

Does the rotten child spoil his companion? Spatial peer effects among children in rural India

CHRISTIAN HELMERS
Santa Clara University and SERC LSE

MANASA PATNAM
CREST (ENSAE)

This paper identifies the effect of neighborhood peer groups on childhood skill acquisition using observational data. We incorporate spatial peer interaction, defined as a child's nearest geographical neighbors, into a production function of child cognitive development in Andhra Pradesh, India. Our peer group definition takes the form of networks, whose structure allows us to identify endogenous peer effects and contextual effects separately. We exploit variation over time to avoid confounding correlated with social effects. Our results suggest that spatial peer and neighborhood effects are strongly positively associated with a child's cognitive skill formation. Further, we explore the effect of peer groups in helping to provide insurance against the negative impact of idiosyncratic shocks to child learning. We find that the data reject full risk-sharing, but cannot rule out the existence of partial risk-sharing on behalf of peers. We show that peer effects are robust to different specifications of peer interactions and investigate the sensitivity of our estimates to potential misspecification of the network structure using Monte Carlo experiments.

Christian Helmers: chelmers@scu.edu

Manasa Patnam: manasa.patnam@ensae.fr

The authors thank the editor Orazio Attanasio, two anonymous referees, Quy-Toan Do, Markus Eberhardt, Marcel Fafchamps, Steve Gibbons, Xavi Gine, Paul Glewwe, Markus Goldstein, Manos Kitsios, Henry Overman, Hanan Jacoby, Stefan Klöpper, Pramila Krishnan, Owen Ozier, Aureo de Paula, Pia Pinger, Biju Rao, and Matthias Schündeln for helpful discussions and comments. We also benefitted from the comments of participants at the DECRG Applied Micro and Poverty Seminar, a seminar at Goethe University Frankfurt, the Midwest Development Conference 2010, the 7th Annual Conference of the Verein für Socialpolitik Research Committee on Development Economics, the 25th Annual Congress of the European Economic Association 2010, the NEUDC Conference 2010, and the Econometric Society World Congress 2010 in Shanghai. We especially thank Sandhya Rao and Anamika Arora for help related to the Geographic Information System (GIS). The authors acknowledge Young Lives for providing household and GIS data for the purpose of this research. Young Lives (www.younglives.org.uk) is a long-term international research project that investigates the changing nature of childhood poverty. Young Lives is core-funded by U.K. aid from the Department for International Development (DFID) for the benefit of developing countries. Substudies are funded by the Bernard van Leer Foundation, the Inter-American Development Bank (in Peru), the International Development Research Centre (in Ethiopia), and the Oak Foundation. The views expressed are those of the authors. They are not necessarily those of, or endorsed by, Young Lives, the University of Oxford, DFID, or other funders.

Copyright © 2014 Christian Helmers and Manasa Patnam. Licensed under the [Creative Commons Attribution-NonCommercial License 3.0](https://creativecommons.org/licenses/by-nc/3.0/). Available at <http://www.qeconomics.org>.

DOI: [10.3982/QE192](https://doi.org/10.3982/QE192)

KEYWORDS. Children, peer effects, cognitive skills, India.

JEL CLASSIFICATION. C21, O15, R23.

1. INTRODUCTION

There is a sizeable economic literature on cognitive and noncognitive skill formation of children (Todd and Wolpin (2003), Cunha and Heckman (2007)). Using a production function framework, this literature investigates the determinants of the child's cognitive and non-cognitive skills creation. The most recent advances in this literature attribute important roles to self-productivity and cross-productivity of cognitive and noncognitive skills as well as to parental investment (Cunha and Heckman (2008)).¹

At the same time, there is an important literature on peer and neighborhood effects (Borjas (1995), Hoxby (2000), Sacerdote (2001), Hanushek, Kain, Markman, and Rivkin (2003), Durlauf (2004), Lin (2010)). In this literature, individual outcomes are influenced by spatial instead of market interactions of individuals, that is, the probability of observing an individual behaving in a certain way is a function of either some characteristics of the individual's environment (neighborhood effects) or depends directly on the prevalence of this type of behavior among her peers (peer effects).²

Durlauf (2004) listed three specific channels through which such neighborhood and peer effects are mediated. First, psychological factors can affect a child's desire to behave like others, for example, purely imitative behavior. Second, interdependencies in the constraints children face motivate similar behavior because the costs associated with a given behavior depend on whether others behave in the same way, for example, the reduction of stigma arising from deviant behavior. Third, behavior of other children may alter the child's subjective perceptions regarding the returns to such behavior, for example, expected income from an additional year of schooling. Intuitively, all channels depend on the existence of contact between individuals. The probability of contact and its intensity may be a function of geographical distance between individuals, family or friendship ties, and so forth. Regardless of the channel, in the presence of peer effects, children are directly influenced by actions and characteristics of their peers. Therefore, peer effects may be an important determinant of a child's cognitive and noncognitive skills development.

The identification of peer effects encounters well known problems laid out in Manski (1993). Manski (1993) listed three effects that need to be distinguished in the analysis of peer effects. The first type are endogenous effects, which arise from an individual's propensity to behave in some way as a function of the behavior of the group; the second are so-called contextual effects, which represent the propensity of an individual to behave in some way as a function of the exogenous characteristics of his peer group. The third type are so-called correlated effects, which arise due to factors that are common

¹Self-productivity refers to any effect of past periods' cognitive/noncognitive skills on current period's cognitive/noncognitive skills, respectively, while cross-productivity refers to any effect of past periods' cognitive/noncognitive skills on current period noncognitive/cognitive skills.

²Our definitional distinction between neighborhood and peer effects is somewhat arbitrary as Durlauf and Ioannides (2010) noted in their article for the *New Palgrave Dictionary of Economics* that "[t]erms like social interactions, neighbourhood effects, social capital and peer effects are often used as synonyms."

among individuals who belong to the same group and compel them to behave in a similar manner. For example, children within the same village may behave similarly because they face a common institutional arrangement. This means that there are group-level unobservables that may have a direct effect on observed outcomes, that is, disturbances may be correlated across individuals in a group. The main empirical challenges, therefore, consist in (i) disentangling *contextual* effects (i.e., the influence of exogenous peer characteristics on a child's observed outcome) and *endogenous* effects (i.e., the influence of peer outcomes on a child's outcome), and (ii) distinguishing between *social* effects (i.e., exogenous and endogenous effects) and *correlated* effects (i.e., children in the same peer group may behave similarly because they share a common environment). Such correlated effects can also include sorting of households, that is, an endogenous location choice by households. The identification problem explains why existing work that looks at children's and teenagers' cognitive outcomes incorporating neighborhood effects, such as Brooks-Gunn, Duncan, Klebanov, and Sealand (1993), McCulloch and Joshi (2001), and Ainsworth (2002), only accounts for contextual effects and assumes the absence of endogenous effects.

In this paper, we use observational data from the Young Lives (YL) project for Andhra Pradesh, India, to examine neighborhood-level peer influences on child cognitive development by estimating a production function of a child's cognitive ability, accounting for endogenous and contextual peer effects. We regard our empirical specification, which explicitly allows children to be influenced by and learn from their peers, as a step forward toward a more realistic model of skill formation.

A common justification for neglecting peer effects in the analysis of child skill formation in the existing literature is the lack of appropriate data. The most commonly used data set in this line of research, the U.S. National Longitudinal Survey of Youth (NLSY), is the result of stratified sampling, which justifies the assumption of independence of children within the data set. Even if information that revealed the identity of a child's peers were available in the survey data, these peers would most likely not have been included in the sample. The nature of the available data, therefore, severely limits the ability to investigate the potential impact of peer effects on skill formation.

In principle, the same applies to the YL data used in our analysis, which justified treating children as independent units in earlier work. However, we show that in our data, surveyed households are located in close vicinities within villages due to the small overall size of the surveyed rural villages in Andhra Pradesh. The presence of this clustered spatial pattern allows us to employ geographical proximity between children to identify spatial peer effects on child cognitive outcomes. Our main identifying assumption is, therefore, that peer effects arise through spatial contiguity between children: children who live next door to each other are more likely to interact and influence each other than are children at the other ends of the village. We thus construct a child's peer group based on geographical proximity of other similarly aged children *within* the same village using GIS location data.³ The resulting structure of peer groups enables

³Children in our sample are of the same age, which means that peer influences are reciprocal and contemporaneous, which is distinct from role-model influences that could emerge if younger children imitate the behavior of older individuals (Durlauf (2004)).

us to disentangle contextual and endogenous effects despite the lack of experimental data. We use mainly longitudinal variation in our data to avoid confounding social effects with correlated effects, but also employ a number of alternative identification approaches. For estimation, we use Kelejian and Prucha's (2007) spatial nonparametric heteroscedasticity and autocorrelation consistent (SHAC) instrumental variable (IV) estimator. The SHAC approach allows us to remain agnostic regarding the structure of the spatial dependence in the residuals and to allow for heteroscedasticity of any arbitrary form.

Our results suggest that a child's geographical neighbors are positively associated in a statistically significant and economically important way with a child's own production of cognitive skills between age 8 and 12. Furthermore, we find that contextual effects appear to have little influence on cognitive achievement gains.

The fact that we observe only a random sample of children within villages, in principle, does not undermine identification of peer effects because children have not been selected into the sample based on their location within villages. However, sampling means that we only observe a fraction of a child's true peer network, which implies that our assumed network structure might be misspecified. Regardless of sampling, even if the population of children were observed within villages, misspecification of peer links between children might still occur given the lack of information on actual peer interaction. We investigate the implications of the potential misspecification of our proximity-based peer network structure and demonstrate the robustness of spatial peer estimates to different forms of misspecification due to both sampling and different assumptions about the true population peer network structure. Nevertheless, we only have observational data for our analysis, which means that despite our identification strategy and robustness checks, our results may not warrant a causal interpretation.

As an additional contribution, we use the augmented skill production function to examine the relevance of peer groups in assisting children to recover from shocks. We investigate whether the presence of a peer group helps insure children of shock-affected households against an adverse impact to their cognitive achievement gain. Many studies have found that economic, health, or climatic shocks to a household have a negative impact on child schooling and health. This is because in such circumstances households typically tend to underinvest in education or health related expenditure for their children. While a considerable amount of the literature is devoted to examining risk-sharing and informal insurance arrangements of households (Townsend (1994), Gertler and Gruber (2002), De Weerd and Dercon (2006)), there is little evidence to show how children from such households find support to cope with adversities that significantly compromise investment in their education and health (Tominey (2009), Ginja (2009)). By utilizing detailed data on idiosyncratic household related shocks, we find that the negative effect of a shock on child cognitive achievement becomes statistically insignificant after incorporating peer effects, which we interpret as evidence that peer groups provide partial insurance. Moreover, our analysis suggests that this peer insurance effect applies only to boys.

Our results contribute to the empirical literature on childhood skill formation by providing evidence for the presence and importance of peer and neighborhood effects in the formation of children's cognitive skills in a rural developing country context. Our evidence also suggests an important role for peers in insuring children against adverse idiosyncratic shocks. Moreover, we contribute to the existing peer effects literature by providing an example of how to identify endogenous and contextual peer effects without the need for data from a controlled randomized experiment by using GPS location data that are routinely collected in household surveys. Our research design may, therefore, be applicable in any context in which peer effects are mediated through spatial proximity and location data are available. It can be applied to study peer effects in other contexts both on other outcome variables and populations of interest. Our ability to separate social effects from correlated effects through longitudinal variation as well as the network structure allows us to investigate the potential bias from conflating social effects and peer group-level unobservables. Our results point to a sizeable downward bias in the coefficient on endogenous peer effects if correlated effects are ignored.

From a policy perspective, understanding the role of social interactions and peer effects in shaping childhood skill formation is important, as policy interventions targeting only a subset of children of a population may influence outcomes of other children who are not directly included in the intervention.⁴ Because of the bidirectional nature of peer effects, their presence also implies social multiplier effects that magnify the impact of policy interventions (Manski (1993), Bobonis and Finan (2009)). As noted by Durlauf (2004), peer effects can also lead to persistence in poverty as neighborhoods can get locked in bad equilibria that are enforced over time by the mutually reinforcing character of peer effects. Therefore, improved understanding of the role of peer interaction, in particular in a developing country context, may contribute to the design of novel interventions aimed at improving children's cognitive skill production and thus success in later life.

The paper is organized as follows. Section 2 discusses our identification strategy, which is translated into the specification of our empirical model presented in the same section. The SHAC instrumental variables estimator used in our analysis is presented in Section 3. The data used are described in Section 4. Section 5 discusses our results and reports our robustness checks; Section 6 concludes.

2. EMPIRICAL STRATEGY

Our analysis of the YL data for Andhra Pradesh, India, reveals close geographical proximity of the surveyed households within communities. Communities represent an administratively defined geographical area: a neighborhood in (semi)urban areas and a village in rural areas. As a result, we use the terms “communities” and “villages” interchangeably. As an example, Figure 1 shows the map of a sample village in Andhra Pradesh. The figure suggests that groups of households are located close to each other within the vil-

⁴See Fafchamps and Vicente (2009) for evidence of such “diffusion” effects in the context of political awareness campaigns.



FIGURE 1. GIS data map: example of a sample village.

lage; the circle on the map has a radius of 250 meters and it encircles almost the entire village.⁵ In fact, the median distance between households within the networks used to define peer interaction, which we discuss further below, is 126 meters.⁶ This short distance is striking in light of the fact that most households are located in rural areas.

The spatial proximity of households allows us to identify surveyed households' geographical neighbors that we use to define each child's peers. The clustering is important, because the close geographical proximity of households allows us to reasonably argue that households interact as neighbors.⁷ But note that spatial clustering among sample households is not a feature of the sampling technique. Our data are a random sample of

⁵Due to data confidentiality agreements, we are unable to disclose the exact locations of sample households. Instead, we report detailed statistics on the proximity of sample households throughout the paper.

⁶The average distance within networks is 447 meters. These figures refer to networks defined as a child's five nearest neighbors.

⁷We obtain some descriptive information on how children spend their time from the YL child-level questionnaire at age 12. We know, for example, that the median amount of time that children spend playing with their peers is 4 hours a day, which leaves ample room for neighborhood-based peer interaction.

households *within* communities (i.e., households have not been selected into the sample based on their location within communities; see Section 4). Hence, we can assume that the observed spatial distribution of observed households is representative of the true underlying spatial distribution of households in the population (more discussion is provided in Section 2.3). This allows us to use a neighborhood-based definition of social interactions.⁸ In addition, we are able to construct our measure of peer effects based on nearest neighbor networks that represent a child’s peer group, since only children of the same age are included in the sample.

These networks are used to estimate the effect that a child’s spatial peer group has on the child’s own cognitive skill formation.⁹ We denote children as i ($i = 1, \dots, n$), y_{it} denotes the cognitive achievement of child i in period t , and \mathbf{x}_{it} is a $1 \times K$ vector of child and household characteristics (we denote vectors with bold lowercase letters and matrixes with bold capital letters). Each child has a peer group P_i of size n_i . By assumption, child i is excluded from P_i . Peer groups are formed within communities (which correspond to villages in rural areas). This means that the maximally connected network of each child is defined at the level of the community C to which the child belongs.

We distinguish between two types of effects that child i ’s peers can have on i ’s cognitive skills y_i : (i) the mean outcome of her peer group (*endogenous effects*: $\frac{\sum_{j \in P_i} y_{jt}}{n_i}$) as well as (ii) the mean characteristics of her peer group (*contextual effects*: $\frac{\sum_{j \in P_i} x_{jt}}{n_i}$). Our regression model, therefore, is

$$y_{it} = \alpha + \beta \frac{\sum_{j \in P_i} y_{jt}}{n_i} + \gamma \mathbf{x}_{it} + \delta \frac{\sum_{j \in P_i} x_{jt}}{n_i} + \theta_i + \phi_{Ct} + \varphi_{Pt} + u_{it}, \tag{1}$$

where α is a constant, β captures endogenous effects, δ captures contextual effects, γ captures child i ’s own, caregiver, and household characteristics, θ_i captures child-specific effects, and ϕ_{Ct} and φ_{Pt} represent *correlated effects* that subsume a range of possibly time-varying community (ϕ_{Ct}) and peer-group (φ_{Pt}) level unobservables that are correlated with the dependent and independent variables included in the analysis. As discussed in detail below, these effects may, for example, include the potentially endogenous sorting of households into geographical locations. We do not require the residuals u_{it} to be homoscedastic or normally distributed.

There are two main challenges to the identification of model (1): (i) the separate identification of endogenous β and contextual δ effects, and (ii) distinguishing social effects, that is, endogenous and contextual effects, from correlated effects. In this section, we explain these two conceptually distinct identification issues and show how we

⁸Examples of other studies that study peer and neighborhood interaction based on geographical proximity include Topa (2001), Conley and Topa (2007), Bayer, Ross, and Topa (2008), who defined location more broadly at the census tract level, and Goux and Maurin (2007), who used survey data on adjacent households. However, in contrast to our data, in Goux and Maurin (2007), close proximity of sampled households is the outcome of the sampling technique.

⁹We assume that interactions between children occur exclusively through social interactions and are thus unrelated to market interactions, which appears to be a reasonable assumption in our setting.

address them with the available data, which consist of observations on individual children at two points in time and a social network structure that remains unchanged over time.¹⁰

2.1 Identification of endogenous peer effects

The identification of endogenous and contextual peer effects is notoriously difficult as explained by Manski (1993) and Moffitt (2001) (for a summary of the literature, see also Blume and Durlauf (2006)). Manski noted that within a linear framework without additional information, it is impossible to infer from the observed mean distribution of a sample whether average behavior within a group affects the individual behavior of members of that group. The expected mean outcome of a peer group and its mean characteristics are perfectly collinear due to the simultaneity induced by social interaction. This fundamental identification problem, termed *reflection problem* by Manski, means that within a linear-in-means model, identification of peer effects depends on the functional relationship in the population between the variables that characterize peer groups and those that directly affect group outcomes.

Lee (2007) was first to show formally that the spatial autoregressive model (SAR) specification, widely used in the spatial econometrics literature, can be used to disentangle endogenous and exogenous effects. In a SAR model, identification of endogenous and contextual effects is possible if there is sufficient variation in the size of peer groups within the sample. Bramoullé, Djebbari, and Fortin (2009) (BDF henceforth) proposed an encompassing framework in which Manski's mean regression function and Lee's SAR specification arise as special cases. BDF showed that endogenous and exogenous effects can be distinguished through specific network structures, for example, the presence of intransitive triads within a network. Intransitive triads describe a structure in which individual i interacts with individual j but not with individual k , whereas j and k interact.

2.1.1 Peer interaction structure Our identification strategy relies on this insight. We construct a peer interaction structure where children interact with each other on the basis of their spatial proximity. We denote the network of peer interactions, in the form of an *adjacency matrix*,¹¹ as \mathbf{W} . We define \mathbf{W} using two different strategies described below; identification conditions implied by both strategies are discussed in Section 2.1.2.

¹⁰The assumption of a fixed network structure is made on the basis that networks are defined according to the location of the households in which children live. Since households in our sample do not move during the two observed time periods, the network structure is fixed (see Section 4). Generally, sample attrition in the YL data for Andhra Pradesh is very low, with 1.29% between the two survey rounds for the "older" cohort of children used in our analysis (Outes-Leon and Dercon (2008)).

¹¹A common way to represent connectivity of network graphs is through a $n \times n$ binary symmetric matrix called an adjacency matrix (n denotes the number of observations on the dependent variable $y = (y_1, \dots, y_n)$). The adjacency matrix is nonzero for entries whose row-column indices correspond to a link between two units and is zero for those that have no links. By definition, a unit is not a neighbor to itself, which means the diagonal of \mathbf{W} contains only zeroes. Operations on the adjacency matrix yield additional information about the network such as degree, clustering, and so forth. For more on adjacency matrices and properties of network graphs, see Kolaczyk (2009).

We illustrate each strategy using an example adjacency matrix of five children who belong to two different communities. Children 1, 2, and 3 belong to community C1, while children 5 and 4 belong to community C2.

| | | C1 | | | C2 | |
|----|---|----------|---|----------|----|---|
| | | 1 | 2 | 3 | 4 | 5 |
| C1 | 1 | 0 | 1 | 0 | 0 | 0 |
| | 2 | 1 | 0 | 1 | 0 | 0 |
| | 3 | 0 | 1 | 0 | 0 | 0 |
| C2 | 4 | 0 | 0 | 0 | 0 | 1 |
| | 5 | 0 | 0 | 0 | 1 | 0 |

(a) KNN

| | | C1 | | | C2 | |
|----|---|----------|---|----------|----|---|
| | | 1 | 2 | 3 | 4 | 5 |
| C1 | 1 | 0 | 1 | 1 | 0 | 0 |
| | 2 | 1 | 0 | 1 | 0 | 0 |
| | 3 | 1 | 1 | 0 | 0 | 0 |
| C2 | 4 | 0 | 0 | 0 | 0 | 1 |
| | 5 | 0 | 0 | 0 | 1 | 0 |

(b) Community

(a) **W**: *K nearest neighbor (KNN) classification*: Our first peer interaction structure, KNN, uses a distance-based definition of neighbors, where K refers to the number of neighbors in a location. Distances are computed by the Euclidean distance between global positioning satellite (GPS) locations of households. Therefore, under this approach, the set of “neighbors” for child i includes the K children characterized by the shortest distance to child i within each community. We set $K = 5$.¹² That means w_{ij} , which is element (i, j) of **W**, is equal to $1/5$ if child j is in the set of five nearest neighbors of child i and zero otherwise (this implies $\sum_{j=1}^n w_{ij} = 1$ for all i).¹³ In our stylized example adjacency matrix (a) above, child 2 is a nearest neighbor to both child 3 and child 1, whereas child 1 and child 3 are not nearest neighbors. The choice of K poses a problem similar to the well known modifiable areal unit problem in spatial econometrics (Openshaw (1983)). We are thus careful to check the robustness of our results to modifications in the definition of K (see Section 5.3). Using this method, we drop households that are not a nearest neighbor to any other household in the sample.¹⁴ Depending on the number of nearest neighbors used in our definition of **W**, this leads us to drop a small number of households, which causes slight variations in the sample size across specifications (see Section 5). Given that the households are a random sample

¹²The child-level questionnaire at age 12 asks children to indicate the number of friends they have. On average, children report having 7 friends, with a standard deviation of 4.8. During the first sampling round, the prevalence of 1-year-old children in the population was estimated to be 2%. This meant only villages with at least 5000 inhabitants were selected among the sample villages to ensure that the target sample size of 100 children for the “younger” cohort was met (Himaz, Galab, and Reddy (2009)). Considering child (1–5 years old) mortality of 21% (Indian National Family Health Surveys 1998–1999), the prevalence of 8-year-old children in the population was most likely lower at around 1.6%. This means that the 50 sampled children of the “older” cohort represent up to 60% of the population of 8-year-old children in each sampled cluster (see also Section 4). This implies that a five nearest neighbor specification appears to be appropriate to capture the share of each child’s friends included in the sample. Nevertheless, for robustness, we also estimate our model using a network definition based on three and seven nearest neighbors (see Section 5.3).

¹³By definition, a child is not a neighbor to himself $w_{ii} = 0$.

¹⁴In fact, for such island households, column sums of the spatial weight matrix **W** are zero. This occurs as a result of specifying a directed network structure, that is, while any node has five nearest neighbors, this does not automatically imply the node itself represents one out of five nearest neighbors to any other node in the network.

of the underlying population, dropping such “island” households should not bias our results (see also Section 2.3.1).

(b) \mathbf{W}^C : *Community classification*: Alternatively, we construct the peer reference group as all children of the same age who belong to the same community, where communities are defined as a neighborhood in (semi)urban areas and as a village in rural areas. This means that the community adjacency matrix, denoted by \mathbf{W}^C , is block diagonal with children interacting with all other children from within their respective communities. In the stylized example adjacency matrix (b) above, all children from community C1 (child 1, child 2, and child 3) are assumed to interact with each other since they belong to the same community.

2.1.2 *Identification conditions* (a) *K nearest neighbor (KNN) classification*: Having defined \mathbf{W} , we can rewrite Equation (1) using matrix notation as (subsuming unobservables θ_i , ϕ_{Ct} , φ_{Pt} , and u_{it} in v_{it})

$$\mathbf{y}_t = \alpha \mathbf{1} + \beta \mathbf{W} \mathbf{y}_t + \gamma \mathbf{x}_t + \delta \mathbf{W} \mathbf{x}_t + \mathbf{v}_t, \quad (2)$$

where \mathbf{y}_t is an $n \times 1$ vector of skill outcomes, $\mathbf{1}$ is an $n \times 1$ vector of ones, $\mathbf{W} \mathbf{y}_t$ now denotes the average peer outcome (endogenous effects), and $\mathbf{W} \mathbf{x}_t$ denotes average peer characteristics (contextual effects). This implies, that the reduced form is given by

$$\mathbf{y}_t = \alpha (\mathbf{I} - \beta \mathbf{W})^{-1} \mathbf{1} + (\mathbf{I} - \beta \mathbf{W})^{-1} (\gamma \mathbf{I} + \delta \mathbf{W}) \mathbf{x}_t + (\mathbf{I} - \beta \mathbf{W})^{-1} \mathbf{v}_t. \quad (3)$$

If we omitted the endogenous effects $\mathbf{W} \mathbf{y}_t$ from Equation (2), the model could be estimated using ordinary least squares (OLS) under the assumption that all covariates are strictly exogenous. However, OLS is biased and inconsistent in the presence of a spatial autoregressive lag (Anselin (1998)). Denoting the variance–covariance matrix of \mathbf{v}_t as $\boldsymbol{\psi}_{v_t}$, it is easy to see that

$$E[(\mathbf{W} \mathbf{y}_t) \mathbf{v}_t] = \mathbf{W} (\mathbf{I} - \beta \mathbf{W})^{-1} \boldsymbol{\psi}_{v_t} \neq 0. \quad (4)$$

Anselin (1998) suggested an approach based on a maximum likelihood (ML) estimator to address the endogeneity problem. To avoid computation accuracy problems in the ML approach noted by Kelejian and Prucha (1999), Kelejian and Prucha (1998) suggested a spatial two-stage least squares estimator (S2SLS). They suggested using a set of instruments for $(\mathbf{x}_t, \mathbf{W} \mathbf{x}_t, \mathbf{W} \mathbf{y}_t)$. From Equation (4), we can see that, ideally, the set of instruments contains linearly independent columns of $[\mathbf{x}_t, \mathbf{W} \mathbf{x}_t, \mathbf{W}^2 \mathbf{x}_t]$. Hence, identification of endogenous and contextual effects is possible if \mathbf{I} , \mathbf{W} , and \mathbf{W}^2 are linearly independent (provided that $\gamma\beta + \delta \neq 0$).

As suggested by BDF, we use the network structure of our peer group to meet this condition. This is the case when the network is characterized by (a small degree of) intransitivity, for example, child i and child j are nearest neighbors, child j and child k are nearest neighbors, but child i and child k are *not* nearest neighbors (see example adjacency matrix (a) above where child 1 is friends with 2 but not 3, whereas 2 is friends with both 1 and 3). This produces a network topology that achieves identification of peer

effects. The network-based intuition of this strategy is straightforward: $\mathbf{W}^2\mathbf{x}_t$ is an identifying instrument for $\mathbf{W}\mathbf{y}_t$, since x_{kt} affects y_{jt} (since k and j are connected and interact with each other), but x_{kt} can only affect y_{it} indirectly through its effect on y_{jt} . Therefore, given our peer network structure, $[\mathbf{x}_t, \mathbf{W}\mathbf{x}_t, \mathbf{W}^2\mathbf{x}_t]$ are valid and informative instruments for endogenous peer effects $\mathbf{W}\mathbf{y}_t$.

(b) *Community classification*: Peer effects are still identified when using the community-based peer interaction matrix \mathbf{W}^C , provided children interact in community-based groups of different sizes (Lee (2007)). The interaction matrix defined at the community level, \mathbf{W}^C , has block diagonal elements C of varying sizes, where all children are interconnected within each community C (see example matrix (b) above). Provided there are at least three communities of different sizes (Davezies, D'Haultfoeuille, and Fougère (2009)), variation in reduced-form coefficients across communities of different sizes ensures identification. This follows from the structure of \mathbf{W}^C , which excludes a given child from his peer group, which in turn creates child-level variation in peer average outcomes and characteristics within communities. This alternative peer-group definition captures both peer and neighborhood effects, and offers a less restrictive way to specify the structure of underlying peer interaction as it avoids having to impose network-based exclusion restrictions and, hence, any assumptions on the number and direction of peer links.

2.2 Correlated and selection effects

In the reduced-form expression (3), coefficients α , β , γ , and δ are identified because of the intransitivity in peer interaction as captured by \mathbf{W} . In addition, we assumed strict exogeneity of \mathbf{x}_t , that is, $\mathbb{E}[\mathbf{v}_t|\mathbf{x}_t] = 0$. However, if \mathbf{v}_t contains child-, peer-group-, and community-level-specific unobservables, identification fails due to such correlated effects.

Correlated effects represented by ϕ_{Ct} and φ_{Pt} in Equation (1) occur when individuals within a community or peer group behave similarly due to the common environment that they face. This problem may arise in our setting, for example, if children within a peer group attend the same school or are subject to the same macroshocks. Selection effects, which can be subsumed under correlated effects, arise when an individual chooses his own peer/reference group, that is, individuals have not been assigned randomly into peer groups; such selection effects can occur also due to child-specific unobservables θ_i . Group formation is endogenous, for example, when popular students interact primarily with other popular students or when households sort themselves into a locality of their choice.¹⁵ In our case, a possible concern is that households sort into neighborhoods based on factors that influence their children's cognitive development. While in the deprived rural setting of Andhra Pradesh, it seems rather unlikely that parents choose the location of a household based on school characteristics, there may still

¹⁵Even when assignment into peer groups is random, members of a peer group may still be exposed to the same environment and (time-varying) common shocks. This means that even in the presence of random peer-group formation, that is, data from a randomized experiment, correlated effects represent a challenge to the identification of peer effects.

TABLE 1. Overview of empirical approaches to address unobservable effects.

| Approach | | Absorbed Unobservables | | | | |
|----------|-----------------------------|------------------------|----------|-------------|-------------|----------------|
| | | μ_i | ϕ_C | φ_P | ϕ_{Ct} | φ_{Pt} |
| 1. | POLS | NO | NO | NO | NO | NO |
| 1(a) | Older siblings ^a | NO | YES | NO | YES | NO |
| 1(b) | Within transformation | NO | YES | NO | NO | NO |
| 1(c) | IV using shocks | NO | YES | YES | YES | YES |
| 2. | FD | YES | YES | YES | NO | NO |
| 2(a) | Older siblings ^a | YES | YES | YES | YES | NO |
| 2(b) | Within transformation | YES | YES | YES | YES | NO |
| 2(c) | IV using shocks | YES | YES | YES | YES | YES |

Notes: POLS, pooled OLS; FD, first-differenced OLS. μ_i denotes child-specific time-invariant unobservables. Correlated effects are denoted as $\phi_C + \varphi_P + \phi_{Ct} + \varphi_{Pt}$, where ϕ_C denotes time-invariant network-level effects, φ_P denotes time-invariant peer-group-specific effects, ϕ_{Ct} denotes time-varying network-specific effects, and φ_{Pt} denotes time-varying peer-group-specific effects.

^aThe approach that uses older siblings' schooling only captures a specific type of ϕ_{Ct} related to changes in community-level schooling.

be other location-specific characteristics that attract households that, for example, attach greater importance to the education of their children.

We employ different approaches to address the issue of correlated and selection effects, which are summarized in Table 1. For illustration, we decompose correlated effects into time-invariant network-specific ϕ_C (in our setting this corresponds to community-specific) and peer-group-level φ_P (in our setting this corresponds to the nearest neighbor network) unobservables, and time-varying network-specific ϕ_{Ct} and peer-group-specific φ_{Pt} unobservables. As shown in Table 1, relying on OLS would assume the absence of child-specific θ_i as well as any community-level or peer-group-level correlated effects.

To account for time-invariant child-specific effects θ_i and correlated effects ϕ_C and φ_P , we use a first-differenced specification following Blume and Durlauf (2006). This means that we explain the change in cognitive skill levels achieved by children between t and $t - 1$ as

$$\Delta y_i = \beta \frac{\sum_{j \in P_i} \Delta y_j}{n_i} + \gamma \Delta x_i + \delta \frac{\sum_{j \in P_i} \Delta x_j}{n_i} + \Delta v_i, \quad (5)$$

where $\Delta y_i = y_{i,t} - y_{i,t-1}$ denotes the difference in cognitive skill levels between periods t and $t - 1$ for child i , $\frac{\sum_{j \in P_i} \Delta y_j}{n_i}$ denotes endogenous peer effects, and $\Delta x_i = x_{i,t} - x_{i,t-1}$ denotes the change in child i 's own characteristics, including parental investment and household characteristics, while $\frac{\sum_{j \in P_i} \Delta x_{j,t}}{n_i}$ denotes the change in child i 's peers' characteristics between t and $t - 1$. Differencing v_{it} means that we eliminate θ_i , ϕ_C , and φ_P from Equation (5).

Table 1 shows that although we are able to difference out all the child-, household-, village, and peer-group-specific effects, correlated effects may still persist if there are

common peer-group or community related time-varying unobservables that affect both the child's as well as her peer group's outcome. For instance, it is possible that more schools were constructed between the two time periods in a particular village, causing growth in education achievement for all children in that village. We address such village-level changes to education by utilizing information on the older siblings of each child. We construct a quasi-cohort data set by pooling information on all the older siblings of each child in each community. We restrict the sample to those children who are up to 2 years older than the reference child in both years. To the extent that school enrollment and continuation is a proxy for student achievement, we calculate the average (highest) grade reached by this subset of older children for each community for both time periods. To capture any time-varying community-level effects that could have a direct impact on the education/schooling of children, we include this variable (in first difference) in our specification. Our objective is to see whether peer effects still continue to hold even after conditioning on these time-varying, community-level effects. This is incorporated as

$$\Delta y_i = \beta \frac{\sum_{j \in P_i} \Delta y_j}{n_i} + \gamma \Delta x_i + \delta \frac{\sum_{j \in P_i} \Delta x_j}{n_i} + \theta \Delta \bar{e}_C + \Delta v_i, \tag{6}$$

where \bar{e}_C represents the change in average educational attainment of older children in the community where the assumption is that $\theta \Delta \bar{e}_C = \phi_{Ct}$. Obviously ϕ_{Ct} only captures schooling-related time-varying common unobservables. That is, while this approach captures to some extent time-varying community-level correlated effects, the required assumption is that the relevant effects are limited to community-wide changes in schooling.

Second, following BDF, we apply a within transformation at the community level to account for correlated effects. In our sample, all communities are maximally connected, that is, links across peer groups form a community-wide network. Therefore, our within transformation absorbs any community-level unobservables. In the case of our community-based peer-group specification, the within transformation applies to the sampling cluster, where a sampling cluster is defined as a collection of communities that are jointly administered at the council level. In both cases, the within transformation is obtained by premultiplying the structural specification of Equation (2) with $(\mathbf{I} - \mathbf{W})$:

$$(\mathbf{I} - \mathbf{W})\mathbf{y}_t = \beta(\mathbf{I} - \mathbf{W})\mathbf{W}\mathbf{y}_t + \gamma(\mathbf{I} - \mathbf{W})\mathbf{x}_t + \delta(\mathbf{I} - \mathbf{W})\mathbf{W}\mathbf{x}_t + (\mathbf{I} - \mathbf{W})\mathbf{v}_t. \tag{7}$$

The reduced form is then given by

$$(\mathbf{I} - \mathbf{W})\mathbf{y}_t = (\mathbf{I} - \beta\mathbf{W})^{-1}(\gamma\mathbf{I} + \delta\mathbf{W})(\mathbf{I} - \mathbf{W})\mathbf{x}_t + (\mathbf{I} - \beta\mathbf{W})^{-1}(\mathbf{I} - \mathbf{W})\mathbf{v}_t, \tag{8}$$

where β and δ are still identified as in (3) if \mathbf{I} , \mathbf{W} , and \mathbf{W}^2 are linearly independent (provided $\beta\gamma + \delta \neq 0$).

Third, we implement an instrumental variable (IV) approach. We have data on household-specific idiosyncratic shocks (see Section 4.4), which we use as instruments for peer-group cognitive outcomes. We use average peer-group child-/household-specific idiosyncratic shocks to instrument for average peer-group cognitive achieve-

ment. The instrument works well under two conditions. First, idiosyncratic shocks should be significantly correlated with child cognitive achievement—as a result, average peer idiosyncratic shocks should affect average peer achievement. Second, the idiosyncratic shocks should be child/household specific and should not contain any information about the cognitive achievement of other children, even those located in the same neighborhood. To this extent, we should not find any effect of average peer idiosyncratic shocks on the target child's cognitive achievement after conditioning on own child idiosyncratic shocks.¹⁶ This IV approach provides an alternative and intuitive way to gauge the effect of correlated effects on our peer effect estimates.

2.3 Misspecification of the peer network structure

Our identification approach relies on the assumption that social interaction is mediated by geographical proximity: children who are located geographically close to each other within villages are more likely to interact than children who are geographically afar. Our empirical approach requires us to fully specify the structure of the social interaction between children. An obvious concern with this approach, which we discuss in this section, is the potential misspecification of the peer network structure, that is, the spatial weight matrix. We distinguish in our discussion between two related issues that may lead to misspecification: (i) data missing at random due to random sampling and (ii) the unknown population network structure.

2.3.1 Sampling A widespread problem associated with the empirics of social network data is that of mismeasurement due to sampling. Sampled data on networks are generally obtained by enumerating links among the sample of individuals who are selected from the population either (i) based on the realization of the dependent variable or a variable correlated with the dependent variable or (ii) through randomization. This would lead to measurement error if there are certain units that exist in the population that are not represented in the sample, but that are connected with some units included in the sample.

Marsden and Hurlbert (1987) discussed the bias that arises when individuals select into the sample based on their realization of the dependent variable or a variable correlated with the dependent variable. The resulting bias is well known in the literature on sample selection (Heckman (1979)). In contrast, when nodes are randomly assigned into the sample, a common assumption (e.g., Hoff (2009), Taskar, Wong, Abbeel, and Koller (2003)) in the estimation of network properties on sampled data is that the missing information on link presence/absence is missing at random. We consider spatial networks where the underlying network structure is based on nearest spatial neighbors. In our data, children were sampled randomly and were not selected into the sample based on the geographical location of the household they belong to within villages. Hence, it is reasonable to assume that the network data on missing links is missing completely at

¹⁶We test this by including peer idiosyncratic shocks in our base specification and find that it has no significant effect on child cognitive achievement conditional on a child's own idiosyncratic and covariate shocks. Subsequently, we show that the instrument is also highly informative as own idiosyncratic shocks are negatively correlated in a statistically significant way with a child's own achievement gains.

random because the units in the analysis were sampled at random from the population. In other words, the probability of observing a missing link depends only on the probability of two units i and j being observed, which is assumed to be random due to sampling. However, Chandrasekhar and Lewis (2012) showed that bias can still arise because sampling may induce networks to be misspecified. This resulting problem is similar to the problem of the unknown population network structure and, therefore, is discussed below.

2.3.2 Misspecification of network structure The definition of interactions between children is based on geographical proximity between children. This is an assumption required in the absence of actual information on interaction patterns between children within villages.¹⁷ This implies that our network structure might only imperfectly approximate the true underlying social network structure. As mentioned above, random sampling may also cause misspecification of networks (Chandrasekhar and Lewis (2012)).

Despite the importance of this issue for identification, the spatial econometrics literature provides hardly any guidance on the magnitude and direction of the bias due to misspecification of the spatial weights matrix and the circumstances under which it is most likely to affect the estimates. Chandrasekhar and Lewis (2012) showed analytically that misspecification implies that the instrument set $[\mathbf{x}, \mathbf{W}\mathbf{x}, \mathbf{W}^2\mathbf{x}]$ is no longer valid because the instrument is based on powers of the misspecified spatial weight matrix. This induces measurement error in the instrument that is by default correlated with the measurement error in $\mathbf{W}\mathbf{y}$.

Páez, Scott, and Volz (2008) used Monte Carlo exercises where they vary the level of spatial autocorrelation and network topology to analyze bias in a SAR model from underspecification of the adjacency matrix (i.e., assuming that a given node's degree is smaller than in the true model) and overspecification (i.e., assuming that a given node's degree is larger than in the true model). For the SAR specification, Páez, Scott, and Volz (2008) found that bias from under-specification is particularly severe when average degree and/or clustering in a network is low (i.e., the spatial weight matrix is sparse) and true underlying spatial autocorrelation is high. Overspecification results in pronounced bias when average degree is low but clustering is high, which means that in networks where there are connected components with few links between components, adding false links results in particularly severe bias. Lee (2009) derived theoretically the bias arising from misspecification of the adjacency matrix in a model with spatial lags as independent variables and provided Monte Carlo results for the SAR model, showing that a misspecified spatial weight matrix causes bias in both maximum likelihood and

¹⁷Actual information on individuals' networks are rare. Some exceptions are Conley and Udry (2010), who have detailed data on self-reported communication networks of farmers in Ghana, and the Add Health database (see Lin (2010) for a description) that incorporates information on friendship links. However, even when data on self-reported networks are available, the resulting network structure might still be misspecified due to perception bias and other potential reasons for individuals (un)intentionally to misreport their social networks. Examples of other studies that rely on geographical proximity to define social interaction in the absence of actual information on individuals' networks include Conley and Topa (2007) and Bayer, Ross, and Topa (2008).

two-stage least squares based estimation. His results suggest that the bias from under-specification points downward, whereas in the case of overspecification, estimates are upward biased. Generally, Lee found bias from overspecification to be lower than bias from underspecification.

To investigate potential implications for our results, we report results from several Monte Carlo experiments in Section 5.3. We investigate the potential bias due to sampling in Section 5.3.1. In Section 5.3.2, we report results from a simulation in which we vary the network topology to assess the bias that arises from the unknown population network structure. We also allow the degree, that is, the number of nearest neighbors, to vary and allow for children to have geographically distant friends (see Section 5.3.3). In our Monte Carlo exercises, we allow for both over- and underspecification of the network as well as a mixture of both. We simulate random deviations from the true underlying network and do not restrict these simulated networks to be above or below the hypothesized population network. This implies that any random draw from the set of simulated networks could pick up either an overspecification or an underspecification.

2.4 Peer effects and insurance against shocks

We are also interested in testing whether peer effects can provide insurance against adverse shocks to skill acquisition. In the given context, we have a rather informal mode of insurance in mind. Children may intuitively rely more on their peers when their own household is affected by an adverse shock, rather than benefit from support mechanisms involving direct transfers. A large literature in psychology (see, for example, Laible, Carlo, and Raffaelli (2000), Hartup (1996)) has documented evidence on peer support among children, particularly in situations when they face risk. More recently, De Giorgi and Pellizzari (forthcoming) proposed and derived conditions for the insurance mechanism to hold in the context of student achievement within classrooms.

We test for insurance by considering risk-sharing as a possible mechanism that gives rise to peer effects. We show in Appendix A.1, using a formal model, how risk-sharing can occur among children interacting within a peer network. In our model, we assume that children are risk-averse and derive positive utility from their cognitive achievement y_i . Achievement is specified as an output of the education production function in which a child's aggregate endowment serves as input. These endowments are subject to fluctuations that depend on the state of nature. Under these conditions, full risk-sharing would imply that a social planner (for example) maximizes the weighted sum of expected utilities (up to their years of schooling) of the $n_i + 1$ children in the network, subject to the network-aggregate endowment constraint by allocating a Pareto efficient weight ω_i to each child. The full risk-sharing implications can be tested empirically because changes in each child's cognitive achievement level are determined by changes in aggregate achievement rather than by idiosyncratic shocks. As shown in Appendix A.1, this allows us to derive the empirical specification to test for the presence of insurance, accounting for idiosyncratic shocks in period t ,

$$\Delta y_i = \beta \frac{\sum_{j \in P_i} \Delta y_j}{n_i} + \gamma \Delta x_i + \delta \frac{\sum_{j \in P_i} \Delta x_j}{n_i} + \eta s_i + v_i, \quad (9)$$

where s_i is a dummy variable that indicates whether the child’s household has experienced an idiosyncratic shock between the two time periods (see Section 4.4). The test for insurance is applied by introducing the idiosyncratic shock as an overspecification of the model, that is, $\hat{\eta} = 0$ is interpreted as evidence in favor of peer-group insurance. The test is, therefore, whether conditional on peer effects, the idiosyncratic shock is orthogonal to our measure of cognitive skill formation. We can interpret this as evidence for peer insurance, because we show in Appendix A.1 that if peer effects act as an insurance mechanism, idiosyncratic household-specific shocks should not affect a child’s skill growth once skill growth across the child’s peers is accounted for.

3. ESTIMATION

While the S2SLS estimator proposed by Kelejian and Prucha (1998) allows consistent estimation of the coefficient associated with the spatial autoregressive term, standard errors require the residuals to be independent and identically distributed (i.i.d.) and homoscedastic. More recently, Kelejian and Prucha (2007) suggested a spatial nonparametric heteroscedasticity and autocorrelation consistent (SHAC) estimator that accommodates heteroscedasticity of unknown form and spatial autocorrelation in the residuals. Contrary to Conley (1999), the proposed estimator does not require the spatial process to be stationary, which is essential in the context of SAR models as specified in Equation (5) above.¹⁸

To see how the S2SLS estimator is implemented, rewrite Equation (5) in a drastically simplified form, omitting spatial lags in covariates \mathbf{x} for expositional simplicity, as

$$\mathbf{y} = \mathbf{Z}\zeta + \mathbf{u}, \tag{10}$$

where $\mathbf{Z} = [\mathbf{W}\mathbf{y}, \mathbf{x}]$ and $\zeta = \begin{pmatrix} \beta \\ \gamma \end{pmatrix}$. We use the matrix $\mathbf{H} = [\mathbf{x}, \mathbf{W}\mathbf{x}, \mathbf{W}^2\mathbf{x}]$ to instrument for $\mathbf{W}\mathbf{y}$. The estimator is then given by

$$\hat{\zeta} = (\hat{\mathbf{Z}}\mathbf{Z})^{-1}\hat{\mathbf{Z}}'\mathbf{y}, \tag{11}$$

where $\hat{\mathbf{Z}} = \mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\mathbf{Z}$. Denote the residuals obtained from this S2SLS estimator as \mathbf{u}_n . Kelejian and Prucha (2007) assumed that the disturbance process can be described as

$$\mathbf{u}_n = \mathbf{R}_n\boldsymbol{\epsilon}_n, \tag{12}$$

where $\boldsymbol{\epsilon}_n$ is an $n \times 1$ vector of innovations and \mathbf{R}_n is an $n \times n$ nonstochastic matrix with unknown elements. Note that vectors and matrices are denoted by n as they may depend on the sample size. Kelejian and Prucha (2007) assumed that \mathbf{R}_n is nonsingular and the row and column sums of \mathbf{R}_n and \mathbf{R}_n^{-1} are bounded uniformly in absolute value by some constant c_R , where $0 < c_R < \infty$. The corresponding variance–covariance (VC) matrix is defined as

$$\psi_{ij,n} = n^{-1}\mathbf{H}'_n\boldsymbol{\Sigma}_n\mathbf{H}_n, \tag{13}$$

¹⁸As noted by Kelejian and Prucha (2007), it suffices that cross-sectional units have different numbers of neighbors to obtain a nonstationary spatial process. Considering the discussion in Lee (2007), this may even be a desired feature in the data to be able to disentangle endogenous and contextual effects.

where \mathbf{H}_n is a $n \times p_h$ nonstochastic matrix of instruments defined above and $\boldsymbol{\Sigma}_n = \mathbf{R}_n \mathbf{R}'_n$ denotes the VC matrix of \mathbf{u}_n , where row and column sums of $\boldsymbol{\Sigma}_n$ are also uniformly bounded. Let $d_{ij,n}^*$ represent the distance between the observations i and j ; d_n is the bandwidth, which is assumed variable and is set to the maximum distance for each observation. Spatial dependence is introduced through a kernel function that is a real continuous and symmetric function that defines weights for covariances as $K(d_{ij,n}^*/d_n)$ with $d_{ij,n} \geq 0$ and bandwidth $d_n > 0$. Whenever $d_{ij,n} \geq d(n)$, the kernel is equal to zero. We choose a plug-in bandwidth based on the distance to child i 's K nearest neighbors as discussed above.¹⁹ Since the choice of the kernel is usually of little importance in the implementation of nonparametric estimators, we choose the standard Epanechnikov kernel. Then the (r, s) th element of the true VC matrix Ψ_n of the SHAC estimator is given by

$$\hat{\psi}_{rs,n} = n^{-1} \sum_{i=1}^n \sum_{j=1}^n h_{ir,n} h_{js,n} \hat{u}_{i,n} \hat{u}_{j,n} K(d_{ij,n}^*/d_n). \quad (14)$$

Our theoretical justification for the specification of the model in Equation (5) allows us to treat any potential spatial autocorrelation in the error term as nuisance. Hence, we do not need to impose any particular functional form on the spatial autocorrelation in the error term, which corroborates our choice of the SHAC estimator.²⁰

In theory, all successive lags of the independent variables $[\mathbf{W}\mathbf{x}_t, \mathbf{W}^2\mathbf{x}_t]$ serve as valid instruments. However, in practice, it remains unclear whether each one of them or indeed all of them jointly are significantly correlated with the endogenous variable. This means that we potentially face a weak instrument problem that arises in spatial settings largely because spatially lagged variables, which are used as instruments, tend to be highly correlated with each other. This induces a certain degree of multi-collinearity among instruments, resulting in a jointly weak instrument set (Gibbons and Overman (2012)). To address this issue, we also estimate Equation (2), by selecting an optimal instrument set from our given extensive list of instruments rather than using the entire instrument vectors. We apply LASSO (least absolute shrinkage and selection operator) techniques (Caner and Fan (2010), Belloni, Chernozhukov, and Hansen (2010)) to choose and construct the optimal instrument set. We use a modified LASSO method proposed by Efron, Hastie, Johnstone, and Tibshirani (2004) that involves using the least angle regression (LARS) model selection algorithm to choose instruments for the first-stage regressions. The method implements a computationally efficient LASSO that constrains the sum of the absolute regression coefficients such that only a subset of the covariates have nonzero values. Using this method, we substantially reduce the dimension of our instruments, choosing only those variables that are strongly correlated with the endogenous regressors.

¹⁹Our choice of the plug-in bandwidth seems appropriate given our modeling choice of peer interaction based on a child's K nearest neighbors. As a potential alternative, see Lambert, Florax, and Cho (2008) for a discussion of data-driven bandwidth selection in the context of the SHAC estimator.

²⁰To implement the SHAC estimator, we use the *spdep* (Bivand et al. (2011)) and *sphet* (Piras (2010)) packages in R. To implement the LASSO, we use the LARS package in R.

4. DATA

We use data from the India part of the YL project. The YL is a long-term study of childhood poverty being carried out in Ethiopia, India (in the state of Andhra Pradesh), Peru, and Vietnam. The survey consists of tracking two cohorts of children over a 15-year period. Currently data from two rounds of data collection are available. In round 1, 2000 children aged around 1 (the “younger” cohort) and 1000 children aged around 8 (the “older” cohort) were surveyed in 2002. Following up, round 2 involved tracking the same children and surveying them in 2006 at ages 5 and 12, respectively.

The sample of children is representative of the three regions of Andhra Pradesh: Rayalseema, Coastal Andhra, and Telangana. The sampling process was fourfold. First, six districts were selected based on the classification of poor/nonpoor given by their relative levels of development. In the second stage, 20 sampling clusters (corresponding to the Indian administrative unit *mandal*) within these districts were identified based on the same classification. Subsequently, one community was randomly selected from approximately four to five communities that comprise a sampling cluster, and households within the selected community were also selected randomly. Finally, the questionnaires were administered to around 100 1-year-old and 50 8-year-old children in these communities. Data were collected through household, caregiver, child, and community questionnaires.

In [Helmert and Patnam \(2011\)](#), we analyzed the formation of both cognitive and non-cognitive skills, paying particular attention to self-productivity (any effect of past periods’ cognitive/noncognitive skills on current period’s cognitive/noncognitive skills) as well as cross-productivity effects (any effect of past periods’ cognitive/noncognitive skills on current period noncognitive/cognitive skills). We have found statistically and economically significant evidence for self-productivity for cognitive skills and cross-productivity effects of cognitive skills on noncognitive skills. However, we have not found any evidence of self-productivity for noncognitive skills or of noncognitive skills affecting cognitive skills. We therefore focus our analysis of spatial peer effects on the formation of cognitive skills. Moreover, we omit the use of many noncognitive inputs in the production function that the data allow us to use precisely for this reason. Since we are interested in the determinants of the evolution of cognitive skills over time, we can only use the older cohort of children because for the younger cohort there is no information on children’s cognitive skill levels at age 1. This means we only use information for the older cohort of children to analyze the determinants of their cognitive skill formation between ages 8 and 12. For a more detailed description of the data set, see [Helmert and Patnam \(2011\)](#). Table 2 shows some summary statistics for the variables used in our analysis.

4.1 Location and peer effects

To construct geographical distances between households/children, we collated various geography variables from two GIS files: the Taluk map of Andhra Pradesh, which provides digitized Taluk (administrative boundary) polygons, and a household location map that contains, as a point feature class, detailed GPS locations of every household/child in the YL data set. The latter was overlaid with the Taluk map to identify

TABLE 2. Summary statistics.

| Variable | No. Obs. | Median | Mean | Std. Dev. |
|--|----------|--------|--------|-----------|
| CH writing score (age 8) | 731 | 2 | 2.098 | 0.691 |
| CH writing score (age 12) | 731 | 3 | 2.640 | 0.580 |
| CH reading score (age 8) | 731 | 3 | 3.046 | 1.047 |
| CH reading score (age 12) | 731 | 4 | 3.660 | 0.780 |
| Δ CH writing score | 731 | 1 | 0.541 | 0.873 |
| Δ CH reading score | 731 | 0 | 0.621 | 1.114 |
| Δ CH weight ^a | 731 | 10.85 | 12.938 | 10.95 |
| Δ CH height ^b | 731 | 23.1 | 22.602 | 11.23 |
| Δ CH siblings | 731 | 0 | 0 | 0.546 |
| Δ CH years of schooling | 731 | 4 | 3.652 | 1.095 |
| Δ CH work | 731 | 0 | -0.136 | 0.583 |
| Δ CH public school ^c | 731 | 0 | -0.123 | 0.511 |
| Δ HH size | 731 | 0 | -0.350 | 1.551 |
| Δ HH wealth | 731 | 0.429 | 0.053 | 0.012 |
| Δ HH covariate shock | 731 | 0 | 0.440 | 0.496 |
| Δ HH idiosyncratic shock | 731 | 0 | 0.426 | 0.494 |
| Δ VIL years of schooling | 731 | 3.13 | 2.970 | 0.582 |
| Urban area | 731 | 0 | 0.265 | 0.441 |
| Coastal Andhra | 731 | 0 | 0.184 | 0.388 |

Notes: CH, child; HH, household; VIL, village.

^aChild weight is weight for age z -score.

^bChild height is height for age z -score.

^cPublic school is defined as [(yes = 1) at age 12 - (yes = 1) at age 8].

village-level clusters for households. This gives us longitude and latitude information on the location of households and children, and thus allows us to compute the Euclidean distance between households. The distance is used to determine a household's nearest neighbors, which are used to measure spatial peer effects. Note that we have GPS locations of only 750 out of the 1000 sampled households. To assert that there are no systematic differences in characteristics of households/children for which GPS information is and is not available, we implement a number of tests. First, a Kolmogorov–Smirnov test does not reject equality of the outcome distributions with a p -value of 0.322. We also conduct t -tests for differences in means between the sample children and those for whom we have missing GPS information over a range of observable covariates that are included in our empirical specification. The results are reported in Table 3. Barring a few variables, we find no significant differences in both child and household demographic characteristics.

4.2 Cognitive skills

In principle, cognitive skills are unobserved. To proxy them, we use observed measures. Since we use a specification in first differences, we need the same measures at ages 8 and 12. This restricts our choice of possible measures because the survey questionnaires differed between round 1 and round 2. The only measures for cognitive skills that are avail-

TABLE 3. Missing data characteristics.

| Variable | Mean Sample | Mean Missing | Difference | Std. Error (of Difference) |
|--|-------------|--------------|---------------|-------------------------------|
| Δ CH weight ^a | 12.866 | 12.922 | -0.056 | 0.715 |
| Δ CH height ^b | 22.568 | 23.640 | -1.071 | 0.731 |
| Δ CH siblings | 0.0026 | 0.078 | -0.075 | 0.039 |
| Δ CH years of schooling | 3.654 | 3.654 | 0.0004 | 0.079 |
| Δ CH work | -0.1370 | -0.320 | 0.183 | 0.041 |
| Δ CH public school ^c | -0.123 | -0.160 | 0.036 | 0.036 |
| Δ HH size | -0.362 | -0.304 | -0.057 | 0.115 |
| Δ HH wealth | 0.053 | 0.057 | -0.004 | 0.009 |
| Δ HH covariate shock | 0.436 | 0.176 | 0.259 | 0.034 |
| Δ HH idiosyncratic shock | 0.421 | 0.267 | 0.154 | 0.035 |
| Urban area | 0.192 | 0.423 | -0.231 | 0.031 |
| CH male | 0.501 | 0.526 | -0.024 | 0.036 |

Notes: Differences that are significant at 5% are indicated in bold. CH, child; HH, household; VIL, village.

^aChild weight is weight for age z-score.

^bChild height is height for age z-score.

^cPublic school is defined as [(yes = 1) at age 12 - (yes = 1) at age 8].

able for children at ages 8 and 12 are reading and writing test scores.²¹ These tests assess mostly a child's general intelligence and her ability to apply acquired knowledge and skills. These skills are distinct from noncognitive skills, which aim to measure a child's personality traits (Borghans, Duckworth, Heckman, and ter Weel (2008)). Our specification in first-differences accounts for unobserved initial conditions. Such initial conditions due to a child's unobserved endowment are assumed to exert a constant effect over time on the formation of cognitive skills, which means that by taking first differences, we eliminate them from the specification.²² We focus on the change in reading and writing scores as an indicator of child's cognitive development. There are two reasons for doing so. First, Cueto et al. (2009) found evidence that the change in skill development of mathematics for children in India is quite negligible. These authors assessed the technical validity of many of the academic and psychometric tests administered in the YL data and found that by round 2 of data collection, most children in India could already do math and improved only in the other areas (writing and reading). Those children who

²¹The reading item required children to read three letters, one word, and one sentence. The reading item was scored as follows: 1 point if children could read the sentence, 0.66 point if they could read the word, 0.33 point if they could read the letters, and 0 points if they could not read anything or did not respond. The writing item asked the children to write a simple sentence that was spoken out loud by the examiner. This item was scored as follows: 1 point if children could write the sentence without difficulty or errors, 0.5 point if they could write with difficulty or errors, and 0 points if they could not write anything or did not respond (Cueto, Leon, Guerrero, and Munoz (2009)).

²²We recognize that this is a strong assumption as the expression of a child's initial endowment over time may vary as a function of a child's environment. To the degree that changes in child anthropometrics, specifically changes in child height and weight, proxy for time-varying effects of unobserved initial endowments (see Weedon et al. (2007)), this problem is mitigated as we include these variables in the conditioning set (see Section 4.3).

had picked up the mathematical skill early on (a high 87%) continued to do well even after 4 years. Second, we find a huge improvement in writing skills of children between the two rounds. The percentage of children who were able to write without difficulty improved from 51% to 69% in 4 years. Table 2 shows the summary statistics of reading and writing scores at ages 8 and 12 as well as in first differences.

4.3 *Inputs*

Since we implement a specification in first differences, only inputs into the production of cognitive skills that change over time can be included in the conditioning set of variables. This rules out the use of variables such as a child's gender, caste, birth order, and so forth. Yet we also split the sample by gender to investigate differences in the role peers may play to insure children against adverse idiosyncratic shocks.

The set of potential input variables is further restricted because of the differences in the design of the questionnaires in the two surveys. We use child anthropometries, that is, child weight and height, as proxies for child health and unobserved initial conditions whose expression varies over time. In addition, we include the change in the number of a child's siblings. This variable captures changes in parental input as well as potential effects that arise from within-family interaction and household size. In addition, we include a measure of how much time the child spends working. Given that a child divides her time between school, work, and leisure, the change in time spent working captures changes in the time spent at school and spare-time activities. We therefore do not account separately for a change in schooling. As a measure for the quality of a child's schooling, we include a variable that indicates whether the child moved from a public to a private school between ages 8 and 12. As a direct measure of the resources available to a child, we include the change in household assets.²³ Moreover, we include a dummy variable that indicates whether a child's household is located in a (semi-)urban area and a dummy variable for whether the household is located in the coastal area of Andhra Pradesh. These indicator variables capture time-varying location-specific effects.

4.4 *Shocks*

The survey captured detailed, separate information on various shocks faced by the child's household, including economic, climatic, health, and other miscellaneous shocks. The detailed available information allows us to divide and group these shocks into (a) idiosyncratic shocks and (b) covariate shocks. Idiosyncratic shocks include the following events reported by the primary caregiver of a child: sudden shortfall of food, loss of livestock, death or serious illness of household members, job loss in the household, and whether the household was subject to crime, including theft and robberies. Here, we take care to include only those shocks that are specific to any given household and are not correlated with the occurrence of the same shock in other households, such as, for example, the death of a household member (this is validated empirically in

²³We use a wealth index that consists of three components: housing quality, consumer durables such as a refrigerator or a telephone, and services such as electricity or toilets available to the household.

the results section later on). Covariate shocks include natural disasters and calamities such as droughts or floods and crop failures that affect the whole community. We create two shock variables that are dummy variables that indicate whether the child’s household had experienced an idiosyncratic or covariate shock, respectively, between ages 8 and 12.

5. RESULTS

We first provide descriptive evidence for spatial peer effects on child cognitive achievement gains. Figures 2 and 3 present nonparametric plots of a child’s first-differenced reading and writing scores against the average first-differenced scores of her five nearest neighbors. Both graphs provide descriptive evidence that peer effects matter, as they show that a child’s own score gain is increasing in her peer’s performance.

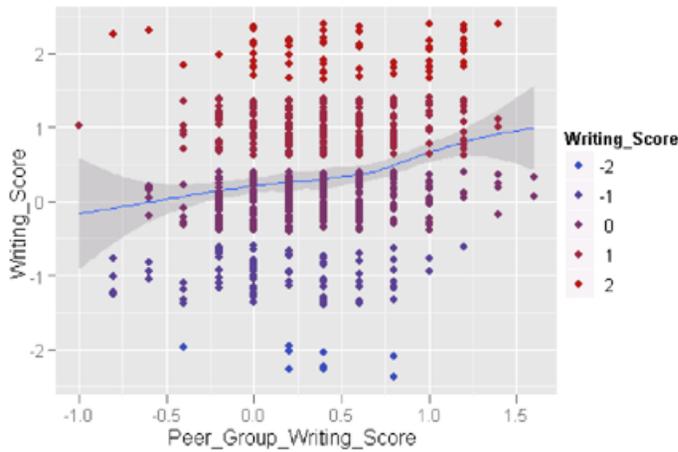


FIGURE 2. Nonparametric plots of Δ own writing score versus Δ peer writing score.

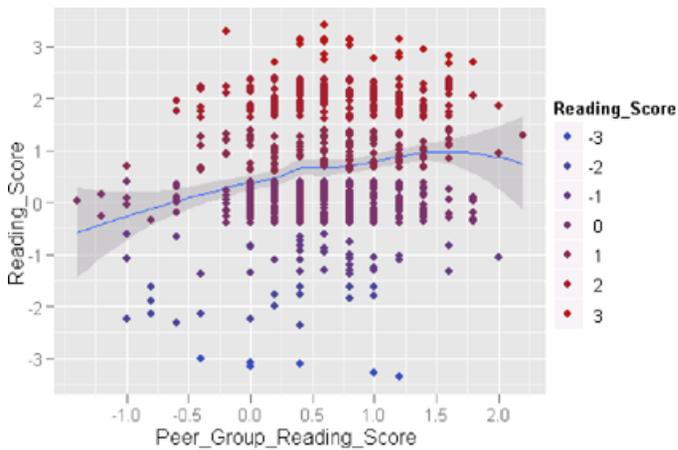


FIGURE 3. Nonparametric plots of Δ own reading score versus Δ peer reading score.

5.1 Peer effects

Table 4 shows the results for writing skills from estimating Equation (5) using OLS and the estimation procedure described in Section 3.

Columns 1 and 2 show OLS results not accounting for potential heteroscedasticity and spatial autocorrelation in the error term. Column 1 does not allow for peer effects; the results show that a number of variables are statistically significant: the change in child height and weight, the change in the time worked, the public school variable, and the indicator for whether the household is located in an urban area. Since the public school variable is equal to 1 if the child switched from private school to public school between ages 8 and 12, 0 if no switch occurred, and -1 if the child switched in the other direction, the positive coefficient suggests that children who switched from private to public school between ages 8 and 12 experienced on average a higher achievement gain.²⁴ Since we use first-differenced variables, the location-specific variables urban and coastal area capture time-variant location-specific unobservables. The fact that the urban indicator variable is statistically significant suggests that this is a relevant concern despite using first differences, that is, that different locations are on different trajectories in terms of child skill development. When we account for peer effects in column 2, we note a positive and statistically significant coefficient associated with endogenous peer effects. An increase of 1 standard deviation of the endogenous effects leads to an increase of a little less than 0.2 standard deviations in the growth of writing skills. The coefficient on the endogenous peer effects as estimated by the OLS specification is not close to 1 in the typical linear-in-means sense because of the nonlinearities induced by the network structure. However, the OLS estimates still remain biased due to simultaneity issues and other unobservable factors. The other covariates that were statistically significant in column 1 remain so.

To test for the presence of spatial dependence in the error term, we apply a Lagrange multiplier (LM) test to the OLS residuals in columns 1 and 2. The null hypothesis is the absence of spatial autocorrelation, which is tested against the presence of autocorrelation captured by a spatial error component (Anselin and Hudak (1992)). The p -value reported in column 2 suggests the presence of spatial dependence in the residuals when

²⁴In our sample, 8% of children move from private to public schools, whereas 20% move from public to private schools. Children who move from private to public schools have on average a higher cognitive skill level at age 8 than children who move from public to private schools. Importantly, children who switch from public to private schools have the same average cognitive skill level as children who stayed in public school, which mitigates concerns over endogeneity of this variable. Switching schools, even within a given school year, is a frequent and common occurrence in India. Woodhead, Frost, and James (2013) documented a general trend in the YL sample toward children switching from public to private schools and, more importantly, found that there is widespread heterogeneity in school quality even within school types. Our result possibly indicates that private-to-public school switchers favor a change to a high(er) quality public school and, therefore, experience a gain in achievement, whereas public-to-private school switchers change to potentially low(er) quality private schools and thus see less progress in cognitive achievement. Ongoing research by Woodhead, Frost, and James (2013) using the YL data explores this possible explanation and analyzes in detail the issue of dynamic school mobility using data from a recently completed school survey on school quality and choices.

TABLE 4. Results for *five* nearest neighbors: writing.

| | Contextual Effects | | | | | | |
|--|---------------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|--------------------------|
| | OLS | | OLS Robust | SAR | SHAC | OLS Robust | SHAC |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| W Δ writing level | – | 0.380** (0.072) | 0.380** (0.076) | 0.840** (0.181) | 0.840** (0.151) | 0.309** (0.082) | 0.764* (0.337) |
| Δ CH weight ^a | –0.011** (0.003) | –0.010** (0.003) | –0.010** (0.003) | –0.009* (0.003) | –0.009* (0.004) | –0.008* (0.003) | –0.007+ (0.004) |
| Δ CH height ^b | –0.007* (0.003) | –0.008* (0.003) | –0.008* (0.003) | –0.009* (0.003) | –0.009* (0.003) | –0.010** (0.003) | –0.010** (0.003) |
| Δ CH siblings | 0.075 (0.057) | 0.068 (0.056) | 0.068 (0.055) | 0.058 (0.057) | 0.058 (0.062) | 0.071 (0.054) | 0.058 (0.061) |
| Δ CH work | –0.153** (0.054) | –0.100+ (0.054) | –0.100+ (0.052) | –0.035 (0.060) | –0.035 (0.062) | –0.081 (0.059) | –0.045 (0.070) |
| Δ public school | 0.156* (0.061) | 0.164** (0.060) | 0.164** (0.060) | 0.174** (0.062) | 0.174** (0.066) | 0.192** (0.064) | 0.204** (0.074) |
| Δ HH assets | 0.006 (0.242) | 0.021 (0.234) | 0.021 (0.250) | 0.040 (0.244) | 0.040 (0.258) | –0.024 (0.254) | –0.019 (0.268) |
| Δ HH shock (covariate) | 0.084 (0.072) | 0.052 (0.071) | 0.052 (0.073) | 0.014 (0.075) | 0.014 (0.085) | 0.027 (0.078) | –0.001 (0.091) |
| Urban area | –0.200* (0.087) | –0.157+ (0.086) | –0.157+ (0.074) | –0.104 (0.091) | –0.104 (0.064) | –0.054 (0.091) | –0.045 (0.072) |
| Coastal Andhra | 0.026 (0.076) | 0.021 (0.074) | 0.021 (0.076) | 0.013 (0.077) | 0.013 (0.058) | 0.106 (0.085) | –0.021 (0.099) |
| W Δ CH weight ^a | | | | | | –0.002 (0.007) | 0.002 (0.007) |
| W Δ CH height ^b | | | | | | 0.014+ (0.007) | 0.015** (0.005) |
| W Δ CH siblings | | | | | | –0.004 (0.118) | –0.081 (0.112) |
| W Δ CH work | | | | | | –0.204* (0.095) | –0.053 (0.148) |
| W Δ public school | | | | | | –0.209* (0.106) | –0.218* (0.098) |
| W Δ HH assets | | | | | | 0.152 (0.481) | 0.139 (0.411) |
| W Δ HH shock (covariate) | | | | | | 0.309* (0.142) | 0.138 (0.199) |
| Breusch–Pagan test ^c | 0.001 | 0.016 | | | | | |
| LM test ^d | 0.456 | 0.000 | | | | | |
| Observations | 731 | 731 | 731 | 731 | 731 | 731 | 731 |

Notes: CH, child; HH, household. Standard errors are given in parentheses (White standard errors for OLS robust). The superscript plus (+) indicates significance at 10%; * at 5%; ** at 1%.

^aChild weight is weight for age z-score.

^bChild height is height for age z-score.

^cp-values (H_0 : constant error variance).

^dp-values (H_0 : no spatial autocorrelation).

allowing for a spatial lag of the dependent variable. We apply the studentized Breusch–Pagan test to check for the presence of heteroscedasticity. The p -values reported in columns 1 and 2 show strong evidence for the presence of heteroscedasticity.

Column 3 accounts for heteroscedasticity by estimating White (1980) robust standard errors. Obviously, the point estimates remain unchanged. Also, the peer effect term remains statistically highly significant. Overall, we find that the standard errors are only marginally changed when computing White (1980) robust standard errors, which is surprising given the strong evidence for the presence of heteroscedasticity shown in columns 1 and 2.

Columns 4 and 5 report the results when using the S2SLS estimator, which accounts for endogenous effects through its instrumental variable approach. The estimates in column 4 assume homoscedasticity and absence of spatial autocorrelation in the residuals of Equation (5), but account for the endogeneity of the spatial lag variable. Hence, the results in column 4 show the bias when ignoring this endogeneity in columns 2 and 3. Column 5, in contrast, reports the results for the SHAC estimator, which is least restrictive in terms of assumptions imposed on the residuals of Equation (5), allowing for both heteroscedasticity and spatial autocorrelation, and is, therefore, our preferred estimator. The results in columns 4 and 5 show a large increase in the coefficient of peer effects. Now an increase of 1 standard deviation in a child's peers has more than twice the effect on the change in writing skills as it had when using OLS. The substantial downward bias found in the peer effects estimate could be explained, to an extent, by the presence of measurement error in own and peer test scores. With measurement error, OLS estimates suffer from attenuation (as in the standard case) and are only a fraction of what the true estimate is. The degree of attenuation is proportional to the signal to noise ratio (Ammermueller and Pischke (2009), Arcidiacono, Foster, and Kinsler (2012)). In our context, using the same data set, Helmers and Patnam (2011) derived and estimated the measurement error for the test scores of 8- and 12-year-old children. These authors estimated a measurement error of around 0.46 for writing scores, which means that if the true estimate is 0.8 (our IV estimate), then the OLS estimate should be a fraction 0.46 of this. Our estimated OLS estimates are around 0.36–0.38 and largely correspond to this level of attenuation. Our chosen instruments potentially alleviate bias from the measurement error under the plausible assumption that the exogenous characteristics of third degree neighbors are uncorrelated with the measurement error/noise.

Columns 6 and 7 report the results when also accounting for contextual effects. Column 6 reports OLS results, whereas column 7 reports the SHAC estimates.²⁵ Again, we note a substantial difference in the coefficients associated with endogenous effects between OLS and the spatial two-step estimator. The coefficient obtained for the S2SLS estimator suggests that a 1 standard deviation increase in peer skill growth increases writing skill formation by around 0.37 standard deviations, which is a sizeable effect. Among contextual effects, only the change in height of a child's peers as well as whether her peers have switched between public and private school are statistically significant.

²⁵We computed the corresponding variance inflation factors to investigate the potential presence of multi-collinearity, but there is no evidence for this.

Cooley (forthcoming) provided a detailed discussion on the specification and interpretation of contextual effects in the classroom/child learning context. She argued that when the child is able to choose her own effort with a view to increase outcomes, it is unclear whether increasing peer exogenous characteristics will have a positive or a negative effect. This is because higher values of peer exogenous characteristics might reduce own outcome values if there are positive spillovers from endogenous peer effects, which we condition on. For instance, consider a child whose writing scores are increasing in her peers' writing scores as well her own school choice (switching to public school). If all her peers decide to switch to public school and increase their effort, then, controlling for the child's own school switch and her peers' achievement levels, we would expect to see a decrease in own achievement levels. This is because the child will be inclined to reduce effort so as to take advantage of the positive peer spillovers. Overall, the precise interpretation of contextual effects having accounted for endogenous effects is at best ambiguous.

Table 5 reports the corresponding results for the change in reading scores. In column 1, we report again OLS results, ignoring both endogenous and contextual effects. Among the covariates, the writing scores, the public school dummy variable, and the indicator for whether the household is located in an urban area are statistically significant. In addition, also the variable that indicates whether a household is located in the coastal area is statistically significant. In column 2, we add endogenous peer effects. The coefficient is positive and statistically significant: a 1 standard deviation increase in a child's peers' skill growth is associated with an increase of a fifth of a standard deviation of the child's own skills. The LM test for spatial autocorrelation in the residuals is flatly rejected in column 2. Similarly, the Breusch–Pagan test strongly suggests a nonconstant variance of the residuals. Column 3 reports OLS results with robust standard errors where again the standard errors do not differ dramatically. In columns 4 and 5, we report the results from using the S2SLS estimator. As for writing scores, the coefficient associated with peer effects increased markedly. In column 5, for our preferred estimation method, the SHAC, a standard deviation of a child's peers' skill change increases her own skill growth by slightly more than a 0.5 standard deviation.²⁶ Columns 6 and 7 show the coefficients when contextual effects are included in the model. The coefficients for peer effects drop slightly for both estimators, OLS and SHAC, and we find only a child's peers' change in weight and the public school variable to be statistically significant among the contextual effects.

In the remainder of this section, we discuss a range of additional results that employ different modifications of the basic specification so as to investigate the importance of potentially omitted unobservables in driving our results. However, considering the great

²⁶The magnitude of the endogenous peer effect is greater than 1. In general, dynamic spatial models with autoregressive coefficients greater than 1 are not stable in the long run (Babcock, Engberg, and Greenbaum (2005)). More recently, Lee and Yu (2007) showed that the autoregressive coefficient can be close to 1 even in cross-sectional data if it captures dynamic effects via the spatial lag. These authors argued that if the true data generating process is a spatial dynamic panel data process with spatial cointegration, then cross-sectional estimates of the autoregressive coefficients can be close to 1.

TABLE 5. Results for *five* nearest neighbors: reading.

| | Contextual Effects | | | | | | |
|-----------------------------------|-------------------------------|------------------------------------|------------------------------------|------------------------------------|------------------------------------|------------------------------------|--|
| | OLS | | OLS Robust | SAR | SHAC | OLS Robust | SHAC |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| W Δ writing level | – | 0.392** (0.071) | 0.392** (0.076) | 1.111** (0.313) | 1.111** (0.290) | 0.365** (0.069) | 0.729⁺ (0.378) |
| Δ CH weight ^a | –0.007 (0.005) | –0.006 (0.005) | –0.006 (0.006) | –0.004 (0.005) | –0.004 (0.007) | –0.004 (0.006) | –0.005 (0.006) |
| Δ CH height ^b | –0.002 (0.005) | –0.004 (0.004) | –0.004 (0.006) | –0.005 (0.005) | –0.005 (0.007) | –0.004 (0.006) | –0.006 (0.007) |
| Δ CH siblings | –0.070 (0.076) | –0.082 (0.074) | –0.082 (0.071) | –0.103 (0.080) | –0.103 (0.075) | –0.077 (0.071) | –0.090 (0.069) |
| Δ CH work | –0.067 (0.072) | –0.044 (0.071) | –0.044 (0.076) | 0.003 (0.078) | 0.003 (0.081) | –0.024 (0.085) | –0.009 (0.093) |
| Δ public school | 0.145 ⁺ (0.082) | 0.160* (0.080) | 0.160 ⁺ (0.086) | 0.186* (0.086) | 0.186* (0.092) | 0.191* (0.088) | 0.208* (0.099) |
| Δ HH assets | –0.121 (0.323) | 0.001 (0.317) | 0.001 (0.322) | 0.225 (0.353) | 0.225 (0.354) | –0.009 (0.327) | 0.087 (0.332) |
| Δ HH shock (covariate) | 0.024 (0.097) | –0.009 (0.095) | –0.009 (0.100) | –0.070 (0.105) | –0.070 (0.123) | –0.046 (0.109) | –0.076 (0.128) |
| Urban area | –0.241* (0.117) | –0.201 ⁺ (0.115) | –0.201 ⁺ (0.109) | –0.127 (0.130) | –0.127 (0.097) | –0.052 (0.129) | –0.065 (0.114) |
| Coastal Andhra | –0.239* (0.101) | –0.168 ⁺ (0.100) | –0.168 ⁺ (0.098) | –0.039 (0.121) | –0.039 (0.089) | –0.047 (0.111) | –0.043 (0.079) |
| W Δ CH weight ^a | | | | | | –0.007 (0.010) | –0.012 ⁺ (0.010) |
| W Δ CH height ^b | | | | | | 0.001 (0.010) | –0.004 (0.010) |
| W Δ CH siblings | | | | | | 0.096 (0.164) | 0.098 (0.115) |
| W Δ CH work | | | | | | –0.220* (0.127) | –0.148 (0.143) |
| W Δ public school | | | | | | –0.285* (0.138) | –0.276* (0.136) |
| W Δ HH assets | | | | | | –0.496 (0.632) | –0.394 (0.530) |
| W Δ HH shock (covariate) | | | | | | 0.374 ⁺ (0.191) | 0.218 (0.241) |
| Breusch–Pagan test ^c | 0.000 | 0.000 | | | | | |
| LM test ^d | 0.460 | 0.000 | | | | | |
| Observations | 731 | 731 | 731 | 731 | 731 | 731 | 731 |

Notes: CH, child; HH, household. Standard errors are given in parentheses (White standard errors for OLS robust). The superscript plus (⁺) indicates significance at 10%; * at 5%; ** at 1%.

^aChild weight is weight for age z -score.

^bChild height is height for age z -score.

^c p -values (H_0 : constant error variance).

^d p -values (H_0 : no spatial autocorrelation).

similarity in the results found in Tables 4 and 5, we limit this discussion to the results obtained for writing test scores.²⁷

Table 6 reports results when accounting for potentially omitted unobservables by augmenting the specification with the village-level schooling variable constructed using children's older siblings or the within-network transformation as described in Section 2.2. In addition, the table contains results for an IV estimation in which we instrument endogenous peer effects with idiosyncratic shocks.

Column 1 shows SAR results when allowing for peer effects and controlling for time-varying unobservables that are correlated with peer effects, such as unobserved changes in the availability or quality of schools. We note that the peer effect coefficient remains nearly unchanged compared to the results reported in column 4 of Table 4. The schooling of older siblings averaged at the village level is statistically not significantly different from zero, providing no evidence for a bias of the coefficient associated with endogenous effects due to time-varying schooling-related unobservables. In column 2, we also explore the possibility of nonlinear effects due to such correlated effects by interacting the schooling of older siblings with the peer effect variables, but find no significant effect of the interaction terms.

Column 3 reports the results when using a within-transformation that accounts for unobservables at the network level.²⁸ The magnitude of the peer effect coefficient increases slightly relative to columns 1 and 2. However, interpretation of the effect is difficult because under the within-transformation peer effects represent the deviation of a child's gain from the average gain of the community-level network. Columns 4 and 5 report results for the IV approach that is based on idiosyncratic shocks. In a first stage, we use average peer idiosyncratic shocks as an instrument for the change in peer-group-level cognitive skills. The exclusion restriction is that an idiosyncratic shock hitting child i affects her peers only indirectly through the shock's impact on the cognitive achievement gain of child i . This is a credible assumption given the idiosyncratic nature of the shocks.²⁹ To validate this identification strategy, we include average peer idiosyncratic shocks in our contextual effects specification shown in column 7 of Table 4, while controlling for own idiosyncratic shocks, and find a statistically insignificant effect associated with it.³⁰ The results in column 4 show that idiosyncratic shocks affect cognitive skill growth adversely in a statistically significant way, which suggests that the instrument is also informative.³¹ When we look at the results in column 5, we note the sim-

²⁷The corresponding results when using reading test scores as the dependent variable are broadly similar to those shown for writing test scores and are available on request from the authors.

²⁸Since the within-transformation eliminates a considerable amount of variation in the data, we choose to estimate the transformed model assuming that the error process follows a known SAR(1) process using a generalized method of moments (GMM) estimator proposed by Kelejian and Prucha (2010).

²⁹Note that the use of shocks as an instrument would hold no power in the presence of full peer-group-based risk-sharing.

³⁰The coefficient of $\mathbf{W}\Delta\text{HH}$ shock (idiosyncratic) is -0.004 with a standard error of 0.179 . In unreported results, we also regress the peer idiosyncratic shock directly on the outcome variable without including endogenous effects and also find a statistically insignificant effect.

³¹This accords with evidence based on the YL data for Ethiopia that show a strong negative effect of idiosyncratic shocks in the form of the death of a child's mother on reading and writing skills between ages 8 and 12 (Himaz (2009)).

TABLE 6. Results for *five* nearest neighbors: alternative specifications (writing).

| | IV Using Peer Idiosyncratic Shocks | | | | | | |
|-------------------------|------------------------------------|--------------------|---------------------|--------------------|--------------------|-----------------------|---------------------------|
| | | | | W Shocks | | W ² Shocks | W & W ² Shocks |
| | SAR | SAR | SAR-LD [†] | First Stage | Second Stage | Second Stage | Second Stage |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| WΔwriting level | 0.832** (0.192) | 0.822+ (0.449) | 0.969* (0.460) | | 0.873* (0.377) | 0.833* (0.332) | 0.854* (0.346) |
| ΔCH shock (idsync.) | -0.119+ (0.068) | -0.108 (0.069) | | -0.051 (0.034) | -0.116 (0.091) | -0.119 (0.086) | -0.117 (0.089) |
| WΔCH shock (idsync.) | | | | -0.327* (0.128) | | | |
| ΔCH weight ^a | -0.009* (0.004) | -0.009* (0.004) | -0.009* (0.004) | -0.004* (0.002) | -0.009+ (0.005) | -0.009+ (0.005) | -0.009+ (0.005) |
| ΔCH height ^b | -0.009* (0.004) | -0.010* (0.004) | -0.012** (0.004) | 0.001 (0.002) | -0.009* (0.004) | -0.009* (0.004) | -0.009* (0.004) |
| ΔCH siblings | 0.057 (0.057) | 0.054 (0.058) | 0.062 (0.060) | 0.017 (0.029) | 0.056 (0.066) | 0.057 (0.065) | 0.056 (0.066) |
| ΔCH work | -0.028 (0.061) | -0.012 (0.059) | -0.029 (0.068) | -0.108* (0.047) | -0.022 (0.070) | -0.028 (0.065) | -0.025 (0.067) |
| Δpublic school | 0.183** (0.062) | 0.174** (0.062) | 0.204** (0.069) | -0.017 (0.032) | 0.183** (0.069) | 0.183** (0.069) | 0.183** (0.069) |
| ΔHH assets | 0.058 (0.243) | 0.070 (0.246) | -0.019 (0.279) | -0.067 (0.140) | 0.059 (0.244) | 0.058 (0.243) | 0.059 (0.244) |
| ΔHH shock (covariate) | 0.048 (0.078) | 0.068 (0.071) | 0.027 (0.085) | 0.128** (0.035) | 0.044 (0.077) | 0.048 (0.078) | 0.046 (0.077) |
| ΔCYS ^c | -0.010 (0.026) | -0.015 (0.033) | | 0.010 (0.036) | -0.010 (0.019) | -0.010 (0.019) | -0.010 (0.019) |
| ΔCYS × WΔwriting level | | 0.029 (0.140) | | | | | |
| Urban area | -0.119 (0.091) | | | -0.185* (0.083) | -0.114 (0.088) | -0.119 (0.084) | -0.116 (0.086) |
| Coastal Andhra | 0.012 (0.076) | | | -0.005 (0.075) | 0.011 (0.043) | 0.012 (0.043) | 0.011 (0.043) |
| Observations | 720 | 720 | 720 | 720 | 720 | 720 | 720 |
| First-stage <i>F</i> | 17.772 | 11.352 | 6.794 | | 25.801 | 25.433 | 13.539 |

Notes: CH, child; HH, household; VIL, village. The dagger (†) refers to the local-differenced specification using the STSLS estimator (Kelejian and Prucha (1998)). Standard errors are given in parentheses; clustered standard errors (by village) for IV results. The superscript plus (+) indicates significance at 10%; * at 5%; ** at 1%.

^aChild weight is weight for age *z*-score.

^bChild height is height for age *z*-score.

^cCYS is community years of schooling.

ilarity of the magnitude of the peer effects coefficient with respect to the coefficients obtained using the S2SLS estimator. This is not that surprising given that the S2SLS uses all variables included in the first stage in column 4 in its instrument set—with the exception of the shock variable. Nevertheless, this finding lends further credibility to our choice of the spatial two-step estimator as our preferred estimator.

In the next section, Section 5.2, we describe and test for peer insurance and find evidence for partial risk-sharing within peer groups. This means that risk-sharing is heterogeneous in the sense that some groups do insure (girls) whereas others do not (boys). As a result, the power or the informativeness of the instrumental variable approach comes from the heterogeneity in the implementation of insurance. However, the presence of treatment heterogeneity, that is, shocks that affect only a subset of the peer group due to selective risk-sharing, does not invalidate our IV strategy.³² Both the composition of peer groups and the distribution of peer shocks are orthogonal to the target child's skill formation. The former is due to the differencing of the time-invariant neighborhood sorting effects and the latter is due to the idiosyncratic nature of the shocks. As this heterogeneity derives from the interaction of these two effects, the overall validity of peer shocks as an instrument should be satisfied. Nevertheless, since we do not investigate in detail the determinants of partial insurance, we offer additional robustness checks around this instrument. We further strengthen the validity of the instrument by using higher order neighbors as additional instruments, that is, shocks that are one step further removed from child i , as suggested by Bramoullé, Djebbari, and Fortin (2009). Columns 6 and 7 provide additional results when we use the average idiosyncratic shocks of peers of peers as an instrument, either on its own (column 6) or in combination with the average idiosyncratic shocks of peers (column 7). The results are largely unchanged relative to those obtained from using only average idiosyncratic shocks of peers as an instrument.

To infer the direction of the different sources of potential biases due to correlated effects, we provide a comparison of different estimators described in Table 1. We show estimates obtained for each approach in isolation in Table 7. To explore bias from time-invariant unobservables, we also report estimates from a pooled OLS (POLS) model in column 1. We also implement our two-step optimal instrument method for each specification and report the corresponding first-stage F -statistic along with coefficient estimates from each strategy. Overall, we find that correlated effects, especially spatial sorting, cause a negative bias.

The coefficient on peer effects from the first-differenced (FD) specification (0.826) is larger than that obtained from the POLS model (0.643). As we account for time-varying correlated effects, like village (using a within-differenced transformation) or schooling of older siblings, the peer effect coefficient increases. The first-stage F -statistic of the optimal instrument set, selected using the LARS method described in Section 3, are all mostly over 17. Finally, we note that the estimates obtained from an instrumental variable strategy using idiosyncratic shocks of peers are largely consistent with results obtained using the spatial estimators.

³²However, it does mean we identify only local average treatment effects for the subset of children that insures against idiosyncratic shocks.

TABLE 7. Comparison of estimators.

| | Panel Random Effects | | First Differenced | |
|--------------------------------|----------------------|------------------|--------------------|------------------|
| | Coefficient | | Coefficient | |
| | (1) | Cragg–Donald F | (2) | Cragg–Donald F |
| Base | 0.653** (0.146) | 20.418 | 0.826** (0.191) | 17.667 |
| Base + within-transformation | 0.741** (0.188) | 22.760 | 0.969** (0.459) | 6.800 |
| Base + older sibling schooling | 0.569** (0.206) | 10.009 | 0.831** (0.192) | 17.632 |
| IV using shocks | 0.621** (0.222) | 42.565 | 0.882* (0.386) | 25.456 |

Notes: Base refers to the specification in Equation (2) (omitting contextual effects) for results shown in column 1 and Equation (5) for the results shown in column 2. IV refers to instrumental variable estimates. See also Table 1. The superscript plus (+) indicates significance at 10%; * at 5%; ** at 1%.

We now use our second peer group classification, which is based on children belonging to the same community, that is, a child's peer reference group consists of all other children in the sample who belong to the same community. Table 8 reports results from the community-based classification. Since we rely only on variation in group size for identification, we include only endogenous effects in the specification at the community level, that is, we assume that $\delta = 0$ in Equation (2), to avoid a potential problem of weak instruments.

The results are similar to those obtained using the neighborhood-based peer-group classification. Peer effects are positive and significant across all different specifications.

Overall, these results provide strong evidence for peer effects to matter for the development of cognitive skills. Moreover, endogenous effects appear to be much more important economically than are contextual effects and we find evidence for the presence of time-varying correlated effects.

5.2 Insurance test

Next, we test to see whether peer groups can help children cope with idiosyncratic, adverse shocks. Table 9 reports results from including idiosyncratic shocks using both the five nearest neighbor (5NN) network definition (column 1) and the community-level network definition (column 2).

Overall, we find mixed evidence. For the nearest neighbor peer group, we find that although idiosyncratic shock has a negative, marginally significant effect on child cognitive achievement, the p -value for a joint Wald test of full risk-sharing ($\beta = 1$ and $\eta = 0$) is insignificant. Conversely, when considering community-based peer groups, we find that the effect of shocks on child achievement, although negative, is not significant. The

TABLE 8. Results for community: writing.

| | OLS | SAR | SAR | SAR-LD ^a |
|-----------------------------|---------------------|---------------------|---------------------|---------------------|
| | (1) | (2) | (3) | (4) |
| W Δwriting level | 0.374** (0.079) | 0.667** (0.209) | 0.665** (0.209) | 0.809** (0.252) |
| ΔCH weight ^b | -0.011** (0.004) | -0.010** (0.004) | -0.010** (0.004) | -0.010* (0.004) |
| ΔCH height ^c | -0.009* (0.004) | -0.009* (0.004) | -0.009* (0.004) | -0.009* (0.004) |
| ΔCH siblings | 0.078 (0.054) | 0.075 (0.054) | 0.076 (0.054) | 0.078 (0.055) |
| ΔCH work | -0.107* (0.054) | -0.073 (0.059) | -0.072 (0.059) | -0.040 (0.062) |
| Δpublic school | 0.161** (0.059) | 0.166** (0.060) | 0.166** (0.060) | 0.178** (0.063) |
| ΔHH assets | -0.036 (0.231) | -0.080 (0.232) | -0.077 (0.233) | -0.125 (0.249) |
| ΔHH shock (covariate) | 0.070 (0.071) | 0.048 (0.072) | 0.050 (0.072) | 0.048 (0.080) |
| Δvillage years of schooling | -0.014 (0.062) | | -0.010 (0.062) | |
| Urban area | -0.118 (0.085) | -0.078 (0.089) | -0.078 (0.089) | |
| Coastal Andhra | 0.041 (0.074) | 0.033 (0.074) | 0.034 (0.074) | |
| Observations | 748 | 748 | 748 | 748 |

Notes: CH, child; HH, household. Standard errors are given in parentheses (White standard errors for OLS robust). The superscript plus (+) indicates significance at 10%; * at 5%; ** at 1%.

^aChild weight is weight for age z-score.

^bChild height is height for age z-score.

^cSAR-LD refers to the local-differenced specification using the S2SLS estimator (Kelejian and Prucha (1998)).

p-value for the Wald test for full risk-sharing, however, is 0.087, suggesting that the presence of risk-sharing is rejected at the 10% level.³³

To investigate potential underlying heterogeneity in peer effects and insurance, we split the sample by gender. In rural India, children are often treated differently depending on their gender and caste. In particular, girls maybe be more restricted in the ability to move around freely outside of the household and, therefore, may be exposed to less interaction with their peers. Another reason to split the sample is the possibility that

³³This result of partial/constrained risk-sharing also justifies our use of idiosyncratic shocks as an instrument for endogenous peer effects (Table 6, column 4). The use of these shocks as an instrument is estimated for the full sample (both boys and girls), implying that there is some power in the first stage, since we find that shocks are strongly negatively correlated with growth in achievement within the five nearest neighbors peer groups and for girls in particular (see below).

TABLE 9. Testing for insurance (writing).

| | 5NN | Community | 5NN, Boys | 5NN, Girls |
|---|--------------------|---------------------|--------------------|---------------------|
| | (1) | (2) | (3) | (4) |
| W Δ writing level | 0.839** (0.182) | 0.647** (0.197) | 0.742* (0.294) | 0.853** (0.253) |
| Δ HH shock (idsync.) | -0.118+ (0.068) | -0.105 (0.066) | -0.043 (0.103) | -0.231** (0.089) |
| W Δ HH shock (idsync.) | -0.011 (0.143) | -0.126 (0.155) | -0.161 (0.206) | 0.040 (0.171) |
| Δ CH weight ^a | -0.009* (0.004) | -0.011** (0.004) | -0.016* (0.006) | -0.004 (0.006) |
| Δ CH height ^b | -0.009* (0.004) | -0.009** (0.004) | -0.011+ (0.006) | -0.012* (0.006) |
| Δ CH siblings | 0.056 (0.057) | 0.073 (0.054) | -0.000 (0.092) | 0.110 (0.068) |
| Δ CH work | -0.026 (0.059) | -0.058 (0.057) | -0.036 (0.090) | -0.128 (0.080) |
| Δ public school | 0.183** (0.062) | 0.174** (0.060) | 0.207* (0.091) | 0.163+ (0.085) |
| Δ HH assets | 0.057 (0.244) | -0.081 (0.232) | 0.021 (0.341) | -0.118 (0.341) |
| Δ HH shock (covariate) | 0.049 (0.079) | 0.090 (0.077) | 0.124 (0.118) | 0.071 (0.102) |
| Δ community years of schooling | -0.010 (0.026) | - | -0.019 (0.043) | -0.031 (0.034) |
| Urban area | -0.120 (0.098) | -0.121 (0.098) | -0.064 (0.142) | -0.109 (0.133) |
| Coastal Andhra | 0.011 (0.077) | 0.020 (0.074) | 0.100 (0.118) | -0.015 (0.101) |
| <i>p</i> -value for Wald test: $\beta = 1$ and $\eta = 0$ | 0.182 | 0.088 | 0.636 | 0.032 |
| Observations | 720 | 748 | 368 | 359 |

Notes: CH, child; HH, household; VIL, village. Standard errors are given in parentheses. The superscript plus (+) indicates significance at 10%; * at 5%; ** at 1%.

^aChild weight is weight for age *z*-score.

^bChild height is height for age *z*-score.

boys tend to interact more with boys and girls tend to interact with girls. Table 9 contains the results from splitting the sample into these categories.³⁴ The most interesting finding is that the idiosyncratic shock negatively affects skill formation of girls, whereas the effect is insignificant for boys. We find the strongest evidence to support the full insurance hypothesis among boys. The *p*-value for a joint Wald test of full risk-sharing is 0.636 and the effect of shocks on boys, conditioning on peer effects, is largely insignif-

³⁴This means we construct peer groups by gender, allowing boys to interact only with boys and girls to interact only with girls.

icant. We fail, therefore, to reject the presence of full insurance among the subsample of boys, but are still unclear whether this result has much significance given the much smaller sample size in this case.

The fact that we find that peer effects tend not to insure girls against adverse shocks lends support to previous findings in the literature suggesting the presence of widespread discrimination against girls in developing countries. Björkman-Nyqvist (2013), for instance, found that negative income shocks have large negative and highly significant effects on female enrollment in primary schools, whereas the effect on boys' enrollment is smaller and only marginally significant. It would be interesting to see whether this effect plays out differentially in rural as compared to urban areas. Given the limited sample size, it is not possible to carry out this exercise with the present data, but we highlight this as a matter for future research.

5.3 Robustness

In this section, we conduct various robustness checks to assess the robustness of our estimates to different sorts of bias. Our key identifying assumption centers around the specification of the peer network structure. This section, therefore, investigates the sensitivity of our results to different assumptions regarding the peer network structure.

5.3.1 Monte Carlo experiment: Repeated sampling First, we provide Monte Carlo evidence to investigate the sensitivity of our results to misspecification of the social network structure due to the fact that we have data only on a random sample of children. In this experiment, we simulate data representing fictitious villages by generating 260 data points and assign them random locations in space. Based on the locations of the units in our population, we build our peer reference group based on the envisioned spatial structure, that is, five nearest neighbors, and compare them to estimates obtained from drawing random samples of size 10–100%. We assess both the bias of the estimator under each sample size and the incidence of Type 1 and Type 2 errors.

Table 10 reports results from these simulations. For each sample size, we report the mean squared error (MSE), the mean sample estimate, and the rejection rates. The results show that the quality of the estimator suffers greatly at sample sizes less than 50%, with the MSE being above 1 (barring when $\beta = 0$). When $\beta = 0$ (i.e., when there is no spatial influence in the population), the rejection frequency is well below 10%. This implies that there is, at most, a 10% chance of committing a Type 1 error, that is, falsely detecting a spatial effect when it does not exist in the population. When $\beta \neq 0$, we find reject rates of over 99% for sample sizes above 40%. This means that for these sample sizes, there is less than 1% chance that we will commit a Type 2 error, that is, reject the presence of a spatial effect when in fact a true spatial effect exists in the population.

Our results provide some evidence that estimates based on a sample structure of networks are likely to be robust to sampling error for sampling rates of above 0.8 at high parameter values. Furthermore, we find that sampling induces a consistent downward

TABLE 10. Simulations: repeated sampling.

| Size | $\beta = 0$ | | | $\beta = 0.1$ | | | $\beta = 0.2$ | | |
|------|---------------|-------|-------|---------------|-------|-------|---------------|-------|-------|
| | Mean | RMSE | RR | Mean | RMSE | RR | Mean | RMSE | RR |
| 0.10 | -0.004 | 0.094 | 8.32 | 0.006 | 0.129 | 7.04 | 0.014 | 0.213 | 8.08 |
| 0.20 | 0.000 | 0.062 | 5.96 | 0.017 | 0.105 | 7.88 | 0.036 | 0.179 | 13.92 |
| 0.30 | 0.000 | 0.050 | 5.28 | 0.028 | 0.090 | 11.40 | 0.061 | 0.153 | 30.12 |
| 0.40 | 0.000 | 0.043 | 5.60 | 0.039 | 0.076 | 18.12 | 0.085 | 0.126 | 57.24 |
| 0.50 | 0.000 | 0.038 | 4.84 | 0.051 | 0.064 | 31.76 | 0.108 | 0.103 | 80.64 |
| 0.60 | 0.000 | 0.036 | 5.64 | 0.060 | 0.053 | 46.52 | 0.127 | 0.083 | 95.24 |
| 0.70 | 0.000 | 0.032 | 5.68 | 0.072 | 0.044 | 64.04 | 0.146 | 0.064 | 99.12 |
| 0.80 | -0.001 | 0.030 | 5.12 | 0.080 | 0.035 | 80.12 | 0.164 | 0.048 | 99.92 |
| 0.90 | 0.000 | 0.028 | 4.96 | 0.090 | 0.029 | 91.36 | 0.183 | 0.032 | 99.96 |
| 1.00 | 0.000 | 0.026 | 4.44 | 0.100 | 0.026 | 96.96 | 0.200 | 0.025 | 100 |
| Size | $\beta = 0.3$ | | | $\beta = 0.4$ | | | $\beta = 0.5$ | | |
| | Mean | RMSE | RR | Mean | RMSE | RR | Mean | RMSE | RR |
| 0.10 | 0.021 | 0.305 | 10.08 | 0.029 | 0.401 | 10.80 | 0.047 | 0.494 | 13.36 |
| 0.20 | 0.059 | 0.256 | 22.76 | 0.089 | 0.330 | 34.60 | 0.120 | 0.405 | 44.20 |
| 0.30 | 0.103 | 0.210 | 55.40 | 0.148 | 0.268 | 75.28 | 0.202 | 0.320 | 86.76 |
| 0.40 | 0.140 | 0.171 | 86.00 | 0.195 | 0.218 | 95.64 | 0.266 | 0.251 | 99.20 |
| 0.50 | 0.168 | 0.142 | 96.96 | 0.240 | 0.172 | 99.68 | 0.321 | 0.193 | 100 |
| 0.60 | 0.200 | 0.110 | 99.68 | 0.279 | 0.132 | 100 | 0.368 | 0.145 | 100 |
| 0.70 | 0.228 | 0.082 | 100 | 0.315 | 0.095 | 100 | 0.407 | 0.104 | 100 |
| 0.80 | 0.255 | 0.056 | 100 | 0.346 | 0.064 | 100 | 0.445 | 0.067 | 100 |
| 0.90 | 0.279 | 0.035 | 100 | 0.378 | 0.036 | 100 | 0.476 | 0.036 | 100 |
| 1.00 | 0.301 | 0.023 | 100 | 0.403 | 0.022 | 100 | 0.504 | 0.021 | 100 |

(Continues)

bias in the estimates at all sample sizes and at all spatial parameter values. We have estimated a coverage rate of 60% of our sample in the population of 8–12-year-old children in each village. Given that our spatial effects estimates range from 0.7 to 0.9, we conclude that our results could be potentially biased downward. However, there is less than a 5% chance that we have falsely detected the true spatial effect.

5.3.2 Monte Carlo experiment: Assessing parameter bias using networks with controlled topology Next we explore bias that arises from empirical misspecification of the true, albeit unknown, network structure. To do this, we use a controlled network topology driven Monte Carlo experiment similar to Páez, Scott, and Volz (2008), discussed in Section 2.3 and outlined in Appendix A.2. As in Section 5.3.1, we estimate a population model and compare it to the distribution of sample estimates, but this time we do not impose any specific network structure on the population. Instead, we allow the network structure to be chosen at random, controlling only two topological network parameters: network density and spatial dependence.

The simulation is carried out in two steps. First, we simulate several random networks in the population given a specific combination of network topology parameter

TABLE 10. *Continued.*

| Size | $\beta = 0.6$ | | | $\beta = 0.7$ | | | $\beta = 0.8$ | | |
|------|---------------|-------|-------|---------------|-------|-------|---------------|-------|-------|
| | Mean | RMSE | RR | Mean | RMSE | RR | Mean | RMSE | RR |
| 0.10 | 0.063 | 0.597 | 15.20 | 0.108 | 0.712 | 20.12 | 0.143 | 0.858 | 20.12 |
| 0.20 | 0.154 | 0.481 | 53.28 | 0.212 | 0.542 | 67.20 | 0.294 | 0.592 | 67.20 |
| 0.30 | 0.270 | 0.357 | 93.76 | 0.354 | 0.387 | 96.80 | 0.464 | 0.398 | 96.80 |
| 0.40 | 0.350 | 0.271 | 99.76 | 0.449 | 0.279 | 99.96 | 0.578 | 0.259 | 99.96 |
| 0.50 | 0.410 | 0.206 | 100 | 0.518 | 0.204 | 100 | 0.648 | 0.178 | 100 |
| 0.60 | 0.466 | 0.149 | 100 | 0.575 | 0.142 | 100 | 0.701 | 0.118 | 100 |
| 0.70 | 0.510 | 0.102 | 100 | 0.619 | 0.096 | 100 | 0.738 | 0.078 | 100 |
| 0.80 | 0.547 | 0.064 | 100 | 0.656 | 0.058 | 100 | 0.770 | 0.045 | 100 |
| 0.90 | 0.580 | 0.034 | 100 | 0.685 | 0.031 | 100 | 0.792 | 0.025 | 100 |
| 1.00 | 0.606 | 0.020 | 100 | 0.708 | 0.019 | 100 | 0.809 | 0.017 | 100 |

| Size | $\beta = 0.9$ | | |
|------|---------------|-------|-------|
| | Mean | RMSE | RR |
| 0.10 | 0.305 | 0.932 | 34.36 |
| 0.20 | 0.437 | 1.084 | 86.68 |
| 0.30 | 0.643 | 0.366 | 99.24 |
| 0.40 | 0.746 | 0.209 | 100 |
| 0.50 | 0.803 | 0.131 | 100 |
| 0.60 | 0.844 | 0.079 | 100 |
| 0.70 | 0.869 | 0.048 | 100 |
| 0.80 | 0.888 | 0.028 | 100 |
| 0.90 | 0.901 | 0.017 | 100 |
| 1.00 | 0.909 | 0.014 | 100 |

Note: RR is the Rejection Rates in percentages.

values. We do this for all combinations within a specified range of parameter values. For each random network, we estimate the population model and compare it to the distribution of sample estimates. The sample estimates are based on a model whose network structure is given by the five nearest neighbors definition. Hence, in essence, we compare the peer effect estimate obtained from the population with a random network structure to the estimate obtained from a sample with a five nearest neighbor network structure over a range of two network topology parameters. We reproduce results for only $\beta = 0.8$ and for a sample size of 0.6. Figure 4 shows this in the form of a surface plot of the MSE along with the relevant network parameters associated with each MSE value. The figure shows that the MSE is lowest when the spatial dependence in the true social network is high and network density is low. The quality of the five nearest neighbor sample estimator suffers with decreasing spatial dependence as captured in the true weighting scheme.

To place these results in the context of our data, we map the spatial topology of a random sample village in our data. To do this, we analyze the spatial pattern of the sample village incident point data. We make use of a variant of Ripley’s K -function statistic

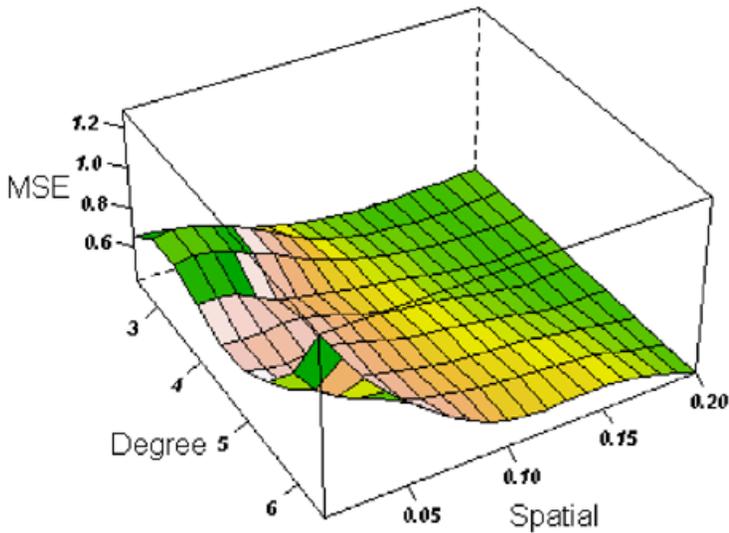


FIGURE 4. Monte Carlo misspecification. *Note:* The surface plot maps the distribution of the mean squared error over two network topology parameters: network density parameter (degree) and spatial dependence parameter (spatial).

that evaluates a given spatial distribution in relation to complete spatial randomness.³⁵ When the observed K -value is larger than the expected K -value for a particular distance, the distribution is more clustered than a random distribution at that distance. When the observed K -value is smaller than the expected K , the distribution is more dispersed than a random distribution at that distance (for further details, see Boots and Getis (1988)). Figure 5 plots the distribution of both observed and expected K -values along with the confidence intervals. The figure shows that the observed K -value is larger than the higher confidence envelope, implying that spatial clustering for *smaller distances* is statistically significant. At larger distances (exceeding 2 kilometers), the observed K -value is smaller than the lower confidence envelope; therefore, spatial dispersion for that distance is statistically significant. This shows that the optimal spatial network involves specifying relatively few neighbors so as to capture high spatial clustering.³⁶ In fact, Figure 6 indicates that the spatial autocorrelation for a 5NN network specification in our data is around 0.1. Figure 4 thus shows that the MSE is relatively low for spatial autocorrelation of this magnitude.

Thus, since we find the presence of substantial spatial autocorrelation in our estimates and specify a sparse network structure, these Monte Carlo results suggest that the

³⁵A variant of the K -function is given by $L(d) = \sqrt{\frac{A \sum_{i=1}^n \sum_{j=1, j \neq i}^n k(i, j)}{\pi n(n-1)}}$, where d is the distance, n is equal to the total number of features, A represents the total area of the features, and $k(i, j)$ is the weight. The weight will be equal to 1 when the distance between i and j is less than d , and will equate to 0 otherwise. We specify 10 distance bands to capture the correct level of clustering.

³⁶The maximum nearest neighbor distance in our five nearest neighbor specification is 1.5 km, which incorporates all levels of spatial clustering.

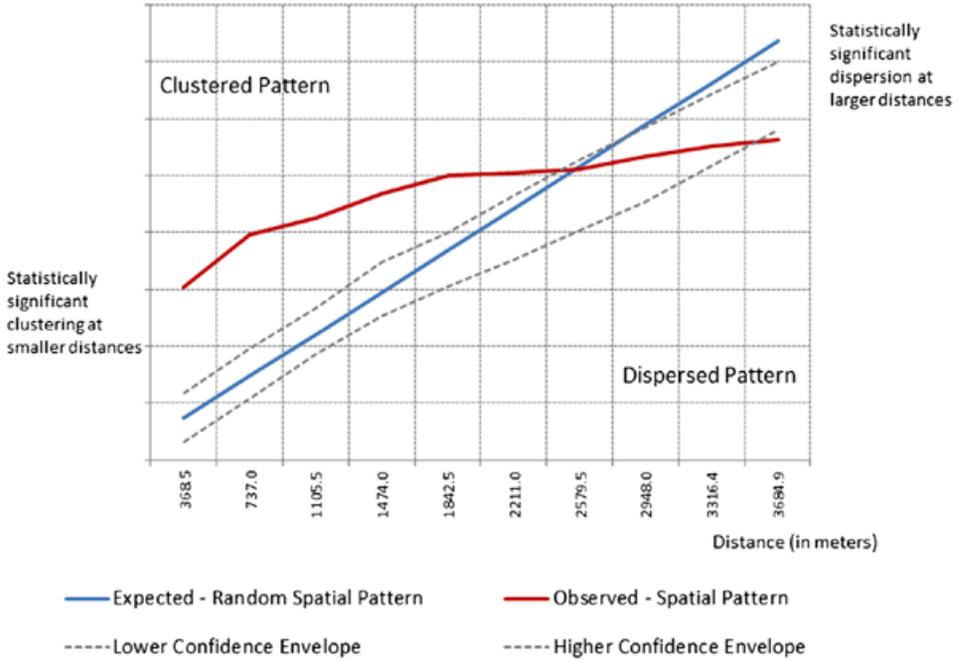


FIGURE 5. Spatial topology of a sample village. *Note:* The graph plots the spatial pattern of a sample village. The line plots the distribution of K -values calculated based on Ripley's K -function, which evaluates a given spatial distribution in relation to complete spatial randomness. The x -axis represents the distances at which the observed K -value is larger or smaller than the expected K -value. For more details, see Boots and Getis (1988).

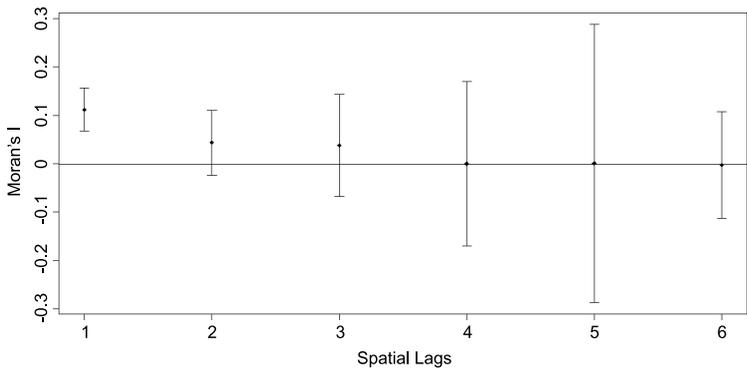


FIGURE 6. Spatial autocorrelation plot for child writing. *Note:* The correlogram plots the spatial autocorrelation in child writing scores (y) for each successive spatial lag, that is, among children considered as first-order neighbors, second-order neighbors, etcetera. We use Moran's I statistic to measure autocorrelation, which is defined as $I = \frac{y'Wy}{n^{-1}y'y[\text{tr}((W'+W)W)^{0.5}]}$, where W is constructed using the 5NN definition.

empirical bias resulting from potential misspecification of the sample network structure might be relatively minor.

5.3.3 Inclusion of long ties and network size To check that our peer effects are not driven solely by the choice of a five nearest neighbor peer-group structure, we provide additional results from (a) varying the size of neighbor groups and (b) relaxing the nearest neighbor assumption by also allowing children to have geographically distant friends.

(a) *K nearest neighbors*: In the network data, the size of a child's peer group is restricted to some arbitrary number as administered in the survey. Therefore, it is difficult to see how the results would change if the survey had recorded more or less peers for the same sample of respondents. In our analysis, we rely on the construction of peer groups *after* data collection. As described earlier, we use the method of *K* nearest neighbors to construct peer groups for each child. Initially, we restricted this set to five. Our results remain largely unchanged when considering groups of three and seven. Table 11 reports results from varying the network size.

We find that the coefficient on peer effects is large, positive, and statistically significant for nearly all specifications. The magnitude of the coefficient, however, falls

TABLE 11. Results for *three* and *seven* nearest neighbors and long ties: writing.

| | SAR | SHAC |
|--|--------------------|--------------------|
| | (1) | (2) |
| W Δ writing level (3NN) | 0.532** (0.238) | 0.532** (0.212) |
| W Δ writing level (7NN) | 0.961** (0.186) | 0.961** (0.161) |
| W Δ writing level (distance decay) ^a | 0.836** (0.240) | 0.836** (0.034) |
| W Δ writing level (inclusion of long ties) ^b | 0.683** (0.173) | 0.683** (0.151) |
| Observations | 751 | 751 |

Notes: Specification as in Table 4. Standard errors are given in parentheses (White standard errors for OLS robust). ** indicates significance at 1%.

^aSpatial weights are defined as $w_{ij} = \frac{1/d_{ij}}{\sum_{r=1}^n (1/d_{ir})}$, where d_{ij} denotes the geographical distance between i and j .

^bThe specification incorporates long ties in the network structure. For each child i , the entire sample is divided into three location-based sets: children within a 2 km radius surrounding child i 's location, all other children in child i 's community outside of the 2 km radius, and all other children outside the community but within the sampling cluster to which the community belongs. Links between child i and five other children from these three distance bands are drawn in two steps. First we randomly select one of the three distance bands based on probabilities assigned to each set: 65% for children within the 2 km radius, 34% for children outside of the 2 km radius but within the community, and 1% for all other children within a given sampling cluster. After selecting the distance band, we select a child with uniform probability without replacement from the set and assign it to the target child as a link. We repeat this exercise 200 times and report the average point estimate from these replications.

when we use three nearest neighbors. When using seven nearest neighbors, the magnitude increases slightly relative to the results obtained when using five nearest neighbors. This might reflect the findings by Lee (2009) discussed above. Lee (2009) suggested that the bias from underspecification of the true network structure points downward, whereas that from overspecification points upward relative to the estimate obtained from the true network structure. These results, therefore, provide additional support for our choice to construct peer groups based on a five nearest neighbor structure.

(b) *Distance decay and long ties*: Our assumption is that children interact with other children who are geographically closest to them. In our KNN specification, this means we defined $w_{ij} = 1/n_i$, where n_i is the number of child i 's neighbors. A simple test of this assumption is provided in Figure 6, where we show the spatial correlogram for six spatial lags (first-order neighbor, second-order neighbor, etc.) with respect to the dependent variable, changes in child writing scores. The graph suggests that peer interactions are indeed local. The spatial autocorrelation coefficient is positive and significant for first degree neighbors, but declines very rapidly for higher order neighbors (lags), with the greatest decline from the first to the second lag.

To test this more formally and to assess how peer effects depend on geographical distance, we estimate an alternative specification in which we define spatial weights as $w_{ij} = \frac{1/d_{ij}}{\sum_{r=1}^n (1/d_{ir})}$, where d_{ij} denotes the geographical distance between i and j . This specification allows also for distant children to exert peer effects, but the strength of their effects decays, that is, $w_{ij} \rightarrow 0$ as $d_{ij} \rightarrow \infty$. Table 11 shows the corresponding estimates, which are very close to the estimates obtained when using a five nearest neighbor definition of peer interaction.

While these estimates as well as the spatial correlogram suggest highly localized peer interaction, there is still the possibility that with small probability, children at distant locations, for example, at other ends of a village, interact with each other. To account for the possibility of such "long ties" in our network structure, we create networks using the algorithm described in Conley and Topa (2007). For each child i , we divide our entire sample into three location-based sets: children residing within a 2 km radius surrounding child i 's location, all other children who live in child i 's community outside of the 2 km radius, and all other children outside the community but within the sampling cluster to which the community belongs. We draw links between child i and five other children from these three distance bands in two steps. We first randomly select one of the three distance bands based on probabilities assigned to each set: 65% for children within the 2 km radius, 34% for children outside of the 2 km radius but within the community, and 1% for all other children within a given sampling cluster. After selecting the distance band, we select a child with uniform probability without replacement from the set and assign it to the target child as a link. We repeat this exercise 200 times and report the average point estimate from these replications.

A potential problem with allowing for long ties even with only small probability is that the cross section could potentially exhibit strong dependence. This is because the incorporation of long ties drastically reduces average (network) path lengths between any two children in the network, invalidating standard inference procedures (Conley and Topa (2007)). We address this issue in two ways. Our approach is similar to the

common factor model approach used to deal with cross-sectional dependence in large (macro) panels that account for the presence of unobserved common factors by augmenting the estimation equation with cross-section averages (Pesaran (2006)) or extracted principal components (Kapetanios (2007), Bai and Ng (2004)) that capture unobserved factors. First, we use spatial filtering techniques to account for long-range spatial dependence given our network specification that includes long ties. Using a two-step approach developed by Getis and Griffith (2002), we decompose the vector of exogenous variables into two components: a part that contains the spatial autocorrelation and a part from which the spatial autocorrelation has been filtered out.³⁷ We include both the filtered and the nonfiltered components of the vector of exogenous variables in the specification, while retaining the endogenous peer effect variables.³⁸ Second, we account for a global spatial trend by including a smoothed natural cubic spline of each child's location coordinates (Fahrmeir and Kneib (2008), Currie and Durban (2002)). To validate our method, we plot the spatial correlogram of residuals obtained from using OLS (Figure 7) not accounting for global spatial dependence and compare it to the spatial correlogram of the residuals obtained from an OLS model that includes filtered variables as well as smoothed splines of spatial coordinates (latitude/longitude) (Figure 8). The residual plot for the model that ignores strong spatial dependence in Figure 7 shows that spatial autocorrelation in the residuals is present at least up to the second lag. Once

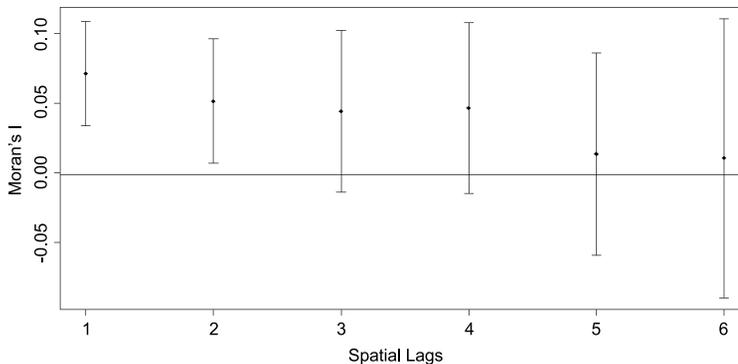


FIGURE 7. Spatial autocorrelation plot for residuals *not* accounting for global dependence. *Note:* The correlogram plots the spatial autocorrelation in the residuals (\mathbf{v}) from a simple OLS specification, not accounting for global dependence, for each successive spatial lag, that is, among children considered as first-order neighbors, second-order neighbors, etcetera. We use Moran's I statistic to measure autocorrelation, which is defined as $I = \frac{\mathbf{v}'\mathbf{W}\mathbf{v}}{n^{-1}\mathbf{v}'\mathbf{v}[\text{tr}((\mathbf{W}' + \mathbf{W})\mathbf{W})]^{0.5}}$, where \mathbf{W} is constructed using the long ties algorithm defined in Section 5.3.3.

³⁷The filtered observation is given by $x_i^* = \frac{x_i(\mathbf{W}_i/n_i)}{G_i}$, where G_i is the Getis–Ord local spatial autocorrelation statistic (Getis and Ord (1992)), which measures the concentration of a given variable within a neighborhood and is defined as $G_i = \frac{\sum_{j=1}^N w_{ij}x_j}{\sum_{j=1}^N x_j}$ ($j \neq i$). The nonfiltered component of the observation x_i is thus $x_i - x_i^*$.

³⁸A different approach to spatial filtering involves eigenvector decomposition of residuals (Tiefelsdorf and Griffith (2007)). However, a limitation of this approach in our context is that we are unable to recover the spatial autoregressive parameter since it is filtered out along with the residuals.

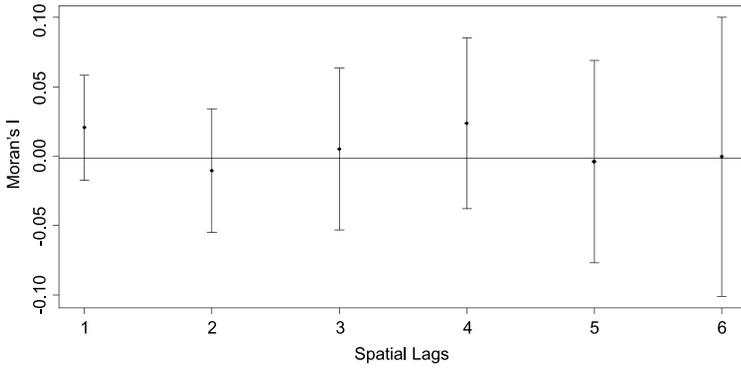


FIGURE 8. Spatial autocorrelation plot for residuals accounting for global dependence. *Note:* The correlogram plots the spatial autocorrelation in the residuals (\mathbf{v}) from a simple OLS specification, accounting for global dependence (by filtering and smoothing), for each successive spatial lag, that is, among children considered as first-order neighbors, second-order neighbors, etcetera. We use Moran's I statistic to measure autocorrelation, which is defined as $I = \frac{\mathbf{v}'\mathbf{W}\mathbf{v}}{n^{-1}\mathbf{v}'\mathbf{v}[\text{tr}((\mathbf{W}+\mathbf{W})^0.5)]}$, where \mathbf{W} is constructed using the long ties algorithm defined in Section 5.3.3.

we account for strong spatial dependence through filtering and the spatial trend, the correlogram in Figure 8 shows no evidence of any remaining spatial dependence.

Table 11 shows that the corresponding point estimate is 0.68, which is slightly lower than the estimates based on the five nearest neighbor network shown in Table 4. This can be interpreted as evidence that children interact primarily with their geographically closest neighbors; allowing for distant friendships with some small probability appears to dilute the peer effect estimate.

5.3.4 Network scramble Finally, we provide a falsification test for our network-based identification strategy. All our results indicate the presence of positive and significant peer effects based on a specific distance-based peer interaction network of a child. We now show that such a result is not obtained from considering *just any random* peer group network. In essence, we validate the strength and significance of the actual observed network by ruling out the presence of peer effects within randomly generated networks. Our objective is to demonstrate that no statistically significant peer interaction is found among children who have been assigned randomly to a peer network. This is a test for our identifying assumption that geographical proximity mediates peer effects. To test this, we randomly assign each child in the sample five nearest neighbors and estimate the model in Equation (5) using the S2SLS estimator, employing the randomly generated spatial weight matrix. Again, we limit ourselves to cognitive skills measured as writing skills as the results carry over to reading skills. We repeat this exercise 300 times, each time generating random five nearest neighbor networks. This exercise is similar to the methodology of *exact test* or *permutation test*, where a distribution of the test statistic under the null hypothesis is obtained by calculating all possible values of the test statistic by rearranging labels of the observed data points (Good (2000)).

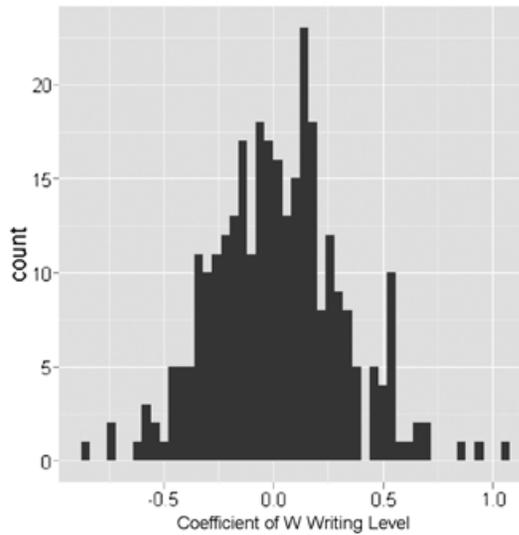


FIGURE 9. Network scramble: histogram of point estimates of endogenous peer effects.

The histogram in Figure 9 shows the empirical distribution of the point estimates obtained from the 300 replications. The mean estimate is 0.019 with a standard error of 0.296, which means that we cannot reject the null hypothesis that peer effects are equal to zero. Moreover, the majority of point estimates are statistically insignificant. For peer effects associated with each of the 300 iterations, we find that we are unable to reject the null that the coefficient (point estimate) is different from zero for 275 coefficients out of 300.

Figure 10 plots the joint distribution of coefficients along with their standard errors obtained from randomizing the network 300 times. The grey dotted ellipse in the figure encloses the area under which coefficient estimates are statistically insignificant. We observe that most point estimates lie within this area. This shows that repeated experiments with different randomized networks produce statistically insignificant peer effects on average. Hence, this exercise corroborates our approach to constructing nearest neighbor peer networks based on geographical proximity.

6. CONCLUSION

In this paper, we analyze the formation of cognitive skills of children in Andhra Pradesh, India, allowing for spatial peer effects. Making use of the specific nature of our data set (i.e., available data on spatial proximity of households), we define a child's peers as her nearest neighbors in terms of geographical distance. Exploiting intransitivity within the networks formed by nearest neighbors, we are able to address Manski's reflection problem. Using first differences allows us to identify contextual and endogenous effects separately and to avoid confounding social effects with unobserved heterogeneity. We use a number of additional model specifications to rule out that time-varying unobservables that are correlated with peer effects drive these results. For our preferred estima-

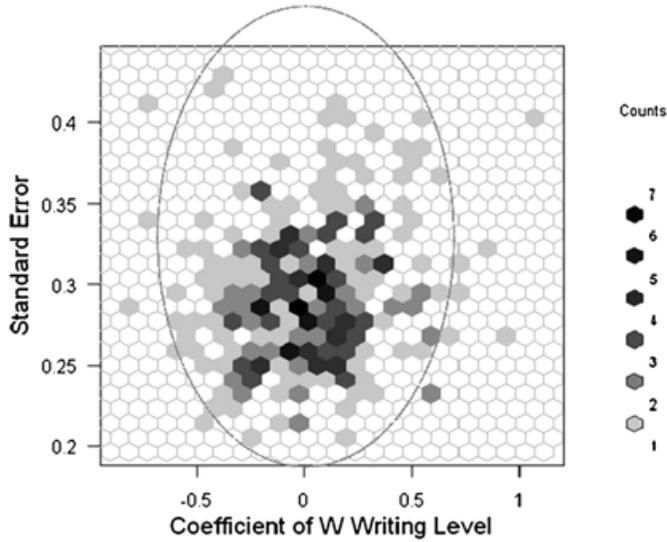


FIGURE 10. Network scramble: lattice map of peer effect coefficients and standard errors. *Note:* The figure shows the joint distribution of coefficients along with their standard errors obtained from randomizing the network 300 times. Coefficients estimates that fall within the area enclosed by the grey dotted ellipse are statistically insignificant, that is, not different from zero.

tion method, which is also the least restrictive one, we find that an increase of 1 standard deviation in the growth of the cognitive achievement of a child’s peers increases cognitive achievement of the child by 0.4 standard deviations. This is a sizeable effect and suggests that peer effects are an important determinant of skill formation thus far neglected in the literature on child skill formation. While we find robust evidence for spatial peer effects, our data do not reveal the mechanisms of these social effects.

We also find evidence for peer effects to provide insurance for children of shock-affected households against adverse effects to their cognitive skill acquisition. Interestingly, this result does not hold when we estimate the model for a sample containing only girls; we interpret this as evidence suggestive of girls being less able to cope with negative idiosyncratic shocks through their peer support.

The strong peer effects that we find in the paper generate a social multiplier of 3.3–4 (for our preferred estimates between 0.7 and 0.8). Other studies that investigate peer effects in education also document evidence on positive peer spillovers. Peer effect estimates in those studies range between 0.2 and 0.6.³⁹ Evidence on peer effects in developing countries shows relatively large peer effects. For instance, [Bobonis and Finan \(2009\)](#) find peer effects in enrollment rates ranging between 0.75 and 0.5, effects being much stronger for poor children. [Duflo, Dupas, and Kremer \(2011\)](#) also found, similar to our results, that a 1 standard deviation increase in average peer test score would increase the test score of a student by 0.445 standard deviations. The magnitude of their peer effects ranges from 0.4 to 0.8, taking higher values for lower quantities of student achievement.

³⁹For an excellent review of peer effects in education and related studies, see [Epple and Romano \(2011\)](#).

Our study, therefore, reinforces findings in the existing literature of large and positive peer effects in developing countries.

APPENDIX

A.1 Risk-sharing: Empirical implications

We describe below the information structure, preferences and endowments of children interacting in networks and derive risk-sharing implications corresponding to the empirical specification in Equation (9). To begin with, we assume that all children share common information at any given time t . This is represented by events $\sum_{\tau=1}^S$, each occurring with a probability $\pi(s_{\tau t})$ such that $\sum_{\tau=1}^S \pi(s_{\tau t}) = 1 \forall t$ (Mace (1991)).

We express the expected utility of child i up to T periods of schooling as

$$\sum_{t=0}^T \lambda^t \sum_{\tau=1}^S \pi(s_{\tau t}) U[y_{it}(s_{\tau t}), b_{it}(s_{\tau t})], \quad (15)$$

where y_{it} denotes academic achievement in event τ at time t , b_{it} is a taste shifter or a preference shock, and λ is the discount factor. Here, the utility of the child i depends positively on her school performance y_i . Children are risk-averse and share the same coefficient of constant absolute risk aversion σ . As in Mace (1991), preferences are homothetic and are both time and state separable. The instantaneous utility function is given by

$$U[y_{it}(s_{\tau t}), b_{it}(s_{\tau t})] = -\frac{1}{\sigma} \exp[-\sigma(y_{it} - b_{it})], \quad \sigma > 0. \quad (16)$$

Achievement is specified as an output of the education production function in which a child's aggregate endowment g_{it} serve as inputs. The input function g_i is determined by a combination of factors, including effort (e_{it}) and time (hours) spent studying (h_{it}):

$$y_{it} = g_{it}(e_{it}, h_{it}; x_{it}). \quad (17)$$

Endowments fluctuate depending on the state $s_{\tau t}$. At any given point of time, they can be characterized as

$$g_{it}(s_{\tau t}) = \tilde{g}_{it} + \phi_{it}(s_{\tau t}) + \varepsilon_{it}(s_{\tau t}), \quad (18)$$

where \tilde{g}_{it} is the deterministic component of the total endowment, while $\phi_{it}(s_{\tau t})$ and $\varepsilon_{it}(s_{\tau t})$ are aggregate and idiosyncratic shocks to endowment, respectively.

To consider the major implication of risk-sharing, we can imagine a social planner who allocates a Pareto weight ω_i to each child within the network and maximizes the weighted sum of expected utilities of the $n_i + 1$ children in the network. The Pareto efficient allocation can also be arrived at by some initial bargaining process where weights could depend on all the variables that determine the outside option of each child, such as paying for private tuition, availability of older siblings to supervise and tutor, and

so forth (Dercon and Krishnan (2000)). Taking the case of the social planner, full risk-sharing within the local peer groups would imply that the planner chooses an allocation subject to the total endowment constraint each state and time period:

$$\sum_{i=1}^{n_i+1} \omega_i \sum_{t=0}^T \lambda^t \sum_{\tau=1}^S \pi(s_{\tau t}) U[y_{it}(s_{\tau t}), b_{it}(s_{\tau t})] \tag{19}$$

$$\text{s.t. } \sum_{i=1}^{n_i+1} y_{it}(s_{\tau t}) = \sum_{i=1}^{n_i+1} g_{it}(s_{\tau t}). \tag{20}$$

The Lagrangian can be written as

$$\mathcal{L} = \sum_{i=1}^{n_i+1} \omega_i \sum_{t=0}^{\infty} \lambda^t \sum_{\tau=1}^S \pi_t U[y_{it}, b_{it}] + \mu_t \left(\sum_{i=1}^{n_i+1} g_{it} - \sum_{i=1}^{n_i+1} y_{it} \right). \tag{21}$$

The first-order conditions of the Lagrangian with respect to y_{it} and y_{jt} , where $j \neq i$, are

$$\omega_i \lambda^t \pi_t U'[y_{it}, b_{it}] - \mu_t = 0, \tag{22}$$

$$\omega_j \lambda^t \pi_t U'[y_{jt}, b_{jt}] - \mu_t = 0. \tag{23}$$

The two equations above yield the condition

$$\frac{U'[y_{it}, b_{it}]}{U'[y_{jt}, b_{jt}]} = \frac{\omega_j}{\omega_i}. \tag{24}$$

Substituting for the functional form of the utility function and taking logs, we can rewrite Equation (24) as

$$\ln \left\{ \frac{\exp[-\sigma(y_{it} - b_{it})]}{\exp[-\sigma(y_{jt} - b_{jt})]} \right\} = \ln \left(\frac{\omega_j}{\omega_i} \right). \tag{25}$$

Rearranging and solving for y_{it} ,

$$y_{it} = y_{jt} + b_{it} - b_{jt} + \frac{1}{\sigma} (\ln \omega_i - \ln \omega_j). \tag{26}$$

A similar expression can be obtained with respect to any of the $J - 1 = n_i$ agents in the network of person i (De Weerd and Dercon (2006)):

$$y_{it} = y_{j+1t} + b_{it} - b_{j+1t} + \frac{1}{\sigma} (\ln \omega_i - \ln \omega_{j+1}), \tag{27}$$

$$y_{it} = y_{j+2t} + b_{it} - b_{j+2t} + \frac{1}{\sigma} (\ln \omega_i - \ln \omega_{j+2}), \tag{28}$$

$$\dots \tag{29}$$

$$y_{it} = y_{J-1t} + b_{it} - b_{J-1t} + \frac{1}{\sigma} (\ln \omega_i - \ln \omega_{J-1}). \tag{30}$$

Summing across the $J - 1$ pairwise combinations of agents and dividing by (n_i) , we get

$$y_{it} = \frac{\sum_{j=0}^{n_i} y_{jt}}{n_i} + b_{it} - \frac{\sum_{j=0}^{n_i} b_{jt}}{n_i} + \frac{1}{\sigma} \left(\ln \omega_i - \frac{\sum_{j=0}^{n_i} \ln \omega_j}{n_i} \right). \tag{31}$$

We eliminate the time-invariant Pareto weights by taking first differences:

$$\Delta y_{it} = \Delta \left[\frac{\sum_{j=0}^{n_i} y_{jt}}{n_i} \right] + \Delta b_{it} - \Delta \left[\frac{\sum_{j=0}^{n_i} b_{jt}}{n_i} \right]. \tag{32}$$

We empirically test for risk-sharing across peer networks by including aggregate peer achievement and time-varying demographic characteristics of the children as well as their peers, which are assumed to capture the changes in taste shifters (b_{it} , $(\sum_{j=0}^{n_i} b_{jt})/n_i$) as specified in Equation (32). Equation (32) implies that under full insurance, child achievement responds to aggregate risk but not to idiosyncratic risk. Therefore, to examine whether full risk-sharing holds, we introduce a measure of unanticipated idiosyncratic shock s_i as an overspecification of the econometric model and test to see if $\eta = 0$. The regression specification is given by

$$\Delta y_i = \beta \frac{\sum_{j \in P_i} \Delta y_j}{n_i} + \gamma \Delta x_i + \delta \frac{\sum_{j \in P_i} \Delta x_j}{n_i} + \eta s_i + v_i. \tag{33}$$

A.2 Monte Carlo setup and results

This section outlines the simulation procedure. We first draw samples of diffract sizes from the true population and then simulate random networks with controlled network topologies. In the first simulation exercise, we consider a true spatial network \mathbf{W}^S with $n = 260$ children residing in three villages. We do this by generating 260 data points and assign them random locations in space. Based on the locations of the units in our population, we build our peer reference group based on the envisioned spatial structure, that is, five nearest neighbors, where \mathbf{W}^S is constructed by taking the five nearest neighbors of each child within the population. We then draw several stratified samples (by village) of different sizes from the population. Data for the experiment are generated using the autoregressive model

$$\mathbf{y}_n = \beta \mathbf{W}_n^S \mathbf{y}_n + \gamma_0 \mathbf{l}_n + \gamma_1 \mathbf{x}_n + u_n, \tag{34}$$

where \mathbf{W}_n^S is an $n \times n$ weights matrix, \mathbf{l}_n is an n -dimensional column vector of ones, $x_{n,i}$ is independently generated from a uniform distribution over the range $[0, 10]$ for $i = 1, \dots, n$, and $u_{i,n}$ are i.i.d. $N(0, \sigma^2 = 0.5)$. The intercept term and the coefficient on the

independent variable are set to 2.0 and 1.0, respectively. The y_n is obtained using the reduced form of Equation (34):

$$\mathbf{y}_n = (\mathbf{I} - \beta \mathbf{W}_n^S)^{-1} (\gamma_0 \mathbf{1}_n + \gamma_1 \mathbf{x}_n + u_n). \tag{35}$$

A set of 50 draws is obtained for the error terms and independent variable \mathbf{x} . The sample levels are $s = 1.0$ (complete network) to $s = 0.1$ (10% sample). The true spatial network is sampled 50 times at each sampling level to give $50 \times 50 = 2500$ independent replications R . For each replication, we estimate Equation (34) with the sample spatial network \hat{W}^S that is obtained by taking the five nearest neighbors of each child within the *sample*. We use a generalized spatial two-stage least squares (2SLS) estimator as proposed in Kelejian and Prucha (1998). To assess the quality of the estimators in the simulations, we calculate the root mean squared error (RMSE), which is the square root of the mean squared error. The MSE for the parameter β is calculated as (Florax and Rey (1995))

$$\text{MSE}(\beta) = \Sigma_r \frac{(\hat{\beta}_r - \bar{\beta})^2}{R} + \left[\frac{\Sigma_r (\beta_s - \hat{\beta}_r)}{R} \right]^2. \tag{36}$$

Additionally, we also test for the presence of the spatial autoregressive parameter, which means we test for Type 1 error when the true peer effect is zero (the percentage of times a test would reject the null hypothesis) and for Type 2 error when the true peer effect is not equal to zero (the percentage of times a test would fail to reject the null hypothesis). We implement a likelihood ratio test (χ^2 -distributed with 1 degree of freedom) defined as (Páez, Scott, and Volz (2008))

$$\text{LR} = 2(L_{\text{SAR}}^* - L_{\text{OLS}}^*), \tag{37}$$

where L_{SAR}^* is the log-likelihood function of the SAR model evaluated using maximum likelihood methods.⁴⁰ The test is computed at the critical value of 0.05 and the frequency of rejections for each test are stored. We calculate the rejection rates when $\beta = 0$ to assess the chances of falsely identifying the social influence process. Similarly, we calculate acceptance rates when $\beta \neq 0$ to assess the chances of failing to detect the social influence process.

Finally, in the second experiment, we consider the true population network \mathbf{W}^P , which is only partly a function of the spatial distribution. We simulate several random networks in the population given a specific combination of network topology parameter values for all combinations within a specified range of parameter values. For each random network, we estimate the population model and compare it to the distribution of sample estimates whose network structure is given by the five nearest neighbors definition.

⁴⁰There are minor differences in the estimates obtained from the ML estimator in comparison to those obtained using the 2SLS estimator. Therefore, we believe that the test results are informative regardless of the choice of the estimator.

REFERENCES

- Ainsworth, J. (2002), "Why does it take a village? The mediation of neighborhood effects on educational achievement." *Social Forces*, 81, 117–152. [69]
- Ammermueller, A. and J. Pischke (2009), "Peer effects in European primary schools: Evidence from the progress in international reading literacy study." *Journal of Labor Economics*, 27, 315–348. [92]
- Anselin, L. (1998), *Spatial Econometrics: Methods and Models*. Kluwer Academic, Dordrecht. [76]
- Anselin, L. and S. Hudak (1992), "Spatial econometrics in practice: A review of software options." *Regional Science and Urban Economics*, 22, 509–536. [90]
- Arcidiacono, P., G. Foster, and J. Kinsler (2012), "Estimating spillovers using panel data, with an application to the classroom." *Quantitative Economics*, 3 (3), 421–470. [92]
- Babcock, L., J. Engberg, and R. Greenbaum (2005), "Wage spillovers in public sector contract negotiations: The importance of social comparisons." *Regional Science and Urban Economics*, 35, 395–416. [93]
- Bai, J. and S. Ng (2004), "A PANIC attack on unit roots and cointegration." *Econometrica*, 72, 1127–1177. [108]
- Bayer, P., S. Ross, and G. Topa (2008), "Place of work and place of residence: Informal hiring networks and labor market outcomes." *Journal of Political Economy*, 116, 1150–1196. [73, 81]
- Belloni, A., V. Chernozhukov, and C. Hansen (2010), "LASSO methods for Gaussian instrumental variables models." Preprint. Available at arXiv:1012.1297. [84]
- Bivand, R., L. Anselin, O. Berke, A. Bernat, M. Carvalho, Y. Chun, C. Dormann, S. Dray, R. Halbersma, N. Lewin-Koh et al. (2011), "spdep: Spatial dependence: Weighting schemes, statistics and models." R package version. [84]
- Björkman-Nyqvist, M. (2013), "Income shocks and gender gaps in education: Evidence from Uganda." *Journal of Development Economics*, 105, 237–253. [101]
- Blume, L. and S. Durlauf (2006), "Identifying social interactions: A review." In *Methods in Social Epidemiology*, 287, Jossey-Bass, San Francisco. [74, 78]
- Bobonis, G. and F. Finan (2009), "Neighborhood peer effects in secondary school enrollment decisions." *Review of Economics and Statistics*, 91, 695–716. [71, 111]
- Boots, B. and A. Getis (1988), *Point Pattern Analysis*, Vol. 8. Sage Publications, Newbury Park, California. [104, 105]
- Borghans, L., A. Duckworth, J. Heckman, and B. ter Weel (2008), "The economics and psychology of personality traits." *Journal of Human Resources*, 43, 972–1059. [87]
- Borjas, G. (1995), "Ethnicity, neighborhoods, and human-capital externalities." *American Economic Review*, 85 (3), 365–390. [68]

- Bramoullé, Y., H. Djebbari, and B. Fortin (2009), "Identification of peer effects through social networks." *Journal of Econometrics*, 150, 41–55. [74, 97]
- Brooks-Gunn, J., G. Duncan, P. Klebanov, and N. Sealand (1993), "Do neighborhoods influence child and adolescent development?" *American Journal of Sociology*, 99 (2), 353–395. [69]
- Caner, M. and Q. Fan (2010), "The adaptive Lasso method for instrumental variable selection." Technical report, North Carolina State University. [84]
- Chandrasekhar, A. and R. Lewis (2012), "Econometrics of sampled networks." Technical report, MIT. [81]
- Conley, T. (1999), "GMM estimation with cross sectional dependence." *Journal of Econometrics*, 92, 1–45. [83]
- Conley, T. and G. Topa (2007), "Estimating dynamic local interaction models." *Journal of Econometrics*, 140, 282–303. [73, 81, 107]
- Conley, T. and C. Udry (2010), "Learning about a new technology: Pineapple in Ghana." *American Economic Review*, 100, 35–69. [81]
- Cooley, J. (forthcoming), "Can achievement peer effect estimates inform policy? A view from inside the black box." *Review of Economics and Statistics*. [93]
- Cueto, S., J. Leon, G. Guerrero, and I. Munoz (2009), "Psychometric characteristics of cognitive development and achievement instruments in Round 2 of Young Lives." Technical Note 15, Young Lives. [87]
- Cunha, F. and J. Heckman (2007), "The technology of skill formation." *American Economic Review*, 97, 31–47. [68]
- Cunha, F. and J. Heckman (2008), "Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation." *Journal of Human Resources*, 43, 738–782. [68]
- Currie, I. and M. Durban (2002), "Flexible smoothing with P-splines: A unified approach." *Statistical Modelling*, 2, 333–349. [108]
- Davezies, L., X. D'Haultfoeuille, and D. Fougère (2009), "Identification of peer effects using group size variation." *Econometrics Journal*, 12, 397–413. [77]
- De Giorgi, G. and M. Pellizzari (forthcoming), "Understanding social interactions: Evidence from the classroom." *The Economic Journal*. [82]
- De Weerd, J. and S. Dercon (2006), "Risk-sharing networks and insurance against illness." *Journal of Development Economics*, 81, 337–356. [70, 113]
- Dercon, S. and P. Krishnan (2000), "In sickness and in health: Risk sharing within households in rural Ethiopia." *Journal of Political Economy*, 108, 688–727. [113]
- Duflo, E., P. Dupas, and M. Kremer (2011), "Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya." *American Economic Review*, 101, 1739–1774. [111]

Durlauf, S. (2004), "Neighborhood effects." In *Handbook of Regional and Urban Economics*, Vol. 4, 2173–2242, Elsevier, Amsterdam. [68, 69, 71]

Durlauf, S. and Y. Ioannides (2010), "Social interactions." *Annual Review of Economics*, 2, 451–478. [68]

Efron, B., T. Hastie, I. Johnstone, and R. Tibshirani (2004), "Least angle regression." *The Annals of Statistics*, 32, 407–499. [84]

Epple, D. and R. Romano (2011), "Peer effects in education: Survey of the theory and evidence." In *Handbook of Social Economics*, Vol. 1, 1053–1163, Elsevier, Amsterdam. [111]

Fafchamps, M. and P. Vicente (2009), "Political violence and social networks: Experimental evidence from a Nigerian election." Working paper, CSAE. [71]

Fahrmeir, L. and T. Kneib (2008), "On the identification of trend and correlation in temporal and spatial regression." In *Recent Advances in Linear Models and Related Areas*, 1–27, Springer, Heidelberg. [108]

Florax, R. and S. Rey (1995), "The impacts of misspecified spatial interaction in linear regression models." In *New Directions in Spatial Econometrics*, 111–135, Springer, Berlin. [115]

Gertler, P. and J. Gruber (2002), "Insuring consumption against illness." *American Economic Review*, 92 (1), 51–70. [70]

Getis, A. and D. Griffith (2002), "Comparative spatial filtering in regression analysis." *Geographical Analysis*, 34, 130–140. [108]

Getis, A. and J. Ord (1992), "The analysis of spatial association by use of distance statistics." *Geographical Analysis*, 24, 189–206. [108]

Gibbons, S. and H. Overman (2012), "Mostly pointless spatial econometrics?" *Journal of Regional Science*, 52 (2), 172–191. [84]

Ginja, R. (2009), "Income shocks and investments in human capital." Working paper, University College London. [70]

Good, P. I. (2000), *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses*, second edition. Springer Series in Statistics. Springer, Berlin. [109]

Goux, M. and E. Maurin (2007), "Close neighbours matter: Neighbourhood effects on early performance at school." *Economic Journal*, 117, 1193–1215. [73]

Hanushek, E., J. Kain, J. Markman, and S. Rivkin (2003), "Does peer ability affect student achievement?" *Journal of Applied Econometrics*, 18, 527–544. [68]

Hartup, W. (1996), "The company they keep: Friendships and their developmental significance." *Child Development*, 67, 1–13. [82]

Heckman, J. (1979), "Sample selection bias as a specification error." *Econometrica*, 47, 153–161. [80]

- Helmers, C. and M. Patnam (2011), "The formation and evolution of childhood skill acquisition: Evidence from India." *Journal of Development Economics*, 95, 252–266. [85, 92]
- Himaz, R. (2009), "The impact of parental death on schooling and subjective well-being: Evidence from Ethiopia using longitudinal data." Working Paper 44, Young Lives. [95]
- Himaz, R., S. Galab, and P. Reddy (2009), "Young Lives Round 2 survey report initial findings: Andhra Pradesh, India." Paper, MPRA. [75]
- Hoff, P. (2009), "Multiplicative latent factor models for description and prediction of social networks." *Computational & Mathematical Organization Theory*, 15, 261–272. [80]
- Hoxby, C. (2000), "Peer effects in the classroom: Learning from gender and race variation." Technical report. [68]
- Kapetanios, G. (2007), "Dynamic factor extraction of cross-sectional dependence in panel unit root tests." *Journal of Applied Econometrics*, 22, 313–338. [108]
- Kelejian, H. and I. Prucha (1998), "A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances." *Journal of Real Estate Finance and Economics*, 17, 99–121. [76, 83, 96, 99, 115]
- Kelejian, H. and I. Prucha (1999), "A generalized moments estimator for the autoregressive parameter in a spatial model." *International Economic Review*, 40, 509–533. [76]
- Kelejian, H. and I. Prucha (2007), "HAC estimation in a spatial framework." *Journal of Econometrics*, 140, 131–154. [70, 83]
- Kelejian, H. and I. Prucha (2010), "Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances." *Journal of Econometrics*, 157, 53–67. [95]
- Kolaczyk, E. (2009), *Statistical Analysis of Network Data: Methods and Models*. Springer, New York. [74]
- Laible, D., G. Carlo, and M. Raffaelli (2000), "The differential relations of parent and peer attachment to adolescent adjustment." *Journal of Youth and Adolescence*, 29, 45–59. [82]
- Lambert, D., R. Florax, and S. Cho (2008), "Bandwidth selection for spatial HAC and other robust covariance estimators." Working paper. [84]
- Lee, L. (2007), "Identification and estimation of econometric models with group interactions, contextual factors and fixed effects." *Journal of Econometrics*, 140, 333–374. [74, 77, 83]
- Lee, L. and J. Yu (2007), "Near unit root in spatial autoregressive model." Report, Ohio State University. [93]
- Lee, S. (2009), "Bias from misspecified spatial weight matrices in SAR models: Theory and simulation studies." Report, Department of Economics, San Francisco State University. [81, 107]

- Lin, X. (2010), "Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables." *Journal of Labor Economics*, 28, 825–860. [68, 81]
- Mace, B. (1991), "Full insurance in the presence of aggregate uncertainty." *Journal of Political Economy*, 99 (5), 928–956. [112]
- Manski, C. (1993), "Identification of endogenous social effects: The reflection problem." *Review of Economic Studies*, 60, 531–542. [68, 71, 74]
- Marsden, P. and J. Hurlbert (1987), "Small networks and selectivity bias in the analysis of survey network data." *Social Networks*, 9, 333–349. [80]
- McCulloch, A. and H. Joshi (2001), "Neighbourhood and family influences on the cognitive ability of children in the British National Child Development Study." *Social Science & Medicine*, 53, 579–591. [69]
- Moffitt, R. (2001), "Policy interventions, low-level equilibria, and social interactions." In *Social Dynamics*, 45–82, Brookings Institution, Washington, D.C. [74]
- Openshaw, S. (1983), *The Modifiable Areal Unit Problem*. Geo Books, Norwich. [75]
- Outes-Leon, I. and S. Dercon (2008), "Survey attrition and attrition bias in Young Lives." Technical Note 5, Young Lives. [74]
- Páez, A., D. Scott, and E. Volz (2008), "Weight matrices for social influence analysis: An investigation of measurement errors and their effect on model identification and estimation quality." *Social Networks*, 30, 309–317. [81, 102, 115]
- Pesaran, M. (2006), "Estimation and inference in large heterogeneous panels with a multifactor error structure." *Econometrica*, 74, 967–1012. [108]
- Piras, G. (2010), "spht: Spatial models with heteroskedastic innovations in R." *Journal of Statistical Software*, 35 (1). [84]
- Sacerdote, B. (2001), "Peer effects with random assignment: Results for Dartmouth roommates." *Quarterly Journal of Economics*, 116, 681–704. [68]
- Taskar, B., M. Wong, P. Abbeel, and D. Koller (2003), "Link prediction in relational data." In *Neural Information Processing Systems*, Vol. 15, MIT Press, Cambridge, Massachusetts. [80]
- Tiefelsdorf, M. and D. Griffith (2007), "Semiparametric filtering of spatial autocorrelation: The eigenvector approach." *Environment and Planning A*, 39, 1193–1221. [108]
- Todd, P. and K. Wolpin (2003), "On the specification and estimation of the production function for cognitive achievement." *Economic Journal*, 113, F3–F33. [68]
- Tominey, E. (2009), "The timing of parental income on child outcomes: The role of permanent and transitory shocks." Working paper, University College London. [70]
- Topa, G. (2001), "Social interactions, local spillovers and unemployment." *Review of Economic Studies*, 68, 261–295. [73]

Townsend, R. (1994), “Risk and insurance in village India.” *Econometrica*, 62, 539–591. [70]

Weedon, M., G. Lettre, R. Freathy, C. Lindgren, B. Voight, J. Perry, K. Elliott, R. Hackett, C. Guiducci, B. Shields et al. (2007), “A common variant of HMGA2 is associated with adult and childhood height in the general population.” *Nature Genetics*, 39, 1245–1250. [87]

White, H. (1980), “A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity.” *Econometrica*, 48, 817–838. [92]

Woodhead, M., M. Frost, and Z. James (2013), “Does growth in private schooling contribute to education for all? Evidence from a longitudinal, two cohort study in Andhra Pradesh, India.” *International Journal of Educational Development*, 33 (1), 65–73. [90]

Submitted September, 2011. Final version accepted June, 2013.