_i^{\gamma} \left(\prod_{j \neq i} S_j^{g_{ij}} \right)^{\delta} \left(\prod_{h=1}^{H} Z_{ih}^{\zeta_h} \right) \tag{1}$$

[2]Some authors prefer the use of the term *sociomatrix* to denote this algebraic object.

The set of parameters $\boldsymbol{\beta} = (\beta_1, \dots, \beta_K)'$, γ and $\boldsymbol{\zeta} = (\zeta_1, \dots, \zeta_H)'$ in (1) represent, respectively, the elasticities of ordinary choice variables, individual effort and invariant factors on Y_i. Parameter δ denotes instead the overall strength of the "spillovers" that the effort of other players, who are connected to i in the network, exert on the latter. However, the intensity of specific bidirectional spillovers flowing between any two players i and j is proportional to both the strength of their relationship g_{ij} and to δ. Moreover, parameter α is a scaling factor. The description of the model is completed by including a cost expression, which we assume to be linear for all choice variables. In particular, the player-specific vector $\mathbf{p}_i = (P_{i1}, \dots, P_{iK})'$ collects the marginal and average costs faced by player i for every additional unit of any ordinary choice variable X_{ik}. Similarly, Q_i is the analogous cost associated to the effort variable. The resulting utility or "profit" of each player is given as follows.

$$U_i(\mathbf{x}_i, \mathbf{s}, \mathbf{t}_i) = \exp(\alpha + \varepsilon_i) \left(\prod_{k=1}^{K} X_{ik}^{\beta_k} \right) S_i^{\gamma} \left(\prod_{j \neq i} S_j^{g_{ij}} \right)^{\delta} \left(\prod_{h=1}^{H} Z_{ih}^{\zeta_h} \right) - \sum_{k=1}^{K} P_{ik} X_{ik} - Q_i S_i \tag{2}$$

Notice that players heterogeneity is completely specified by the invariant characteristics, the cost factors and the luck component. We group these factors, except luck, into what we call with some abuse of terminology "types" $\mathbf{t}_i = (\mathbf{z}_i, \mathbf{p}_i, Q_i) \in \mathbb{R}_+^{K+H+1}$.

Expression (2) looks familiar as a firm profit function featuring R&D spillovers, where $\mathbf{z}_i$ represent idiosyncratic properties of firms, $\mathbf{x}_i$ are conventional (variable) inputs like capital and labor, while S_i is some measure of R&D investment. However, it can be equally useful for studying other types of social interactions, such as peer effects in the classroom. In that context $\mathbf{z}_i$ would represent specific and *invariant* pupil characteristics (like parental background, socio-economic status, ability), $\mathbf{x}_i$ would refer to *variable* and *costly* activities that may influence grades, such as extramural tutoring or training for standardized tests, and finally S_i is the effort independently made by each student towards tests and grades, e.g. on assigned homeworks. Notice that (2) is concave in each of its choice variables.

In the model's objective function, the only variable causing direct spillovers is effort S_i. Since an increase in peer effort is an incentive to increase individual effort, effort S_i is itself a strategic complement. This is a common feature of other models of peer effects in the classroom (Calvó-Armengol et al., 2009; Blume et al., 2015; Pereda Fernández, 2015). The main difference with the cited works is the assumption that the outcomes of interest, say grades, are determined by a Cobb-Douglas function. Because of this assumption, the model features cross-elasticities between the effort variable S_i and both other choice variables X_{ik} and invariant characteristics Z_{ih}. This improves on the realism of the model: it accommodates circumstances whereby more skilled pupils have stronger incentives to put effort towards grades. It similarly allows for those who receive, say, extramural tutoring, to be more encouraged to do well.

These typical real world instances are not allowed in the other theoretical frameworks of peer effects. In the model by Calvó-Armengol et al., on the one hand, there are two different "effort" variables. One interacts with student-specific covariates, but induces no spillovers. The other one does cause spillovers, as it enters the objective function multiplicatively together with the analogous choice variables of peers, but does not interact with covariates. This artificial distinction, while making the model more tractable, has no clear underlying economic rationale and restricts its realism. The model by Blume et al., on the other hand, features only one effort variable which interacts with covariates. In their model, strategic complementarities enter students' utility via a "loss" function of individual effort's distance from average peer effort.[3] However, the ultimate outcomes implicitly and deterministically depend on the sole effort, which again trades off realism for tractability.

1.2 Equilibrium Properties

The timing of the game is as follows: 1) Nature draws all players' types $\mathbf{t}_i$, which are common knowledge; 2) players make their choices of S_i and $(X_{i1}, \ldots, X_{iK})$; 3) Nature draws all players' luck component ε_i, unknown *ex-ante*; 4) payoffs are paid out. Because of the luck component, we apply the Bayes-Nash solution concept in order to identify equilibria of the game. Noticeably, the luck term does not affect the model substantively. Without it, the solution could be expressed as a more straightforward Nash equilibrium, and it would read almost identically. However, the luck component – in the form of an error term – is necessary for an appropriate empirical counterpart of the present theoretical setup. Furthermore, using the Bayes-Nash solution concept allows for easier comparison of the equilibrium properties with those obtained with altered assumptions, e.g. when types $\mathbf{t}_i$ are not common knowledge.

[3]This is motivated by the need of peers to conform to "social norms" of typical behavior, and not on technological complementarities in effort like in Calvó-Armengol et al. and in the present model. Their assumptions rule out multiple equilibria without restrictions on parameter values.

The equilibrium relationships are most easily expressed with logarithmic variables. We denote them with lowercase italic letters, unlike uppercase italic variables in levels. All vectors and matrices collecting logged quantities in all their elements are denoted, respectively, by lowercase and uppercase roman (that is, not bold) letters.[4] A Bayes-Nash equilibrium, in particular, is the collection of all players' log-effort choices vector s^*, and of K vectors x_k^* for the other "logged" choice variables, such that the utility of all players is maximized *in expectation* relative to alternative individual choices.

Proposition 1. *One Bayes-Nash equilibrium strategy profile can be expressed as*

$$\mathrm{s}^* = \frac{1}{1-\gamma-\sum_{k=1}^{K}\beta_k}\left(\mathbf{I}-\frac{\delta}{1-\gamma-\sum_{k=1}^{K}\beta_k}\mathbf{G}\right)^{-1}(\bar{\mathrm{s}}+\mathrm{Z}\boldsymbol{\zeta}) \tag{3}$$

$$\mathrm{x}_k^* = \bar{\mathrm{x}}_k + \mathrm{s}^* \quad \forall k = 1,\dots,K \tag{4}$$

where $\bar{\mathrm{x}}_1,\dots,\bar{\mathrm{x}}_K$ *and* $\bar{\mathrm{s}}$ *are N-dimensional vectors: their i-th elements read as follows.*

$$\bar{x}_{ik} = \log\beta_k - \log\gamma - (p_{ik}-q_i) \quad \forall k = 1,\dots,K$$

$$\bar{s}_i = \alpha + \sum_{k=1}^{K}\beta_k\bar{x}_{ik} + \log\gamma - q_i + \log\mathbb{E}\left[\exp(\varepsilon_i)\right]$$

Furthermore, if $\delta+\gamma+\sum_{k=1}^{K}\beta_k<1$ *and* $\mathbf{G}\iota<\iota$ *this equilibrium is unique.*

Proposition (1) extends standard results from the literature on network games to the setup at hand. In particular, equation (3) expresses the equilibrium logarithmic "effort" as a function of the individual types of all (indirectly connected) members of the network, as given by the vector $\bar{\mathrm{s}}+\mathrm{Z}\boldsymbol{\zeta}$. Player's i choice of equilibrium effort is also affected by the types of players who are not directly connected to him, as they indirectly influence other intermediate nodes in the network. The relative influence of a player on others is given by the individual measures of Bonacich centrality with parameter $\vartheta \equiv \delta/\left(1-\gamma-\sum_{k=1}^{K}\beta_k\right)$.[5] This result is due to strategic complementarities, and is first encountered in the cited work by Calvó-Armengol et al.

[4]Hence, $\mathrm{y} = (y_1,\dots,y_N)'$, $\mathrm{s} = (s_1,\dots,s_N)'$, $\mathrm{x}_k = (x_{1k},\dots,x_{Nk})'$, $\mathrm{z}_h = (z_{1h},\dots,z_{Nh})'$ and, similarly, $\mathrm{X} = \left(\mathrm{x}_1' \ \dots \ \mathrm{x}_K'\right)'$ and $\mathrm{Z} = \left(\mathrm{z}_1' \ \dots \ \mathrm{z}_H'\right)'$.

[5]Bonacich centrality is a recursive, descriptive measure of how much a node is "central" in the network, in the sense of occupying a prominent position in the network and appearing frequently in paths. Being recursive, Bonacich centrality is most easily (albeit somehow inaccurately) summarized as a measure of how much a node is connected to other central nodes.

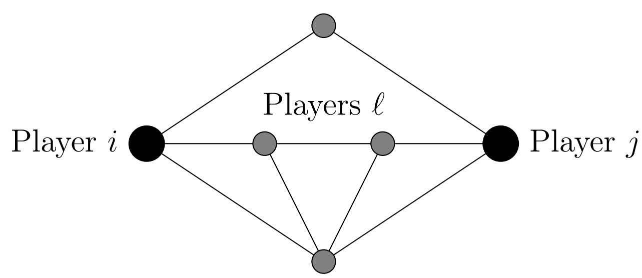
The set of equations expressed by (4) follow from the constrained maximization of Cobb-Douglas functions. What is relevant to consider in this context is that since the optimal values of "ordinary" choice variables are themselves functions of s^*, they are also affected by network externalities and are therefore a function of other players' types. Finally, the equilibrium uniqueness result under $\vartheta < 1$ (ruling out "explosive" solutions) is analogous to others from games with strategic complementarities.

The next result, along with the associated corollaries, is crucial for the identification strategy proposed in the paper.

Proposition 2. *Suppose that the vector of individual types $\mathbf{t}_i$ is independently drawn from a generic probability distribution function $\mathcal{F}$. Then, for any two unconnected players i, j, $i \neq j$ with $g_{ij} = 0$, in equilibrium it holds that*

$$\mathbb{C}\text{ov}\left(S_i, S_j \middle| \{S_\ell \in \mathbf{s} : g_{i\ell} \neq 0 \vee g_{j\ell} \neq 0\}\right) = 0 \tag{5}$$

that is, conditional on the optimal effort choices of all direct links of both players i and j, the covariance between the respective equilibrium effort choices is zero.



Graph 1: Illustration of Proposition 2

The intuition behind this result is illustrated by Graph 1. It represents a generic subgraph featuring two players i and j, that are indirectly connected through other players indexed ℓ. In equilibrium, their effort choices are reciprocally but indirectly influenced, because they both affect the equilibrium effort of the ℓ players. Thus, S_i and S_j are both functions of $(\mathbf{t}_i, \mathbf{t}_j)$, and in equilibrium their covariance is nonzero. However, by conditioning upon the equilibrium effort choices of the ℓ players, that source of mutual dependence is removed. Hence, if $(\mathbf{t}_i, \mathbf{t}_j)$ are independent, the conditional covariance expressed in (5) is restricted to be zero.[6]

What follows are two straightforward extensions of Proposition 2.

[6]By using more complex notation, expression (5) can be alternatively formulated by conditioning

Corollary 1. *Under the hypotheses of Proposition 2, in equilibrium:*

$$\mathbb{C}\text{ov}\left(\mathbf{x}_i, \mathbf{x}_j \middle| \{S_\ell \in \mathbf{s} : g_{i\ell} \neq 0 \vee g_{j\ell} \neq 0\}\right) = 0$$

that is, the ordinary *choice variables of players i and j are conditionally independent.*

The intuition of this result is simple: individual equilibrium effort is a sufficient statistic of network effects in the model. Since "ordinary" choice variables, in equilibrium, are functions of individual effort and individual characteristics (types) as per the set of equations (4), one can extend the result from Proposition 2 to the X_{ik} variables, as long as types are independent.

Corollary 2. *Under the hypotheses of Proposition 2, in equilibrium:*

$$\mathbb{C}\text{ov}\left(Y_i, Y_j \middle| \{S_\ell \in \mathbf{s} : g_{i\ell} \neq 0 \vee g_{j\ell} \neq 0\}\right) = 0$$

that is, the outcomes of players i and j are conditionally independent.

This corollary follows straightforwardly from (1), Proposition 2 and Corollary 1. It constitutes the theoretical motivation for the covariance restriction that we introduce in the paper's econometric model in order to identify δ, when this is otherwise impossible because of endogeneity problems (e.g. due to unobserved, omitted variables).

1.3 Discussion

In principle, the model can be generalized to allow for multiple choice variables to be sources of spillovers. In the case of firms, for example, physical capital may cause negative externalities in the form of pollution. In the case of students, instead, extramural activities like sports might have additional social benefits by, say, encouraging healthy behavior. The empirical implications of the model would be very similar, and are more easily illustrated with only one "spillover" variable. Furthermore, unlike much of the literature on peer effects in education, the model presented in this paper

on a *minimal set* of nodes that indirectly link i with j. There is generally a multiplicity of such minimal sets. For example, out of the four grey, ℓ nodes from Graph 1, one can construct two minimal sets by excluding either node located in the middle of the straight, horizontal path connecting i with j: in fact, conditioning on one of the two makes conditioning on the other redundant. This also shows that it is not necessary to condition on all immediate neighbors of i or j for the covariance restriction to hold. However, we prefer to express it as stated in the text, because it more closely resembles the resulting moment condition that we employ in the econometric model.

does not allow for "contextual effects." These are the *direct* externalities that are caused by invariant characteristics, like the socio-economic status of peers. Reduced form models of social interactions routinely incorporate contextual effects with an implicit "technological" interpretation.

We do not rule out *a priori* the existence of multiple sources of spillovers or contextual effects. However, we exclude them from the model in order to illustrate that they are not empirically identified when "effort" S_i is simultaneously unobserved and complementary to all covariates. In order to appreciate this, consider the workhorse empirical model that is typically employed in the literature on peer effects: a generalized linear-in-means model (Bramoullé, Djebbari, and Fortin, 2009)[7]:

$$\mathrm{y} = \eta \iota + \lambda \mathbf{G} \mathrm{y} + \mathrm{Z} \boldsymbol{\mu} + \mathbf{G} \mathrm{Z} \boldsymbol{\nu} + \boldsymbol{\epsilon} \tag{6}$$

where we denote the set of individual-level covariates with the matrix Z, in analogy with the "invariant characteristics" of our model. However, with the opposite aim of avoiding confusion with the latter we also use, in order to label parameters, a set of greek letters that is unusual for the literature on peer effects.[8]

In equation (6) η is a constant term, $\boldsymbol{\mu}$ are the direct effects of covariates on the outcome y, while $\boldsymbol{\nu}$ denotes contextual effects, also called "exogenous" spillover effects as per Manski's (1993) classification. Under the same taxonomy, instead, coefficient λ denotes the so-called "endogenous" effect, which represents peers' reciprocal influences on equilibrium behavior. The prevalent microfoundation of the endogenous effect is a theoretical model which assumes that some unobserved variable, call it effort, exerts spillovers across peers.[9] Since effort is unobserved, the endogenous effect can only be recovered through the reduced-form dependence of outcomes across the network. This gives potentially rise to the "reflection" problem: a particular instance of simultaneity first identified by Manski.

In the literature on peer effects, one does not typically distinguish different types of covariates. However, the linear-in-means model can naturally accomodate both

[7] In a linear-in-means model applied upon a network structure, the adjacency matrix $\mathbf{G}$ is typically row-normalized, although this affects little else beyond the interpretation of coefficients.

[8] Covariates are typically denoted with the letter x ($\mathbf{x}, \mathrm{X}$) in this literature. Moreover, the set of parameters $(\eta, \lambda, \boldsymbol{\mu}, \boldsymbol{\nu})$ corresponds to the set $(\alpha, \beta, \boldsymbol{\gamma}, \boldsymbol{\delta})$ from most studies, hence model (6) would read, say in Bramoullé et al., similarly to $\mathrm{y} = \alpha \iota + \beta \mathbf{G} \mathrm{y} + \mathrm{X}' \boldsymbol{\gamma} + \mathbf{G} \mathrm{X}' \boldsymbol{\delta} + \boldsymbol{\epsilon}$.

[9] Prime examples are the already cited works by Calvó-Armengol et al. (2009) and Blume et al. (2015), or appropriate variations of them.

the "invariant characteristics" and the "ordinary choice variables" we discussed above, together with the associated contextual effects. Such an augmented model would read as

$$\mathrm{y} = \eta\boldsymbol{\iota} + \lambda\mathbf{G}\mathrm{y} + \mathrm{X}\boldsymbol{\mu}_x + \mathbf{G}\mathrm{X}\boldsymbol{\nu}_x + \mathrm{Z}\boldsymbol{\mu}_z + \mathbf{G}\mathrm{Z}\boldsymbol{\nu}_z + \boldsymbol{\epsilon} \tag{7}$$

where subscripts x and z are useful to distinguish the parameter sets $\boldsymbol{\mu}, \boldsymbol{\nu}$ of direct and "contextual" effects by covariate type. Suppose that the vector of residuals $\boldsymbol{\epsilon}$ is orthogonal to both the covariates and the network structure, that is $\mathbb{E}\left[\boldsymbol{\epsilon} \middle| \mathbf{G}, \mathrm{X}, \mathrm{Z}\right] = 0$. Then, model (7) is known to be identified, in the sense that the reflection problem is solved, as long as $\mathbf{I}$, $\mathbf{G}$ and $\mathbf{G}^2$ are linearly independent (Bramoullé et al., 2009). With the following result, however, we show that there is a more fundamental identification problem that is related to the economic assumptions about the context of interest.

Proposition 3. *Consider a set of data* $(\mathrm{y}, \mathrm{X}, \mathrm{Z})$ *observed in some network* $\langle N, \boldsymbol{G}\rangle$. *There are always two specific restrictions from two different models that are observationally equivalent in generating such data:*

1. *a combination of:* i. *a restriction on the set of parameters* $(\eta, \lambda, \boldsymbol{\mu}_x, \boldsymbol{\nu}_x, \boldsymbol{\mu}_z, \boldsymbol{\nu}_z)$ *with* $\lambda \neq 0$, $\boldsymbol{\nu}_x \neq 0$, $\boldsymbol{\nu}_z \neq 0$; *and* ii. *a realization of the vector* $\boldsymbol{\epsilon}$, *that together relate to* $(\mathrm{y}, \mathrm{X}, \mathrm{Z})$ *via the relationships expressed by the statistical model* (7);

2. *a combination of:* iii. *a restriction on the set of parameters* $(\alpha, \boldsymbol{\beta}, \gamma, \delta, \boldsymbol{\zeta})$, iv. *a sequence of types* $\mathbf{t}_1, \dots, \mathbf{t}_N$, v. *a sequence of* effort *values* s *satisfying equations* (3-4), *and* vi. *a realization of the vector* $\boldsymbol{\varepsilon}$, *that together relate to* $(\mathrm{y}, \mathrm{X}, \mathrm{Z})$ *via the following model.*

$$\mathrm{y} = \alpha\boldsymbol{\iota} + \mathrm{X}\boldsymbol{\beta} + \gamma\mathrm{s} + \delta\mathbf{G}\mathrm{s} + \mathrm{Z}\boldsymbol{\zeta} + \boldsymbol{\varepsilon} \tag{8}$$

This result formally states the idea that statistical "contextual effects" ($\boldsymbol{\nu} \neq 0$) might just be the byproduct of more fundamental and complex behavioral patterns of the economic agents under analysis, rather than "technological" determinants of individual outcomes. In particular, if individual characteristics influence individual behavior in the form of effort, which in turn affects peer effort and viceversa, individual outcomes end up being *indirectly correlated* to peer characteristics. Hence, a naïve estimation of model (6) – even if the "endogenous" effect is algebraically identified – would suffer from an omitted variable bias problem. This affects in particular the estimation of $\boldsymbol{\nu}$: some parameters would be estimated different from zero even if contextual effects do not actually exist in reality.

In more practical terms, Proposition 3 should be seen as an encouragement, for the sake of research on network/peer effects and social interactions, to shift the focus of empirical analyses from the algebraic identification of different types of "effects" to the careful observation and economic analysis of group behavior. More work should be devoted to the collection of data, and in particular to the measurement of individual effort – say, the time devoted by pupils on their homework. In the empirical application of this paper we demonstrate the relevance of including direct measures of "effort" in models of network effects. We consider an important economic setting where this variable is readily available: the analysis of R&D spillovers between firms.

2 Econometrics

In this section we develop the econometric methodology for the estimation our model under the condition that the "effort" spilling-over variable is observable. We focus on a case in which identification is both not straightforward and of practical interest: that is, when some individual-specific covariates (Z_{ih} in our model) are unobserved. This section is splitted in three parts: one about identification, another one about estimation, and a final one on possible extensions of the model.

2.1 Identification

If one assumes (8) to be the true model and $(\mathrm{X}, \mathrm{Z}, \mathrm{s})$ are all observed, identification is trivial provided that the error term is exogenous. In fact, it is easy to see that $\mathbb{E}[\boldsymbol{\varepsilon} \mid \mathrm{X}, \mathrm{Z}, \mathrm{s}] = 0$ given the assumptions and timing of the model.[10] The error term $\boldsymbol{\varepsilon}$ is interpreted as the combination of all unexpected factors affecting agents' outcomes y, or equivalently as luck. Econometrically, it also incorporates measurement error of the outcome y. Model (8) can be easily estimated via OLS.

Things are different if some of the variables in $(\mathrm{X}, \mathrm{Z}, \mathrm{s})$ are not observed. If some of the ordinary choice variables X are missing, the model is not identified: OLS would estimate all parameters inconsistently. In order to address this, the econometrician

[10] For the error term to be exogenous it is necessary that agents are unable to anticipate it, so that choice variables do not reflect its value. In the game from the previous section, the "luck component" is realized after players have chosen their strategies, hence it is ex-ante unknown.

would need as many appropriate "external" instruments as there are parameters of interest. Yet, if such instruments are absent but effort s is still observed, the estimation of a model augmented with contextual effects (albeit biased) would not recover non-existent sources of spillovers. If instead s is missing, the identification problem is more fundamental. As Proposition 3 highlights, in this case the econometrician is not able, by estimating linear models, to disentangle different spillover mechanisms.

When some of the invariant characteristics Z are unobserved, the model is also affected by problems of the omitted variable type. In intuitive terms, if some individual invariant factors – e.g. firm management quality, child ability – are missing, the spillovers parameter δ is not identified because of cross-complementarity effects in equilibrium. In such cases, however, the theoretical restrictions implied by the model suggest strategies to identify *all* the parameters except, of course, some elements in $\boldsymbol{\zeta}$. In particular, δ is identified via empirical analogues of the covariance restrictions from Proposition 2 and associated corollaries. Moreover, we use instruments motivated by the equilibrium conditions (4) to identify all but one parameters in $(\boldsymbol{\beta}, \gamma)$. This method is both simple and theoretically sound, yet not common in the practice of estimating production functions. Hence, we can bound the number of necessary "external" restrictions to one, which serves to identify the last remaining parameter.

In the rest of this section we illustrate our approach to identification. To simplify the exposition, we suppose that *all* invariant characteristics are unobserved and denote their parameter-weighted sum as $\boldsymbol{\omega} \equiv \mathrm{Z}\boldsymbol{\zeta}$; we assume without loss of generality that $\mathbb{E}[\boldsymbol{\omega}] = 0$. This way, model (8) would read as follows.

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

Models akin to (9) have been employed in studies on the evaluation of R&D spillovers by, among the others, Lychagin et al. (2010), Manresa (2014), Zacchia (2015).[11] Here $\boldsymbol{\omega}$ is an additional, endogenous, error term ($\mathbb{E}[\boldsymbol{\omega} \mid \mathrm{X}, \mathrm{s}, \mathbf{G}\mathrm{s}] \neq 0$). In the context of production functions it is popularly referred to as the "productivity shock", which is simultaneous to both conventional inputs and R&D. However, as we have discussed,

[11]Model (9) and the cited studies have in common that the "spillovers variable," which multiplies parameter δ, is expressed as the sum of the logarithms of peers' R&D. Parameter δ is interpreted as the elasticity of a homogeneous, simultaneous change of peer R&D on individual outcomes. However, while easier to relate to models of peer effects, such a specification is neither traditional nor the most common in the literature on R&D spillovers. Other studies, most notably Jaffe (1986) and Bloom et al. (2013), define the spillover variable as the logarithm of the sum of peers' R&D.

model (9) is equally suitable to other settings, like social interactions in the classroom. There, $\boldsymbol{\omega}$ would represent unobserved factors such as pupil ability. It is worth noticing that unlike the present case, traditional "additive" statistical models of peer effects – like the linear-in-means model – are not challenged by the omission of unobserved factors such as ability. Here, the threat to identification comes from the fact that individual ability affects individual effort which, in turn, affects peer effort.

2.1.1 Identification of δ

We show first how we obtain identification of the spillovers parameter δ. To this end, we introduce some technical assumptions.

Assumption 1. *The random variable ε_i is i.i.d. with mean zero, variance σ_ε^2 and finite higher order moments ($\mathbb{E}\left[\varepsilon_i^{4+d}\right] < \infty$ for any $d > 0$). Independence implies that $\mathbb{C}\text{ov}\left(\varepsilon_i, \varepsilon_j\right) = 0\ \forall i, j$. Moreover, ε_i is exogenous with respect to both individual characteristics and the network topology: $\mathbb{E}\left[\boldsymbol{\varepsilon} \mid \mathrm{X}, \mathrm{s}, \mathbf{G}\mathrm{s}, \mathbf{G}, \boldsymbol{\omega}\right] = 0$.*

This assumption states some usual regularity conditions and formalizes the exogeneity property of the shock ε_i. However, the underlined fact that its draws are pairwise independent is economically substantive. In fact, this excludes the possibility that different agents – or observations – share any individual idiosyncratic factor or "luck." Therefore, any exogenous "common shock" that is shared by multiple agents must enter the term ω_i, which is endogenous in the model. This is more interesting for the purpose of analyzing social and economic interactions, since endogenous common shocks are usually considered a major threat, if not the main threat (Angrist, 2014), to the identification of spillover effects.[12]

Assumption 2. *The random variable ω_i is i.i.d. with mean zero, variance σ_ω^2 and finite higher order moments ($\mathbb{E}\left[\omega_i^{4+d}\right] < \infty$ for any $d > 0$). Independence implies that $\mathbb{C}\text{ov}\left(\omega_i, \omega_j\right) = 0\ \forall i, j$.*

Technical conditions aside, this corresponds to the assumption of independent types under which Proposition 2 holds. We rule out, for the moment, even the possibility of exogenous "common shocks" in ω_i. Hence, we circumscribe the identification problem to the endogenous reflection of unobserved characteristics in equilibrium. We later

[12]In linear-in-means models, "common shocks" are unobserved factors that are correlated across peers and affect individual outcomes. As such, they might bias the estimation of the spillover effects.

relax this hypothesis and show that our identification results still hold after modifying the moment conditions appropriately. Armed with this assumption, we can derive the empirical analogues of the covariance restriction from Proposition 2 and its corollaries.

Lemma 1. *In the model, under Assumption 2, there are two empirical analogues of the covariance restriction from Proposition 2, and they are linearly dependent:*

$$\mathbb{E}\left[\left(\mathrm{s}-\frac{\delta}{1-\gamma-\boldsymbol{\beta}'\iota}\left(\mathbf{G}\mathrm{s}-\bar{\mathrm{s}}\right)\right)'\mathbf{G}\left(\mathrm{s}-\frac{\delta}{1-\gamma-\boldsymbol{\beta}'\iota}\left(\mathbf{G}\mathrm{s}-\bar{\mathrm{s}}\right)\right)\right]\propto\mathbb{E}\left[\boldsymbol{\omega}'\mathbf{G}\boldsymbol{\omega}\right]=0 \tag{10}$$

$$\mathbb{E}\left[\left(\mathrm{s}-\frac{\delta}{1-\gamma}\left(\mathbf{G}\mathrm{s}-\mathrm{X}\boldsymbol{\beta}-\bar{\bar{\mathrm{s}}}\right)\right)'\mathbf{G}\left(\mathrm{s}-\frac{\delta}{1-\gamma}\left(\mathbf{G}\mathrm{s}-\mathrm{X}\boldsymbol{\beta}-\bar{\bar{\mathrm{s}}}\right)\right)\right]\propto\mathbb{E}\left[\boldsymbol{\omega}'\mathbf{G}\boldsymbol{\omega}\right]=0 \tag{11}$$

where $\bar{\bar{\mathrm{s}}}=\left(\boldsymbol{\alpha}+\log\boldsymbol{\gamma}+\log\mathbb{E}\left[\exp\left(\varepsilon_i\right)\right]\right)\boldsymbol{\iota}+\mathrm{q}$.

Lemma 1 can be easily shown via manipulation of the First Order Conditions in (3-4).[13] Equations (10-11) are not easily implemented empirically, because they require knowledge of individual cost factors $\mathbf{p}_i$ and q_i. However, they help very well conveying the intuition behind the identification of δ. Consider first that vector $\bar{\mathrm{s}}$ (or, alternatively, $\bar{\bar{\mathrm{s}}}+\mathrm{X}\boldsymbol{\beta}$) is a sufficient statistic of all exogenous factors that determine equilibrium effort s – except $\boldsymbol{\omega}$. Thus, if the latter is independently distributed across agents, controlling for $\bar{\mathrm{s}}$ (or $\bar{\bar{\mathrm{s}}}+\mathrm{X}\boldsymbol{\beta}$) should result in the residual equilibrium effort to be uncorrelated across agents as well, if not for the equilibrium response to spillovers. Hence, the excess covariance would capture said spillovers, allowing to identify δ.

Lemma 2. *In the model, under Assumptions 1-2, the empirical analog of the covariance restriction expressed by Corollary 2 is*

$$\begin{aligned}\mathbb{E}\left[\left(\mathrm{y}-\boldsymbol{\alpha}\boldsymbol{\iota}-\mathrm{X}\boldsymbol{\beta}-\gamma\mathrm{s}-\boldsymbol{\delta}\mathbf{G}\mathrm{s}\right)'\mathbf{G}\left(\mathrm{y}-\boldsymbol{\alpha}\boldsymbol{\iota}-\mathrm{X}\boldsymbol{\beta}-\gamma\mathrm{s}-\boldsymbol{\delta}\mathbf{G}\mathrm{s}\right)\right]&=\\=\mathbb{E}\left[\left(\boldsymbol{\omega}+\boldsymbol{\varepsilon}\right)'\mathbf{G}\left(\boldsymbol{\omega}+\boldsymbol{\varepsilon}\right)\right]&=0\end{aligned} \tag{12}$$

which is linearly dependent to both restrictions (10) and (11), separately.

[13]The difference between equations (10-11) is that the latter, unlike the former, includes the X variables explicitly, rather than substituting them as functions of s into equation (3). Since the X variables and the spillover effect are complementary, making this substitution "enhances" the equilibrium response to peer effort by a factor of $1/\left(1-\gamma-\boldsymbol{\beta}'\iota\right)>1/\left(1-\gamma\right)$, accounting for the elasticity β_k of each k-th ordinary choice variable on the final outcome Y_i.

To see that condition (12) is linearly dependent to both (10) and (11) is immediate: by Assumption 1, $\mathbb{E}\left[(\boldsymbol{\omega}+\boldsymbol{\varepsilon})'\,\mathbf{G}\,(\boldsymbol{\omega}+\boldsymbol{\varepsilon})\right] = \mathbb{E}\left[\boldsymbol{\omega}'\mathbf{G}\boldsymbol{\omega}\right] + \mathbb{E}\left[\boldsymbol{\varepsilon}'\mathbf{G}\boldsymbol{\varepsilon}\right] = \mathbb{E}\left[\boldsymbol{\omega}'\mathbf{G}\boldsymbol{\omega}\right] = 0$. Intuitively, if the equilibrium cross-correlation in the final outcomes y_i is a byproduct of the cross-correlation in the strategic variables, by controlling for these as well as for spillover effects the residual covariance between connected players in the network would fall to zero. The discussion above about identification also applies in this case, with final outcomes replacing equilibrium effort (which is now among the factors to be controlled for). Condition (12), unlike (10)-(11), is easier to implement empirically as it requires data on final outcomes rather than on input prices, and the former are much more common to be available than the latter at the individual level. Hence, we advocate the use of (12) for practical purposes.

2.1.2 Identification of $(\boldsymbol{\beta}, \gamma)$

A problem with any proposed covariance restriction is that it requires knowledge, or consistent estimation, of the parameters set $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma)$. Absent external instruments, these parameters (except $\boldsymbol{\alpha}$) cannot be identified from conditional moments of the first order of (9) because $\boldsymbol{\omega}$ is endogenous. There is a vast literature on the estimation of production functions.[14] The most popular methods (Olley and Pakes, 1996; Levinsohn and Petrin, 2003) all require panel data. Here, we propose a simple solution to the endogeneity problem, which requires only a single external exclusion restriction, has no stronger data requirements than most other approaches, and yet works with cross-sectional data. Specifically, we argue that the model's equilibrium properties are sufficient to identify K of the $K+1$ parameters $(\boldsymbol{\beta}, \gamma)$. Hence, exact identification can be achieved via an additional, "external" exclusion restriction for the last $K+1$-st parameter of the model.

Consider the set of equations (4), which are derived from taking logarithms of the firm maximization problem's First Order Conditions. They express the K equilibrium input ratios X_{ik}/S_i – the marginal rates of substitution – as functions of structural parameters and input prices. In particular, these are not functions of the unobservable characteristics $\boldsymbol{\omega} = \mathrm{Z}\boldsymbol{\zeta}$. This is a specific property of Cobb-Douglas and more generally CES production functions. For this class of functional forms, in fact, the elasticity of any input is constant even controlling for the elasticity of other complementary inputs. Hence, when calculating the marginal rate of substitution

[14] For a survey see Van Beveren (2012).

between any two inputs in equilibrium, the effect of an additional input cancels out both at the numerator and at the denominator. It follows that, in statistical terms, the equilibrium input ratios are statistically uncorrelated with the unobserved error term: $\mathbb{E}\left[\mathrm{x}_k - \mathrm{s} \middle| \boldsymbol{\omega}\right] = 0$ for every $k = 1, \ldots, K$. Hence, under the assumptions of the model, the data offer K natural moment conditions of the following form.

$$\mathbb{E}\left[(\mathrm{x}_k - \mathrm{s})' (\mathrm{y} - \alpha\boldsymbol{\iota} - \mathrm{X}\boldsymbol{\beta} - \gamma\mathrm{s} - \delta\mathbf{G}\mathrm{s})\right] = \mathbb{E}\left[(\mathrm{x}_k - \mathrm{s})' (\boldsymbol{\omega} + \boldsymbol{\varepsilon})\right] = 0 \quad \forall k = 1, \ldots, K \tag{13}$$

Notice that for the k-th equation in (13), the instrument is relevant for both β_k and γ, as $\mathbb{E}\left[(\mathrm{x}_k - \mathrm{s})' \mathrm{x}_k\right] > 0$ and $\mathbb{E}\left[(\mathrm{x}_k - \mathrm{s})' \mathrm{s}\right] < 0$.

However, separate identification of *one* parameter in $(\boldsymbol{\beta}, \gamma)$ is still necessary. To this end, the model offers no further "internal" solutions. In fact, trying to use an additional input ratio as an instrument – say $\mathbb{E}\left[X_{ik'}/X_{ik''}\right]$ – would not work: the resulting moment condition would be perfectly collinear with the two from (13) that involve, respectively, the k'-th and the k''-th inputs. Nor would using the First Order Condition (3) work to identify γ because of, again, an instance of exact simultaneity.[15] Hence, an "external" exclusion restriction is necessary. Any exogenous instruments that correlates with either a choice variable X_{ik} or individual effort S_i would suffice to identify either β_k or γ. Here we focus, for illustratory purposes, on the case where some instrument of S_i is available. Suppose that the cost of effort, q_i, is a function $q_i = q(c_i)$ of an additional variable c_i, which varies at the agent level,[16] and that is exogenous to both types of error terms: $\mathbb{E}\left[\boldsymbol{\omega} + \boldsymbol{\varepsilon} \middle| \mathrm{c}\right] = 0$. Clearly, relevance holds by (3): $\mathbb{E}\left[\mathrm{c}'\mathrm{s}\right] > 0$. Then, γ can be identified via a natural exclusion restriction.

$$\mathbb{E}\left[\mathrm{c}' (\mathrm{y} - \alpha\boldsymbol{\iota} - \mathrm{X}\boldsymbol{\beta} - \gamma\mathrm{s} - \delta\mathbf{G}\mathrm{s})\right] = \mathbb{E}\left[\mathrm{c}' (\boldsymbol{\omega} + \boldsymbol{\varepsilon})\right] = 0 \tag{14}$$

This permits identification of all parameters of interest $(\alpha, \boldsymbol{\beta}, \gamma, \delta)$.[17]

[15]This is best seen by substituting (3) into (9): one gets $\mathrm{y} - \mathrm{s} = \left[-\log\gamma + \log\mathbb{E}\left[\exp\left(\varepsilon_i\right)\right]\right]\boldsymbol{\iota} + \mathrm{q} + \boldsymbol{\varepsilon}$. On the right hand side there are no observable variables interacting with the parameter of interest γ; moreover $\log\gamma$ and $\log\mathbb{E}\left[\exp\left(\varepsilon_i\right)\right]$ cannot even be separately identified via first moments of $\mathrm{y} - \mathrm{s}$.

[16]Input prices are obvious candidates for instrumenting inputs, although price variation at the firm level is seldom available (for a recent countexample in a study on the returns of R&D see Doraszelski and Jaumandreu (2013)). A common critique of using prices as instruments is that, in presence of firm market power, they are themselves endogenous. For our purposes, the observation of the exact price of an input faced by each agent is not necessary. Only an exogenous component of the price, differing by agent, is needed (in the case of firms, tax credits for either R&D or other types of investment are prime options).

[17]If an exogenous instrument for effort S_i were not available, but instead an instrument for some

2.1.3 Power of Identification

An important result of the literature on the estimation of peer and spillover effects is that identification hinges on the topology of social interaction. Perhaps most notably, Bramoullé et al. (2009) show that identification of a model like (6) is possible as long as matrices $\mathbf{I}$, $\mathbf{G}$ and $\mathbf{G}^2$ are linearly independent. This is equivalent to say that the patterns of interaction between agents should not be uniform within well-specified but hermetically closed "groups." In network parlance, agents must form connections that are not perfectly-overlapping: *open triads*[18] should be frequent in the network. It is easy to verify that the same requirement applies to the identification strategy that we propose. In particular, covariance restrictions (10), (11) and (12) would not be relevant moments if agents interact in perfectly overlapping groups.

Even if identification is algebraically possible, however, it can still be weak. On the one hand, if agents form few bonds between each other there might be few open triads to properly identify spillover effects. On the other hand, a similar conclusion can be drawn in the opposite scenario, when the network is very tight and most nodes are connected to all the others. In this sense, we are less optimistic than Blume et al. (2015) are, relative to the claim that spillover effects can almost always be identified as long as perfect group regularity is broken. The breaks may, in fact, result in very weak instruments. To the best of our knowledge no studies on spillover effects report measures, based on the network topology, to validate the quality of identification. We advocate the use of two measures in combination: *i.* the ratio of the open triads count over $\binom{N}{3}$ from the network census, and *ii.* the R^2 coefficient from the regression

$$\operatorname{vec}\left(\mathbf{G}^2\right) = a + b \cdot \operatorname{vec}\left(\mathbf{G}\right) + c \cdot \operatorname{vec}\left(\mathbf{I}\right) + \mathbf{e} \tag{15}$$

where (a, b, c) are parameters and $\mathbf{e}$ is an error term. It is easy to see that if $\mathbf{I}$, $\mathbf{G}$ and $\mathbf{G}^2$ are linearly dependent, the R^2 coefficient from (15) is equal to 1. Hence, for identification to be powerful it is desirable for this R^2 to be as low as possible. For

other input X_{iq} were instead, identification is still achieved. The reason is that in this case γ is identified via the q-th condition in (13), since the instrument $\mathrm{x}_q - \mathrm{s}$ is also relevant for effort. The relative precision of parameter estimate can, however, be affected. Moreover, an instrument c for effort can also be employed to construct an additional moment condition aimed at identifying δ. In fact the vector $\mathbf{G}\mathrm{c}$, while still being exogenous, would be a good predictor of $\mathbf{G}\mathrm{s}$. In this case, we would be able to obtain overidentification.

[18] An open triad is a triple of any three nodes in a network, such that two of them are unconnected to each other, but are both connected to the thid node.

illustratory purposes, in Appendix B we relate this measure of "connections overlap" to the eleven different types of tetrads – the different configurations of one network's any four nodes – and show the R^2 measure associated to each of them. Notably, $\mathrm{R}^2 \neq 1$ only for those tetrads that contain at least one open triad.

Given the fact that identification should not only be mathematically possible, but also statistically powerful, we find it necessary to establish conditions for identification which are stronger than just linear independence of $\mathbf{I}$, $\mathbf{G}$ and $\mathbf{G}^2$. We formally state the notion that the network shall be "neither be too sparse nor too dense, but just fine" by making the following, non mutually exclusive assumptions.

Assumption 3. (Sparsity) *The number of connections that each node can have is bounded,* i.e.

$$\lim_{N\to\infty} \sum_{j=1}^{N} \mathbb{I}\left[g_{ij} > 0\right] < \overline{G}$$

for some $\overline{G} < N$ and for all $i = 1, \ldots, N$.

Assumption 4. (Density) *Asymptotically, each node $i = 1, \ldots, N$ belongs to at least one open triad. That is, for sufficiently large N there exist two nodes j, ℓ with $i \neq j \neq \ell$ such that one and only one element of the triple $(g_{ij}, g_{j\ell}, g_{\ell i})$ is zero.*

Assumption 3 places bounds on how many connections a node can have: therefore, we rule out the possibility that some nodes are connected to all other nodes. Hence, the network is not "too dense." Assumption 4, instead, poses another restriction on the network topology: that is, each node must be member of at least one open triad. This way, the network cannot be "too sparse" either. We denote the total number of open triads in the network as T_N. We make the following observation about one implication of these two assumptions combined.

Proposition 4. *Let assumptions 3 and 4 hold. Then, the growth rate of the number of open triads equals the growth rate of the number of nodes,* i.e. $O\left(\frac{N}{T_N}\right) = 1$.

With the number of open triads in the network growing, asymptotically, as fast as the number of nodes, the identification strategy we propose is guaranteed to be adequately powerful for practical purposes. In the Appendix we show how this result relates to the GMM estimator we are about to discuss. In addition, we argue that the same intuition applies more generally, in particular to the IV approaches for the identification of linear in means models *á la* Bramoullé et al. (2009).

2.2 Estimation

We estimate the parameter set $\theta = (\alpha, \beta, \gamma, \delta)$ jointly using GMM. We define $\mathrm{m}(\theta)$ as the vector of all the moment conditions we need for identification: the covariance restriction (12), the K moment conditions (13) employing input ratios as instruments, the external exclusion restriction from (14), and finally the moment condition of the form $\mathbb{E}[\omega + \varepsilon] = 0$ for the identification of the constant term α. The sample analogue of $\mathbb{E}[\mathrm{m}_i(\theta)]$, written as $\bar{\mathrm{m}}(\theta)$, reads as

$$\overline{\mathrm{m}}_N(\theta) = \frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix} e_i(\theta) \\ (x_{1i} - s_i)\, e_i(\theta) \\ \dots \\ (x_{Ki} - s_i)\, e_i(\theta) \\ c_i e_i(\theta) \\ \sum_{j\neq i} g_{ij} e_i(\theta)\, e_j(\theta) \end{pmatrix} \tag{16}$$

where $e_i(\theta) = y_i - \alpha - \sum_{k=1}^{K} \beta_k x_{ki} - \gamma s_i - \delta \sum_{j\neq i} g_{ij} s_j$. The GMM estimator is

$$\hat{\theta}_{GMM} = \arg\min_{\theta} \overline{\mathrm{m}}_N'(\theta)\, \mathbf{W}_N \overline{\mathrm{m}}_N(\theta) \tag{17}$$

where $\mathbf{W}_N$ is some GMM weight matrix.[19] In order to characterize the asymptotic properties of the GMM estimator in (17), we maintain the following regularity conditions:

Assumption 5. $\theta_0 \equiv (\alpha, \beta', \gamma, \delta)'$ *is in the interior of the parameter space* $\boldsymbol{\Theta}$*, which is a bounded subspace of* $\mathbb{R}^{K+3}$.

Assumption 6. $(\overline{x}_{1i}, \dots, \overline{x}_{Ki}, c_i)$ *are distributed over a bounded subspace of* $\mathbb{R}^{K+1}$.

Assumption 7. $\|\mathbf{W}_N - \mathbf{W}_0\| = o_P(1)$ *uniformly in* θ.

Assumption 8. $\|\nabla_\theta \mathrm{m}_i(\theta)\| \leq \overline{M}$*, for some* $\overline{M}$ *such that* $\overline{M} < \infty$*,* $\forall \theta \in \boldsymbol{\Theta}$.

Assumption 9. $\mathbb{E}[\nabla_\theta \mathrm{m}_i(\theta_0)]' \mathbf{W}_0 \mathbb{E}[\nabla_\theta \mathrm{m}_i(\theta_0)]$ *exists and is invertible.*

These assumptions allow to derive the following result, whose proof is provided in Appendix A.

[19] As it is well known, in the case of exact identification the choice of the weight matrix is irrelevant. However, we maintain this use of notation in order to accommodate the case of overidentification, which arises when using an additional instrument of the type $\sum_{j\neq i} g_{ij} c_j$.

Proposition 5. *Under Assumptions 1-9, the estimator $\hat{\boldsymbol{\theta}}_{GMM}$ given by equation (17) is a consistent estimator of $\boldsymbol{\theta}_0$ and its asymptotic distribution is given by*

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}}_{GMM}-\boldsymbol{\theta}_0\right)\xrightarrow{d}\mathcal{N}\left(0,\mathbf{V}_{\boldsymbol{\theta}}\right)$$

with $\mathbf{V}_{\boldsymbol{\theta}}$, $\mathbf{V}_0$, $\mathbf{M}_0$ and $\mathbf{W}_0$ defined as follows.

$$\begin{aligned}
\mathbf{V}_{\boldsymbol{\theta}} &\equiv \left(\mathbf{M}_0'\mathbf{W}_0\mathbf{M}_0\right)^{-1}\mathbf{M}_0'\mathbf{W}_0\mathbf{V}_0\mathbf{W}_0\mathbf{M}_0\left(\mathbf{M}_0'\mathbf{W}_0\mathbf{M}_0\right)^{-1}\\
\mathbf{V}_0 &\equiv \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{m}_i\left(\boldsymbol{\theta}_0\right)\mathrm{m}_j\left(\boldsymbol{\theta}_0\right)'\\
\mathbf{M}_0 &\equiv \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\frac{\partial\mathrm{m}_i\left(\boldsymbol{\theta}_0\right)}{\partial\boldsymbol{\theta}'}\\
\mathbf{W}_0 &\equiv \lim_{N\to\infty}\mathbf{W}_N
\end{aligned}$$

2.3 Extensions

In this section, we show how the proposed estimation procedure can be appropriately modified to accommodate assumptions or types of data that are different from those maintained so far. We explore first how the covariance restriction for the identification of δ can be modified under the hypothesis of "common shocks." We then turn our attention to how panel data offer further options for identification, that are related to typical identifying assumptions for the estimation of production functions.

2.3.1 Common Shocks

Suppose that instead of Assumption 2, the following one holds.

Assumption 10. *The random variable ω_i has mean zero and the following variance-covariance structure*

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\left(\omega_i\right) &= \phi_0^2\sum_{j\neq i}g_{ij}^2\\
\mathbb{C}\mathrm{ov}\left(\omega_i,\omega_j\right) &= \phi_0^2 g_{ij}^2
\end{aligned}$$

that is, draws of ω_i are not i.i.d. but are pairwise correlated for connected agents. Moreover, ω_i has finite higher order moments ($\mathbb{E}\left[\omega_i^{4+d}\right]<\infty$ for any $d>0$).

This is a simple way to introduce "common shocks" in the model. Notice that the variance-covariance structure specified in Assumption 10 reflects a precise microfoundation. In particular, if connected nodes share similar characteristics that are specific to their link – if any two nodes i and j get an i.i.d. "common shock" υ_{ij} with mean zero, variance ϕ_0^2 and finite higher order moments for all i, j – writing $\omega_i = \sum_{j \neq i} g_{ij} \upsilon_{ij}$ results in the given structure.

With Assumption 10, we introduce an "exogenous source of endogeneity" into the model. Without common shocks, the spillover term in (9) is endogenous since it reflects the unobserved characteristics of individual agents, because of the equilibrium response to complementarities. With common shocks, in addition, it is endogenous as two connected agents reflect unobserved factors that are *ex ante* (exogenously) similar. Clearly, common shocks invalidate the moment condition for the identification of δ (12). However, it is possible by re-formulate the latter appropriately.[20]

Lemma 3. *In the model, under Assumptions 1 and 10, the following zero covariance condition holds:*

$$\mathbb{E}\left[(\mathrm{y} - \alpha \iota - \mathrm{X}\boldsymbol{\beta} - \gamma \mathrm{s} - \delta \mathbf{G}\mathrm{s})' \mathbf{H} (\mathrm{y} - \alpha \iota - \mathrm{X}\boldsymbol{\beta} - \gamma \mathrm{s} - \delta \mathbf{G}\mathrm{s})\right] = \\ = \mathbb{E}\left[(\boldsymbol{\omega} + \boldsymbol{\varepsilon})' \mathbf{H} (\boldsymbol{\omega} + \boldsymbol{\varepsilon})\right] = 0 \quad (18)$$

where $\mathbf{H}$ *is the matrix of "indirect connections" of second degree, and each of their elements* h_{ij} *for* $i, j = 1, \ldots, N$ *is defined as follows.*

$$h_{ij} = \begin{cases} \sum_{\ell=1}^{N} g_{i\ell} g_{\ell j} & \text{for } g_{ij} \neq 0 \text{ and } i \neq j \\ 0 & \text{otherwise} \end{cases}$$

If one is concerned about common shocks, the covariance restriction in the GMM problem can be substituted for an empirical analog of equation (18). This version of the covariance restriction in 18 exploits the fact that "indirect connections" do not share any common shocks under Assumption 10. However, thanks to common shocks, the relevance condition still holds. In fact, it can be seen that

$$\frac{\partial}{\partial \delta} \mathbb{E}\left[\mathrm{e}'(\boldsymbol{\theta}) \mathbf{H} \mathrm{e}(\boldsymbol{\theta})\right] = -2\mathbb{E}\left[\mathrm{s}' \mathbf{G}\mathbf{H} \mathrm{e}(\boldsymbol{\theta})\right] > 0$$

[20] This can be motivated by an appropriate re-formulation of Proposition 2 in the context of the analytical model. We show this in the Appendix.

as peer effort still correlates with the unobserved shocks of indirect friends.

Suppose that, instead, common shocks also affect nodes that are reciprocally located at network distances higher than 1. If, for example,

$$\mathbb{C}\text{ov}\left(\omega_i, \omega_j\right) = \phi_1^2 g_{ij}^2 + \phi_1^2 \sum_{\ell=1}^{N} g_{i\ell} g_{\ell j} \tag{19}$$

then also "indirect friends" at distance two are hit by common shocks.[21] In such a case, clearly, also covariance restriction (18) is invalidated. The solution is intuitive: to use, instead of $\mathbf{F}$, a matrix whose entries are non-zero only for indirect connections of third degree (call it $\mathbf{J}$). One may be concerned that, by adopting this approach to address the problem of common shocks, the relevance of the covariance restriction is increasingly reduced the higher is the degree of the indirect connections which are used for constructing the "meat" matrix. This problem is analogous to the one faced by Zacchia (2015) when developing, in a similar setting, an IV strategy based on the effort choices S_i of indirect connections of higher degree. In that context the higher is the degree of distance, the weaker is potentially the associated instrument. More generally, the issue also bears similarities to that of "higher and weaker lags" in the Arellano-Bond approach (and subsequent extensions) for estimating dynamic models in panel data. Hence, good applied practice would be to present more estimates that are based on different definitions of the covariance restriction.

2.3.2 Panel Data

With panel data over $t = 1, \ldots, T$ periods, our model reads as

$$\mathrm{y}_t = \boldsymbol{\alpha} + \mathrm{X}_t \boldsymbol{\beta} + \gamma \mathrm{s}_{t-s} + \delta \mathbf{G} \mathrm{s}_{t-s} + \boldsymbol{\omega}_t + \boldsymbol{\varepsilon}_t \tag{20}$$

where $\boldsymbol{\alpha}$ is an unrestricted vector, $s \in \mathbb{N}$, and $\boldsymbol{\omega}_t$ presents some degree of persistence in time. A common assumption in the literature on production functions is that the unobserved shocks follows a Markov Process: in particular, an AR(1) process of the type $\boldsymbol{\omega}_t = \rho \boldsymbol{\omega}_{t-1} + \boldsymbol{\xi}_t$, with $\rho \in [0, 1]$ and $\boldsymbol{\xi}_t$ is an i.i.d. "innovation" of the shock.

[21] Equation (19) can be motivated on a different microfoundation of common shocks. Suppose that every node receives a "fundamental" i.i.d. shock ϖ_i with mean zero, variance ϕ_1^2 and finite higher order moments. If these shocks are aggregated as $\omega_i = \sum_{j \neq i} g_{ij} \varpi_j$ – that is, if the network reflects node similarities – then the covariance structure of ω_i is given by (19). In this case, cross-correlation extends up to two degrees of separation because of two nodes' similarities with their common links.

A central feature of (20) is that when $s > 0$ (most typically $s = 1$), individual effort has a "lagged" effect. This is, for example, the most sensible assumption in the case of R&D, whose effects on firm-level outcomes like productivity are known to pan out with some delay between the initial investment and its realizations.

The theoretical analysis of our model in a dynamic setting offers no particularly relevant insight relative to the static case, and it is best left for an appendix. However, it is important to remark that as usual with panel data, there are potentially additional "internal" exclusion restrictions that can be exploited for identification. Most of the literature on the estimation of production functiosn is based on some central assumption of the following type: because of specific timing in input choices, coupled with adjustment costs,[22] some input variable is orthogonal to the "innovation" component of the error term. In particular, with "forward looking investment" the current capital stock is orthogonal to the innovation component of $\boldsymbol{\omega}_t$:

$$\mathbb{E}\left[\boldsymbol{\xi}_t \middle| \mathrm{x}_{k't}\right] = 0 \tag{21}$$

for some k'-th choice variable. For $s > 0$, however, another more natural exclusion restriction is available in the model:

$$\mathbb{E}\left[\boldsymbol{\xi}_t \middle| \mathrm{s}_{t-s}\right] = 0 \tag{22}$$

that is, past effort is orthogonal to the innovation of the unobserved shock, even if not to its lagged values.[23] Both moment conditions (21) and (22) can be exploited for identification instead of an "external" instrument in a transformed model like

$$\tilde{\mathrm{y}}_t = \tilde{\boldsymbol{\alpha}} + \tilde{\mathrm{X}}_t \boldsymbol{\beta} + \gamma \tilde{\mathrm{s}}_{t-s} + \boldsymbol{\delta} \mathbf{G} \tilde{\mathrm{s}}_{t-s} + \boldsymbol{\xi}_t + \tilde{\boldsymbol{\varepsilon}}_t \tag{23}$$

where the tilde applied on any vector a_t denotes the transformation $\tilde{\mathrm{a}}_t \equiv \mathrm{a}_t - \rho \mathrm{a}_{t-1}$ (this nests the first differences or "between" transformation for $\rho = 1$). The idea is similar to that in Blundell and Bond (2000), but restricted to only one input variable.

[22]Adjustment costs are a built-in feature of stock variables. Physical capital, for example, is thought to follow a "law of motion" of the type $K_t = (1 - d_1) K_{t-1} + I_t$ where d_0 is the depreciation parameter and I_t is yearly investment in capital goods. The fact that capital cannot be freely, fully adjusted from one year to another, is equivalent to the presence of adjustment costs. Under the knowledge capital paradigm (Griliches, 1979), R&D is also treated as a stock variable, following $S_t = (1 - d_2) S_{t-1} + R_t$ where R_t is yearly R&D investment.

[23]This idea is natural in the case of R&D, see e.g. Doraszelski and Jaumandreu (2013).

3 Monte Carlo

We run a Monte Carlo simulation in order to evaluate the effectiveness of our proposed estimation procedure. First, given the sample size that we set at $N = 200$, we generate a random unweighted and undirected network $\mathbf{G}$ such that $\mathbb{P}(g_{ij} = g_{ji} = 1) = 0.02$ and, conversely, $\mathbb{P}(g_{ij} = g_{ji} = 0) = 0.98$ for any pair (i, j) with $i \neq j$. Given $\mathbf{G}$, the data are generated according to equations (3), (4) and (9), with $\boldsymbol{\omega} \equiv \mathrm{Z}\boldsymbol{\zeta}$. We set parameters $(\alpha, \beta_1, \beta_2, \gamma, \delta) = (2, 0.15, 0.65, 0.05, 0.01)$ and randomly draw the following variables (we set $\boldsymbol{\varepsilon} = 0$ for simplicity).

$$\begin{aligned}
\boldsymbol{\omega} &\sim \mathcal{N}\left(0, 0.02^2\right) \\
\mathbf{p}_1 &\sim \mathcal{N}\left(-1, 0.02^2\right) \\
\mathbf{p}_2 &\sim \mathcal{N}\left(-1, 0.06^2\right) \\
\mathbf{q} &\sim \mathcal{N}\left(-1, 0.05^2\right)
\end{aligned}$$

The number of repetitions is $R = 1000$. The results are shown in Table 1.

Table 1: Monte Carlo Results

	(1)	(2)	(3)	(4)	θ_0
α	1.798	2.192	2.121	1.387	2
	(0.0522)	(0.4603)	(0.4261)	(0.0026)	
β_1	0.259	0.046	0.082	0.507	0.15
	(0.0159)	(0.1358)	(0.1278)	(0.0010)	
β_2	0.564	0.735	0.706	0.335	0.65
	(0.0102)	(0.0872)	(0.0819)	(0.0006)	
γ	0.052	0.046	0.047	0.092	0.05
	(0.0006)	(0.0014)	(0.0014)	(0.0003)	
δ	0.008	0.012	0.011	0.004	0.01
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	

Notes: In every column, for each parameter we report mean point estimates and standard deviations (in parentheses) across repetitions. Columns 1, 2 and 3 report the estimates of the GMM estimator where we set the "meat" matrix of the covariance restrictions as $\mathbf{G}$, $\mathbf{H}$ and $\mathbf{J}$, respectively (see Section 2.3.1). Column 4 reports the estimates of the naive 2SLS estimator employing $(\iota, \mathrm{x}_1, \mathrm{x}_2, \mathrm{s}, \mathbf{G}\mathrm{s}, \mathbf{G}^2\mathrm{s})$ as instruments.

4 Conclusion

We have the following plans for to complete and polish the present paper:

1. to refine the Monte Carlo simulation;

2. to add an empirical application about R&D spillovers;

3. to extend the analysis of the common shocks problem and its solutions.

References

Angrist, J. D. (2014). The perils of peer effects. *Labour Economics 30*, 98–108.

Bloom, N., M. Schankerman, and J. Van Reenen (2013). Identifying technology spillovers and product market rivalry. *Econometrica 81*(4), 1347–1393.

Blume, L. E., W. A. Brock, S. N. Durlauf, and R. Jayaraman (2015). Linear Social Interactions Models. *Journal of Political Economy 123*(2), 444–496.

Blundell, R. and S. Bond (2000). Gmm estimation with persistent panel data: an application to production functions. *Econometric reviews 19*(3), 321–340.

Bramoullé, Y., H. Djebbari, and B. Fortin (2009). Identification of peer effects through social networks. *Journal of econometrics 150*(1), 41–55.

Calvó-Armengol, A., E. Patacchini, and Y. Zenou (2009). Peer Effects and Social Networks in Education. *The Review of Economic Studies 76*(4), 1239–1267.

Doraszelski, U. and J. Jaumandreu (2013). R&D and productivity: Estimating endogenous productivity. *The Review of Economic Studies 80*(4), 1338–1383.

Graham, B. S. (2008). Identifying social interactions through conditional variance restrictions. *Econometrica 76*(3), 643–660.

Griliches, Z. (1979). Issues in assessing the contribution of research and development to productivity growth. *The Bell Journal of Economics 10*(1), 92–116.

Jaffe, A. B. (1986). Technological opportunity and spillovers of R&D: evidence from firms' patents, profits and market value. *American Economic Review 76*(5), 984–1001.

Kelejian, H. H. and I. R. Prucha (2001). On the asymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics 104* (2), 219–257.

Lee, L.-f. (2007). Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. *Journal of Econometrics 140* (2), 333–374.

Levinsohn, J. and A. Petrin (2003). Estimating production functions using inputs to control for unobservables. *The Review of Economic Studies 70* (2), 317–341.

Lychagin, S., J. Pinkse, M. E. Slade, and J. Van Reenen (2010). Spillovers in space: does geography matter? Forthcoming in *The Journal of Industrial Economics*.

Manresa, E. (2014). Estimating the Structure of Social Interactions Using Panel Data. Mimeo.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies 60* (3), 531–542.

Olley, S. G. and A. Pakes (1996). The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica 64* (6), 1263–1297.

Pereda Fernández, S. (2015). Social spillovers in the classroom: Identification, estimation and policy analysis. Mimeo.

Van Beveren, I. (2012). Total factor productivity estimation: A practical review. *Journal of Economic Surveys 26* (1), 98–128.

Zacchia, P. (2015). Knowledge Spillovers through Networks of Scientists. mimeo.

Appendix A Mathematical Proofs

A.1 Proof of Proposition 1

An *individual strategy profile* is a combination of effort and ordinary choice variables $\mathbf{b}_i = (S_i, X_{1i}, \ldots, X_{Ki}) \in \mathbb{R}_+^{K+1}$. A *strategy profile* is the collection of all individual strategy profiles $\mathbf{b} = (\mathbf{b}_1, \ldots, \mathbf{b}_N)$. A Bayes-Nash Equilibrium of the game described in Section 1 is a strategy profile $\mathbf{b}^*$ satisfying

$$\mathbb{E}\left[U_i\left(\mathbf{b}_i^*, \mathbf{b}_{-i}^*; \mathbf{t}_i\right)\right] \geq \mathbb{E}\left[U_i\left(\mathbf{b}_i, \mathbf{b}_{-i}^*; \mathbf{t}_i\right)\right] \tag{A.1}$$

for all $\mathbf{b}_i \in \mathbb{R}_+^{K+1}$, for all $i = 1, \ldots, N$ and for $U_i(\cdot)$ defined as in (2). Notice that: *i.* the only component of $\mathbf{b}_{-i}^*$ affecting the individual utility of player i *directly* is other players' effort (spillovers) $\mathbf{s}_{-i}^* = \left(S_1^*, \ldots, S_{i-1}^*, S_{i+1}^*, \ldots, S_N^*\right)$; *ii.* the only element of stochastic uncertainty in the model is "luck" ε_i, which is unknown even to individual receivers of the shock. Hence, the game does not differ in a substantive way from a single choice variable game of complete information. However, we prefer to maintain the current setup and frame the solution as a Bayes-Nash Equilibrium in order to keep consistency with: *1.* the empirical counterpart of the model, *2.* the solution concept that would be adopted in those extensions of the model assuming that the vector of types $\mathbf{t} = (\mathbf{t}_1, \ldots, \mathbf{t}_N)$ is not common knowledge.

Individual expected utility reads as

$$\begin{aligned}
&\mathbb{E}\left[U_i\left(\mathbf{b}_i, \mathbf{b}_{-i}; \mathbf{t}_i\right)\right] = \\
&= \mathbb{E}\left[\exp\left(\alpha + \varepsilon_i\right)\left(\prod_{k=1}^{K} X_{ki}^{\beta_k}\right) S_i^{\gamma}\left(\prod_{j\neq i} S_j^{g_{ij}}\right)^{\delta}\left(\prod_{h=1}^{H} Z_{hi}^{\zeta_h}\right)\right] - \sum_{k=1}^{K} P_{ki}X_{ki} - Q_iS_i \\
&= \exp\left(\alpha\right)\left(\prod_{k=1}^{K} X_{ki}^{\beta_k}\right) S_i^{\gamma}\left(\prod_{j\neq i} S_j^{g_{ij}}\right)^{\delta}\left(\prod_{h=1}^{H} Z_{hi}^{\zeta_h}\right)\mathbb{E}\left[\exp\left(\varepsilon_i\right)\right] - \sum_{k=1}^{K} P_{ki}X_{ki} - Q_iS_i
\end{aligned} \tag{A.2}$$

with corresponding First Order Conditions

$$\frac{\partial U_i(\cdot)}{\partial S_i} = \gamma\bar{A}\left(S_i^*\right)^{-1} - Q_i = 0 \tag{A.3}$$

$$\frac{\partial U_i(\cdot)}{\partial X_{qi}} = \beta_q\bar{A}\left(X_{qi}^*\right)^{-1} - P_{qi} = 0 \quad \text{for } q = 1, \ldots, K \tag{A.4}$$

where

$$\bar{A} \equiv \exp\left(\alpha\right)\left(\prod_{k=1}^{K}\left(X_{ki}^*\right)^{\beta_k}\right)\left(S_i^*\right)^{\gamma}\left(\prod_{j=1}^{N} S_j^{g_{ij}}\right)^{\delta}\left(\prod_{h=1}^{H} Z_{ih}^{\zeta_h}\right)\mathbb{E}\left[\exp\left(\varepsilon_i\right)\right]$$

thus, by combining (A.3) with each of the K conditions in (A.4), one obtains the K marginal rates of technical substitution between effort and each one of the ordinary choice variables:

$$\frac{X_{ki}^*}{S_i^*} = \frac{\beta_k}{\gamma}\frac{Q_i}{P_{ki}} \quad \text{for } k = 1, \ldots, K \tag{A.5}$$

corresponding, upon taking logarithms, to the K equilibrium conditions in (4).

By assuming simultaneous optimizing behavior by all players, substituting (A.5) into (A.3), rearranging and taking logarithms one gets for each player $i = 1, \ldots, N$

$$s_i^* = \frac{1}{1 - \gamma - \sum_{k=1}^K \beta_k}\left(\bar{s}_i + \sum_{h=1}^H \zeta_h z_{hi} + \delta \sum_{j \neq i} g_{ij} s_j^*\right) \tag{A.6}$$

or, in matrix form

$$\mathrm{s}^* = \frac{1}{1 - \gamma - \sum_{k=1}^K \beta_k}\left(\bar{\mathrm{s}} + \mathrm{Z}\boldsymbol{\zeta} + \delta \mathbf{G}\mathrm{s}^*\right) \tag{A.7}$$

which shows how the model can be treated as game of complete information with a single strategic variable. Equation (A.7), in fact, is the expression of the F.O.C.s of the effort variable S_i^* given optimal choices of the $(X_{1i}^*, \ldots, X_{Ki}^*)$ variables. Since the problem is concave (and thus second order conditions are unnecessary), any fixed point s^* of (A.7) allows to retrieve one Bayes-Nash Equilibrium by calculating x_k^* via (4) for $k = 1, \ldots, K$. A closed form solution of the equilibrium can be thus obtained from (A.7) and it is shown in the statement of the Proposition as equation (3). Such a closed form solution can be calculated as long as matrix

$$\mathbf{J} \equiv \mathbf{I} - \frac{\delta}{1 - \gamma - \sum_{k=1}^K \beta_k}\mathbf{G} = \mathbf{I} - \vartheta\mathbf{G}$$

is invertible, which happens to be the case for all but one value of ϑ (that is, matrix $\mathbf{J}$ is invertible almost surely).

To demonstrate conditional uniqueness of the equilibrium, notice that if $\vartheta < 1$ and $\mathbf{G}\iota < \iota$, then (A.7) can be represented as a contraction in $\mathbb{R}_+^N$ with contraction constant $\vartheta \cdot \max_i \sum_{j \neq i} g_{ij}$. Hence, by the Bonacich fixed point Theorem, the associated Bayes-Nash Equilibrium is unique. Notice that the two conditions stated in the text are sufficient but not necessary: in fact, the only necessary condition for uniqueness is that $\vartheta \cdot \max_i \sum_{j \neq i} g_{ij} \in [0, 1)$. However, it is conceptually worth remarking the fact that *both* ϑ and $\max_i \sum_{j \neq i} g_{ij} \in [0, 1)$ must be "small enough" (they are nonnegative by definition) so that their product is smaller than 1. In fact, the two conditions have two different economic interpretations: *1.* $\vartheta < 1 \Leftrightarrow \sum_{k=1}^K \beta_k + \gamma + \delta < 1$ implies decreasing returns to scale from the outcome production technology; *2.* $\mathbf{G}\iota < \iota \Leftrightarrow \max_i \sum_{j \neq i} g_{ij} < 1$ rules out the existence of "highly connected" players, who are very central in the network. Together, the two conditions bound the strategic incentives for the rise of "explosive" multiple equilibria in the game.

A.2 Proof of Proposition 2; Associated Corollaries

Define $\mathrm{v} \equiv \left(1 - \gamma - \sum_{k=1}^{K} \beta_k\right)^{-1} (\bar{\mathrm{s}} + \mathrm{Z}\boldsymbol{\zeta})$. By equation (A.7) one can write:

$$s_i^* = v_i + \vartheta \sum_{\ell \neq i} g_{i\ell} s_\ell^*$$

for any player $i = 1, \ldots, N$. Recall that types $\mathbf{t}_i$ are drawn from an i.i.d. distribution $\mathcal{F}$; since v_i is a function of types $v_i = v(\mathbf{t}_i)$, it is itself drawn from an i.i.d. univariate probability distribution $\mathcal{G}$ that is function of $\mathcal{F}$: $\mathcal{G} = v(\mathcal{F})$. Finally, denote $\mathcal{H}$ as the probability distribution function of s_i^*: since s_i^* is a function of the entire vector v as per (3), draws of s_i^* are not i.i.d. even if individuals types are.

Let us adopt the shorthand notation: $\mathbb{S} = \{s_\ell^* \in \mathbf{s} : g_{i\ell} \neq 0 \vee g_{j\ell} \neq 0\}$ to denote the set of all direct links of both players' i and j as in the Proposition statement. Define $h_{ij}(\cdot)$, $g_{ij}(\cdot)$ and $g(\cdot)$ as the (joint) probability density functions of $\left(s_i^*, s_j^*\right)$, (v_i, v_j), and v_i, respectively. The analysis of the conditional density function

$$\begin{aligned} h_{ij}\left(s_i^*, s_k^* \middle| \mathbb{S}\right) &= h_{ij}\left(v_i + \vartheta \sum_{\ell \neq i} g_{i\ell} s_\ell^*, v_j + \vartheta \sum_{\ell \neq j} g_{j\ell} s_\ell^* \middle| \mathbb{S}\right) \\ &= g_{ij}\left(v_i, v_j \middle| \mathbb{S}\right) \\ &= g\left(v_i \middle| \mathbb{S}\right) \cdot g\left(v_j \middle| \mathbb{S}\right) \end{aligned}$$

reveals it to be equivalent to the product of two independent draws from $\mathcal{G}$. Hence, conditional on $\mathbb{S}$, s_i^* and s_j^* are independent, implying that S_i^* and S_j^* are too.

Corollaries 1 and 2

Given Proposition 2, pairwise independence of types across players, and (A.5), Corollary 1 follows straightforwardly. Similarly, Corollary 2 is deduced from the definition of outcomes Y_i in (1), Proposition 2, Corollary 1 and the assumption of i.i.d. types.

A.3 Proof of Proposition 3

First, re-write (7) as

$$\mathrm{y} = (\mathbf{I} - \lambda \mathbf{G})^{-1} (\eta \boldsymbol{\iota} + \mathrm{X}\boldsymbol{\mu}_x + \mathbf{G}\mathrm{X}\boldsymbol{\nu}_x + \mathrm{Z}\boldsymbol{\mu}_z + \mathbf{G}\mathrm{Z}\boldsymbol{\nu}_z + \boldsymbol{\epsilon}) \tag{A.8}$$

and recall that in equilibrium

$$\mathrm{s} = \frac{1}{1-\gamma} \left(\mathbf{I} - \frac{\delta}{1-\gamma}\mathbf{G}\right)^{-1} (\bar{\bar{\mathrm{s}}} + \mathrm{X}\boldsymbol{\beta} + \mathrm{Z}\boldsymbol{\zeta}) \tag{A.9}$$

where $\bar{\bar{\mathrm{s}}} = (\alpha + \log \gamma + \log \mathbb{E}[\exp(\varepsilon_i)])\, \boldsymbol{\iota} + \mathrm{q}$.

Substitute (A.9) into (8): as a result

$$
\begin{aligned}
\mathrm{y} = & \alpha\iota + \mathrm{X}\boldsymbol{\beta} + \left(\frac{\gamma}{1-\gamma}\mathbf{I} + \frac{\delta}{1-\gamma}\mathbf{G}\right)\left(\mathbf{I} - \frac{\delta}{1-\gamma}\mathbf{G}\right)^{-1}\left(\bar{\bar{\mathrm{s}}} + \mathrm{X}\boldsymbol{\beta} + \mathrm{Z}\boldsymbol{\zeta}\right) + \mathrm{Z}\boldsymbol{\zeta} + \boldsymbol{\varepsilon} \\
= & \left(\mathbf{I} - \frac{\delta}{1-\gamma}\mathbf{G}\right)^{-1}\left(\mathbf{I} - \frac{\delta}{1-\gamma}\mathbf{G}\right)\left[\frac{1}{1-\gamma}\left(\gamma\bar{\bar{\mathrm{s}}} + \mathrm{X}\boldsymbol{\beta} + \mathrm{Z}\boldsymbol{\zeta} + \boldsymbol{\varepsilon}\right) + \alpha\iota\right] + \\
& \frac{\delta}{1-\gamma}\mathbf{G}\left(\mathbf{I} - \frac{\delta}{1-\gamma}\mathbf{G}\right)^{-1}\left(\bar{\bar{\mathrm{s}}} + \mathrm{X}\boldsymbol{\beta} + \mathrm{Z}\boldsymbol{\zeta}\right)
\end{aligned}
$$

which is equivalent to (7) for the following set of restrictions

$$
\begin{aligned}
\lambda &= \frac{\delta}{1-\gamma} \\
\eta &= \frac{\alpha}{1-\gamma} + \frac{\gamma}{1-\gamma}\left(\log\gamma + \log\mathbb{E}\left[\exp\left(\varepsilon_i\right)\right]\right) \\
\boldsymbol{\mu}_x &= \frac{1}{1-\gamma}\boldsymbol{\beta} \\
\boldsymbol{\mu}_z &= \frac{1}{1-\gamma}\boldsymbol{\zeta} \\
\boldsymbol{\nu}_x &= -\frac{\delta}{\left(1-\gamma\right)^2}\boldsymbol{\beta} \\
\boldsymbol{\nu}_z &= -\frac{\delta}{\left(1-\gamma\right)^2}\boldsymbol{\zeta}
\end{aligned}
$$

while $\boldsymbol{\epsilon}$ is determined residually.

A.4 Proof of Proposition 4

We first demonstrate that, because of Assumption 3, there is a bound on the total number of open triads for which any node i can be the central one: this bound is $\binom{\overline{G}}{2}$. Let i be connected to $\overline{G}$ other nodes, which is its maximum degree as per Assumption 3. If none of these nodes are connected to each other, then i is central to exactly $\binom{\overline{G}}{2}$ triads. If at least two nodes among i's links are connected to each other, or if i is connected to $G < \overline{G}$ nodes, it is easy to verify that the number of open triads for which i is the central node is actually smaller than $\binom{\overline{G}}{2}$.

Notice that the sum over all nodes $i = 1, \ldots, N$ of the total number of open triads in which i is the central node is actually equal to the overall number of open triads T_N in the network. Therefore, it follows that T_N is bounded *above* by $\binom{\overline{G}}{2}N$, and thus $T_N = O\left(N\right)$. However, by Assumption 4 every node is present in at least one triad, hence the number of triads is also bounded *below* by $\frac{N}{3}$, so that $N = O\left(T_N\right)$. Consequently, $\frac{N}{T_N} = O\left(1\right)$, demonstrating our claim.

A.5 Proof of Proposition 5

We demonstrate first a preliminary, ancillary result.

Lemma A.1. *Let $a_{ij,N}$ and $b_{i,N}$ be given by*

$$a_{ij,N} = \begin{bmatrix} 0_{K+2} \\ g_{ij} \end{bmatrix} \tag{A.10}$$

$$b_{i,N} = \begin{bmatrix} 1 \\ \overline{x}_i \\ c_i \\ 0 \end{bmatrix} \tag{A.11}$$

then, under Assumptions 3 and 6, the following inequalities hold.

$$\frac{1}{N}\sum_{i=1}^{N} |a_{ij,N}| < \infty \tag{A.12}$$

$$\frac{1}{N}\sum_{i=1}^{N} |b_{i,N}|^{2+\eta} < \infty \tag{A.13}$$

To see this, consider that

$$\begin{aligned} \frac{1}{N}\sum_{i=1}^{N} |a_{ij,N}| &= \frac{1}{N}\sum_{i=1}^{N} \begin{pmatrix} 0_{K+2} \\ |g_{ij}| \end{pmatrix} \\ &\leq \frac{1}{T_N} \begin{pmatrix} 0_{K+2} \\ N\overline{G} \end{pmatrix} < \infty \end{aligned}$$

where the inequality follows by Assumption 3. Similarly,

$$\begin{aligned} \frac{1}{N}\sum_{i=1}^{N} |b_{i,N}|^{2+\eta} &= \frac{1}{N}\sum_{i=1}^{N} \begin{pmatrix} 1 \\ |\overline{x}_i|^{2+\eta} \\ |c_i|^{2+\eta} \\ 0 \end{pmatrix} \\ &\leq \begin{pmatrix} 1 \\ \max_i |\overline{x}_i|^{2+\eta} \\ \max_i |c_i|^{2+\eta} \\ 0 \end{pmatrix} < \infty \end{aligned}$$

where, instead, the inequality holds by Assumption 6.

Armed with this result, we move to the actual demonstration of Proposition 5. Throughout the proof we use the matrix $\frac{\partial \mathrm{m}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}$, which is the horizontal concatenation of the following matrices:

$$\frac{\partial \mathrm{m}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\alpha}} = -\begin{bmatrix} 1 \\ \overline{x}_i \\ c_i \\ \sum_{j\neq i} g_{ij}\left(e_i(\boldsymbol{\theta}) + e_j(\boldsymbol{\theta})\right) \end{bmatrix}$$

$$\frac{\partial \mathrm{m}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} = -\begin{bmatrix} x_i' \\ \overline{x}_i x_i' \\ c_i x_i' \\ \sum_{j\neq i} g_{ij}\left(e_i(\boldsymbol{\theta}) x_j' + e_j(\boldsymbol{\theta}) x_i'\right) \end{bmatrix}$$

$$\frac{\partial \mathrm{m}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma}} = -\begin{bmatrix} s_i \\ \overline{x}_i s_i \\ c_i s_i \\ \sum_{j\neq i} g_{ij}\left(e_i(\boldsymbol{\theta}) s_j + e_j(\boldsymbol{\theta}) s_i\right) \end{bmatrix}$$

$$\frac{\partial \mathrm{m}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\delta}} = -\begin{bmatrix} \sum_{j\neq i} g_{ij} s_j \\ \overline{x}_i \sum_{j\neq i} g_{ij} s_j \\ c_i \sum_{j\neq i} g_{ij} s_j \\ \sum_{j\neq i} g_{ij}\left(e_i(\boldsymbol{\theta}) \sum_{h=1}^{N} g_{ih} s_h + e_j(\boldsymbol{\theta}) \sum_{l=1}^{N} g_{jl} s_l\right) \end{bmatrix}$$

Denote, as per typical conventions, $\boldsymbol{\theta}_0$ as the vector of the true parameter values, and $\boldsymbol{\theta}$ as any possible value in the associated bounded space $\boldsymbol{\Theta}$. We can rewrite the vector of moments conditions $\overline{\mathrm{m}}_N(\boldsymbol{\theta})$ as:

$$\begin{aligned} \overline{\mathrm{m}}_N(\boldsymbol{\theta}) &= \frac{1}{N}\sum_{i=1}^{N} \mathrm{m}_i(\boldsymbol{\theta}) \\ &= \frac{1}{N}\sum_{i=1}^{N}\sum_{j\neq i} a_{ij,N} e_i(\boldsymbol{\theta}) e_j(\boldsymbol{\theta}) + \frac{1}{N}\sum_{i=1}^{N} b_{i,N} w_i(\boldsymbol{\theta}) \end{aligned}$$

where $a_{ij,N}$ and $b_{i,N}$ are defined in (A.10) and (A.11). By Assumptions 1 and 2, and thanks to Lemma A.1, it is possibly to apply lemmas A.3 and A.4 in Lee (2007) and Proposition 1 in Kelejian and Prucha (2001). Let

$$\begin{aligned} e_j(\boldsymbol{\theta}) &= y_i - \boldsymbol{\alpha}_0 - \sum_{k=1}^{K} \boldsymbol{\beta}_{k0} x_{ik} - \boldsymbol{\gamma}_0 s_i - \boldsymbol{\delta}_0 \sum_{j\neq i} g_{ij} s_j \\ &\quad + (\boldsymbol{\alpha}_0 - \boldsymbol{\alpha}) + \sum_{k=1}^{K} (\boldsymbol{\beta}_{k0} - \boldsymbol{\beta}_k) x_{ik} - (\boldsymbol{\gamma}_0 - \boldsymbol{\gamma}) - (\boldsymbol{\delta}_0 - \boldsymbol{\delta}) \sum_{j\neq i} g_{ij} s_j \\ &\equiv \omega_j + d_j(\boldsymbol{\theta}) \end{aligned}$$

hence, $\overline{\text{m}}_N(\boldsymbol{\theta})$ can be further decomposed as follows.

$$\overline{\text{m}}_N(\boldsymbol{\theta}) = \frac{1}{T_N}\sum_{i=1}^{N}\sum_{j\neq i} a_{ij,N}e_ie_j + \frac{1}{T_N}\sum_{i=1}^{N} b_{i,N}e_i + \frac{1}{T_N}\sum_{i=1}^{N}\sum_{j\neq i} a_{ij,N}e_id_j(\boldsymbol{\theta})$$
$$+\frac{1}{T_N}\sum_{i=1}^{N}\sum_{j\neq i} a_{ij,N}d_i(\boldsymbol{\theta})e_j + \frac{1}{T_N}\sum_{i=1}^{N}\sum_{j\neq i} a_{ij,N}d_i(\boldsymbol{\theta})d_j(\boldsymbol{\theta}) + \frac{1}{T_N}\sum_{i=1}^{N} b_{i,N}d_i(\boldsymbol{\theta})$$

Thanks to A.1 and Lemma A.4 in Lee (2007), we have

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{j\neq i} a_{ij,N}e_id_j(\boldsymbol{\theta}) = o_P(1)$$
$$\frac{1}{N}\sum_{i=1}^{N}\sum_{j\neq i} a_{ij,N}d_i(\boldsymbol{\theta})e_j = o_P(1)$$

uniformly in $\boldsymbol{\theta}$. By Lemma A.1 here and Lemma A.3 in Lee (2007),

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{j\neq i} a_{ij,N}\omega_i\omega_j = \frac{1}{N}\mathbb{E}\left[\sum_{i=1}^{N}\sum_{j\neq i} a_{ij,N}\omega_i\omega_j\right] + o_P(1) = o_P(1)$$

uniformly in $\boldsymbol{\theta}$. Furthermore, by the weak Law or Large Numbers,

$$\frac{1}{N}\sum_{i=1}^{N} b_{i,N}\omega_i = o_P(1)$$

uniformly in $\boldsymbol{\theta}$. Since $\text{m}_i(\boldsymbol{\theta})$ is a quadratic function of $\boldsymbol{\theta}$ and $\boldsymbol{\Theta}$ is bounded, $\overline{\text{m}}_N(\boldsymbol{\theta})$ is uniformly equicontinuous on $\boldsymbol{\Theta}$, which together with the identification condition of $\boldsymbol{\theta}$ implies us identifiable uniqueness. Convergence of $\hat{\boldsymbol{\theta}}_{GMM}$ to $\boldsymbol{\theta}_0$ finally follows from uniform convergence of $Q_N(\boldsymbol{\theta})$ and identifiable uniqueness.

By Assumption 5, the minimization of $Q_N(\boldsymbol{\theta}) = \overline{\text{m}}_N(\boldsymbol{\theta})'\mathbf{W}_N\overline{\text{m}}_N(\boldsymbol{\theta})$ implies that

$$\nabla_{\boldsymbol{\theta}}Q_N\left(\hat{\boldsymbol{\theta}}\right) = 2\nabla_{\boldsymbol{\theta}}\overline{\text{m}}_N\left(\hat{\boldsymbol{\theta}}\right)'\mathbf{W}_N\overline{\text{m}}_N\left(\hat{\boldsymbol{\theta}}\right) = 0 \tag{A.14}$$

and since

$$\text{m}_N\left(\hat{\boldsymbol{\theta}}\right) = \overline{\text{m}}_N(\boldsymbol{\theta}_0) + \nabla_{\boldsymbol{\theta}}\overline{\text{m}}_N\left(\overline{\boldsymbol{\theta}}\right)\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)$$

for some $\overline{\boldsymbol{\theta}} \in \left[\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0\right]$ by the Mean Value Theorem, substitution into (A.14) gives

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) = \left[\nabla_{\boldsymbol{\theta}}\overline{\text{m}}_N\left(\hat{\boldsymbol{\theta}}\right)'\mathbf{W}_N\nabla_{\boldsymbol{\theta}}\overline{\text{m}}_N\left(\overline{\boldsymbol{\theta}}\right)\right]^{-1}\nabla_{\boldsymbol{\theta}}\overline{\text{m}}_N\left(\hat{\boldsymbol{\theta}}\right)'\mathbf{W}_N\sqrt{N}\overline{\text{m}}_N(\boldsymbol{\theta}_0)$$

after rearranging terms.

By Theorem 1 in Kelejian and Prucha (2001):

$$\sqrt{N}\overline{\mathrm{m}}_N\left(\boldsymbol{\theta}_0\right) \xrightarrow{d} \mathcal{N}\left(0, \mathbf{V}_0\right)$$

and by the uniform Law of Large Numbers:

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\nabla_{\boldsymbol{\theta}}\overline{\mathrm{m}}_N\left(\boldsymbol{\theta}\right) - \mathbb{E}\left[\nabla_{\boldsymbol{\theta}}\mathrm{m}_i\left(\boldsymbol{\theta}\right)\right]\right\| = o_P\left(1\right)$$

moreover, by Assumption 7

$$\left\|\mathbf{W}_N - \mathbf{W}_0\right\| = o_P\left(1\right)$$

uniformly in $\boldsymbol{\theta}$, and by Assumption 8

$$\begin{aligned}\left|\nabla_{\boldsymbol{\theta}}\overline{\mathrm{m}}_N\left(\boldsymbol{\theta}\right)\right| &\leq \overline{M}\\ \left|\mathbb{E}\left[\nabla_{\boldsymbol{\theta}}\mathrm{m}_i\left(\boldsymbol{\theta}\right)\right]\right| &\leq \overline{M}\end{aligned}$$

then it follows that

$$\begin{aligned}&\left|\nabla_{\boldsymbol{\theta}}\overline{\mathrm{m}}_N\left(\boldsymbol{\theta}\right)'\mathbf{W}_N\nabla_{\boldsymbol{\theta}}\overline{\mathrm{m}}_N\left(\boldsymbol{\theta}\right) - \mathbb{E}\left[\nabla_{\boldsymbol{\theta}}\mathrm{m}_i\left(\boldsymbol{\theta}\right)\right]'\mathbf{W}_0\mathbb{E}\left[\nabla_{\boldsymbol{\theta}}\mathrm{m}_i\left(\boldsymbol{\theta}\right)\right]\right|\\ &\qquad\leq \left|\nabla_{\boldsymbol{\theta}}\overline{\mathrm{m}}_N\left(\boldsymbol{\theta}\right)'\mathbf{W}_N\nabla_{\boldsymbol{\theta}}\overline{\mathrm{m}}_N\left(\boldsymbol{\theta}\right) - \mathbb{E}\left[\nabla_{\boldsymbol{\theta}}\mathrm{m}_i\left(\boldsymbol{\theta}\right)\right]'\mathbf{W}_N\nabla_{\boldsymbol{\theta}}\overline{\mathrm{m}}_N\left(\boldsymbol{\theta}\right)\right|\\ &\qquad\quad + \left|\mathbb{E}\left[\nabla_{\boldsymbol{\theta}}\mathrm{m}_i\left(\boldsymbol{\theta}\right)\right]'\mathbf{W}_N\nabla_{\boldsymbol{\theta}}\overline{\mathrm{m}}_N\left(\boldsymbol{\theta}\right) - \mathbb{E}\left[\nabla_{\boldsymbol{\theta}}\mathrm{m}_i\left(\boldsymbol{\theta}\right)\right]'\mathbf{W}_N\mathbb{E}\left[\nabla_{\boldsymbol{\theta}}\mathrm{m}_i\left(\boldsymbol{\theta}\right)\right]\right|\\ &\qquad\quad + \left|\mathbb{E}\left[\nabla_{\boldsymbol{\theta}}\mathrm{m}_i\left(\boldsymbol{\theta}\right)\right]'\mathbf{W}_N\mathbb{E}\left[\nabla_{\boldsymbol{\theta}}\mathrm{m}_i\left(\boldsymbol{\theta}\right)\right] - \mathbb{E}\left[\nabla_{\boldsymbol{\theta}}\mathrm{m}_i\left(\boldsymbol{\theta}\right)\right]'\mathbf{W}_0\mathbb{E}\left[\nabla_{\boldsymbol{\theta}}\mathrm{m}_i\left(\boldsymbol{\theta}\right)\right]\right|\\ &\qquad\leq \left|\nabla_{\boldsymbol{\theta}}\overline{\mathrm{m}}_N - \mathbb{E}\left[\nabla_{\boldsymbol{\theta}}\mathrm{m}_i\left(\boldsymbol{\theta}\right)\right]'\right|\left|\mathbf{W}_N\right|\left|\nabla_{\boldsymbol{\theta}}\overline{\mathrm{m}}_N\left(\boldsymbol{\theta}\right)\right|\\ &\qquad\quad + \left|\mathbb{E}\left[\nabla_{\boldsymbol{\theta}}\mathrm{m}_i\left(\boldsymbol{\theta}\right)\right]'\right|\left|\mathbf{W}_N\right|\left|\nabla_{\boldsymbol{\theta}}\overline{\mathrm{m}}_N\left(\boldsymbol{\theta}\right) - \mathbb{E}\left[\nabla_{\boldsymbol{\theta}}\mathrm{m}_i\left(\boldsymbol{\theta}\right)\right]\right|\\ &\qquad\quad + \left|\mathbb{E}\left[\nabla_{\boldsymbol{\theta}}\mathrm{m}_i\left(\boldsymbol{\theta}\right)\right]'\right|\left|\mathbf{W}_N - \mathbf{W}_0\right|\left|\mathbb{E}\left[\nabla_{\boldsymbol{\theta}}\mathrm{m}_i\left(\boldsymbol{\theta}\right)\right]\right|\\ &\qquad= o_P\left(1\right)\end{aligned}$$

uniformly in $\boldsymbol{\theta}$. This, together with consistency of $\hat{\boldsymbol{\theta}}$ and Assumption 9, allows to finally establish that

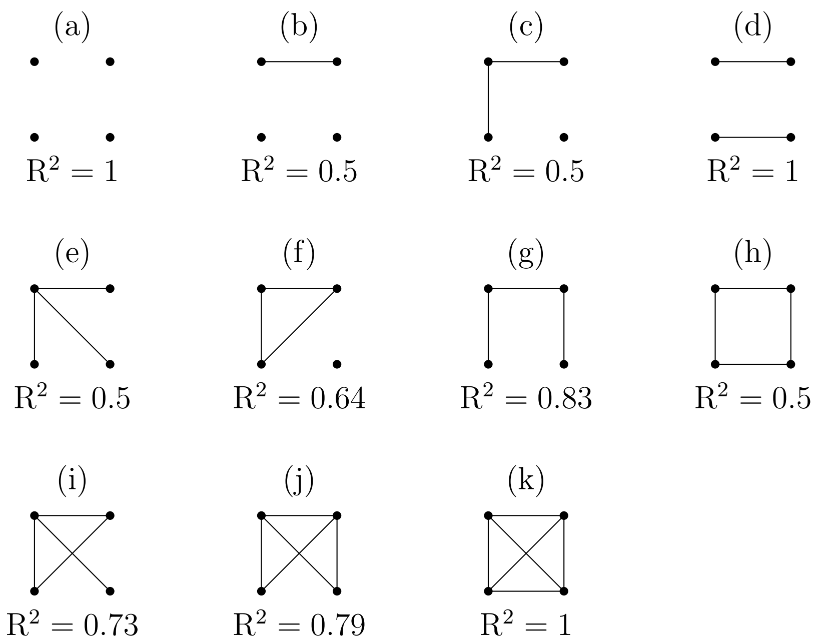
$$\sqrt{N}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) = \left[\mathbf{M}_0'\mathbf{W}_0\mathbf{M}_0\right]^{-1}\mathbf{M}_0'\mathbf{W}_0\sqrt{N}\overline{\mathrm{m}}_N\left(\boldsymbol{\theta}_0\right) + o_P\left(1\right) \xrightarrow{d} \mathcal{N}\left(0, \mathbf{V}_{\boldsymbol{\theta}}\right)$$

given matrices $\mathbf{V}_{\boldsymbol{\theta}}$, $\mathbf{V}_0$, $\mathbf{M}_0$, and $\mathbf{W}_0$ as defined in the statement of the Proposition.

Appendix B Tetrad Analysis

In Graph 2 we present a breakdown of the eleven different types of tetrad (a tetrad is any network or subset of a network, composed by four nodes) which are distinguished by the different configuration of their edges. Notice that the order of the nodes does not matter for the purpose of identifying a specific tetrad type. For example, given any set of four nodes labeled (A,B,C,D), tetrad (b) represents all six possible networks characterized by a single edge connecting any two nodes, whether we represent it graphically as a single vertical, horizontal or diagonal segment – wherever placed.

We report the measure of R^2 associated with each type of tetrad, calculated by running regression (15). Graph 2 shows how high values of the R^2 coefficient, which generally reflects the degree of *symmetry* of one tetrad, is generally associated with very dense or very sparse networks. In particular, $R^2 = 1$ either for the "empty" tetrad (a), for the "totally connected" tetrad (k), or for the otherwise symmetric tetrad (d).



Graph 2: The 11 Types of Tetrad