ELSEVIER

Journal of Public Economics 89 (2005) 797–819

JOURNAL OF PUBLIC ECONOMICS

www.elsevier.com/locate/econbase

The effect of overcrowded housing on children's performance at school

Dominique Goux[a,b,*], Eric Maurin[c,d]

[a]*INSEE, Timbre F230, Division emploi, 18, boulevard Adolphe Pinard, 75675 Paris Cedex 14, France*
[b]*ENS, France*
[c]*CREST, France*
[d]*CEPR, France*

Received 19 November 2001; received in revised form 11 June 2004; accepted 25 June 2004
Available online 25 September 2004

Abstract

This paper provides estimates of the causal effect of living in an overcrowded home on performance at school in France. Our identification strategy relies on the fact that the size and housing conditions of families vary with the sex composition of the siblings. In particular, large families in which the two youngest children are (by descending age) a boy and a girl tend to live less often in overcrowded housing than the other families. French parents seem to be more reluctant about bringing up their children in the same room when they are not of the same sex, especially when the youngest one is a girl. We build on these results to develop several econometric analyses of the effect of overcrowding on schooling outcomes using variables describing the sex composition of the siblings as instrumental variables. These different strategies reveal that the very strong statistical relationship between housing conditions and academic failure is plausibly one of cause and effect. Children in large families perform much less well than children in small families, but our IV estimates suggest that this is mostly due to the fact that they live in more overcrowded homes.

Keywords: School; Children; Overcrowded housing

* Corresponding author. INSEE, Timbre F230, Division emploi, 18, boulevard Adolphe Pinard, 75675 Paris Cedex 14, France. Tel.: +33 1 41 17 5442; fax: +33 1 41 17 6163.

E-mail address: dominique.goux@insee.fr (D. Goux).

0047-2727/$ - see front matter
doi:10.1016/j.jpubeco.2004.06.005

1. Introduction

Children from poor families do not do as well and leave school earlier than children from rich families. These are well-known facts that no longer need to be validated. The interpretation of these facts, however, is still the subject of great controversy. Consequently, public policies that could help reduce inequalities in educational opportunities remain poorly defined.

One basic issue is whether increasing financial aid to the poorest families represents a good means for improving their children's performance at school. A number of studies argue that parental income, as such, does not have any impact on children's performance at school. According to these studies, the link between poverty and academic failure is not one of cause and effect. They stress that increasing financial aid to poor families would have no effect on the inequalities between children from rich and poor families.[1]

Another important issue concerns the impact of targeted aid, aimed to directly improve the living conditions of poor children. Even if financially assisting the parents of the poorest families would not have any effect on their children's schooling, aid aimed at specifically improving children's access to medical care or quality of housing could have a very important and positive effect on children's development and performances at school.[2]

In this paper we try to contribute to this second debate. We focus on one aspect of children's living conditions, which we suspect to be of particular importance—the amount of personal space they have at home for their activities. More specifically, we try to evaluate the impact of the number of persons per room on the probability of being held back in primary or junior high school. This does not mean measuring the overall effects of parental income, but the effects of a very particular potential use of parental income—the spending allotted for housing so that children do not have to live in an overcrowded space. The underlying issue is to understand whether public policies favoring quality housing for low-income families could also serve as a vehicle for improving the performances of their children and equal opportunities at school.

To shed light on this issue, we have used the French Labor Force surveys, which were conducted each year by the French National Institute for Statistics and Economic Studies (hereafter, INSEE) between t=1990 and t=2000. These surveys have provided us with large samples of 15-year-olds (i.e., born at $t-15$) with information on whether they have been held back a grade in elementary school or in junior high school,[3] as well as on their family size and on how many people there are per room in their home. This data set makes possible to analyze the impact that overcrowded housing has on adolescents' situation at

[1] See Mayer (1997), Blau (1999) and Shea (2000).

[2] The absence of a direct effect of parental income on children's performance does not imply that an improvement in poor children's living conditions would not have a positive effect on their success at school. The absence of an income effect can just as well mean that the parents receiving an income supplement have other priorities than to improve the conditions related to their children's success at school.

[3] To repeat a grade is plausibly one of the most direct indicator of performance at school in France. The recent Program for International Student Assessment (PISA) conducted by the OECD shows that the 15-year-old French adolescents who are late at school obtain much lower scores in math, reading or science than normal-age adolescents. The difference represents about 1.14 standard deviation of the score in math, 1.26 SD in reading, 1.17 SD in science (see Murat and Rocher, 2003).

school using very large samples of 15-year-old adolescents. Given that only one-third of the LFS samples are renewed each year, we can also construct subsamples of adolescents born at $t-15$ with information on their situation at school at t and $t+1$. Hence, our data set also makes possible to evaluate the impact of housing conditions at t on the probability of moving up to the next grade at $t+1$.

To identify the causal effect of the housing conditions, we have consecutively used two different sets of instruments. The first set is constructed from the available information on the sex composition of the siblings. Families in which the two oldest children are the same sex are more likely to have a third child than the other families. In addition, conditional on having a third child, families in which the two youngest children are (by descending age) a boy and a girl tend to live less often in overcrowded housing than the other families. There exists no significant correlation between the sex of the two youngest siblings and the final size of families with three children or more, however. All in all, the sex composition of the two oldest children may be interpreted as a random shock affecting the probability of having a third child, while the sex composition of the two youngest children in large families may be considered as a random shock affecting the subsequent housing conditions of these families. These findings plausibly reflects that French parents (a) prefer large or mixed-gender families to small, same-sex ones but, (b) are more reluctant about bringing up their children in the same room when they are not the same sex, especially when the youngest one is a girl. Given these facts, we have focused on families with two or more children and constructed two basic instrumental variables. The first one S_0 is a dummy indicating that the two oldest children are the same sex. The second one S_1 is a dummy which is defined on the subsample of large families and which indicates that the two youngest children are (by descending age) a boy and a girl. First-stage regressions confirm that S_0 affects significantly family size while S_1 affects significantly overcrowding in families with three or more children. Our basic identification assumption is that this is the main channel through which the sex composition of the siblings (as described by S_0 and S_1) affects performances at school.[4] Within this framework, we have developed several strategies for identifying the effect of overcrowding on educational outcomes. The most basic strategy focuses on the subsample of families with three or more children and identifies the effect of overcrowding using S_1 as an instrumental variable. Another basic strategy uses the full sample of families with two or more children and identifies jointly the effects of family size and overcrowded housing using jointly S_0 and S_1 as instrumental variables.[5] The second strategy relies on the sex composition of the youngest and the oldest children to identify the effect of housing conditions while the first strategy relies on the sex composition of the youngest children only. Comfortingly, the different strategies provide very similar estimates of the effect of overcrowding on educational outcomes.

[4] As discussed below, sex differences between the oldest children have already been used in other contexts, most notably for estimating the impact of family size on mothers' labour supply (see Angrist and Evans, 1998 or Rosenzweig and Wolpin, 2000).

[5] When we work on the full sample of families with two or more children, we have to redefine S_1 by standardizing it and interacting it with a dummy indicating that the family has three or more children.

The second set of instruments is constructed from the available information on the parents' place of birth. Our data show that parents' places of birth affect the housing conditions of families even after controlling for the socioeconomic level of parents. Specifically, it is possible to divide metropolitan France into large areas such that (a) the average socioeconomic level of parents does not vary with respect to the areas where they were born, but such that (b) the average number of persons per room varies significantly with respect to the areas where the parents were born. Generally speaking, parents born in urban areas tend—ceteris paribus—to live in more overcrowded housing than parents born in nonurban areas. We have constructed dummies indicating the areas where the parents were born and we have used these dummies as instrumental variables. The identification assumption is that this is the main channel through which parents' place of birth affects school performances.

To anticipate the rest of the paper, our main empirical findings may be summarized as follows. First, a very significant correlation exists between children's situation at school and housing conditions, even after controlling for family size or family socioeconomic status. The median of the distribution of the number of persons per room is 1. About 59% of the adolescents living in families above this median are late at school, a proportion that is about 30 percent points higher than it is for adolescents living in families below the median. In addition, regarding the 15-year-olds who are still normal age, those who are living in the more overcrowded housing are much less likely to move up to the next grade than the other normal-age children. About 37% of the normal-age adolescents living in the more overcrowded housing (i.e., above the median) do not move up to the next grade against only about 21% of those living in the less overcrowded housing (i.e., below the median).

Second, the causal effect of overcrowding is probably even larger than what the raw correlation suggests. The IV estimates of overcrowding effect are significantly greater than the OLS estimates regardless of whether we use instruments built from the information on the sex composition of the siblings or on the parents' place of birth and regardless of whether we assume that family size is exogenous or endogenous. In addition, the results remain very much the same when we add a measurement of family permanent income as a supplementary control variable. Our data only provide an indirect and potentially rough measurement for housing conditions. The downward biases which affect the OLS estimates plausibly correspond to biases that arise from measurement errors. All in all, our data provide an array of findings that suggest that overcrowded housing is an important way in which parental poverty affects children's outcomes. Children in large families perform much less well than children in small families, but our IV estimates suggest that this is mostly due to the fact that they live in more overcrowded homes.

The paper is organized in the following way. In the next section, we present an overview of medical, sociological and sociopsychological literature, which describes the impact of overcrowded housing on the health and behavior of individuals. In Section 3, we develop a model for parental behavior, making it possible to define what is meant by the causal impact of overcrowding on school performance, as well as the econometric strategies that make it possible to identify that impact. In Sections 4 and 5, we describe the data and methods used, and the econometric results are presented in Section 6.

2. The effects of overcrowded housing: an overview of the literature

The sociological and social psychological literature has long been interested in the problems caused by overcrowded housing.[6] Empirically, the degree of overcrowding is measured by the number of persons per room. Theoretically, the problems caused by lack of living space are conceptualized as the consequences (a) of an excess of interactions, stimulations and demands from the people living in the immediate area, and (b) of a lack of intimacy and the possibility of being alone. People who live in overcrowded housing suffer from not being able to control outside demands. It is impossible for them to have the necessary minimum amount of quiet time they need for their personal development. One of the most convincing sociological studies on this subject is perhaps that of Gove et al. (1979). Using American data, the authors establish the existence of a very clear correlation between the number of persons per room and individuals' mental and physical health.[7]

Medical literature has also shown great interest in the health of people living in overcrowded conditions, i.e., in houses and/or apartments that are too small for their families. It has been well established that individuals living or having lived in such conditions are sick more often than others, particularly due to respiratory insufficiency and pulmonary problems[8] (Britten et al., 1987, Rasmussen et al., 1978, Mann et al., 1992). In general, people who grow up in overcrowded housing die at a younger age than others (Coggon et al., 1993, Deadman et al., 2001), most notably of cancer (Barker et al., 1990).

The medical literature gives many reasons for these health problems and their persistence. Living in an overcrowded space is a source of stress and favors illnesses linked to anxiety. The members of a family living in a crowded space also transmit their infections to one another more easily, weakening their immune systems. Living in an overcrowded space puts people at greater risk to problems linked to poor ventilation and hygiene conditions, such as poisoning caused by the smoking of one or more family members (see the survey by Prescott and Vestbo, 1999).

With overcrowded housing, occupants' health at greater risk and their capacity for intellectual concentration being decreased, it is clear that a lack of space is a potentially unfavorable factor for children's success at school. To our knowledge, however, no study that analyzes the nature and intensity of the links between available living space and children's success at school exists in the economic literature. The work published in the

[6] Since the 1960s, experiments carried out on groups of rats have brought to light the very serious behavioral and social problems that occur in animals when the size of their vital living space is modified.

[7] In addition, the authors establish that the number of persons per room is a good measurement for feelings of excessive outside demands and lack of private time. They also show that the quality of care given to children, and more generally, the quality of the relationship between parents and their children, tends to deteriorate when the number of individuals per room increases. Gove et al.'s (1979) results are obtained from American data, but they compare rather well with Chombart de Lauwe's (1956) seminal results based on French data, which also establish a statistical relationship between the number of persons per room and the frequency of social pathologies.

[8] At greater risk due to unhygienic conditions, they suffer more often than others from appendicitis inflammation. According to Coggon et al. (1991), the drop in appendicitis inflammation cases observed since the beginning of the 1960s in Anglesey is linked more to the decrease in the number of overcrowded housing than to the improvement in the housing's modern conveniences. In addition, Fuller et al. (1993) established a link between the degree of overcrowded housing and the probability of mental health problems through analyzing data from Thailand.

sociological and medical literature corresponds essentially to the analysis of statistical correlations. Given that housing and health problems probably share common unmeasured determinants, these statistical correlations do not necessarily correspond to relations of cause and effect. The meaning of the results obtained from this literature is unclear.

In the next section, we will develop an economic model of family behavior that makes it possible to define what we mean by the causal effect of overcrowding on children's performance at school. This model will also help us to define econometric strategies that make it possible to identify this effect.

3. Theoretical framework and econometric model

In this section, we develop a model for family behavior that describes the simultaneous determination of the number of persons per room and the probability of academic failure. Our purpose is to define what is meant by the effects of overcrowding on schooling and to clarify the conditions that make it possible to econometrically identify this effect. Our model is based on the following assumptions:

(H1). The academic abilities of a child (noted Q_i for the child i) depend on the exogenous characteristics of the child measured in the survey (x_i), the characteristics unmeasured in the survey (u_i), but also on the total number of children in the family (N_i) as well as the amount of space available for each member in the family home (L_i). The underlying assumption is that children do better at school when (1) they have a quiet room for studying, and (2) have parents who do not have to divide their resources between too many children. To stay within a simple framework, we assume that Q_i can be log-linearly decomposed,

$$\ln Q_i = \alpha \ln L_i + \beta N_i + \gamma x_i + u_i. \tag{3.1}$$

(H2). A child experiences academic failure and repeats a grade in elementary school and/or middle school if his/her scholastic abilities Q_i are lower than a minimum aptitude threshold, which depends only on his/her relative age within his/her age group[9] (written a_i). The assumption is that at a given level of ability, a child is more vulnerable to being held back if he/she was born at the end of the year, meaning that he/she is among the youngest of his/her age group. In noting E_i as the dummy variable with a value of 1 when the child i is failing, we postulate that there exists an intercept Q_0 and a parameter θ, such that we can write:

$$E_i = 1 \Leftrightarrow \ln Q_i + \theta a_i < Q_0. \tag{3.2}$$

(H3). The final number of children (N_i) and the available space for each children (L_i) in the family home are chosen sequentially by parents, in the spirit of Rosenzweig and Wolpin (2000). The corresponding two-period model is developed in Appendix A. At the

[9] In france, two children belong to the same age group(i.e., are in the same year of school) if they were born in the same year.

beginning of the first period, each family consists in two parents and N_0 children, N_0 being exogenously determined by parents' preferences. It corresponds to the minimum number of children that they want. In the empirical application, we assume N_0=2. During the first period, the parents chose the final size of the family in light of the characteristics (denoted S_0) of their N_0 first children. For instance, parents with two girls and with strong preferences for mix-gender families are—ceteris paribus—more likely to have a third child than parents who do not care about the sex composition of their children. During the second period, the parents chose the space available for each child in light of the final size of family and of the final characteristics of children[10] (S_1). They compare the costs of a larger house with the benefits of having a separate room for each child. Assuming for instance that parents are less reluctant about bringing up two children in the same room when the youngest one is a girl and the oldest one is a boy, the number of room per persons should be on average less important in families where the last children are a boy and a girl than in other families.

Within this framework, the decisions made by the parents lead them (1) to chose a final number of children $N=N(R,q,S_0)$ as a function of income R, prices q and oldest children's characteristics S_0 and (2) to express a housing demand $L=L(R,q,S_0,S_1)$ as a function of income, prices, and children's characteristics. With these notations, our purpose is to determine the impact α of L_i on academic failure E_i when N_i, x_i, a_i and u_i are kept constant. Using Eqs. (3.1) and (3.2), the corresponding model can be written as:

$$E_i = 1 \Leftrightarrow \alpha \ln L_i + \beta N_i + \gamma x_i + a_i + u_i < 0. \tag{3.3}$$

where the coefficients are normalized so that the impact of the relative age is equal to 1 (i.e., θ=1). By convention, the intercept Q_0 is included in the group of exogenous variables x_i.

If the unobserved factors of academic failure u_i could be assumed independent from L_i and N_i, the identification of α would not cause any particular problem. The problem is that these factors may be correlated with family resources (most notably, R) and, consequently, with N_i and L_i. In this scenario, it is unclear whether the correlations observed between E_i and L_i reflect the causal effect of L_i on E_i, or the fact that the two variables L_i and E_i both vary simultaneously with u_i. To avoid this kind of problem, it is necessary to observe instrumental variables that affect academic failure E_i only insofar as they affect the size of the family and its housing conditions. Within our theoretical framework, such instruments typically correspond to variables[11] describing S_0 and/or S_1. Assuming that these two variables affect schooling only insofar as they affect overcrowding and family size, they provide us with plausible instrumental variables for identifying the effect of L and N on E.

[10] Notice that, by construction, S_1=S_0 when the parents chose not to have a third child, while S_1 can be written $S(S_0,s)$ when there is a third child, where s denotes the characteristics of the third child. During the first period, the parents make their decision based on expectations about the potential realization of s.

[11] Any variable determining the preferences for kids and uncorrelated with R may provide alternative plausible instrumental variables.

4. Data and variables

The data used come from the French annual Labor Force Surveys that were carried out between 1990 and 2000 by the French Statistical Office (INSEE). Each survey corresponds to a sample of about 80,000 households, representative of the population of French households (sampling rate 1/300). Each member of the household who is 15 or older is surveyed, with the cut-off age being December 31, of the year preceding the one the survey is conducted. These surveys make it possible to construct a large sample of 15-year-olds (i.e., responding in t, born in $t-15$), and to analyze the links that exist between their housing conditions and situations at school.

Specifically, our analysis will be carried out using the sample representative of those individuals who were born in $t-15$, observed in the Labor Force Surveys conducted in t=1990,. . .,2000, living in two-parent families with at least two children. This sample contains about 19,000 observations.

An interesting feature of the French Labor Force Surveys is that only one-third of the sample is renewed each year. For each t, we can construct a subsample of adolescents born in $t-15$ with information on their situation at school at t and t+1. This subsample (N=5793) makes possible to analyze the links between the housing conditions at t and the probability of repeating a grade at t+1 (i.e., being in the same grade at t and t+1).

For each 15-year-old respondent, the Labor Force Survey gives (a) their sex, date of birth and the grade they are in at the time of the survey; (b) the number of persons and the number of rooms in their home; (c) their parents' wages and occupations (which makes it possible to code their families' socioeconomic level using the French Occupational Prestige Scale); (d) their parents' places of birth and (e) the number, sex and birth date of the other children living in the home.

Respondents of year t born in $t-15$ are in at least the ninth grade if they have not repeated a year. Thus, our measurement for "having repeated a grade in elementary school and/or middle school" is simply a dummy variable that equals 1 if they are not yet in the ninth grade. For respondents that are tracked for 2 years, our measurement for "repeating a grade" equals 1 if they are in the same grade at t and t+1.

Table 1
The Labour Force Surveys' samples: descriptive statistics

Variables	Families with two children or more		Families with three children or more	
	Mean	Standard deviation	Mean	Standard deviation
Being held back	0.43	0.50	0.50	0.50
No. of persons/room	1.09	0.45	1.24	0.50
Three or more children	0.54	0.50	–	–
Father SES	−0.09	0.94	−0.24	0.96
Boy	0.52	0.50	0.52	0.50
No. of observations	19,499		10,570	

Source: Labour Force Surveys, 1990–2000, INSEE.

Field: Children who were born in $t-15$ and surveyed in t living in an intact family.

Note: Father's socioeconomic status (SES) corresponds to father's position on the French occupational prestige scale (Chambaz et al., 1998).

Knowing the number of rooms (NR) and the number of persons (NP), it is possible to evaluate the number of children per room (NP/NR) for each home. Our econometric work has mostly consisted in regressing our dummy variable for academic failure on this measurement of the space per person in the home, using family size and family income as control variables. Table 1 provides standard descriptive statistics.

5. Instrumental variables and estimation method

For the rest of this paper, our purpose will be to identify the parameter α that appears in Eq. (3.3). The basic issue is to find instrumental variables which affect the performance at school only insofar as they affect overcrowding. We have used in turn two different sets of instruments.

The first set is constructed from the information available on the sex differences between the siblings.[12] Table 2 shows the mean number of persons per room by family size and sex composition of the siblings.[13] It reveals that the proportion of large families (i.e., with three or more children) is significantly more important in families where the two oldest children are the same sex. Furthermore, for the group of families with three or more children, the mean number of persons per room is significantly less important in families where the two youngest children are a boy and a girl (by descending age) than in other families. In contrast, for the group of families with three children or more, the final number of children does not vary significantly with the sex composition of the youngest children. These findings plausibly reflect that parents (a) prefer large or mixed-gender families to small and same-sex ones and (b) are more reluctant about bringing up two children in the same room when they are not the same sex and when the youngest is a girl. In addition, these results are consistent with our theoretical framework when S_0 is interpreted as a dummy indicating whether the two oldest children are the same sex and S_1 is interpreted as a dummy indicating that the two youngest children are a boy and a girl (by descending age). Table 3 shows the corresponding first-stage regressions which confirm what raw statistics suggest. Specifically, S_1 has a significant impact on the number of persons per

[12] There is controversy in the literature about whether sex composition has, as such, an effect on education (see Butcher and Case, 1994, Hauser and Kuo, 1998, Kaestner, 1997, Conley, 2000, Rosenzweig and Wolpin, 2000). Butcher and Case (1994) argue that, for girls, the presence of any sisters reduces educational attainment, while they find no effect of sex composition among boys. Hauser and Kuo (1998) argue that Butcher and Case overinterpret weak evidence and state that there is no convincing support for any hypothesis suggesting an effect of the gender composition of the sibship. Notice that if sex composition matters only for girls, then sex composition should be a valid instrument in a sample of boys. As discussed below, our IV estimates are very similar regardless of whether we focus on the sample of boys or on the total sample.

[13] The survey on Formation et Qualification Professionnelle conducted in 2003 by the French statistical office provides an information for each French women on the number, sex and date-of-birth of her children. We have checked that when we focus on women aged 40 or more (i.e., having finished having children) and with two children or more, we obtain almost exactly the same proportion of families where the two oldest children are two boys, a boy and a girl, etc. as in our Table 2. Boys are more likely than girls and also we have checked that a second birth is slightly more likely when the first one is a boy.

Table 2
Overcrowded housing by family size and sex composition of the siblings

Sex of the siblings	Families with two or more children			Families with three or more children		
	Fraction of sample	Mean no. of persons per room	Fraction with three or more children	Fraction of sample	Mean no. of persons per room	Mean no. of children
Two oldest siblings:						
Boy–boy	0.280 (0.002)	1.101 (0.006)	0.587 (0.005)	0.299 (0.003)	1.228 (0.008)	3.72 (0.025)
Girl–girl	0.231 (0.002)	1.101 (0.006)	0.582 (0.006)	0.243 (0.003)	1.241 (0.009)	3.75 (0.025)
Girl–boy	0.238 (0.002)	1.078 (0.007)	0.514 (0.005)	0.222 (0.003)	1.241 (0.011)	3.77 (0.025)
Boy–Girl	0.251 (0.002)	1.078 (0.007)	0.524 (0.005)	0.238 (0.003)	1.242 (0.011)	3.73 (0.025)
Two youngest siblings:						
Boy–boy	0.261 (0.002)	1.101 (0.006)	0.556 (0.005)	0.263 (0.003)	1.246 (0.009)	3.74 (0.025)
Girl–girl	0.226 (0.002)	1.098 (0.007)	0.573 (0.005)	0.234 (0.003)	1.243 (0.010)	3.77 (0.025)
Girl–boy	0.256 (0.002)	1.094 (0.006)	0.548 (0.005)	0.253 (0.003)	1.249 (0.010)	3.76 (0.025)
Boy–girl	0.257 (0.002)	1.067 (0.006)	0.536 (0.005)	0.250 (0.003)	1.213 (0.009)	3.70 (0.025)

Source: Labour Force Surveys, 1990–2000, INSEE.
Interpretation: For the group of families with two or more children, the mean of the number of persons per room is 1.101 when the two oldest children are two boys.

room in large families,[14] while S_0 has a significant impact on the probability of having a third child. Our basic identifying assumptions is that it is the only channel through which the sex composition (as measured by S_1 and S_0) affects the probability of being held back a grade. A reduced-form regression confirms that the probability of being held back a grade is less important in large families where the two last children are (by descending age) a boy and a girl than in other families (Table 3, column 2). Comfortingly, the dummies describing the sex composition of the two youngest children have the same effects on overcrowding as on performance at school.

Within this framework, two basic strategies may be adopted to identify the effect of overcrowded housing. The first one focuses on large families and uses S_1 as an instrumental variable. The second one considers all families and uses jointly S_0 and S_1 to identify jointly the effect of family size and the effect of overcrowding. The first strategy relies on S_1 only to identify the effect of housing conditions, while the second one relies on S_0 and S_1. As discussed below, the two strategies provide similar results.

Angrist and Evans (1998) have already used the same-sex instrument (i.e., S_0) in order to identify the effect of family size on labor supply in the United States. They find that family size has a significant negative impact on the time spent at work by mothers and their IV estimates are smaller than their OLS estimates.[15] The downward bias which affects their OLS estimates suggests that—holding family size constant—the same-sex instrument is correlated with the time spent at work by mothers. Assuming that the time spent at work by mothers affects schooling outcomes, this finding is potentially

[14] When we work on the full sample of families with two children or more, we use a standardized version of S_1 (i.e., zero mean) interacted with a dummy indicating that the family is large (i.e., three children or more).

[15] From a statistical viewpoint, the difference between their OLS and IV impacts is not significant at standard level, however.

Table 3
First stage regressions: the impact of the differences in sex between the siblings on overcrowded housing and family size

	Families with three or more children		Families with two or more children		
	No. of persons per room	Being held back	No. of persons per room	No. of persons per room	(No. of children>2)
Sex composition of the two youngest siblings					
Boy–girl	−0.034 (0.001)	−0.024 (0.012)	−0.027 (0.010)	−0.023 (0.008)	0.002 (0.019)
Girl–boy	−0.001 (0.014)	−0.005 (0.013)	−0.003 (0.010)	−0.007 (0.008)	0.001 (0.014)
Girl–girl	−0.009 (0.015)	−0.011 (0.014)	−0.002 (0.012)	−0.013 (0.009)	0.001 (0.015)
Boy–boy	Ref.	Ref.	Ref.	Ref.	Ref.
(No. of children>2)	–	–		0.193 (0.002)	
Sex composition of the two oldest siblings					
Boy–girl	–	–	−0.017 (0.010)	–	−0.067 (0.011)
Girl–boy	–	–	−0.028 (0.010)	–	−0.079 (0.010)
Girl–girl	–	–	−0.010 (0.013)	–	−0.014 (0.013)
Boy–boy	–	–	Ref.	–	–
No. of observations	10,570	10,570	19,499	19,499	19,499

Source: Labour Force Surveys, 1990–2000, INSEE.
Field: Children who were born in $t-15$ and surveyed in t living in an intact family.
Note: All models include an intercept, 10-year dummies and a dummy indicating the sex of the child as supplementary independent variables. Models are estimated using OLS. When we work on the sample of families with two or more children (three last columns), the dummies indicating the sex composition of the two youngest children are standardized (zero mean) and interacted with a dummy indicating that the number of children is greater than 2.

problematic for us, especially when we rely on both S_0 and S_1 to identify the effect of overcrowding.

To further explore this issue, we have examined the links between our same-sex instrument and different measurement of mothers' labor supply. We found that—holding family size constant—the sex composition of the sibship has no significant effect on the probability of being out of the labor force nor on the number of hours spent at work by mothers. Raw statistics show that the proportion of mothers' out of the labor force is actually almost exactly the same in same-sex and mixed-gender families (about 45% in two children families and 73.5% in larger families). As it turns out, our basic instrumental variable is not correlated with this potential component of the residual.[16] In addition, we have checked that our basic IV estimates are unaffected by the introduction of a dummy

[16] The difference between the correlations observed in France and the United States plausibly reflects the institutional differences between the two countries. In the US case, the downward bias, which affects the OLS impact of family size on labor supply plausibly, reflects the fact that labor-supply and family-size decisions both depend on the ability to earn high wages. American women have no statutory paid parental leave and are guaranteed the right to return to employment following a birth for only a relatively short period of employment interruption compared to their European counterpart. In France, given the employment protection rules and the proportion of jobs available in the public sector, the family-size decisions are plausibly much less correlated with the ability to perform well on the labor market.

indicating that the mother is in the labor force as a supplementary control variable (the regressions are available on request).

The second set of instruments has been constructed from the available information on the mothers' and fathers' place of birth. French metropolitan area is divided into 96 elementary administrative subdivisions (*départements*). For each household, the survey provides us with the département where the different members of the household were born. Table 4 provides the corresponding first stage regressions. Our groups are defined so as to be uncorrelated with our measurement of father's socioeconomic status. We end up with four groups of départements for the mothers' place of birth and four groups for the fathers'. They confirm that significant differences in family size and housing conditions exist according to these variables. For instance, parents born in the Parisian region or in one of the large French cities tend ceteris paribus to live in more overcrowded housing than parents born in less urban areas. We interpret parents' place of birth as proxies for the housing conditions that parents' have experienced during their early childhood. We interpret the correlation between the parents' place of birth and current housing conditions as reflecting the fact that decisions on housing conditions are to some extent determined by early childhood experience. The identifying assumption is

Table 4
First stage regressions: the impact of parents' places of birth on overcrowding and family size

Independent variables	Dependent variables	
	No. of persons per room	(No. of children>2)
Father's place of birth		
Reference (Parisian suburbs)	Ref.	Ref.
Paris and large cities	−0.03 (0.01)	0.013 (0.015)
Regions P2	−0.08 (0.01)	−0.001 (0.014)
Regions P3	−0.03 (0.01)	−0.031 (0.011)
Mother's place of birth		
Reference (Parisian suburbs)	Ref.	Ref.
Paris and large cities	0.01 (0.01)	0.043 (0.015)
Regions M2	0.01 (0.01)	−0.001 (0.015)
Regions M3	−0.04 (0.01)	−0.034 (0.009)
No. of observations	14,599	14,599
R^2	0.02	0.01

Source: Labour Force Surveys, 1990–2000, INSEE.

Field: Children who were born in $t-15$ and surveyed in t living in an intact family such that the information on parents' place of birth is available.

Note: All models include an intercept, 10-year dummies and a dummy indicating the sex of the child as supplementary independent variables. All models are estimated using OLS.

Definition of the dummies indicating mother's place of birth: Reference regions correspond to departments 92, 91, 78, 93, 94, 95, 77, 06, 11, 14, 20, 24, 28, 32, 37, 41, 52, 53, 55, 58, 62, 76, 80, 83, 89, 97, 01, 29, 40, 43, 51, 54, 73, 82, 81, Paris and large cities correspond to departments 75, 59, 60, 13, Regions M2 correspond to departments 04, 15, 16, 23, 33, 35, 46, 61, 72, 74, 79, 84, Regions M3 correspond to the remaining departments.

Definition of the dummies indicating father's places of birth: Reference regions correspond to departments 92, 91, 78, 93, 94, 95, 77, 05, 06, 11, 24, 28, 47, 52, 53, 62, 68, 80, 89, 97, Paris and large cities correspond to departments 75, 59, 60, 13, Regions P2 correspond to departments 01, 07, 18, 25, 26, 29, 31, 42, 4, 47, 50, 51, 56, 85, and Regions P3 to remaining departments.

that this is the main channel through which parents' place of birth affects children's performances.

5.1. Estimation method

For the sake of simplicity, this paper relies on linear probability models only. In Appendix B, we build on Lewbel (2000) to state conditions under which such linear probability models provide consistent estimates for the structural parameters of binary response models.[17] One such condition is the existence of a continuous, exogenous and uniformly distributed explanatory variable (denoted v). Once such a regressor exists, the structural parameters of interest (here, α) may be estimated through the linear regression of the binary dependent variable on all the other regressors.

For our case, a natural candidate for v is the relative age of the adolescent in his/her age group, i.e., within the cohort of adolescents who were born the same year as he/she was. This variable is continuous and it is reasonable to assume that it satisfies the exogenous conditions introduced by Lewbel. As shown in Maurin (2002), this variable is definitely a factor of being held back a year at school: children born at the end of the year—the youngest in their age group—are clearly held back much more often than children born at the beginning of the year. Regarding our sample, about 48.5% of the 15-year-olds who were born during the last quarter of the year (i.e., October, November or December) have been held back a grade, a proportion that is about 11 percent points larger than it is for adolescents born during the first quarter of the year.

6. Results

Before moving on to the more sophisticated analysis, we will show our basic findings through a simple tabulation. More specifically, Table 5 shows that the probability of being held back a grade increases very significantly with the number of children per room, regardless of the size and the socioeconomic level of the families under consideration.

The median of the distribution of the number of persons per room is 1. About 59% of the adolescents living in families above this median have been held back a grade, a proportion that is more than 29 points higher than it is for adolescents living in families below the median. Generally speaking, there exist much more differences in the probability of being held back between overcrowded and nonovercrowded families than between large and small families.

6.1. Overcrowding and the probability of being held back: a causal analysis

The rest of this section discusses IV and OLS estimates of linear regression models relating the probability of being held back a grade to the number of persons per room. Table 6 focuses on families with three or more children. Model (1) shows the basic OLS

[17] Angrist (2001) provides other strategies for accommodating binary endogenous regressors.

Table 5
Fraction of 15-year-olds who have been held back a grade and probability of repeating a grade at age 15, by family size, father's socioeconomic status and number of persons per room

Types of families	Proportion of 15-year-olds who have been held back (in %)		Probability of repeating a grade at age 15 (in %)	
	More than one person per room	Less than one person per room	More than one person per room	Less than one person per room
Family size				
Two children	48.6 (1.6)	29.4 (0.7)	34.1 (3.0)	21.2 (1.0)
Three or more	60.4 (0.6)	29.8 (1.0)	37.7 (1.4)	20.4 (1.5)
Family income				
Low	63.9 (0.7)	41.9 (1.0)	41.6 (1.6)	27.1 (1.5)
High	46.8 (1.1)	23.3 (0.6)	26.9 (2.1)	17.3 (0.9)
All families	58.7 (0.6)	29.6 (0.6)	37.1 (1.3)	20.9 (0.8)

Source: Labour Force Surveys, 1990–2000, INSEE.
Field: Children who were born in $t-15$ and surveyed in t living in an intact family with two or more children.
Note: Relatively rich (poor) families are families which socioeconomic level is above (below) the median of the distribution. Relatively small (large) families are families with two (more than two) children. The value of 0.5 represents the median of the distribution of the number of children per room.
Interpretation: The probability of being held back is 63.9% in relatively poor and overcrowded families.

estimate ($\hat{\alpha}$=0.20). It is significant at standard levels and quite consistent with raw statistics. According to this OLS estimate, a one standard deviation (SD) increase in the number of persons per room yields an increase of about 18% of a SD of our measurement of educational outcomes. This result is valid under the assumption that errors in our measurement of the space available for each person are negligible and that no unmeasured factors simultaneously explain the number of persons per room and the probability of being held back at school. Model (2) corresponds to the reestimation of the OLS models when the number of persons per room is considered endogenous, and when its effect is identified using the sex of the two youngest siblings (as described by S_1) as instrumental

Table 6
The impact of the number of persons per room on the probability of being held back a grade: the case of large families

Independent variables	Dependent variable: to have been held back a grade			
	All children		Boys only	
	OLS (1)	IV (2)	OLS (3)	IV (4)
No. of persons per room	0.20 (0.01)	0.63 (0.39)	0.19 (0.01)	0.38 (0.15)
Male	0.11 (0.01)	0.11 (0.01)	–	–
No. of observations	10,570	10,570	5473	5473
R^2	0.06	0.02	0.04	0.01

Source: Labour Force Surveys, 1990–2000, INSEE.
Field: Children born in $t-15$, surveyed at t, living in an intact family with at least three children.
Note: The dependent variable corresponds to a dummy variable with value 1 when the child is behind at school. In models (2) and (4), the number of persons per room is assumed endogenous. The instruments are dummies indicating the sex composition of the two youngest children. All models include an intercept and a set of year dummies.

variable. This IV model provides an estimate ($\hat{\alpha}$=0.63) which is less well estimated (i.e., significant at the 9% level only), but larger than the OLS estimate.

As discussed above, there exists controversy about whether the sex composition of the siblings affects the educational attainment of girls (see, e.g., Butcher and Case, 1994, Hauser and Kuo, 1998). To address this issue, we have reestimated models (1) and (2) on the subsample of boys [Table 6, models (3) and (4)]. Comfortingly, we obtained qualitatively similar results. The IV effect ($\hat{\alpha}$=0.38) is not as large, but better estimated on the subsample of boys than on the full sample (i.e., significant at the 2% level).

Our IV estimates are about 0.5. How large these coefficients are? One way to answer this question is to compute the fraction of the educational performance gap between children in small and large families that can be explained (according to our IV estimates) by overcrowded housing per se. The average number of children per room is 0.91 in two children families and 1.24 in families with three children or more. Given this fact, the difference in the probability of being held back a grade between children in large and small families (i.e., 15 percent points) mostly reflect the impact of the differences in housing conditions [i.e., $0.5\times(1.24–0.91)$]. To put it differently, a 1 SD increase in the proportion of overcrowded housing (i.e., 0.45) would ceteris paribus lead to an increase of about 45% (i.e., $0.45\times0.5/0.5$) of an SD of the proportion of children behind at school.

Generally speaking, the downward biases that affect the OLS effect of overcrowding suggest that some unobserved factors simultaneously explain the number of persons per room and the performance at school. In addition, our data only provide an indirect and potentially rough measurement for the housing conditions that children have experienced during their early childhood. The downward biases that affect the OLS estimates may correspond to biases that arose from measurement errors.

To probe the robustness of our results, we have reestimated these models on the full sample of families with a measurement of family size as a supplementary explanatory factor (Table 7). Section (a) considers jointly boys and girls, while section (b) focuses on the subsample of boys. Models (1) and (2) show the basic OLS estimates. Comfortingly, they are very similar to the OLS estimates in Table 6, regardless of whether we control for family size or not. Model (3) shows the IV estimate when family size is assumed exogenous and when the effect of the number of persons per room is identified using dummies indicating the sex of the two youngest children.[18] Unsurprisingly, the estimate obtained with this instrument is very close to the IV estimate in Table 6. The source of identification is the same. Model (4) shows the IV estimates when the space per person and the size of the family are both assumed endogenous. We use dummies indicating whether the two oldest children are the same sex as supplementary instrumental variables. Interestingly, the IV impact of the number of persons per room remain large ($\hat{\alpha}$=0.60) and significantly different from zero at standard level. In addition, the effect of family size in model (4) is now very small and not significantly different from zero, which suggests that most of its positive OLS effect is due to its correlation with the space per person in the home and with the omitted determinant of performance at school.

[18] Given that we work on the full sample of families, the dummy is standardized (zero mean) and interacted with a dummy indicating that the family is large (i.e., three children or more).

Table 7
The impact of the number of persons per room on the probability of being held back a grade

Independent variables	Dependent variable: to have been held back a grade			
	OLS (1)	OLS (2)	IV (3)	IV (4)
Section (a): all children				
No. of persons per room	0.23 (0.01)	0.19 (0.01)	0.60 (0.35)	0.60 (0.36)
No. of siblings >2	–	0.08 (0.01)	−0.04 (0.11)	−0.04 (0.07)
Male	0.11 (0.01)	0.11 (0.01)	0.11 (0.01)	0.11 (0.01)
No. of observations	19,499	19,499	19,499	19,499
R^2	0.06	0.07	0.04	0.03
Section (b): boys only				
No. of persons per room	0.21 (0.01)	0.18 (0.01)	0.38 (0.15)	0.40 (0.14)
No. of siblings >2	–	0.09 (0.01)	0.02 (0.05)	0.01 (0.05)
No. of observations	10,079	10,079	10,079	10,079
R^2	0.04	0.05	0.03	0.01

Source: Labour Force Surveys, 1990–2000, INSEE.
Field: Children born in $t-15$, surveyed at t, living in an intact family with at least two children.
Note: The dependent variable corresponds to a dummy variable with value 1 when the child is behind at school. In model (3), the number of persons per room is assumed endogenous and its effect is identified using the same instrument as in Table 6, i.e., dummies indicating the sex composition of the youngest children. Given that we work on the full sample of adolescents, this dummy is standardized (zero mean) and interacted with a dummy indicating that the number of children is greater than 2. In model (4), the number of persons per room and the dummy with value 1 when the number of children is greater or equal to 3 are both assumed endogenous. We use dummies indicating whether the two oldest children are two boys or two girls as supplementary instrumental variables. All models include an intercept and a set of year dummies.

The section (b) of Table 7 shows the reestimation of models (1)–(4) on the subsample of boys. The results are very similar to the results obtained in section (a) (or in the two last columns of Table 6). The IV effect of overcrowding in the model (5) of section (b) is not as large as in section (a), but better estimated.

All in all, we end up with the same basic diagnosis when we use jointly S_1 and S_0 as instrumental variables as when we use S_0 only: the space per person has a very significant impact on performance at school.

6.2. An alternative set of instrumental variables

Table 8 proposes a reestimation of models (4) and (5) in Table 7 using a completely different set of instruments for identifying the impacts of family size and overcrowded housing. The new instruments are constructed from the available information on the fathers' and mothers' place of birth.

Most interestingly, the IV estimates obtained using these new instruments are significantly larger than the OLS estimates and very close to (and not statistically different from) the IV estimates obtained using the first set of instruments.[19]

[19] We have checked that the results are not affected by the introduction of the current region of residence as a supplementary control variable.

Table 8
The impact of the number of persons per room on the probability of being held back a grade: an alternative set of instruments

Independent variables	Dependent variable: to have been held back a grade	
	IV (1)	IV (2)
Section (a): all children		
No. of children per room	0.41 (0.13)	0.62 (0.16)
No. of children>2	0.03 (0.03)	−0.47 (0.16)
Male	0.12 (0.01)	0.12 (0.01)
No. of observations	14,599	14,599
R^2	0.04	0.02
Section (b):boys only		
No. of children per room	0.41 (0.17)	0.53 (0.20)
No. of children>2	0.03 (0.04)	−0.30 (0.20)
No. of observations	7568	7568
R^2	0.04	0.01
Section (c)		
No. of children per room	0.41 (0.11)	0.39 (0.12)
No. of children>2	0.01 (0.03)	−0.01 (0.03)
Male	0.12 (0.01)	0.12 (0.01)
Father's SES	−0.11 (0.01)	−0.12 (0.01)
No. of observations	19,499	19,499
R^2	0.11	0.10

Source: Labour Force Surveys, 1990–2000, INSEE.
Field: Children born in $t-15$, surveyed at t, living in an intact family with at least two children and such that the information on parents' place of birth is available.
Note: The dependent variable corresponds to a dummy variable with value 1 when the child is behind at school. In model (1), the number of persons per room is assumed endogenous. In model (2), the dummy with value 1 when the number of children is greater or equal to 3 and the number of persons per room is both assumed endogenous. In both cases, we use a set of dummies indicating parents' place of birth as instrumental variables. In section (c), we use a set of dummies indicating the sex composition of the two oldest and two youngest siblings (same as in Table 7) as supplementary instrumental variables. All models include an intercept and a set of year dummies.

The section (c) of Table 8 corresponds to the case where we introduce a measurement of family permanent income as a supplementary control variable. The measurement of parental income corresponds to the position of the father's occupation on the French Occupational Prestige Scale.[20] We have used the two sets of instruments jointly in order to improve the precision of our estimates.[21] The results remain very similar to those from section (a), which is consistent with the fact that our instruments are not significantly correlated with income.

[20] For more details on the construction of this variable and on its strong correlation with family income, see Chambaz et al. (1998) and Maurin (2002). In Goux and Maurin (2001), we provide additional regressions using more direct measurement for family income. In addition, we develop two-sample instrumental variable techniques (in the spirit of Angrist and Krueger, 1992, or Currie and Yelowitz, 2000) to better estimate its effect. With the results obtained being very close to the results in Table 5, we do not report these analyses.

[21] The results are very similar when we use separately the sex of the children (or the places of birth of parents) as instrumental variables, but the effects are less well estimated.

Table 9
An alternative dependent variable: the impact of overcrowded housing on the probability of repeating a grade

Independent variables	Dependent variable: to repeat a grade		
	OLS (1)	IV (2)	IV (3)
No. of children per room	0.17 (0.02)	0.28 (0.08)	0.20 (0.08)
No. of siblings>2	0.02 (0.01)	−0.01 (0.02)	−0.00 (0.02)
Male	0.06 (0.01)	0.06 (0.01)	0.06 (0.01)
Month of birth	−0.04 (0.02)	−0.04 (0.02)	−0.02 (0.02)
Father SES	–	–	−0.08 (0.01)
No. of observations	5793	5793	5793
R^2	0.02	0.01	0.05

Source: Labour Force Surveys, 1990–2000, INSEE.
Field: Children born in $t-15$, surveyed at t and t+1, on time at school and in an intact family with at least two children at t.
Note: The dependent variable is "to be in the same grade at t and t+1". Both family size and overcrowding are considered endogenous in models (2) and (3). The set of instruments corresponds to the reunion of the sets used in Tables 6 and 7 (i.e., dummies indicating the sex composition of the siblings and dummies indicating parents' places of birth).

6.3. An alternative dependent variable

The dependent variable analyzed in the previous subsections is whether a 15-year-old child has ever been held back a grade. This is a cumulative outcome, but the regressors are measured as of the survey date. This raises measurement error problems, which are perhaps only partially overcome by our IV estimation strategy. In this subsection, we consider the subsample of adolescents who are on time at t and for which we have information on their grade at t and t+1. We focus on the probability of repeating a grade at t+1, i.e., being in the same grade at t+1 and t. The interesting feature of this outcome is that it is noncumulative and measured after the regressors.

To begin with, the two last columns of Table 5 confirms that the probability of repeating a grade is significantly more important in families above the median of the distribution of the number of persons per room than in families below the median (i.e., 1). Table 9 goes a step further and presents an econometric evaluation of the causal effect of the number of children per room at t on the probability of repeating a grade[22] at t+1.

The subsample of 15-year-olds that are on time is not representative of the total population of 15-year-olds, and we have to control for the biases that could arise from endogenous selection. The simplest method is to introduce a supplementary control variable, which is correlated with the probability of being on time at 15, but uncorrelated with the current probability of moving up to the next grade.[23] To address this issue, we

[22] The minimum age for leaving school being 16, our dependent variable may slightly underestimate the actual proportion of adolescents who do not move up to the next grade. Generally speaking, it is because of this age limit that our paper focuses on 15-year-olds.

[23] An alternative strategy is to consider all adolescents surveyed at t and t+1, to add a dummy indicating whether they are on time at t as a supplementary explanatory variable and to use the date-of-birth within the year as an instrumental variable for identifying the effects of being on time at t. This strategy provides us with similar estimates of the overcrowding effect. The drawback of this approach is that it assumes that the effects of overcrowding are the same for children who have already been held back as they are for those who have not.

have used the date-of-birth within the year as a supplementary control variable. The underlying assumption is that the date-of-birth within the year affects the probability of repeating a grade at early stages in schooling only.[24]

The OLS specification in Table 9 confirms that adolescents who live in overcrowded homes are much less likely to move up to the next grade than other adolescents, even after controlling for family size. Model (2) proposes a reestimation of model (1) using jointly the sex of the siblings and the place of birth of parents as instrumental variables.[25] The IV overcrowding effect is significant at standard level and close to the OLS effect. Model (3) shows that the result still holds true when we control for family permanent income. According to these estimates, a 1 SD increase (i.e., 0.35) in the proportion of adolescents aged 15 living in an overcrowded housing would ceteris paribus lead to an increase of about 16% (i.e., $0.35\times0.2/0.44$) of an SD of the proportion of adolescents aged 15 who do not move up to the next grade. All in all, we come to the same conclusion regardless of whether we focus on a cumulative or noncumulative outcome. The space available for each child in the home is an important factor of performance at school at each stage of the schooling career.

7. Single room and diploma

To probe the robustness of our results, we have also used a retrospective survey conducted in 1997 by the French National Statistical Office in addition to the LFS. The sample of this retrospective survey consists of about 1000 individuals, representative of the French male population, aged 20–40. The respondents describe their schooling career as well as their housing conditions during childhood. Specifically, they indicate, (1) whether they dropped out of school before earning a diploma, and (2) whether they had their own room at the age of 11. The advantage of this survey is that it gives more direct information on respondents' housing conditions during their childhood and makes it possible to identify the potential long-term effects on educational achievement. The disadvantage of this survey is that it is much smaller than the Labor Force Surveys and does not allow for as precise an identification of the structural parameters. When we restrict the analysis to the individuals who had at least one brother or sister, the sample only contains a little over 600 individuals.

Table 10 presents the distribution of the respondents from the 1997 retrospective survey according to family size, year of birth, father's occupation and housing conditions during childhood. The table also describes the variations in the probability of leaving school without a diploma according to the same criteria. These simple tabulations confirm that the probability of dropping out from school without a diploma is greater for older generations

[24] As shown in Maurin (2002), the date of birth within the year has a very strong effect on the probability of being held back in primary school. In this section, we assume that this variable influences schooling transitions in primary school only.

[25] The estimated impact are larger when we use only the dummies indicating the sex composition of the siblings as instrumental variables, but the estimates are less well estimated (i.e., significant at the 10% level).

Table 10
The impact of overcrowding on the probability of dropping out of school without a diploma

	No. of observations	Fraction without diploma	Net effects
Family size:			
Three or more children	359	32.6	0.45 (0.19)
1 or 2	276	18.9	Ref
Date of birth:			
Born after 1964	357	23.5	0.76 (0.19)
Born before 1964	258	31.8	Ref
Father's occupation:			
Manual worker	279	33.8	0.54 (0.20)
Nonmanual worker	356	21.6	Ref
Overcrowding:			
Own room at 11	274	18.9	−0.58 (0.20)
No own room at 11	341	33.4	Ref

Source: Survey on Educational and Occupational Career, 1997, INSEE.
Note: The third column provides the results of a probit regression where the dependent variable is "To have dropped from school without a diploma" and where the independent variables are an intercept, a dummy for "own room at age 11", a dummy for "Three of more children", a dummy for "Father manual worker" and date of birth.

than for recent generations, for large families than for those with only one or two children, and finally, for blue-collar families than for white-collar families. The correlation is also very clear between the housing conditions during childhood and the probability of dropping out of school before earning a diploma. Close to 56% of the respondents said that they did not grow up having their own room. One-third of these individuals dropped out of school before earning a diploma, meaning a rate of academic failure twice that of other children.

A multivariate regression confirms that individuals who have their own room during childhood had ceteris paribus a much smaller probability than the others of dropping out of school before earning a diploma, even after controlling for the father's occupation and the number of siblings (Table 10, third column). Generally speaking, these supplementary investigations tend to confirm the diagnosis obtained using the Labor Force Surveys.[26]

8. Conclusion

Several results have come from our analysis. First, we found a very clear correlation between housing conditions during childhood and early performance at school. Specifically, the probability of being held back a grade in primary or junior high school increases very significantly with the number of persons per room in the home.

[26] We reestimated the effect of overcrowding using the sex composition of the siblings as instrumental variables. We obtained very imprecise results, which is not surprising given the small size of the sample.

The result holds true regardless of the size of the family or the socioeconomic status of the parents. Second, we found that the sex composition of the two oldest siblings influences the final size of the family while the sex composition of the youngest siblings in large families influences the subsequent housing conditions in these families. We build on these results to develop an econometric analysis of the effect of overcrowding on schooling outcomes using variables describing the sex composition of the siblings as instrumental variables. These different analyses reveal that the statistical relationships between housing conditions and academic failure are plausibly one of cause and effect. Altogether, we have an array of findings that indicate that public policy favoring the access of modest households to larger dwellings could have a substantial effect on educational outcomes. Further research is necessary to really define a housing policy that could affect the poorest children's school performance. This research must rely on in-depth analysis of the effects of existing public policies that favor the housing of low-income families.

Appendix A

We assume that parents make their decisions on the final size of the family (N_1) and on overcrowding (L) in two steps. At the beginning of the first period, they have N_0 children whose characteristics are S_0. N_0 and S_0 are assumed exogenous. During this first period, parents chose the final size of the family in light of S_0. During the second period, parents chose whether each child has his own room (L=1) in light of the final number and characteristics of their siblings (i.e., N_1 and S_1). Notice that, by construction, S_1=S_0 if N_1=N_0, and S_1=(S_0,s) otherwise, where s denotes the characteristics of the children born during the first period.

Let $U_2(C_2,L;\ S_1,N_1)$ denote parental utility during the second period, where C_2 represent family consumption. During the second period, parents maximize U_2 with respect to L and C_2 subject to the budget constraint: $C_2+qL+pN_1=R$, where p captures the cost of a supplementary child and q the cost of a larger house. Within this framework, we observe L=1, if and only if,

$$U_2(R-q-pN_1,1;S_1,N_1)>U_2(R-pN_1,0;S_1,N_1),$$

meaning if and only if the welfare gain due to L=1 is greater than the welfare loss due to the cost of increasing the number of rooms. Let us denote $L(S_1,N_1,R,q,p)$ the corresponding optimal dummy variable and $U_2^*(S_1,N_1,R,q,p)$ the corresponding indirect utility.

If $U_1(C_1,N_1;S_0,N_0)$ denotes the one-period flow of utility during the first period, then parents chose N_1 and C_1 in order to maximise the expected discounted flows of utility, $U_1(C_1,N_1;S_0,N_0)+\delta E_s[U_2^*(S_0,s,N_1,R,q,p)]$, where δ is the discount rate, subject to $C_2+pN_1=R$. The optimal number of children can be written $N_1(S_0,R,q,p)$. All in all, we end up with a model where family size depends on the characteristics of the oldest siblings while overcrowding depends on family size and on the characteristics of all the siblings.

Appendix B

Consider the following binary response model,

$$y = (\alpha_0 + \alpha_1 x + v + \varepsilon > 0)$$

where y is a binary dependent variable (typically "to have repeated a grade"), x a potentially endogenous regressor ("overcrowded housing"), v a continuous, exogenous regressor ("date of birth within the year") and ε an error term. α_0 is an intercept and α_1 the parameter of interest. The coefficient of v is normalized to 1. Within this framework, assume (1) that there exist an instrumental variable z which is correlated with x, but uncorrelated with the error term ε [i.e., $E(z\varepsilon)$=0] and (2) that v is large support {i.e., Supp(v)=[v_L,v_H]$\sqsupseteq$Supp($-x\beta$-ε)} and uniformly distributed conditional on x and z [i.e., $f_v(v|x,z)=1$], (3) the conditional distribution of ε does not depend on v [i.e., $f_\varepsilon(\varepsilon|x,z,v)=f_\varepsilon(\varepsilon|x,z)$].

This setting corresponds to a straightforward adaptation of the setting introduced by Lewbel (2000), and we know from Lewbel (2000) that α_1 can be consistently estimated though a simple linear regression of y on x using z as an instrumental variable. Specifically, we have,

$$E(y|x,z) = \int_{-\varepsilon|x,z} \int_{v_L}^{v_H} y f_\varepsilon(\varepsilon|x,z)\mathrm{d}\varepsilon\mathrm{d}v = \int_{-\varepsilon|x,z} f_\varepsilon(\varepsilon|x,z)\left(\int_{-(\alpha+x\beta+\varepsilon)}^{v_H} \mathrm{d}v\right)$$

$$= \int_{-\varepsilon|x,z} (v_H + \alpha_0 + \alpha_1 x + \varepsilon) f_\varepsilon(\varepsilon|x,z)\mathrm{d}\varepsilon = v_H + \alpha_0 + \alpha_1 x + E(\varepsilon|x,z),$$

which implies that

$$E(zy) = E(z(v_H + \alpha_0 + \alpha_1 x)$$

which proves that α_1 (and v_H+α_0) can be defined as the probability limit of the IV regression of y on x. Put differently, once there exists an exogenous, large-support, continuous regressor the parameters of the standard binary response model can be estimated through straightforward linear regression. In a recent contribution, Magnac and Maurin (2002) have shown that the large-support assumption can be replaced by a symmetry assumption on the tail of the conditional distribution of ε.

References

Angrist, J., 2001. Estimation of limited-dependent variable models with dummy endogenous regressors: simple strategies for empirical practice. Journal of Business and Economics Statistics 2001 (19), 2–16.

Angrist, J., Krueger, A., 1992. The effect of age at school entry on educational attainment: an application of instrumental variables with moments from two samples. Journal of the American Statistical Association 418, 328–336.

Angrist, J., Evans, W., 1998. Children and their parents' labor supply: evidence from exogenous variations in family size. American Economic Review 88, 450–477.

Barker, D.J., Coggon, D., Osmond, C., Whickham, C., 1990. Poor housing in childhood and high rates of stomach cancer in England and Wales. British Journal of Cancer 61 (4), 575–578.

Blau, D., 1999. The effect of income on child development. The Review of Economics and Statistics LXXXI (2), 261–277.

Britten, N., Davies, J.M., Colley, J.R., 1987. Early respiratory experience and subsequent cough and peak expiratory flow rate in 36 year old men and women. British Journal of Medicine 294, 1317–1320.

Butcher, K., Case, A., 1994. The effect of sibling sex composition on women's education and earnings. The Quarterly Journal of Economics 109 (3), 531–563.

Chambaz, C., Maurin, E., Torelli, C., 1998. L'évaluationsociale des professions en France. Revue Française de Sociologie 39 (1), 177–199.

Chombart de Lauwe, P.H., 1956. La vie quotidienne des familles ouvrières, Paris. CNRS.

Coggon, D., Barker, D.J., Cruddas, M., Olivier, R.H., 1991. Housing and appendicitis in Anglesey. Journal of Epidemiology and Community Health 45, 244–246.

Coggon, D., Barker, D.J., Inskip, H., Wield, G., 1993. Housing in early life and later mortality. Journal of Epidemiology and Community Health 47, 345–348.

Conley, D., 2000. Sibship sex composition: effects on educational attainment. Social Science Research (29), 441–457.

Currie, J., Yelowitz, A., 2000. Are public housing projects good for kids? Journal of Public Economics 75, 99–124.

Deadman, D.J., Gunnel, D., Smith, G.D., Frankel, D., 2001. Childhood housing conditions and later mortality in the Boy Orr cohort. Journal of Epidemiology and Community Health 55, 10–15.

Fuller, T.D., Edwards, J.N., Sermsri, S., Vorakitphokatorn, S., 1993. Housing, stress, and physical well-being: evidence from Thailand. Social Science Medicine 36 (11), 1417–1428.

Goux, D., Maurin, E., 2001. The impact of overcrowded housing on children's performances at school. CEPR Discussion Papers, vol. 3818.

Gove, W., Hughes, M., Galle, O., 1979. Overcrowding in the home: an empirical investigation of its possible pathological consequences. American Sociological Review 44, 59–80.

Hauser, R.M., Kuo, H.H.O., 1998. Does the gender composition of sibship affects women's educational attainment? Journal of Human Resources (33), 644–657.

Kaestner, R., 1997. Are brothers really better? Sibling sex composition and educational achievement revisited. Journal of Human Resources 32 (2), 250–284.

Lewbel, A., 2000. Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables. Journal of Econometrics 97, 145–177.

Magnac, T., Maurin, E., 2002. Identification and information in monotone binary models. Unpublished Manuscript, CREST, Malakoff.

Mann, S.L., Wadsworth, M.E., Colley, J.R., 1992. Accumulation of factors influencing respiratory illness in members of a national birth cohort and their offspring. Journal of Epidemiology and Community Health 46, 286–292.

Maurin, E., 2002. The impact of parental income on early schooling transitions: a reexamination using data over three generations. Journal of Public Economics 85, 301–332.

Mayer, S., 1997. What Money Can't Buy: Family Income and Children's Life Chances. Harvard University Press, Cambridge MA.

Murat, F., Rocher, T., 2003. La Place du Projet Professionnel dans les Inégalités de Réussite Scolaire à 15 ans, in France Portrait Social, 2002–2003. INSEE, Paris, pp. 101–121.

Prescott, E., Vestbo, J., 1999. Socioeconomic status and chronic obstructive pulmonary disease. Thorax 54, 737–741.

Rasmussen, F.V., Borchsenius, L., Winslow, J.B., Ostergaard, E.R., 1978. Associations between housing conditions, smoking habits and ventilatory lung function in men with clean jobs. Scandinavian Journal of Respiratory Disease 59, 264–276.

Rosenzweig, M., Wolpin, K., 2000. Natural Experiment in Economics. Journal of Economics Literature 38 (4), 827–874.

Shea, J., 2000. Does parents' money matter? Journal of Public Economics 77, 155–184.