

# Semiparametric Estimation of Peer Effects in Classrooms: Evidence for Brazilian Schools in 2003<sup>1</sup>

Cristine Campos de Xavier Pinto<sup>◇</sup>

INITIAL DRAFT: December 2006

THIS DRAFT: February 17, 2010

## Abstract

This paper proposes a semiparametric methodology to estimate an education production function with peer effects. The advantage of this methodology is that we estimate and identify peers effects under weak assumptions, in particular without imposing a functional form for the production function. We assume that student's achievement is a function of student's quality and peer's quality. Student's quality is defined as a linear combination of student's characteristics, and peer's quality is a symmetric function of this single index in each classroom. Peer effects are identified as the marginal derivative of the production function in relation to peer's quality. We propose a three step procedure to estimate the objects of interest. In the first step, we use a generalized version of the rank regression proposed by Abrevaya (2000) to estimate the parameters inside the index that defines quality. In the second and third steps, we use the control function approach proposed by Newey, Powell and Vella (1999). This methodology is applied to estimate peer effects in the last year of elementary school in Brazil. Using how students were allocated to classrooms as a vector of instruments, we find evidence that peer effects are positive for the all the students, independent of their own quality. In addition, students with an average quality have a higher marginal benefit from peer's quality than a low quality students. The results also show that student's achievement is monotonically increased with student's quality.

JEL CLASSIFICATION: C14, C31, I20, I21

KEY WORDS: Peer effects, semiparametric approach, rank estimator, control function

---

<sup>1</sup>I would like to thank you Bryan Graham for advice and support. I am grateful to Jim Powell for helpful discussion and suggestions. Participants at Berkeley Econometrics Reading Group, Berkeley Development Lunch and FaCE seminar provide helpful comments. I thank José Francisco Soares and the Consórcio de Informações Sociais (CIS) for providing access to the data. All the usual disclaimers apply.

<sup>◇</sup> FACE/Cedeplar/UFMG

Avenida Antonio Carlos, 6627 , Sala 3121- Campus Pampulha - CEP 31270-901 - Brazil

E-MAIL: [cristine@cedeplar.ufmg.br](mailto:cristine@cedeplar.ufmg.br).

WEB: <http://cristine.berkeley.edu>.

## 1 Introduction

The effect of peer characteristics on student achievement has been a subject of interest in social science for a long time. Many authors emphasize the importance of having a better understanding of the education production function, and of the relative importance of each input into this function. In economics, the interest in peer effects is due to the large social multipliers that can be created from student interactions in classrooms. For example, in a classroom, the achievement of students can improve due to interaction with their classmates; students can teach one another, improving their performances. Convincing evidence of peer effects also helps to define the optimal way to allocate students to classrooms. Social programs try to integrate students with different characteristics into classrooms as a way of improving student outcome and behavior. For example, in the US, given the large disparities between achievement of white and nonwhite students, desegregation is pointed to as a way to raise nonwhite achievement and decrease the gap between white and nonwhite performances in school. However, there are many theoretical models that argue that the optimal way to improve students achievement is to segregate students by type (Lazear, 2001).

It is hard to find convincing evidence of peer effects in classrooms. The baseline model for estimating peer effects is the linear-in-means model. The reduced form of this model associates a student outcome with his own characteristics and the characteristics of his peers. Characteristics of peers are equal to the average characteristics of the students in the group. This model has many shortcomings. As pointed out by many authors (Manski, 2003; Moffit, 2001), there are many ways in which endogeneity can arise in this model. The most common one is that individuals can self-select into groups, creating "endogenous membership". Parents can put pressure on schools to assign their children to classrooms with students with similar backgrounds. This self-selection of children to classrooms generates unobservable variables that are correlated with peer characteristics. Endogeneity can also arise from measurement errors in the variable that measures peer characteristics (Sacerdote, 2001). Another drawback of the linear-in-means model is its linearity. In this model, the outcome is affected linearly by the mean of the characteristics of the students in the group. If students are reallocated to different groups, but the mean of student characteristics in each group is kept fixed, the peer effects are zero even though the distribution of peers inside each group has changed. Moreover, if the mean of student characteristics in one group is increased, it must be decreased in another group. In this

linear-in-means model, the gains in one group are compensated by the loss in another group. No matter how peers are allocated among groups, the total achievement will be the same.

This paper develops a semiparametric methodology to estimate the education production function and peer effects. This methodology generalizes the linear-in-means model in two different ways. First, it estimates peer effects without imposing a functional form for the production function; student outcome varies with peer characteristics in nonlinear ways. In addition, it deals with endogeneity using a control function approach. This way of dealing with endogeneity is not new in the literature. Many papers use an instrumental variable approach to deal with endogeneity (Cooley, 2006; Evans, Oates and Schwab, 1992). As pointed out by Arnott and Rowse (1987), the effect of peer characteristics on student outcome depends on the curvature of the education production function. Different from parametric models, this methodology provides a way to estimate peer effects free from the bias caused by assuming the wrong functional form for the production function. It estimates peer effects controlling for endogeneity, but with few parametric restrictions on the format of the production function.

The production function estimated in this paper has peer quality and student quality as inputs. Student quality is defined as a linear combination of student characteristics (sex, race, family background, etc.) and peer quality is a symmetric function of this single index in the classroom. Peer effects are identified as the average marginal derivative of the education production function with respect to peer quality. They represent the marginal effect of an increase in peer quality on student achievement, while keeping the other inputs fixed. In the semiparametric model proposed in this chapter, peer effects can vary with student quality. This is an important advantage of this model in relation to the linear-in-means model, since the marginal benefit of being in a high quality classroom can vary by student quality.

To identify and estimate the production function and its derivative, we propose a three step procedure. In the first step, we estimate the parameters inside the linear index that defines student quality, using a generalized version of the rank estimator proposed by Abrevaya (2000). The main assumption is that student achievement is monotonic in student quality, which implies that students with high quality should have a higher outcome than students with low quality. Under this monotonicity assumption, the rank of student quality will be correlated with the rank of student achievement in a classroom, and the rank estimator will maximize this correlation in all the groups. In this first step, we use the within

classroom variation to estimate student quality. Peer quality is defined as a symmetric function of this index in the classroom.

The second and third steps deal with the endogeneity presented in models that estimate peer effects. Parents can self-select their children into classrooms based on preferences, student characteristics, teacher characteristics, etc. Even controlling for classroom and school characteristics, there are unobservable components that are correlated with peer quality. To deal with the endogeneity of peer quality, we assume the existence of suitable instrumental variables and follow the control function approach proposed by Newey, Powell and Vella (1999).

There is a growing literature that proposes alternative methods to estimate peer effects that overcome the shortcomings of the linear-in-means model. Based on the framework of the linear-in-means model, Graham (2008) develops a method that estimates the magnitude of social interactions based on excess variance contrast. Using the data from the Tennessee class size reduction experiment, Project Star, this method is applied to study the variation in achievement of students in kindergarten. The author shows that peer characteristics explain a large amount of the individual-level variation of test scores. Sacerdote (2001) demonstrates the importance of peer effects using data from an experiment in which individuals were randomly assigned to their peers. Recently, other nonparametric and semiparametric methods have been proposed. Graham, Imbens and Ridder (2009) develop a model to estimate the effect on the average outcome of changing the allocation rule of an input among heterogeneous firms, while keeping the marginal distribution of the other inputs fixed. In another paper, Graham, Imbens and Ridder (2008) propose a structural model in which there are two types of individuals, "low" and "high". In this model, individual outcomes are a function of the fraction of high types in their social groups. Within this framework, the authors identify measures of social spillovers that are estimated using nonparametric methods. In both paper, the authors don't deal with endogeneity, but they use the selection on observables approach. They assume that there is a vector of latent variable such that controlling for these variables, the unobservable components are independent of peer and student characteristics. Cooley (2006) uses a flexible parametric model to estimate the education production function and peer effects. Using a quantile regression, she estimates a structural function that relates student achievement with peer characteristics and peer achievement. The quantile regression allows student achievement, at different points in its distribution, to respond differently to peer characteristics. In a recent paper, Duflo, Dupas and Kremer (2009) estimates the benefits of tracking, using a

random experiment in Kenya. The authors find that students at all levels of the achievement's distributions benefit from tracking. Students on tracking schools score 0.14 higher on average than students in nontracking schools.

The key contribution of this new methodology is to provide a method to estimate the education production function and peer effects under weak assumptions about the functional form of this function and the distribution of unobservable components. In addition, using a single index to define peer quality, we can use a flexible definition of peer quality (in the sense that we can include as many peer characteristics as we want to define quality of the peers), while the dimensionality of the model is kept under control. One of the drawbacks of this methodology is that since we don't impose enough structure in our model, we cannot predict the optimal way to allocate students to classrooms. Graham, Imbens and Ridder (2008) find the optimal way to allocate students to classrooms in a structural model in which there are two types of individual, "low" and "high". In this model, there are two different production functions depending on the type of individual, and one of the inputs in these production functions is the percentage of high types in the social group of the individual. These authors show that the allocation rule that maximizes individual's outcome is a function of the curvature of the production function and of the distribution of types in the population. Since we don't assume a specific format for the production function, our estimators are unable to directly guide any reallocation policy. However, this model will be able to tell which students benefit more from their peers, since in this model peer effects vary with student quality. This model provides insights into the benefit of aggregating good students or bad students into good classrooms, and this information can help the design of reallocation policies within a school. Another drawback of this methodology is that, with the production function unrestricted, we cannot obtain estimators that converge at the standard parametric rate. Assuming that objective function maximized in the first stage is very smooth, the rate of convergence of our estimators can get close to the standard rates for a series estimator.

This new methodology is applied to estimate peer effects in classrooms in Brazil. We use data provided by the Brazilian National Evaluation System of Basic Education (SAEB) in 2003 to estimate the average production function and peer effects for students in the last year of elementary school in public and private schools. We estimate these parameters using the semiparametric methodology developed, and for comparison purposes, we estimate a linear-in-means model. Using the way students were allocated to classrooms as a vector of instruments, peer quality as the average of the single index in a classroom and a production

function that relates math test scores with student and peer quality, we find peer effects that are positive for all students, except the ones at the bottom of the distribution of achievement. In addition, the average production function increases monotonically with student quality, although the relationship between peer quality and student achievement is not clear. It appears that test scores of average students increase with peer quality, but test scores of high quality students decrease with peer quality in very low quality groups. The results obtained using the semiparametric methodology highlight the limitations of the linear-in-means model, namely, that it does not allow peer effects to vary with student quality, and only captures the average effect.

This paper is divided in seven sections. Section 2 describes the general model considered in this paper, and describe the estimands of interest, the average production function and peer effects. In the next section, we describe the assumptions necessary to identify these estimands, and in section 4 we state our estimation procedure. In section 5, we discuss the rate of convergences and the asymptotic distribution of the estimators. In section 6, we present applied exercise in which we estimate the average production function and peer effects for students in the last year of elementary school in Brazil in 2003. The last section summarizes our main findings and presents some possible extensions.

## 2 Model

The baseline model for estimating peer effects is the linear-in-means model. The reduced form of this model relates the outcome of the individual with his own characteristics and the characteristics of his peer group

$$Y_{ig} = X'_{ig}\beta + \bar{X}'_g\gamma + \alpha_g + \varepsilon_{ig}$$

where  $Y_{ig}$  is the outcome of individual  $i$  in group  $g$ ;  $X_{ig}$  represents the characteristics of the individual,  $\bar{X}_g$  is the mean of these characteristics in group  $g$ ,  $\varepsilon_{ig}$  represents the individual heterogeneity and  $\alpha_g$  represents the correlated effects at the level of the group. The two elements,  $\alpha$  and  $\varepsilon$ , are not observable. Many authors (Brock and Durlauf (2000), Glaeser and Scheinkman (2002), Graham (2004)) derive this equation as a social equilibrium in a decision problem in which the individual maximizes his utility, which depends on his own characteristics, a common environment shared by the individuals in his peer group, and the stock of social capital in his peer group. Another way to interpret this equation is

to consider the individual's characteristics, the environment shared by the group and the stock of social capital in group  $g$  as inputs into the production function of the outcome  $Y_{ig}$  (Becker and Murphy (2000), Lazear (2001), Cooley (2006)).

We generalize the linear-in-means model by estimating this production function without imposing a functional form. Peer effects are characterized by the marginal effect of peer characteristics on the outcome of interest. Our model assumes that there is an education production function  $H(\cdot)$  that relates a student's achievement with his own quality and the quality of his peer

$$Y_{ig} = H(Q_{ig}, Q_g, \alpha_g, \varepsilon_{ig})$$

where  $Q_{ig}$  represents the quality of student  $i$  in group  $g$ ;  $Q_g$  is the quality of group  $g$ . In this model, a group is defined as a certain classroom in certain school. We impose a "single index" restriction that says that student quality is a linear combination of student characteristics, and peer quality is a symmetric function this index in group  $g$ . Let  $M_g$  represents the number of individuals in group  $g$ <sup>2</sup>,  $Q_{ig}(\beta) = X'_{ig}\beta$  and  $Q_g(\beta) = f(X'_{1g}\beta, X'_{2g}\beta, \dots, X'_{M_g g}\beta)$ . The vector  $X$  includes characteristics of the students, such as sex, age, family background, etc. This "single index" assumption is a reduction dimensionality that simplifies identification and estimation in a model that does not impose a functional form for the production function and includes peer characteristics as one of its inputs. This assumption allows us to use a flexible definition of peer quality in the sense that we can use as many characteristics as we need to define peer quality, but keep the dimensionality of the vector of inputs inside this production function under control. For simplicity, we define  $Q_{ig} \equiv Q_{ig}(\beta)$  and  $Q_g \equiv Q_g(\beta)$ .

Following Graham and Hahn (2003), we understand this model as a quasi-panel, where the dimension that goes to infinity is the number of groups, defined as classrooms, and the dimension that is fixed is the number of students in each classroom. To be able to identify and estimate the parameters of interest, we impose some restrictions:

---

<sup>2</sup>Some applied works use the "leave-own-out" mean,  $X_{g(-j)} = \frac{\sum_{j \neq i} X_{jg}}{M_g - 1}$ , in which the mean is computed without taking in account the quality of student  $j$ ; and  $X_{g(-j)}$  is used in the place of  $X_g$  in the expression for  $Y_{jg}$ . When  $M_g = 2$ , using the "leave-own-out" is the same as using the total mean in our methodology. Some authors (Graham and Hahn, 2003) argue that the "leave-own-out" is appropriate for small groups, when all members are observed.

For large groups, the difference between the "leave-own-out" and the mean used in the main text is irrelevant. In this methodological paper, we choose not to use the "leave-own-out" for simplicity. The methodology proposed in this paper can be applied to the "leave-own-out" mean case with mild modifications.

**Assumption 2.1** (ADDITIVITY) :  $H$  is additively separable <sup>3</sup>,

$$Y_{ig} = H(Q_{ig}, Q_g) + \alpha_g + \varepsilon_{ig} \text{ for all } i = 1, \dots, M_g; g = 1, \dots, N$$

where  $Q_{ig} = X'_{ig}\beta$  and  $Q_g = \bar{X}'_g\beta$ .

This assumption is restrictive, but it simplifies identification and estimation in the semi-parametric case. By imposing the restriction that unobservable components are additive in this model, we assume that achievement is a linear function of unobservable components, and there are no interactions between the unobservable and observable components.

**Assumption 2.2** (SYMMETRY):  $f(X_{1g}, X_{2g}, \dots, X_{M_gg})$  can be written as

$$f(X'_{1g}\beta, X'_{2g}\beta, \dots, X'_{M_gg}\beta) = f(X_{1g}, X_{2g}, \dots, X_{M_gg})' \beta$$

where  $f(X_{1g}, X_{2g}, \dots, X_{M_gg})$  is a symmetric function with  $\mathbb{E}[\|f(X_{1g}, \dots, X_{M_gg})\|^2] < \infty$ .

This assumption restricts the set of functions that are covered in this paper, but it simplifies the derivation of the asymptotic properties of the estimator. This assumption covers the standard case of the literature in which  $Q_g(\beta)$  is just the average of the singly index, and it allows us to go away from representing peer effects as just averages. For example, it allows us to use the standard deviation of  $X$ ,  $Q_g(\beta) = \frac{1}{M_g} \sum_{i=1}^{M_g} (X_{ig} - \bar{X}_g)' \beta$ .

**Assumption 2.3** (STRICT EXOGENEITY CONDITIONAL ON  $\alpha_g$ ) :

$$\mathbb{E}[\varepsilon_{ig} | Q_{1g}, \dots, Q_{M_gg}, \alpha_g] = 0 \text{ for all } i = 1, \dots, M_g; g = 1, \dots, N$$

**Assumption 2.4** (CONTROL FUNCTION) : There is a vector of instrumental variables  $A_g$  such that,

$$Q_g = \Psi(A_g, W_g) + v_g$$

where

- $W_g$  represent the characteristics of group  $g$

---

<sup>3</sup>As pointed by Rau (2006), this assumption is very restrictive, since it imposes the constraint that  $\frac{\partial H(Q_{ig}, Q_g, \alpha_g, \varepsilon_{ig})}{\partial Q_{ig}} = \frac{\partial H(Q_{jg}, Q_g, \alpha_g, \varepsilon_{ig})}{\partial Q_{jg}}$  for all pairs of  $(i, j)$  and that  $\frac{\partial H(Q_{ig}, Q_g, \alpha_g, \varepsilon_{ig})}{\partial Q_g} = \frac{\partial H(Q_{jg}, Q_g, \alpha_g, \varepsilon_{ig})}{\partial Q_g}$  for all  $g$ . This model does not allow interactions between the error and the covariates, which means that it does not incorporate the "sorting effect".



- $\mathbb{E}[v_g | A_g, W_g] = 0$
- $\mathbb{E}[\alpha_g | v_g, A_g, W_g] = E[\alpha_g | v_g, W_g]$

Under these assumptions<sup>4</sup>:

$$\begin{aligned}\mathbb{E}[\alpha_g | Q_g, A_g, W_g] &= \mathbb{E}[\alpha_g | v_g, A_g, W_g] \\ &= \mathbb{E}[\alpha_g | v_g, W_g] = \Gamma(v_g, W_g)\end{aligned}$$

Assumption 2.3 imposes that  $Q_{ig}$  is strictly exogenous conditional on the correlated effects,  $\alpha_g$ . In other words, conditional on the correlated effects, individual quality is uncorrelated with individual heterogeneity. This assumption is standard in the panel data literature, and is called strict exogeneity conditional on a latent variable. In this literature, it is common to assume that all the endogeneity comes from the fact that unobservable group components are correlated with the covariates<sup>5</sup>. In assumption 2.4, we state the endogeneity problem of peer effects models. This assumption establishes that the unobservable correlated effect is correlated with peer quality,  $\mathbb{E}[\alpha_g | Q_g] \neq 0$ . As emphasized before, this correlation can come from the fact that students are usually non-randomly assigned to classrooms. Students can be assigned to classrooms based on teacher and parents preference, on unobservable characteristics of the students and classrooms, etc. In this model, we deal with endogeneity by following a control function approach. We assume that there is a vector of instruments ( $A_g$ ) that allow us to separate out the endogenous part of peer quality from the exogenous one. Assumption 2.4 imposes that the conditional expectation of  $\alpha_g$  is independent of the vector of instruments, but dependent on the control variable ( $v_g$ ) and the characteristics of the group ( $W_g$ ). In other words, the correlated effects ( $\alpha_g$ ) are related to peer quality only through  $v_g$  and  $W_g$ . Notice that this assumption is different than the usual exclusion assumption in a control function approach. In this case, the conditional expectation of  $\alpha_g$  is not only a function of  $v_g$ , but also of the characteristics of the group. We assume a "local" exclusion restriction in the sense that the vector of instruments impacts the outcome of interest ( $Y_{ig}$ ) only through peer quality, after we control for group characteristics (for a certain subpopulation) and the unobservable component of peer quality. Conditional on  $v_g$  and  $W_g$ , variations in peer quality are stochastically independent of variations in the correlated effects ( $\alpha_g$ ) and we can separate

---

<sup>4</sup>Notice that we can replace these two assumptions by  $(v_g, \alpha_g, \varepsilon_{ig}) \perp A_g | W_g$  and  $(\alpha_g, \varepsilon_{ig}) \perp W_g$ .

<sup>5</sup>If the exogeneity assumption does not hold, in the semiparametric case, the estimation of the average structural function  $H(Q_{ig}, Q_g)$  and the endogeneity function  $\Gamma(v_g, W_g)$  are going to be biased.

out the effect of peer quality on the outcome of interest from the effect of the correlated effects.

In assumption 2.4,  $\Gamma(v_g, W_g)$  represents the function that allows to control for the endogeneity of peer quality. This function is called control function, and represents the conditional expectation of  $\alpha_g$ . The conditional expectation of  $\alpha_g$  is a function of the unobservable component of peer quality and of the characteristics of the group.

In the next assumption, we assume that the observations are *i.i.d* across groups.

**Assumption 2.5** *An iid sample  $\{(Y_{g1}, \dots, Y_{gM_g}, X_{g1}, \dots, X_{gM_g}, W_g, A_g, \alpha_g) : g = 1, \dots, N\}$  is drawn from the population with distribution  $F_0$ .*

We estimate and identify two objects of interest, the average production function and peer effects. In the notation of Blundell and Powell (2001), the average production function is the average structural function (ASF),

$$\begin{aligned}\Phi_1(q_i, q) &= \mathbb{E}[H(q_i, q) + \alpha_g + \varepsilon_{ig}] \\ &= H(q_i, q)\end{aligned}$$

where  $(q_i, q)$  represent fixed levels of the random variables  $(Q_{ig}, Q_g)$ .

This function allows us to evaluate the average student's achievement at certain levels of inputs. This function can be used to evaluate each student achievement at a fixed level of student quality and peer quality.

Peer effects are defined as the average marginal effect of peer quality on student achievement, and it is given by the marginal productivity of  $Q_g$ . The peer effects equal to the average marginal derivative,

$$\Phi_2(q_i) = \frac{\partial \mathbb{E}[H(q_i, Q_g) + \alpha_g + \varepsilon_{ig}]}{\partial Q_g} = \mathbb{E}\left[\frac{\partial}{\partial Q_g} H(q_i, Q_g)\right]$$

where the second inequality holds by interchanging differentiation and expectation, and by assumptions 2.1, 2.3 and 2.4. The expectation is taken over the marginal distribution of  $Q_g$ .

Peer effects give the average effect on student achievement of an increase in peer quality by one unit, for a specific value of student quality. An equivalent expression for the average derivative is  $\lim_{\Delta Q_g \rightarrow 0} \frac{\mathbb{E}[(H(q_i, Q_g + \Delta Q_g) + \alpha_g + \varepsilon_{ig}) - (H(q_i, Q_g) + \alpha_g + \varepsilon_{ig})]}{\Delta Q_g}$ . Suppose that the Principal in a certain school allocated two students with the same quality to two classrooms that

have the same characteristics, except that one classroom has peer quality  $q_g$  and the other has peer quality  $q_g + 1$ . In this experiment, the average treatment effect is the difference between the achievements of the two students. This average treatment effect is basically the average impact of a one unit increase in peer quality on student achievement, which is similar to average marginal effect of peer quality on student outcome.

We should interpret this second estimand carefully. If this estimator is positive for an individual with a high quality, but is negative for an individual with low quality, it means that an increase in the quality of the group increases the achievements high quality students, but decreases low quality students' achievements. We might interpret these results as evidence that high quality types benefit more from an increase in peer quality, and hence segregation by type is optimal. However, this analysis is not correct. As pointed by Arnott and Rowse (1987) and Graham, Imbens and Ridder (2007), the allocation that maximizes the average outcome also depends on the curvature of the production function and the marginal distribution of types (high quality or low quality) in the population. The optimal way to allocate students to classrooms depends on the distribution of quality in the school and on the degree of complementarity between student quality and peer quality.

We propose a semiparametric methodology to identify and estimate the parameters of interest without assuming a functional form for  $H(Q_{ig}, Q_g)$ . Instead of a parametric form for  $H(Q_{ig}, Q_g)$ , we impose monotonicity of  $H(Q_{ig}, Q_g)$  in relation to  $Q_{ig}$  and place restrictions on the distribution of  $\varepsilon_{ig}$ . To estimate the average production function and peer effects, we propose a three step procedure. In the first step, we need to construct the "single index" that defines student quality. Using the within-group variation, we identify and estimate the parameters  $\beta$  that are in the linear combination that defines  $Q_{ig}$ ,  $Q_{ig} = X'_{ig}\beta$ . These parameters are estimated using a generalized version of rank regression. The idea is that in each classroom, we rank the students based on their quality. Based on the monotonicity assumption imposed on  $H(Q_{ig}, Q_g)$ , we expect that students with high quality have high achievement. In other words, we expect that inside a group, if  $Q_{ig} > Q_{jg}$ , then  $Y_{ig} > Y_{jg}$ . The parameters  $\beta$  will maximize the rank correlation between  $Y_{ig}$  and  $Q_{ig}$  for individuals in the same group. When we compare individuals within each group, we subtract off the correlated effects and use the within group variation to estimate  $\beta$ . The identification of  $\beta$  is based not only on monotonicity of  $H(Q_{ig}, Q_g)$ , but also on the assumption that the individual heterogeneities are stationary inside the group. The stationary assumption imposes that the conditional distribution function of the individual unobservables are the same for all the individuals inside the group, conditional

on the vector of covariates and the uncorrelated effects. This assumption implies that the distribution of  $\varepsilon$  does not change when we move among the individuals inside a group. If we assume that  $\varepsilon_{igs}$  are exchangeable, in the sense that the joint distribution of  $\varepsilon_{igs}$  does not change under permutation, this stationarity assumption holds. Notice that this assumption is weaker than assuming that the individual heterogeneities ( $\varepsilon_{igs}$ ) are *i.i.d* inside the group. If assumption 2.1 does not hold, and the unobservable components of this model ( $\alpha_g$  and  $\varepsilon_{ig}$ ) are non-additive elements, the stationary assumption needs to be replaced by the *i.i.d* assumption. Using this estimator of  $\beta$ , we can get the single indexes that represent student quality and peer quality.

The next two steps are based on the control function approach presented in Newey, Powell and Vella (1999). In the second step, we estimate  $\Psi(A_g, W_g)$  using a series approach. In this case, we use a vector of approximating functions of  $(A_g, W_g)$  to estimate the nonparametric regression of  $Q_g$  on  $(A_g, W_g)$ , using  $\widehat{Q}_g$  estimated in the first step in the place of the true  $Q_g$ . Using assumption 2.4, the residual  $\widehat{v}_g$  of this regression is given by  $\widehat{v}_g = \widehat{Q}_g - \widehat{\Psi}(A_g, W_g)$ . In the last step, we use a vector of approximating functions of  $(Q_{ig}, Q_g, W_g, v_g)$  to estimate the nonparametric regression of  $Y_{ig}$  on  $Q_{ig}, Q_g, W_g, v_g$ , using the  $\widehat{Q}_{ig}, \widehat{Q}_g, \widehat{v}_g$  obtained in the previous steps in the place of the true values. The advantage of this series approach is that we can impose our additivity assumption (assumption 2.1). We use a vector of approximating functions such that each term in the vector depends either on  $(\widehat{Q}_{ig}, \widehat{Q}_g)$  or on  $(\widehat{v}_g, W_g)$ , but not on both. In this case, the estimator of the average structural function ( $H(Q_{ig}, Q_g)$ ) can be recovered by pulling out the components that depend only on  $(\widehat{Q}_{ig}, \widehat{Q}_g)$ , and the estimator of  $\Gamma(v_g, W_g)$  can be constructed by the terms that depend only on  $(\widehat{v}_g, W_g)$ .

The drawback of this procedure is that our estimators converge to a rate that is slower than the usual parametric rate,  $\sqrt{N}$ . In spite of this disadvantage, this procedure is one of the few methodologies that estimates the peer effects without imposing a functional form for the function that relates achievement to student, peer, school and teacher characteristics.

### 3 Identification

In this model, we first need to identify the parameters inside the index that define student's quality. To identify  $\beta$ , we use a rank estimator that is a generalized version of the rank estimator proposed by Abrevaya (2000). When we obtain peer and student quality, we can identify the average structural function and peer effects by imposing additional assumptions

about the behavior of  $Q_{ig}$  and  $Q_g$ .

### 3.1 Identification of $\beta$ using rank regression

To be able to identify  $\beta$  using a rank estimator, we need to impose the following assumptions:

**Assumption 3.1** (MONOTONICITY) :  $H(Q_{ig}, Q_g) : \mathbb{R}^q \rightarrow \mathbb{R}$  is strictly increasing in the first argument  $Q_{ig}$ .

This assumption is key for identification of  $\beta$ , and it implies that inside each group, the rank of  $X_i'\beta$  and the rank of  $Y_i$  are positively correlated.

**Assumption 3.2** (STATIONARY) :  $\varepsilon_{ig}$ s are stationary inside each group conditional on  $\mathbf{X} = [X_1, \dots, X_{M_g}]$  and  $\alpha_g$ , with positive density almost everywhere, which implies that:  $F_{\varepsilon_{ig}-\varepsilon_{jg}|\mathbf{X},\alpha} = F_{\varepsilon_{jg}-\varepsilon_{ig}|\mathbf{X},\alpha}$  for all  $(i, j)$  in group  $g$ .

**Assumption 3.3** : (a) Let  $\Delta X_{ij} = X_{ig} - X_{jg}$ . For all  $i, j$ , the support of the distribution of  $\Delta X$  is not contained in any proper linear subspace of  $\mathbb{R}^k$ , where  $k = \dim(X)$ .

(b) For all  $i, j$ , the elements in  $\Delta X$  can be rearranged such that  $\beta_1 \neq 0$  and for almost all values of  $\Delta \tilde{X}_{ijg} = (\Delta X_{ij,2}, \dots, \Delta X_{ij,k})$ ,  $\Delta X_{ij,1}$  has everywhere positive Lebesgue measure conditional on  $\Delta \tilde{X}_{ijg}$  and  $Y_{ig} \neq Y_{jg}$ .

**Assumption 3.4** :  $|\beta_1| = 1$  and  $\tilde{\beta} = (\beta_2, \dots, \beta_k)'$  is contained in a compact subset  $\tilde{\beta}$  of  $\mathbb{R}^{k-1}$ .<sup>6</sup>

**Assumption 3.5** : Let  $c_{ijg} = r_{ig} \cdot r_{jg}$ , where  $r_{ig}$  is an indicator that equals 1 if  $\{Y_{ig}, X_{ig}\}$  is observed, and 0 otherwise; and  $r_{jg}$  is an indicator that equals 1 if  $\{Y_{jg}, X_{jg}\}$  is observed, and 0 otherwise.  $c_{ijg}$  is independent of  $(Y_{1g}, X_{1g}, Y_{2g}, X_{2g}, \dots, Y_{M_g}, X_{M_g})$  for all  $g$  and  $\Pr[c_{ijg} > 0] > 0$  for some  $(i, j)$  in each  $g$ .

Assumption 3.2 imposes stationarity in the distribution of the individual's heterogeneity. Assumptions 3.2, 3.3 and 3.4 are regularity conditions needed for identification of  $\beta$ .

---

<sup>6</sup>Notice that with this normalization, we define quality of the students in terms of the unit of the characteristic associated with the normalized coefficient. For example, if we normalize the coefficient associated with household income, the quality of the student is defined in terms of income.

Assumption 3.3(a) is a rank condition and assumption 3.3(b) guarantees that  $\Delta X_{ij}$  contains a continuous element with a coefficient that is different from zero. Assumption 3.4 is a normalization of  $\beta$ . Since  $H(\cdot)$  is not specified, the scale and location of  $\beta$  are not identified. To fix the location, we need to normalize one of the components of the vector  $\beta$  to have an absolute value equal 1. Assumption 3.5 is necessary to deal with the fact that we observe different numbers of individuals in each group, like an incomplete panel.

The rank estimator is based on a simple idea proposed by Han (1987), that the observations corresponding to each individual can be ranked against each other inside the group. If  $X'_{ig}\beta > X'_{jg}\beta$ , then, by assumption 3.1, it is likely that  $Y_{ig}$  is bigger than  $Y_{jg}$ . This comparison will subtract off the correlated effects, since this effect is the same for individuals in the same group. The parameter  $\beta$  will maximize the rank correlation between  $Y_i$  and  $Q_i$  for individuals in group  $g$ , and identification and estimation of  $\beta$  will be based on the "within-group" variation.

In this paper, we use a smooth version of the traditional rank estimator proposed by Horowitz (1992),

$$S_N(b, \sigma_G) = \frac{1}{N} \sum_{g=1}^N \sum_{i=1}^M \sum_{j=i+1}^M c_{ijg} \left\{ [\mathbf{1}(Y_{ig} > Y_{jg}) - \mathbf{1}(Y_{ig} < Y_{jg})] \mathbf{K} \left( \frac{\Delta X'_{ijg} b}{\sigma_N} \right) \right\}, \quad (1)$$

where  $\sigma_N \rightarrow 0$  (as  $N \rightarrow \infty$ ) and  $\mathbf{K}(v)$  is a continuous function of the real line into itself that converges to the indicator function as  $N \rightarrow \infty$ , and satisfies:

**K1**  $|\mathbf{K}(v)| < C$  for some finite  $C$  and all  $v$  in  $\mathbb{R}$

**K2**  $\lim_{v \rightarrow -\infty} \mathbf{K}(v) = 0$  and  $\lim_{v \rightarrow \infty} \mathbf{K}(v) = 1$ .

and  $M = \max(M_g)$  and  $c_{ijg}$  is the indicator defined in assumption 3.5<sup>7</sup>.

**Theorem 3.1** *Let assumptions A4.1-A4.5 hold. Define  $\tilde{b} = (b_2, \dots, b_k)'$  and  $\Delta \tilde{X}_{ij} = (\Delta X_{ij2}, \dots, \Delta X_{ijk})$ . Let  $b_N$  be a solution to*

$$\max_{b: |b_1|=1, \tilde{b} \in \tilde{\mathcal{B}}} S_N(b, \sigma_N)$$

---

<sup>7</sup>This trick of using an indicator function to control the number of observations in each group is due to Charlier, Melenberg and van Soest, 1995.

then,

$$b_N \rightarrow \beta \text{ almost surely.}$$

To show strong consistency, we extend the proof of Theorem 2 in Abrevaya (2000). The idea is to show that  $\beta$  is the limit maximizer of the non-smooth limit objective function. Then we can use the properties of  $\mathbf{K}(\cdot)$  to show that at the limit,  $S_N(b)$  can be approximated by  $S_N(b, \sigma_G)$ .

### 3.2 Identification of the ASF and Peer Effects

In the first step, we identified  $\beta$ , so we could get  $Q_{ig}$  and  $Q_g$ . From now on, we are going to assume that student quality and peer quality are known. To identify the parameters of interest, we need to impose additional assumptions.

**Assumption 3.6**  $\mathbb{E}[\alpha_g | Q_{1g}, \dots, Q_{M_g g}, W_g, A_g] = \mathbb{E}[\alpha_g | Q_g, W_g, A_g]$

This last assumption holds if conditional on peer, school and classroom characteristics, and on the vector of instruments, the correlated effect is independent of individual characteristics. This assumption says that the conditional distribution of the correlated effects only depends on the quality of the student in group  $g$  through a symmetric function of the quality of the students in the group<sup>8</sup>. This is a dimension reduction that is needed because our instrumental variable only varies at the group level. Assumption 3.6 is similar to dimension reduction assumptions presented in panel data models. For example, Mundlak uses a similar hypothesis to estimate a linear "random effect" model. In this model, the author assumes that  $\alpha_g$  is a linear function of  $Q_{ig}$  and a unobservable component,  $\omega_{ig}$ ; and averaging this equation over  $i$  for a given  $g$ , the conditional expectation of  $\alpha_g$  is only function of  $Q_g$  and  $\omega_g$ . Altonji and Matzkin (2003) also use a dimension reduction assumption to identify a local average response function in a nonseparable panel data model. They assume that the conditional distribution of the unobservable component is exchangeable in the covariates. In this case, the conditional distribution is invariant to permutations of the covariates, and can be expressed as a function of exchangeable functions of the covariates, for example the mean, the product, etc.

---

<sup>8</sup>First note that this assumption holds if  $\alpha_g \perp Q_{ig} | Q_g, W_g$ . Assumption 3.6 does not seem very restrictive. For example, in the case that  $Q_g = \frac{1}{N} \sum_{i=1}^{M_g} X_{ig}$ , it holds under the assumption that  $X_{ig}$ s are independent with normal distributions with the same mean and variance. In addition,  $\alpha_g$  is also normally distributed, with  $Cov[\alpha_g, X_{ig}] = a > 0$  for all  $i$ .

Under assumptions 2.1, 2.3, 2.4 and 3.6,

$$\begin{aligned}\mathbb{E} [Y_{ig} | Q_{ig}, \dots, Q_{M_{gg}}] &= H(Q_{ig}, Q_g) + \mathbb{E} [\alpha_g | Q_{ig}, \dots, Q_{M_{gg}}] \\ &= H(Q_{ig}, Q_g) + \Gamma(v_g, W_g)\end{aligned}$$

**Assumption 3.7** (RANK AND SUPPORT CONDITIONS) : *For all the points in the support of  $(Q_{ig}, \dots, Q_{M_{gg}}, Q_g)$ , the support of  $(v_g, W_g)$  conditional on  $(Q_{ig}, \dots, Q_{M_{gg}}, Q_g)$  is equal to the support of  $(v_g, W_g)$ .*

This assumption guarantees that there is no functional relationship between the random vector  $(Q_{ig}, Q_g)$  and  $(v_g, W_g)$ . Under this condition, as proved in Powell and Vella (1999), we can identify  $H(Q_{ig}, Q_g)$  up to a constant. To be able to identify the level of  $H(Q_{ig}, Q_g)$ , we need to impose some location restriction on the distribution of  $\alpha_g$  and  $\varepsilon_{ig}$ . We assume that  $\mathbb{E}[\alpha_g] = \mathbb{E}[\varepsilon_{ig}] = 0$  (Assumption 2.3).

**Assumption 3.8** :  *$H(Q_{ig}, Q_g)$  and  $\Gamma(v_g, W_g)$  are continuously differentiable.*

**Assumption 3.9** :  *$Q_g$  has a continuous distribution given  $(v_g, W_g)$ .*

These last two assumptions are necessary to identify the average marginal derivative at a specific point.

Under assumptions 2.1, 2.3, 2.4, 3.6 and 3.7, and using the fact that  $\mathbb{E}[\alpha_g] = 0$ , we can identify the average production function,  $H(q_i, q)$ .

Assumptions 3.8 and 3.9 guarantee that the derivative of the function that represents student achievement,  $H(q_i, Q_g) + \alpha + \varepsilon_i$ , is a well-defined object. Assumptions 3.6, 3.7 and 2.4 are also necessary for the identification of this derivative. We can express the average marginal derivative as an explicit function of the data distribution,



$$\begin{aligned}
\Phi_2(q_i) &= \int \left[ \frac{\partial (H(q_i, Q_g) + \alpha + \varepsilon_i)}{\partial Q} \right] \cdot f(\alpha, \varepsilon, Q) \, dm(\varepsilon) \, dm(\alpha) \, dm(Q) \\
&= \int \frac{\partial H(q_i, Q_g)}{\partial Q_g} \cdot f(\varepsilon|v, W, \alpha, Q) f(\alpha|v, W, Q) f(Q|W, v) \\
&\quad f(v, W) \, dm(\varepsilon) \, dm(\alpha) \, dm(Q) \, dm(v, W) \\
&\quad + \int \frac{\partial \alpha}{\partial Q_g} \cdot f(\varepsilon|v, W, \alpha, Q) f(\alpha|v, W, Q) f(Q|W, v) \\
&\quad f(v, W) \, dm(\varepsilon) \, dm(\alpha) \, dm(Q) \, dm(v, W) \\
&= \mathbb{E} \left[ \frac{\partial}{\partial Q_g} H(q_i, Q_g) \right],
\end{aligned}$$

where the second equality follows by assumptions 2.3, 2.4 and 3.6. In addition, the object in the last equality is well-defined by assumptions 3.8, 3.7 and 3.9.

Under assumptions 2.1, 2.3, 2.4, 3.6, 3.8 and 3.9, we can identify  $\mathbb{E} \left[ \frac{\partial (H(q_i, Q_g) + \alpha_g + \varepsilon_{ig})}{\partial Q_g} \right]$ .

#### 4 Estimation

We propose a three step estimation procedure. In the first step, we estimate the vector  $b_N$  by finding the global maximum of the function  $S_N(b, \sigma_N)$ . With this vector  $b_N$ , we can construct the indexes for quality of each student and quality of his peer,  $\widehat{Q}_{ig} = X'_{ig} b_N$  and  $\widehat{Q}_g = f(X_{1g}, X_{2g}, \dots, X_{M_g g})' b_N$ . In the second step, we use a series estimator to approximate  $\Psi(A_g, W_g)$ , and construct the vector of residuals,  $\widehat{v}_g = \widehat{Q}_g - \widehat{\Psi}(A_g, W_g)$ . In the third step, we use a series estimator to approximate the conditional expectation of  $Y$ ,  $\mathbb{E}[Y|Q_{ig}, Q_g, W_g, A_g] = H(Q_{ig}, Q_g) + \Gamma(v_g, W_g)$ , using  $\widehat{v}_g$ ,  $\widehat{Q}_{ig}$  and  $\widehat{Q}_g$  in the place of the true values.

Since the function  $S_N(b, \sigma_N)$  has many local maximums, conventional algorithms tend to fail in this case, since even when these algorithms converge, there is no guarantee that they have found the global maximum, as opposed to the local maximum. To find the global maximum of  $S_N(b, \sigma_N)$ , we suggest a global search algorithm, called simulated annealing. This algorithm is an iterative search procedure that moves in all directions, avoiding the local maximums encountered during the interactions<sup>9</sup>. The advantage of this

---

<sup>9</sup>For a description of this algorithm, see Corana, Marchesi, Martini and Ridella (1987) and Goffe, Ferrier and Rogers (1994).

algorithm is that it searches for the maximum on the entire surface of the function; it optimizes the function while moving uphill and downhill, which allows it to escape from local maximums and find the global maximum. Before doing this maximization, we need to choose the smooth parameters in the function  $S_N(b, \sigma_N)$ : the bandwidth ( $\sigma_N$ ) and the kernel ( $K(\cdot)$ ). The kernel needs to be a bounded function that has limits equal to 0 or 1, but does not need to be a distribution function. Different kernels have been proposed in the literature, depending on how much we want to smooth the function. The choice of the bandwidth is less simple than the choice of the kernel, since there are no optimal rules to choose bandwidths in the literature. Horowitz (1992) suggests a plug-in method. This method chooses the bandwidth that minimizes the asymptotic mean square error of the rank estimator. In the next section, we are going to derive the asymptotic theory for our rank estimator and show how this plug-in method works.

In the second and third steps, we take  $\hat{\beta}$  as given, and estimate  $\hat{\Psi}(A_g, W_g)$  and  $\Lambda(\hat{Q}_{ig}, \hat{Q}_g, W_g, \hat{v}_g) = H(\hat{Q}_{ig}, \hat{Q}_g) + \Gamma(\hat{v}_g, W_g)$  using a series estimator. The idea is to approximate each one of the functions by a series of independent variables, in such a way that the number of terms in the series expands with the sample size. For the second step, define  $Z_g \equiv [A'_g, W'_g]'$ , consider a positive integer  $L_1$ , and let  $r^{L_1}(Z) = (r_{1L_1}(Z), \dots, r_{L_1L_1}(Z))'$  be a vector of approximating functions. In this case,  $\hat{\Psi}(A_g, W_g)$  is the predicted value of a regression of  $\hat{Q}_g$  on  $r_g = r^{L_1}(Z)$ ,

$$\hat{\Psi}(z) = r^{L_1}(z)' \hat{\gamma}, \quad \hat{\gamma} = (R'R)^{-1} R'(\hat{q}_1, \dots, \hat{q}_N)', \quad R = [r_1, \dots, r_N]'$$

The residuals from this regression are going to be used in the next step of the estimation process,  $\hat{v}_g = \hat{Q}_g - \hat{\Psi}(A_g, W_g)$ .

In the last step, we define  $P^L(T) = (p_{1L}(T), \dots, p_{LL}(T))'$  as the vector of approximating functions of  $T_{ig} = [Q_{ig}, Q_g, W_g, v_g]$  such that each term  $p_{lL}(T)$  depends either on  $(Q_{ig}, Q_g)$  or on  $(v_g, W_g)$ , but not on both. Notice that we are imposing the additive structure stated in assumption 2.1 in this vector of approximating functions. In this case, an estimator of  $H(Q_{ig}, Q_g)$  can be constructed using the terms that depend only on  $(Q_{ig}, Q_g)$ , and an estimator for  $\Gamma(v_g, W_g)$  will be based on the terms that depend only on  $(v_g, W_g)$ . To avoid small denominators in the OLS estimation due to outliers in  $t$ , we assume nonrandom trimming of the group level variables<sup>10</sup>. Define  $T_g = (Q_g, W_g, v_g)$ , the nonrandom

<sup>10</sup>Nonrandom trimming deletes the extreme observations, and moves the goalpost a little.

trimming function is defined as

$$\tau(T_g) = \prod_{j=1}^{d_t} \mathbf{1}\{a_j < t_{gj} < b_j\},$$

where  $a_j, b_j$  are finite constants and  $t_{gj}$  is the  $j$ th component of  $t_g$ .

This trimming function  $\tau(T_g)$  allows us to exclude groups with large values of  $t$  that can distort the results of a polynomial series estimator. In addition, it helps to deal with denominators close to zero in the derivation of the asymptotic distribution of the estimators. We allow trimming at the group level. If we trim at the individual level, we are excluding individuals from the group, but you are keeping the peer quality that we estimated in step one. In this case, the interpretation of peer effects changes. To avoid this problem, we trim the variables at the group level.

Replacing the true values of  $(v_g, Q_{ig}, Q_g)$  by their estimated values obtained in the previous steps, we obtain  $\widehat{T}_{ig} = [\widehat{Q}_{ig}, \widehat{Q}_g, W_g, \widehat{v}_g]$ ,  $\widehat{T}_g = [\widehat{Q}_g, W_g, \widehat{v}_g]$  and  $\widehat{\tau}_g = \tau(\widehat{T}_g)$ . We also let  $r_{ig}$  be the indicator variable that equals 1 if  $\{Y_i, X_i\}$  is observed in group  $g$ . As in the previous step,  $\Lambda(\widehat{T}_{ig})$  is the predicted value of a regression of  $Y_{ig}$  on  $\widehat{p}_{ig} = p^L(\widehat{T}_{ig})$ . Using the observations with  $\widehat{\tau}_g = 1$  and  $r_{ig} = 1$ ,

$$\begin{aligned} \widehat{\Lambda}(\widehat{t}) &= p^L(\widehat{t})' \widehat{\pi} & \widehat{P} &= [r_{11}\widehat{\tau}_1\widehat{p}_{11}, \dots, r_{M1}\widehat{\tau}_1\widehat{p}_{M1}, \dots, r_{1N}\widehat{\tau}_N\widehat{p}_{1N}, \dots, r_{MN}\widehat{\tau}_N\widehat{p}_{MN}] \\ \widehat{\pi} &= (\widehat{P}'\widehat{P})^{-1} \widehat{P}'Y' & Y &= (r_{11}Y_{11}, \dots, r_{M1}Y_{M1}, \dots, r_{1N}Y_{1N}, \dots, r_{MN}Y_{MN}). \end{aligned}$$

By collecting the terms that depend only on  $(\widehat{Q}_{ig}, \widehat{Q}_g)$ , we construct the estimator of the Average Structural Function. In addition, by taking the average of the derivative of this function in relation to  $Q_g$ , we estimate the peer effects. Suppose  $p_{1l}(\widehat{t}) = 1$ , and the first  $L_l$  terms  $p_{1L}(\widehat{t})$  depend only on  $(\widehat{Q}_{ig}, \widehat{Q}_g)$ , and the remaining terms depend only on  $(W_g, \widehat{v}_g)$ . In this case, the estimators can be constructed as

$$\begin{aligned} \widehat{\Phi}_1(q_i, q) &= \widehat{c}_h + \sum_{j=2}^{L_l+1} \widehat{\pi}_j p_j(q_i, q) \\ \widehat{\Phi}_2(q_i) &= \frac{1}{N} \sum_{g=1}^N \frac{\partial \left[ \sum_{j=2}^{L_l+1} \widehat{\pi}_j p_j(q_i, \widehat{Q}_g) \right]}{\partial Q_g} \end{aligned}$$

Notice that in the second estimator, we average over  $g$ , the sample of groups. As noticed by Newey, Powell and Vella (1999),  $\Phi_1(q_i, q)$  is defined up to a constant, since the first term

in the polynomial (the constant) is a common term between  $H(\widehat{Q}_{ig}, \widehat{Q}_g)$  and  $\Gamma(\widehat{v}_g, W_g)$ . For example, in this case, the estimated value of  $\Gamma(\widehat{v}_g, W_g)$  is  $\widehat{c}_\Gamma + \sum_{j=2}^{L_l+1} \widehat{\pi}_j p_j(\widehat{v}_g, W_g)$ , where  $\widehat{c}_\Gamma + \widehat{c}_h = \widehat{\pi}_1$ . To be able to separate  $\widehat{c}_\Gamma$  from  $\widehat{c}_h$ , we need to impose another restriction,  $\Gamma_0(\bar{v}, \bar{w}) = \bar{\Gamma}$ <sup>11</sup>. In this case, we can choose  $\widehat{c}_\Gamma = \bar{\Gamma} - \sum_{j=L_l+2}^L \widehat{\pi}_j p_j(\bar{v}, \bar{w})$ , where  $\bar{\Gamma} = \frac{\sum_{g=1}^N \sum_{j=L_l+2}^L \widehat{\pi}_j p_j(v_g, w_g)}{N}$  and  $\widehat{c}_h = \widehat{\pi}_1 - \widehat{c}_\Gamma$ .

In the estimation of  $\beta$ , we need to maximize  $S_N(\sigma_N, b_N)$  imposing the restriction that the coefficient associated with one of the continuous covariates is equal to one, which implies that the scale of quality of the student is defined in terms of the student characteristic associated with the normalized coefficient. Depending on the variable that is normalized, the estimated value of peer quality changes, and consequently  $\widehat{\Phi}_2(q_i)$  changes.  $\widehat{\Phi}_2(q_i)$  will represent the marginal effect of peer quality on student outcome in units of the variable associated with the normalized coefficient<sup>12</sup>.

## 5 Inference

In this section, we provide the asymptotic distribution of our estimators, and necessary conditions for inference<sup>13</sup>. Since we propose a three step procedure to estimate the average structural function and the peer effects, the rate of convergence and standard error of these estimators will depend on the asymptotic behavior of the rank estimator used in the first step and the series estimator used in the other two steps. Before obtaining the asymptotic theory for our estimator, we derive the asymptotic theory for our rank estimator and the rates of convergence for the series estimators.

<sup>11</sup>For a better explanation of this condition, see Newey, Powell and Vella (1999).

<sup>12</sup> $\widehat{\Phi}_2(q_i)$  will tell us if quality of the peers has a positive or a negative impact on the outcome of a given student with quality  $q_i$ ; however the size of the the marginal effect of peer quality on student outcome will depend on which student characteristic has its coefficient normalized to 1. In this sense, the size of this estimator does not provide an accurate measure of the marginal effect of peer quality. To have an idea of peer effects sizes, we could compare this estimator with a third one,

$$\Phi_3 = \mathbb{E} \left[ \frac{\partial H(Q_i, Q_g)}{\partial Q_i} \right],$$

where  $\Phi_3$  represents the average marginal effect of student's quality on the outcome.  $\frac{\Phi_2}{\Phi_3}$  will give us an idea how the size of the marginal effect of peer quality compares to the size of the marginal effect of student quality. Note that in this case, the expectatio is taken over the joint distribution of  $(q_{i1}, \dots, q_{M_g N})$ .

<sup>13</sup>The proofs of all theorems in this part are in the technical appendix that is available upon request.

## 5.1 Asymptotic Distribution: Rank Estimator

Horowitz (1992) derives the asymptotic distribution of the smooth maximum score estimator for the case of a cross-sectional data set. Extension of the asymptotic results of Horowitz for a smoothed maximum score estimator in the case of panel data is straightforward (see Charlier et al (1995), etc.). In this section, we extend the results of Horowitz (1992) and Charlier et al (1995) to get the asymptotic properties of the rank estimators proposed above. The key difference is that the assumption on smoothness of the marginal distribution of  $\varepsilon_{ig}$  used by Horowitz (1992) is going to be replaced by smoothness of the joint distribution of the residuals.

The proofs of the theorems stated in this section are straightforward extensions of the proofs in Horowitz (1992). To be able to derive the asymptotic distribution of our rank estimator, we need to make the following assumptions:

**Assumption 5.1** : *The components of  $\Delta\tilde{X}_{ij}$  and of the matrices  $\Delta\tilde{X}_{ij}\Delta\tilde{X}'_{kl}$  and  $\Delta\tilde{X}_{ij}\Delta\tilde{X}'_{ij}\Delta\tilde{X}_{kl}\Delta\tilde{X}'_{kl}$  have finite first absolute moments for all  $(i, j)$ ,  $(k, l)$  such that  $i < j$  and  $k < l$ .*

**Assumption 5.2** :  $\frac{\log N}{N\sigma_N^4} \rightarrow 0$  as  $N \rightarrow \infty$

**Assumption 5.3** : (a)  $\mathbf{K}$  is twice differentiable everywhere,  $|\mathbf{K}'(\cdot)|$  and  $|\mathbf{K}''(\cdot)|$  are uniformly bounded, and each of the following integrals over  $(-\infty, \infty)$  is finite:  $\int [\mathbf{K}'(v)]^4 dv$ ,  $\int [\mathbf{K}''(v)]^2 dv$ ,  $\int [v^2\mathbf{K}''(v)]^2 dv$ ; (b) For some integers  $h \geq 2$  and each integer  $s$  ( $0 \leq s \leq h$ ),  $\int |v^s\mathbf{K}'(v)| dv < \infty$ , and

$$\int v^s\mathbf{K}'(v) dv = \begin{cases} 0 & \text{if } s < h \\ d \neq 0 & \text{if } s = h \end{cases}$$

(c) For any integer  $s$  between 0 and  $h$ , any  $\eta > 0$ , and any sequence  $\{\sigma_N\}$  converging to 0,

$$\lim_{N \rightarrow \infty} \sigma_N^{s-h} \int_{|\sigma_N v| > \eta} |v^s\mathbf{K}'(v)| dv = 0$$

$$\lim_{N \rightarrow \infty} \sigma_N^{-1} \int_{|\sigma_N v| > \eta} |\mathbf{K}''(v)| dv = 0$$

**Assumption 5.4** : Let  $Z_{ij} \equiv \Delta X'_{ij}\beta$  and  $f(Z_{ij} | \Delta\tilde{X}_{ij})$  be the distribution of  $Z_{ij}$  conditional on  $\Delta\tilde{X}_{ij}$ . For each integer  $1 \leq s \leq h-1$ , all  $Z_{ij}$  in a neighborhood of 0, almost

every  $\Delta\tilde{X}_{ij}$ , and some  $C < \infty$ ,  $f^{(s)}\left(Z_{ij}|\Delta\tilde{X}_{ij}\right) \equiv \frac{\partial^s f(Z_{ij}|\Delta\tilde{X}_{ij})}{\partial Z_{ij}^s}$  exists and is a continuous function of  $Z$  satisfying  $\left|f^{(s)}\left(Z_{ij}|\Delta\tilde{X}_{ij}\right)\right| < C$ . In addition,  $\left|f\left(Z_{ij}|\Delta\tilde{X}_{ij}\right)\right| < C$  for all  $Z$  and almost every  $\Delta\tilde{X}_{ij}$ ; and  $\left|f\left(Z_{ij}, Z_{lk}|\Delta\tilde{X}_{ij}, \Delta\tilde{X}_{kl}\right)\right| < C$  for all  $(Z_{ij}, Z_{kl})$  and almost every  $(\Delta\tilde{X}_{ij}, \Delta\tilde{X}_{kl})$ .

**Assumption 5.5** Let  $L\left(Z_{ij}, \Delta\tilde{X}_{ij}\right) \equiv 1 - 2F_{\varepsilon_i - \varepsilon_j}\left(Z_{ij}, \Delta\tilde{X}_{ij}\right)$ .<sup>14</sup> For each integer  $1 \leq s \leq h-1$ , all  $Z$  in a neighborhood of 0, almost every  $\Delta\tilde{X}_{ij}$ , and some  $C < \infty$ ,  $F_{\varepsilon_i - \varepsilon_j}^{(s)}\left(Z_{ij}, \Delta\tilde{X}_{ij}\right) \equiv \frac{\partial^s F_{\varepsilon_i - \varepsilon_j}\left(Z_{ij}, \Delta\tilde{X}_{ij}\right)}{\partial Z^s}$  exists and is a continuous function of  $Z_{ij}$  satisfying  $\left|F_{\varepsilon_i - \varepsilon_j}^{(s)}\left(Z_{ij}, \Delta\tilde{X}_{ij}\right)\right| < C$ .

**Assumption 5.6** :  $\tilde{\beta}$  is an interior point of  $\tilde{\mathcal{B}}$

To find the asymptotic distribution of  $b_N$ , Horowitz (1992) proposes a Taylor series expansion of  $S_N(b_N, \sigma_N)$  around  $\beta$ . Using the assumptions above, he can prove the existence of the matrices in this Taylor expansion and find the right rate of convergence. Under assumption 5.4,  $S_N(b)$  is twice differentiable with respect to  $\tilde{b}$ . Define

$$S_{1N}(b, \sigma_N) \equiv \frac{\partial S_N(b, \sigma_N)}{\partial \tilde{b}}$$

$$S_{2N}(b, \sigma_N) \equiv \frac{\partial^2 S_N(b, \sigma_N)}{\partial \tilde{b} \partial \tilde{b}'}$$

Let  $b_N \equiv (b_{1N}, \tilde{b}'_N)'$  denote the solution of the maximization problem of  $S_N(b_N, \sigma_N)$  as  $N \rightarrow \infty$ , with probability approaching 1,  $b_{1N} = \beta_1 = \pm 1$  and  $S_{1N}(b_N, \sigma_N) = 0$ . By the mean value theorem, for  $b_N^* \in (b_N, \beta)$

$$S_{1N}(b_N, \sigma_N) = S_{1N}(\beta, \sigma_N) + S_{2N}(b_N^*, \sigma_N) (\tilde{b}_N - \tilde{\beta}) = 0.$$

Suppose that there is a real function  $\rho(N)$  such that  $\rho(N) S_{1N}(b_N, \sigma_N)$  converges in distribution as  $N \rightarrow \infty$ . Suppose that  $S_{2N}(b_N^*, \sigma_N)$  converges to a nonsingular, nonsto-

---

<sup>14</sup>Notice that in this case, we need to change assumption 9 from Horowitz (1992), since  $H(\cdot)$  is not a linear function.

chastic matrix  $S_2$ . In this case,

$$0 = \rho(N) S_{1N}(\beta, \sigma_N) + S_{2N}(b_N^*, \sigma_N) \rho(N) (\tilde{b}_N - \tilde{\beta})$$

$$\rho(N) (\tilde{b}_N - \tilde{\beta}) = -S_2^{-1} \rho(N) S_{1N}(\beta, \sigma_N) + o_p(1).$$

Hence,  $\rho(N) (\tilde{b}_N - \tilde{\beta})$  is distributed asymptotically as  $-S_2^{-1} \rho(N) S_{1N}(\beta, \sigma_N)$ .

**Assumption 5.7** : *The matrix  $S_2$  is negative definite.*

**Theorem 5.1** *Let assumptions 3.1-3.5 and 5.1-5.7 hold for some  $h \geq 2$ , and let  $\{b_N\}$  be a sequence of solutions to the following maximization problem*

$$\max_{b: |b_1|=1, \tilde{b} \in \tilde{\mathcal{B}}} S_N(b, \sigma_N)$$

where

$$S_N(b, \sigma_N) = \frac{1}{N} \sum_{g=1}^N \sum_{i=1}^M \sum_{j=i+1}^M c_{ijg} \left\{ [\mathbf{1}(Y_{ig} > Y_{jg}) - \mathbf{1}(Y_{ig} < Y_{jg})] \mathbf{K} \left( \frac{\Delta X'_{ijg} b}{\sigma_N} \right) \right\}.$$

By Theorem 3.1,  $b_{N1} \rightarrow \beta_1$  a.s.

(i) If  $N\sigma_N^{2h+1} \rightarrow \infty$  as  $N \rightarrow \infty$ ,  $\sigma_N^{-h} (\tilde{b}_N - \tilde{\beta}) \rightarrow_p - (S_2)^{-1} S_1$

(ii) If  $N\sigma_N^{2h+1}$  has a finite limit  $\lambda$  as  $N \rightarrow \infty$ ,

$$(N\sigma_N)^{\frac{1}{2}} (\tilde{b}_N - \tilde{\beta}) \rightarrow_d \mathcal{N}_{k-1} \left( -\lambda^{\frac{1}{2}} S_2^{-1} S_1, S_2^{-1} D S_2^{-1} \right)$$

(iii) The optimal rate of convergence in distribution for the remaining parameters  $\tilde{b}_N$  is obtained for  $\sigma_N = \left(\frac{\lambda}{N}\right)^{\frac{1}{2h+1}}$  with  $0 < \sigma < \infty$  (fixed). Then

$$N^{\frac{h}{2h+1}} (\tilde{b}_N - \tilde{\beta}) \xrightarrow{d} \mathcal{N}_{k-1} \left( -(\lambda^*)^{\frac{h}{2h+1}} S_2^{-1} S_1, (\lambda^*)^{-\frac{1}{2h+1}} S_2^{-1} D S_2^{-1} \right),$$

where  $(k-1)$  vector  $S_1$  and the  $(k-1) \times (k-1)$  matrices  $D$  and  $S_2$  are defined by

$$\begin{aligned}
S_1 &\equiv -2 \sum_{i=1}^{M-1} \sum_{j=i+1}^M \alpha_{S_1} \sum_{s=1}^h \left\{ [s!(h-s)!]^{-1} \right. \\
&\quad \left. \cdot \mathbb{E} \left[ F_{\tilde{\varepsilon}_{i-\varepsilon_j}}^{(s)} \left( 0 | \Delta \tilde{X}_{ij} \right) f^{(h-s)} \left( 0 | \Delta \tilde{X}_{ij} \right) \Delta \tilde{X}_{ij} \right] \right\} \Pr [c_{ij} = 1] \\
D &\equiv \sum_{i=1}^{M-1} \sum_{j=i+1}^M \alpha_D \mathbb{E} \left[ L \left( 0, \Delta \tilde{X}_{ij} \right) \Delta \tilde{X}_{ij} \Delta \tilde{X}_{ij}' f \left( 0 | \Delta \tilde{X}_{ij} \right) \right] \Pr [c_{ij} = 1] \\
S_2 &\equiv 2 \sum_{i=1}^{M-1} \sum_{j=i+1}^M \mathbb{E} \left[ \Delta \tilde{X}_{ij} \Delta \tilde{X}_{ij}' F_{\tilde{\varepsilon}_{i-\varepsilon_j}}^{(1)} \left( 0 | \Delta \tilde{X}_{ij} \right) f \left( 0 | \Delta \tilde{X}_{ij} \right) \right] \Pr [c_{ij} = 1]
\end{aligned}$$

with

$$\alpha_{S_1} = \int_{-\infty}^{\infty} v^h \mathbf{K}'(v) dv, \quad \alpha_D = \int_{-\infty}^{\infty} [\mathbf{K}'(v)]^2 dv.$$

Let  $\Omega$  be any nonstochastic, positive semidefinite matrix such that  $S_1' S_2^{-1} \Omega S_2^{-1} S_1 \neq 0$ .

Define  $MSE = \lim_{N \rightarrow \infty} \mathbb{E} \left[ N^{\frac{2h}{2h+1}} \left( \tilde{b}_N - \tilde{\beta} \right)' \Omega \left( \tilde{b}_N - \tilde{\beta} \right) \right]$ . Then for a given  $\Omega$ , the MSE minimizing value for  $\lambda$  is  $\lambda = \lambda^* = \left[ \frac{\text{trace}(S_2^{-1} \Omega S_2^{-1} D)}{2h S_1' S_2^{-1} \Omega S_2^{-1} S_1} \right]$ .

As in the case of the smoothed maximum score estimator, under assumptions 5.1-5.7, the bias of our rank estimator is  $O\left(\sigma_N^{-h}\right)$  and the variance is  $O\left(\left(N\sigma_N\right)^{-1}\right)$ , and the fastest rate of convergence is  $N^{-\frac{h}{1+2h}}$ . This rate can be attained if the bandwidth is proportional to  $N^{-\frac{1}{1+2h}}$ ,  $\sigma_N \propto N^{-\frac{1}{1+2h}}$ . This rate of convergence is slower than  $\sqrt{N}$ . By using high order kernels (choosing  $h$  large enough), this rate of convergence can be arbitrarily close to  $\sqrt{N}$ . As argued in Horowitz (1992), for  $h = 1$ , the rate of convergence is  $N^{\frac{1}{3}}$ , and the limit distribution is unknown. For  $h \geq 2$ , the limiting distribution of the estimator is given by theorem 5.1. The asymptotic bias and a covariance matrix of the rank estimator can be consistently estimated using the consistent estimators for  $S_1$ ,  $S_2$  and  $D$  as stated in the next theorem.

Horowitz (1992) uses the consistent estimators of  $S_1$ ,  $S_2$  and  $D$  to find the optimal bandwidth. The plug in method proposed by Horowitz (1992) is based on the results of theorem 5.1. This theorem shows that the optimal bandwidth is given by  $\sigma_N^* = \left(\frac{\lambda^*}{N}\right)^{\frac{1}{2h+1}}$ , where  $\lambda^*$  minimizes the asymptotic mean square error. The plug-in method obtains the



optimal bandwidth by getting a random sequence that consistently estimated  $\lambda^*$ . The consistent estimator of  $\lambda^*$  is based on the consistent estimators of  $S_1$ ,  $S_2$  and  $D$ .

**Theorem 5.2** Let  $\tilde{b}_N$  be a consistently smoothed maximum score estimator based on  $\sigma_N = O\left(N^{-\frac{1}{2h+1}}\right)$ . For  $b \in \{-1, 1\}$   $X \tilde{\mathcal{B}}$  defined

$$s_{ijg}^1(b, \sigma) = c_{ijg} K' \left( \frac{\Delta X'_{ij} b}{\sigma} \right) [\mathbf{1}(Y_{ig} > Y_{jg}) - \mathbf{1}(Y_{ig} < Y_{jg})] \left( \frac{\Delta X'_{ij}}{\sigma} \right)$$

Let  $\sigma_N^* = O\left(N^{-\frac{\delta}{2h+1}}\right)$ , where  $0 < \delta < 1$ . Then,

- (a)  $\hat{S}_{1N} \equiv (\sigma_N^*)^{-h} S_{1N}^1(\tilde{b}_N, \sigma_N^*)$  converges in probability to  $S_1$
- (b)  $\hat{D}_N \equiv \left(\frac{\sigma_N}{N}\right) \sum_{g=1}^N \sum_{i=1}^{M_g-1} \sum_{j=i+1}^{M_g} s_{ijg}^1(\tilde{b}_N, \sigma_N) s_{ijg}^1(\tilde{b}_N, \sigma_N)'$  converges in probability to  $D$
- (c)  $S_{2N}(b_N, \sigma_N)$  converges in probability to  $S_2$ .

Theorem 5.1 states that if  $\sigma_N \propto N^{-\frac{h}{2h+1}}$ , the asymptotic bias of  $N^{\frac{h}{2h+1}}(\tilde{b}_N - \tilde{\beta})$  is  $-(\lambda^*)^{\frac{h}{2h+1}} S_2^{-1} S_1$ . Using the results of Theorem 5.2, we can consistently estimate the asymptotic bias of the estimator by  $-(\lambda_N^*)^{\frac{h}{2h+1}} S_{2N} \hat{S}_{1N}$ , and provide an asymptotically unbiased rank estimator,

$$\hat{b}_N = \tilde{b}_N + \left(\frac{\lambda_N}{N}\right)^{\frac{h}{2h+1}} S_{2N}(\tilde{b}_N, \sigma_N)^{-1} \hat{S}_{1N}$$

where  $\lambda_N = \left[ \frac{\text{trace}(S_{2N}^{-1} \Omega S_{2N}^{-1} \hat{D}_N)}{2h \hat{S}'_{1N} S_{2N}^{-1} \Omega S_{2N}^{-1} \hat{S}_{1N}} \right]$ .

Another way to get an asymptotically unbiased estimator is to undersmooth. If  $N\sigma_N^{2h+1} \rightarrow 0$ , when  $N \rightarrow \infty$ , the rank estimator is asymptotically unbiased. If we set  $\sigma_N \propto N^{-\varpi}$ , where  $\varpi > \frac{1}{2h+1}$ , then  $N\sigma_N^{2h+1} \rightarrow 0$ . Horowitz (2002) shows that the errors in levels of  $t$  and  $X^2$  tests for the smoothed maximum score estimator are larger with the bandwidth that minimizes the rate of convergence of maximum score estimator than with undersmoothing. As in nonparametric regression, the bandwidth that maximizes the rate of convergence is not optimal for testing. The motivation for undersmoothing is that it is less dangerous when one is constructing a confidence interval than when one is worried about point estimation (Pagan and Ullah, 1999).

## 5.2 Convergence Rates: Series Estimator

The rates of convergence for the series estimators in step2 and 3 of the control function approach were derived in Newey (1994) and Newey, Powell and Vella (1999). However, in this section, the rates of convergence are going to differ from the ones found by these authors, because we need to take into account the error that comes from the estimation in the first step.

Using the results in Newey (1994) and Newey, Powell and Vella (1999), we can find the mean-squared and uniform rates of convergence under the following assumptions.

**Assumption 5.8** :  $\text{Var}[Y|X, Z], \mathbb{E}[Y_{ig}, Y_{jg}|X_{ig}, X_{jg}, Z_g], \mathbb{E}[X_{ig}X_{jg}]$  and  $\mathbb{E}\left[\|f(X_{1g}, \dots, X_{Ng})\|^2|Z\right]$  are bounded.

**Assumption 5.9** : (i) For every  $L_1$  there is a nonsingular constant matrix  $B_1$  such that  $R^{L_1}(Z) = B_1 r^{L_1}(Z)$ ; (ii) the smallest eigenvalue of  $\mathbb{E}[R^{L_1}(Z_i)R^{L_1}(Z_i)']$  is bounded away from zero uniformly in  $L_1$ ; (iii) there is a sequence of constants  $\zeta_0(L_1)$  satisfying  $\sup_{Z \in \mathcal{Z}} \|R^{L_1}(Z)\| \leq \zeta_0(L_1)$ ; (iv) For a vector of functions  $\Psi(Z)$ , define the vector of partial derivatives  $\partial^\mu \Psi(Z) = \frac{\partial^{|\mu|} \Psi(Z)}{\partial z_1^{\mu_1} \dots \partial z_r^{\mu_r}}$  and let  $|\Psi|_d = \max_{|\mu| \leq d} \sup_{z \in \mathcal{Z}}$

$|\partial^{|\mu|} \Psi(Z)|$ . For any integer  $d \geq 0$  there are  $\vartheta_1, \gamma_{L_1}$  such that  $|\Psi_0 - r^{L_1} \gamma_{L_1}|_d = O(L_1^{-\vartheta_1})$  as  $L_1 \rightarrow \infty$ .

**Assumption 5.10** : Define  $\mathcal{T} = \{t : \tau_g(t) = 1\}$ . (i)  $t(Q, Q_g, W, \Psi - Q_g)$  is Lipschitz in  $Q, Q_g$  and  $\Psi$ ; (ii)  $t(Q, Q_g, W, \Psi(A, W) - Q_g)$  is continuously distributed with bounded density in  $\mathcal{T}$ , and  $\mathcal{T}$  is contained in the interior of support of  $t$ ; (iii) for every  $L$  there is a singular constant matrix  $B$  such that  $P^L(t) = B p^L(t)$ ; (iv) the smallest eigenvalue of  $\mathbb{E}[\tau(t) P^L(t) P^L(t)']$  is bounded away from zero uniformly in  $L$ ; (v) for each nonnegative integer  $d$ , there is  $\xi_d(L)$  such that  $\max_{\|\mu\| \leq d} \sup_{t \in \mathcal{T}} \|\partial^{|\mu|} P^L(t)\| \leq \xi_d(L)$ ; (vi) there is  $\vartheta$  and  $\gamma_L$  such that  $|\Phi_0 - P^L \gamma_L|_d = O(L^{-\vartheta})$ .

Assumption 5.8 is a standard assumption in the literature about series estimators. It assumes bounded second moments for  $Y$  and  $X$ . Assumptions 5.9 and 5.10 are important in the derivation of the rates of convergence for our series estimator. Part (iii) of assumption 5.9 and part (v) of assumption 5.10 control the bias of the series estimators; and the other assumptions deal with the magnitude of the series terms and the second moment of the estimators.

**Theorem 5.3** *If assumptions 5.8 and 5.9 hold with  $d = 0$ , and  $\frac{\zeta_0(L_1)^2 L_1}{N} \rightarrow 0$ , then*

$$\int \left[ \Psi_0(z) - \widehat{\Psi}(z) \right]^2 dF_0(Z) = O_p \left( \frac{1}{N\sigma_N} + \sigma_N^{2h} + \frac{L_1}{N} + L_1^{-2\vartheta_1} \right).$$

*In addition, if assumptions 5.8 and 5.9 hold with  $d \geq 0$ , and  $\frac{\zeta_0(L_1)^2 L_1}{N} \rightarrow 0$ , then*

$$\left| \widehat{\Psi} - \Psi_0 \right|_d = O_p \left( \zeta_d(L_1) \left[ \frac{1}{\sqrt{N}\sigma_N} + \sigma_N^h + \frac{L_1^{\frac{1}{2}}}{N^{\frac{1}{2}}} + L_1^{-\vartheta_1} \right] \right).$$

This theorem gives the mean square and uniform rates of convergence for the function estimated in the second step,  $\widehat{\Psi}(A_g, W_g)$ . These rates are equal to the rate obtained by Newey (1997), assuming that  $\beta$  is known  $\left( \frac{L_1}{N} + L_1^{-2\vartheta_1} \right)$ , plus the rate for our rank estimator  $\left( \sigma_N^{2h} + \frac{1}{N\sigma_N} \right)$ . Notice that  $L_1^{-2\vartheta_1}$  and  $\sigma_N^{2h}$  correspond to bias terms, and  $\frac{1}{N\sigma_N}$  and  $\frac{L_1}{N}$  to variance terms. If we choose  $L_1$  and  $h$  that make the rate of bias and variance equal, we have  $L_1 = N^{\frac{1}{1+2\vartheta_1}}$  and  $\sigma_N = N^{-\frac{1}{1+2h}}$ . In this case, the mean square of convergence is given by  $O_p \left( \max \left( N^{-\frac{2\vartheta_1}{1+2\vartheta_1}}, N^{-\frac{2h}{1+2h}} \right) \right)$ . Notice that if  $\vartheta_1 > h$ , the rate of convergence of the rank estimator dominates, and in this case the rate is slower than the one found by Newey (1997). If  $h > \vartheta_1$ , the rate of convergence of the series estimator dominates and we are back to the rate found by Newey (1997).

The next theorem provides the rate of convergence for the function estimated in the third step,  $\Lambda \left( \widehat{Q}_{ig}, \widehat{Q}_g, W_g, \widehat{v}_g \right) = H \left( \widehat{Q}_{ig}, \widehat{Q}_g \right) + \Gamma \left( \widehat{v}_g, W_g \right)$ .

**Theorem 5.4** *If assumptions 5.8 and 5.10 hold with  $d = 0$ ,  $\frac{\xi_0(L)^2 L}{N} \rightarrow 0$  and*

$$\left( L^{\frac{1}{2}} M \xi_1(L) + \xi_o(L)^2 \zeta_0(L_1) M \right) \left[ \sigma_N^h + \frac{1}{\sqrt{N}\sigma_N} + \frac{L_1^{\frac{1}{2}}}{N^{\frac{1}{2}}} + L_1^{-\vartheta_1} \right] \rightarrow 0, \text{ then}$$

$$\begin{aligned} \int \tau(t) \left[ \widehat{\Lambda}(t) - \Lambda_0(t) \right]^2 dF_0(w) \\ = O_p \left( \frac{ML}{N} + ML^{-2\vartheta} + \frac{ML_1}{N} + ML_1^{-2\vartheta_1} + \frac{M}{N\sigma_N} + M\sigma_N^{2h} \right). \end{aligned}$$

*If assumptions 5.8 and 5.10 hold with  $d \geq 0$ ,  $\frac{\xi_0(L)^2 L}{N} \rightarrow 0$  and*

$$\left( L^{\frac{1}{2}} M \xi_1(L) + \xi_o(L)^2 \zeta_0(L_1) M \right) \left[ \sigma_N^h + \frac{1}{\sqrt{N} \sigma_N} + \frac{L_1^{\frac{1}{2}}}{N^{\frac{1}{2}}} + L_1^{-\vartheta_1} \right] \rightarrow 0, \text{ then}$$

$$\sup_{T \in \mathcal{T}} \left| \widehat{\Lambda}(t) - \Lambda_0(t) \right|_d = O_p \left( M^{\frac{1}{2}} \xi_d(L) \left[ \frac{L^{\frac{1}{2}}}{\sqrt{N}} + L^{-\vartheta} + \frac{L_1^{\frac{1}{2}}}{\sqrt{N}} + L_1^{-\vartheta_1} + \frac{1}{\sqrt{N} \sigma_N} + \sigma_N^h \right] \right).$$

This theorem shows that the rate of convergence for  $\widehat{\Lambda}(t)$  is equal to the sum of three different rates: the rate we get for  $\widehat{\Lambda}(t)$ , assuming that we know  $Q_{ig}$ ,  $Q_g$  and  $v_g$ ,  $\left(\frac{KM}{N} + MK^{-2\alpha}\right)$ , the rate for the series estimator in step 2, assuming that we know  $\beta$ ,  $\left(\frac{ML}{N} + ML^{-2\alpha_1}\right)$ , and the rate of our rank estimator,  $\left(\frac{M}{N\sigma_N} + M\sigma_N^{2h}\right)$ . This rate is different from the one provided by Newey, Powell and Vella (1999) since we need to take into account that some of the independent variables were estimated in step 1.

Similar to theorem 5.3,  $M\sigma_N^{2h}$ ,  $ML_1^{-2\vartheta_1}$  and  $ML^{-2\vartheta}$  correspond to bias terms, and  $\frac{LM}{N}$ ,  $\frac{ML_1}{N}$  and  $\frac{M}{N\sigma_N}$  correspond to variance terms. If we choose  $L$ ,  $L_1$  and  $\sigma_N$  that make the bias and variance terms equal, we get  $L = N^{\frac{1}{1+2\vartheta}}$ ,  $L_1 = N^{\frac{1}{1+2\vartheta_1}}$  and  $\sigma_N = N^{-\frac{1}{1+2h}}$ , and the rate of convergence is equal to  $O_p \left( \max \left( N^{-\frac{2\vartheta}{1+2\vartheta}}, N^{-\frac{2\vartheta_1}{1+2\vartheta_1}}, N^{-\frac{2h}{1+2h}} \right) \right)$ . Notice that if  $h < \vartheta_1$  and  $h < \vartheta$ , the rate of convergence of the rank estimator dominates, and the rate of converge obtained in Theorem 5.4 is slower than the rate obtained by Newey, Powell and Vella (1999). If  $h > \vartheta_1$  or  $h > \vartheta$ , the rate of convergence only differs from the one obtained by Newey, Powell and Vella (1999) by the presence of  $M$ . Recall that  $M = \max(M_g)$ , and it is a fixed and small number, and does not affect the rate of convergence.

### 5.3 Inference: ASF and Average Marginal Derivative

In this section, we show asymptotic normality for the average structural function and the average marginal derivative of this function in relation to peer quality. In addition, we show that these estimators are not root- $N$  consistent, but the rate of convergence of these estimators will depend on the rate of convergence of the rank estimator used in the first step of the estimation procedure. We construct the standard error for the estimators taking into account the errors that come from the first and second steps of the estimation procedure. Our estimators are linear functions of  $\widehat{\Lambda}(t)$ , and we can extend the results of Newey, Powell and Vella (1999) for asymptotic distributions of estimators that are a linear

functionals of  $\widehat{\Lambda}(t)$ . Linearity implies that we can write both estimators as linear functions of the Ordinary Least Squares coefficient obtained in the third step of the estimation procedure.

The estimator of the average structural function can be represented as

$$\begin{aligned}\widehat{\Phi}_1(q_i, q) &= \widehat{c}_h + \sum_{j=2}^{L_i+1} \widehat{\pi}_j p_j(q_i, q) \\ &= \sum_{j=1}^L \widehat{\pi}_j \cdot p_j(q_i, q, \bar{v}, \bar{w}) - \sum_{j=L_i+2}^L \widehat{\pi}_j \cdot \left( \frac{\sum_{g=1}^N P_j(v_g, w_g)}{N} \right) \\ &= a_1(p^L(t))' \widehat{\pi} = a_1(\widehat{\Lambda}(t)),\end{aligned}$$

where

$$a_1(p^L(t)) = p^L(q_i, q, \bar{v}, \bar{w}) - \left[ \frac{\mathbf{0}_{K+1 \times 1}}{\sum_{g=1}^N p_j(v_g, w_g)} \right].$$

The estimator of the peer average is the average marginal derivative of the student outcome in relation to peer quality and can be represented as

$$\begin{aligned}\widehat{\Phi}_2(q_i) &= \frac{1}{N} \sum_{g=1}^N \frac{\partial \left[ \sum_{j=1}^L \widehat{\pi}_j p_j(q_i, \widehat{Q}_g, w_g, v_g) \right]}{\partial Q_g} \\ &= a_2 \left( P^L(q_i, \widehat{Q}_g) \right)' \widehat{\pi},\end{aligned}$$

where

$$a_2 \left( P^L(q_i, \widehat{Q}_g) \right) = \frac{1}{N} \sum_{g=1}^N \frac{\partial P^L(q_i, \widehat{Q}_g, w_g, v_g)}{\partial Q_g}.$$

Since the average structural function is a linear combination of the OLS coefficients in the third step, we can use the Delta Method to find the asymptotic variance of this estimator,

$$\widehat{V}_{\widehat{\Phi}_1} = a_1(p^L(t))' \widehat{V}_{\widehat{\pi}} a_1(p^L(t)).$$

The estimator for peer effects is also a linear combination of  $\widehat{\pi}$ . However, the coefficients in this linear combination are functions of the rank estimator recovered in the first step, since in this estimator we are estimating the distribution function of peer quality, and the estimator of peer quality is a function of the rank estimator. For the estimator of

the marginal derivative, we need to deal with the approximation of  $a_2 \left( P^L \left( q_i, \widehat{Q}_g \right) \right)$  to  $\mathbb{E} \left[ \frac{\partial P^L(q_i, Q_g, w_g, v_g)}{\partial Q_g} \right]$  before applying the Delta Method, and the variance term will account for the fact that we are using an estimator for the marginal distribution of peer quality, then

$$\begin{aligned} \widehat{V}_{\widehat{\Phi}_2} &= a_2 \left( P^L \left( q_i, \widehat{Q}_g \right) \right) \widehat{V}_{\widehat{\pi}} a_2 \left( P^L \left( q_i, \widehat{Q}_g \right) \right)' + a_2^2 \left( P^L \left( q_i \right) \right) \widehat{V}_{\widehat{b}_N} a_2^2 \left( P^L \left( q_i \right) \right)' \\ &\quad + a_2 \left( P^L \left( q_i, \widehat{Q}_g \right) \right) \widehat{V}_{\widehat{\pi}}^{\frac{1}{2}} \widehat{V}_{\widehat{b}_N}^{\frac{1}{2}} a_2^2 \left( P^L \left( q_i \right) \right)' + a_2^2 \left( P^L \left( q_i \right) \right) \widehat{V}_{\widehat{\pi}}^{\frac{1}{2}} \widehat{V}_{\widehat{b}_N}^{\frac{1}{2}} a_2 \left( P^L \left( q_i, \widehat{Q}_g \right) \right)' \end{aligned}$$

where  $\widehat{V}_{\widehat{b}_N}$  is the variance of the rank estimator and

$$a_2^2 \left( P^L \left( q_i \right) \right) = \frac{1}{N} \sum_{g=1}^N \frac{\partial^2 P^L \left( q_i, \widehat{Q}_g, w_g, v_g \right)' \widehat{\pi}}{\partial Q_g^2} \cdot f \left( X_{1g}, X_{2g}, \dots, X_{M_g g} \right).$$

Notice that both variance estimators,  $\widehat{V}_{\widehat{\Phi}_1}$  and  $\widehat{V}_{\widehat{\Phi}_2}$ , are functions of the variance of the OLS estimator obtained in the third step,  $\widehat{\pi}$ . To find the variance  $\widehat{\pi}$ , we just do a Taylor expansion of the first order condition used to obtain  $\widehat{\pi}$ . This Taylor expansion allows us to write  $\widehat{\pi} - \pi_0$  as a linear combination of the rank estimator recovered in the first step of the estimation procedure, and the OLS coefficient obtained in the series estimation of  $\widehat{\Psi} \left( A_g, W_g \right)$ . To obtain the asymptotic distribution of  $\widehat{\pi}$ , we normalize the terms in the expansion by the rate of convergence of the rank estimator, since this estimator converges at a slower rate than residuals of the OLS estimation<sup>15</sup>. We can show that the variance of

<sup>15</sup>Another way to find the asymptotic distribution of  $\widehat{\pi} - \pi_0$  is to normalize each term by its own rate of convergence, and construct an adaptive estimator for the variance. This estimator of the variance is a generalization of the estimator we proposed in this section, and it is subject of future research.

$\sqrt{N\sigma_N}(\hat{\pi} - \pi_0)$  is

$$\begin{aligned} \widehat{V}_{\hat{\pi}}^{\frac{1}{2}} = & \left[ \frac{1}{N} \sum_{g=1}^N \sum_{i=1}^M r_{ig} \tau_{ig} P^L(t_{ig}) P^L(t_{ig})' \right]^{-1} \cdot \left\{ \left( \frac{1}{N} \sum_{g=1}^N \sum_{i=1}^M r_{ig} \tau_{ig} \frac{\partial \Lambda_0(t_{ig})}{\partial \beta} \cdot P^L(t_{ig}) \right) \right. \\ & + \left( \frac{1}{N} \sum_{g=1}^N \sum_{i=1}^M r_{ig} \tau_{ig} P^L(t_{ig}) \cdot \frac{\partial \Lambda_0(t_{ig})}{\partial \Psi} \cdot r^{L_1}(z_g) \right) \cdot \left( \frac{1}{N} \sum_{g=1}^N \sum_{i=1}^M r^{L_1}(z_g) r^{L_1}(z_g)' \right)^{-1} \\ & \left. \cdot \left( \frac{1}{N} \sum_{g=1}^N \sum_{i=1}^M r^{L_1}(z_g) f(x_{1g}, x_{2g}, \dots, x_{M_g g})' \right) \right\} \cdot V_{b_N}^{\frac{1}{2}}. \end{aligned}$$

To state the results of asymptotic normality of the average structural function and the estimator of peer effects, we need to impose some regularity conditions. The first condition assumes differentiability of  $\Lambda_0(t)$ . This condition assures that the derivatives of  $\Lambda_0(t)$  can be approximated by  $P^L(t)$ , which is important for the consistency of terms in the covariance matrices.

**Assumption 5.11**  $\Lambda_0(t)$  is twice continuously differentiable in  $t \equiv (q, q_g, w_g, \Psi - q_g)$  with bounded first and second derivatives.

The second condition restricts the rates of  $L_1$  and  $L$ , so the terms inside the covariance matrix converge.

**Assumption 5.12**  $\zeta(L_1)$  and  $\xi_d(L)$  are bounded away from zero as  $L_1$  and  $L$  grow, and

$$\begin{aligned} \sqrt{N\sigma_N} L^{-\vartheta} &\rightarrow 0 \text{ and } \sqrt{\sigma_N N} L_1^{-\vartheta_1} \rightarrow 0 \\ \frac{\zeta(L_1)^2 L_1}{N\sigma_N} &\rightarrow 0 \text{ and } \frac{\xi(L)^2 M^2 L}{N} \rightarrow 0 \\ \frac{(L\xi_1(L)^2 + \xi_0(L)^4 \zeta_0(L_1)^2) \cdot L_1^2 \cdot L}{N} &\rightarrow 0 \\ \frac{(\zeta(L_1)^4 \cdot \xi_1(L)^2 + L_1 \xi_1(L)^2) \cdot L_1^2}{N} &\rightarrow 0 \\ \frac{\xi_0(L)^2 \cdot \xi_1(L)^2 \cdot (L_1 \cdot L + L_1^2)}{N\sigma_N} &\rightarrow 0. \end{aligned}$$

Suppose that  $\sigma_N$ ,  $L_1$  and  $L$  are chosen in order to make the rates of bias and variance equal,  $\sigma_N \propto N^{-\frac{1}{2h+1}}$ ,  $L_1^{\frac{1}{1+2\vartheta_1}}$  and  $L^{\frac{1}{1+2\vartheta}}$ . In this case,  $\sqrt{N\sigma_N} L^{-\vartheta} \rightarrow 0$  and  $\sqrt{\sigma_N N} L_1^{-\vartheta_1} \rightarrow 0$  hold if  $h < \vartheta$  and  $h < \vartheta_1$ . These two side conditions hold if the rate of convergence of the rank estimator dominates the rate of convergence of the series estimator. These conditions impose that the rate of the bias of the series estimator shrinks faster than  $\frac{1}{\sqrt{N\sigma_N}}$ , which

is the rate of convergence of the variance of the rank estimator. Since we undersmooth,  $N\sigma_N^{2h+1} \rightarrow 0$ , or use the bias-corrected rank estimator, we don't need to control for the bias that comes from the first step. Given that, these conditions are very similar to the necessary conditions stated by Newey, Powell and Vella (1999) to find the asymptotic distributions of series estimators. Assumption 5.12 assures that the center of the limiting distribution is at zero.

**Theorem 5.5** *If assumptions 3.1-3.5, 5.1-5.7, 5.8-5.10, 5.11*

*and 5.12 hold, and if  $N\sigma_N^{2h+1} \rightarrow 0$  or  $\sigma_N = \left(\frac{\lambda}{N}\right)^{\frac{1}{2h+1}}$  and  $\widehat{b}_N = \widetilde{b}_N + \left(\frac{\lambda_N}{N}\right)^{\frac{1}{2h+1}} S_{2N}(\widetilde{b}_N, \sigma_N)^{-1} \widehat{S}_{1N}$ , then*

$$\sqrt{N\sigma_N} V_{\widehat{\Phi}_i}^{\frac{1}{2}}(\widehat{\Phi}_i - \Phi_{i0}) \rightarrow \mathcal{N}(0, 1) \quad \text{and} \quad \sqrt{N\sigma_N} \widehat{V}_{\widehat{\Phi}_i}^{\frac{1}{2}}(\widehat{\Phi}_i - \Phi_{i0}) \rightarrow \mathcal{N}(0, 1) \quad \text{for } i = 1, 2,$$

where

$$\begin{aligned} V_{\widehat{\Phi}_1}^{\frac{1}{2}} &= a_1 (P^L(t))' \cdot \left[ \sum_{i=1}^M \Pr[r_{ig} = 1] \cdot \mathbb{E}[\tau_{ig} P^L(t_i)' P^L(t_i)] \right]^{-1} \\ &\quad \left\{ \sum_{i=1}^M \Pr[r_{ig} = 1] \cdot \mathbb{E} \left[ \tau_{ig} \frac{\partial \Lambda_0(t_{ig})}{\partial \beta} \cdot P^L(t_{ig}) \right] \right. \\ &\quad + \sum_{i=1}^M \Pr[r_{ig} = 1] \cdot \mathbb{E} \left[ \tau_{ig} P^L(t_{ig}) \cdot \frac{\partial \Lambda_0(t_{ig})}{\partial \Psi} \cdot r^{L_1}(Z_g) \right] \\ &\quad \left. \cdot \mathbb{E} [r^{L_1}(Z_g)' r^{L_1}(Z_g)]^{-1} \cdot \mathbb{E} [f(X_{1g}, X_{2g}, \dots, X_{M_gg})' r^{L_1}(Z_g)] \right\} \cdot V_{\widetilde{b}_N}^{\frac{1}{2}} \end{aligned}$$

$$\text{with } a_1(P^L(t)) = p^L(q_i, q, \bar{v}, \bar{w}) - \left[ \frac{\mathbf{0}_{K+1 \times 1}}{\sum_{g=1}^N p_j(v_g, w_g)} \right].$$



$$\begin{aligned}
V_{\Phi_2}^{\frac{1}{2}} = & \left\{ a_2 (P^L (q_i))' \cdot \left( \sum_{i=1}^M \Pr [r_{ig} = 1] \cdot \mathbb{E} [\tau_{ig} P^L (t_i)' P^L (t_i)] \right)^{-1} \right. \\
& \cdot \left[ \sum_{i=1}^M \Pr [r_{ig} = 1] \cdot \mathbb{E} \left[ \tau_{ig} \frac{\partial \Lambda_0 (t_{ig})}{\partial \beta} \cdot P^L (t_{ig}) \right] \right. \\
& \left. \left. + \sum_{i=1}^M \Pr [r_{ig} = 1] \cdot \mathbb{E} \left[ \tau_{ig} P^L (t_{ig}) \cdot \frac{\partial \Lambda_0 (t_{ig})}{\partial \Psi} \cdot r^{L_1} (Z_g) \right] \right] \right. \\
& \cdot \mathbb{E} [r^{L_1} (Z_g)' r^{L_1} (Z_g)]^{-1} \cdot \mathbb{E} [f (X_{1g}, X_{2g}, \dots, X_{M_gg})' r^{L_1} (Z_g)] \left. + \mathbb{E} \left[ \frac{\partial^2 \Lambda_0 (t_{ig})}{\partial Q_g^2} \cdot f (X_{1g}, X_{2g}, \dots, X_{M_gg}) \right] \right\} \cdot V_{b_N}^{\frac{1}{2}} \\
& \text{with } a_2 (P^L (q_i)) = \mathbb{E} \left[ \frac{\partial P^L (q_i, Q_g, w_g, v_g)}{\partial Q_g} \right].
\end{aligned}$$

Theorem 5.5 states the asymptotic normality that determines the large sample confidence intervals for our estimators. In addition, this theorem shows that the estimators are not root-N consistent, but the rate of convergence depends on rate of the rank estimator and on how fast  $\widehat{V}_{\Phi_i}$  goes to infinity. There is not a complete characterization of the convergence rates of series estimators in the literature, although this theorem shows that our estimators will converge at a slower rate than the standard series estimators that are linear functionals of  $\Lambda_0 (t)$ , since we need to take into account the error that comes from the rank estimator in the first step. To center the distribution at zero, we need to remove the asymptotic bias of the rank estimator by undersmoothing or by using the bias-corrected estimator.

## 6 Empirical Application

In Brazil, the low quality of the public schools and the high drop out rates are related to poverty and high inequality rates. Some authors, Soares (2004) and Fletcher (1997), argue that to be able to improve the educational system in Brazil and increase the level of education it is necessary to understand which are the inputs in the production function of education and what is the impact of each one on the average achievement of the students. In the past, the analysis of the relationship between the inputs and outputs of the educational process in Brazil was very restricted, since the only measure of achievement available was the number of years of schooling completed by the student. This measure does not include

any information about the quality of the education obtained by the student, and does not allow comparisons among students that have the same number of years of schooling.

With the establishment of a new educational policies in the early 1990's, evaluation systems were created for the different levels of education: SAEB, for those finishing elementary schools, middle school or high school, and the Undergraduate National Exam, for those finishing college. Since the implementation of these evaluation systems, many researchers have been investigating the impact of different inputs (student characteristics, school infrastructure, teacher experience, teacher education, etc.) on student test scores. Using SAEB (2005), Fletcher (1997) uses a linear hierarchical model to determine which factors affect the math proficiency of students in the last year of middle school. This author finds that race, sex and the socioeconomic level<sup>16</sup> of the students have a significant impact on math proficiency. In addition, he finds that the average of the socioeconomic index in the classroom has the largest impact on student proficiency. Albernaz, Ferreira and Franco (2002) estimate the production function of education in Brazil for students in the last year of middle school using SAEB 1999. Using a linear hierarchical model, they find that the average of the socioeconomic index in the classroom is responsible for most of the variance in math test scores among classrooms. They also find evidence that in classrooms with a high average socioeconomic index, the effect of student's own socioeconomic index is smaller than in classrooms with a low average index. Using SAEB 2001, Franco et al (2004) find evidence that the average of the socioeconomic index in the classroom increases math test scores for students in the last year of elementary school by 5 points on average, and decreases by 3 points the coefficient associated with the socioeconomic index of the student. They find that students with a high socioeconomic index benefit more from "good" classrooms than students with a low index. "Good" classrooms are defined as the ones in which teachers assign and grade homework, and with a low percentage of students that are repeating the same grade. In Brazil, peer's characteristics seem to be an important input into the production function.

In addition to the evaluation systems, other policies were proposed in the early 90's. The Law of the Basis of National Education ("Lei de Diretrizes e Bases da Educação Nacional" - LDB) established in 1996 and the National Plan of Education ("Plano Nacional

---

<sup>16</sup>Since the data provided by SAEB does not have a measure of household income, they construct a measure of socioeconomic status of the students that is used as one of the student characteristics in the production function for education. This measure is an aggregated index that includes access to some consumer durable goods, like television, radio, telephone, computers, etc..

de Educação - PNE) established in 2001 changed the allocation system for educational resources and proposed new policies to guarantee high quality education for all kids between 6 and 14 years old in Brazil. According to the LDB, federal government is responsible not only for evaluating the education provided by private and public schools and colleges, but also for implementing a curriculum reformulation. State governments are responsible for regulating and controlling schools that offer basic education, including private ones. The state government is responsible for transferring educational resources to schools, and for establishing mechanisms to ensure that these resources are used properly. Decisions related to pedagogical, administrative and financial matters are made at the school level. The Principal, with the community<sup>17</sup>, has autonomy to decide how to allocate resources to classrooms, including financial resources, textbooks, teachers and students. In this section, we model this decision problem at the school level.

In a previous decision problem, the parents decide which schools their kids can attend based on the characteristics of the schools ( $X_s$ ) and costs (like the distance of the school from the house, tuition in the case of private schools, etc.). Based on the supply of students, the school selects which individuals are going to attend its classes. We assume that the school chooses an admission rule that maximizes the total quality of the students in the school. Examples of admission rules are: select students that live in the neighborhood, admission exam, etc. Let  $N$  be the total number of classrooms in the school, and  $M_g$  the number of students in classroom (group)  $g$ . The total quality of the students in the school is equal to the sum of the quality index for all students in the school,  $Q_s = \sum_{g=1}^N \sum_{i=1}^{M_g} Q_i$ . The total quality of the students in school is determined in this previous decision problem and is a function of school characteristics ( $X_s$ ) and the admission rule chosen by the school ( $A_s$ ),  $Q_s \equiv Q(A_s, X_s)$

Given the total quality of the students in the school, the Principal, with the community, allocates resources and students to classrooms. For simplicity, we assume that the school only has students in the last year of elementary school, and the size of the classrooms is fixed. Each classroom has the same number of students, which is equal to the total number

---

<sup>17</sup>Community is defined as teachers, staff and some parents that participate in school activities. The Principal can create a council to make decisions, or he can decide by himself and get the approval of the community. Some decisions are made by the Principal himself, or by the Principal and teachers, like decisions on pedagogical programs and how to allocate students and teachers to classrooms. The majority of the decisions at the school level are made by the Principal and the community. For simplicity, we are going to describe the decision problem, we are going to refer as the main agent in the decision problem as "the school".

of students in the school ( $T$ ) divided by number of classrooms,  $M_g = \frac{T}{N}$ . We assume that the school chooses the allocation rule that maximizes the total achievement of the students subject to financial constraints and the education production function, represented by  $H(\cdot)$ . This production function relates student's achievement with student's quality and peer's quality. The school solves the following decision problem

$$\max_{A_g, A_t, B_g} \sum_{g=1}^N \sum_{i=1}^{M_g} \mathbb{E}[Y_{ig} | q_{ig}, q_g, v_g, b_g, t_g, M_g, X_s, A_g]$$

subject to

$$Y_{ig} = H(Q_{ig}, Q_g) + \alpha_g + \varepsilon_{ig}$$

$$Q_g \equiv Q_g(A_g, f(Q_i(A_s, X_s)))$$

$$t_g \equiv t_g(A_t, f(X_t))$$

$$\sum_{g=1}^N w(t_g) + \sum_{g=1}^N P_B B_g = R$$

where  $B_g$  is the total of material received by classroom  $g$  (including textbooks, teaching material, etc),  $t_g$  is the quality of the teacher in group  $g$ , which is a function of the way teachers are allocated to classrooms ( $A_t$ ) and the distribution of teacher characteristics in the school ( $X_t$ ),  $w(t_g)$  is the salary of the teacher,  $R$  is the total amount of money that the school can spend,  $A_g$  represents the allocation rule that assigns students to classrooms,  $f(Q_i(A_s, X_s))$  represents the distribution of student quality in the school, which is a function of total quality of the students in the school,  $Q_g$  represents the average quality of the students in classroom  $g$ ,  $Q_{ig}$  is the quality of individual  $i$  in group  $g$ , and  $Y_{ig}$ , the achievement of student  $i$  in group  $g$ . There are three unobservable components in this model,  $\alpha_g$ ,  $\varepsilon_{ig}$  and  $v_g$ . In this case,  $\alpha_g$  and  $\varepsilon_{ig}$  represent the correlated effects and the individual heterogeneity in the production function of achievement, and  $v_g$  represents a noisy signal of the correlated effects  $\alpha_g$ .

According to this decision problem, the school maximizes the total achievement of the students in the school by choosing the teacher, the quantity of material, and group composition in each classroom. The quality of the teacher and the average of the quality of the students in the classroom is determined by the allocation rules chosen by the school. By choosing the student and teacher allocation rules, the school determines  $Q_g$  and  $t_g$ . One

important assumption of this model is that conditional on school and classroom characteristics, the student allocation rule only affects total achievement through its effect on the quality of the peers.

The solution of this model leads to the triangular system of equations used in this application,

$$Y_{ig} = H(Q_{ig}, Q_g) + \alpha_g + \varepsilon_{ig} \text{ for } i = 1, \dots, M_g \text{ and } g = 1, \dots, N \quad (2)$$

$$Q_g = \Psi(A_g, W_g) + v_g \text{ for } g = 1, \dots, N \quad (3)$$

where

(i)  $\mathbb{E}[\varepsilon_{ig} | Q_{1g}, \dots, Q_{M_g g}, \alpha_g] = 0$

(ii)  $\mathbb{E}[v_g | A_g, W_g] = 0$

(iii)  $\mathbb{E}[\alpha_g | v_g, A_g, W_g] = \mathbb{E}[\alpha_g | v_g, W_g] = \Gamma(v_g, W_g)$ .

In this system of equations, the vector  $W_g = [X_s, X_t, M_g, A_s, A_t, B_g]$  includes school characteristics, classroom characteristics (including teacher characteristics), the admission rule and teacher allocation rule. If this applied exercise,  $f(X_{1g}, X_{2g}, \dots, X_{M_g g}) = \frac{1}{N} \sum_{i=1}^{M_g} X_i$ .

Notice that the assumption that  $\mathbb{E}[\alpha_g | v_g, A_g, W_g] = \mathbb{E}[\alpha_g | v_g, W_g]$  holds if controlling for school and classroom characteristics and the other allocation rules,  $\alpha_g$  varies independently of the way students are allocated to classrooms. Conditional on the characteristics of the groups and the other allocation rules, all the productivity effects of student allocation operate through their effect on peer composition. This exclusion restriction can be violated in many situations. Suppose that there is an input that the Principal can observe, but the econometrician does not know about, like a measure of teacher quality, and this input is correlated with the way students were allocated to classrooms. In this case, even when we control for the characteristics of schools and classrooms, the student allocation rule will be correlated with  $\alpha_g$ , and the exclusion restriction will be violated. Or suppose that there is an unobservable political process (for example parent's pressuring the school to select some allocation rule) that affects the decision made by the school, but does not have a clear relationship with school and classroom characteristics. As before, the exclusion restriction will be violated. In this model, we assume that the assignment of students is determined entirely by the allocation rule that is known by the econometrician.

This very simple decision problem deals with the main source of selection bias presented in models that estimate peer effects in classrooms, non-random assignment of the students to classroom, by assuming that the assignment of students to classrooms is only determined by the allocation rule chosen by the school, and that this allocation rule is chosen based on the observable characteristics of the school, teachers and students in this school. We investigate the validity of our exclusion restriction in section 6.3. For now, we assume that exclusion restriction is valid for public and private schools in Brazil.

### **6.1 Description of the Data: SAEB 2003**

The Brazilian National Evaluation System of Basic Education (SAEB) is a biannual survey conducted by the National Institute for Educational Studies and Research (INEP). SAEB evaluates students in the last year of each education level (elementary school, middle school and high school) in two subjects: Portuguese and math. In this survey, students not only take standard exams in Portuguese and math, but also fill a questionnaire that contains information about their socioeconomic status, behavior towards learning and parent participation in the educational process. Teachers and Principals also answer contextual questionnaires on teaching practices, management and socioeconomic background. In addition, this survey collects information about the infrastructure in the school, for example, availability of textbooks to the students, if the classrooms have air circulation and enough light, etc.

In this applied exercise, we are going to use data from SAEB 2003. SAEB collects information from a sample of students in urban school in all states in Brazil. In 2003, students in the last year of elementary schools in rural schools with more than 10 students in this grade were added to the SAEB sample.

One interesting aspect about the SAEB 2003 is that it includes information on how students and teachers are allocated to classrooms, which is important data in the analysis of peer effects. In addition, it has information regarding admission rules used by the Principal to select the students that are going to attend the school. However, this data does not include an important characteristic of students' family backgrounds, household income. To deal with this lack of information, we follow the literature and construct an aggregate index that represents the socioeconomic index of the students. This aggregate index is created by a Principal Component Analysis of twelve items in the questionnaire. These items are related to the existence of some consumer durable goods in the household

and to the infrastructure of the household. A detailed description of this index can be found in Appendix A.

In this application, we focus on math test scores for students in the last year of elementary school. We focus on students in the last year of elementary school because peer effects are usually stronger in classrooms in elementary school than in classrooms in high school. The size of classrooms in elementary school is usually smaller than in high school, and kids less than 10 years old do not have a network outside the classroom; they largely interact with their classmates, teachers and their parents. Since students learn Portuguese not only at school, but also at home through reading habits, talks with their parents, etc., there may be some unobservable components at the individual level that affect Portuguese test scores, but not math test scores. To avoid dealing with the impact of these unobservable components on student achievement, we focus on math test scores.

The sample used in this application has 20,137 students allocated in 2,687 classrooms in 2,117 schools. The math test scores for the students will be the measure of achievement for students in their last year of elementary school. We use sex, age, race, parent education, the socioeconomic index and a measure of student background in school as student's characteristics. The vector  $X$  includes a dummy variable that equals 1 for females; a variable for age which is a discrete variable that varies from 8 until 15; a dummy that equals 1 if the student is white; an aggregate index that represents the socioeconomic level of the student; a variable for parent schooling which is a discrete variable that represents the number of years of schooling for the parent with the most education; and a dummy that equals 1 if the student attended kindergarten<sup>18</sup>.

The vector of controls ( $W$ ) includes school and classroom characteristics. We consider the following school and classroom characteristics: location (region in Brazil, and rural or urban areas), type of school (dummy variable that equals 1 for private, and 0 for public schools), characteristics of the teachers in each classroom (years of schooling, experience (measured by the number of years as a teacher), gender and race), homework assignment (dummy variable that equals 0 if the teacher does not assign homework, 1 if the teacher

---

<sup>18</sup>In a previous version of this paper, we include parent attendance at school meetings as a student characteristic. However, this variable does not have a significant impact on student achievement, and we decide to exclude it from the analysis. In the vector  $X$ , we use the maximum years of schooling between the father and mother as the measure of parent education because of the great number of missing values in the variables that define education of the mother and the education of the father. Parental education is a discrete variable. Suppose that  $E$  indicates years of schooling. In this case,

$E = \max(E_{mother}, E_{father})$  and equals 1 if  $E = 0$ , 2 if  $E < 4$ , 3 if  $E = 4$ , 4 if  $4 < E < 8$ , 5 if  $E = 8$ , 6 if  $8 < E < 11$ , 7 if  $E = 11$ , 8 if  $12 < E < 16$ , 9:  $E \geq 16$ .

assigns homework, but does not grade it, and 2 if the teacher assigns and grades the homework), total number of students in the classroom, the principal's characteristics (education, experience (measured by the number of years as a principal in the school), age, race and gender), a variable that measures if the classrooms have light and air circulation, and a variable that equals 1 if the classroom has the materials necessary for teaching (eraser, chalk, etc.).

We also add to this vector the admission rule and teacher allocation rule chosen by the school. According to SAEB, the school can select students using five criterion: an admission exam, lottery, proximity of student house to school, first come first served, and other criterion. In this data, 4.34% of the schools use an admission exam, 1.13% use lottery, 22.48% choose their students based on neighborhood, 23.25% give the spots to the first students in line, 13.22% select the students using another criterion and 35.59% have no criterion to select the students<sup>19</sup>. The school can allocate teachers to classrooms based on seven criterion: respect the preferences of the teachers, more experienced teachers are assigned to students that learn faster, more experienced teachers are assigned to students that learn slower, keep the same teacher with the same classroom, switch teachers among grades, hold a lottery or use another criterion. 23.11% of schools in the sample respect the preference of the teachers, 20.08% switch teachers among grades and 10.5% do not have a specific way to allocate teachers to classrooms.

Following the simple model presented in the previous section, the vector of instruments ( $A_g$ ) will contain dummy variables that will represent how the school allocates students to classrooms. There are four ways that the school can assign students to classrooms<sup>20</sup>: integration by age, segregation by age, integration by score, segregation by score. In the data, 42.30% of the schools use integration by age, 8.28% of the schools use integration by score, 8.99% of the schools use segregation by age, 15.84% use segregation by score and 24.59% have no assigned rule. It's interesting to note that the percentage of private schools that adopt each one of the allocation rules is the same as the percentage of public schools.

Table 1 presents the summary statistics of some of variables used in this applied exercise. Figure 1 shows the distribution of the math test scores in this data. The standard test was designed in such a way that the average student in the last year of elementary school

---

<sup>19</sup>It is interesting to note that 54% of the rural schools do not adopt any criterion to select the students, and 39% of the private schools. In addition, there are no private schools in the sample that use the lottery criterion.

<sup>20</sup>The survey asked the Principal if he used one of this 4 options and if he has no assigned rule.



should have a score of 250. This graph shows that the distribution of test scores is shifted to the left, with few students with test scores above 250. Test scores vary from 72 to 369, and the average test score is 193.

Figure 2 shows how test scores vary with student characteristics (sex, race, parent's education, and if the student attended kindergarten). This graph shows that males and whites have higher math test scores on average than females and non-whites. In addition, students that attended kindergarten or have parents with a college degree have a higher achievement on average than other students.

Figure 3 shows how test scores vary with some school characteristics (location, and if it is public or private). This graph shows a huge gap between test scores in public and private schools. This result was expected since private schools concentrate students with a "good" family background, and we have evidence that the characteristics of a student's peer group impact the achievement of the student. In this case, "good" family background is measure by higher household income and parents' education. In addition, Figure 3 shows that test scores vary by the location of schools. Students in the Southeast region have higher test scores than students in other regions, and students in urban areas have higher test scores than students in rural areas. There is evidence that the quality of teachers and the infrastructure are better in urban areas than in rural areas, which can explain part of the difference in average test scores among areas in Brazil.

Figure 4 shows how test scores vary with the admission and teacher allocation rules. The schools that choose their students based on an exam have the best test scores. This graph also shows that schools that keep the same teacher with the same classroom have students with higher test scores on average than other schools. Figure 5 shows how test scores vary with the student allocation rule. The schools that segregate students by score among classrooms have better student achievement than the others.

In the next section, we use the data described in this section to estimate the average production function and peer effects.

## 6.2 Results

In this section, we estimate the average production function and the average marginal derivative using the new semiparametric procedure described in section 4. The results obtained using this methodology are compared with the results of a linear-in-means model.

### 6.2.1 Baseline Model: linear-in-means

In this section, we estimate the linear-in-means model using a three step procedure. In this model, the production function is linear in both quality of students and peers. We assume that the function that relates peer quality with student allocation rule is also linear, and that the control function is a linear function of the unobservable component of peer quality ( $v_g$ ) and of the vector of school and classroom characteristics ( $W_g$ ). Thus,

$$Y_{ig} = H(Q_{ig}, Q_g) + \alpha_g + \varepsilon_{ig} = \pi_0 + Q_{ig} + \pi_1 Q_g + \alpha_g + \varepsilon_{ig} \quad (4)$$

$$Q_g = \Psi(A_g, W_g) = \gamma_0 + \gamma'_1 A_g + \gamma'_2 W_g + v_g \quad (5)$$

$$\Gamma(v_g, W_g) = \pi_2 v_g + \pi'_3 W_g \quad (6)$$

with  $Q_{ig} = X'_{ig}\beta$  and  $Q_g = \bar{X}'_g\beta$ .

The parameter that identifies peer effects is  $\pi_1$ . As in the semiparametric model,  $\pi_1$  represents the marginal effect of student quality on student achievement. The average production function is just the linear function,  $H(Q_{ig}, Q_g) = \pi_0 + Q_{ig} + \pi_1 Q_g$ , evaluated at specific values of peer and student quality.

The parameters of this model are identified and estimated using a three step procedure. In the first step, we use the within-group variation to identify and estimate  $\beta$ . In this step, we construct the "within variation" version of equation 4 by subtracting the average achievement in each classroom,  $\bar{Y}_g = \frac{\sum_{i=1}^N Y_{ig}}{N} = \frac{\sum_{i=1}^N H(Q_{ig}, Q_g, \alpha_g, \varepsilon_{ig})}{N}$ ,

$$Y_{ig} - \bar{Y}_g = \beta' (X_{ig} - \bar{X}_g) + \varepsilon_{ig} - \varepsilon_g. \quad (7)$$

Equation 7 does not include the correlated effect ( $\alpha_g$ ), and we can identify  $\beta$  using the assumption that  $\mathbb{E}[\varepsilon_{ig} | Q_{1g}, \dots, Q_{M_g g}, \alpha_g] = 0$ . This equation is estimated by Ordinary Least Squares. In this step, we estimate  $\hat{\beta}$ , and recover the indexes that represent quality of the student and quality of his peers,  $Q_{ig} = X'_{ig}\hat{\beta}$  and  $Q_g = \bar{X}'_g\hat{\beta}$ .

To identify the other parameters in this linear-in-means model, we use the between-group variation. To deal with the endogeneity of  $Q_g$ , we use the control function approach proposed by Newey, Powell and Vella (1999). In the last two steps, we identify the parameters in equations 5 and 6 based on the following assumptions:  $\mathbb{E}[v_g | A_g, W_g] = 0$  and  $\mathbb{E}[\alpha_g | v_g, A_g, W_g] = \mathbb{E}[\alpha_g | v_g, W_g] = \Gamma(v_g, W_g)$ . In the second step, we estimate equation 5 by Ordinary Least Squares, using the peer quality estimated in the first step,  $\hat{Q}_g$ . In this

step, we recover the residuals of the OLS regression,  $\hat{v}_g$ . In the last step, we use  $\hat{Q}_g$  and  $\hat{v}_g$  obtained in first and second steps to estimate the following equation by OLS,

$$\bar{Y}_g = \pi_0 + (1 - \pi_1)\hat{Q}_g + \pi_2\hat{v}_g + \pi_3'W_g + \varphi_g.$$

In the last step, we obtain the estimated value of peer effects,  $\hat{\pi}_1$ , and the other coefficients in the average production function. The standard deviation of these coefficients is obtained using a GMM approach.

Table 2 presents the results of the first step. This table shows that all the coefficients, except for the socioeconomic index, have a significant impact on student test scores. The sign of the coefficients are the same as the ones found in the literature that estimate linear models using SAEB. Males and whites have higher math test scores on average than females and nonwhites, and parent education and kindergarten attendance have a positive impact on student outcome. Table 4 presents the results for the second and third steps of this parametric procedure. The first column of this table shows that the student allocation rules have a significant impact on peer quality, which indicates that there is a strong relationship between the allocation system and peer quality. The coefficients of the allocation rules are positive. Since the exclusion category is allocation based on segregation by age, these results indicate that the schools that integrate by age or allocate students based on score have better peer groups on average than the schools that segregate by age. The second column of table 4 shows the results obtained in the third step of the estimation procedure. This result indicates that peer quality has a positive effect on student achievement, but this effect is not significant. One of the limitations of this baseline model is that it does not allow peer effects to vary with peer quality. It may be the case that high quality students benefit more from an increase in peer quality than a student with low quality, or vice versa. The results of the semiparametric model presented in the next section will address this point.

### 6.2.2 Flexible Functional Form

In the semiparametric model, we estimate the average production function and peer effects without imposing a functional form for the production function ( $H(Q_{ig}, Q_g)$ ). We estimate the triangular system of equations represented by equations 2 and 3, using the semiparametric methodology described in section 4.

In the first step of this procedure, we estimate the parameters in the index that defines

student quality ( $\beta$ ), by finding the global maximum of the function,  $S_N(b, \sigma_N)$ . Since this function has many local maximums, we use a global search algorithm, called simulated annealing, to maximize the function  $S_N(b, \sigma_N)$ . This function has two smooth parameters as arguments: kernel ( $\mathbf{K}(\cdot)$ ) and the bandwidth ( $\sigma_N$ ). In this applied exercise, we use a fourth order kernel ( $h = 4$ )<sup>21</sup>. The bandwidth is selected using the plug-in method proposed by Horowitz (1992). This method chooses the bandwidth that minimizes the asymptotic mean square error of the rank estimator. We are going to refer to this bandwidth as the optimal ( $\sigma_N^*$ ). With this bandwidth, we can calculate the asymptotically bias-corrected rank estimator. In addition, we consider two other values for the bandwidth, half of the optimal bandwidth ( $0.5 \cdot \sigma_N^*$ ) and 75% of this optimal bandwidth ( $0.75 \cdot \sigma_N^*$ ). When we use these two different bandwidths, we are undersmoothing, and the asymptotic bias of the estimator goes to zero. In order to satisfy assumptions (vii) and (viii), we normalize the coefficient associated with the socioeconomic index to 1. This random variable is the only one that attains more than five discrete values, being close to be continuously distributed with a positive measure over the entire support. Notice that with this normalization, the quality index is defined in the units of the socioeconomic index, since we are normalizing all the coefficients in the index by the coefficient associated with socioeconomic index. The magnitude of the index that estimates peer quality will change depending on which coefficient has been normalized, thus the magnitude of peer effects will change. The magnitude of the peer effects represents the marginal effect of peer quality on student achievement in units of the socioeconomic index. In this optimization, we use 103,667 combinations of  $(y_{ig}, y_{jg})$  for  $i \neq j$ .

Table 2 shows the values for the asymptotically bias-correct rank estimator, and the values obtained using the other two bandwidths,  $0.5 \cdot \sigma_N^*$  and  $0.75 \cdot \sigma_N^*$ . In all the three cases, the coefficients have the same signs as in the parametric case. Age and female have a negative impact on test scores, while white, parent schooling and kindergarten attendance have a positive impact on student achievement. In the estimation with the optimal bandwidth, all the coefficients are significant. However, when we undersmooth, the standard deviation of the coefficients increases. Using  $0.75 \cdot \sigma_N^*$  as the bandwidth, we

---

<sup>21</sup>We use the integral of fourth order kernel for nonparametric estimation proposed by Muller (1984),

$$K_4(v) = \begin{cases} 0 & \text{if } v < -1 \\ \left(\frac{105}{64}\right) (1 - 5v^2 - 7v^5 - 3v^7) & \text{if } -1 \leq v \leq 1 \\ 1 & \text{if } v > 1 \end{cases}$$

get coefficients with large sizes, but standard deviations that are not very different from the ones obtained using the optimal bandwidth. When we decrease the bandwidth to  $0.5 \cdot \sigma_N^*$ , the sizes of the coefficients decrease, getting close the values of the asymptotically unbiased estimator, but the variance explodes. Since in this applied exercise we are interested in point estimation, we use the results obtained using the optimal bandwidth.

Notice that the coefficients in the parametric case are similar to the ones obtained with the optimal bandwidth. Table 3 compares the estimated values obtained for student quality and peer quality in the parametric and semiparametric cases. This table shows that the maximum value and the median of the quality distribution is almost the same in the parametric and semiparametric cases. However, the values obtained in the parametric case have higher standard deviations and higher minimums than in the nonparametric case. It appears that in the parametric case, we obtain more negative values for quality. In both cases, the mean and median of the quality distribution is negative, which indicates that a large part of the distribution of student quality is concentrated in the negative part of its support.

In the second and third steps of the estimation procedure, we use series approximation to estimate the conditional expectation of peer quality and the conditional expectation of the outcome. In this applied exercise, we use polynomial approximation functions. Let  $\mu = (\mu_1, \dots, \mu_{d_1})$  be a vector of nonnegative integers and  $Z^\mu \equiv \prod_{j=1}^{d_1} Z_j^{\mu_j}$ . For a sequence  $(\mu(l))_{l=1}^\infty$  of vectors with the same dimension as  $Z$ , a power series approximation in the second step is

$$r^{L_1}(Z) = \left( Z^{\mu(1)}, \dots, Z^{\mu(L_1)} \right)'.$$

Analogously, to define a power series approximation in the third step, let  $(\mu(l))_{l=1}^\infty$  denote a sequence of vectors with the same dimension as  $T$ , such that for each  $l$ ,  $T^{\mu(l)}$  depends only on  $(Q_{ig}, Q_g)$  or on  $(v_g, W_g)$ , but not on both. In this case, the power series approximation in the third step is

$$p^L(T) = \left( T^{\mu(1)}, \dots, T^{\mu(L)} \right)'.$$

In both steps, we use orthogonal polynomials. We just replace  $Z^{\mu(\cdot)}$  and  $T^{\mu(\cdot)}$  by a product of univariate polynomials with the same order. These univariate polynomials are orthogonal<sup>22</sup>. As pointed out by Newey, Powell and Vella (1999), orthogonal polynomials

---

<sup>22</sup>In this estimation, we use the simplest orthogonal polynomials in the interval  $[-1, 1]$ , Legendre Poly-

may reduce colinearity. To use orthogonal polynomials, we need to normalize our variables to have values between  $-1$  and  $1$ . With this normalization, students with a good background will have quality close to  $1$ , and values at the bottom of the distribution of quality will be around  $-1$ .

In the third step, we use nonrandom trimming to deal with outliers. We exclude from our samples classrooms with a very low quality or very high quality. We restrict our sample to classrooms with quality between  $-0.2$  and  $0.8$ . This restricted sample corresponds to  $99\%$  of the classrooms in the original sample. In the new sample, we have  $18,981$  students allocated among  $2,662$  classrooms.

The number of nonlinear terms in the polynomials is chosen by standard least squares cross-validation<sup>23</sup>. We first choose the number of terms that will approximate  $\Psi(A_g, W_g)$ . Then, we get the residuals of the model estimated in the second step, and choose the number of terms in the polynomial series approximation of  $\Lambda(\hat{t})$  by cross-validation, using the vector  $(\hat{q}_{ig}, \hat{q}_g, w_g, \hat{v}_g)$  as the vector of independent variables. A third order polynomial minimizes the CV criterion for the second step, and a polynomial of order four minimizes the CV criterion for the third step.

Using the results obtained in the series approximation of  $\Psi(A_g, W_g)$  and  $\Lambda(\hat{t})$ , we obtain the estimated average production function and the peer effects. Figure 6 plots the average production function against peer quality and student quality. This graph shows that test scores are monotonically increasing with student quality for students with quality greater than or equal to  $-0.5$ . However, tests scores do not have a monotonic relationship with peer quality. This graph indicates that test scores decrease with peer quality for students with very low quality, but increase with peer quality for students with quality above  $0.2$ . To gain a better understanding of the relationship between test scores and peer

---

nomial. The components of a univariate legendre polynomial are:  $P_0(x) = 1$ ,  $P_1(x) = x$ ,  $P_2(x) = \frac{3x^2-1}{2}$ ,  $P_3(x) = \frac{5x^3-3x}{2}$ ,  $P_4(x) = \frac{35x^4-30x^2+3}{8}$ ,  $P_5(x) = \frac{63x^5-70x^3+15x}{8}$ ,  $P_6(x) = \frac{231x^6-315x^4+105x^2-5}{16}$ . To use these orthogonal polynomials, we need to normalize our independent variables in such a way that they have values between  $[-1, 1]$ . We do that by applying the following normalization,

$$x^* = \frac{2x - a - b}{b - a}$$

where  $x \in [a, b]$  and  $x^* \in [-1, 1]$ .

<sup>23</sup>In this case, we choose  $L_1$  and  $L$  that minimize the estimated sum of the predicted error squared,  $\frac{\sum_{g=1}^N (\Psi(z_g) - \hat{\Psi}_{-g}(z_g))^2}{N}$  and  $\frac{\sum_{g=1}^N \sum_{i=1}^{M_g} (\Lambda(\hat{t}_i) - \hat{\Lambda}_{-i}(\hat{t}_i))^2}{N}$ . We include one component of the polynomial each time, and calculate the predicted error. We include as many terms as possible until we have a rank deficient matrix of covariates and cannot fit the model by OLS. For the values found in the cross validation procedure, contact the author.

quality, we fix the quality of the students at three different levels. One level represents a high quality student and is equal to the upper quartile of the distribution of student quality, another characterizes a low quality student and is equal to the lower quartile of the distribution of student quality, and a third represents the average student and is equal to the mean of the quality distribution<sup>24</sup>. Figure 7 shows the average production function at these three values of student quality, and includes confidence intervals. This figure shows that test scores do not vary much with peer quality. For a low quality student, test scores decrease with peer quality. However, this negative slope is not significant. For an average student, test scores increase with peer quality, however the effect of peer quality on student outcome is small. For a high quality student, test scores decrease with peer quality for groups with quality lower than 0.5, but increase with peer quality in high quality groups.

In Figure 8, we fix peer quality at three levels, and analyze how test scores vary with student quality. As in figure 7, we fix peer quality at three levels: at the upper quartile, at the lower quartile and at the mean of the distribution of quality<sup>25</sup>. This figure shows that tests scores are monotonically increasing with student quality for students with quality above -0.6, and this relationship does not change with peer quality. Figure 9 plots peer effects against student quality. This graph shows that peer effects are positive for students with quality greater than -0.4, and increase with student quality for students with quality lower than 0.7. For students with quality larger than 0.7, peer effects decrease with student quality, however this descending slope is not significant. This graph indicates that the marginal benefit of an increase in peer quality is larger for students with average quality than for students with low quality.

In summary, these graphs show that test scores do not have a monotonic relationship with peer quality, but are monotonically increasing with student quality. An increase in peer quality has a positive effect on tests scores; however, this effect is bigger for an average student. As we argued before, the results obtained in this section cannot define the optimal way to allocate students in the last year of elementary school to classrooms in schools in Brazil. The optimal way to allocate students to classrooms depends on the distribution of student quality in the school. In a school with a large contingent of average students, segregation by type can be optimal, since the marginal benefits obtained by the average

---

<sup>24</sup>In this restricted sample, the upper quartile of the distribution of student's quality is 0.6470, the lower quartile is -0.4863 and the average -0.4863.

<sup>25</sup>The upper quartile of peer's quality distribution is 0.7392, the lower quartile is -0.1305 and the mean corresponds to 0.2602.

quality students can compensate the marginal losses of a small contingent of low quality students. However, the opposite will happen in a school with a large fraction of low types.

When we compare the results obtained in this section with the linear-in-means model, we can see that the linear-in-means model does not take into account nonlinearities in the production function that are important in the estimation of peer effects. In particular, using this semiparametric methodology, we obtain peer effects that are an increasing function of student quality.

One of the assumptions used to identify the average structural function and the average derivative is the support condition, assumption 3.7. This assumption establishes that we can move the control variables over their entire support. Since our instrumental variable is a vector of dummy variables, this condition is violated. One way to deal with this problem is to restrict the format of the function  $\Psi(A_g, W_g)$  in a such a way that we can identify the parameters of interest under discrete support. Another way to deal with that is to examine in which range of  $(Q_{ig}, Q_g)$  the support condition is identified. In this range, we may identify some bounds for the parameters of interest, instead of point identifying the average structural function and the marginal derivative. Both of these approaches will be developed in future work.

### 6.3 Discussion: Exclusion Restriction

The student allocation rule is a valid instrument under two conditions: 1) peer quality is correlated with the allocation system; 2) if, conditional on school characteristics (including the admission rule and teacher allocation rule) and classroom characteristics, the correlated effects ( $\alpha_g$ ) vary independently of the way students are allocated to classrooms. The idea behind this instrument is that the correlation between the unobservable correlated effects ( $\alpha_g$ ) and peer' quality ( $Q_g$ ) arises from the fact that students are not randomly selected into classrooms. The allocation rule allows us to control for the way students are selected into classrooms if, conditional on school and classroom characteristics, these allocation rules only impact student achievement through their effect on group composition. We cannot test this exclusion restriction, but in this section we analyze situations in which this restriction may be violated, and examine the impact of this violation on the results obtained in the previous section.

As we argue in the previous sections, if there is an input that is not observable by the econometrician, but is observable by the Principal, and this input is correlated with



the decisions made at the school, then the exclusion restriction is violated. To check this hypothesis, we examine how other decisions made at the school level vary with the student allocation rule. The idea is that if there is some unobservable component that is related to the student allocation rule, this component should impact not only the student allocation rule but also other decisions made by the school. For example, if some parents can influence the school's decision, they will pressure the school not only to have their kids in high quality peer groups, but also to put their kids in classrooms that have better teachers, to provide tutoring for students that fall behind and can be a bad influence in the classrooms, etc. If there are no unobservable variables, conditional on school characteristics, we shouldn't observe a pattern in the decisions made at the school level.

Table 5 shows how other decisions made at the school level (school admission, teachers allocation rule, etc.) vary with how students were allocated to classrooms. In the last column of this table, we present a Wald test for the null hypothesis that decisions made at the school level shouldn't be statistically different among the student allocation rules. This table shows that only 58% of the schools that have no specific criterion to allocate students to classrooms provide training for the teachers, while 68% of the schools that allocated students based on segregation by age provide training for teachers. The difference in the percentage of schools that provide training programs for teachers among the allocation rules is statistically significant. We can also see that 61% of the schools that do not have a student allocation rule does not have an admission rule, while more than 50% of the schools that choose one of the allocation criteria have an admission rule. Again, the difference in the percentage of Principals that select each one of the admission rules among the allocations rules is significant. 24% of the schools that have no specific criterion to allocate students into classrooms also have no specific way to allocate teachers to classrooms, while 28% of the schools that choose to segregate students by score allocate teachers to classrooms based on teacher's preferences. In addition, only 25% of the schools that have no student's allocation rule have a program to reduce drop out, while 41% of the schools that segregate students by age have a program to reduce drop out. This table indicates that the schools that have no specific way to allocate students to classrooms are also the ones without programs to reduce drop out or grade repetition. Without controlling for school characteristics, we cannot rule out the hypothesis that there are some unobservable components that can affect all the decisions made at the school level. For example, a school that suffers pressure to improve the learning process in the school (due to parental pressure, or high ability of the teachers, etc.) segregates students by score, allocates teachers according to their preferences, has a

program to reduce grade repetition, and provides tutoring for the students.

Given the evidence in this table, we run a robustness test. In the vector of controls, we include whether the school provides training for teachers, whether the school has a drop-out program, whether it has a grade-repetition program and whether school provides tutoring for students( $W$ )<sup>26</sup>. Figure 10 presents the average production function obtained using a control function that includes these variables in the school's characteristics. As in Graph 6, test scores are increasing with student quality. However, different from Graph 6, this graph shows that test scores are increasing with peer quality. Figure 11 shows the estimated peer effects. Peer effects are positive and increase with student quality. In this graph, peer effects are larger than in Graph 9. These graphs indicate that if there is a bias in our estimation due to an unobservable component that is correlated with the decisions made at the school level, this bias is underestimating the marginal effect of peer quality on student achievement; although the sign of peer effects is always positive.

We believe that the regulations are stronger in public schools than in private schools in Brazil. In private schools, parents and teachers may have the power to reallocate student and teacher to classrooms, and the decision rule made at the school level may not be binding. In this case, the exclusion restriction is violated. To investigate this possibility, we restrict our sample to public schools and estimate the average production function and peer effects. The drawback of restricting the sample to public school is that if high quality students are concentrated in private schools, we are missing an important part of the distribution of student quality.

In the SAEB sample, we have 14,149 students allocated into 1,769 classrooms in public schools. Public schools correspond to 66% of the classrooms represented in the data. As we expect, the distribution of quality obtained using the public school sample is different from the one obtained using the full sample. Student quality varies from -35.13 to 12.05, with a standard deviation of 7.01 and a mean of -7.71. Peer quality varies from -34.84 to 4.94, with a standard deviation of 5.27 and a mean of -8.24. Comparing these values with those obtained in Table 3, we confirm that the very good students are concentrated

---

<sup>26</sup>We didn't add these variables in the estimation in the previous section, because these variables can come from different decisions problems. The Principal can decide if he is going to provide training for the teachers, tutoring, etc. after he allocated students to classrooms, and consequently these programs could be induced by the allocation rules, and shouldn't impact peer quality. For example, perhaps if the Principal chooses how to allocate students and teachers to classrooms at the beginning of the school year, but the Principal decides to provide tutoring for the students in the middle of the school year. This decision shouldn't impact peer quality.

in private schools. The maximum value of student quality obtained for public schools is lower than the maximum obtained for the full sample. Figure 12 plots the average production function for public schools against peer quality and student quality. Graph 12 is very different from Graph 6, which was obtained with the full sample. In the sample of public schools, test scores increase in student quality for students with quality above 0.6, but decrease in student quality for other students. In addition, tests scores increase with peer quality for groups with quality below 0, but decrease with peer quality for high quality groups. These results may be related to the fact that public schools concentrated the bad students, and we are probably underestimating the effect of student quality on test scores. Figure 13 shows the peer effects obtained using the sample of public schools. Peer effects are positive for students with quality greater than -0.4, and increase with peer quality. Although, the effect of peer quality on test scores is not significant for high quality students.

In summary, we have evidence that peers effects are positive for students in the last year of elementary school in Brazil, and that student test scores increase with student quality for average and high quality students. These results hold under different robustness checks.

## 7 Final Remarks

In this paper, we propose a semiparametric methodology to estimate peer effects in classrooms. In this semiparametric methodology, we assume that student achievement is a function of the student quality, a single index of student characteristics and peers quality, a symmetric function of this index in the group. This methodology generalizes the linear models used to estimate peer effects in different ways. First, it allows peer effects to vary with student's quality, allowing students at the upper tail of the distribution of quality to respond differently from students at the bottom of the distribution. Second it controls for "membership endogeneity" using a control function approach

The semiparametric methodology is applied to estimate the education production function and peer effects for students in the last year of elementary school in Brazil, using the information in the Brazilian National Evaluation System of Basic Education (SAEB) in 2003. Using the way students were allocated to classrooms in each school as the vector of instruments, we find that student test scores are a monotonic increasing function of student quality. In addition, we find evidence that peer effects are positive for students with quality greater than -0.4, and the students with an average quality have a higher marginal

benefit from peer quality than the student with low quality.

Although this semiparametric methodology provides a flexible way to estimate peer effects, it has some shortcomings. One of the drawbacks of this methodology is that our estimator is not root-N consistent, and converges at a slower rate than the standard series estimators used by Newey, Powell and Vella (1999). The rank estimator used in the first step of the estimation procedure converges at a slower rate than root-N, which impacts the rate of convergence of the estimators of the average structural function and peer effects. If we impose conditions on the structural function  $H(\cdot)$  that are stronger than just the monotonicity on  $Q_i$ , we may improve the rates of convergence in the first step. There is a trade off between rates of convergence and flexibility of  $H(\cdot)$ . This new methodology focuses on providing a semiparametric methodology assuming a very flexible  $H(\cdot)$ .

Another drawback of this methodology is that we are assuming additivity of the unobservable components. As pointed out by many authors (Newey and Imbens (2006), Altonji and Manski (2005), etc.), this assumption is restrictive, and it does not allow interactions between the observable and unobservable components in the model. These interactions between unobservables and observables are important to motivate, from the economic perspective, the endogeneity presented in a model with peer effects. There is a growing literature focused on semiparametric and nonparametric identification and estimation in settings in which the disturbances are nonadditivity components of the model. For example, Altonji and Manski (2005) consider a panel data model with nonadditive disturbances, Chester (2002) looks at local identification, Imbens and Newey (2002) focus on a control function approach. The semiparametric model proposed in this chapter fits into the framework of Imbens and Newey (2002). Identification of the average structural function and peer effects can be achieved replacing the conditional expectation assumptions with independence assumptions and the stationarity condition of the individual heteroskedasticity by independence.

Another possible extension of the semiparametric methodology proposed in this paper is to use a more flexible and interesting definition of peer quality. The idea is to define peer quality as an exchangeable function of the single indexes that defined student's quality in the group. In this case, peer effect can be defined as moments of the distribution of student's quality. This approach will provide insights about how the inequality of types inside a classroom may affect peer effects, which is important in determining which is the optimal way to allocate students to classrooms.

An ongoing work is a set of Monte Carlo simulations that provide insights about the

small sample properties of our estimator and how sensitive the estimator is to the choice of the smooth parameters.

## References

- Abrevaya, Jason (2000). "Rank estimation of a generalized fixed-effects regression model", *Journal of Econometrics*, 95:1-23.
- Albernaz, Ângela, Francisco Ferreira and Creso Franco (2002). "Qualidade e Equidade na Educação Brasileira", mimeo, PUC-Rio.
- Arnott, Richard and John Rowse (1987). "Peer Group Effects and Educational Attainment", *Journal of Public Economics*, 32:287-305.
- Brock, William A. and Steven N. Durlauf (2001). "Interaction-based Models", *Handbook of Econometrics* 5: 3297-3380. Amsterdam: North-Holland.
- Blundell, Richard and James L. Powell (2001). "Endogeneity in Nonparametric and Semiparametric Regression Models", *CEMMAP Working Paper*, CWP09/01.
- Cavanagh, Christopher and Robert Sherman (1998). "Rank estimators for monotonic index models", *Journal of Econometrics*, 84:351-381.
- Charlier, E, B. Melenberg and A. H. O. van Soest (1995). "A smoothed maximum score estimator for the binary choice panel data model with an application to the labour force participation", *Statistica Neerlandica*, 49(3): 324-342.
- Cooley, Jane (2006). "Desegregation and the achievement gap: do diverse peer help?", *mimeo*.
- Evans, Willian, Wallace Oats and Robert Schwad (1992). "Measuring Peer Group Effects: A Study of Teenage Behavior", *Journal of Political Economy*, 100(5): 966-991.
- Franco, Creso, Isabela Ortigano et al. "Qualidade e Equidade em Educação: Reconsiderando o Significado dos Fatores Intra-escolares", mimeo.
- Graham, Bryan and Jinyong Hahn (2003). "Identification and estimation of the linear in means model of social interactions", *Economics of Letters* 88 (1): 1 - 6
- Graham, Bryan (2005). "Identifying Social Interactions through Excess Variance Contrasts". *mimeo*.

- Graham, Bryan; Guido W. Imbens and Geert Ridder (2006a). "Complementary and Aggregate Implications of Assortive Matching: A Nonparametric Analysis", *mimeo*.
- Graham, Bryan; Guido W. Imbens and Geert Ridder (2006b). "Measuring the average outcome and inequality effects of segregation in the presence of social spillovers", *mimeo*.
- Graham, Bryan; Guido W. Imbens and Geert Ridder (2006c). "Complementarity and Optimal Allocation of Inputs ", *mimeo*.
- Han, Aaron K. (1987). "Non-parametric analysis of a generalized regression model: the maximum rank correlation estimator", *Journal of Econometrics*, 35:303-316.
- Hausman, Jerry and William Taylor (1981). "Panel data and unobservable individuals effects", *Econometrica*, 49(6): 1377-1398.
- Horowitz, Joel (1992), "A Smoothed Maximum Score Estimator for Binary Response Model", *Econometrica* 60(3): 505-531.
- Hoxby, Caroline (2000). "Peer Effects in the Classroom: Learning from Gender and Race Variation ", *mimeo*.
- Imbens, Guido and Whitney Newey (2002). "Identification and Estimation of the Triangular Simultaneous Equations Models Without Additivity ", *NBER Technical Working Paper* 285.
- Lazear, Edward (2001). "Education Production ", *Quarterly Journal of Economics*, 96(3):777-802.
- Manski, Charles F. (1975). "Maximum Score estimation of the stochastic utility model of choice", *Journal of Econometrics*, 3:205-228.
- Manski, Charles F. (1993). "Identification of Endogenous Social Effects: The Reflection Problem", *The Review of Economic Studies*, 60(1993):531-542.
- Newey, Whitney K. and Daniel McFadden (1994). "Large sample estimation and hypothesis testing", *Handbook of Econometrics*, volume IV.
- Newey, Whitney K., James Powell and Francis Vella (1999). "Nonparametric Estimation of Triangular Simultaneous Equations Models ", *Econometrica*, 67(3): 565-603.

Rau, Tomas (2006). "Semiparametric Estimation of microeconomic models with endogenous regressor and sorting ", *mimeo*.

Sacerdote, Bruce (2001). "Peer Effects with Random Assignment Results for Dartmouth Roommates", *The Quarterly Journal of Economics*, May 2001: 681-704.

Sherman, Robert (1993). "The limiting distribution of the maximum rank correlation estimator", *Econometrica*, 61(1): 123-137.

Soares, José Franciso (2004). "Qualidade e Equidade na Educação Básica Brasileira: A Evidência do SAEB 2001", *Archivos Análíticos de Políticas Educativas*, 12(38).



# Appendix

## A Socioeconomic Index

The aggregate socioeconomic index used in the empirical analysis is created by a Principal Component analysis. We assume that there is a linear model that relates the items of a questionnaire with a number of latent factors,

$$X - \mu = RF + \varrho$$

where  $X$  is the item in the questionnaire,  $\mu$  is mean of  $X$ ,  $R$  is the matrix of weights and  $F$  is the factor. In this case,  $F$  are orthogonal variables that are independent of  $\varrho$ , which implies that  $\mathbb{E}[F] = 0$  and  $\text{Cov}[F] = I$ . In addition, we define  $\mathbb{E}[\varrho] = 0$  and  $\text{Cov}[\varrho] = \Xi$ .

Since  $F$  and  $\varrho$  are independent,

$$\Sigma = \text{Cov}[X] = RR' + \Xi$$

Using the principal components analysis, we decompose  $\Sigma \simeq U\Lambda U^T + \Xi$ , where  $\Lambda$  is a diagonal matrix with eigenvalues, and  $U$  is the matrix with eigenvectors.

The factors are extracted by the principal component analysis through the following steps:

1. Get the matrix  $X$
2. Subtract the mean of each column of  $X$
3. Calculate the covariance matrix of  $X$
4. Calculate the eigenvector and eigenvalues of the covariance matrix
5. Choosing the highest eigenvector that is associated with the highest eigenvalue. This eigenvector is the basis of the principal component. The principal component is the square root of the eigenvalue times the eigenvector.

To construct this single index, we use 12 items of the questionnaire: number of TVs in the household (0, 1, 2, 3, 4 or more), number of radios in the household (0, 1, 2, 3, 4 or more), a dummy variable that is equal to 1 if the household has VHS or DVD, if the household has freezer (0 for none, 1 for freezer without fridge, 2 for freezer with fridge)<sup>27</sup>, a dummy variable that is equal to 1 if the household has washing machine, a dummy that is equal to 1 if the household has vacuum cleaner, number of cars in the household (0, 1, 2 or more), if the household has a computer (0 for none, 1 for computer without internet, 2 for computer with internet), if the household has a maid (0 for none, 1 for a maid that goes to the house less than once a week, 2 for a maid that goes to the house more than once a week), number of bathrooms in the household (0, 1, 2, 3 or more), number of bedrooms per person in the household (0 for >3, 1 for (2,3], 2 for [2,1), 3 for <=1).

---

<sup>27</sup>The construct of this variables uses 2 items of the questionnaire.

Table 1: Summary Statistics

<b>Variable</b>	<b>Maximum</b>	<b>Minimum</b>	<b>Mean</b>	<b>Std</b>
PROFIC(Y)	369.98	72.19	193.12	46.46
<b>Student Characteristics (X)</b>				
FEMALE	1	0	0.49	0.50
AGE	15	8	10.59	1.19
WHITE	1	0	0.42	0.49
Socioeconomic INDEX (NSE)	9.01	-6.49	0.43	3.41
PARENT EDUCATION	9	1	6.04	2.55
No education	1	0	0.02	0.15
Elementary School (Incomplete)	1	0	0.07	0.25
Elementary School	1	0	0.16	0.36
Middle School (Incomplete)	1	0	0.08	0.26
Middle School	1	0	0.12	0.33
High School (Incomplete)	1	0	0.04	0.19
High School	1	0	0.16	0.37
College (Incomplete)	1	0	0.07	0.26
College	1	0	0.28	0.45
PRESCHOOL	1	0	0.83	0.38
<b>School Characteristics (W)</b>				
URBAN	1	0	0.96	0.20
PRIVATE	1	0	0.29	0.46
NUMBER OF STUDENTS IN THE CLASSROOM	56	2	29.13	8.28

NOTE: In this table, the mean of the characteristics is taken over all students in all schools. Parent education and Region include the categories listed below the variables.

Table 2: First Stage

VARIABLES	SEMIPARAMETRIC CASE			PARAMETRIC CASE
	$0.5 \cdot \sigma_N$	$0.75 \cdot \sigma_N$	$\sigma_N$	
Female	-6.191 (16.9997)	-11.3735 (1.6536)**	-2.821 (1.1539)**	-5.9969 (0.4949)**
Age	-8.5294 (24.7368)	-10.0904 (1.1529)**	-3.4956 (1.0971)**	-5.0402 (0.2605)**
White	2.7799 (7.5403)	2.2631 (1.2278)	0.9427 (0.3946)**	1.8234 (0.5432)**
NSE	1	1	1	0.1196 (0.1145)
Parent schooling	2.7799 (4.163)	1.2152 (0.1813)**	0.6296 (0.2462)**	0.8979 (0.1263)**
Preschool	8.8737 (24.2689)	13.4491 (1.3842)**	4.1595 (1.8638)**	9.4393 (0.7204)**
Bandwidth	0.2476	0.3714	0.4134	-
N. Observations	103,667	103,667	103,667	20,137

NOTE: To obtain these maximum score estimators, we use the annealing algorithm, using the values for the initial parameters proposed by Coronal et al (1987). In the semiparametric case, we use a kernel of order 4 and the bandwidth was selected by Horowitz's plug-in method ( $\sigma_N$ ). The values in the parentheses represent the standard errors using the formula for the asymptotic variance of the smoothed estimator in the nonparametric case and the GMM asymptotic variance in the parametric case. In this table, \*\* means significant at 5%, \* means significant at 10%.

Table 3: Student's Quality and Peer's Quality

Variable	Student Quality ( $\hat{q}_{ig}$ )			
	$0.5\sigma_N$	$0.75\sigma_N$	$\sigma_N$	PARA METRIC
<b>Mean</b>	-8.04	-22.11	-5.91	-7.06
<b>Median</b>	-5.54	-19.17	-5.29	-5.18
<b>Std. Dev.</b>	17.67	17.90	7.75	9.46
<b>Min</b>	-77.67	-96.91	-36.18	-46.14
<b>Max</b>	36.69	25.11	15.82	15.33
<b>N. Obs</b>	20137			

Variable	Peer quality ( $\hat{q}_g$ )			
	$0.5\sigma_N$	$0.75\sigma_N$	$\sigma_N$	PARA METRIC
<b>Mean</b>	-8.25	-22.40	-6.01	-7.06
<b>Median</b>	-7.54	-20.82	-6.04	-6.25
<b>Std. Dev.</b>	14.68	14.31	6.63	6.57
<b>Min</b>	-77.39	-96.62	-35.90	-46.11
<b>Max</b>	25.52	12.37	9.68	9.97
<b>N. Obs</b>	2687			

NOTE: In the estimation of the average structural function and peer effects, we use the estimated values of quality obtained using the optimal bandwidth ( $\sigma_N$ )

Table 4: Parametric Case

	<b>Second Step</b>	<b>Third Step</b>
<b>DEPENDENT VARIABLE</b>	<b>peers' quality</b>	<b>Classroom's Test Score</b>
<b>ALLOCATION RULE</b>		
Integration by age	0.5300 (0.0633)**	—
Integration by score	0.3113 (0.0832)**	—
Segregation by Score	0.4607 (0.0734)**	—
No criterion	0.4415 (0.0688)**	—
PEER QUALITY	—	4.0460 (14.281)
$v_g$	—	8.2209 (12.6073)
N. Observations	2687	2687
R-squared	0.4393	0.4168

NOTES: In this regression we include a vector of controls that includes: school characteristics (region, if rural or urban, if private or public, number of students per classroom, if the classroom has air circulation and light, if the school has access to teaching material), teacher characteristics (sex, age, race, experience, education, if the teacher assigns and grades homework), Principal characteristics (sex, age, race, experience, education). We also include the admission rule used by the school and the way teachers were allocated to classroom as a vector of instruments. In the second step, the exclusion category among the student allocation rules is integration by age. We calculate the standard deviation using the GMM approach explained in a footnote of the chapter 3. In this table, \*\* means significant at 5%, \* means significant at 10%.

Table 5: Allocation Rule vs School's decision

Variables	Mean of the Variables					Test
	None	Age-Int	Score-Int	Age-Seg	Score-Seg	P-value
SCHOOL						
<b>Teacher's Training</b>	0.58	0.67	0.65	0.68	0.64	0.02
<b>Pedagogical Project</b>	0.89	0.89	0.94	0.93	0.96	0.00
<b>Design of Pedagogical Project</b>						
Suggest by Education Institute	0.09	0.11	0.11	0.13	0.09	0.46
Made by Principal	0.03	0.01	0.03	0.01	0.03	0.01
Made by Principal, review by teachers	0.12	0.11	0.16	0.12	0.09	0.25
Suggest by the teachers	0.07	0.09	0.10	0.10	0.18	0.00
Made by Principal and teachers	0.56	0.56	0.52	0.54	0.56	0.89
<b>Admission Rule</b>						
None	0.61	0.36	0.43	0.42	0.28	0.00
Exam	0.02	0.05	0.06	0.03	0.03	0.02
Lottery	0.00	0.01	0.00	0.00	0.02	0.00
Neighborhood	0.13	0.15	0.17	0.18	0.27	0.00
First come	0.12	0.29	0.19	0.27	0.21	0.00
Other	0.11	0.14	0.15	0.10	0.19	0.03
<b>Teacher's Allocation Rule</b>						
None	0.24	0.08	0.07	0.07	0.07	0.00
Teacher's preference	0.20	0.20	0.21	0.17	0.28	0.05
Good teachers + Good Students	0.02	0.05	0.09	0.07	0.03	0.00
Good teachers + Bad Students	0.05	0.10	0.12	0.10	0.09	0.00
Same teacher in the same class	0.15	0.17	0.19	0.17	0.12	0.11
Teachers switch grades	0.17	0.23	0.17	0.27	0.16	0.00
Lottery	0.01	0.01	0.00	0.02	0.01	0.00
Other	0.17	0.16	0.16	0.13	0.25	0.02
<b>Program to</b>						
Reduce drop out	0.25	0.35	0.28	0.41	0.32	0.00
Reduce grade repetition	0.46	0.64	0.55	0.64	0.64	0.00
Help learning	0.81	0.86	0.80	0.80	0.89	0.00
<b>Choose of books</b>						
Teachers	0.45	0.51	0.40	0.52	0.55	0.00
Principal, but ask teacher's opinion	0.13	0.12	0.14	0.11	0.10	0.41
Principal+Pedagogical coordinator	0.01	0.01	0.00	0.01	0.00	0.00
Principal	0.01	0.00	0.01	0.01	0.01	0.75
Outside Instution	0.02	0.02	0.01	0.02	0.03	0.17
Don't know	0.07	0.07	0.08	0.07	0.09	0.72
No books	0.43	0.59	0.59	0.58	0.31	0.05

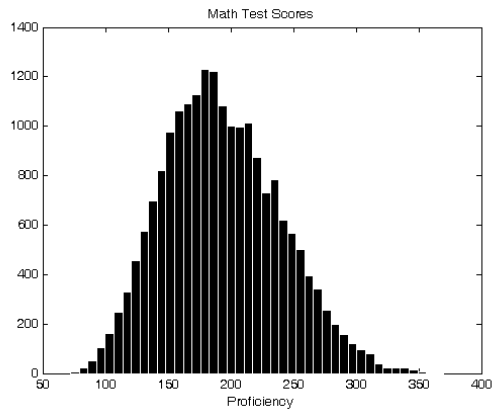


Figure 1: Test Scores

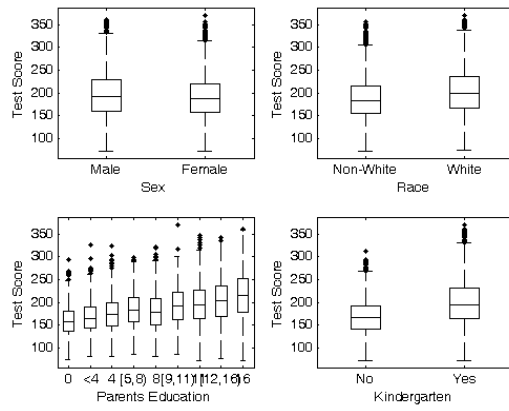


Figure 2: Test Scores vs Students characteristics

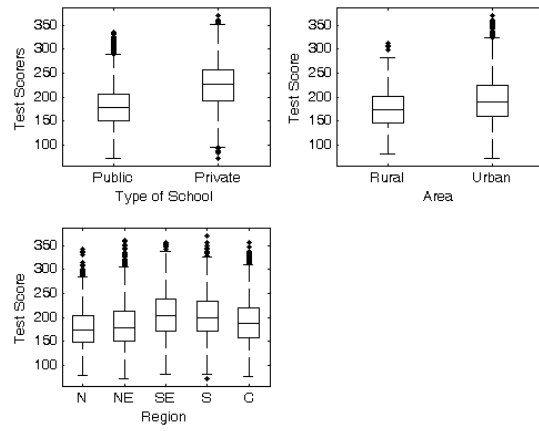


Figure 3: Test Scores vs School Characteristics

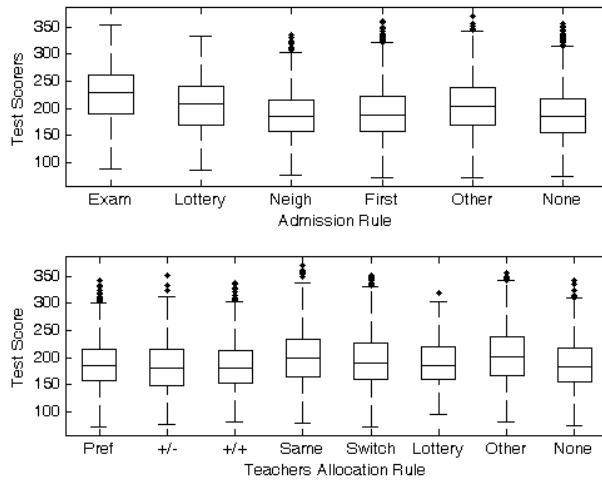


Figure 4: Test Scores vs Allocation Rules



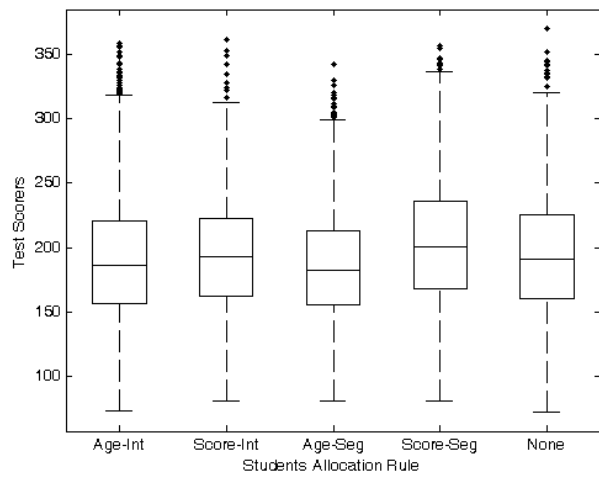


Figure 5: Test Scores vs Students Allocation

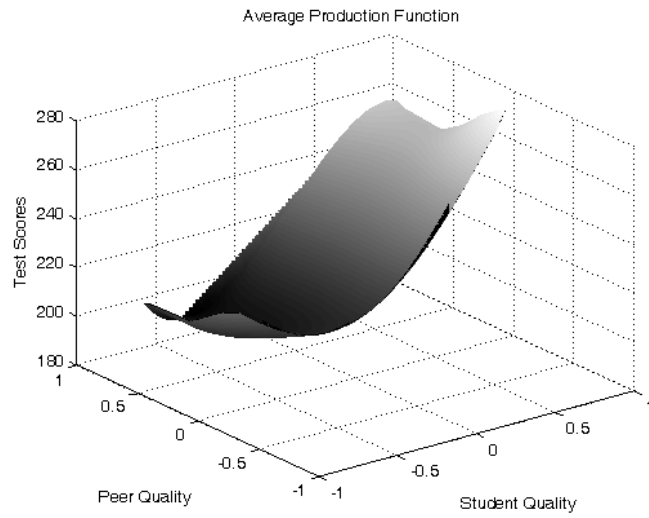


Figure 6: Average Production Function

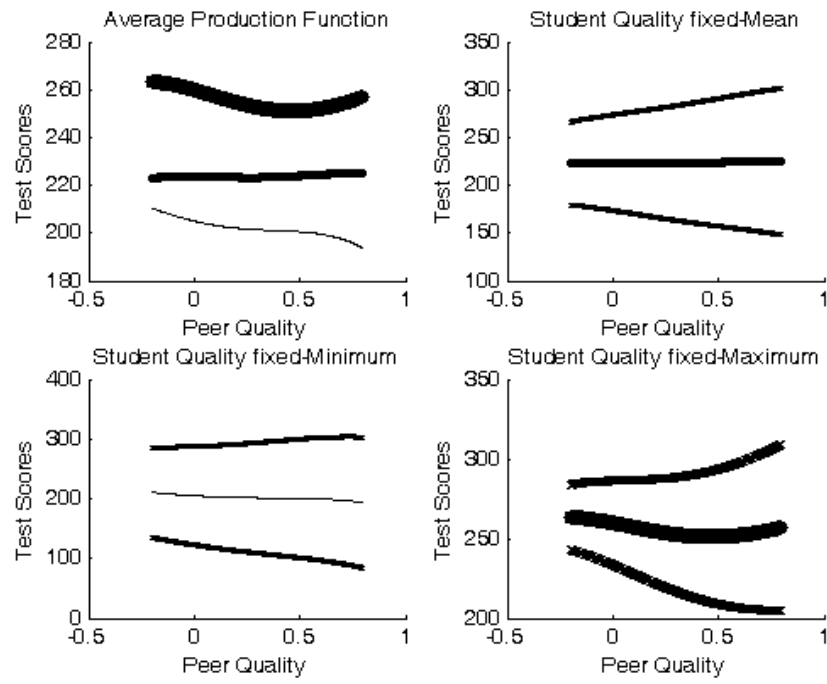


Figure 7: Average Production Function vs Peers Quality

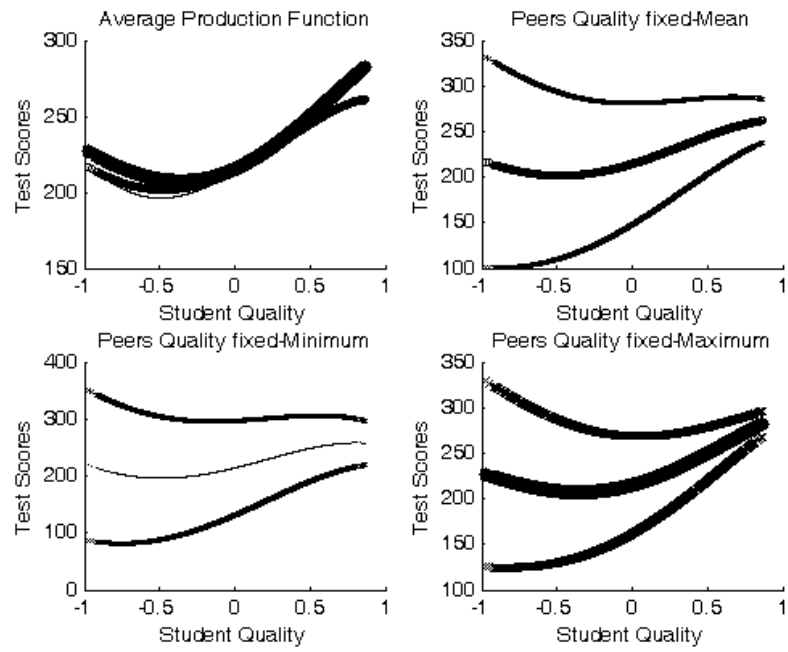


Figure 8: Average Production Function vs Student Quality

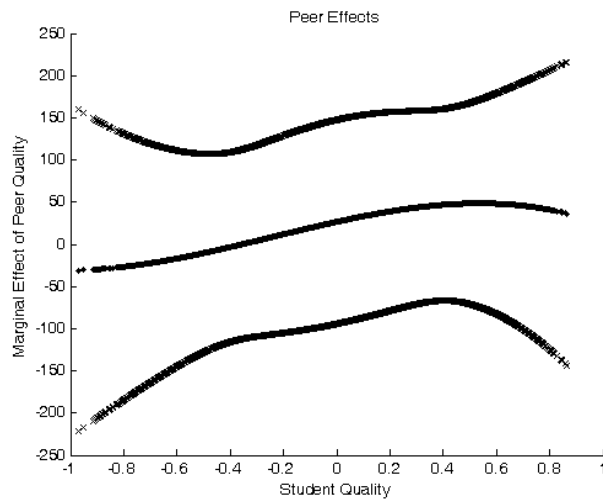


Figure 9: Peer Effects

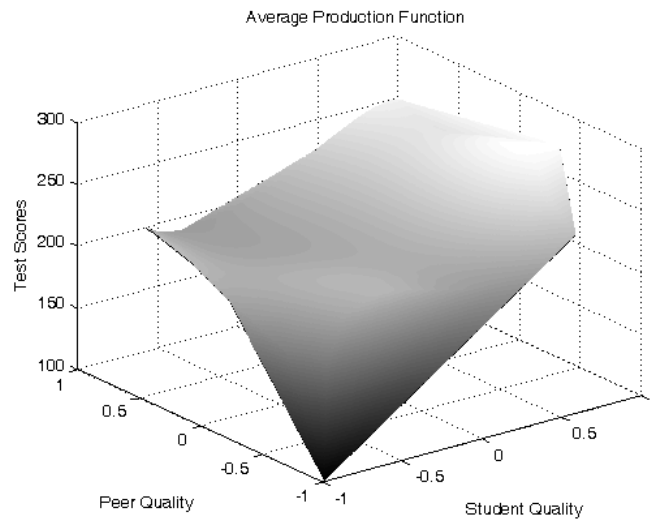


Figure 10: Average Production Function (with more controls)

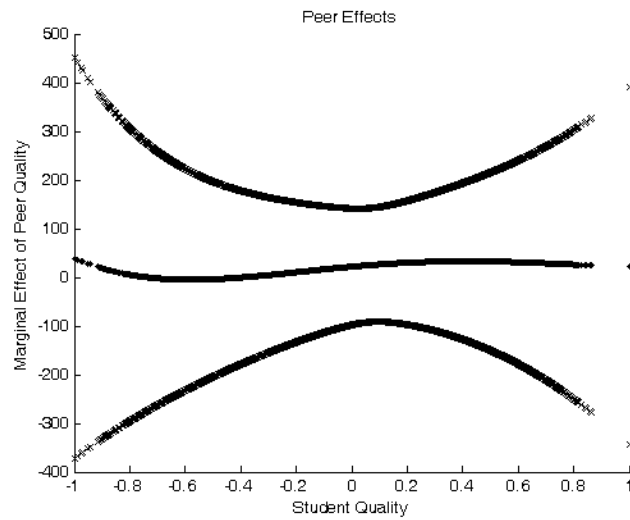


Figure 11: Peer Effects (with more controls)

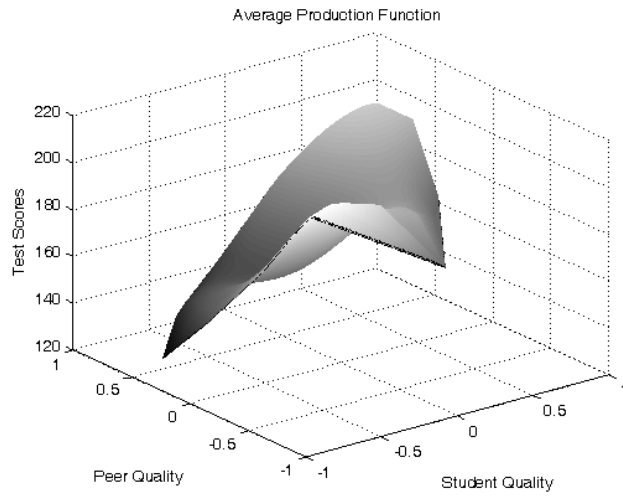


Figure 12: Average Production Function (only Public Schools)

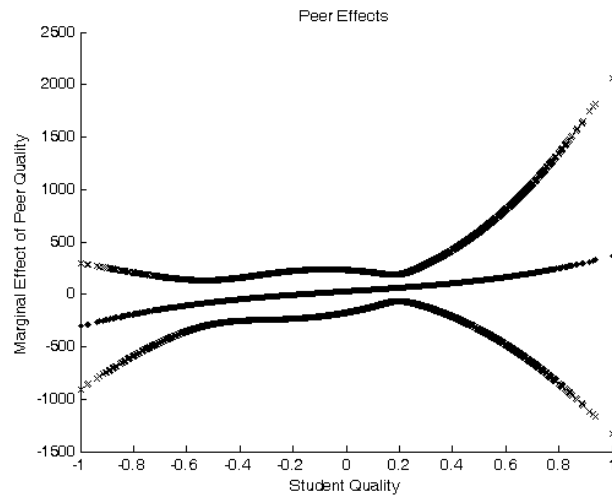


Figure 13: Peer Effects (only Public Schools)