The University of Queensland
AUSTRALIA
CREATE CHANGE

Discussion Paper Series
UQ School of Economics

Identification of Peer Effects using Panel Data
2020-12-03
Marisa Miraldo, Carol Propper, Christiern Rose

Discussion Paper No.639

Identification of Peer Effects using Panel Data

Marisa Miraldo, Carol Propper, Christiern Rose*

Abstract

This paper provides new identification results for panel data models with contextual and endogenous peer effects. Contextual effects operate through individuals' time-invariant unobserved heterogeneity. Identification hinges on a conditional mean restriction requiring exogenous mobility of individuals between groups over time. Some networks governing peer interactions preclude identification. For these cases we propose additional conditional variance restrictions. We conduct a Monte-Carlo experiment to evaluate the performance of our method and apply it to surgeon-hospital-year data to study take-up of minimally invasive surgery. A standard deviation increase in the average time-invariant unobserved heterogeneity of other surgeons in the same hospital leads to a 0.12 standard deviation increase in take-up. The effect is equally due to endogenous and contextual effects.

Key words: Peer effects, panel data, networks, identification, innovation, health-care

1. Introduction

This paper provides new identification results for panel data models of peer effects. We consider *contextual* peer effects, through which outcomes depend on peers' time-invariant unobservable (to the researcher) heterogeneity, and *endogenous* peer

*Miraldo and Propper: Imperial College Business School, Rose: University of Queensland. Corresponding author: Christiern Rose, School of Economics, University of Queensland, 4072, Australia, christiern.rose@uq.edu.au. This project is part of the Health Foundation's Efficiency Research programme. The Health Foundation is an independent charity working to improve the quality of healthcare in the UK. We are thankful for their financial support. Propper acknowledges financial support from the ERC. Hospital Episode Statistics datasets were provided by the Health and Social Care Information Centre (HSCIC, now NHS Digital).

effects through which outcomes are simultaneously determined. Our framework also allows for *correlated* effects, modelled as unobserved peer group heterogeneity which is permitted to be correlated with individual heterogeneity in an unrestricted manner. This permits, for example, that high outcome individuals be systematically located in high outcome peer groups. We derive identification results which can be applied to general network structures governing peer interactions and are straightforward to verify in practice. Identification hinges primarily on a conditional mean restriction requiring exogenous mobility of individuals between groups over time. Not all patterns of mobility suffice to identify the peer effects, and we provide identifying and non-identifying examples. If there are both contextual and endogenous effects, our identification results show that for some networks there do not exist identifying mobility patterns.[1] For these cases we propose additional conditional variance restrictions.

We conduct a Monte-Carlo experiment to compare the performance of the conditional mean based fixed-T consistent non-linear least squares estimator of the contextual effects (Arcidiacono et al., 2012) with the fixed-T consistent non-linear least squares estimator of the endogenous and contextual effects resulting from the conditional variance restrictions. Both estimators perform well if there are contextual effects only but the conditional mean estimator (in addition to requring weaker identifying assumptions) is more efficient. With endogenous effects the conditional mean estimator does not accurately estimate the reduced form parameters, whereas the conditional variance estimator accurately estimates both endogenous and contextual effects. Increasing the number of time periods, the rate of mobility and the richness of the network data (e.g. social network data) improves the performance of both estimators.

We use the matched surgeon-patient-hospital-year data construced by Barrenho et al. (2019) to illustrate our method by studying surgeons' take-up of minimally invasive (keyhole) surgery for colorectal cancer surgery in the English National Health Service. Colorectal cancer is the third most common cancer worldwide (Arnold et al., 2017). In England, it accounts for 10% of cancer deaths and is the most expensive cancer to treat (Laudicella et al., 2016). Our empirical innovation is to allow for contextual effects operating through surgeons' latent propensity to take up the technique (i.e. their time-invariant unobserved heterogeneity). We find positive and statistically significant endogenous and contextual effects in take-up. Our results suggest that a standard deviation increase in the average latent take-up of other surgeons in the same hospital leads to a net increase in take-up of 0.12 standard deviations. Endogenous and contextual peer effects each make up

[1] We consider a model of endogenous effects in which outcomes are simultaneously determined, which is appropriate for our empirical application.

approximately half of the net effect.

Our work contributes to the literature on identification of peer effects. To date, research has largely focussed on settings in which contextual effects operate through exogenous observable individual characteristics. That is, in addition to endogenous and correlated effects, outcomes depend on individual and peers' exogenous observables (Manski, 1993; Moffitt et al., 2001; Lee, 2007; Bramoullé et al., 2009; Calvó-Armengol et al., 2009; Davezies et al., 2009; De Giorgi et al., 2010; Goldsmith-Pinkham and Imbens, 2013; Blume et al., 2015; De Paula, 2017; Cohen-Cole et al., 2018; Bramoullé et al., 2019). This strand of the literature focuses on cross-section data, for which identification prospects depend on cross-sectional variation in peer groups and the extent to which exogenous individual characteristics determine outcomes. If there is sufficient variation in peer groups, the exogenous characteristics of peers-of-peers can be used as instrumental variables for peers' outcomes. In contrast, our panel data allow us to capture contextual effects operating through individual unobserved heterogeneity and to identify endogenous effects without needing to observe exogenous characteristics. Our method can be applied with a panel comprising only outcome and group membership data. Another strand of the literature considers peer effects operating through individual unobservables using cross-section data. The most closely related papers are Graham (2008) and Rose (2017). Our conditional variance restrictions are panel data counterparts of those of Rose (2017). Relative to these papers, we allow for unobserved group level heterogeneity to be arbitrarily correlated with individual heterogeneity, and, though our results can be applied to any network structure, we show that panel data can be used to identify endogenous peer effects for the linear-in-means network which, in the absence of additional restrictions, precludes identification using cross-section data (Manski, 1993; Rose, 2017; Bramoullé et al., 2009).[2]

The most closely related strand of the literature focuses on panel data models of peer effects. Key contributions are Mas and Moretti (2009) and Arcidiacono et al. (2012), both of which study models of contextual peer effects comprising the average unobserved time-invariant heterogeneity of the other members of the peer group. We allow additionally for a general network structure (e.g. a social network), for simultaneously determined outcomes (endogenous effects) and for peer group level unobserved heterogeneity (correlated effects). Arcidiacono et al. (2012) consider a variant of endogenous effects through which outcomes depend on the expected (as opposed to realised) outcomes of others. This implies that there is no simulataneity in outcomes, which is appropriate for the authors' application to peer effects in education. Arcidiacono et al. (2012) show that identification and fixed-T consistent

[2] The linear-in-means network is one through which peer effects are constructed using the group average.

estimation can be attained based on an instrumental variables approach which requires that there exist observable exogenous individual characteristics which vary over time and determine an individual's own outcome but not those of their peers. In our empirical application we expect that surgeons directly observe and base their take-up on the take-up of other surgeons in the same hospital, leading to a model of endogenous effects with simultaineity, for which the estimator of Arcidiacono et al. (2012) cannot be shown to be consistent (Arcidiacono et al., 2012). Instead, we propose a fixed-T estimator based around conditional variance restrictions which does not require exogenous instrumental variables. This approach is in the same spirit as Graham (2008) and Rose (2017).

Beyond the peer effects literature our work can be viewed as extending the canonical worker-firm fixed effects framework of Abowd et al. (1999) in labour economics to allow for within-firm interactions of workers. That is, in addition to worker and firm heterogeneity, wages may depend on the composition of other workers in the firm as well as their wages. Such spillovers would be expected to operate in firms in which workers work in teams Mas and Moretti (2009).

We proceed as follows. In Section 2, we present our baseline model. In Section 3 we provide identification results based on a conditional mean restriction and apply them to two examples. In Section 4 we consider endogenous effects and provide additional identification results based on a conditional variance restriction. In Section 5 we consider a more general form of correlated effect, which is permitted to be group-year specific as opposed to group specific. In Sections 6-7 we conduct a Monte-Carlo experiment and apply our method to the NHS data. In Section 8 we conclude. All proofs are provided in the Appendix.

1.1. Notation

If A is a strictly positive integer, we denote by $[A] = \{1, 2, ..., A\}$. If $\mathbf{A}$ and $\mathbf{B}$ are $M \times P$ and $M \times Q$ matrices, we denote the $M \times (P + Q)$ matrix obtained by concatenating $\mathbf{A}$ and $\mathbf{B}$ by $(\mathbf{A}, \mathbf{B})$. If element ij of $\mathbf{A}$ is $\mathbf{A}_{ij}$, we write $\mathbf{A} = (\mathbf{A}_{ij})_{i\in[M],j\in[P]}$. We use $\mathbf{I}_M$ for the M dimensional identity, ι_M for the $M \times 1$ vector of ones and $\mathbf{0}$ to denote a matrix of zeros. If its dimensions are ambiguous we write $\mathbf{0}_{M,P}$ to denote a $M \times P$ matrix of zeros. We use $\mathbf{1}(\cdot)$ to denote the indicator function.

2. Model

We first present the model and identification results with contextual effects only. Endogenous effects are considered in Section 4. There are N individuals and M

groups observed for T years. In each year, every individual is in exactly one group. Every group has at least one individual in at least one year. For individual $i \in [N]$ in year $t \in [T]$ in group $g(i,t) \in [M]$ of size $N_{g(i,t)}$, the outcome y_{it} is

$$y_{it} = \alpha_i + \rho \sum_{j=1}^{N} \mathbf{G}_{ijt}\alpha_j + \gamma_{g(i,t)} + \epsilon_{it}, \tag{2.1}$$

and hence comprises an individual effect, a contextual effect, a group level correlated effect and a disturbance. Stacking by individual in year t yields

$$\mathbf{y}_t = (\mathbf{I}_N + \rho\mathbf{G}_t)\,\alpha + \mathbf{C}_t\gamma + \epsilon_t \tag{2.2}$$

where $\mathbf{y}_t$ and ϵ_t are $N \times 1$, $\mathbf{C}_t$ is the $N \times M$ matrix of group membership indicators in year t and $\mathbf{G}_t = (\mathbf{G}_{ijt})_{(i,j)\in[N]^2}$. The matrix $\mathbf{G}_t$ summarises the structure of interactions governing the peer effect, and can evolve over time as individuals move between groups. Unless otherwise stated $\mathbf{G}_t$ is unrestricted and can be interpreted as the adjacency matrix of a weighted, directed network linking N individuals. In particular, it need not be the case that $\mathbf{G}_{ijt} = 0$ when $g(i,t) \neq g(j,t)$, nor when $i = j$, nor when $\mathbf{G}_{jit} = 0$. A typical example of $\mathbf{G}_t$ is the *linear-in-means* network

$$\overline{\mathbf{G}}_t = \left(\mathbf{1}(g(i,t) = g(j,t))N_{g(i,t)}^{-1}\right)_{(i,j)\in[N]^2}.$$

The linear-in-means network implies that the contextual effect is the average individual effect over all individuals in the same group. This is a natural choice when only group membership indicators are available, which we use to illustrate our identification results. If more detailed data on the structure of within-group or between-group interactions are available, this information can be incorporated into $\mathbf{G}_t$. Stacking (2.2) by year yields $\mathbf{y} = (\mathbf{J} + \rho\mathbf{G})\,\alpha + \mathbf{C}\gamma + \epsilon$, where $\mathbf{y}$ and ϵ are $NT \times 1$, $\mathbf{C} = (\mathbf{C}_1', \mathbf{C}_2', \ldots, \mathbf{C}_T')'$, $\mathbf{J} = (\mathbf{I}_N, \mathbf{I}_N, \ldots, \mathbf{I}_N)'$ and $\mathbf{G} = (\mathbf{G}_1', \mathbf{G}_2', \ldots, \mathbf{G}_T')'$. Since $\sum_{i=1}^{N} \mathbf{J}_{ki} = \sum_{f=1}^{M} \mathbf{C}_{kf} = 1$ for $k = 1, ..., NT$, α and γ are only separately identifiable up to a normalization. Using the normalization $\gamma_M = 0$, we obtain

$$\mathbf{y} = (\mathbf{J} + \rho\mathbf{G})\,\alpha + \mathbf{D}\gamma + \epsilon$$

where, $\mathbf{D}$ comprises the first $M - 1$ columns of $\mathbf{C}$ and from this point forwards $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_{M-1})'$. Our identification results depend on three sources of variation of variation in $\mathbf{J}$, $\mathbf{G}$ and $\mathbf{D}$:

(i) **Mobility between groups over time**

$$\exists (i,j,s,t) \in [N]^2 \times [T]^2 : \quad \mathbf{G}_{ijs} \neq \mathbf{G}_{ijt};$$

(ii) **Heterogeneous intra-group interactions**

$$\exists (i,j,k,l,t) \in [N]^4 \times [T] : \quad g(i,t) = g(j,t) = g(k,t) = g(l,t), \quad \mathbf{G}_{ijt} \neq \mathbf{G}_{klt};$$

(iii) **Existence of inter-group interactions**

$$\exists (i,j,t) \in [N]^2 \times [T] : \quad g(i,t) \neq g(j,t), \quad \mathbf{G}_{ijt} \neq 0$$

Mobility serves to separate the individual and correlated effects (Abowd et al., 1999) but also to separate these from the contextual effect. This is because, if an individual moves from one group to another she ceases to be a peer of others in her previous group and becomes a peer of others in her new group. Heterogeneous intra-group interactions lead to within-group variation in the contextual effect, separating it from the correlated effect. Existence of inter-group interactions also separates the contextual effect from the correlated effect.

3. Identification

We first study identification of α, γ, ρ based on the moment condition

$$\mathbb{E}[\mathbf{y}|\mathbf{G}, \mathbf{D}, \alpha, \gamma] = (\mathbf{J} + \rho \mathbf{G})\,\alpha + \mathbf{D}\gamma, \tag{3.1}$$

which implies exogenous mobility of individuals between groups over time (Abowd et al., 1999). Since (3.1) is non-linear in the parameters, establishing identification is non-trivial, depending both on the properties of the $NT \times 2N + M - 1$ matrix $(\mathbf{J}, \mathbf{G}, \mathbf{D})$ and on the values of α and ρ. For example, it is clear that ρ is not identified when $\alpha = \mathbf{0}$. Moreover, if the rows of $\mathbf{G}$ sum to one (e.g. the linear-in-means network), α is not identifed when $\rho = -1$, as explained below. That identification is non-uniform over the parameter space is common in models of peer effects. For example, in models with endogenous and contextual effects, identification fails whenever these exactly offset one another such that there is no net peer effect (Bramoullé et al., 2009; Rose, 2017). As in Abowd et al. (1999) and Arcidiacono et al. (2012), mobility of individuals between groups over time is necessary for identification. In the absence of mobility, $(\mathbf{J}, \mathbf{G}, \mathbf{D})$ has rank at most N, so (3.1) yields N linearly independent equations in $N + M$ unknowns. For the same reason, we also require $T \geq 2$. The formal identification result below makes use of the within-group annihilator for the correlated effects, given by

$$\mathbf{W} = \mathbf{I}_{NT} - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}$$

and a decomposition of vectors $\mathbf{V} = (\mathbf{V}_1', \mathbf{V}_2')'$ which lie in the null-space of $(\mathbf{WJ}, \mathbf{WG})$ such that $\mathbf{V}_1$ and $\mathbf{V}_2$ are both $N \times 1$.

Proposition 1
ρ is identified if there also does not exist a vector $\mathbf{V} = (\mathbf{V}_1', \mathbf{V}_2')'$ in the null-space of $(\mathbf{WJ}, \mathbf{WG})$ and scalars δ_1 and $\delta_2 \neq 0$ verifying $\delta_1\mathbf{V}_1 + \delta_2\mathbf{V}_2 = \alpha$. Otherwise ρ is not identified. α, γ are identified if additionally $(\mathbf{J} + \rho\mathbf{G}, \mathbf{D})$ has full column rank.

Notice that we do not require that $(\mathbf{J}, \mathbf{G}, \mathbf{D})$ has have full column rank since there is no requirement that $\mathbf{V} = \mathbf{0}$ be the only vector in the null-space of $(\mathbf{WJ}, \mathbf{WG})$. This is because $(\mathbf{J}, \mathbf{G}, \mathbf{D})$ has $2N + M - 1$ columns but there are only $N + M$ unknowns. Requiring $(\mathbf{J}, \mathbf{G}, \mathbf{D})$ to have full column rank is too strong an assumption because it rules out $T = 2$ if $M > 1$. Proposition 1 shows that ρ is not identified for some values of α. Clearly, the identification condition for ρ cannot be verified in practice because α is not observed. Instead, the researcher can ask how 'large' is the set of values of α for which ρ is not identified. This is the notion of generic identification (see Lewbel (2019)). For this purpose we can use the following Corollary.

Corollary 1.1
If $\text{rank}(\mathbf{WJ}, \mathbf{WG}) \geq N + 1$ *the set of α for which ρ is not identified is a measure zero subset of $\mathbb{R}^N$.*

Corollary 1.1 means that ρ is generically identified if $\text{rank}(\mathbf{WJ}, \mathbf{WG}) \geq N + 1$. This is because $\mathbf{V}$ lies in a subspace of $\mathbb{R}^{2N}$ of dimension at most $N - 1$, hence $\delta_1\mathbf{V}_1 + \delta_2\mathbf{V}_2$ lies in a subspace of $\mathbb{R}^N$ of dimension at most $N - 1$. In practice, if $\text{rank}(\mathbf{WJ}, \mathbf{WG}) \geq N + 1$ the researcher can be confident that ρ is identified because it is overwhelmingly 'unlikely' that the identification condition in Proposition 1 is violated. Given identification of ρ, the second part of Proposition 1 establishes identification of α, γ using the standard rank condition for linear models.

We now consider two concrete cases with a view to making explicit which values of α preclude identification of ρ. Since $\mathbf{V} = \mathbf{0}$ is always in the null-space of $(\mathbf{WJ}, \mathbf{WG})$, ρ is not identified when $\alpha = \mathbf{0}$. If $(\mathbf{WJ}, \mathbf{WG})$ (or equivalently $(\mathbf{J}, \mathbf{G}, \mathbf{D})$) has full column rank, then $\mathbf{V} = \mathbf{0}$ is the only possibility and we have the following.

Corollary 1.2
ρ, α and γ are identified when $(\mathbf{J}, \mathbf{G}, \mathbf{D})$ has full column rank and $\alpha \neq \mathbf{0}$.

A common empirical setting arises when the rows of $\mathbf{G}$ sum to one (e.g. the linear-in-means network), implying that $(\mathbf{J}, \mathbf{G}, \mathbf{D})$ cannot have full column rank. If there are no further linear relationships among the columns of $(\mathbf{J}, \mathbf{G}, \mathbf{D})$ then one can apply the following Corollary.

Corollary 1.3
If $\mathbf{G}\iota_N = \iota_{NT}$*,* α, ρ *and* γ *are identified if* $\text{rank}(\mathbf{J}, \mathbf{G}, \mathbf{D}) = 2N + M - 2$*, there exists* $(i, j) \in [N]^2$ *such that* $\alpha_i \neq \alpha_j$ *and* $\rho \neq -1$*.*

Corollary 1.3 can be shown using the decomposition $(\mathbf{WJ}, \mathbf{WG}) = \mathbf{SR}$ where $\mathbf{S}$ is the $NT \times 2N - 1$ full rank matrix formed by concatenating $\mathbf{WJ}$ and the first $N - 1$ columns of $\mathbf{WG}$ and

$$\mathbf{R} = (\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'(\mathbf{WJ}, \mathbf{WG}) = \begin{pmatrix} \mathbf{I}_N & \mathbf{0}_{N,N-1} & \iota_N \\ \mathbf{0}_{N-1,N} & \mathbf{I}_{N-1} & -\iota_{N-1} \end{pmatrix}$$

From the structure of $\mathbf{R}$, it is immediate that vectors $\mathbf{V}$ in the null-space of $(\mathbf{WJ}, \mathbf{WG})$ are of the form $\mathbf{V}_1 = \kappa\iota_N$, $\mathbf{V}_2 = -\mathbf{V}_1$ for $\kappa \in \mathbb{R}$. Applying Proposition 1 yields identification of ρ if there exists $(i, j) \in [N]^2$ such that $\alpha_i \neq \alpha_j$. If this condition is violated then $\alpha = a\iota_N$ for some $a \in \mathbb{R}$, and $(\mathbf{J} + \rho\mathbf{G})\alpha = (1 + \rho)a\iota_{NT}$, so only $(1 + \rho)\alpha$ is identifiable. The reason for this is that there is no variation in the individual effects, implying no variation in any weighted average of peer individual effects. Identification of α, γ requires that $\text{rank}(\mathbf{J} + \rho\mathbf{G}, \mathbf{D}) = N + M - 1$. This is violated only when $\rho = -1$, in which case it has rank $N - 1$. Hence α, γ is identified when also $\rho \neq -1$. If $\rho = -1$ then for any $a \in \mathbb{R}$ one has $(\mathbf{J} + \rho\mathbf{G})(\alpha + a\iota_N) = (\mathbf{J} + \rho\mathbf{G})\alpha$, implying that only $\alpha_i - \alpha_j$ is identified for $(i, j) \in [N]^2$.

To further demonstrate Proposition 1, consider the following simple examples with $N = M = T = 2$ using the linear-in-means network. The linear-in-means network does not exhibit heterogeneous intra-group interactions nor between group-interactions, hence identification depends only on mobility of individuals between groups over time.

3.1. Example 1: An Identifying Mobility Pattern

Suppose that individuals one and two are respectively in groups A and B in the first year. In the second year, individual one remains in group A and individual two moves from group B to group A. Under the linear-in-means network, this mobility pattern yields

$$(\mathbf{J}, \mathbf{G}, \mathbf{D}) = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1/2 & 1/2 & 1 \\ 0 & 1 & 1/2 & 1/2 & 1 \end{pmatrix}, (\mathbf{WJ}, \mathbf{WG}) = \begin{pmatrix} 1/3 & -1/3 & 1/3 & -1/3 \\ 0 & 1 & 0 & 1 \\ 1/3 & -1/3 & -1/6 & 1/6 \\ -2/3 & 2/3 & -1/6 & 1/6 \end{pmatrix}$$

which respectively have rank 4 and 3. Since $\text{rank}(\mathbf{J}, \mathbf{G}, \mathbf{D}) = 2N + M - 2$, by Corollary 1.3, α, ρ and γ are identified if $\alpha_1 \neq \alpha_2$ and $\rho \neq -1$. Since $\text{rank}(\mathbf{WJ}, \mathbf{WG}) = N + 1$, Corollary 1.1 states that ρ is generically identified. This is because the subset of $\mathbb{R}^2$ such that $\alpha_1 = \alpha_2$ has measure zero. For the intuition, consider the underlying system of equations

$$\begin{aligned} \mathbb{E}[y_{11}|\mathbf{G}, \mathbf{D}, \alpha, \gamma] &= \alpha_1 + \rho\alpha_1 + \gamma_1 & \mathbb{E}[y_{21}|\mathbf{G}, \mathbf{D}, \alpha, \gamma] &= \alpha_2 + \rho\alpha_2 \\ \mathbb{E}[y_{12}|\mathbf{G}, \mathbf{D}, \alpha, \gamma] &= \alpha_1 + \rho(\alpha_1 + \alpha_2)/2 + \gamma_1 & \mathbb{E}[y_{22}|\mathbf{G}, \mathbf{D}, \alpha, \gamma] &= \alpha_2 + \rho(\alpha_1 + \alpha_2)/2 + \gamma_1 \end{aligned}$$

When individual two moves groups, individual one obtains a new peer, hence the contextual effect on individual one changes from $\rho\alpha_1$ in the first year to $\rho(\alpha_1 + \alpha_2)/2$ in the second year, whilst the individual and correlated effect are unchanged. The change in the contextual effect is given by the expected change in the outcome of individual one between the first and second year

$$\mathbb{E}[y_{11}|\mathbf{G}, \mathbf{D}, \alpha, \gamma] - \mathbb{E}[y_{12}|\mathbf{G}, \mathbf{D}, \alpha, \gamma] = \rho(\alpha_1 - \alpha_2)/2$$

To identify ρ we need to identify $\alpha_1 - \alpha_2$. We can use year two, in which both individuals are in the same group, hence have the same correlated and contextual effects. This means that any difference in their expected outcomes is due to differences in their individual effects, so

$$\mathbb{E}[y_{12}|\mathbf{G}, \mathbf{D}, \alpha, \gamma] - \mathbb{E}[y_{22}|\mathbf{G}, \mathbf{D}, \alpha, \gamma] = \alpha_1 - \alpha_2$$

If $\alpha_1 \neq \alpha_2$ we can divide the first equation above by the second to identify ρ. If $\rho \neq -1$ the remaining parameters are identified by

$$\begin{aligned} \gamma_1 &= (\mathbb{E}[y_{22}|\mathbf{G}, \mathbf{D}, \alpha, \gamma] - \mathbb{E}[y_{21}|\mathbf{G}, \mathbf{D}, \alpha, \gamma]) - (\mathbb{E}[y_{11}|\mathbf{G}, \mathbf{D}, \alpha, \gamma] - \mathbb{E}[y_{12}|\mathbf{G}, \mathbf{D}, \alpha, \gamma]) \\ \alpha_1 &= (1 + \rho)^{-1}(\mathbb{E}[y_{11}|\mathbf{G}, \mathbf{D}, \alpha, \gamma] - \gamma_1), \quad \alpha_2 = (1 + \rho)^{-1}\mathbb{E}[y_{21}|\mathbf{G}, \mathbf{D}, \alpha, \gamma] \end{aligned}$$

3.2. Example 2: A Non-identifying Mobility Pattern

Now modify Example 1 such that both individuals move in the second year. This means that the individuals are in different groups in the first year and swap groups in the second year, implying that $\mathbf{G} = \mathbf{J}$. The null-space of $(\mathbf{WJ}, \mathbf{WG})$ comprises any vector $\mathbf{V} = (\mathbf{V}_1', \mathbf{V}_2')' = (\mathbf{U}', -\mathbf{U}')'$ for $\mathbf{U} \in \mathbb{R}^2$. For any value of $\alpha \in \mathbb{R}^2$, there clearly exists $\mathbf{U} \in \mathbb{R}^2$ and scalars δ_1 and $\delta_2 \neq 0$ such that $(\delta_1 - \delta_2)\mathbf{U} = \alpha$, so by Proposition 1, ρ is not identified. Note also that the generic identification condition in Corollary 1.1 is violated since $(\mathbf{WJ}, \mathbf{WG})$ has rank $2 < N + 1 = 3$, and the identification condition in Corollary 1.3 is violated since $(\mathbf{J}, \mathbf{G}, \mathbf{D})$ has rank

$3 < 2N + M - 2 = 4$. For the intuition, consider the underlying system of equations

$$\mathbb{E}[y_{11}|\mathbf{G}, \mathbf{D}, \alpha, \gamma] = \alpha_1 + \rho\alpha_1 + \gamma_1 \quad \mathbb{E}[y_{21}|\mathbf{G}, \mathbf{D}, \alpha, \gamma] = \alpha_2 + \rho\alpha_2$$
$$\mathbb{E}[y_{12}|\mathbf{G}, \mathbf{D}, \alpha, \gamma] = \alpha_1 + \rho\alpha_1 \quad \mathbb{E}[y_{22}|\mathbf{G}, \mathbf{D}, \alpha, \gamma] = \alpha_2 + \rho\alpha_2 + \gamma_1$$

The correlated effect γ_1 is identified by individual one moving groups ($\gamma_1 = \mathbb{E}[y_{11}|\mathbf{G}, \mathbf{D}, \alpha, \gamma] - \mathbb{E}[y_{12}|\mathbf{G}, \mathbf{D}, \alpha, \gamma]$) but the remaining equations are only sufficient to identify $(1 + \rho)\alpha$. The reason for this is that there is no variation in the peers of either individual because the individuals are never in the same group in the same year. Mobility of individuals between groups over time is insufficient for identification in this example because it does not induce changes in peers. This contrasts with Abowd et al. (1999), for which group-swapping would be sufficient to identify $\alpha_1, \alpha_2, \gamma_1$ because $\rho = 0$ is imposed.

3.3. Estimation

The conditional mean restriction (3.1) suggests the non-linear least squares estimator obtained by solving

$$\min_{\alpha,\rho,\gamma} (\mathbf{y} - (\mathbf{J} + \rho\mathbf{G})\alpha - \mathbf{D}\gamma)'(\mathbf{y} - (\mathbf{J} + \rho\mathbf{G})\alpha - \mathbf{D}\gamma). \tag{3.2}$$

Arcidiacono et al. (2012) study (3.2) under the *linear-in-others'-means* network

$$\widetilde{\mathbf{G}}_t = \Big(\mathbf{1}(g(i,t) = g(j,t), i \neq j)(N_{g(i,t)} - 1)^{-1}\Big)_{(i,j)\in[N]^2}$$

and with a slightly different specification of the correlated effect, which is allowed to vary over time but is restricted to be the same across multiple groups.[3] Under additional assumptions, including $\mathbb{E}[\epsilon_{it}\epsilon_{js}] = 0$ for all $i \neq j, t \neq s$, $\mathbb{E}[\epsilon_{it}^2|g(i,t)] = \mathbb{E}[\epsilon_{jt}^2|g(j,t)]$ for all $i, j, t : g(i,t) = g(j,t)$, $\mathbb{E}[\epsilon_{it}\alpha_j] = 0$ for all i, j, t and $\rho < \min_{i,t} N_{g(i,t)}$, Arcidiacono et al. (2012) show that the non-linear least squares estimator of ρ is consistent and asymptotically normal for fixed $T \geq 2$. The authors also provide a computationally tractable algorithm which is feasible for large N. The algorithm iterates between estimation of α, ρ and γ, and exploits the fact that the model is linear in its remaining parameters when either α or ρ is fixed. With suitable modifications, their estimator could be applied for other network structures and specifications of the correlated effect. We do not pursue this theoretically because, as pointed out by Arcidiacono et al. (2012), it is unclear how to obtain a fixed-T consistent estimator in models with endogenous effects in which outcomes

[3] We consider a time-varying correlated effect in Section 5.

are simulataneously determined, which is relevant for our application.[4] We now consider such a model of endogenous effects.

4. Endogenous effects

With endogenous effects, the outcome equation is

$$y_{it} = \alpha_i + \psi \sum_{j=1}^{N} \mathbf{G}_{ijt} y_{jt} + \rho \sum_{j=1}^{N} \mathbf{G}_{ijt} \alpha_j + \gamma_{g(i,t)} + \epsilon_{it},$$

or in stacked form

$$\mathbf{y} = \psi \mathbf{F} \mathbf{y} + (\mathbf{J} + \rho \mathbf{G})\alpha + \mathbf{D}\gamma + \epsilon,$$

where ψ is a scalar parameter and $\mathbf{F}$ is a $NT \times NT$ block diagonal matrix with blocks $\mathbf{G}_1, \mathbf{G}_2, ..., \mathbf{G}_T$. Throughout this section we suppose that the rows of $\mathbf{G}$ sum to one ($\mathbf{G}\iota_N = \iota_{NT}$), that there are no inter-group interactions ($\mathbf{G}_{ijt} = 0$ if $g(i,t) \neq g(j,t)$) and $|\psi| < 1$. These are standard assumptions maintained in almost all papers concerning identification of peer effects, and are made to ensure that the reduced form exists and the reduced form correlated effect is proportional to the structural correlated effect. The reduced form is

$$\mathbf{y} = (\mathbf{I}_{NT} - \psi \mathbf{F})^{-1}(\mathbf{J} + \rho \mathbf{G})\alpha + \mathbf{D}(1-\psi)^{-1}\gamma + (\mathbf{I}_{NT} - \psi \mathbf{F})^{-1}\epsilon.$$

with conditional mean $\mathbb{E}[\mathbf{y}|\mathbf{G}, \mathbf{D}, \alpha, \gamma] = (\mathbf{I}_{NT} - \psi \mathbf{F})^{-1}(\mathbf{J} + \rho \mathbf{G})\alpha + \mathbf{D}(1-\psi)^{-1}\gamma$. We now modify Proposition 1 to allow for endogenous effects, making use of $\mathbf{H} = ((\mathbf{G}_1^2)', (\mathbf{G}_2^2)', ..., (\mathbf{G}_T^2)')'$ vectors $\mathbf{V} = (\mathbf{V}_1', \mathbf{V}_2', \mathbf{V}_3')$ in the null-space of $(\mathbf{WJ}, \mathbf{WG}, \mathbf{WH})$ such that $\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3$ are all $N \times 1$.

Proposition 2
ψ and ρ are identified if $\psi + \rho \neq 0$ and there does not exist a vector $\mathbf{V} = (\mathbf{V}_1', \mathbf{V}_2', \mathbf{V}_3')$ in the null-space of $(\mathbf{WJ}, \mathbf{WG}, \mathbf{WH})$ and scalars $\delta_1, \delta_2, \delta_3, \delta_4$ satisfying $\delta_1 \mathbf{V}_1 - \mathbf{V}_2 = (\delta_1 - \delta_2)\alpha$,

[4] Consistency hinges on the reduced form error being uncorrelated across individuals, which is violated if outcomes are simultaneously determined. Arcidiacono et al. (2012) consider an alternative specification of endogenous effects more suited to their application to peer effects in education, in which outcomes depend on the expected (as opposed to realised) outcome of others conditional on observable characteristics and α, but not on ϵ. In this model, y_{it} depends on ϵ_{it} but not on ϵ_{jt} for $j \neq i$. This means that the reduced form errors are uncorrelated across individuals (assuming that the structural errors are uncorrelated), so that the reduced form can be consitently estimated for fixed T by non-linear least squares. Given an instrumental variable z_{it} which determines y_{it} but not y_{jt} for $j \neq i$, Arcidiacono et al. (2012) give an example in which both ρ and the endogenous effect can be recovered from the reduced form.

$\delta_3\mathbf{V}_1 + \mathbf{V}_3 = (\delta_3 - \delta_4)\alpha$ and either $\delta_1 \neq \delta_2$ or $\delta_3 \neq \delta_4$. Otherwise, ψ and ρ are not identified. α, γ are identified if additionally $(\mathbf{J} + \rho\mathbf{G}, \mathbf{D})$ has full column rank.

The identification conditions are similar in form to those of Proposition 1. The additional requirement that $\psi + \rho \neq 0$ is needed because $\psi + \rho = 0$ implies $\mathbb{E}[\mathbf{y}|\mathbf{G}, \mathbf{D}, \alpha, \gamma] = \mathbf{J}\alpha + \mathbf{D}(1-\psi)^{-1}\gamma$, such that ψ and ρ are not separately identifiable. This is because the endogenous effect exactly offsets the contextual effect, yielding no net peer effect. Similar conditions are used throughout the literature, and can be found in Bramoullé et al. (2009) and Rose (2017), among others. With endogenous effects, generic identification of ρ and ψ is attained whenever $\mathbf{V}$ lies in a subspace of $\mathbb{R}^N$ of dimension strictly less than N. A sufficient condition which can be verified in practice is that $(\mathbf{WJ}, \mathbf{WG}, \mathbf{WH})$ has rank at least equal to $2N + 1$.

Under the linear-in-means network ρ and ψ are not separately identifiable without additional restrictions. This is because $\mathbf{H} = \mathbf{G}$. In this case the conditional mean is

$$\mathbb{E}[\mathbf{y}|\mathbf{G}, \mathbf{D}, \alpha, \gamma] = \left(\mathbf{J} + \frac{\psi + \rho}{1-\psi}\mathbf{G}\right)\alpha + \mathbf{D}\frac{\gamma}{1-\psi}$$

so that only $(\psi + \rho)(1-\psi)^{-1}$ and $\gamma(1-\psi)^{-1}$ are identifiable.[5] To solve the identification problem, one could use an instrumental variables approach, supposing that there exists a time-varying instrument z_{it} which is uncorrelated with ϵ_{it} and directly determines y_{it} but not y_{jt} for all $(i, j) \in [N]^2, j \neq i$ (Arcidiacono et al., 2012). Even if such an instrumental variable is available, there remains the fixed-T estimation for models with simultaneity discussed in Section 3.3. We now consider an alternative based on the conditional variance

$$\begin{aligned}\mathbb{V}[\mathbf{Wy}|\mathbf{G}, \mathbf{D}] =& \mathbf{W}(\mathbf{I}_{NT} - \psi\mathbf{F})^{-1}(\mathbf{J} + \rho\mathbf{G})\mathbb{V}[\alpha|\mathbf{G}, \mathbf{D}](\mathbf{J} + \rho\mathbf{G})'(\mathbf{I}_{NT} - \psi\mathbf{F}')^{-1}\mathbf{W} \\ &+ \mathbf{W}(\mathbf{I}_{NT} - \psi\mathbf{F})^{-1}\mathbb{V}[\epsilon|\mathbf{G}, \mathbf{D}](\mathbf{I}_{NT} - \psi\mathbf{F}')^{-1}\mathbf{W}\end{aligned}$$

and the restrictions

$$\begin{aligned}\mathbb{COV}[\epsilon_{it}, \epsilon_{js}|\mathbf{G}, \mathbf{D}] &= \begin{cases}\sigma^2(g(i,t)) & i = j, t = s \\ 0 & \text{otherwise}\end{cases} \\ \mathbb{COV}[\alpha_i, \alpha_j|\mathbf{G}, \mathbf{D}] &= \begin{cases}\sigma_\alpha^2 & i = j \\ 0 & \text{otherwise}\end{cases}.\end{aligned} \tag{4.1}$$

[5]The conditional mean takes the same form as (3.1), hence the results of Section 3 can be applied to establish identification of $(\psi + \rho)(1-\psi)^{-1}$, α and $\gamma(1-\psi)^{-1}$. In Example 1, α, $(\psi + \rho)(1-\psi)^{-1}$ and $\gamma(1-\psi)^{-1}$ are identified if $\rho \neq -1$ and $\alpha_1 \neq \alpha_2$. In Example 2, one is only able to identify $(1+\rho)(1-\psi)^{-1}\alpha$ and $\gamma(1-\psi)^{-1}$.

Our approach is similar in spirit to Graham (2008) and Rose (2017), which consider cross-section data. The restriction on the variance of ϵ requires that the transitory shocks experienced by individuals in the same group be uncorrelated and have equal variance. Uncorrelatedness implies that outcomes are correlated only due to the correlated effect and peer effects, whereas equality of variance implies that within a group, no individuals are subject to systematically larger shocks than others. Importantly, $\mathbb{COV}[\alpha_i, \gamma_{g(j,t)}|\mathbf{G}, \mathbf{D}]$ is unrestricted for all i, j and t. This allows for sorting of individuals to groups based on their time-invariant heterogeneity. The assumption on the variance of α is used so as to avoid estimating α. The conditional variance restriction facilitates identification, does not require exogenous instrumental variables, has computational advantages relative to the non-linear least squares estimator in (3.2),[6] and permits fixed-T estimation with endogenous effects. For brevity, we present only the identification result for the linear-in-means case. In the Appendix we derive an identification result which applies to general network structures and under weaker conditional variance restrictions. In general, we require only that $\mathbb{COV}[\epsilon_{it}, \epsilon_{js}|\mathbf{G}, \mathbf{D}] = 0$ if $g(i,t) \neq g(j,t)$. The stronger restriction in (4.1) is necessary for the linear-in-means network, which is well known to be the most challenging case (Manski, 1993; Bramoullé et al., 2009).

Proposition 3
If $\mathbf{G} = \overline{\mathbf{G}}$, $\psi, \rho, \sigma_\alpha^2$ *are identified if* $\mathbf{W}, \mathbf{WJJ'W}, \mathbf{W}(\mathbf{JG'} + \mathbf{GJ'})\mathbf{W}, \mathbf{WGG'W}$ *and* $\mathbf{WFW}$ *are linearly independent.*

In contrast to Proposition 2 there is no requirement that $\psi + \rho \neq 0$. This is because the endogenous effect operates through the outcomes, and hence propagates variation both in α and ϵ, whilst the contextual effect operates through α alone. Since they operate through different channels, the peer effects can be separated under restrictions on the within-group variance of ϵ such as those in (4.1). The identification condition in Proposition 3 fails if $T = 1$. In this case $\mathbf{F} = \mathbf{G} = \mathbf{G}' = \mathbf{GG}'$ and $\mathbf{J} = \mathbf{I}_N$ so that $\mathbf{W}(\mathbf{JG}' + \mathbf{GJ}')\mathbf{W} = 2\mathbf{WGG'W} = 2\mathbf{WFW}$. This is in agreement with Rose (2017), which shows that the conditional variance restrictions in (4.1) are insufficient for identification with cross-section data.[7]

Due to their simplistic nature, the identification condition in Proposition 3 fails in both Example 1 and Example 2. However, it typically holds in more realistic examples. For example, identification is restored in Example 1 if a third individual is added and individual three is in group B in both years.

[6] In contrast to (3.2), there is no need to minimize over the N-dimensional parameter α.

[7] Graham (2008) shows that contextual effects can be identified with cross-section data under additional restrictions, including on $\mathbb{COV}[\alpha_i, \gamma_{g(j,t)}|\mathbf{G}, \mathbf{D}]$.

4.1. Estimation

For estimation, we modify the approach of Rose (2017). For simplicity of exposition we discuss estimation for the linear-in-means network and suppose that $\sigma^2(m) = \sigma^2$ for all $m \in [M]$.[8] In this case we have

$$\mathbb{V}[\mathbf{Wy}|\mathbf{G},\mathbf{D}] = \sigma^2\mathbf{W} + \frac{\sigma^2\psi(2-\psi)}{1-\psi}\mathbf{WFW} + \sigma_\alpha^2\mathbf{WJJ}'\mathbf{W} + \frac{\sigma_\alpha^2(\psi+\rho)}{1-\psi}\mathbf{W}(\mathbf{JG}'+\mathbf{GJ}')\mathbf{W} + \frac{\sigma_\alpha^2(\psi+\rho)^2}{(1-\psi)^2}\mathbf{WGG}'\mathbf{W}. \tag{4.2}$$

If the identification condition in Proposition 3 is satisfied, the five reduced form parameters in (4.2) can be consistently estimated for fixed $T \geq 2$ by regressing the lower-triangular elements of $\mathbf{WyWy}'$ on the lower triangular elements of the five matrices on the right-hand side. One can then solve to obtain consistent estimators of $\psi, \rho, \sigma_\alpha^2$ and σ^2. An alternative which we find works better in practice is to use non-linear least squares to estimate the structural parameters directly.[9]

5. Time-varying Correlated Effects

If correlated effects are time-varying, then $\gamma_{g(i,t)}$ is replaced by $\gamma_{g(i,t)t}$, in which case $\mathbf{C}$ is a $NT \times MT$ block diagonal matrix with blocks $\mathbf{C}_1, \mathbf{C}_2, ..., \mathbf{C}_T$, $\mathbf{D}$ comprises the first $MT-1$ columns of $\mathbf{C}$ and γ is $(MT-1)\times 1$. All Propositions then apply as stated. The parameters are not identifiable under the linear-in-means network because the peer effect varies only at the group-year level and cannot be separated from the correlated effect. In the absence of further restrictions, identification can be attained only when there are heterogenous intra-group interactions or there exist inter-group interactions.

6. Monte-Carlo Experiment

We study the performance of the non-linear least squares estimators based on conditional mean and conditional variance restrictions. We consider the cases with contextual effects only and with both contextual and endogenous effects. Arcidiacono et al. (2012) prove that the conditional mean estimator is consistent

[8] If $\sigma^2(m)$ is unrestricted the first two terms on the right-hand side of (4.2) become group-specific.. See the proof of Proposition 3 for the derivation of (4.2).

[9] It is possible that non-linear least squares obtains a local minimum. We did not encounter this problem in our Monte-Carlo experiment nor in our empirical application. In practice we use least-squares estimation of (4.2) to obtain starting values for non-linear least squares.

and asymptotically normal for ρ when T is fixed, there are no correlated effects,[10] it is known that $\psi = 0$ and the network is linear-in-others'-means. Our experiment also studies its properties with correlated effects, for networks other than linear-in-others'-means, and when $\psi \neq 0$. The design is tailored to our empirical application, with $N = 500$, $M = 100$, $T = 8$ and annual mobility rate $p = 0.05$ (see summary statistics in Table 3). We use the parameter values $\rho = 0.3$, $\psi = 0.2$ and $\sigma_\alpha^2/\sigma^2 = 3/2$, which are taken from our estimates for the linear-in-others'-means network with correlated effects (see the final column of Table 4). We also consider designs with $T = 2$, $\psi = 0$ and $p = 0.1$.

The data generating process is as follows. In the first year, all individuals are randomly assigned to groups of size five. In each subsequent year, each individual moves group with probability p, in which case she draws a new group with uniform probability over all groups. The expected group size is five for all groups in all years. The networks we consider are the linear-in-means network, the linear-in-others'-means network and a social network in which links are formed as follows. In the first year, each individual draws two links uniformly over other individuals in the same group. In each subsequent year, links persist whilst individuals remain in the same group. If an individual loses a link(s) due to mobility, a replacement link(s) is drawn uniformly over the other individuals in the group with whom there is not already a link. If there are three or fewer individuals in the group, each individual is linked to all other individuals. Links need not be reciprocal.[11] If there exists a link between i and j in year t then $\mathbf{G}_{ijt}$ is equal to the inverse of the number of links that i has in year t. Otherwise $\mathbf{G}_{ijt} = 0$. We take $\alpha_i \sim \mathcal{N}(0, 1)$ and $\epsilon_{it} \sim \mathcal{N}(0, 3/2)$ to be i.i.d., and consider cases in which $\gamma = \mathbf{0}$ is imposed and in which γ_m is the mean of $(\alpha_j)_{t \in [T], g(j,t) = m}$ for $m \in [M - 1]$. We simulate 500 datasets for each experiment. Every dataset verifies the relevant identification condition.

Table 1 considers a design with contextual effects only. The conditional mean and conditional variance estimators perform well in estimating the contextual effect, though the conditional mean estimator is more efficient in all cases other than for the social network with correlated effects. As is the case in the empirical application, the endogenous peer effect is more precisely estimated than the contextual effect. Including correlated effects decreases the precision of both. For the conditional variance estimator, the reduced form effect is estimated with higher precision than the contextual effect. This is because the estimators of ψ and ρ are negatively correlated with one another. Comparing network structures we observe that the parameters are best estimated for the social network and worst for the linear-in-means

[10] The authors argue that time-varying correlated effects could be incorporated provided that the effect is common across multiple peer groups.

[11] If an individual is in a group of size 1 they have no link. As in the empirical application, observations with no links are omitted from the estimation sample.

Table 1: *Monte-Carlo Results* ($\psi = 0, \rho = 0.3$)

Linear-in-means	Conditional Mean		Conditional Variance	
Endogenous peer effect ($\psi = 0$)			-0.0011	-0.0049
			(0.0357)	(0.0432)
Contextual peer effect ($\rho = 0.3$)	0.2982	0.3040	0.3048	0.3321
	(0.1230)	(0.1686)	(0.1439)	(0.3353)
Reduced form peer effect (= 0.3)			0.3017	0.3184
			(0.1187)	(0.2807)
Linear-in-others'-means				
Endogenous peer effect ($\psi = 0$)			-0.0008	-0.0024
			(0.0288)	(0.0295)
Contextual peer effect ($\rho = 0.3$)	0.3004	0.3024	0.3018	0.3033
	(0.0943)	(0.1265)	(0.1064)	(0.2063)
Reduced form peer effect (= 0.3)			0.3003	0.2991
			(0.0896)	(0.1902)
Social network				
Endogenous peer effect ($\psi = 0$)			-0.0011	0.0007
			(0.0387)	(0.0340)
Contextual peer effect ($\rho = 0.3$)	0.3008	0.2365	0.3020	0.3076
	(0.0662)	(0.1841)	(0.0712)	(0.0898)
Correlated Effects	No	Yes	No	Yes
Individuals (N)	500	500	500	500
Groups (M)	100	100	100	100
Years (T)	8	8	8	8
Mobility Rate (p)	0.05	0.05	0.05	0.05

Notes: For each parameter and design we report the mean and standard deviation (in parenthesis) over 500 datasets. If there are no correlated effects this information is treated as known to the researcher. For the conditional mean estimator we impose $\gamma = \mathbf{0}$. For the conditional variance estimator we use $\mathbf{W} = \mathbf{I}_{NT}$. The reduced form peer-effect for the linear-in-means network is $(\psi + \rho)(1 - \psi)^{-1} = 0.3$. The counterpart for the linear-in-others'-means network is $(\psi + \rho)(1 - \psi(\psi + N_{g(i,t)} - 2)/(N_{g(i,t)} - 1))^{-1}$ for individual i in year t in group $g(i,t)$. We report simulation results for $N_{g(i,t)} = 5$, which is the mean group size, with true value equal to 0.3. We do not report a reduced form effect for the Social Network because it is individual-specific, depending on the network structure in their group. Correlated effects are time-invariant.

network. This is because the social and linear-in-others'-means networks exhibit heterogeneous intra-group interactions, whereas linear-in-means relies entirely on mobility between groups for identification. Though Arcidiacono et al. (2012) only establish fixed-T consistency and asymptotic normality of the conditional mean

Table 2: *Monte-Carlo Results ($\psi = 0.2, \rho = 0.3$)*

Linear-in-means	Conditional Mean		Conditional Variance	
Endogenous peer effect ($\psi = 0.2$)			0.1993	0.1944
			(0.0297)	(0.0411)
Contextual peer effect ($\rho = 0.3$)	0.7559	0.8937	0.3037	0.3401
	(0.1819)	(0.2430)	(0.1454)	(0.3853)
Reduced form peer effect (= 0.6250)			0.6260	0.6501
			(0.1486)	(0.3814)
Linear-in-others'-means				
Endogenous peer effect ($\psi = 0.2$)			0.1993	0.1971
			(0.0260)	(0.0283)
Contextual peer effect ($\rho = 0.3$)	0.6748	0.7637	0.3014	0.3064
	(0.1107)	(0.1335)	(0.1101)	(0.2384)
Reduced form peer effect (= 0.5952)			0.5949	0.5946
			(0.1079)	(0.2499)
Social network				
Endogenous peer effect ($\psi = 0.2$)			0.1980	0.2034
			(0.0391)	(0.0419)
Contextual peer effect ($\rho = 0.3$)	0.5564	0.5340	0.3021	0.3073
	(0.0630)	(0.1857)	(0.0789)	(0.0978)
Correlated Effects	No	Yes	No	Yes
Individuals (N)	500	500	500	500
Groups (M)	100	100	100	100
Years (T)	8	8	8	8
Mobility Rate (p)	0.05	0.05	0.05	0.05

Notes: For each parameter and design we report the mean and standard deviation (in parenthesis) over 500 datasets. If there are no correlated effects this information is treated as known to the researcher. For the conditional mean estimator we impose $\gamma = \mathbf{0}$. For the conditional variance estimator we use $\mathbf{W} = \mathbf{I}_{NT}$. The reduced form peer-effect for the linear-in-means network is $(\psi + \rho)(1 - \psi)^{-1} = 0.625$. The counterpart for the linear-in-others'-means network is $(\psi + \rho)(1 - \psi(\psi + N_{g(i,t)} - 2)/(N_{g(i,t)} - 1))^{-1}$ for individual i in year t in group $g(i, t)$. This lies between 0.5208 ($N_{g(i,t)} = 2$) and 0.6250 ($N_{g(i,t)} \to \infty$.) We report simulation results for $N_{g(i,t)} = 5$, which is the mean group size, with true value equal to 0.5952. We do not report a reduced form effect for the Social Network because it is individual-specific, depending on the network structure in their group. Correlated effects are time-invariant.

estimator for linear-in-others'-means without correlated effects, it also appears to perform well for the other network structures and with correlated effects. The exception is for the social network with correlated effects.

Table 2 considers a design with contextual and endogenous effects. In this design the best case for the conditional mean estimator is that it accurately recovers the reduced form parameter. For the linear-in-means network, the reduced form parameter is $(\psi + \rho)(1 - \psi)^{-1} = 0.625$. Table 2 suggests upwards bias in estimation of the reduced form parameter. This conclusion holds across all variations of T and p, as well as with and without correlated effects (see Table 6 in the appendix). It is also true for the linear-in-others'-means network, for which the reduced form effect is group specific, ranging between 0.521 for groups of size two and 0.625 as the group size grows large.[12] Arcidiacono et al. (2012) suggest that this is due to the dependence in the reduced form errors induced by a model of endogenous effects with simultaneity. The conditional variance estimator performs similarly to the case of contextual effects only, though the contextual effect is estimated with marginally lower precision. Tables 5 and 6 in the appendix respectively report results with and without endogenous effects for all networks, with and without correlated effects and all combinations of $T \in \{2, 8\}$ and mobility rate $p \in \{0.05, 0.10\}$. The performance of the estimators is better for larger T, larger p, without correlated effects, and for the social network.

7. Application

We study surgeons' take-up of minimally invasive (keyhole) surgery for colorectal cancer in the English National Health Service (NHS). Our data are from Barrenho et al. (2019), comprising a panel of NHS surgeons and their take-up of keyhole surgery from 2001 to 2008. The data are obtained by merging Hospital Episodes Statistics, which provides patient level diagnosis and treatment information for all patients in the English NHS with NHS Workforce Statistics and the General Medical Council register. This provides matched patient-surgeon-hospital-year data, which is collapsed into a surgeon-hospital-year panel. We refer the reader to Barrenho et al. (2019) for further institutional details and a more detailed description of the construction of the data.

A surgeon's take-up in a given year is measured by the fraction of eligible colorectal cancer sugeries which they performed via keyhole surgery in that year.[13] In 2001, only 1% of eligible surgeries were keyhole, increasing to 25% by 2008. We apply our model for surgeon i in hospital $g(i, t)$ in year t, and consider a surgeon's peers to be those other surgeons located in the same hospital in the same year. To

[12] See the notes under Table 2 for the form of the reduced form effect.

[13] Keyhole surgery is suitable for some, but not all, patients. The sample of patients considered is restricted to those for which the surgeon has a choice between keyhole and the alternative open surgery. We apply our methods to take-up in deviations from the year t mean to control for common trends over time.

Table 3: *Summary Statistics for Estimation Sample*

Year		01	02	03	04	05	06	07	08	All
Surgeons observed		324	339	361	369	374	379	383	358	475
Take-up	Mean	0.01	0.02	0.02	0.04	0.07	0.13	0.17	0.25	0.09
	S.D.	0.04	0.04	0.06	0.11	0.14	0.20	0.23	0.26	0.18
	Min	0	0	0	0	0	0	0	0	0
	Max	0.29	0.31	0.35	1	0.74	1	1	1	1
Moved	Mean	-	0.13	0.06	0.03	0.03	0.03	0.05	0.03	0.05
Peers changed	Mean	-	0.59	0.59	0.68	0.72	0.75	0.69	0.53	0.64
Hospitals observed		73	68	67	66	65	64	64	63	81
No. of surgeons	Mean	4.44	4.99	5.39	5.59	5.75	5.92	5.98	5.68	5.45
	S.D.	1.99	2.43	2.63	2.57	2.57	2.57	2.59	2.31	2.50
	Min	2	2	2	2	2	2	2	2	2
	Max	11	12	14	14	13	12	12	12	14

Notes: 'Take-up' is the fraction of colorectal cancer surgeries performed by keyhole. 'Moved' is a binary indicator for a surgeon being located in a different hospital in year t than in year $t-1$. 'Peers changed' is a binary indicator for a surgeons' peers being different in year t than in $t-1$. 'No. of surgeons' is the number of surgeons located in a hospital.

be included in the estimation sample, surgeons must be observed at least twice, hospitals must be in the largest connected component of the graph of mobility of surgeons between hospitals,[14] and hospital-year pairs must have at least two surgeons. The resulting unbalanced panel comprises $2,887$ observations of $N = 475$ surgeons over $T = 8$ years across $M = 81$ hospitals. Table 3 summarises the estimation sample.

Our identification results are applicable if there is exogenous mobility of surgeons between hospitals over time. We expect this to be the case because all hospitals have the required technology for keyhole surgery, the NHS is a public system with similar working conditions and renumeration nationwide,[15] and colorectal cancer forms only a small part of surgeons' workloads. The leading reason for mobility is to relocate closer to the pre-medical school family home (Goldacre et al., 2013). These arguments support the conditional moment restriction in (3.1). We also posit that there is limited correlation in ϵ_{it} for surgeons in the same hospital

[14]This is because even without peer effects, hospital effects are only identifiable for hospitals in the connected set. To find these hospitals, we first construct an undirected graph between hospitals for which there is a (reciprocal) link between two hospitals if a surgeon ever moves from one to the other between 2001 and 2008.

[15]There is a small private sector in England that mainly provides care for planned procedures for which there are long waiting lists. Private sector provision for cancer is limited.

because patients are allocated to surgeons in a quasi-random fashion due to a two week waiting time guarantee.[16] This argument supports the conditional variance restriction in (4.1). Since we do not restrict $\mathbb{COV}[\alpha_i, \gamma_{g(i,t)}]$, we do not rule out high take-up surgeons being systematically located in high take-up hospitals, though we do rule out their mobility between hospitals being driven by contemporaneous take-up shocks.

We now discuss the extent of identifying variation in the data. Each year, around 5% of surgeons move hospitals (see Table 3). Though only a few surgeons move hospital in any given year, our panel is relatively long, comprising 8 years in total. Moreover, each move changes the peer groups of all those in the hospital left and all those in the hospital joined. The median number of surgeons in a hospital-year pair in our estimation sample is 5, hence a surgeon moving from one median sized hospital to another changes the peer groups of 10 surgeons. For this reason, despite only 5% of surgeons moving, 64% of surgeons' peers change in a typical year (see Table 3). These changes in peer groups can generate large fluctuations in average peer take-up because peer groups are small. In sum, though there are few moves, each one can lead to large changes in average peer take-up for around 10 surgeons.

For every specification we estimate below, our estimation sample verifies the relevant identification condition. The rank of $(\mathbf{WJ}, \mathbf{WG})$ is 720 for the linear-in-means network and 932 for the linear-in-others'-means network. Since $N = 475 < 720$, by Corollary 1.1, ρ is generically identified by the conditional mean restriction in (3.1) when $\psi = 0$ is imposed. In the presence of endogenous effects the conditional mean restriction cannot identify ψ and ρ under the linear-in-means network. In contrast, by Proposition 3, the conditional variance restriction in (4.1) identifies ψ and ρ because the matrices $\mathbf{W}, \mathbf{WJJ'W}, \mathbf{W}(\mathbf{JG'} + \mathbf{GJ'})\mathbf{W}, \mathbf{WGG'W}$ and $\mathbf{WFW}$ are linearly independent. Under linear-in-others'-means, the matrix $(\mathbf{WJ}, \mathbf{WG}, \mathbf{WH})$ has rank 1268, which exceeds $2N + 1 = 951$, so that ψ and ρ are generically identified by the conditional mean restriction. Moreover, there exists hospital $m \in [M]$ such that all of the matrices in Proposition 4 are linearly independent. This means that the weaker conditional variance restriction in (A.1) identifies ψ and ρ.

Table 4 reports the estimation results. Beginning with the linear-in-means network, we find evidence of positive peer effects in take-up. The conditional mean estimator suggests a reduced form peer-effect of close to one. The conditional variance estimator finds a similar reduced form effect, which is not statistically distinguishable from one at any conventional level. Using the conditional variance estimator we can decompose the reduced form effect into an endogenous and contextual effect. We find that both are positive and statistically distinguishable

[16]Cancer patients are allocated to surgeons based primarily on the availability of a surgeon within two-weeks of urgent referral.

Table 4: *Peer Effects in Take-up of Laparoscopic Resection for Colorectal Cancer*

Linear-in-means	Conditional Mean		Conditional Variance	
Endogenous peer effect (ψ)			0.1613	0.1364
			(0.0673)	(0.0654)
Contextual peer effect (ρ)	1.0009	0.9440	0.6200	0.9673
	(0.1677)	(0.2580)	(0.1716)	(0.3554)
Reduced form peer effect			0.9315	1.2781
			(0.1966)	(0.3551)
Surgeon effect variance (σ_α^2)			0.0069	0.0073
			(0.0009)	(0.0011)
Error variance (σ^2)			0.0141	0.0135
			(0.0016)	(0.0014)
Linear-in-others'-means				
Endogenous peer effect (ψ)			0.1714	0.2174
			(0.0560)	(0.0651)
Contextual peer effect (ρ)	0.7986	0.6454	0.3984	0.2894
	(0.0751)	(0.1126)	(0.1095)	(0.2470)
Reduced form peer effect			0.6594	0.6087
			(0.1268)	(0.2406)
Surgeon effect variance (σ_α^2)			0.0090	0.0096
			(0.0013)	(0.0015)
Error variance (σ^2)			0.0148	0.0139
			(0.0016)	(0.0014)
Correlated Effects	No	Yes	No	Yes
Surgeons (N)	475	475	475	475
Hospitals (M)	81	81	81	81
Years (T)	8	8	8	8
Observations	2,887	2,887	2,887	2,887

Notes: Standard errors are reported in parentheses and are clustered by hospital. The reduced form peer-effect for the linear-in-means network is $(\psi + \rho)(1 - \psi)^{-1}$. The counterpart for the linear-in-others'-means network is $(\psi + \rho)(1 - \psi(\psi + N_{g(i,t)} - 2)/(N_{g(i,t)} - 1))^{-1}$ for surgeon i in year t in hospital $g(i,t)$. We report this for $N_{g(i,t)} = 5$, which is the median number of surgeons in a hospital-year in our estimation sample. Standard errors for the reduced form effect are obtained using the Delta method. Correlated effects are time-invariant.

from zero. Though the contextual effect is larger in magnitude than the endogenous effect, the standard errors are large enough so that we cannot rule out equality. To interpret the effect magnitude, consider the effect on a surgeon's take-up of a standard deviation increase in the average α_j in their hospital, which, for the median hospital with five surgeons is estimated to be approximately $1 \times \sqrt{0.007/5} \approx 0.04$. The the standard deviation of take-up in 2008 is 0.26, hence the effect is a 0.15 of a 2008 standard deviation increase in take-up. Our conditional variance results suggest that the majority of this change is attributable to the contextual effect, which makes up around three quarters of the reduced form effect.

Moving on to the linear-in-others'-means network, we also find positive peer effects, though of a smaller magnitude and more precisely estimated than linear-in-means. This is because a surgeon is not treated as their own peer, which is more natural in settings with small peer groups. Under linear-in-others'-means, the reduced form peer effect depends on the size of the peer group. Table 4 reports the reduced form effect for the median hospital with five surgeons. Our estimation results with correlated effects suggest that the reduced form peer effect is around 0.6. For a surgeon in the median sized hospital, a standard deviation change in the average α_i of the other four surgeons is estimated to lead to $0.6 \times \sqrt{0.009/4} \approx 0.03$ increase in take-up, which is smaller than for the linear-in-means network, equating to a 0.12 standard deviation increase. The linear-in-others'-means results point to a more equal role for the endogenous and contextual effects. With correlated effects, half of the reduced form effect is due to the endogenous effect. Though all our estimates are qualitatively similar, we consider the linear-in-others'-means model with correlated effects to provide the most credible results, which we use to design our Monte-Carlo experiment.

Barrenho et al. (2019) also find positive endogenous peer effects, with a baseline estimates ranging between 0.49 and 0.62 (see Table 6 in Barrenho et al. (2019)). The most comparable specification is that based on the linear-in-others'-means network with correlated effects at the hospital level, for which we find a smaller endogenous effect of 0.22 and a reduced form effect of 0.61. Barrenho et al. (2019) do not allow for contextual effects operating through surgeons' time-invariant heterogeneity, which we find to be positive though not statistically significant. This implies that the authors' estimate of the endogenous effect, which is similar to our reduced form effect, might also be picking up the contextual effect.

8. Conclusion

This paper provides new identification results for panel data models with contextual peer effects, endogenous peer effects and correlated effects. Our results suggest that these channels can typically be separated provided either that there is sufficient mobility in the data or that the network data are sufficiently detailed. In specific cases, such as for the linear-in-means network, additional conditional variance restrictions may be necessary to separate the peer effects. In practice, the researcher can always specific the linear-in-others'-means network, which has better identification prospects.

References

Abowd, J. M., F. Kramarz and D. N. Margolis, "High wage workers and high wage firms," *Econometrica* 67 (1999), 251–333.

Arcidiacono, P., G. Foster, N. Goodpaster and J. Kinsler, "Estimating spillovers using panel data, with an application to the classroom," *Quantitative Economics* 3 (2012), 421–470.

Arnold, M., M. S. Sierra, M. Laversanne, I. Soerjomataram, A. Jemal and F. Bray, "Global patterns and trends in colorectal cancer incidence and mortality," *Gut* 66 (2017), 683–691.

Barrenho, E., M. Miraldo, C. Propper, C. Rose et al., "Peer and Network Effects in Medical Innovation: The Case of Laparoscopic Surgery in the English NHS," Technical Report, HEDG, c/o Department of Economics, University of York. Latest version: https://drive.google.com/file/d/1xB9lbYkAzxrGCk_0CTuYNPN_NWGqv_nt/view?usp=sharing, 2019.

Blume, L. E., W. A. Brock, S. N. Durlauf and R. Jayaraman, "Linear social interactions models," *Journal of Political Economy* 123 (2015), 444–496.

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

———, "Peer effects in networks: A survey," (2019).

Calvó-Armengol, A., E. Patacchini and Y. Zenou, "Peer effects and social networks in education," *The Review of Economic Studies* 76 (2009), 1239–1267.

Cohen-Cole, E., X. Liu and Y. Zenou, "Multivariate choices and identification of social interactions," *Journal of Applied Econometrics* 33 (2018), 165–178.

Davezies, L., X. d'Haultfoeuille and D. Fougère, "Identification of peer effects using group size variation," *The Econometrics Journal* 12 (2009), 397–413.

De Giorgi, G., M. Pellizzari and S. Redaelli, "Identification of social interactions through partially overlapping peer groups," *American Economic Journal: Applied Economics* 2 (2010), 241–75.

De Paula, A., "Econometrics of network models," in *Advances in economics and econometrics: Theory and applications, eleventh world congress* (Cambridge University Press Cambridge, 2017), 268–323.

Goldacre, M., J. Davidson, J. Maisonneuve and T. Lambert, "Geographical movement of doctors from education to training and eventual career post: UK cohort studies," *Journal of the Royal Society of Medicine* 106 (2013), 96–104.

Goldsmith-Pinkham, P. and G. W. Imbens, "Social networks and the identification of peer effects," *Journal of Business & Economic Statistics* 31 (2013), 253–264.

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

Laudicella, M., B. Walsh, E. Burns and P. C. Smith, "Cost of care for cancer patients in England: evidence from population-based patient-level data," *British journal of cancer* 114 (2016), 1286–1292.

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

Lewbel, A., "The identification zoo: Meanings of identification in econometrics," *Journal of Economic Literature* 57 (2019), 835–903.

Manski, C. F., "Identification of endogenous social effects: The reflection problem," *The Review of Economic Studies* 60 (1993), 531–542.

Mas, A. and E. Moretti, "Peers at work," *American Economic Review* 99 (2009), 112–45.

Moffitt, R. A. et al., "Policy interventions, low-level equilibria, and social interactions," *Social dynamics* 4 (2001), 6–17.

Rose, C. D., "Identification of peer effects through social networks using variance restrictions," *The Econometrics Journal* 20 (2017), S47–S60.

Appendix

Conditional Variance Restrictions for a General Network Structure

In the general network case identification can be attained without making restrictions on the within-group variance of ϵ. This is in contrast to the linear-in-means network.[17] One can use

$$\mathbb{COV}[\epsilon_{it}, \epsilon_{js}|\mathbf{G}, \mathbf{D}] = 0 \quad g(i,t) \neq g(j,s),$$
$$\mathbb{COV}[\alpha_i, \alpha_j|\mathbf{G}, \mathbf{D}] = \begin{cases} \sigma_\alpha^2 & i = j \\ 0 & \text{otherwise} \end{cases} \tag{A.1}$$

instead of (4.1), which allows for unrestricted within-group variance of ϵ. We make use of the following definition from Rose (2017), reprinted here for convenience.

Definition *Consider L matrices of the same dimension $\mathbf{A}_1, ..., \mathbf{A}_L$. The matrix $\mathbf{A}_l$ ($l \in \{1, ..., L\}$) is maximally linearly independent of $\mathbf{A}_1, ..., \mathbf{A}_L$ if $\lambda_l = 0$ for all $\lambda \in \mathbb{R}^L$ such that $\sum_{l=1}^{L} \lambda_l \mathbf{A}_l = \mathbf{0}$.*

Note that linear independence of $\mathbf{A}_1, ..., \mathbf{A}_L$ is equivalent to each of the L matrices being maximally linearly independent. In what follows we require only that a strict subset be maximially linearly independent, which is weaker than linear independence. Identification hinges on the covariance terms for outcomes of observations in different groups, which are non-zero provided that there is mobility between groups. To extract these covariance terms, we use $\mathcal{E}_m \subseteq [NT]$ to denote the indices of the observations in group $m \in [M]$ and define $\mathbf{E}_m$ as the matrix constructed from rows $\mathcal{E}_m$ of $\mathbf{I}_{NT}$. Pre-multiplying any conformable matrix by $\mathbf{E}_m$ extracts rows $\mathcal{E}_m$, and post-multiplying by $\mathbf{E}_m'$ extracts columns $\mathcal{E}_m$. We also use $\mathbf{E}_{-m}$ to extract the rows in the complement of $\mathcal{E}_m$, which are the observations corresponding to all groups other than m.

Proposition 4
$\psi, \rho, \sigma_\alpha^2$ are identified if $\psi + \rho \neq 0$ and there exists $m \in [M]$ such that $\mathbf{E}_m\mathbf{W}\mathbf{J}\mathbf{J}'\mathbf{W}\mathbf{E}_{-m}'$, $\mathbf{E}_m\mathbf{W}(\mathbf{J}\mathbf{G}' + \mathbf{G}\mathbf{J}')\mathbf{W}\mathbf{E}_{-m}'$ and $\mathbf{E}_m\mathbf{W}(\mathbf{J}\mathbf{H}' + \mathbf{H}\mathbf{J}')\mathbf{W}\mathbf{E}_{-m}'$ are maximially linearly independent from $\mathbf{E}_m\mathbf{W}\mathbf{J}\mathbf{J}'\mathbf{W}\mathbf{E}_{-m}'$, $\mathbf{E}_m\mathbf{W}(\mathbf{J}\mathbf{G}' + \mathbf{G}\mathbf{J}')\mathbf{W}\mathbf{E}_{-m}'$, $\mathbf{E}_m\mathbf{W}(\mathbf{J}\mathbf{H}' + \mathbf{H}\mathbf{J}')\mathbf{W}\mathbf{E}_{-m}'$, $\mathbf{E}_m\mathbf{W}(\mathbf{G}\mathbf{H}' + \mathbf{H}\mathbf{G}')\mathbf{W}\mathbf{E}_{-m}$, $\mathbf{E}_m\mathbf{W}\mathbf{G}\mathbf{G}'\mathbf{W}\mathbf{E}_{-m}'$, and $\mathbf{E}_m\mathbf{W}\mathbf{H}\mathbf{H}'\mathbf{W}\mathbf{E}_{-m}'$.

Identically to Proposition 2, $\psi + \rho \neq 0$ is required so that the endogenous and contextual effects do not exactly offset one another. This contrasts with Proposition 3, which concerns the linear-in-means network, for which $\psi + \rho \neq 0$ is not required. The reason for this is that Proposition 3 additionally restricts the within-group variance of ϵ because restrictions on the between-group variance alone are insufficient for identification. Since the endogenous effects operate both through ϵ and α, whereas the contextual effects operate only through α, under restrictions on

[17] See the end of the proof of Proposition 3 for an explanation of why restrictions on the within-group variance are necessary for the linear-in-means model.

the within-group variance of ϵ there are no values of ψ and ρ such that the two peer effects exactly offset one another in the reduced form variance.

Proofs

Proof of Proposition 1 Denote $\theta = (\alpha, \rho, \gamma)$. Under the conditional moment restriction we have

$$\mathbb{E}[\mathbf{W}\mathbf{y}|\mathbf{G}, \mathbf{D}, \alpha, \gamma] = \mathbf{W}\mathbf{J}\alpha + \mathbf{W}\mathbf{G}\rho\alpha$$

Suppose that there is $\overline{\theta}$ satisfying the conditional moment restriction in (3.1). Then we have

$$(\mathbf{WJ}, \mathbf{WG}) \begin{pmatrix} \alpha - \overline{\alpha} \\ \rho\alpha - \overline{\rho\alpha} \end{pmatrix} = \mathbf{0}$$

Let $\mathbf{V} = (\mathbf{V}_1', \mathbf{V}_2')$ be a vector in the null-space of $(\mathbf{WJ}, \mathbf{WG})$. Then for some $\mathbf{V}$ we have

$$\mathbf{V}_1 = \alpha - \overline{\alpha}$$
$$\mathbf{V}_2 = \rho\alpha - \overline{\rho\alpha}$$

which implies $\overline{\rho}\mathbf{V}_1 - \mathbf{V}_2 = (\overline{\rho} - \rho)\alpha$. If $\overline{\rho} \neq \rho$ then this is equivalently expressed as

$$\frac{\overline{\rho}}{\overline{\rho} - \rho}\mathbf{V}_1 - \frac{1}{\overline{\rho} - \rho}\mathbf{V}_2 = \alpha$$

If there does not exist $\mathbf{V}$ in the null-space of $(\mathbf{WJ}, \mathbf{WG})$ and scalars δ_1 and $\delta_2 \neq 0$ verifying $\delta_1\mathbf{V}_1 + \delta_2\mathbf{V}_2 = \alpha$ then, by contradiction we have $\overline{\rho} = \rho$. If $\overline{\rho} = \rho$ then we have

$$(\mathbf{J} + \rho\mathbf{G}, \mathbf{D}) \begin{pmatrix} \alpha - \overline{\alpha} \\ \gamma - \overline{\gamma} \end{pmatrix} = \mathbf{0}$$

so $\overline{\alpha} = \alpha, \overline{\gamma} = \gamma$ if $(\mathbf{J} + \rho\mathbf{G}, \mathbf{D})$ has full column rank.

Proof of Proposition 2 Denote $\theta = (\psi, \rho, \alpha, \gamma)$. Under the conditional moment restriction we have

$$\mathbb{E}[\mathbf{y}_t|\mathbf{G}, \mathbf{D}, \alpha, \gamma] = (\mathbf{I}_N - \psi\mathbf{G}_t)^{-1}(\mathbf{I}_N\alpha + \mathbf{G}_t\rho\alpha + \mathbf{D}_t\gamma)$$

for $t \in [T]$. The inverse of $(\mathbf{I}_N - \psi \mathbf{G}_t)$ exists because $\mathbf{G}_t \iota_N = \iota_N$ and $|\psi| < 1$. Since $\mathbf{G}_t \iota_N = \iota_N$ and $\mathbf{G}_{ijt} = 0$ if $g(i,t) \neq g(j,t)$ we have $(\mathbf{I}_N - \psi \mathbf{G}_t)^{-1} \mathbf{D}_t \gamma = \mathbf{D}_t (1-\psi)^{-1} \gamma$, hence

$$\mathbb{E}[\mathbf{y}_t | \mathbf{G}, \mathbf{D}, \alpha, \gamma] = (\mathbf{I}_N - \psi \mathbf{G}_t)^{-1} (\mathbf{I}_N + \rho \mathbf{G}_t) \alpha + \mathbf{D}_t (1-\psi)^{-1} \gamma$$

Suppose that there is $\overline{\theta}$ satisfying the conditional moment restriction. Then we have

$$\begin{aligned}
&(\mathbf{I}_N - \psi \mathbf{G}_t)^{-1} (\mathbf{I}_N + \rho \mathbf{G}_t) \alpha + \mathbf{D}_t (1-\psi)^{-1} \gamma = \\
&(\mathbf{I}_N - \overline{\psi} \mathbf{G}_t)^{-1} (\mathbf{I}_N + \overline{\rho} \mathbf{G}_t) \overline{\alpha} + \mathbf{D}_t (1-\overline{\psi})^{-1} \overline{\gamma}
\end{aligned}$$

Pre-multiplying both sides by $(\mathbf{I}_N - \psi \mathbf{G}_t)(\mathbf{I}_N - \overline{\psi} \mathbf{G}_t)$ and rearranging yields

$$\begin{aligned}
&\mathbf{I}_N (\alpha - \overline{\alpha}) + \mathbf{G}_t \left((\rho - \overline{\psi}) \alpha - (\overline{\rho} - \psi) \overline{\alpha} \right) + \mathbf{G}_t^2 \left(-\overline{\psi} \rho \alpha + \psi \overline{\rho \alpha} \right) \\
&+ \mathbf{D}_t \left((1 - \overline{\psi}) \gamma - (1 - \psi) \overline{\gamma} \right) = \mathbf{0}
\end{aligned}$$

Stacking for $t \in [T]$ yields

$$\mathbf{J}(\alpha - \overline{\alpha}) + \mathbf{G} \left((\rho - \overline{\psi}) \alpha - (\overline{\rho} - \psi) \overline{\alpha} \right) + \mathbf{H} \left(-\overline{\psi} \rho \alpha + \psi \overline{\rho \alpha} \right) + \mathbf{D} \left((1 - \overline{\psi}) \gamma - (1 - \psi) \overline{\gamma} \right) = \mathbf{0}$$

and applying $\mathbf{W}$ on the left yields

$$(\mathbf{WJ}, \mathbf{WG}, \mathbf{WH}) \begin{pmatrix} \alpha - \overline{\alpha} \\ (\rho - \overline{\psi}) \alpha - (\overline{\rho} - \psi) \overline{\alpha} \\ -\overline{\psi} \rho \alpha + \psi \overline{\rho \alpha} \end{pmatrix} = \mathbf{0}$$

Let $\mathbf{V} = (\mathbf{V}_1', \mathbf{V}_2', \mathbf{V}_3')$ be a vector in the null-space of $(\mathbf{WJ}, \mathbf{WG}, \mathbf{WH})$. Then for some $\mathbf{V}$,

$$\begin{aligned}
\mathbf{V}_1 &= \alpha - \overline{\alpha} \\
\mathbf{V}_2 &= (\rho - \overline{\psi}) \alpha - (\overline{\rho} - \psi) \overline{\alpha} \\
\mathbf{V}_3 &= -\overline{\psi} \rho \alpha + \psi \overline{\rho \alpha}
\end{aligned}$$

Eliminating $\overline{\alpha}$ yields

$$\begin{aligned}
&(\overline{\rho} - \psi) \mathbf{V}_1 - \mathbf{V}_2 = (\overline{\rho} - \psi + \overline{\psi} - \rho) \alpha \\
&\psi \overline{\rho} \mathbf{V}_1 + \mathbf{V}_3 = (\psi \overline{\rho} - \overline{\psi} \rho) \alpha
\end{aligned}$$

If there does not exist $\mathbf{V}$ in the null-space of $(\mathbf{WJ}, \mathbf{WG}, \mathbf{WH})$ and scalars $\delta_1, \delta_2, \delta_3, \delta_4$ satisfying $\delta_1 \mathbf{V}_1 - \mathbf{V}_2 = (\delta_1 - \delta_2) \alpha$, $\delta_3 \mathbf{V}_1 + \mathbf{V}_3 = (\delta_3 - \delta_4) \alpha$ and either $\delta_1 \neq \delta_2$ or

$\delta_3 \neq \delta_4$ then we have

$$\overline{\rho} - \psi + \overline{\psi} - \rho = 0 \tag{A.2}$$

$$\psi\overline{\rho} - \overline{\psi}\rho = 0 \tag{A.3}$$

Solving (A.2) for $\overline{\psi}$ and injecting into (A.3) yields $(\overline{\rho} - \rho)(\psi + \rho) = 0$, hence $\overline{\rho} = \rho$ provided that $\psi + \rho \neq 0$. Injecting into (A.2) yields $\overline{\psi} = \psi$. If $\overline{\rho} = \rho$ and $\overline{\psi} = \psi$ then we have

$$(\mathbf{J} + \rho\mathbf{G}, \mathbf{D}) \begin{pmatrix} \alpha - \overline{\alpha} \\ \gamma - \overline{\gamma} \end{pmatrix} = \mathbf{0}$$

so $\overline{\alpha} = \alpha, \overline{\gamma} = \gamma$ if $(\mathbf{J} + \rho\mathbf{G}, \mathbf{D})$ has full column rank.

Proof of Proposition 3 Denote $\theta = (\psi, \rho, \sigma_\alpha^2)'$. We have the reduced form

$$\mathbf{y} = (\mathbf{I}_{NT} - \psi\mathbf{F})^{-1}(\mathbf{J} + \rho\mathbf{G})\alpha + \mathbf{D}(1-\psi)^{-1}\beta + (\mathbf{I}_{NT} - \psi\mathbf{F})^{-1}\epsilon$$

Since $\mathbf{G} = \overline{\mathbf{G}}$, we have $\mathbf{G}_t = \mathbf{G}_t^2$ for $t \in [T]$ and $\mathbf{F} = \mathbf{F}^2$, hence $(\mathbf{I}_{NT} - \psi\mathbf{F})^{-1} = \mathbf{I}_{NT} + \frac{\psi}{1-\psi}\mathbf{F}$ and

$$\mathbf{W}\mathbf{y} = \mathbf{W}\left(\mathbf{J} + \frac{\psi+\rho}{1-\psi}\mathbf{G}\right)\alpha + \mathbf{W}\left(\mathbf{I}_{NT} + \frac{\psi}{1-\psi}\mathbf{F}\right)\epsilon$$

with conditional variance

$$\begin{aligned}
\mathbb{V}[\mathbf{W}\mathbf{y}|\mathbf{G},\mathbf{D}] &= \sigma_\alpha^2\mathbf{W}\left(\mathbf{J} + \frac{\psi+\rho}{1-\psi}\mathbf{G}\right)\left(\mathbf{J} + \frac{\psi+\rho}{1-\psi}\mathbf{G}\right)\mathbf{W} \\
&\quad + \mathbf{W}\left(\mathbf{I}_{NT} + \frac{\psi}{1-\psi}\mathbf{F}\right)\Sigma\left(\mathbf{I}_{NT} + \frac{\psi}{1-\psi}\mathbf{F}\right)\mathbf{W} \\
&= \sigma_\alpha^2\mathbf{W}\mathbf{J}\mathbf{J}'\mathbf{W} + \frac{\sigma_\alpha^2(\psi+\rho)}{1-\psi}\mathbf{W}(\mathbf{J}\mathbf{G}' + \mathbf{G}\mathbf{J}')\mathbf{W} + \frac{\sigma_\alpha^2(\psi+\rho)^2}{(1-\psi)^2}\mathbf{W}\mathbf{G}\mathbf{G}'\mathbf{W} \\
&\quad + \mathbf{W}\Sigma\mathbf{W} + \frac{\psi(2-\psi)}{(1-\psi)^2}\mathbf{W}\mathbf{F}\Sigma\mathbf{F}\mathbf{W}
\end{aligned}$$

where $\Sigma = \mathbb{V}[\epsilon|\mathbf{G},\mathbf{D}]$. Consider group $m \in [M]$ and define $\mathcal{E}_m \subseteq [NT]$ to be the indices of the observations in group m. Then we can define $\mathbf{E}_m$ as the matrix constructed from rows $\mathcal{E}_m$ of $\mathbf{I}_{NT}$. Pre-multiplying any conformable matrix by $\mathbf{E}_m$ extracts rows $\mathcal{E}_m$, and post-multiplying by $\mathbf{E}_m'$ extracts columns $\mathcal{E}_m$. Consider first the within-group conditional variance for group m, given by

$$\sigma_\alpha^2 \mathbf{E}_m \mathbf{W} \mathbf{J} \mathbf{J}' \mathbf{W} \mathbf{E}_m' + \frac{\sigma_\alpha^2(\psi+\rho)}{1-\psi} \mathbf{E}_m \mathbf{W}(\mathbf{J}\mathbf{G}' + \mathbf{G}\mathbf{J}')\mathbf{W}\mathbf{E}_m'$$
$$+ \frac{\sigma_\alpha^2(\psi+\rho)^2}{(1-\psi)^2} \mathbf{E}_m \mathbf{W}\mathbf{G}\mathbf{G}'\mathbf{W}\mathbf{E}_m' + \sigma^2(m)\mathbf{E}_m\mathbf{W}\mathbf{E}_m' + \frac{\sigma^2(m)\psi(2-\psi)}{(1-\psi)^2}\mathbf{E}_m\mathbf{W}\mathbf{F}\mathbf{W}\mathbf{E}_m'$$

and the between group conditional variance, given by

$$\sigma_\alpha^2 \mathbf{E}_m \mathbf{W} \mathbf{J} \mathbf{J}' \mathbf{W} \mathbf{E}_{-m}' + \frac{\sigma_\alpha^2(\psi+\rho)}{1-\psi} \mathbf{E}_m \mathbf{W}(\mathbf{J}\mathbf{G}' + \mathbf{G}\mathbf{J}')\mathbf{W}\mathbf{E}_{-m}' + \frac{\sigma_\alpha^2(\psi+\rho)^2}{(1-\psi)^2} \mathbf{E}_m \mathbf{W}\mathbf{G}\mathbf{G}'\mathbf{W}\mathbf{E}_{-m}'$$

where $\mathbf{E}_{-m}$ extracts the rows $\mathcal{E}_m^c$, which denotes the complement of $\mathcal{E}_m$. Notice first that the between-group variance alone is insufficient to identify ψ and ρ. Only $(\psi+\rho)(1-\psi)^{-1}$ is identifiable based on $\mathbb{COV}[\epsilon_{it}, \epsilon_{jt}|\mathbf{G}, \mathbf{D}] = 0$ for all $(i, j, t) \in [N]^2 \times T : g(i, t) \neq g(j, t)$ alone. However, under the additional restrictions on the within-group variance in (4.1), we obtain two additional terms which allow additionally for identification of $\psi(2-\psi)(1-\psi)^{-2}$. Now suppose that there is $\overline{\theta}$ satisfying the conditional variance restrictions in (4.1). Then, if $\mathbf{W}, \mathbf{W}\mathbf{J}\mathbf{J}'\mathbf{W}, \mathbf{W}(\mathbf{J}\mathbf{G}' + \mathbf{G}\mathbf{J}')\mathbf{W}, \mathbf{W}\mathbf{G}\mathbf{G}'\mathbf{W}$ and $\mathbf{W}\mathbf{F}\mathbf{W}$ are linearly independent then there exists m such that $\overline{\sigma}^2(m) = \sigma^2(m)$, $\overline{\sigma}_\alpha^2 = \sigma_\alpha^2$ and

$$\frac{\psi+\rho}{1-\psi} = \frac{\overline{\psi}+\overline{\rho}}{1-\overline{\psi}}$$
$$\frac{\psi(2-\psi)}{(1-\psi)^2} = \frac{\overline{\psi}(2-\overline{\psi})}{(1-\overline{\psi})^2}$$

These two equations have two solutions given by $\overline{\psi} = \psi, \overline{\rho} = \rho$ and $\overline{\psi} = 2-\psi, \overline{\rho} = (2\psi + 3\rho - \psi\rho - 2)(1+\psi)^{-1}$. Since $|\psi| < 1$ we have $|2-\psi| > 1$, hence the second solution is infeasible. Hence we obtain $\overline{\theta} = \theta$.

Proof of Proposition 4 Denote $\theta = (\psi, \rho, \sigma_\alpha^2)'$. Under the conditional variance restriction (4.1) we have

$$\mathbb{V}[\mathbf{y}|\mathbf{G},\mathbf{D}] = \sigma_\alpha^2(\mathbf{I}_{NT}-\psi\mathbf{F})^{-1}(\mathbf{J}+\rho\mathbf{G})(\mathbf{J}+\rho\mathbf{G})'(\mathbf{I}_{NT}-\psi\mathbf{F}')^{-1} + \frac{1}{(1-\psi)^2}\mathbf{D}\Sigma_\gamma\mathbf{D}'$$
$$+ (\mathbf{I}_{NT}-\psi\mathbf{F})^{-1}\Sigma(\mathbf{I}_{NT}-\psi\mathbf{F}')^{-1} + \frac{1}{1-\psi}(\mathbf{I}_{NT}-\psi\mathbf{F})^{-1}(\mathbf{J}+\rho\mathbf{G})\Sigma_{\alpha\gamma}\mathbf{D}'$$
$$+ \frac{1}{1-\psi}\mathbf{D}\Sigma'_{\alpha\gamma}(\mathbf{J}+\rho\mathbf{G})'(\mathbf{I}_{NT}-\psi\mathbf{F}')^{-1} + \frac{1}{1-\psi}(\mathbf{I}_{NT}-\psi\mathbf{F})^{-1}\Sigma_{\epsilon\gamma}\mathbf{D}'$$
$$+ \frac{1}{1-\psi}\mathbf{D}\Sigma'_{\epsilon\gamma}(\mathbf{I}_{NT}-\psi\mathbf{F}')^{-1}$$

where Σ encodes the conditional variance of ϵ, $\Sigma_{\alpha\gamma}$ encodes the conditional covariance terms for α,γ and similarly for $\Sigma_{\epsilon\gamma}$. Suppose that there is $\overline{\theta}$ satisfying the conditional variance restriction in (A.1). Then we have

$$\sigma_\alpha^2(\mathbf{I}_{NT}-\psi\mathbf{F})^{-1}(\mathbf{J}+\rho\mathbf{G})(\mathbf{J}+\rho\mathbf{G})'(\mathbf{I}_{NT}-\psi\mathbf{F}')^{-1} + \frac{1}{(1-\psi)^2}\mathbf{D}\Sigma_\gamma\mathbf{D}'$$
$$+ (\mathbf{I}_{NT}-\psi\mathbf{F})^{-1}\Sigma(\mathbf{I}_{NT}-\psi\mathbf{F}')^{-1} + \frac{1}{1-\psi}(\mathbf{I}_{NT}-\psi\mathbf{F})^{-1}(\mathbf{J}+\rho\mathbf{G})\Sigma_{\alpha\gamma}\mathbf{D}'$$
$$+ \frac{1}{1-\psi}\mathbf{D}\Sigma'_{\alpha\gamma}(\mathbf{J}+\rho\mathbf{G})'(\mathbf{I}_{NT}-\psi\mathbf{F}')^{-1} + \frac{1}{1-\psi}(\mathbf{I}_{NT}-\psi\mathbf{F})^{-1}\Sigma_{\epsilon\gamma}\mathbf{D}'$$
$$+ \frac{1}{1-\psi}\mathbf{D}\Sigma'_{\epsilon\gamma}(\mathbf{I}_{NT}-\psi\mathbf{F}')^{-1} =$$
$$\overline{\sigma}_\alpha^2(\mathbf{I}_{NT}-\overline{\psi}\mathbf{F})^{-1}(\mathbf{J}+\overline{\rho}\mathbf{G})(\mathbf{J}+\overline{\rho}\mathbf{G})'(\mathbf{I}_{NT}-\overline{\psi}\mathbf{F}')^{-1} + \frac{1}{(1-\overline{\psi})^2}\mathbf{D}\overline{\Sigma}_\gamma\mathbf{D}'$$
$$+ \epsilon(\mathbf{I}_{NT}-\overline{\psi}\mathbf{F})^{-1}\overline{\Sigma}(\mathbf{I}_{NT}-\overline{\psi}\mathbf{F}')^{-1} + \frac{1}{1-\overline{\psi}}(\mathbf{I}_{NT}-\overline{\psi}\mathbf{F})^{-1}(\mathbf{J}+\overline{\rho}\mathbf{G})\overline{\Sigma}_{\alpha\gamma}\mathbf{D}'$$
$$+ \frac{1}{1-\overline{\psi}}\mathbf{D}\overline{\Sigma}'_{\alpha\gamma}(\mathbf{J}+\overline{\rho}\mathbf{G})'(\mathbf{I}_{NT}-\overline{\psi}\mathbf{F}')^{-1} + \frac{1}{1-\overline{\psi}}(\mathbf{I}_{NT}-\overline{\psi}\mathbf{F})^{-1}\overline{\Sigma}_{\epsilon\gamma}\mathbf{D}'$$
$$+ \frac{1}{1-\overline{\psi}}\mathbf{D}\overline{\Sigma}'_{\epsilon\gamma}(\mathbf{I}_{NT}-\overline{\psi}\mathbf{F}')^{-1}$$

Pre-multiplying both sides by $\mathbf{W}(\mathbf{I}_{NT}-\psi\mathbf{F})(\mathbf{I}_{NT}-\overline{\psi}\mathbf{F})$ and post-multiplying by $(\mathbf{I}_{NT}-\psi\mathbf{F}')(\mathbf{I}_{NT}-\overline{\psi}\mathbf{F}')\mathbf{W}$ yields

$$\sigma_\alpha^2\mathbf{W}(\mathbf{I}_{NT}-\overline{\psi}\mathbf{F})(\mathbf{J}+\rho\mathbf{G})(\mathbf{J}+\rho\mathbf{G})'(\mathbf{I}_{NT}-\overline{\psi}\mathbf{F}')\mathbf{W} + \mathbf{W}(\mathbf{I}_{NT}-\overline{\psi}\mathbf{F})\Sigma(\mathbf{I}_{NT}-\overline{\psi}\mathbf{F}')\mathbf{W} =$$
$$\overline{\sigma}_\alpha^2\mathbf{W}(\mathbf{I}_{NT}-\psi\mathbf{F})(\mathbf{J}+\overline{\rho}\mathbf{G})(\mathbf{J}+\overline{\rho}\mathbf{G})'(\mathbf{I}_{NT}-\psi\mathbf{F}')\mathbf{W} + \mathbf{W}(\mathbf{I}_{NT}-\psi\mathbf{F})\overline{\Sigma}(\mathbf{I}_{NT}-\psi\mathbf{F}')\mathbf{W}$$

Rearranging yields

$$(\sigma_\alpha^2 - \overline{\sigma}_\alpha^2)\mathbf{WJJ'W} + (\sigma_\alpha^2(\rho - \overline{\psi}) - \overline{\sigma}_\alpha^2(\overline{\rho} - \psi))\mathbf{W}(\mathbf{JG'} + \mathbf{GJ'})\mathbf{W}$$
$$+ (\overline{\sigma}_\alpha^2\psi\overline{\rho} - \sigma_\alpha^2\overline{\psi}\rho)\mathbf{W}(\mathbf{JH'} + \mathbf{HJ'})\mathbf{W} + (\overline{\sigma}_\alpha^2(\overline{\rho} - \psi)\psi\overline{\rho} - \sigma_\alpha^2(\rho - \overline{\psi})\overline{\psi}\rho)\mathbf{W}(\mathbf{GH'} + \mathbf{HG'})\mathbf{W}$$
$$+ (\sigma_\alpha^2(\rho - \overline{\psi})^2 - \overline{\sigma}_\alpha^2(\overline{\rho} - \psi)^2)\mathbf{WGG'W} + (\sigma_\alpha^2\overline{\psi}^2\rho^2 - \overline{\sigma}_\alpha^2\psi^2\overline{\rho}^2)\mathbf{WHH'W} + \mathbf{W}(\Sigma - \overline{\Sigma})\mathbf{W}$$
$$+ \mathbf{W}(\mathbf{F}(\psi\overline{\Sigma} - \overline{\psi}\Sigma) + (\psi\overline{\Sigma} - \overline{\psi}\Sigma)\mathbf{F'})\mathbf{W} + \mathbf{WF}(\overline{\psi}^2\Sigma - \psi^2\overline{\Sigma})\mathbf{F'W} = \mathbf{0}$$

Since the within-group variance in ϵ conditional on $\mathbf{G}$ is not restricted by (A.1), identification hinges on between-group variance in the outcomes. Consider the covariance terms for the outcomes of group m with those in all other groups, for which we have the equations

$$(\sigma_\alpha^2 - \overline{\sigma}_\alpha^2)\mathbf{E}_m\mathbf{WJJ'WE}'_{-m} + (\sigma_\alpha^2(\rho - \overline{\psi}) - \overline{\sigma}_\alpha^2(\overline{\rho} - \psi))\mathbf{E}_m\mathbf{W}(\mathbf{JG'} + \mathbf{GJ'})\mathbf{WE}'_{-m}$$
$$+ (\overline{\sigma}_\alpha^2\psi\overline{\rho} - \sigma_\alpha^2\overline{\psi}\rho)\mathbf{E}_m\mathbf{W}(\mathbf{JH'} + \mathbf{HJ'})\mathbf{WE}'_{-m}$$
$$+ (\overline{\sigma}_\alpha^2(\overline{\rho} - \psi)\psi\overline{\rho} - \sigma_\alpha^2(\rho - \overline{\psi})\overline{\psi}\rho)\mathbf{E}_m\mathbf{W}(\mathbf{GH'} + \mathbf{HG'})\mathbf{WE}'_{-m}$$
$$+ (\sigma_\alpha^2(\rho - \overline{\psi})^2 - \overline{\sigma}_\alpha^2(\overline{\rho} - \psi)^2)\mathbf{E}_m\mathbf{WGG'WE}'_{-m}$$
$$+ (\sigma_\alpha^2\overline{\psi}^2\rho^2 - \overline{\sigma}_\alpha^2\psi^2\overline{\rho}^2)\mathbf{E}_m\mathbf{WHH'WE}'_{-m} = \mathbf{0}$$

where $\mathbf{E}_m$ and $\mathbf{E}'_{-m}$ are defined in the proof of Proposition 3. If there exists $m \in [M]$ such that $\mathbf{E}_m\mathbf{WJJ'WE}'_{-m}$, $\mathbf{E}_m\mathbf{W}(\mathbf{JG'} + \mathbf{GJ'})\mathbf{WE}'_{-m}$ and $\mathbf{E}_m\mathbf{W}(\mathbf{JH'} + \mathbf{HJ'})\mathbf{WE}'_{-m}$ are maximially linearly independent from the other matrices then we have $\sigma_\alpha^2 = \overline{\sigma}_\alpha^2$ and

$$\overline{\rho} - \psi + \overline{\psi} - \rho = 0 \tag{A.4}$$
$$\psi\overline{\rho} - \overline{\psi}\rho = 0 \tag{A.5}$$

Solving (A.4) for $\overline{\psi}$ and injecting into (A.5) yields $(\overline{\rho} - \rho)(\psi + \rho) = 0$, hence $\overline{\rho} = \rho$ provided that $\psi + \rho \neq 0$. Injecting into (A.4) yields $\overline{\psi} = \psi$.

Table 5: *Monte-Carlo Results: Contextual Effects Only ($\psi = 0, \rho = 0.3$)*

Linear-in-means, $T = 8$	Conditional Mean				Conditional Variance			
Endogenous peer effect ($\psi = 0$)					0.00	0.00	0.00	0.00
					(0.04)	(0.04)	(0.04)	(0.04)
Contextual peer effect ($\rho = 0.3$)	0.30	0.30	0.30	0.30	0.30	0.33	0.30	0.31
	(0.12)	(0.17)	(0.09)	(0.12)	(0.14)	(0.34)	(0.16)	(0.22)
Reduced form peer effect (= 0.3)					0.30	0.32	0.30	0.30
					(0.12)	(0.28)	(0.12)	(0.19)
Linear-in-others'-means, $T = 8$								
Endogenous peer effect ($\psi = 0$)					0.00	0.00	0.00	0.00
					(0.03)	(0.03)	(0.03)	(0.03)
Contextual peer effect ($\rho = 0.3$)	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.31
	(0.09)	(0.13)	(0.07)	(0.09)	(0.11)	(0.21)	(0.11)	(0.15)
Reduced form peer effect (= 0.3)					0.30	0.32	0.30	0.30
					(0.09)	(0.19)	(0.09)	(0.13)
Social network, $T = 8$								
Endogenous peer effect ($\psi = 0$)					0.00	0.00	0.00	0.00
					(0.04)	(0.03)	(0.04)	(0.03)
Contextual peer effect ($\rho = 0.3$)	0.30	0.24	0.30	0.29	0.30	0.31	0.29	0.29
	(0.07)	(0.18)	(0.05)	(0.06)	(0.07)	(0.09)	(0.07)	(0.08)
Linear-in-means, $T = 2$	Conditional Mean				Conditional Variance			
Endogenous peer effect ($\psi = 0$)					-0.01	-0.02	-0.01	-0.02
					(0.08)	(0.10)	(0.09)	(0.12)
Contextual peer effect ($\rho = 0.3$)	0.56	0.23	0.42	0.30	0.24	0.22	0.34	0.30
	(1.24)	(0.69)	(0.67)	(0.56)	(0.52)	(1.14)	(0.26)	(0.79)
Reduced form peer effect (= 0.3)					0.22	0.16	0.32	0.23
					(0.50)	(0.96)	(0.18)	(0.61)
Linear-in-others'-means, $T = 2$								
Endogenous peer effect ($\psi = 0$)					-0.01	-0.02	-0.02	-0.03
					(0.07)	(0.08)	(0.11)	(0.13)
Contextual peer effect ($\rho = 0.3$)	0.40	0.22	0.35	0.29	0.21	0.16	0.32	0.25
	(0.56)	(0.63)	(0.37)	(0.52)	(0.67)	(1.68)	(0.40)	(0.54)
Reduced form peer effect (= 0.3)					0.20	0.13	0.30	0.21
					(0.65)	(1.54)	(0.32)	(0.46)
Social network, $T = 2$								
Endogenous peer effect ($\psi = 0$)					-0.01	-0.01	0.00	0.00
					(0.06)	(0.06)	(0.06)	(0.06)
Contextual peer effect ($\rho = 0.3$)	0.32	0.71	0.31	0.37	0.33	0.34	0.30	0.30
	(0.35)	(0.50)	(0.27)	(0.66)	(0.18)	(0.30)	(0.15)	(0.22)
Correlated Effects	No	Yes	No	Yes	No	Yes	No	Yes
Mobility Rate (p)	0.05	0.05	0.10	0.10	0.05	0.05	0.10	0.10

Notes: For each parameter and design we report the mean and standard deviation (in parenthesis) over 500 datasets. If there are no correlated effects this information is treated as known to the researcher. For the conditional mean estimator we impose $\gamma = \mathbf{0}$. For the conditional variance estimator we use $\mathbf{W} = \mathbf{I}_{NT}$. The reduced form peer-effect for the linear-in-means network is $(\psi + \rho)(1 - \psi)^{-1} = 0.3$. The counterpart for the linear-in-others'-means network is $(\psi + \rho)(1 - \psi(\psi + N_{g(i,t)} - 2)/(N_{g(i,t)} - 1))^{-1}$ for individual i in year t in group $g(i,t)$. We report simulation results for $N_{g(i,t)} = 5$, which is the mean group size, with true value equal to 0.3. We do not report a reduced form effect for the Social Network because it is individual-specific, depending on the network structure in their group. Correlated effects are time-invariant.

Table 6: *Monte-Carlo Results: Contextual and Endogenous Effects* ($\psi = 0.2, \rho = 0.3$)

Linear-in-means, $T = 8$	Conditional Mean				Conditional Variance			
Endogenous peer effect ($\psi = 0.2$)					0.20	0.19	0.20	0.20
					(0.03)	(0.04)	(0.04)	(0.04)
Contextual peer effect ($\rho = 0.3$)	0.76	0.89	0.76	0.83	0.30	0.34	0.30	0.31
	(0.18)	(0.24)	(0.13)	(0.18)	(0.15)	(0.39)	(0.17)	(0.24)
Reduced form peer effect (= 0.625)					0.63	0.65	0.62	0.62
					(0.15)	(0.38)	(0.15)	(0.24)
Linear-in-others'-means, $T = 8$								
Endogenous peer effect ($\psi = 0.2$)					0.20	0.20	0.20	0.20
					(0.03)	(0.03)	(0.03)	(0.03)
Contextual peer effect ($\rho = 0.3$)	0.67	0.76	0.67	0.72	0.30	0.31	0.30	0.30
	(0.11)	(0.13)	(0.08)	(0.10)	(0.11)	(0.24)	(0.11)	(0.16)
Reduced form peer effect (= 0.595)					0.59	0.59	0.59	0.59
					(0.11)	(0.25)	(0.10)	(0.16)
Social network, $T = 8$								
Endogenous peer effect ($\psi = 0.2$)					0.20	0.20	0.20	0.20
					(0.04)	(0.04)	(0.04)	(0.03)
Contextual peer effect ($\rho = 0.3$)	0.56	0.53	0.56	0.56	0.30	0.31	0.29	0.29
	(0.06)	(0.19)	(0.05)	(0.05)	(0.08)	(0.10)	(0.08)	(0.08)
Linear-in-means, $T = 2$	Conditional Mean				Conditional Variance			
Endogenous peer effect ($\psi = 0.2$)					0.19	0.18	0.19	0.18
					(0.07)	(0.09)	(0.07)	(0.10)
Contextual peer effect ($\rho = 0.3$)	2.00	0.83	1.66	0.93	0.24	0.22	0.34	0.31
	(3.28)	(0.62)	(2.43)	(0.36)	(0.53)	(1.21)	(0.27)	(0.86)
Reduced form peer effect (= 0.625)					0.53	0.44	0.65	0.54
					(0.63)	(1.25)	(0.23)	(0.79)
Linear-in-others'-means, $T = 2$								
Endogenous peer effect ($\psi = 0.2$)					0.19	0.18	0.19	0.19
					(0.06)	(0.07)	(0.07)	(0.07)
Contextual peer effect ($\rho = 0.3$)	1.11	0.94	1.07	0.97	0.18	0.16	0.31	0.21
	(0.64)	(0.52)	(0.45)	(0.28)	(0.64)	(1.40)	(0.28)	(0.57)
Reduced form peer effect (= 0.595)					0.48	0.41	0.59	0.46
					(0.74)	(1.52)	(0.27)	(0.58)
Social network, $T = 2$								
Endogenous peer effect ($\psi = 0.2$)					0.19	0.21	0.20	0.20
					(0.06)	(0.09)	(0.06)	(0.06)
Contextual peer effect ($\rho = 0.3$)	0.70	0.67	0.71	0.74	0.33	0.35	0.31	0.29
	(0.30)	(0.39)	(0.23)	(0.26)	(0.20)	(0.39)	(0.17)	(0.24)
Correlated Effects	No	Yes	No	Yes	No	Yes	No	Yes
Mobility Rate (p)	0.05	0.05	0.10	0.10	0.05	0.05	0.10	0.10

Notes: For each parameter and design we report the mean and standard deviation (in parenthesis) over 500 datasets. If there are no correlated effects this information is treated as known to the researcher. For the conditional mean estimator we impose $\gamma = \mathbf{0}$. For the conditional variance estimator we use $\mathbf{W} = \mathbf{I}_{NT}$. The reduced form peer-effect for the linear-in-means network is $(\psi + \rho)(1 - \psi)^{-1} = 0.625$. The counterpart for the linear-in-others'-means network is $(\psi + \rho)(1 - \psi(\psi + N_{g(i,t)} - 2)/(N_{g(i,t)} - 1))^{-1}$ for individual i in year t in group $g(i, t)$. This lies between 0.5208 ($N_{g(i,t)} = 2$) and 0.6250 ($N_{g(i,t)} \to \infty$.) We report simulation results for $N_{g(i,t)} = 5$, which is the mean group size, with true value equal to 0.5952. We do not report a reduced form effect for the Social Network because it is individual-specific, depending on the network structure in their group. Correlated effects are time-invariant. In designs with endogenous effects the conditional mean estimator of ρ diverges on rare occasions. For such designs we report the mean and standard deviation for the truncated distribution lying between -100 and 100.