Neighborhood E¤ects on Performances at School.¤

Dominique Goux (INSEE)
and Eric Maurin (CREST and CEPR).

October 28, 2003

Abstract: To repeat a grade in elementary school or junior-high school is the most direct indicator of early performances at school in France. At the end of junior high-school, about 40% of French adolescents have been held back a grade and our paper provides estimates of neighborhood e¤ects on this indicator of performances at school. We develop three di¤erent strategies in order to address the endogeneity issues. The …rst approach uses the available information on the socioeconomic status of families who have just moved into a neighborhood. French families are bound to send their children to schools located in the administrative district where they live. Information on the families who move into a neighborhood are information on the quality of the neighborhood and may be used to control for the e¤ect of omitted neighborhood attributes. Our second strategy builds on the fact that - in France- the distribution of families across public-sector housing is under the control of administrative authorities and can be shown exogenous. The true neighborhood e¤ects can plausibly be evaluated by focusing on the sample of public-sector neighborhoods. Our last approach uses instrumental variables. In France, one of the determinant of having repeated a grade is the date-of-birth within the year. Children born at the beginning of the year are more mature and less likely to repeat a grade than children born near the end of the year. Within this context, the distribution of children's date-of-birth in the neigborhood is a plausible instrument for identifying the e¤ect of the proportion of children who have repeated a grade in the neighborhood. Generally speaking, our three strategies provide consistent …ndings. The estimated neighborhood e¤ects are signi…cant and almost as large as the direct e¤ect of parental education.

¤Previous versions of this paper were presented at the Conference in memory of Zvi Griliches (Paris, 2003) and at the CEPR Conference on Changing Condition in Education (Uppsala, 2003) and at the CREST (Paris, 2003). This reasearch has bene…ted from a grant of the French Commissariat General au Plan (n± 7-2002). We thank Per Johansson and Gerard Van den Berg for their comments on earlier versions of this paper.

1 Introduction

The assumption that children's outcomes are in‡uenced by the neighborhood in which they grow up forms the basis of a large and still growing literature in social sciences. Neighborhood e¤ects correspond to externalities which are plausibly not internalized by families and, as such, represent a key issue for public policies. If childrens' performances are a¤ected by the performances of the other children in the neighborhood, then public policies targeted at disadvantaged neighborhoods may increase jointly the overall level of performances at school and the degree of equality of opportunities across children from di¤erent social background.

The statistical evidence showing the causal e¤ect of neighborhoods on individual outcomes remain nevertheless very weak and the issue is still very controversial (Ginther et al, 2000). The lack of consensus is plausibly due to the fact that beliefs about the role of social contexts correspond to deep ideological orientations. Judgements about the role of social contexts are judgements about the actual importance of individual responsability and, as such, represent one key ingredient of social justice evaluation. Another important reason for persisting controversies is that no consensus exists on how to identify and estimate the causal e¤ect of social contexts.

When empirically evaluating neighborhood e¤ects, social scientists typically construct average measures of neighborhoods' characteristics and add these measures as supplementary regressors into models which already include individual-level regressors. The speci…cation of the neighborhood-level and individual-level variables varies widely across studies and the estimated neighborhood e¤ects are far from consistent among the existing studies (see Ginther et al., 2000). One e¤ect appears rather consistently, however, which is the e¤ect of schooling outcomes in the neighborhood on the outcome of the child. Using the terminology introduced by Manski (1993), the most signi…cant neighborhood e¤ect seems the endogenous one.

Glaeser and Scheinckman (2001) provide an interesting description of the physical, learning, stigma or taste-related mechanisms that may generate such positive social interactions. Regarding behaviors at school, there exist some obvious channels through which children's behaviors may be in‡uenced by the behaviors of the other children in the neighborhood. Working hard in school might be less di¢cult for a children if his friends in the neighborhood are also working hard, both because they can help him learn and because they are not available for other leisure activities. Also, it is plausible that children and adolescents just enjoy behave as their peers. The stigma associated with working hard in school is plausibly less important when everybody is working hard, and vice versa.

From an empirical viewpoint, most existing studies identify the neighborhood e¤ects through assuming that the average characteristics of neighbors are exogenous. There are several problems with this methodology, especially when the focus is on estimating endogenous neighborhood e¤ects, i.e. the in‡uence of neighbors'outcomes on individual outcome (see Manski, 1993).

The …rst issue is that neighborhood membership may be endogenous. Parents who chose (or are constrained) to live in the same neighborhood are likely to have the same values and behaviors. A second issue is that there may be omitted variables characterizing each particular neighborhood (typically the quality of schools) which a¤ect simulatenously the performances of all the children in the neighborhood. Given these facts, children in the same neighborhood may perform similarly because their parents share similar unobserved characteristics or because they bene…t from similar schools and teachers, not because they in‡uence each other.

A third crucial issue is that if a child's performances are in‡uenced by the performances of the other children in the neighborhood, his/her performances also a¤ect the performances of his/her neighbors. In such a case, a child's performances and the performances of his/her neighbors are determined simultaneously and standard estimators are a¤ected by simultaneity biases.

This paper proposes new estimates of neighborhood e¤ects on performances at school using new French data and identifying strategies to overcome the endogeneity issues. To repeat a grade in elementary school or junior-high school is the most direct indicator of early performances at school in France. At the end of junior high-school, about 45% of French adolescents aged 15 have been held back a grade[1] and our paper provides an analysis of this indicator of performances at school. The data come from the 12 French Labor Force Surveys (LFS) conducted each year between 1991 and 2002. The very interesting feature of the French LFS is that the basic sampling units consist of groups of households (aires). A typical LFS consists of a representative sample of about 3,500 aires. Within each aire, all the households are surveyed and, within each household, all the persons aged 15 or more are surveyed. All in all, a typical LFS is large representative sample of persons aged 15 of more who can be grouped in small neighborhoods of about 20/30 households (i.e., 40/60 individuals). For each family surveyed, we know whether the housing belong to the public sector and/or whether the family just moved into it. For each adolescent aged 15, we have detailed information on his/her individual situation and on the situation of the persons living in his/her neighborhood. Given these data, the basic purpose of the paper is to identify the e¤ect of the aires' attributes (proportion of non-french or non-educated persons, proportion of children who are behind at school) on adolescents'probability of being behind at school.

1.1 Identi…cation strategies

Our …rst iden…cation strategy relies on the information available on families who have just moved into a neighborhood. This information makes possible to test whether (and identify why) neighborhood membership is endogenous. Also it provides a measurement for the unobserved quality of the neighborhoods.

To begin with, our data show that there exist a very signi…cant correlation between the probability of being late at school which characterizes children

[1] In France, shooling is compulsory until age of 16. It is the reason why we focus on 15 years' old.

who move into a neighborhood and the probability of being late which characterizes the children who already live in the neighborhood. It con…rms that the neighborhood membership is endogenous, meaning that families who chose (or are constrained) to live together are similar with respect to some important determinant of performances at school. Most interestingly, we …nd that this correlation disappears when we control for parental education and nationality. As it turns out, most of the endogeneous membership issue comes from the fact that families with similar education and nationality tend to live together and that these family characteristics are determinants of performances at school. In our case, neutralizing the biases that may arise from the endogenous neighborhood membership amounts to controlling for parental education and nationality.

The available information on families who have just moved into a neighborhood makes possible to reduce the bias generated by endogenous membership. Also, it provides a mean to reduce the e¤ect of the omitted neighborhood attributes. Speci…cally, the socio-economic status of the families who move into a neighborhood provides a direct information on the quality of the neighborhood and may be used as control variable to better identify the true contextual e¤ects.

This assumption is based on the fact that - in France - the administration de…nes a local district for each public school and families are bound to send their children to schools located in the district where they live. Our data con…rms that there is a competition for the best neighborhood and that only high-income families can a¤ord high-quality neighborhood, meaning neighborhoods where the probability of being late at school is low. Hence, the variables describing the permanent income of the families who move into a neighborhood (i.e., the parents' education and occupation) provide a plausible measurement for the characterictics of the neighborhood which are unobserved in the Labor Force Surveys, but which belong to the set of determinants of performances at school. These variables can be used as control function to reduce the e¤ects of the unobserved quality of neighborhoods.

All in all, our …rst strategy consists in adding two sets of control variables to the baseline regression of children's outcomes on neighborhood attributes, speci…cally (a) control variables describing the education and nationality of the parents, which plausibly reduce the e¤ect of endogenous neigborhood membership and (b) control variables describing the average educational and occupational level of the families who have just moved into the neighborhood, which plausibly reduce the e¤ect of omitted neighborhood attributes. Within this framework, neighborhood e¤ects are identi…ed through comparing families with similar education and nationality, living in neighborhoods similar with respect to the socioeconomic level of families who move into them, but di¤erent with respect to the actual social composition of their neighborhood (and in particular with respect to the rate of children who are late at school).

Generally speaking, this …rst strategy provides estimates of neighborhood e¤ects which are signi…cantly di¤erent from zero and relatively large. For instance, living in an aire where the parents are high-school dropouts increases ceteris paribus the log odds of being late at school by about 0.8 (i.e. multiplies

the odds of being late by about two and a half).

Our second identifying strategy rely on a comparison between families which housing belong to the public sector (HLM, 20% of the population) and families living in the private sector. The idea is that families who live in a public housing are less sorted across neighborhoods than families living in the private sector and that neighborhood e¤ects can be directly identi…ed through focusing on public sector families[2].

In France, any family is eligible for a public sector housing (HLM) provided that it is actually allowed to live in France and that its income per unit of consumption is su¢ciently low. Public housing are managed by several di¤erent types of administrative authorities and -in general- eligible families apply simultaneously to the di¤erent possible channels. The number of eligible families is about three times as large as the number of available public housing and the waiting lists are very long. Families have often to wait for several years before a decision is made. The key point is that the decision to allocate a given housing to a given family is made by the administration, meaning families have no direct control on the speci…c neighborhood to which they are allocated. We have checked that there exist no signi…cant correlation between the probability of being late at school which characterizes children who move into a public housing and the probability of being late which characterizes the children who already live in the corresponding neighborhood[3]. It con…rms that families are not really sorted across public-sector neighborhood according to the characteristics which a¤ect schooling outcomes and that the neighborhood membership can plausibly be considered as exogenous in the public sector. Most interestingly, the estimates of neighborhood e¤ects obtained on the subsample of public housing do not depend on whether we control for individual characteristics and are very close to the estimates obtained through our …rst identifying approach.

Our two …rst identi…cation strategies are designed to reduce the endogenous membership issue as well as the omitted attributes problem. As discussed below, the main problem with these approaches is that they do not make it possible to adress the simultaneity issue and to separate the endogenous from the exogenous neighborhood e¤ects.

Our third identifying strategy tries to adress this issue by developing an instrumental variable approach. Speci…cally, our third approach relies on the variations across neighborhoods in the distribution of childrens' date-of-birth within the year[4]. In France, the date at which a child has to begin primary school is determined by his/her year of birth. French children who were born

[2] This approach is close in spirit from Edin et al. (2003). They use the fact that Swedish government authorities distribute refugee immigrants across locales in a way that may be assumed exogenous to identify the e¤ect of living in ethnic enclaves on labour market outcomes.

[3] In practice, one cannot exclude that public managers actually sort eligible families across neighborhoods according to some of the characteristics that they observe. As it turns out, these characteristics do not seem to be correlated with schooling outcomes.

[4] Angrist and Krueger (1991) have already used individual season-of-birth to identify the impact of compulsory schooling on earnings. The idea is that individuals born at the beginning of the year start school ar an older age and are therefore allowed to drop out after completing less schooling than individuals born near the end of the year.

the same year (and who are normal age) are in the same class. Within this context, the date-of-birth within the year determines the relative level of maturity within the class and, as such, is a direct determinant of the probability of being held back a grade. As a matter of fact, we have checked that the proportions of adolescents in a neighborhood born in January, February,..., December are direct determinants of the proportion of adolescents who are held back a grade : the larger the share of adolescents born at the end of the year, the larger the proportion of adolescents in the neighborhood who have been held back a grade. The identifying assumption is that it is the only channel through which the distribution of dates of birth in a neighborhood actually a¤ects performances at school in the neighborhood. Under this assumption, the distribution of date-of-birth of the other adolescents in the neighborhood can be used as an instrumental variable to identify the endogenous neighborhood e¤ect.

This last strategy provide estimates of neighborhood e¤ects which are signi…cant and consistent with the two previous one. The paper is organized as follows. The next section provides an overlook of the literature, section 3 describes the data, section 4 shows the result of the approaches which build on the available information on families who have just moved into a neigborhood as well as on the available information on public sector housing. The following section shows the results of our IV strategy and the last section concludes.

2 Background

The issue of neighborhood e¤ects on children'outcomes has a long history among social scientist, including economists[5]. In their recent contribution, Ginther et al. (2000) present a summary for 17 studies conducted by economists and sociologists since 1980. These studies provide empirical evaluations of the effect of neighborhood attributes on several di¤erent outcomes, including years' of schooling, high-school completion, teenage fertility. These studies typically construct average measure of neighborhoods' characteristics (average family income, racial and demographic composition of the neighborhood, etc.) and add these measures as supplementary regressors into models which already include individual level regressors (typically family income). A discussed by Ginther et al., the econometric speci…cations and the estimated e¤ects vary a lot across studies. The most robust …nding is perhaps that the endogenous contextual e¤ects are more signi…cant than the exogenous ones. Put di¤erently, an individual's outcome seems to be more a¤ected by the average outcome in the neighborhood than by any other characteristics of the neighborhood.

Most of these studies regress children's outcomes on neigborhoods' attributes using standard OLS or (when the outcome is binary) standard probit models. This strategy typically ignores the potential endogeneity of neighborhoods' attributes. There are several problems with this methodology, as discussed by

[5] A related literature is on peer-group e¤ects in education. This literature typically explores the e¤ect of school racial composition on performance at school (see the survey by Scho…eld, 1995).

Manski (1993) or more recently by Mo¢t (2001).

First of all, if a child is a¤ected by the other children of the neighborhood, he also a¤ects his neighbors. Given this fact, his neighbors' average performances are a function of his own performances. All in all, individuals' outcome and neighbors' average outcomes are determined simultaneously and neighbors' average outcomes cannot be assumed exogenous. A related issue is that there may be omitted variables characterizing each particular neighborhood (typically the quality of schools in the neighborhood) which a¤ect simulatenously the performances of the individuals and their neighbors'performances.

These simultaneity and correlated-unobservables problems can be soved provided that there exist an individual attribute (say, v) which a¤ect individuals' outcomes, but which is such that its average level across neighbors (E(v j n); where n denotes the neighborhood)) does not a¤ect individuals' outcomes. In such a case, its average level can be used as an instrumental variable for identifying the endogenous neighborhood e¤ect (note sur les policy interventions).

In their early attempt to adress these issues, Case and Katz (1991) assume that all the available variables describing neighbors' parents were exogenous. Using the terminology of Manski (1993), it amounts assuming that the only signi…cant contextual e¤ect is the endogenous one, which is obviously debatable.

Another problem is that parents chose their neighborhood and parents who are likely to have the same values and behaviors may chose to live close to one another. This issue of endogenous neighborhood membership is also potentially treatable using exogenous neighbor's characteristics as instruments for neighbors' performances. Another solution is to provide a model for the choice of neighborhood.

In their study of the e¤ects of school characteristics on children's outcomes, Evans, Oates and Schwab (1992) assume that parents take their choice of the metropolitan area as given when chosing a school for their children, but their location within the metropolitan area as a decision variable. Within this framework, the unemployment rate or the median income characterizing the metropolitan area can be used as instrumental variables for identifying the e¤ect of the characteristics of children's schools on childrens'performances. The problem with this approach is that - once the characteristics of the metropolitan area actually a¤ect the average quality of schools - it is not clear why the "good" parents will not move towards the "good" metropolitan area.

In their analysis of ghettos, Cutler and Glaeser (1997) are faced with the problem that abler black may chose to move from highly segregated cities to less segregated ones. To adress this issue, they construct an instrumental variable using the available (census) information on where individuals were living …ve years previously. They implicitely assume that the past level of segregation has no e¤ect on the current outcome, which is again debatable.

As emphasized by Mo¢t (2001), the theoretical literature on social interactions has run considerably ahead of empirical testing[6] and the issue of …nding convincing instruments and econometric strategies for identifying neighborhood

[6] See e.g. Brock and Durlauf (2001).

e¤ects is still largely unsolved.

3 Data and Variables

The datasets used in this paper come from the 12 French Labor Force Surveys (LFS) conducted each year between 1991 and 2002. One of interesting feature of the French LFS is that the basic sampling units consist of groups of about 20 adjacent households (aires). More speci…cally, a typical LFS consists of a representative sample of about 3,500 aires. Each aire is surveyed three consecutive year and then replaced by an adjacent new aire. Each year, within each aire, all the households are surveyed and, within each household, all the persons aged 15 or more are surveyed. The French statistical o¢ce (INSEE) has chosen this sampling strategy in order to reduce the travelling expenses of the investigators who are in charge of the survey. Furthermore, it has been shown that such a sampling design tend to improve the response rates. People are more likely to accept being surveyed by an INSEE's investigators when they know that other persons in the neigborhood have already been surveyed. The main drawback of relying on a sample of neighborhood is that it tends to increase the standard deviation of estimators.

For each respondent, we have standard information on his age, sex, nationality, family situation, place of birth, education, labor market situation (unemployed, out of the labor force, employed). Also we know whether the respondent was already living in his current house one year before or whether he just moved into the neighborhood. For respondents who are still in the educational system, we know their current grade. In particular, we know the grade of each of the 15 years' old respondents which implies that we know whether s/he has been held back a grade in elementary or middle school. Speci…cally, respondents of year t born in t ¡ 15 are in at least the ninth grade if they have not repeated a grade. Our measurement for "having repeated a grade" is simply a dummy which equals 1 if they are not yet in the ninth grade.

We have extracted the subsample of respondents who are 15 years old, who were already living in the aire one year before the survey and such that there exist at least one other adolescent aged 15 observed in the aire ($N = 25;000$). Table 1 provides the distribution of the adolescents of this sample according to the number of other adolescents aged 15 observed in their aire. Table 2 provides the basic descriptive statistics for these adolescents.

We have also constructed the subsample of respondents who are 15 years old, who were already living in their current house one year before, but such that there exist at least one other 15 years old in the neighborhood who just move into the neighborhood ($N = 5;000$).

In a second step, we have constructed for each respondent several variable describing the average characteristics of the families of the other adolescents in the aire, namely the proportion of high-school dropouts, the proportion of college graduates, the proportion of non-french among the adults of these families. Using the terminology of Manski (1993), the impact of these neighborhood vari-

ables correspond to exogenous contextual e¤ects. For each respondent, using the information on all the adolescents aged 15 observed in his/her aire, we have also calculated the probability of having repeated a grade in primary or junior high-school in his/her aire. Following Manski, the impact of this variable will correspond to endogenous social e¤ects.

For about 15,000 respondents the available surveys provide us with information not only on the aire in which they currently live, but also on the adjacent aire which was previously in the survey and which has been replaced by the current one. Given this fact, we are able ot construct aire-level information not only for the aire in which the respondent lives, but also for the adjacent aire which was previously in the survey. For each respondent, this provides us with information on the lag value of E(y j n); i.e., on the probability of having repeated a grade at age 15 for individuals who belong to the same neighborhood, but who are a couple of years older than the respondent.

Let us emphasize that, for each respondent, the di¤erent aire-level indicators are constructed using only the information on the individuals who do not belong to the family of the respondent.

4 The Identi…cation of Neighborhood E¤ects Using Additional Controls

Our …rst basic empirical strategy for identifying the neighborhood e¤ects builds on the information available on families who have just moved into a neigbourhood. To begin with, this information makes possible to text whether similar families tend to live close to one antother and, as a consequence, to adress the issue of endogenous neighbourhood membership.

Table 3 focuses on children who have just moved into a neighborhood and shows the result of the regression of a dummy indicating whether they are late at school on the probability of being late at school which characterizes the neighborhood into which they move[7](column 1, Table 3). It reveals a very signi…cant correlation between the two variables. It con…rms that families who chose to (or are constrained to) live close to one another are similar with respect to some important individual determinants of performances at school.

The second column of Table 3 shows the results of the same regression after controlling for the parents' education and nationality. These variables provide a measurement for parental human capital and, as such, a measurement for one of the children's key ressource during their early schooling career[8]. Most interestingly, the estimated e¤ect of the probability of being late at school becomes much weaker than when we do not control for these family characteristics. It is not signi…cantly di¤erent from zero anymore. The estimated e¤ect becomes

[7] To keep thing as simple as possible, we use standard logistic model. We have checked that the results were very much the same when we use linear probability model or more sophisticated semi-parametric technique, such as those developped by Lewbel (2000).

[8] Even though it is not clear whether the impact of parents'cultural ressources is direct through direct transmission or whether it is through increasing parents' permanent income.

even less signi…cant when we control for the average education and nationality of families who were already in the neighborhood (column 3). Once we control for the education and nationality of both the families and the neighborhood, there is no correlation left between the unobserved determinants of the situation at school of children who move into a neighborhood and the unobserved determinants of the situation at school of the children who are already living in the neighborhood.

This basic result suggests that - in our case- most of the endogenous membership problem comes from the facts that (i) families with similar levels of education and similar nationality tend to live close to one another[9] and (ii) parents' education and nationality capture one of the key determinant of children's performances at school.

Given this fact, the next question is whether neighborhood attributes have an e¤ect on children performances at school even after controlling for the endogenous neighborhood membership e¤ect, i.e., in our case, even after controlling for parents' education and nationality. In Table 4, we focus on the basic sample of adolescents who have been living in their neighborhood for more than one year and we analyze the links between their individual situation at school and the situation of their neighbors, before and after controlling for parents' education and nationality. When we do not control for these individual factors, we …nd a very large and signi…cant endogenous neighborhood e¤ect. The estimated impact on the log odd (.94) is even larger than the corresponding impact in Table 1 (.69) .

When we do control for parents' education and nationality, the estimated endogenous neighborhood e¤ect is divided by two, but remains large and signi…cant. The impact of living in an aire where the other adolescents are late at school rather than in an aire where they are normal age is a 0.44 increase in the log odds of being late at school (i.e. a 50% increase in the odds of being late at school). An important part of the observed correlation between the performances of children and the performances of their neighbors seems to be due to endogenous neigborhood membership, but an important part is not explained by this phenomenum and consistent with the existence of signi…cant neighborhood e¤ects.

4.1 Identi…cation using Information on Families who Have Just Moved into a Neighborhood

Endogenous neighborhood membership is not the only issue that we have to adress when it comes to identify neighborhood e¤ects. The results of the regressions in Table 4 may still be biased because of omitted variables characterizing each particular neighborhood (typically the quality of schools) and a¤ecting simultaneously the performances of all the children in the neighborhood. To

[9] We have checked that the standard deviation of the distribution of non-French persons (or college graduates) across the aires of the LFS is three times smaller than the standard deviation that we would observe if non-French persons (college graduates) were randomly distributed across neighborhoods.

adress this issue, it is again possible to use the available information on the families who have just moved into a neighborhood, speci…cally the information on their permanent income. The idea is that their permanent income provides an information on the unobserved quality of the neighborhood and may be used as a control function to purge out the e¤ect of this unobserved factor (a more formal presentation is given in Appendix A).

This assumption is based on the fact that -in France- the administration de…nes a local district for each public school and families are bound to send their children in the local district where they live. In other words, chosing a place where to live amounts to chosing the quality of schools available for the children and the social composition of these schools. Also, assuming that families actually prefer settling and living in high-quality neighborhoods, there is a competition for the best neighborhoods and only high-income families can a¤ord moving in high-quality neighborhood. As a matter of fact, our data show a very signi…cant correlation between the variables measuring the permanent income of families who have just moved into a neighborhood and the probability of being late at school which characterizes the neighborhood into which they move. It con…rms that the socioeconomic status of families who have just moved into a neighborhood is an indicator of the the quality of the corresponding schooling context and may be used as a control function to reduce the e¤ect of endogenous unobserved neigborhood attributes.

Table 5 shows the results of the same regressions as in Table 4 when we add the average education, nationality and occupational status of the families who have just moved into the neighborhood as supplementary control variables. These regressions con…rms that neighborhood attributes, as such, have a significant e¤ect on children's performances at school, even after controlling for the permanent income of families who move into the neighborhood. According to model (4) in Table 5, living in a neighborhood where parents are non-educated increases the log odds of being late at school by about 0.8 (meaning a 120% increase in the odds) . Also, according to model (5), a signi…cant part of this e¤ect re- ects a purely endogenous e¤ects, meaning it is due to the fact that non-educated parents are more likely to have children late at school than the other parents.

4.2 Private versus Public Housing

Our second econometric strategy builds on a comparison between families whose housing belong to the public sector (HLM, 20% of the population) and families living in the private sector.

In France, any family is eligible for a public sector housing (HLM) provided that it is actually allowed to live in France and that its income per unit of consumption is below a threshold (about 30,000 euros for a four persons family, in 2002) which depend on the region and which is updated at the beginning of each year. Eligible families can apply for a housing in any given city (commune) where such public programs exist, regardless of their current place of living or nationality.

Public housing are managed by several di¤erent types of administrative authorities and -in general- eligible families apply simultaneously to the di¤erent possible channels. The waiting list are very long and families have often to wait for several years before a decision is made. The key point is that the decision to allocate a given family to a given housing is made by the administration and families have no direct control on the speci…c housing to which they are allocated [10]. Given this fact, the sorting of families across public-sector neighborhoods is plausibly much more exogenous than across private-sector neighborhoods[11]. To test this assumption, we have performed the same regression as in Table 1 column 1, but for families moving into a public-sector housing on the one hand (see …rst column of Table 6a), and, on the other hand, for families moving into a private-sector housing (…rst column of Table 6b). Regarding the private-sector families, the regression con…rms that there exists a very signi…cant correlation between children's probability of being late at school and the probability of being late which characterize the neighborhood into which they move. Most interestingly, this correlation does not exist for families moving into a public-sector housing. No correlation exists between the children's probability of being late at school which characterizes families moving into a public sector neighborhood and the probability of being late which characterizes the neighborhood into which they move. The public-sector population is actually not sorted across neighborhoods according to factors which a¤ect early performances at school.

Given this reality, the attributes of public-sector neighborhood represent determinants of performances at school which can plausibly be assumed exogenous (or, at least, more exogenous than private-sector attributes). Table 6a focuses on families living in the public sector and con…rms that the e¤ects of neighborhood attributes do not depend on whether we control for family characteristics (which is not the case when we focus on families in the private sector). Furthermore, the estimated neighborhood e¤ects have the same order of magnitude as in Table 5.

Our two …rst identi…cation strategies are designed to reduce the biases generated by the endogenous membership as well as by the omitted attributes. The main problem with these approaches is that they do not make it possible to adress the simultaneity issue.

Assuming that the performances of a child are actually a¤ected by the performances of the other children in the neighborhood then s/he also a¤ects his/her neighbors. Hence, by construction, the performances at school of a child and his/her neighbors are determined simultaneously and are jointly a¤ected by the same unobserved residuals. This issue cannot be adressed by simply adding supplementary regressors to some baseline regression. Our third identifying

[10] In theory, when deciding to allocate a given family to a given housing, public managers should not take into account the social composition of the corresponding neighborhood. In practice, one cannot exclude that public managers actually sort eligible families across neighborhoods according to some of the basic characteristics that they observe, such as ethnicity.

[11] In Sweden, the distribution of refugee immigrants across locales is under the control of the administration and may be considered as exogenous. In a recent contribution, Edin, Fredriksson and Aslund (2003) use this fact to evaluate the impact of living in ethnic enclaves.

strategy tries to adress this issue through developing a new instrumental variable approach.

5 An instrumental variable strategy

Generally speaking, the simultaneity problem (as well as the omitted variable and the endogenous membership problems) can be soved provided that there exist an individual attribute (say, v) which a¤ect individuals' outcomes, but which is such that its average level across neighbors (E(v j n); where n denotes the neighborhood)) does not a¤ect individuals' outcomes. In such a case, its average level E(v j n) can be used as an instrumental variable for identifying the endogenous neighborhood e¤ect.

To better understand why it is the case, let v represent this speci…c individual attribute and assume that the outcome under consideration y can be written,

$$y = \alpha E(y \mid n) + \beta v + \varepsilon \tag{1}$$

where E(y j n) is the endogenous contextual e¤ect (i.e., the probability of being held back in neighborhood n), ² represents the residuals and ® is the main parameter of interest. The identifying assumption is that v and ² are uncorrelated conditional on n, meaning that E(v j n) is uncorrelated with the other determinants of schooling performances (²): Using Manski's terminology, it amounts assuming that there exists no exogenous contextual e¤ects associated with v. After averaging conditional on n, Equation (1) yields,

$$E(y \mid n) = \frac{\beta}{1 - \alpha} E(v \mid n) + \frac{1}{1 - \alpha} E(\varepsilon \mid n);$$

Given that E(v j n) and E(² j n) are uncorrelated, it shows that E(v j n) is a valid instrumental variable for identifying the e¤ect of E(y j n) in equation (1): It is uncorrelated with ², but correlated with E(y j n):

Notice that this strategy would still be valid in the presence of additional contextual e¤ects (with the exception of E(v j n)) and other individual e¤ects. An interesting case is when the lagged value of children's average outcome belongs to the set of explanatory factors, i.e. when children have been a¤ected not only by the children of their own birth cohort, but also by the children from the previous birth cohort. Denote $E_{g-1}(y \mid n))$ the average outcome in the neighborhood for the cohort of children born before the individual's generation (i.e., cohort $g - 1$) and assume that y can be written,

$$y = \alpha_g E_g(y \mid n) + \alpha_{g-1} E_{g-1}(y \mid n)) + \beta v + \varepsilon \tag{2}$$

where v is uncorrelated with ²: Using the same argument as above, the endogenous e¤ects α_g and α_{g-1} can be identi…ed through IV regressions with $E_g(v \mid n)$ and $E_{g-1}(v \mid n)$ as instrumental variables.

Regarding the choice of v, one obvious candidate is the date-of-birth within the year. For each given cohort, children born near the end of the year are less

mature than children born near the beginning and there exists ample evidence showing that French children born near the end of the year are much more likely to be held back a grade (see, e.g. Maurin, 2002). Within this context, the identifying assumption is that the distribution of date-of-birth across the children living in the neighborhood has no direct e¤ects on children's outcome[12].

5.1 Results

Table 7 shows the results of the …rst stage regression which con…rms that the proportions of children in the neighborhood born in January,..., December are actually signi…cantly correlated with the proportion of children late at school in the neighborhood[13]. Table 8 shows that the basic IV estimated impact (.17) is signi…cant at the 6% level and not signi…cantly smaller than the impact estimated by OLS (.22). These IV estimates correspond to linear probability models. Lewbel (2000) and Magnac and Maurin (2003) specify the conditions under which such semi-parametric IV estimators are consistent[14].

Table 9 shows the results of the OLS and IV approaches when we estimate simultaneously the e¤ects of the present and past proportions of adolescents behind at school. As discussed above, the past proportion is constructed using information on an aire which is adjacent to the aire in which the respondent lives and which was previously in the sample. This variable is a proxy for the proportion of 15 years old who were behind at school in the neighborhood when the respondent was about three years younger. The instrumental variables used for identifying the two endogenous e¤ects correspond to the present and past proportions of adolescents born in January, ..., December. These regressions con…rm a signi…cant e¤ect of the present proportion of adolescents behind at school and reveals a marginally signi…cant e¤ect of the past proportion. The estimated e¤ect (.35) suggests that a one standard-deviation increase in the proportion of adolescents late at school in the neigborhood increases ceteris paribus the probability of being late by about 20% of a SD.

[12] Given that the season-of-birth a¤ects infection deseases during infancy (with a peak in fall and winter) the proportion of children in the neighborhood who were born, say, in fall or winter has plausibly a ceteris paribus e¤ect on children's risks of being contaminated during infancy. In theory, it may cause a direct correlation between our instrument and the outcome under consideration. It is worth emphasizing, however, that the e¤ect of our instrument on the probability of having been held back is monotone and that the main di¤erence in the probability of being held back a grade is between children born in December and children born in January or February, meaning between children born in winter. Put di¤erently, the correlation between our instrument and the endogenous regressor is clearly not driven by seasonal variations of infectious deseases. On the impact of season-of-birth on infectious deceases during infancy (and other outcomes), see Doblhammer (2003).

[13] Let us again emphasize that these proportions are constructed without using the date-of-birth of the respondent, but using only the information on the date of birth of the other adolescents living in his/her aire.

[14] One such condition is the existence of an explanatory factor which may be assumed exogenous and uniformly distributed. The-date-of-birth within the year is a good candidate : it a¤ects the outcome under consideration, it can be assumed exogenous and a Kolmogorov test do not show any signi…cant di¤erences between its distribution and the uniform distribution.

It is worth emphasizing that the estimated endogenous neighborhood e¤ect is larger and better estimated on the subsample of respondents for which we have information on lag neighborhood attributes than on the total sample of respondents. This result holds true regardless of whether we use the past neighborhood attributes as additional regressor or not (see Table 9, column 4). The data on lag neighborhood attributes are not available for the aires surveyed at the beginning of the period under consideration and the subsample used in Table 9 corresponds mostly to aires surveyed after 1994. The aires surveyed at the beginning of period are typically those that we observe only one or two times, meaning those for which the measurement of the probability of being held back a grade is constructed from one or two observation only. It may cause larger measurement errors than in the aires that are surveyed three times and may explain why the neighborhood e¤ect is better identi…ed when we drop out these aires.

Table 10 replicates the regressions of Table 9 on private-sector housing on the one hand and, on the other hand, on public-sector housing. It con…rms the existence of signi…cant endogenous neighborhood e¤ects. The estimated e¤ect is larger in the public sector, perhaps due to the fact that the density of population is more important in public-sector neighborhood and, as a consequence, social interactions more e¤ective. The di¤erences between public and private estimations are not statistically di¤erent at standard level, however.

6 Conclusion

Buiding on the speci…cities of the French institutions, we propose three strategies for identifying the impact of neighborhood characteristics on the probability of being late at school. The …rst strategy uses the fact that French families are bound to send their children to schools which belong to the local district where they live. Given this legal constraint, information on the permanent income of families who move into a neighborhood are information on the unobserved quality of the neighborhood and can be used to control for its e¤ect on performances at school. The second strategy uses the fact that - in France- the distribution of public-sector housing is under the control of administrative authorities and that families settling in public-sector housing are not sorted across neighborhoods according to factors determining performances at school. Given this fact, neighborhood e¤ects can plausibly be identi…ed by focusing on the subsample of public-sector housing. Our third strategy builds on the fact that the date-of-birth within the year is a determinant of the probability of being late at school. We use the distribution of date-of-birth across the adolescents living in the neighborhood to identify neighborhood's in‡uence on children's performances.

Most interestingly, the three strategies provide signi…cant and consistent estimates for the e¤ects of neighborhood attributes on performances at school. Living in a neighborhood where parents are without diploma multiplies ceteris paribus the odds of being late at school by about 2.5, meaning an impact as

negative as the impact of having parents without diploma.

Our paper focuses on the probability of being late at school, which is arguably a very meaningful outcome in the French context. Further researches are needed, however, to explore the neighborhood e¤ects on other outcomes, such as the decision to drop out from school. This question is on our research agenda. We speculate that the neighborhood e¤ects on such decisions are even stronger - in France, at least- than the neighborhood e¤ects on performances stricto censu. Also further researches are needed to better explore the channels through which the children living in the same neighborhood in‡uence each other. It is obviously a key issue for de…ning public policies. In particular, it would be useful to separate the contribution of social interactions at school from the contribution of social interactions during extracurricular activities. Generally speaking, similar evaluations need to be performed in other countries to explore whether (and why) the role of social interactions varies across societies.

References

[1] Angrist J. and A. Krueger, 1991, "Does Compulsory School Attendance A¤ect Schooling and Earnings?", The Quarterly Journal of Economics, vol:106, n±4; Nov. 1991, pp:979-1014.

[2] Borjas G. J.. 1995. «Ethnicity, Neighborhoods, and Human-Capital Externalities ", The American Economic Review, Vol.85 n±3, 365-390.

[3] Brock, W. and S. Durlauf, 2001, "Interaction-Based Models", in Handbook of Econometrics vol:IV, J. Heckman and E. Leamer eds., Amsterdam: Noth-Holland.

[4] Case A. C. et L. Katz, 1991, "The Company You Keep: The E¤ects of Family and Neighborhood on Disadvantaged Youths"», NBER Working Paper n±3705, Cambridge, MA.

[5] Cutler D. et E. Glaeser, 1997, "Are Ghettos Good or Bad? ", Quarterly Journal of Economics, 112 : 827-872.

[6] Doblhammer G., 2003, "The Late Life Legacy of Very Early Life", MPID, Rostock, Working Paper WP-2003-31.

[7] Durlauf S. N. et H. Peiton Young (éditeurs), 2001, Social Dynamics , MIT Press.

[8] Edin P.A, P. Frederiksson and O.Aslund, 2003, "Ethnic Enclaves and The Economic Success of Immigrants: Evidence from a Natural Experiment", Quarterly Journal of Economics, 118 (1) 329-357.

[9] Evans W. N., W. E. Oates et R. Schwab, 1992, "Measuring Peer Group E¤ects: A Study of Teenage Behavior ", Journal of Political Economy, 1000, pp. 966-991.

[10] Ginther D., R. Haveman et B. Wolfe, 2000, " Neighborhood Attributes as Determinants of Children's Outcomes. How Robust are the Relationships ? ", Journal of Human Resources, XXXV.

[11] Glaeser E. L. and J.A. Scheinkman, 2001, "Measuring Social Interaction", in Social Dynamics, Durlauf et Young eds, MIT Press.

[12] Krueger A., 2003, " Economic Considerations and Class Size ", Economic Journal, n± 113, pp. F34-F64.

[13] Lewbel A., 2000, " Semiparametric Qualitative Response Model Estimation with Unknown Heteroscedasticity or Instrumental Variables ", Journal of Econometrics.

[14] Manski C. F., 1993, " Identi…cation of Endogenous Social E¤ects : The Re‡ection Problem ", Review of Economic Studies, 60.

[15] Magnac T. and Maurin E. , 2003, "Identi…cation and Information in Monotone Binary Models", CREST Working Paper 2003-07.

[16] Maurin E., 2002, "The Impact of Parental Income on Early Schooling Transitions: A Re-examination Using Data Over Three Generations", Journal of Public Economics, 85:301-32.

[17] Mo¢ tt R. A., 2001, "Policy Interventions, Low-level Equilibria and Social Interactions", in Social Dynamics, Durlauf et Young eds, MIT Press.

6.1 Appendix A: the control function approach

In France, the administration de…nes a local district for each public school and families are bound to send their children in the local district where they live. In other words, chosing a place where to live amounts to chosing the quality of public schools available for the children and the social composition of these schools. Also, assuming that families actually prefer settling in high-quality neighborhoods, the costs of settling and living in a high-quality neighborhood should be higher than in a low-quality one. Given this fact, the socioeconomic status of families who just move into a neighborhood provide an indicator for the the quality of the corresponding schooling context and may be used as a control function to purge out the e¤ect of unobserved neigborhood qualities.

To be more speci…c, assume that the baseline model can be rewritten,

$$y = \beta E(y \mid n) + \gamma v + \xi_n + u \qquad (3)$$

where v represents the individual factors (typically the date-of-birth within the year), ξ_n represents the unobserved quality of the neighborhood and u an individual idiosyncratic random e¤ect. For the sake of simplicity, assume that there exist only two types of neighborhoods, i.e. low-quality ξ_L and high-quality ξ_H; with $\xi_L = -\xi_H = \xi_0$: At each date, we assume that N families disappears and are replaced by N new families. We denote N_H (N_L) the number of housing

available in high-quality (low-quality) neighborhoods and $°_H$ ($°_L$) the costs of living in a high-quality (low-quality) neighborhood. $°_H$ is market determined while $°_L$ is exoegnous. New families observe the exogenous attributes of the neighborhood $²_n$ and chose $n \in \{L; H\}$ in order to maximize a utility function $U(²_n; C)$ where C is private consumption, subject to $C + °_n = R$, where R is the permanent income. For any given $°$; a family choses to settle in a high-quality neighborhood if an only if its permanent income R satis…es $U(²_H; R - °_H) > U(²_L; R - °_L)$ and the equilibrium price of high-quality neighborhood $°^¤_H$ is such that the number of families satisfying $U(²_H; R - °^¤_H) > U(²_L; R - °_L)$ is N_H (and the number of families satisfying $U(²_H; R - °^¤_H) < U(²_L; R - °_L)$ is N_H). Within this framework, families settle in high-quality neighborhood if and only if their permanent income R is greater than $R^¤$ as de…ned by $U(²_H; R^¤ - °^¤_H) = U(²_L; R^¤ - °_L)$ and there is a one-to-one correspondance between the average income of families who settle in a neighborhood and the unobserved quality of this neighborhood. Given this fact, we can control at each date for the e¤ect of $²_n$ by introducing a measurement for the permanent income of families who just settle in n.

Table 1: The distribution of respondents aged 15 according to the number of other adolescents aged 15 observed in their aire

Nb other adolescents observed in the aire	Nb respondents	%
1	4245	17.4
2	4393	18.0
3	3678	15.1
4	3122	12.8
5	2262	9.3
6	1710	7.0
7 ou +	4958	20.3
Total	*24368*	*100*

Source : LFS, 1991 à 2002, Insee.
Field : children's born in *t-15*.

Table 2: Some Descriptive Statistics

Variables	Mean	Standard Deviation
Having been held back	.42	.49
Proport. Held back	.42	.34
Proport. Parents College graduate	.07	.15
Proport. Parents High-school dropout	.38	.28
Proport. Parents Non-french	.09	.19
Boy	.51	.50
Father non-French	.06	.23
Father High-school dropout	.21	.41
Mother High-school dropout	.27	.44
Father college graduate	.14	.34
Mother college graduate	.15	.35

Source : LFS, 1991 à 2002, Insee.
Field : children's born in *t-15*.

Table 3: Endogenous Neighborhood Membership: an Analysis of the Factors which Determine Jointly Neigborhood Choice and Performances at School

Logistic models	Dependent Variable: To be behind at school		
Independent variables	(1)	(2)	(3)
Neighborhood's characteristics:			
% 15 years old behind at school	0.69 (0.14)	0.25 (0.16)	0.07 (0.17)
% college graduates			-0.50 (0.44)
% high-school dropouts			0.71 (0.28)
% non-French			0.71 (0.38)
Individual's characteristics :			
Intercept	-0.44 (0.21)	-0.27 (0.24)	-0.47 (0.25)
[Boy=1]	0.45 (0.10)	0.47 (0.11)	0.48 (0.11)
[Non-french=1]		0.86 (0.22)	0.50 (0.27)
Father's education:			
- missing		0.42 (0.14)	0.36 (0.14)
- no diploma		0.20 (0.16)	0.05 (0.16)
- vocational		ref	ref
- high-school grad.		-0.27 (0.22)	-0.24 (0.22)
- college grad.		-0.78 (0.20)	-0.81 (0.21)
Mother's education:			
- missing		0.43 (0.23)	0.39 (0.23)
- no diploma		0.60 (0.13)	0.45 (0.14)
- vocational		ref	Ref
- high-school grad.		-0.86 (0.19)	-0.80 (0.19)
- college grad.		-1.12 (0.20)	-1.08 (0.21)
Year dummies	yes	yes	Yes
Likelihood ratio (p)	57.2 (*<0.0001*)	296.8 (*<0.0001*)	319.9 (*<0.0001*)
Number of observations	1 707	1 707	1 707

Source : LFS, 1991 à 2002, Insee.

Field : children's born in *t-15* and who have just moved into a neighborhood.

Reading: Once we control for education and nationality, we find no significant correlation between the probability of being late at school which characterizes the adolescents who into a neighborhood and the probability which characterizes the adolescents already living in the neighborhood.

Table 4: Neighborhood Effects on the Probability of Being Late at School: an Estimation controlling for Endogenous Neighborhood Membership

Logistic models	Dependent variable : To be behind at school			
Independent variables	(1)	(2)	(3)	(4)
***Neighborhood's characteristics**:*				
% 15 years old behind at school	0.94 (0.04)	0.52 (0.04)	0.44 (0.05)	
% college graduates			-0.12 (0.14)	-0.29 (0.14)
% high-school dropouts			0.37 (0.07)	0.55 (0.07)
% non-French			0.03 (0.11)	0.03 (0.11)
Individual's characteristics :				
Intercept	-0.73 (0.06)	-0.60 (0.06)	-0.69 (0.07)	-0..51 (0.08)
[Boy=1]	0.49 (0.03)	0.57 (0.03)	0.57 (0.03)	0.56 (0.03)
[Non-french=1]		0.29 (0.06)	0.27 (0.08)	0.24 (0.07)
Father's education:				
- missing		0.42 (0.04)	0.40 (0.05)	0.42 (0.05)
- no diploma		0.43 (0.04)	0.37 (0.04)	0.42 (0.04)
- vocational		ref	ref	ref
- high-school grad.		-0.48 (0.06)	-0.46 (0.06)	-0.47 (0.06)
- college grad.		-0.84 (0.06)	-0.79 (0.06)	-0.84 (0.06)
Mother's education:				
- missing		0.58 (0.09)	0.57 (0.09)	0.58 (0.09)
- no diploma		0.54 (0.04)	0.48 (0.04)	0.53 (0.04)
- vocational		ref	ref	ref
- high-school grad.		-0.78 (0.05)	-0.75 (0.06)	-0.78 (0.05)
- college grad.		-1.07 (0.06)	-1.02 (0.06)	-1.06 (0.06)
Year dummies	yes	yes	yes	yes
Likelihood ratio (*p*)	1 044 (*<0.0001*)	4 175 (*<0.0001*)	4 207 (*<0.0001*)	4 116 (*<0.0001*)
Number of observations	24 368	24 368	24 368	24 368

Source : LFS, 1991 à 2002, Insee.
Field : children's born in *t-15* and who have been living in the neighborhood for more than one year.
Reading: The endogenous neighborhood effect on the log odds of being late at school remains significant and large (+.44) even after controlling for parental education and nationality and for the average level of education and nationality in the neighborhood.

Table 5: Neighborhood Effects on the Probability of Being Late at School: an Estimation controlling for Endogenous Neighborhood Membership and Omitted Neighborhood Attributes

Logistic models	Dependent variable : To be behind at school			
	(1)	(2)	(3)	(4)
Neighborhood's characteristics:				
% 15 years old behind at school	0.86 (0.11)	0.48 (0.12)	0.33 (0.12)	
% college graduates			-0.15 (0.39)	-0.22 (0.39)
% high-school dropouts			0.67 (0.20)	0.80 (0.20)
% non-French			0.16 (0.26)	0.21 (26)
Individual's characteristics :	no	yes	yes	yes
Characteristics of families who move into the neighborhood	no	yes	yes	yes
Year dummies	yes	yes	yes	yes
Likelihood ratio (*p*)	347 (*<0.0001*)	926 (*<0.0001*)	941 (*<0.0001*)	933 (*<0.0001*)
Number of observations	4 974	4 974	4 974	4 974

Source : LFS, 1991 à 2002, Insee.
Field : children's born in *t-15*, who have been living in the neighborhood for more than one year.
Reading: The effect of the proportion of high-school dropouts among the parents living in the neighborhood on the log odds of being late at school is significant and large (+.80) even after controlling for parental education and nationality and for the average level of education, occupational status, nationality of the families who have just moved into the neighborhood.

Table 6a: Neighborhood Effects on the Probability of Being Late at School: subsample of public-sector housing

Logistic model	Dependent variable: To Be Behind at school		
	Just Move	Already present	Already present
% aged 15 late at school	.05 (.35)	.63 (.09)	.50 (.10)
Individual controls	no	no	yes
Year dummies	yes	yes	Yes
Likelihood (p)	21.2 (0.07)	183.6 (0.00)	388.7 (0.00)
Nb Obs.	388	4626	4626

Source : LFS, 1991 à 2002, Insee.
Field : children's born in *t-15* and who have just moved into a public-sector neighborhood (column 1). Children's born in *t-15* living in a private sector neighborhood and who have been living in the neighborhood for more than one year (column 2 and 3). The different models also include a (boy=1) dummy.
Reading: No correlation exists between the probability of being late which characterizes the families who have just moved into a public sector housing and the probability of being late which characterizes the neighborhood (Table 6a, column 1).

Table 6b : Neighborhood Effects on the Probability of Being Late at School: subsample of private-sector housing

Logistic model	Dependent variable: To Be Behind at School		
	Just Move	Already present	Already present
% aged 15 late at school	.61 (.16)	.75 (.05)	.44 (.05)
Individual controls	no	no	yes
Year dummies	yes	yes	yes
Likelihood (p)	42.0 (0.00)	664.5 (0.00)	3139 (0.00)
Nb Obs.	1319	19 742	19 742

Source : LFS, 1991 à 2002, Insee. Field : : children's born in *t-15* and who have just moved into a private-sector neighborhood (column 1). Children's born in *t-15* living in a private sector neighborhood and who have been living in the neighborhood for more than one year (column 2 and 3).). The different models also include a (boy=1) dummy.

Reading: A very significant correlation exists between the probability of being late which characterizes the families who have just moved into a private sector housing and the probability of being late which characterizes the neighborhood (Table 6b, column 1).

Table 7: IV strategy: first-stage regression

	Dependent variable : neighborhood proportion of 15 years' old who are behind at school
	OLS
Proportion children born in :	
January	-0.16 (0.02)
February	-0.16 (0.02)
March	-0.11 (0.02)
April	-0.14 (0.02)
May	-0.11 (0.02)
June	-0.12 (0.02)
July	-0.10 (0.02)
August	-0.03 (0.02)
September	-0.03 (0.02)
October	-0.07 (0.02)
November	-0.04 (0.02)
December	ref
[Boy=1]	-0.002 (0.004)
Year dummies	Yes
F-test (*p*)	21.1 (*<0.0001*)
R^2	0.02
Number of Observations	24 369

Source : LFS, t=1991 à 2002, Insee.
Field : children's born in *t-15*.
Note: For each respondent age 15, the proportion of children born in January (or February, …, December) is constructed from the available information on the month of birth of the other adolescents observed in his/her *aire*. Similarly the proportion of children late at school is constructed from the available information on the situation at school of the other adolescents observed in the aire. The regression shows that the larger the proportion of children born in January, the smaller the proportion of children late at school.

Table 8: An estimation of endogenous social effects:
an instrumental variable approach

Linear probability model	Dependent variable : To be behind at school	
	OLS	IV
Proportion 15 years' old behind at school	0.22 (0.01)	0.17 (0.09)
[Boy=1]	0.12 (0.01)	0.12 (0.01)
Year dummies	Yes	Yes
R2	0.442	0.434
Nb Obs.	24 369	24 369

Source : LFS, t=1991 à 2002, Insee.
Field : children's born in *t-15*.
Note : The instrumental variables correspond to the neighborhood's proportions of children born in t-15, in January, February, …., December.

Table 9: An estimation of present and past endogenous social effects:
an instrumental variable approach

Linear probability model	Dependent variable : To be behind at school		Dependent variable : To be behind at school	
Independent variables	MCO	IV	MCO	IV
% 15 years'old behind at school	0.21 (0.01)	0.35 (0.11)	0.22 (0.01)	0.29 (0.12)
Past % 15 years' old behind at school	0.08 (0.01)	0.11 (0.11)		
Intercept	0.33 (0.01)	0.26 (0.05)	0.33 (0.01)	0.30 (0.06)
[Boy=1]	0.11 (0.01)	0.11 (0.01)	0.11 (0.01)	0.11 (0.01)
R^2	0.043	0.021	0.041	0.021
Number of Observations	15 293	15 293	15 293	15 293

Source : LFS, t=1991 à 2002, Insee.Field : children's born in *t-15* .
Note : The instrumental variables correspond to the neighborhood's present and past proportions of children born in t-15, in January, February, …., December.
Standard deviation in parentheses**.**

Table 10: Replication of the IV analysis by housing sectors

Linear probability model	IV Private-sector Housing	IV Public-sector Housing
% 15 years'old behind at school	.34 (.12)	.54 (.19)
Past % 15 years' old behind at school.	0.15 (0.13)	-0.12 (.13)
Nb Obs.	12 360	2,930

Source : LFS, t=1991 à 2002, Insee.
Field : children's born in *t-15*.
Note : The instrumental variables correspond to the neighborhood's present and past proportions of children born in t-15, in January, February, …., December. Standard deviation in parentheses**.**