



Misclassification in linear-in-means models: Theory and application to peer effects estimation[☆]

Fanny Valli-Puljic^{a, }, Simone Balestra^{b, c}, Beatrix Eugster^{b, c, d }, *

^a University of Konstanz, Germany

^b University of St. Gallen, Switzerland

^c CESifo, Germany

^d CEPR, UK

ARTICLE INFO

JEL classification:

C18

C31

C51

I21

Keywords:

Measurement error

Misclassification

Spillover effect

Linear-in-means

ABSTRACT

This paper investigates how misclassification affects the estimation of peer effects in linear-in-means (LIM) models. We formally show that the peer effect estimate is biased by an "own effect" due to error in the group average and a smearing effect due to misclassification in the individual regressor. The direction of the total bias depends on how misclassification is distributed across groups. We develop a simple heuristic based on the ratio of the estimated peer effect to the estimated individual effect to help researchers identify the likely direction of the total bias. We verify the theoretical predictions in the setting of peer effects of students with learning disabilities on other students' performance.

1. Introduction

In the last two decades, the literature on spillover effects has grown constantly. A considerable portion of this literature focuses on applied work, whereas a relatively smaller body of research addresses the consequences of endogeneity. Most of this work studies endogeneity problems arising from simultaneity (or reflection problems) and nonrandom group formation (Blume et al., 2011; Manski, 1993), but only a limited number of studies formally investigate endogeneity problems stemming from measurement error in observed peer characteristics (Ammermueller and Pischke, 2009; Feld and Zölitz, 2017).¹

The few existing studies derive implications for settings in which peer variables are based on continuous characteristics measured with classical measurement error. However, much of the peer effect literature relies on binary characteristics that, if measured with error, give rise to misclassification error. The misclassification error and the classical measurement error are intrinsically different: while the latter

is mean-zero and independent of the latent variable, misclassification error is always negatively correlated with the true latent variable and does not average to zero (Aigner, 1973). Because of these fundamental differences, the existing results do not directly apply to much of the empirical literature.

This paper investigates the consequences of misclassification in the linear-in-means (LIM) model both theoretically and empirically. The theoretical analysis characterizes the biases resulting from misclassification and shows a simple heuristic based on the ratio of the estimated peer effect to the estimated individual effect that can be used in practice to infer whether the peer effect is likely underestimated or overestimated, even when misclassification rates are unknown. The empirical analysis validates the theoretical findings and illustrates how to make use of the developed heuristic in practice. By combining theoretical and empirical arguments, this study provides a comprehensive understanding of the implications of misclassification in the context of the LIM model, along with practical guidance for assessing the extent to which misclassification may pose a problem.

[☆] Special thanks go to two anonymous reviewers, the editor Bernd Fitzenberger, Josh Angrist, Giacomo De Giorgi, Albrecht Glitz, Bryan Graham, Michael Jansson, Michael Lechner, Helge Liebert, Michael Knaus, Raphaël Parchet, Michele Pellizzari, Aurélien Sallin, Stefan Sperlich, Aleksey Tetenov, and participants of the NISMES 2022 for their insightful feedback along the way. We gratefully acknowledge financial support from the Swiss National Science Foundation, Switzerland through grant no. 176381. This study was conceived and drafted when Dr. Balestra was employed at the University of St. Gallen; the findings and views in this manuscript do not necessarily reflect the official views or policy of his current employer.

* Correspondence to: Swiss Institute for Empirical Economic Research, University of St. Gallen, Rosenbergstrasse 22, CH 9000 St. Gallen, Switzerland.

E-mail addresses: fanny.valli@uni-konstanz.de (F. Valli-Puljic), simone.balestra@gmail.com (S. Balestra), beatrix.eugster@unisg.ch (B. Eugster).

¹ Moffitt (2001) and Angrist (2014) also partly address measurement error problems in peer effect estimation, but the topic is beyond the scope of both studies.

The theoretical framework is based on the simplest possible form of LIM model, including only the individual binary characteristic (e.g., having a learning disability) and its groupwise average (e.g., proportion of peers with a learning disability), where the individual characteristic is misclassified. We compute the probability limits of both coefficients under two key assumptions: (i) the latent variable and group formation are exogenous, and (ii) the measurement error is non-differential, meaning that the observed variable is not a relevant covariate once the latent variable is controlled for. Under these assumptions, we find that the peer effect estimate can be compromised by two sources of bias denoted as an “own” and a “smearing” effect, respectively.² The own effect originates from the measurement error contained in the group average. The smearing effect arises from the misclassification error contained in the individual characteristic. The two biases are linearly related and depend on two key factors: the misclassification rates and the cross-group distribution of the observed variable and its misclassification error.

Three scenarios can arise. The peer effect is underestimated when misclassification is concentrated in a small number of groups; it is overestimated when a high concentration of misclassification is prevalent across multiple groups; and it is closest to the true effect when the observed variable and its misclassification error are similarly distributed across groups. The existence of these three well-defined regions, together with the linear negative relationship between the total bias in the peer effect and the individual effect, allows us to derive a simple heuristic based on the ratio of the estimated peer effect to the estimated individual effect. This ratio enables researchers to infer the likely direction of bias even when misclassification rates and the cross-group distribution of misclassification error are unknown.

To illustrate our theoretical results, we simulate samples with varying group sizes, numbers of groups, and probabilities of the observed binary variable. The simulations show that, in settings with a large number of groups, both estimators are biased by an attenuation bias only. However, if the number of groups is small, the total bias in the peer effect (but not in the individual effect) can be substantial, leading to either under- or overestimation, and potentially even to sign reversals. The simulations also confirm that the ratio of the estimated peer effect to the estimated individual effect reliably indicates whether the peer effect is under- or overestimated in most empirically relevant settings.

We apply the theoretical findings in an empirical context of peer effects in education. This empirical context is particularly suitable, as it comprises a notable number of examples where binary characteristics are used to study peer effects, often implicitly or explicitly as a proxy variable for disruptive peers.³ We leverage data from Balestra et al. (2022) and apply their same identification strategy, which exploits cross-classroom within school-track-year variation, using three different measures of low academic ability. We select psychologists' diagnoses of learning disabilities as the true measure of low abilities, and rely on teacher referral to the School Psychological Service (SPS) and grade repetition as approximate measures of low abilities.

² The terms “own effect” and “smearing effect” are borrowed from Garber and Klepper (1980), who use this terminology to describe, respectively, the direct and indirect bias components arising from classical measurement error in multivariate linear regression models.

³ Examples of binary, predefined characteristics include gender (Anelli and Peri, 2017; Black et al., 2013; Hoxby, 2000; Lavy and Schlosser, 2011; Whitmore, 2005), race or immigration (Angrist and Lang, 2004; Cho, 2012; Diette and Uwaifo Oyelere, 2014, 2017; Gould et al., 2009; Hanushek et al., 2009; Hoxby, 2000; Hoxby and Weingarth, 2005; Jensen and Rasmussen, 2011), experimental treatment (Chetty et al., 2011; Fafchamps and Mo, 2018), high/low abilities (Lavy and Schlosser, 2011; Bietenbeck, 2020; Balestra et al., 2023; Hill, 2014; Lavy et al., 2012a,b), exposure to troubled families (Kristofersen et al., 2015; Carrell et al., 2018, 2009), change of residence (Gibbons and Telhaj, 2011; Hanushek et al., 2004), and special educational needs indicators (Fletcher, 2009, 2010; Balestra et al., 2022).

The empirical application not only confirms the theoretical predictions but also highlights an additional insight relevant for applied work. The developed diagnostic heuristic is only informative about misclassification if the model is correctly specified, that is, if our two key assumptions hold and no relevant unobserved factors are omitted. Accordingly, we recommend assessing robustness not only of the peer effect estimate, as is common in the literature, but also of the individual effect. As a practical guideline, when both estimates are stable across specifications, the ratio provides a reliable diagnostic of the direction of bias induced by misclassification.

This paper makes three main contributions. First, it extends the measurement error literature in the context of peer effects. Earlier studies focus on classical measurement error, even in settings where the variable of interest is binary, as in Ammermueller and Pischke (2009). We complement this line of work by deriving bias expressions that hold for general additive measurement error and by characterizing their specific form under misclassification, which is inherently different from classical measurement error. This setting is representative of a large share of applied peer effects research, where binary indicators and their groupwise proportions are commonly used (see Cools and Patacchini, 2021; Sacerdote, 2011).

Second, we show how the estimated ratio of the peer effect to the individual effect can be exploited to infer the direction of the bias. In contrast to the classical measurement error setting, in which this ratio is known to be unbiased (Ammermueller and Pischke, 2009), we show that, under misclassification, it follows predictable patterns relative to the true ratio. This allows us to propose a simple heuristic that helps researchers infer where a given peer effect estimate is likely to lie when misclassification is present.

Finally, we extend the literature studying the measurement error problem in linear regression models by showing the consequences of multiple variables measured with nonclassical measurement error in a multivariate linear model. To the best of our knowledge, virtually all existing studies on linear models with multiple mismeasured variables (Griliches, 1986; Garber and Klepper, 1980; Abel, 2018a,b) show the consequences of classical measurement error.

2. Theoretical framework

The first part of this section presents the mathematical derivation of the biases arising from misclassification in a LIM model, while the second part illustrates their simulated distributions.

2.1. Notation

Consider a random sample $\{(Y_{ig}, D_{ig}, W_{ig})\}_{i=1}^N$ of finite size $N < \infty$ where observations are randomly partitioned into $G \geq 2$ different groups. Individuals are indexed by $i \in I = \{1, \dots, N\}$, groups by $g \in \mathcal{G} = \{1, \dots, G\}$, and group sizes by $N_g \in [2, N - 2(G - 1)]$.⁴

Let $D_{ig} \sim \text{Be}(\pi)$ denote the *unobserved* (or latent) binary trait of individual i from group g , and let $D = (D_{ig})_{i=1, \dots, N; g=1, \dots, G}$ be the vector of individual latent traits, ordered by group. Let A be a $N \times G$ diagonal block matrix, where $\text{diag}(A) = (A_1, A_2, \dots, A_G)$, while the off-diagonal elements are 0. Each row corresponds to an individual i , while each column corresponds to a group g , hence elements' of each block, $[A_g] = \mathbb{1}\{i \in g\}$, indicate whether individual i belongs to the group $g \in \mathcal{G}$. This allows us to express the vector of group proportions of trait D_{ig} as $\bar{D} = (A' A)^{-1} A' \cdot D$, where each element of $\bar{D}$ corresponds to the group-level share $\bar{D}_g = \frac{1}{N_g} \sum_{i \in g} D_{ig}$, taking the same value for all $i \in g$. Analogously, we define the *observed* binary trait as $W_{ig} \sim \text{Be}(p)$, and its proportion in a group g as $\bar{W}_g = \frac{1}{N_g} \sum_{i \in g} W_{ig}$,

⁴ The bounds on N_g result from the assumption that each group contains at least two individuals.

with $W = (W_{ig})_{i=1,\dots,N;g=1,\dots,G}$ and $\bar{W} = (A'A)^{-1}A' \cdot W$ denoting their respective vectors.

Following previous work on measurement error in LIM models (Feld and Zölitz, 2017; Ammermueller and Pischke, 2009), we adopt the standard *linear additive measurement error model*, assuming that the observed (proxy) measure W_{ig} relates to the latent variable D_{ig} as

$$W_{ig} = D_{ig} + e_{ig}, \quad (1)$$

where $e_{ig} \in \{-1, 0, 1\}$ represents the misclassification error. Furthermore, we denote misclassification rates with $\nu = P(W_{ig} = 1 | D_{ig} = 0)$ (false positive) and $\eta = P(W_{ig} = 0 | D_{ig} = 1)$ (false negative).

Unlike the case of classical measurement error, where the error is mean-zero and uncorrelated with the latent variable, under additive misclassification error we generally have $E[e_{ig}] \neq 0$ and $Cov(D_{ig}, e_{ig}) \neq 0$, as already shown by Aigner (1973),⁵ which complicates the estimation of peer effects. We address this by deriving bias expressions for peer effect estimators under misclassification, going beyond existing analyses that are restricted to classical measurement error (Ammermueller and Pischke, 2009; Feld and Zölitz, 2017). Note that throughout the paper, we use the term measurement error when referring to results that hold for any additive measurement error (including classical measurement error) and misclassification when a result applies specifically to the binary-variable case.

2.2. Econometric setting

If D_{ig} could be observed, we would estimate the following LIM model:

$$Y_{ig} = \beta_0 + \beta_1 D_{ig} + \beta_2 \bar{D}_g + U_{ig}, \quad (2)$$

where Y_{ig} is the outcome of individual i in group g , and U_{ig} represents the unobservable error term.⁶ Because we are primarily interested in assessing the consequences of using W_{ig} instead of D_{ig} on the estimation of the two parameters, β_1 and β_2 , we consider the following ordinary least square estimators:

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} S_{WW} & S_{W\bar{W}} \\ S_{W\bar{W}} & S_{\bar{W}\bar{W}} \end{pmatrix}^{-1} \begin{pmatrix} S_{YW} \\ S_{Y\bar{W}} \end{pmatrix}, \quad (3)$$

where S_{WW} and $S_{W\bar{W}}$ denote the sample variances of W_{ig} and $\bar{W}_g$, whereas $S_{\bar{W}\bar{W}}$ is their sample covariance. To compute the probability limits of the two estimators, we make two assumptions.

Assumption 1 (*Exogenous Latent Characteristic and Random Group Assignment*). Both the unobserved trait (D_{ig}) and the group composition (A) are orthogonal to the error term U_{ig} :

$$E[U | D, A] = E[U].$$

Assumption 2 (*Non-differential measurement error*). The measurement error contained in the observed measure is **non-differential**, i.e.,

$$E[Y | D, A, W] = E[Y | D, A].$$

⁵ We formally restate these results in section A.2.3 of Online Appendix A.2.

⁶ For simplicity, we analyze a model specified with group averages instead of leave-own-out (LOO) means, which are typically expressed as $\bar{D}_{-ig} = \frac{1}{N_g - 1} \sum_{i \neq j \in g} D_{ig}$. LOO means are used in the applied peer effects literature to attenuate the correlation between the own treatment status and the group average in the case of random group assignment. However, econometrically, it makes little difference, as shown by Angrist (2014). The peer effect estimator only differs by a factor of approximately $\frac{N}{N-1}$ when using LOO means instead of group averages, where $\bar{N} = \sum_{g \in G} m_g N_g$ is the average group size, with $m_g = N_g / N$.

The first assumption implies that the latent trait is predetermined (i.e., determined prior to group formation) and that groups are randomly formed. In other words, there are no omitted variables or *correlated effects* in the sense of Manski (1993) that simultaneously influence the composition of the group and the outcome.⁷ This assumption is fundamental for the causal identification of exogenous peer effects.

The second assumption has two main implications. First, the observed measure and its measurement error are exogenous. Second, the observed measure does not contain additional information about the outcome variable beyond what is already contained in the latent variable (Carroll et al., 2006, p. 36). In other words, W_{ig} is not a relevant covariate once D_{ig} is accounted for. This assumption ensures that the bias arises only from measurement error, and not from systematic correlation between the measurement error and the outcome. Relaxing this assumption only affects the estimated individual effect but not the peer effect, as we show in Online Appendix A.4.

2.3. Bias channels

To derive the probability limits, we start from Eq. (3), substitute Y_{ig} using the population model from Eq. (2), and re-express D_{ig} and $\bar{D}_g$ in terms of W_{ig} and $\bar{W}_g$, respectively, based on Eq. (1). We then compute the relevant sample covariances and variances, obtain their probability limits and apply Slutsky's theorem to obtain the probability limits presented in Proposition 1.⁸

Proposition 1. Consider a LIM model specified as in Eq. (2). If the latter is estimated using an observed binary variable, W_{ig} , which follows an additive misclassification error process as described in Eq. (1), then under Assumptions 1 and 2, the probability limits of the two estimates take the form:

$$plim \hat{\beta}_1 = \beta_1 (1 - \delta_{e-w}) - \beta_2 \delta_{e-\bar{w}}, \quad (4)$$

$$plim \hat{\beta}_2 = -\beta_1 \delta_{e-\bar{w}} + \beta_2 (1 - \delta_{e-\bar{w}}) \quad (5)$$

where

$$\delta_{e-w} = B \frac{r_c}{r_v}, \quad \delta_{e-\bar{w}} = 0, \quad \delta_{e-\bar{w}} = -\frac{B}{r_v} \left(1 - \frac{1 - r_c}{1 - r_v} \right), \quad \delta_{e-\bar{w}} = B \frac{1 - r_c}{1 - r_v}, \quad (6)$$

with $r_v = \frac{E[Var(W_{ig}|g)]}{Var(W_{ig})}$, $r_c = \frac{E[Cov(W_{ig}, e_{ig}|g)]}{Cov(W_{ig}, e_{ig})}$, and $B = \frac{Cov(W_{ig}, e_{ig})}{Var(W_{ig})} = \nu \frac{1-\pi}{p} + \eta \frac{\pi}{1-p}$.

Interpretation. Proposition 1 suggests that estimating a LIM model using a binary, misclassified variable results in two sources of bias, the *own* and the *smearing* effect. The term $(1 - \delta_{e-\bar{w}})$ in Eq. (5) represents the own effect on the peer effect estimate, $(1 - \delta_{e-w})$ in Eq. (4) represents the own effect on the individual effect estimate, whereas $\delta_{e-\bar{w}}$ and δ_{e-w} represent the *smearing effects* on the peer effect and the individual effect estimates, respectively.

Intuitively, the own effect arises from the measurement error in the variable itself, whereas the smearing effect refers to the bias induced by measurement error in another regressor. Formally, all δ 's correspond to OLS coefficients on either W_{ig} or $\bar{W}_g$ in bivariate linear regression models where the dependent variable is either e_{ig} or $\bar{e}_g$.⁹ As such, the own effects reflect the net correlation between a variable and its own measurement error, whereas the smearing effect reflects the net

⁷ Manski (1993) refers to correlated effects as endogeneities arising from common shocks or the formation of endogenous groups (see also Bramoullé et al., 2020).

⁸ Each step is detailed in Sections A.1–A.3 of Online Appendix A.

⁹ Specifically, $e_{ig} = \delta_{e,0} + \delta_{e-w} W_{ig} + \delta_{e-\bar{w}} \bar{W}_g + r_e$ and $\bar{e}_g = \delta_{\bar{e},0} + \delta_{e-w} W_{ig} + \delta_{e-\bar{w}} \bar{W}_g + r_{\bar{e}}$.

correlation between the variable and the other regressor's measurement error.¹⁰

Notably, the smearing effect influences the peer-effect estimate $\hat{\beta}_2$, but not the individual-effect estimate $\hat{\beta}_1$. This result comes from the fact that the error in the group average, $\bar{W}_g$, originates from the error in the individual indicator W_{ig} , and not the other way around. As a consequence, $\bar{W}_g$ can carry part of the variation in e_{ig} , while W_{ig} cannot carry any of the variation in $\bar{e}_g$, which implies $\delta_{\bar{e},w} = 0$.

Importantly, the smearing effect on the peer effect estimate cannot be eliminated simply by replacing group means with leave-own-out averages, since those averages are still based on the same mismeasured individual indicator. Furthermore, it can be shown that Eqs. (4) and (5) hold for any type of additive measurement error and that Equation (4) in Feld and Zölitz (2017) is a special case, namely when the measurement error exhibits the properties of the classical measurement error.

Drivers. When examining the δ 's in (6), one can see that all biases are driven primarily by *three components*: the *overall misclassification error* (captured by B), r_v representing the proportion of the total variance in W_{ig} that comes from within-group variance, and r_c representing the fraction of the total covariance between W_{ig} and e_{ig} coming from the within-group covariance.

The term B is a monotonically increasing function of the two misclassification rates (ν and η) and, more generally, corresponds to the OLS coefficient on the observed variable in a univariate linear regression with the misclassification error as the dependent variable.¹¹ As such, B can be seen as an *accuracy measure* of the observed variable. It takes values between 0 (free of error) and 1 (zero correlation with the latent variable) if we consider only cases for which D_{ig} and W_{ig} are positively correlated.¹² Note that B appears in all δ 's in (6), implying that there is no bias in absence of any measurement error in W_{ig} .

The terms r_v and r_c are, respectively, the ratio of the average within-group variance to the total variance of W_{ig} , and the ratio of the average within-group covariance to the total covariance between W_{ig} and e_{ig} . These two parameters can be interpreted as measures of the representativeness of groups with respect to the overall population. Using the variance decomposition,¹³ we can show that r_v takes values between 0 and 1, with $r_v = 0$ meaning that groups are fully segregated, while $r_v = 1$ is a sign of perfectly identical groups. On the other hand, r_c is bounded below by 0, but can exceed 1. If groups are fully segregated in terms of the proxy variable, then $r_c = 0$. If all groups have the same misclassification error composition, then $r_c = 1$. Finally, if the correlation of W_{ig} and e_{ig} is stronger in larger groups, which have more weight in the across-group average, then r_c can exceed 1.

¹⁰ Formally, for the peer effect estimate, the own effect arises when $Cov(E[\bar{W}_g | W_{ig}], E[\bar{e}_g | W_{ig}]) \neq 0$, whereas the smearing effect arises when $Cov(E[\bar{W}_g | W_{ig}], E[e_{ig} | W_{ig}]) \neq 0$.

¹¹ A comprehensive analysis of the term B is provided in Online Appendix B.

¹² If negative correlations are allowed, then $B \in [0, 2]$, where 2 indicates perfect negative correlation.

¹³ From the variance decomposition it follows that the total variance is the sum of the between-group variance and the average within-group variance

$$\begin{aligned} \text{Var}(W_{ig}) &= \text{Var}(E[W_{ig} | g]) + E[\text{Var}(W_{ig} | g)] \Leftrightarrow \\ \text{Var}(E[W_{ig} | g]) &= \text{Var}(W_{ig})(1 - r_v). \end{aligned}$$

Similarly, the covariance decomposition states that the total covariance equals the sum of the between-group covariance and average within-group covariance:

$$\begin{aligned} \text{Cov}(W_{ig}, e_{ig}) &= \text{Cov}(E[W_{ig} | g], e_{ig}) + E[\text{Cov}(W_{ig}, e_{ig} | g)] \Leftrightarrow \\ \text{Cov}(E[W_{ig} | g], e_{ig}) &= \text{Cov}(W_{ig}, e_{ig})(1 - r_c). \end{aligned}$$

Total bias. How strongly the peer effect estimate is biased depends mainly on the between-group distribution of the misclassification error relative to the observed variable. As an illustration, consider Fig. 1, where we fix r_v at 0.75 (vertical red line), and highlight two key reference points: $r_c = 1$ (triangle), and $r_c = r_v$ (square).

If $r_c = r_v$, the ratio $(1 - r_c)/(1 - r_v)$ is exactly 1, and the peer effect estimate $\hat{\beta}_2$ is biased only by the attenuation factor $|1 - B| < 1$ and there is no smearing effect.¹⁴ In contrast, when $r_c = 1$, there is no attenuation bias on the peer effect, while the smearing effect becomes B/r_v . Note that, for the individual effect estimate $\hat{\beta}_1$, the implications are different. As stated above, the individual effect estimate will never be biased by a smearing effect. But while $r_c = r_v$ leads to the same amount of attenuation bias as for the peer effect estimate, the case with $r_c = 1$ implies an attenuation bias of $(1 - B/r_v)$ on $\hat{\beta}_1$.

Rearranging the expression for $1 - \delta_{\bar{e},w}$ reveals a linear relationship: the smearing effect and the own effect of $\hat{\beta}_2$ increase linearly together, while the own effect of $\hat{\beta}_1$ moves linearly in the opposite direction.¹⁵ Recognizing the systematic relationship between the smearing effect and the own effect allows us to predict the total bias between and beyond the two reference points.

In the region $r_c > 1$ (dotted red line), both the smearing effect and the own effect of $\hat{\beta}_2$ increase, resulting in an overestimation of the peer effect, while $\hat{\beta}_1$ is further attenuated toward zero. By contrast, when $r_c < r_v$ (dashed red line), both components of the peer effect estimate are attenuated and may even become negative, pushing $\hat{\beta}_2$ toward zero or beyond (sign reversal). At the same time, attenuation of the individual effect $\hat{\beta}_1$ is reduced. Finally, in the intermediate region $r_v < r_c < 1$ (solid red line), the peer effect estimate $\hat{\beta}_2$ remains underestimated but lies closer to the true value β_2 . In this region, $r_c > r_v$ mitigates attenuation of the peer effect, while $r_c < 1$ ensures that the smearing effect remains nonzero and partially offsets the remaining bias without inducing overestimation.

Practical implications. A central question is how this insight can be leveraged in applications, given that key components required for bias correction, such as r_c and B , are unobservable.

Note that, if $r_c = r_v$, the ratio of the peer effect to the individual effect is consistently estimable:

$$\text{plim} \frac{\hat{\beta}_2}{\hat{\beta}_1} = \frac{\beta_2}{\beta_1} = k. \quad (7)$$

Although this results generally holds under classical measurement error (Ammermueller and Pischke, 2009), in the case of binary regressors it depends on the relative magnitude of r_c and r_v . This implies, that if there exists a reasonable benchmark for the true ratio β_2/β_1 , observing $\hat{\beta}_2/\hat{\beta}_1$ is enough to infer the region in which r_c lies relative to r_v , and thus whether the peer effect is over- or underestimated.¹⁶ Since both the smearing and own effect of $\hat{\beta}_2$ increase in r_c , while the own effect of $\hat{\beta}_1$ decreases in it, the ratio $\hat{\beta}_2/\hat{\beta}_1$ exceeds the true ratio when $r_c > 1$ (overestimation), falls below it when $r_c < r_v$ (underestimation), and is close to it when $r_v < r_c < 1$.

¹⁴ Note that, if $r_c = r_v$, then $\delta_{\bar{e},w} = 0$ and $\delta_{\bar{e},w} = B$.

¹⁵ Algebraically rearranging the expression for $1 - \delta_{\bar{e},w}$, leads to the following two expressions

$$(1 - \delta_{\bar{e},w}) = (1 - B) + r_v \cdot \delta_{e,w} \quad \text{and} \quad (1 - \delta_{\bar{e},w}) = \frac{1 - B}{1 - r_v} - \frac{r_v}{1 - r_v} \cdot (1 - \delta_{e,w}),$$

indicating that the own effect component of $\hat{\beta}_2$, increases linearly in the smearing effect and decreases linearly in the own effect in $\hat{\beta}_1$.

¹⁶ A benchmark can typically be obtained from the existing peer-effects literature as we will illustrate in the empirical application.

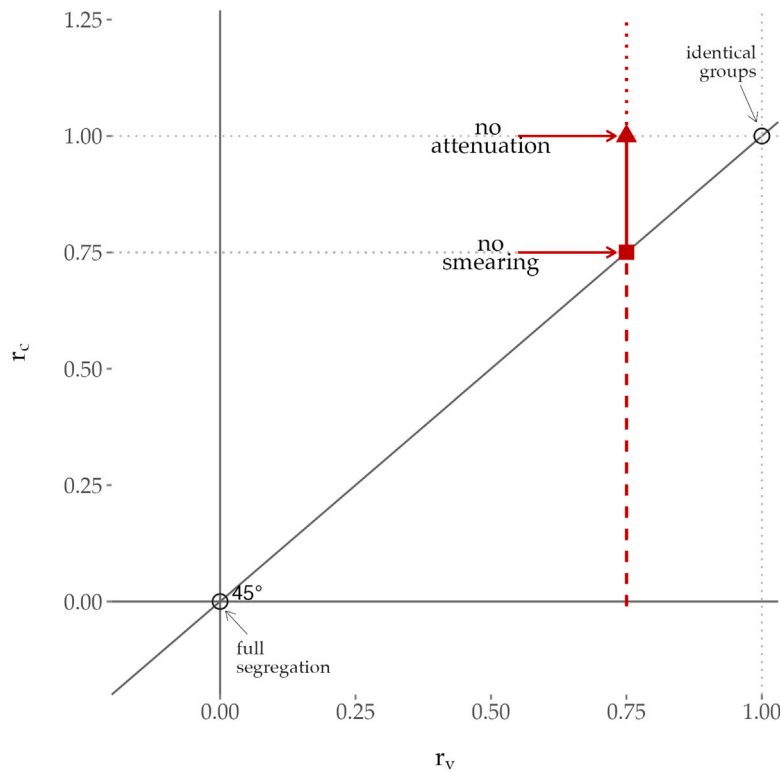


Fig. 1. r_c relative to r_v .

Note: The two empty circles denote the two extreme cases where $r_c = r_v = 0$ (full segregation) and $r_c = r_v = 1$ (perfectly identical groups). The vertical red line is fixed at $r_v = 0.75$, and the different segments denote different scenarios: $r_c > 1$ (dotted), $r_v < r_c < 1$ (solid), and $r_c < r_v$ (dashed). The triangle ($r_c = 1$) and the square ($r_c = r_v$), instead, are the two reference points. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.4. Simulations

We next illustrate these mechanisms using a simulation exercise. We generate nine samples that vary in the number and size of the groups into which observations are partitioned. The three group-size intervals are chosen to reflect those commonly observed in empirical applications, including $N_g \in [15, 35]$, $N_g \in [50, 120]$, and $N_g \in [300, 800]$.¹⁷ We label the three group-size categories as *classes*, *cohorts*, and *schools*. To capture different empirical settings, we vary the number of groups across $G = 120$, $G = 500$, and $G = 2000$. Combining the three group size intervals with the three possible values of G yields nine distinct types of samples.

For each sample, we first generate group sizes by drawing randomly G different values of N_g from the corresponding uniform distribution. We then draw a Bernoulli vector of length $N = \sum_{g=1}^G N_g$ with success probability (sample coverage) $\pi = 0.3$.¹⁸ This variable represents the latent variable D_{ig} .

For the proxy variables, we generate four additional Bernoulli vectors that differ in their success probabilities, $p = \{0.03, 0.3, 0.5, 0.8\}$, while maintaining similar correlations with the pseudo latent variable. All correlation coefficients are strictly positive and lie in the range

¹⁷ For example, these are the sizes reported for classrooms and squadrons in de Gendre and Salamanca (2020), Balestra et al. (2022), Bietenbeck (2020) and Carrell et al. (2009); for smaller cohorts and university courses such as in Hill (2014), and Booij et al. (2017); and for university lectures such as in Feld and Zölitz (2017), or for medium and large cohorts in studies using Add Health data, such as Bifulco et al. (2011) and Cools et al. (2022).

¹⁸ This value is chosen to match the coverage observed in the empirical application.

[0.24, 0.27] (see Table C.1).¹⁹ Holding the correlation with the latent variable fixed allows us to isolate the effect of variation in the success probability.

To obtain the distribution of the two biases in each sample type, we resample each sample 1000 times. In each iteration (and for each proxy indicator), we compute r_v and r_c , as well as the bias terms defined in (6).

Ratios. Mirroring Fig. 1, Fig. 2 displays the distributions of r_c against r_v across different values of p , group sizes, and numbers of groups.²⁰ Each panel corresponds to a different number of groups, each column to a different group size, and values of p are distinguished by both shape and by color, with dark triangles representing $p = 0.03$ and light circles $p = 0.3$. From Fig. 2 we draw three main lessons.²¹

First, for very low coverage levels such as $p = 0.03$, the variability of r_c is substantially higher. This arises because, when p is small, it is mechanically more likely that many groups contain no individuals with $W_{ig} = 1$, while the remaining groups contain relatively larger numbers of such individuals. Consequently, within-group covariances with the misclassification error are zero for many groups, while a few display disproportionately large covariances relative to the population, generating realizations with $r_c > 1$.

Second, this issue becomes less pronounced as the number of groups increases. With more groups, each group receives less weight in the cross-group averages, which narrows the range of possible values for

¹⁹ See Section C.1 of Online Appendix C for more details on the data-generating process.

²⁰ Note that in Tables C.2–C.5 of Online Appendix C, we also report detailed statistics on the distribution of the ratios.

²¹ Results for $p = 0.5$ and $p = 0.8$ are shown in Online Appendix C.3.

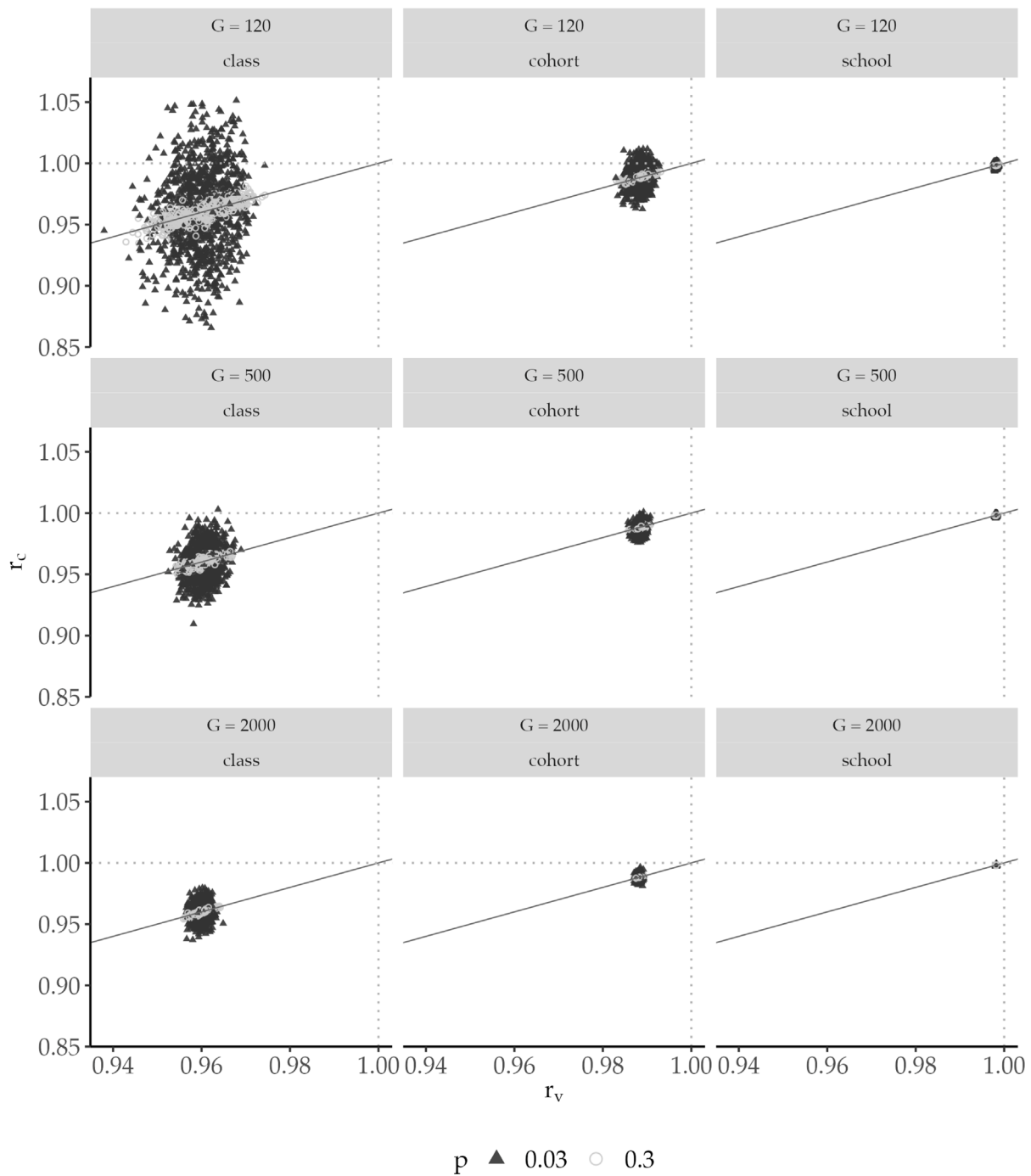


Fig. 2. r_c vs. r_v for different N_g , G , and p .

Note: The black solid line represents the 45 degree line. The dotted vertical line represents the upper bound on r_v , whereas the horizontal dotted line denotes one of the two reference points where $r_c = 1$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

these statistics. As a result, the distributions of both r_c and r_v become more tightly concentrated around the 45-degree line, regardless of group size or the proxy's success probability.

Third, as group size increases, the average values of both ratios converge to 1, while the overall patterns remain similar across the values of p (even if not visually apparent). Larger groups tend to be more representative of the population because they contain more observations. Consequently, the within-group distributions of W_{ig} and e_{ig} become more homogeneous across groups and more closely aligned with the full-sample distribution.

Biases. Fig. 3 illustrates the relationship between the smearing effect and the own effect of the peer-effect estimate $\hat{\beta}_2$. We restrict attention to groups with $N_g \in [15, 35]$, where variation is largest, and because our empirical analysis exploits cross-classroom variation. As in previous figures, we focus on $p = \pi = 0.3$ (circles), and on $p = 0.03$ (triangles), a choice that closely aligns with our empirical application.²²

²² As shown in Figure C.4 of Online Appendix C.4, similar patterns and conclusions can be drawn for $p \in \{0.5, 0.8\}$.

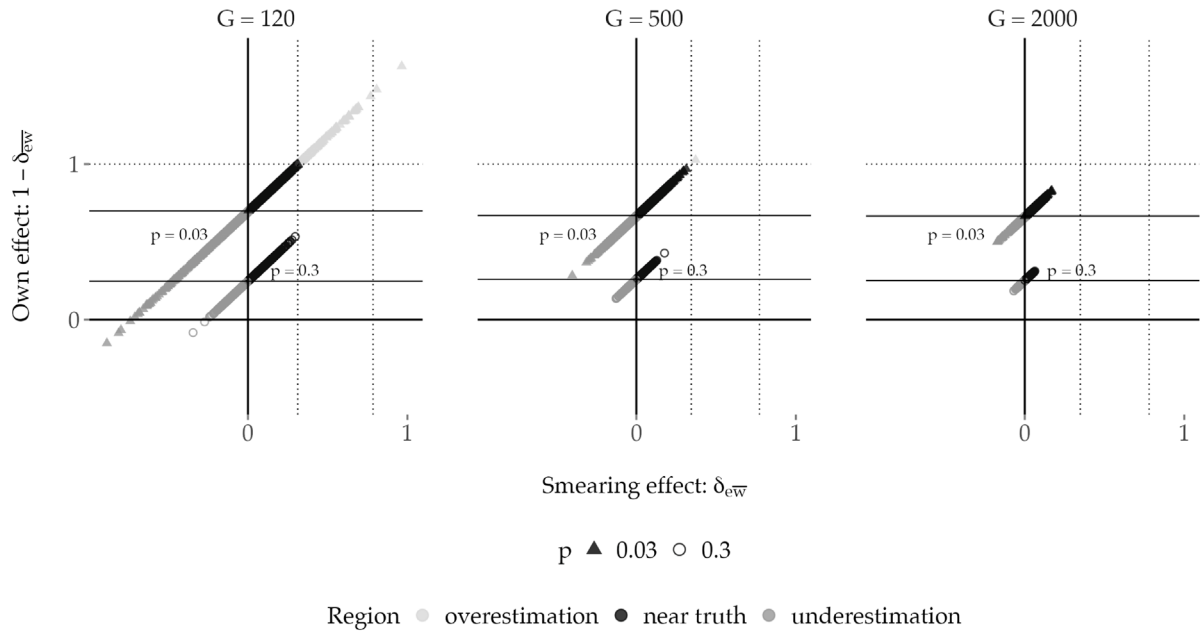


Fig. 3. Smearing vs. own effect on $\hat{\beta}_2$.

Note: The solid horizontal lines denote the reference point at which the smearing effect is zero and the own effect on the peer effect estimate is $1 - B$. The dotted lines (horizontal and vertical) represent the other reference point where the own effect on the peer effect estimate is zero whereas the smearing effect is exactly B/r_v .

The figure confirms the linear relationship between the smearing effect (x-axis) and the own effect (y-axis) for $\hat{\beta}_2$. The three regions of interest are distinguished by different shades of gray: light gray for $r_c \geq 1$ (overestimation of the peer effect), black for $r_v < r_c < 1$ (near-unbiased peer-effect estimate), and darker gray for $r_c \leq r_v$ (underestimation). The solid horizontal lines indicate the case in which the smearing effect is exactly zero and the own effect equals $1 - B$. The dotted horizontal and vertical lines mark the cases in which the own effect equals 1 and the smearing effect equals B/r_v , respectively.

The figure shows that, for very small values of p , both substantial overestimation and substantial underestimation (including sign reversals) can occur under random assignment when the number of groups is limited. These distortions become less likely as the number of groups increases. For a larger success probability, $p = 0.3$, the likelihood of overestimation is essentially zero even with relatively few groups, although substantial underestimation and sign reversals may still occur when the number of groups is small. As the number of groups increases, these distortions disappear, and the total bias converges to the attenuation bias.

Importantly, the group size does not mitigate the probability of either substantial overestimation or underestimation. The same patterns persist for larger group sizes, as shown in Online Appendix C.4.

Estimates. We conclude the simulation exercise by generating *pseudo outcomes* according to the data-generating process $y_{ig} = \beta_1 D_{ig} + \beta_2 \bar{D}_g + u_{ig}$, where $\beta_1 = -0.1$ and $\beta_2 = -0.7$, and $u \sim N(0, 1)$. These parameter values are calibrated to the estimates reported in Balestra et al. (2022), whose data we also draw on in the empirical application. We then estimate $y_{ig} = \hat{\beta}_0 + \hat{\beta}_1 W_{ig} + \hat{\beta}_2 \bar{W}_g + \hat{u}_{ig}$.²³

Fig. 4 reports, in sequence, the 1000 estimates of the individual effect ($\hat{\beta}_1$), the peer effect ($\hat{\beta}_2$), and their ratio ($\hat{\beta}_2/\hat{\beta}_1$). As before, we use triangles to denote estimates obtained using the proxy with $p = 0.03$,

while circles correspond to $p = 0.3$. Different shades of gray indicate the three regions of interest: light gray for $r_c \geq 1$, black for $r_v < r_c < 1$, and dark gray for $r_c \leq r_v$. The yellow star marks the mean across the 1000 estimates. Horizontal dotted lines indicate the true expected value for each of the parameters: $\beta_1 = -0.1$, $\beta_2 = -0.7$, and $\beta_2/\beta_1 = 7$.

For the individual effect $\hat{\beta}_1$, the estimates exhibit little variation and cluster tightly around the attenuated expected value. Sign reversal is unlikely for moderate values of B that are well below its upper bound of 1. The first panel further indicates that, as the number of groups increases, the estimate of the individual effect converges in probability to $\beta_1(1 - B)$.

The results for the peer effect estimate (middle panel) are considerably more nuanced. When the number of groups is small and $r_c \leq r_v$, sign reversals may occur for both proxy variables. With few groups and $r_c \geq 1$, substantial overestimation is also possible, although only for the variable with low success probability. If $r_v < r_c < 1$, peer effect estimates lie closer to the true value than in the other regions. For the low-coverage proxy, the smearing effect can fully offset attenuation and may even induce a slight overestimation. By contrast, for the high-coverage proxy, attenuation cannot be fully eliminated, so the peer effect remains consistently underestimated even within the “near-truth” region. In the limit, as the number of groups increases, both overestimation and sign reversal become increasingly unlikely for either proxy, and the estimated peer effect converges to the lower bound of the “near truth” region, $\beta_2(1 - B)$, as indicated by the yellow marker.

Turning to the ratio $\hat{\beta}_2/\hat{\beta}_1$ (right panel), the simulation results closely align with the theoretical predictions. When $r_c \leq r_v$, the ratio falls below its benchmark and becomes small and negative in cases of sign reversal. When $r_c \geq 1$, it takes very large positive values (around 40% or more above the benchmark), at least for the low-coverage proxy. In the intermediate region $r_v < r_c < 1$, the ratio exceeds the benchmark but remains relatively close to it, lying roughly 15% to 40% above depending on the number of groups. As the number of groups increases, the ratio converges to its benchmark, as indicated by the tightening of both segments.

These results suggest that our heuristic based on the ratio of the peer effect to the individual effect performs well in most settings, except

²³ Given that the patterns observed in Fig. 3 for small groups are very similar to those observed for larger groups in Figures C.2 and C.3, we present results for the two selected proxies with $p \in \{0.03, 0.3\}$ and small group sizes $N_g \in [15, 35]$ only.

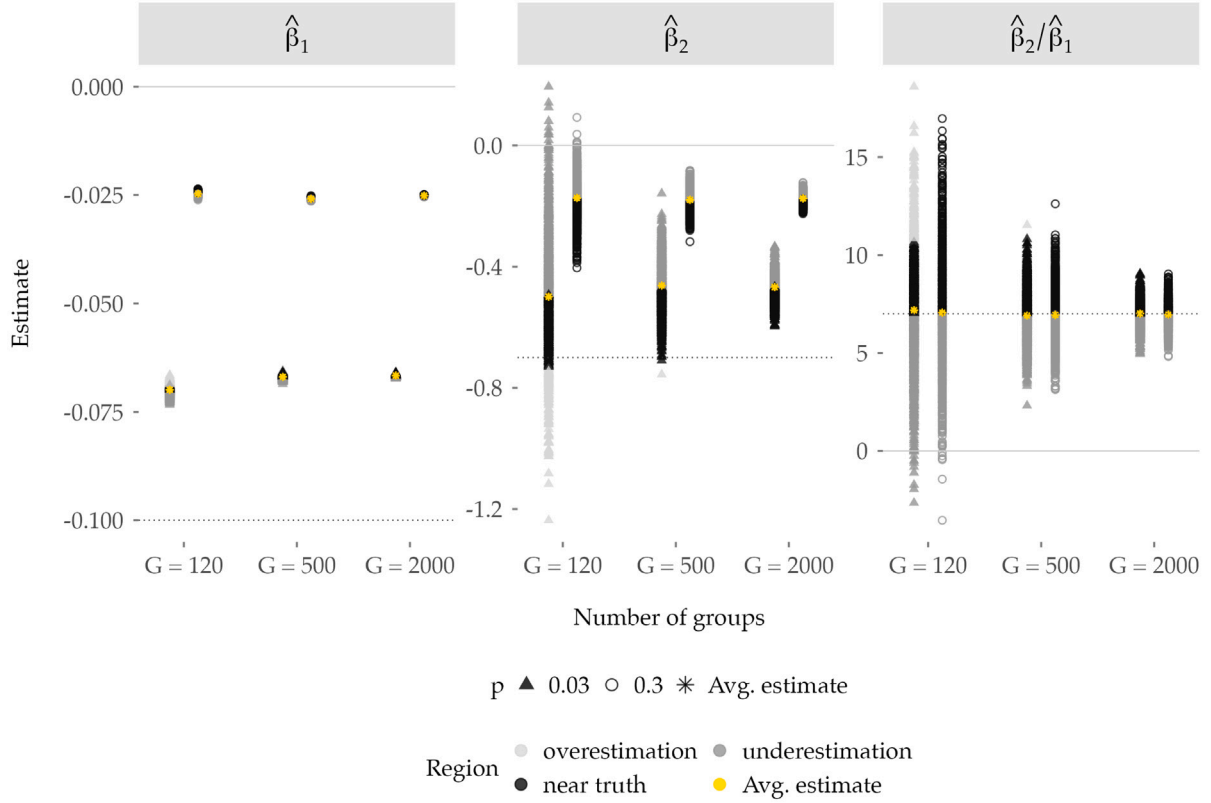


Fig. 4. $\hat{\beta}_2$, and $\hat{\beta}_2/\hat{\beta}_1$ for $N_g \in [15, 35]$.

Note: Horizontal dotted lines indicate the true expected value for each of the parameters: $\beta_1 = -0.1$, $\beta_2 = -0.7$, and $\beta_2/\beta_1 = 7$.

when the success probability is high and the number of groups is small, specifically for $p = 0.3$ and $G = 120$. In this case, the ratio $\hat{\beta}_2/\hat{\beta}_1$ can take very large values which, based on our heuristic, would suggest that the peer effect is overestimated, when in fact it is underestimated, as shown in the middle panel. A plausible explanation for this breakdown is the correlation between the proxy and the latent variable. Although the simulations are calibrated such that all proxies have similar empirical correlations with the latent variable, their maximal attainable correlations depend on p relative to π . For $p = 0.03$, a correlation of 0.24 is close to its theoretical maximum of 0.27, whereas for $p = 0.3$ the same correlation is far below the maximum of one. Consequently, for higher-coverage proxies, large values of $\hat{\beta}_2/\hat{\beta}_1$ may reflect limited proxy–latent correlation rather than genuine overestimation of the peer effect.²⁴

Practical implications. A practical way to assess whether the observed ratio is indicative of overestimation or underestimation is to compute the maximal value of the ratio within the near truth region. The maximal value is reached when $r_c = 1$, and misclassification is minimized (equivalently, when correlation with the latent variable is maximized).

²⁴ The correlation between two binary variables reaches its maximum when $P(W = 1, D = 1) = p$, if $p \leq \pi$, or when $P(W = 1, D = 1) = \pi$, if $p > \pi$. Consequently, the maximum correlation yields: $\max(0, p + \pi - 1) \leq P(W = 1, D = 1) \leq \min(p, \pi)$, and

$$\rho_{W,D}^{\max} = \sqrt{\frac{p(1-\pi)}{\pi(1-p)}} \quad \text{for } p \leq \pi \quad \text{and}$$

$$\rho_{W,D}^{\max} = \sqrt{\frac{\pi(1-p)}{p(1-\pi)}} \quad \text{for } p > \pi.$$

Formally, the upper bound is given by

$$\text{plim} \frac{\hat{\beta}_2}{\hat{\beta}_1} = \frac{k + \frac{B^{\min}}{r_v}}{1 - \frac{B^{\min}}{r_v}}, \quad (8)$$

where $B^{\min}$ is obtained from the relationship below²⁵:

$$B^{\min} = 1 - \rho_{W,D}^{\max} \sqrt{\frac{\pi(1-\pi)}{p(1-p)}}, \quad (9)$$

where $\rho_{W,D}^{\max}$ denotes the maximal feasible correlation between W_{ig} and D_{ig} (see Footnote 24). We illustrate how to implement this calculation in the empirical application.

We conclude with one further observation. The visualization of the relationship in Eq. (9) (see Online Appendix B.2) reveals that, for a fixed correlation with the latent variable, proxies with smaller probability of occurrence generate the lowest attenuation bias when $\pi < 0.5$. When $\pi > 0.5$, the opposite holds: proxies with larger p minimize attenuation bias for the same correlation, whereas at $\pi = 0.5$, both proxies with very small and very large coverage yield the smallest attenuation bias. However, extreme coverages are not inherently preferable. They are advantageous only when there is good reason to believe that their correlation with the latent variable is close to its maximum. In such cases, sparse cross-group variation can generate high values of r_c that offsets attenuation bias even under random group assignment.

²⁵ Note that $B = \text{Cov}(W_{ig}, e_{ig})/\text{Var}(W_{ig})$, implying that $1 - B = \text{Cov}(W_{ig}, D_{ig})/\text{Var}(W_{ig})$, which, if multiplied by $\sqrt{\text{Var}(W_{ig})/\text{Var}(D_{ig})}$ gives:

$$B(p, \rho_{W,D}) = 1 - \rho_{W,D} \sqrt{\frac{\pi(1-\pi)}{p(1-p)}}.$$

3. Empirical application

In this section, we apply the theoretical findings to empirical data, illustrating their practical implications and showing how to make use of the heuristics developed above. Leveraging data from Balestra et al. (2022), we study the peer effects of students with low academic abilities.

3.1. Background, data, and identification

The data contain information on the universe of middle schools from the canton of St. Gallen, the fifth-largest state in Switzerland. The education system in St. Gallen is characterized by three features that are key for the estimation. First, students are reshuffled when transitioning from primary to middle school. Second, in middle school students are assigned to either a higher track (*Sekundarschule*) or a lower track (*Realschule*). The two tracks differ in academic requirements and class size (16–24 students in the lower track and 20–24 students in the higher track), with assignment based on student performance and the primary school teacher's recommendation. Third, regardless of track, all students take a mandatory standardized test ("Stellwerk 8") in core subjects in grade eight. The test is computer-based and administered by the cantonal Department of Education.

We observe test score data for all eighth-grade students from 2008 to 2017. These data contain information on student achievement (standardized test score in math), student characteristics (date of birth, gender, and native German speaker), and composition of schools and classes in grade eight (school, track, classroom).

We merge these data with the examination records from the School Psychological Service (SPS) of the canton of St. Gallen. The SPS data include all students ever registered with the SPS through teacher referral, a psychologist diagnosis for learning disabilities, and an indicator for grade repetition.

We consider three measures of low academic ability: diagnosed learning problems, referral to the SPS, and grade repetition. Among the three variables of interest, we define learning disability diagnoses as the most accurate indicator of low ability,²⁶ while we treat referral to the SPS and grade repetition as proxies. Referrals to the SPS capture all students ever referred by their teachers to the SPS before entering middle school. The most common reasons for referral are learning-related issues and behavioral or socioemotional problems. While most referred students are subsequently diagnosed with learning disabilities, a small fraction is not. Consequently, referrals to the SPS are a less precise measure of low academic ability. Grade repetition, on the other hand, is a highly accurate indicator of low ability, but it captures only the most severe cases of learning problems.

We restrict our sample to only include regular, non-segregated schools and classrooms, and school-track-years with more than one class to ensure the necessary identifying variation, for a total of 40,626 observations and 890 school-track-years.²⁷

Panel A of Table 1 presents descriptive statistics for the three groups of students, while Panel B describes the relationship between the two proxies and the indicator of diagnosed learning disabilities, including their observed correlations, and the two misclassification rates. Panel A shows that the three measures capture different parts of the student population: 28.9% of the students were referred to the SPS, only 1.7% repeated a grade, and 24.3% were diagnosed with learning disabilities. Panel B further indicates that both correlations with the true measure are very close to their maximum. The high correlation is driven in both cases by the high fraction of true positives which amounts to 86% for referred students and 94% for repeaters.

²⁶ By the "most accurate indicator" we mean that we treat learning disability diagnoses as the true measure D_i .

²⁷ Table D.1 in Online Appendix D describes the exact sample selection.

Panels C and D of Table 1 further show that both proxies, as well as the true measure, capture a student population characterized by a higher likelihood of being male, nonnative speaker, slightly older at test, and with lower average test scores, with repeaters scoring the lowest on average. The average class size is approximately 19 students.

To estimate peer effects of low-ability students, we exploit quasi-random variation in classroom composition within school-track-year cells, following Balestra et al. (2022). Importantly, all three measures are predetermined with respect to peer-group assignment in the usual sense (Bietenbeck, 2020; Lavy et al., 2012a), thereby ruling out potential reflection problems (Manski, 1993). We estimate two versions of Eq. (2) (with and without controls):

$$y_{icst} = \hat{\beta}_0 + \hat{\beta}_1 D_{icst} + \hat{\beta}_2 \bar{D}_{cst} + \hat{\beta}_3 X_{ig} + \hat{\theta}_{st} + \hat{u}_{icst}. \quad (10)$$

where y_{icst} is the standardized math test score in grade eight of student i , from classroom c , year t , and school-track s , D_{icst} is the indicator for diagnosed learning disabilities, $\bar{D}_{cst}$ is class c 's proportion of students diagnosed with learning problems, and θ_{st} is the school-track-year fixed effect. Additionally, X_{ig} is a vector of control variables, including gender and nonnative indicators, age at test, and class size.

In line with the theoretical framework, we estimate Eq. (10) replacing the learning disability diagnosis indicator and its classroom proportion with SPS referral and grade repetition and their corresponding classroom shares.

For all three measures, we implement four standard checks to assess the validity of the identification strategy. First, we verify that there is sufficient within school-track-year variation for identification. Second, following the literature (Lavy et al., 2012b; Bifulco et al., 2011; Feld and Zölitz, 2017), we test for selection into classrooms by regressing peer proportions on predetermined individual characteristics. Third, we apply the two randomization tests proposed by Guryan et al. (2009) to assess grouping based on the peer characteristic. Fourth, we conduct a resampling-based test in the spirit of Bifulco et al. (2011) to confirm that the three characteristics are approximately randomly distributed across classrooms within school-track-year cells.

Results from all four tests are reported in Online Appendix D.2. Overall, the tests indicate that peer variation in all three characteristics is sufficiently large and approximately random across classrooms within school-track-year cells, according to the resampling-based evidence of Bifulco et al. (2011). There is no indication of grouping for grade repeaters or systematic selection on observables. By contrast, both randomization tests of Guryan et al. (2009) and the selection-on-observables test suggest possible grouping of students with learning or behavioral problems, which are more prevalent among nonnative speakers and tend to occur in slightly smaller classes. We therefore control for these characteristics in the second specification of the estimation.

3.2. Results

Table 2 reports estimates from Eq. (10). Odd-numbered columns present specifications without controls, while even-numbered columns include individual covariates and class size. The specifications with covariates primarily serve to demonstrate robustness and to show that the theoretical findings continue to hold when covariates and fixed effects are included.

Columns (1) and (2) report regression results for teacher referral without and with covariates, respectively. The estimated individual effect suggests that ever-referred students score, on average, 0.084 standard deviation (SD) lower than non-referred students. The peer effect of teacher-referred students is statistically significantly different from zero. Increasing the proportion of ever-referred students from 0 to 1 is associated with a 0.659 SD reduction in peers' test scores. Given an average class size of 19 students, this corresponds to a decrease of approximately 0.035 SD for each additional ever-referred student.

Table 1

Descriptive statistics.

Source: Own calculations. Data are from the School Psychological Service St. Gallen and the Stellwerk test service provider.

Variable	All students		Teacher-referred		Repeater		Learning disability diagnosed	
	(1)	(2)	(3)	(4)	(5)	(6)	(8)	(9)
	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.
Panel A. Variables of interest								
Teacher-referred	0.289	0.453	1	0	0.976	0.153	1	0
Learning disability diagnosed	0.243	0.429	0.897	0.304	0.959	0.198	1	0
Repeater	0.017	0.129	0.057	0.232	1	0	0.062	0.242
Panel B. Misclassification								
ρ (corr. with learning disability)			0.928		0.209			
ν (false positives)			0.040		0.001			
η (false negatives)			0.000		0.938			
Panel C. Outcome								
Standardized Math test score	0	1	-0.389	1.056	-0.745	0.874	-0.427	1.053
Panel D. Covariates								
Boys	0.497	0.500	0.577	0.494	0.510	0.500	0.574	0.495
Non-native speakers	0.144	0.351	0.182	0.386	0.177	0.382	0.182	0.386
Age at test	14.89	0.710	15.12	0.762	15.62	0.725	15.13	0.762
Class size	19.10	3.552	18.62	3.702	18.12	3.466	18.54	3.722
Observations	40,626		12,162		708		10,912	

Notes: Data contains information from 123 school-tracks, with an average number of 7.24 cohorts observed per school-track.

Table 2

Results with and without misclassification .

Source: Own calculations. Data are from the School Psychological Service St. Gallen and the Stellwerk test service provider.

	Variable of interest					
	Teacher-referred		Repeater		Learning disability diagn.	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Effects on math test score						
Individual effect	−0.084*** (0.009)	−0.085*** (0.009)	−0.159*** (0.026)	−0.069*** (0.026)	−0.094*** (0.009)	−0.092*** (0.009)
Peer effect	−0.659*** (0.115)	−0.598*** (0.112)	−0.730** (0.336)	−0.691** (0.338)	−0.733*** (0.118)	−0.677*** (0.116)
<i>Controls</i>						
Boy		0.308*** (0.008)		0.301*** (0.008)		0.308*** (0.008)
Non-native		−0.176*** (0.013)		−0.173*** (0.014)		−0.175*** (0.013)
Age at test		−0.134*** (0.006)		−0.145*** (0.006)		−0.134*** (0.006)
Class size		0.020*** (0.007)		0.022*** (0.008)		0.019*** (0.007)
Observations	40,626		40,626		40,626	
Panel B. Misclassification statistics:						
$B \left(= \nu \frac{1-\pi}{p} + \eta \frac{\pi}{1-p} \right)$	0.104		0.234			
r_c	0.969		0.974			
r_v	0.960		0.963			
$\hat{\beta}_2/\hat{\beta}_1$	7.845	7.035	4.591	10.014	7.798	7.359

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors are clustered at the school-track-year level and reported in parentheses below the coefficients. All regressions controls for school-track-year fixed effects. Data contains information from 123 school-tracks, with an average number of 7.24 cohorts observed per school-track. Each estimation controls for school-track-year fixed effects. Odd-numbered columns report results from specifications without controls, whereas even-numbered columns report results for specifications with covariates. Note that statistics in Panel B are calculated from non-rounded numbers.

Both the individual and peer effects remain remarkably stable after the inclusion of covariates.

Columns (3) and (4) report the results for grade repetition. The estimated individual effect is larger, corresponding to an average reduction of 0.159 SD in test scores. By contrast, the peer effect is comparable in magnitude to that obtained using teacher referral. In a classroom of 19 students, the presence of one additional repeater reduces peers' test scores by 0.038 SD. The inclusion of covariates substantially attenuates the individual effect, which is expected given that repeaters are systematically older at the time of testing, which implies a high, positive correlation with the covariate age at test.

For comparison, Columns (5) and (6) report regression results for learning-disability diagnoses. On average, students diagnosed with learning disabilities score 0.094 SD lower on the test, and the presence of one additional diagnosed student in a classroom of 19 is associated with a 0.039 SD reduction in peers' test scores. These estimates serve as our benchmark, against which we compare the results obtained using the two proxy variables.

Panel B reports the components underlying the two bias channels in Eq. (5), as well as the ratio of the peer effect to the individual effect, $\hat{\beta}_2/\hat{\beta}_1$. We observe a smaller value of B for teacher referral than for grade repetition, consistent with the stronger correlation reported in Table 1. For both variables, r_v is close to one, suggesting that these

characteristics are distributed across groups in a broadly representative manner. Similarly, for both teacher referral and grade repetition, r_c lies between r_o and 1, indicating a similar cross-group distribution of misclassification and the observed variable. In light of the theoretical framework, this pattern implies attenuation of both peer effect estimates, although slightly milder for grade repetition, as its low coverage allows for a stronger offset of the attenuation.

Similar conclusions emerge when examining the ratio $\hat{\beta}_2/\hat{\beta}_1$. Comparing column (1) with its benchmark in column (5) suggests slightly milder attenuation, as the observed ratio is marginally larger but very close to its benchmark. By contrast, comparing column (2) to column (6) yields a ratio below the benchmark, indicating stronger underestimation. Consistent with these predictions, both peer effect estimates based on teacher referral are smaller than the benchmark, with somewhat stronger attenuation in column (2), as anticipated. Overall, these patterns indicate that the heuristic performs well for this variable.

Applying the same comparison to grade repetition yields a more nuanced picture. The ratio in column (3) lies about 40% below its benchmark in column (5), indicating substantial underestimation, whereas the ratio in column (4) exceeds its benchmark in column (6), suggesting potential overestimation. While the theoretical prediction is confirmed when covariates are included, it does not hold in specifications without controls, as reflected in the corresponding peer effect estimates.

Variable omission and model misspecification are the primary sources of the breakdown of the heuristic. For grade repetition, the estimate of β_1 falls by more than half once controls are included, which explains the much lower ratio in column (3) despite the peer effect estimate being close to its benchmark. Because $\hat{\beta}_1$ enters the denominator, omitting relevant variables that are positively correlated with the characteristic of interest inflates the individual effect and mechanically depresses the ratio, even when attenuation of the peer effect is limited. Consequently, if Assumptions 1 and 2 hold and the model is correctly specified, the heuristic based on the ratio $\hat{\beta}_2/\hat{\beta}_1$ can be used reliably to infer over- or underestimation. If these conditions are violated, however, the heuristic may be confounded by omitted variable bias.

Bounds on $\hat{\beta}_2/\hat{\beta}_1$ in practice. Next, we illustrate how researchers can use information that is typically available in empirical applications to assess whether an estimated peer effect is likely to be over- or underestimated. As shown in the theoretical section, given a plausible value for $k = \beta_2/\beta_1$ and a rough estimate of the population prevalence of the latent variable, π , it is possible to characterize the range of values that the ratio is expected to take within the “near-truth” region for a given proxy with success probability p .

In our application, we set $k = 7.36$ based on column (6) of Table 2, but in practice a comparable benchmark can often be drawn from the existing literature. A common rule of thumb is that increasing the share of peers with a given characteristic by one student in a class of 20 (i.e., a 5% increase) generates a peer effect roughly equal to one third of the corresponding individual effect. This implies $\beta_2 \times 0.05 \approx \beta_1 \times 0.33$, yielding a benchmark ratio of approximately 6.6.

To obtain the upper bound against which the observed ratio can be compared, we use Eq. (8), which requires only a rough estimate of the true prevalence π . This could be coming, e.g., from official statistics. Using $\pi = 0.259$, we obtain $B^{\min} = 0.104$ and an upper bound of $\hat{\beta}_2/\hat{\beta}_1|_{r_c=1, B^{\min}} = 8.57$ for teacher referral, and $B^{\min} = 0.248$ with a corresponding bound of 11.24 for grade repetition.

This upper bound is particularly informative for interpreting the estimates based on grade repetition, for which the heuristic yields nontrivial predictions, while conclusions for the other proxy remain unchanged. The ratio reported in column (4) lies slightly below, but close to, the upper bound of 11.24. In the absence of information on the true peer effect, this would suggest that the estimated peer effect is close to its true value, which is indeed what we find.

Omission of relevant peer characteristics. We conclude this section by noting that peer effect estimates based on teacher referrals are more attenuated than those based on grade repetition, despite their higher prevalence and stronger correlation with the benchmark measure. A plausible explanation is omitted variable bias arising from *omitted peer characteristics*. As discussed earlier, students can be referred to the SPS for various reasons, among which behavioral problems are the second most common. This implies that peer effect estimates based on teacher referrals may partially capture the influence of peers with behavioral difficulties. If the classroom share of students with behavioral problems is a relevant control variable, positively correlated with teacher referral and negatively associated with math performance, then omitting it would lead to a stronger downward bias than that generated by attenuation alone.

Table 3 supports this conjecture. Columns (1)–(2) report the baseline estimates, while columns (3)–(4) add the classroom share of students with behavioral problems to quantify the effect of the omitted peer characteristic. Columns (5)–(6) report an auxiliary regression of the proportion of teacher referred peers on all covariates, including own teacher-referral and behavioral problems. Multiplying the corresponding coefficients from columns (3)–(4) and (5)–(6) implies an omitted variable bias of approximately +0.107 and +0.100, respectively. Since the benchmark peer effect is negative, this positive bias results in a more attenuated peer effect, explaining why estimates based on teacher referrals are more attenuated than those based on grade repetition.

This suggests that when proxy variables capture multiple dimensions, failing to control for the additional group-level dimensions can introduce bias beyond that arising from misclassification alone, with direct implications for the ratio of the two estimates. Gender provides a prominent example, as it is correlated with a wide range of individual and peer characteristics. A practical implication is that researchers should not only include the relevant proportions for these additional dimensions at the group level, as is already common in applied work, but also assess the stability of the ratio of peer to individual effect estimates across specifications. When this ratio, and not only the peer effect, remains stable across specifications, in the spirit of Altonji et al. (2005), the scope of the relevant omitted variables is reduced and the ratio becomes a more reliable object of inference.

4. Conclusions

This paper contributes to the growing literature on spillover effects by providing a comprehensive analysis of misclassification in LIM models, a topic that has received relatively little formal attention despite the widespread use of binary peer characteristics in empirical applications. We formally show that under random group formation and non-differential misclassification error, misclassification generates two distinct sources of bias: an own effect and a smearing effect. The two are positively linearly dependent, and their magnitudes depend on both misclassification rates and the relative distribution of the observed variable and its error across groups. Together, these factors can lead to substantial underestimation (including sign reversal) or overestimation of peer effects in finite samples, even under random group formation. However, whether the peer effect is over- or underestimated follows a predictable pattern that is reflected in the ratio of the peer to the individual effect. Because this ratio is always observed, it provides a simple and informative diagnostic of the likely direction of bias, even when misclassification rates or the distribution of errors across groups are unknown.

Using administrative school data, we validate the theoretical predictions and highlight an additional insight relevant for applied work. The ratio of the estimated peer to individual effect may also reflect omitted variable bias and is therefore informative about misclassification only if the model is correctly specified, that is, if our two key assumptions hold and no relevant unobserved factors are omitted. Accordingly, we recommend assessing robustness not only of the peer effect estimate, as

Table 3

Omitted variable bias.

Source: Own calculations. Data are from the School Psychological Service St. Gallen and the Stellwerk test service provider.

Outcome:	Baseline Math test score		+ beh. problems Math test score		Auxiliary reg. Prop. teacher-referred peers	
	(1)	(2)	(3)	(4)	(5)	(6)
Teacher-referred	−0.084*** (0.009)	−0.085*** (0.009)	−0.084*** (0.009)	−0.085*** (0.009)	0.034*** (0.002)	0.033*** (0.001)
Prop. teacher-referred peers	−0.659*** (0.115)	−0.598*** (0.112)	−0.690*** (0.125)	−0.627*** (0.122)		
Prop. beh. problems			0.154 (0.180)	0.144 (0.167)	0.695*** (0.049)	0.693*** (0.049)
Individual controls	No	Yes	No	Yes	No	Yes
Observations	40,626		40,626		40,626	

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors are clustered at the school-track-year level and reported in parentheses below the coefficients. All regressions controls for school-track-year fixed effects. Data contains information from 123 school-tracks, with an average number of 7.24 cohorts observed per school-track. Each estimation controls for school-track-year fixed effects. Odd-numbered columns report results from specifications without controls, whereas even-numbered columns report results for specifications with covariates.

is common in the literature, but also of the individual effect. As a practical guideline, when both estimates are stable across specifications, the ratio provides a reliable diagnostic of the direction of the bias induced by misclassification.

In contrast to classical measurement error, which implies a simple attenuation bias (Ammermueller and Pischke, 2009; Feld and Zölitz, 2017), we show that misclassification induces a richer and more complex bias structure. Although both forms of measurement error converge to simple attenuation of individual and peer effects in large samples, finite-sample behavior differs markedly in the binary-variable case, as the cross-group distribution of the misclassification error plays an additional role. This distribution has predictable implications for peer and individual effect estimates, which we exploit to develop a practical diagnostic for assessing bias. More broadly, our findings underscore the value of jointly interpreting individual and peer effect estimates when evaluating bias in models with misclassified binary regressors, a setting that is pervasive in administrative and survey data.

CRedit authorship contribution statement

Fanny Valli-Puljic: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Simone Balestra:** Writing – original draft, Supervision, Data curation, Conceptualization. **Beatrix Eugster:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Funding acquisition, Formal analysis, Conceptualization.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.labeco.2026.102887>.

Data availability

Data will be made available on request.

References

- Abel, Andrew, 2018a. Classical Measurement Error with Several Regressors. Working Paper, Wharton School of University of Pennsylvania.
- Abel, Andrew, 2018b. The effects of q and cash flow on investment in the presence of measurement error. *J. Financ. Econ.* 128 (2), 363–377.
- Aigner, Dennis, 1973. Regression with a binary independent variable subject to errors of observation. *J. Econometrics* 1 (1), 49–60.
- Altonji, Joseph G., Elder, Todd E., Taber, Christopher R., 2005. Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. *J. Political Econ.* 113 (1), 151–184.
- Ammermueller, Andreas, Pischke, Jörn-Steffen, 2009. Peer effects in European primary schools: evidence from the progress in international reading literacy study. *J. Labor Econ.* 27 (3), 315–348.
- Anelli, Massimo, Peri, Giovanni, 2017. The effects of high school peers' gender on college major, college performance and income. *Econ. J.* 129 (618), 553–602.
- Angrist, Joshua, 2014. The perils of peer effects. *Labour Econ.* 30, 98–108.
- Angrist, Joshua, Lang, Kevin, 2004. Does school integration generate peer effects? Evidence from Boston's Metco Program. *Am. Econ. Rev.* 94 (5), 1613–1634.
- Balestra, Simone, Eugster, Beatrix, Liebert, Helge, 2022. Peers with special needs: Effects and policies. *Rev. Econ. Stat.* 104 (3), 602–618.
- Balestra, Simone, Sallin, Aurélien, Wolter, Stefan, 2023. High-ability influencers? The heterogeneous effects of gifted classmates. *J. Hum. Resour.* 2 (58), 633–665.
- Bietenbeck, Jan, 2020. The long-term impacts of low-achieving childhood peers: Evidence from Project STAR. *J. Eur. Econ. Assoc.* 18 (1), 392–426.
- Bifulco, Robert, Fletcher, Jason, Ross, Stephen, 2011. The effect of classmate characteristics on post-secondary outcomes: Evidence from the Add Health. *Am. Econ. J.: Econ. Policy* 3 (1), 25–53.
- Black, Sandra, Devereux, Paul, Salvanes, Kjell, 2013. Under pressure? The effect of peers on outcomes of young adults. *J. Labor Econ.* 31 (1), 119–153.
- Blume, Lawrence, Brock, William, Durlauf, Steven, Ioannides, Yannis, 2011. Chapter 18 - Identification of social interactions. In: Benhabib, Jess, Jackson, Matthew, Bisin, Alberto (Eds.), *Handbook of Social Economics*, vol. 1B, Elsevier, Amsterdam, pp. 853–964.
- Booij, Adam, Leuven, Edwin, Oosterbeek, Hessel, 2017. Ability peer effects in university: Evidence from a randomized experiment. *Rev. Econ. Stud.* 84, 547–578.
- Bramoullé, Yann, Djebbari, Habiba, Fortin, Bernard, 2020. Peer effects in networks: A survey. *Annu. Rev. Econ.* 12 (1), 603–629.
- Carrell, Scott, Fullerton, Richard, West, James, 2009. Does your cohort matter? Measuring peer effects in college achievement. *J. Labor Econ.* 27 (3), 439–464.
- Carrell, Scott, Hoekstra, Mark, Kuka, Elira, 2018. The long-run effects of disruptive peers. *Am. Econ. Rev.* 108 (11), 3377–3415.
- Carroll, Raymond, Ruppert, David, Stefanski, Leonard, Crainiceanu, Ciprian, 2006. *Measurement Error in Nonlinear Models: A Modern Perspective*, second ed. In: Chapman & Hall, CRC Press.
- Chetty, Raj, Friedman, John, Hilger, Nathaniel, Saez, Emmanuel, Whitmore Schnazebach, Diane, Yagan, Danny, 2011. How does your kindergarten classroom affect your earnings? Evidence from Project STAR. *Q. J. Econ.* 126 (4), 1593–1660.
- Cho, Rosa Minhyo, 2012. Are there peer effects associated with having English language learner (ELL) classmates? Evidence from the Early Childhood Longitudinal Study Kindergarten Cohort (ECLS-K). *Econ. Educ. Rev.* 31 (5), 629–643.
- Cools, Angela, Fernández, Raquel, Patacchini, Eleonora, 2022. The asymmetric gender effects of high flyers. *Labour Econ.* 79, 102287.
- Cools, Angela, Patacchini, Eleonora, 2021. Chapter 10 - Peer effects in education. In: McCall, Brian (Ed.), *The Routledge Handbook of the Economics of Education*, first ed. Routledge, London, pp. 253–275.
- de Gendre, Alexandra, Salamanca, Nicolás, 2020. On the Mechanisms of Ability Peer Effects. Iza discussion paper no. 13938, Institute of Labor Economics.
- Diette, Timothy, Uwaifo Oyelere, Ruth, 2014. Gender and race heterogeneity: the impact of students with limited English on native students' performance. *Am. Econ. Rev.: Pap. Proc.* 104 (5), 412–417.
- Diette, Timothy, Uwaifo Oyelere, Ruth, 2017. Do limited English students jeopardize the education of other students? Lessons from the North Carolina public school system. *Educ. Econ.* 25 (5), 446–461.
- Fafchamps, Marcel, Mo, Di, 2018. Peer effects in computer assisted learning: Evidence from a randomized experiment. *Exp. Econ.* 21 (2), 355–382.
- Feld, Jan, Zölitz, Ulf, 2017. Understanding peer effects: on the nature, estimation, and channels of peer effects. *J. Labor Econ.* 35 (2), 387–428.
- Fletcher, Jason, 2009. The effects of inclusion on classmates of students with special needs: the case of serious emotional problems. *Educ. Financ. Policy* 4 (3), 278–299.
- Fletcher, Jason, 2010. Spillover effects of inclusion of classmates with emotional problems on test scores in early elementary school. *J. Policy Anal. Manag.* 29 (1), 69–83.

- Garber, Steven, Klepper, Steven, 1980. Extending the classical normal errors-in-variables model. *Econometrica* 48 (6), 1541–1546.
- Gibbons, Stephen, Telhaj, Shqiponja, 2011. Pupil mobility and school disruption. *J. Public Econ.* 95 (9), 1156–1167.
- Gould, Eric, Lavy, Victor, Paserman, Daniele, 2009. Does immigration affect the long-term educational outcomes of natives? Quasi-experimental evidence. *Econ. J.* 119 (540), 1243–1269.
- Griliches, Zvi, 1986. Chapter 25 – Economic data issues. In: Griliches, Zvi, Intriligator, Michael (Eds.), In: *Handbook of Econometrics*, vol. 3, Elsevier, pp. 1466–1514.
- Guryan, Jonathan, Kroft, Kory, Notowidigdo, Matthew J., 2009. Peer effects in the workplace: Evidence from random groupings in professional golf tournaments. *Am. Econ. J.: Appl. Econ.* 1 (4), 34–68.
- Hanushek, Eric, Kain, John, Rivkin, Steven, 2004. Disruption versus Tiebout improvement: The costs and benefits of switching schools. *J. Public Econ.* 88 (9–10), 1721–1746.
- Hanushek, Eric, Kain, John, Rivkin, Steven, 2009. New evidence about Brown v. Board of Education: the complex effects of school racial composition on achievement. *J. Labor Econ.* 27 (3), 349–383.
- Hill, Andrew, 2014. The costs of failure: negative externalities in high school course repetition. *Econ. Educ. Rev.* 43, 91–105.
- Hoxby, Caroline, 2000. Peer Effects in the Classroom: Learning from Gender and Race Variation. NBER Working Paper No. 7867, National Bureau of Economic Research.
- Hoxby, Caroline, Weingarth, Gretchen, 2005. School Reassignment and the Structure of Peer Effects. NBER Conference Paper, National Bureau of Economic Research.
- Jensen, Peter, Rasmussen, Astrid Würtz, 2011. The effect of immigrant concentration in schools on native and immigrant children's reading and math skills. *Econ. Educ. Rev.* 30 (6), 1503–1515.
- Kristoffersen, Jannie Helene Grøne, Krægpøth, Morten Visby, Nielsen, Helena Skyt, Simonsen, Marianne, 2015. Disruptive school peers and student outcomes. *Econ. Educ. Rev.* 45, 1–13.
- Lavy, Victor, Paserman, Daniele, Schlosser, Analia, 2012a. Inside the black box of ability peer effects: evidence from variation in the proportion of low achievers in the classroom. *Econ. J.* 122 (559), 208–237.
- Lavy, Victor, Schlosser, Analia, 2011. Mechanisms and impacts of gender peer effects at school. *Am. Econ. J.: Appl. Econ.* 3 (2), 1–33.
- Lavy, Victor, Silva, Olmo, Weinhardt, Felix, 2012b. The good, the bad, and the average: evidence on ability peer effects in schools. *J. Labor Econ.* 30 (2), 367–414.
- Manski, Charles, 1993. Identification of endogenous social effects: The reflection problem. *Rev. Econ. Stud.* 60 (3), 531–542.
- Moffitt, Robert, 2001. Chapter 3 - Policy interventions, low-level equilibria and social interactions. In: Durlauf, Steven, Young, Peyton, Ansell, Christopher (Eds.), In: *Social Dynamics*, vol. 20, Brookings Institution Press, pp. 45–82.
- Sacerdote, Bruce, 2011. Chapter 4 — Peer effects in education: how might they work, how big are they and how much do we know thus far? In: Hanushek, Eric, Machin, Stephen, Woessmann, Ludger (Eds.), In: *Handbook of the Economics of Education*, vol. 3, Elsevier, pp. 249–277.
- Whitmore, Diane, 2005. Resource and peer impacts on girls' academic achievement: evidence from a randomized experiment. *Am. Econ. Rev.: Pap. Proc.* 95 (2), 199–203.