New Evidence on the Link between Housing Environment and Children's Educational Attainments: The Case of Taiwan[+]

Hsien-Ming Lien[*], Wen-Chieh Wu[**], Chu-Chia Lin[***]

Abstract

There is an extensive literature that posits the hypothesis that a better housing environment enhances a child's educational attainments. However, there is little causal evidence demonstrating the presence of this effect. Using the census files covering the entire population of Taiwan, we examine the effect of housing environment on children's educational attainments. Because the Taiwan census data contain unique address information for every household, we are able to control for unobserved family heterogeneity by comparing a child with his or her peers of the same age cohort in the same neighborhood. After controlling for neighborhood using tens of thousands of area dummies, the chance of high school enrollment for teens (ages 16 and 17) and college enrollment for young adults (ages 19 and 20) is found to be positively correlated with increases in floor space, increases in residential stability, and ownership status, but negatively correlated with increases in building age. In addition, we found that the effect of a child's private space on the chance of school enrollment is nonlinear and different across age and gender. The results are robust even when we account for the potential endogeneity between sibship size and educational outcome using the instrumental variable method.

Keywords: quantity–quality trade-off, housing, educational attainment

JEL classification: R0, I2

[#] We thank David H. Autor, Charles Leung, and Jin-Tan Liu for their comments and suggestions. We are grateful to Directorate General of Budget, Accounting, and Statistics for providing the census data. Supports from National Science Council (NSC-93-2415-H-004-013) and National Health Research Institute (NHRI-EX93-9204PP) for Hsienming Lien are greatly appreciated. The usual disclaimer applies.
[*] Department of Public Finance, National Cheng-Chi University, 64 Sec 2, Zhi-Nan Rd, Taipei, Taiwan; email: hmlien@nccu.edu.tw.
[**] Corresponding author, Department of Public Finance, National Cheng-Chi University, 64 Sec 2, Zhi-Nan Rd, Taipei, Taiwan; email:jackwu@nccu.edu.tw.
[***] Department of Economics, National Cheng-Chi University, 64 Sec 2, Zhi-Nan Rd, Taipei, Taiwan; email: nccut001@nccu.edu.tw.

1. Introduction

One long-standing area of interest in the social sciences is to understand the connection between the family environment and a child's outcome, particularly educational attainments. It is generally believed that a larger family size may negatively affect a child's outcome through resource dilution [e.g., Blake (1981, 1989)]. The best-known economic theory that links family circumstances and a child's educational outcome is perhaps the quantity–quality trade-off model [Becker and Lewis (1973) and Becker and Tomes (1976)]. This theory claims that as parents become richer because of the interaction between quantity and quality in the budget constraint, they demand higher quality children, but not necessarily more children. Thus, a reduction in family size leads to an improvement in a child's schooling.

The majority of early studies confirm this trade-off relationship, with a negative relationship between family size and a child's educational attainments being widely observed in regression results.[1] While this negative correlation is often interpreted as evidence supporting the quantity–quality trade-off theory, the conclusions are facing serious criticism. Most problematic is that the apparent negative relationship is not necessarily indicative of a causal effect. That children raised in larger families have less schooling than those in smaller families is not necessarily because of the sibship size per se, but may reflect the omission of other unobserved characteristics, such as parental preferences, household resources, neighborhood conditions, and quality of schooling. In light of this potential bias, several studies have sought to uncover the *causal* effect of family size on a child's educational outcome using the instrumental variable method (IV) [e.g., Angrist, Lavy, and Schlosser (2005), Caceres (2004), and Conley and Glauber (2005)], or controlling for family fixed effects [e.g., Guo and VanWey (1999) and Black, Devereux, and Salvanes (2005)]. Notably, these studies generally found the coefficient of sibship size becomes insignificant after controlling for unobserved family characteristics.

[1] For a review in the economic literature on the link between family size and children's outcomes, see Schultz (2005). There have also been numerous discussions on this issue in sociology. For details, see Blake (1981, 1989), Powell and Stellman (1993), and Guo and VanWey (1999).

Why are results of OLS estimation so different from those of IV or the family fixed effects model? One likely explanation, as pointed out by Phillips (1999), is that sibship size does not produce a negative impact on a child's educational outcome, but the type of family resources it dilutes does.[2] Furthermore, Goux and Maurin (2005) investigated the effect of household crowdedness on a child's school performance, one key resource likely to affect a child's education. Using exogenous variations of family size and household crowdedness as instruments, they found the importance of sibship size becomes negligible under IV estimation, but the private space each child has is negatively associated with a child's educational attainments. In other words, children in large families perform less well not because of their family size, but because of the smaller private space each child has available to them.

In the same spirit as Goux and Maurin (2005), this paper seeks to explore the underlying relationship between the housing environment and a child's educational attainments. Unlike Goux and Maurin (2005), which controls for unobserved family heterogeneity using instruments, we overcome this difficulty by comparing a child with his or her peers of the same age in the same very small neighborhood: "lin," the smallest government jurisdiction in Taiwan that usually covers less than 0.1 square kilometer. Families residing in the same lin often share similar housing preferences and family incomes. In addition, youths raised in the same lin generally have experienced the same neighborhood effect. Furthermore, under the current regulation, children in the same lin typically attend the same school for compulsory education. Thus, by comparing youths with peers of the same age in the same lin, we control to some extent for unobserved family heterogeneity such as parental preferences, earning potential, neighborhood conditions, and, most importantly, quality of compulsory schooling.

Our data are derived from the census files that cover the entire Taiwanese population, more than 22 million, in the year 2000. The census data not only record detailed family and housing information, but also include *unique* address information for every household. The large sample size, together with detailed address information, allows us to examine the chances of high school enrollment for teens (ages 16 and 17) and college enrollment for young adults (ages 19 and 20) while controlling for family

[2] Black, Devereux, and Salvanes (2005) offer a different explanation: family size itself might have little impact on the quality of every child, but more likely impacts the marginal child through the effect of birth order. In their results, children of higher birth orders are likely to have worse educational attainments.

heterogeneity. After including tens of thousands of area dummies, our results confirm the importance of housing environment in determining a child's educational attainments. Specifically, our estimates show that youths' educational attainment is positively associated with an increase in housing floor space, an increase in residential stability, and ownership status, but negatively related to an increase of building age. The results continue to hold even accounting for the endogeneity between sibship size and a child's education using twin births or sex-composition as instruments.

An important difference between our study and Goux and Maurin (2005) is that we include a wide range of housing variables. Aside from each child's private space, this study also considers a house's floor space, building age, residential stability, and ownership status as various determinants of housing environment. Therefore, the analysis is able to provide a more complete picture about the impact of housing on a child's education. Another key difference is that we obtain a different effect of household crowdedness. While Goux and Maurin (2005) found that a reduction in a child's private space resulted in a negative effect on his or her schooling, our estimates indicate that the effect may be nonlinear: conditional on a household's size, reducing each child's private space does not always lead to an decrease in the chance of school enrollment. Moreover, this crowdedness effect is likely to differ according to the child's gender and age.

Our paper also relates to another line of literature exploring the effect of housing variables on children's outcomes, including tenure status [e.g., Green and White (1997), Boehm and Schlottmann (1999), Aaronson (2000), and Haurin, Parcel, and Haurin (2002)] and residential mobility [e.g., Lee, Oropesa, and Kanan (1994), Green and White (1997), Aaronson (2000), and Haurin, Parcel, and Haurin (2002)].[3] Although some studies have demonstrated the importance of housing environment, few of them controlled for the endogeneity problem caused by various housing variables.[4] To our knowledge, this paper is the first study that investigates the effect on a child's educational attainments of a wide range of housing variables.

---

[3] For a complete review on the tenure status literature, see Haurin, Dietz, and Weinberg (2003).

[4] A number of studies have attempted to control for the endogeneity of housing variables. For instance, Green and White (1997) adopted the bivariate probit model to solve the selection bias problem between tenure decision and schooling, but found no evidence of it. Haurin, Parcel, and Haurin (2002) used a treatment effect model to reduce selection bias. Aaronson (2000) dealt with the endogeneity problem of home ownership and mobility.

The paper is organized as follows. In the next section, we outline the estimation problem and discuss the existing identification strategies as well as our strategies. Section 3 describes the data source, sample selection, and measures of educational attainments, along with the basic statistics of our sample. Section 4 shows results of the basic specification, the effect of area dummies, as well as comparisons with IV estimates. Section 5 offers concluding remarks.

2. Conceptual Framework

A. Parameter of Interest

Let:

(1) $$edu_i = X_i\alpha + \beta N_i + v_i + \varepsilon_i,$$

where edu_i is the child's educational attainments, X_i is a vector of observed characteristics of the child and his or her family (e.g., age, sex, birth order, and father's and mother's education and working status), N_i is a variable of child *i*'s sibship size, v_i is the family-specific unobserved determinant (e.g., parental preferences or quality of schooling), and ε_i represents the idiosyncratic shock that is assumed to be independent across other factors.

The central parameter of interest is β, which is viewed as a measure of the trade-off between quantity and quality of children. Early studies primarily found this coefficient to be negative in OLS estimation and therefore inferred that substantial quality improvements can be gained by controlling for family size. However, the regression results are likely to be confounded by the existing observed factors (e.g., parental education) as well as the unobserved determinants (e.g., quality of schooling). The omitted variable formula suggests that the OLS coefficient from the regression is:

(2) $$\beta_{OLS} = \beta + \frac{\mathrm{cov}(N_i, X_i\alpha)}{\mathrm{var}(N_i)} + \frac{\mathrm{cov}(N_i, v_i)}{\mathrm{var}(N_i)}.$$

Therefore, even if children raised in larger families have less schooling than those in smaller families, the strength of the relationship could be driven by the correlation between sibship size and other observed and unobserved factors, not necessarily the quantity–quality trade-off.

B. Existing Identification Strategy

In light of the potential bias, the existing literature has adopted several methods to uncover the underlying relationship between a child's education and sibship size. Early studies attempted to account for this potential bias by including more controls, such as parental IQ, and better measures of household income. However, adding more controls cannot rule out the possibility of an association between family size, educational attainment, and something immeasurable, such as housing environment, neighborhood conditions, or quality of schooling. As a result, recent studies have taken different approaches to account for unobserved family heterogeneity. For instance, Guo and VanWey (1999) and Black, Devereux, and Salvanes (2005) include the household's dummies, i.e., family fixed effects, to control for the unobserved family-level heterogeneity. Angrist, Lavy, and Schlosser (2005), Caceres (2004), and Conley and Glauber (2005) employ exogenous variations in family size, such as multiple births or preferences of a mixed sibling-sex composition, as instruments to investigate the causal effect of family size on a child's education. Notably, studies using IV estimations or fixed family effects found weaker correlations between family size and a child's education, many of which turn out to be negligible.

The inconsistency of results between OLS and other estimation methods cast doubts over the link between family size and a child's outcome. One likely explanation, as pointed out by Phillips (1999), is that sibship size per se does not affect the child's educational attainments, but the type of resources it dilutes does. Goux and Maurin (2005) extended this line of thought by exploring the impact of a child's private space, one important kind of resource likely to be affected by additional children, on the child's schooling. Specifically, they considered the following equation:

$$(3) \qquad edu_i = X_i\alpha + \beta N_i + \gamma h_i + \nu_i + \varepsilon_i,$$

where h_i is the average number of rooms per person in the household, used as a proxy for a child's private space. Notice that equation (3) also includes the sibship size variable to account for the effect caused by family size. Because sibship size and the child's private space are likely to be endogenous, they employ two instruments, gender of the first two children and of the last two children, respectively, to control for unobserved family heterogeneity. Consistent with previous studies, they found that the coefficient of sibship size becomes insignificant under IV estimation. Interestingly, the coefficient of the average number of rooms per person in IV estimates is significantly negative, suggesting that children in large families perform less well, not because of their family size, but because of the smaller private space available to each child.

C. Our Identification Strategy

In contrast to Goux and Maurin (2005), our study seeks to identify the effect of a variety of housing variables on a child's educational outcome, such that:

$$(4) \qquad edu_i = X_i\alpha + \beta N_i + H_i\gamma + \nu_i + \varepsilon_i.$$

The biggest difference between (3) and (4) is that the housing environment is now a vector of multiple variables (H_i) instead of a single variable (h_i). There are substantial difficulties in using existing identification strategies for this specification. Because these housing variables do not change within a household, including household dummies essentially eliminates the effect of housing environment. Another possible strategy is to find instruments for housing variables, as Goux and Maurin (2005) did for household crowdedness. Nevertheless, controlling for the unobserved heterogeneity in this setting requires us to find many more instruments.

We take a different approach to identify the causal link. Apart from including a detailed set of important variables of a child's family background used in previous studies (e.g., a child's birth order, parental age, work status, and education), we account for unobserved family heterogeneity by adding dummies of the child's residential neighborhood. Our unique data are derived from the census data that collects information on the entire Taiwanese population, with detailed address information.

Therefore, we are able to compare a child with his or her peers of the same age in the same very small neighborhood, the lin. Families residing in the same lin tend to share similar housing preferences and parental incomes, as well as earning potentials. Moreover, youths raised in the same lin generally encounter the same neighborhood effect. Furthermore, youths in the same lin typically attend the same elementary or junior high schools, allowing us to control for the quality of compulsory schooling prior to high school or college. In fact, given Taiwan's current school regulation, it is almost certain that youths in the same lin go to the same school.[5,6] Thus, by controlling for neighborhood fixed effects, we account for the neighborhood effect, quality of schooling, and parental incomes and preferences. Nevertheless, it is still possible that our approach may not fully capture unobserved family heterogeneity. We will discuss this point in the results section.

To be more specific, we estimate the following equation:

$$(5) \qquad edu_i = X_i\alpha + \beta N_i + \gamma H_i + Area_i + \varepsilon_i,$$

where edu_i is a dummy equal to one if child *i*'s highest educational attainment is general high school for teens or general college for young adults, and zero otherwise; H_i is a set of variables on the housing environment, including building age, tenure status, household crowdedness, and residential stability; $Area_i$ is a vector of neighborhood dummies to control for unobserved family heterogeneity; and ε_i is an independent error across various individuals. As discussed earlier, we compare youths residing in the same lin. In Taiwan, the lin is the fourth and smallest level of government jurisdiction, following county, town, and village. As such, the estimation includes tens of thousands of area

---

[5] According to Taiwan Compulsory Education Law, students residing in the same "lin" belong to the same public school district and thus are assigned to the same public elementary or junior high school. For instance, the school district for Beitu Elementary School in Taipei includes every Lin of Central and Da-Tong Villages, 1st–9th and 12th Lin of Chang-An, 2nd Lin of Hot-Spring Village, and 1st–10th Lin of Ching-Jiang Village. For details on the regulations, see http://www.tp.edu.tw/neighbor/elementary/e_beitu.jsp.

[6] One exception is that children hoping to enroll in better elementary or junior high schools may move their registries to relatives or friends residing in better school districts, but continue to live with their parents. In this case, those children are coded as "other relatives" in the households of friends or relatives in the census. Because our data remove children that coreside with other relatives, we expect this proportion to be small in our sample.

dummies. Because of computational complexity, we focus on the linear probability model instead of nonlinear models. Alternative models (e.g., probit and logit), however, yield similar results.

3. Data and Sample

A. Data Source

The data for this study are derived from the 2000 Taiwan census, conducted every 10 years by the Directorate of General Budgeting, Accounting, and Statistics. The Taiwan census files collect information using a detailed questionnaire similar to that used to create the PUMS files for the US censuses (long-form), except that income-related variables are excluded. At each household, the interviewer records each individual's basic demographics (race, sex, age, and marital status), educational attainment, relationship with the head of household, working and employment status within the past two weeks, as well as the industry in which he or she works. In addition, the interviewer records the residence's structure (number of living rooms, bedrooms, kitchens, and bathrooms), tenure status (rent or own), years lived in the residence, and the location from which the family last moved. The residence information is further linked with the housing registry from the Ministry of Interior to ascertain the floor space of the house, the building year, and the major construction material used for the residence. More importantly, the Taiwan census includes a scrambled, but unique, address for every household's residence. As seen below, this unique address information plays an essential role in the analysis.

The advantage of using the Taiwan census is that the files contain the full sample of Taiwan residents, around 22 million in total or 300,000 individuals in most age cohorts. The large sample size, together with the detailed address information, provides a good opportunity to analyze the effect on educational attainment of the housing environment. Ideally, we would examine the link using the final education levels of all family adult respondents and their current housing information. In practice, however, this is not possible because the census files do not record family information of those who no longer reside with their parents and siblings. Obtaining the complete family background is therefore difficult, especially for adult respondents because a large portion of them do not coreside with parents and siblings. Moreover, the census files report only the respondent's relationship with the head of the household, but not with other members. Although we could match their relationships according to each member's age and gender, the identification becomes quite complicated when there are more than three adults in a household (e.g., coresiding with a brother or sister-in-law).

B. Sample Selection

For the purposes of this study, we restrict the sample in several ways. We select households with at least one unmarried child aged between 15 and 20 at the time of the census, of which the eldest sibling is no older than 22. We focus on the younger sample to reduce the bias resulting from incomplete family information. We restrict the sample to ages over 15 because compulsory schooling in Taiwan ends at junior high school (9^{th} grade). To avoid mistakes arising from matching parents, we keep only nuclear families in the sample, eliminating households that live with grandparents, relatives, or other friends. Furthermore, we drop households in which children are raised by a single parent to reduce complications because different family structures may also affect a child's education. Finally, we include only samples that have stayed in the residence for at least three years because the housing effect usually takes a longer time to materialize.

To demonstrate the impact of exclusion criteria, Table 1 lists the observed number of youths aged from 15 to 20 for each selection criteria. The first column lists the total number of youths in the census by age cohort. As indicated by these numbers, the number of respondents peaks at the age of 19 and then gradually declines as their age rises; this pattern is consistent with the number of births between 1980 and 1985 (ages 15 to 20 in 2000) in Taiwan.[7] The vast majority of youths, particularly younger ones, coreside with their families. This can be seen from the difference between the first and the second columns, which shows the number of youths who live with at least one adult aged 35 or older. Nevertheless, more and more youths, especially those older than 20 years, choose to live alone for either marriage or work reasons. That youths live alone for other reasons may increase the risks of matching complete family information, a point we will return to later.

---

[7] The number of respondents obtained from the census data is very close to the birth numbers between 1970 and 1975; the difference is less than 3 percent in every age cohort.

The largest reduction in sample size occurs when restricting the sample to nuclear families. This is not surprising because about 67% of the elderly in Taiwan coreside with their children.[8] Among these nuclear families, around 20% of the youth do not have valid parental information: either they are growing up in single-parent families (around 60% are single mothers) or are no longer coresiding with both parents. Another 10–20% are removed because of the age restriction of the eldest sibling; the older the respondent, the more likely they are to be removed by this age constraint. Finally, around 7% are eliminated because they have stayed in the current residence for less than three years. The final sample size consists of a little over one third of the original sample. Still, we have around 100,000 respondents in each age cohort.

C. Measure of Educational Attainment

Before describing our analysis sample, it is important to first discuss our measures of educational attainment. Previously used measures include the highest completed level of education [Boehm and Schlottmann (1999), Angrist, Lavy and Schlosser (2005), Black, Devereux and Salvanes (2005)], private school attendance [Conley and Glauber (2005)], held back in school grade [Conley and Glauber (2005), Goux and Maurin (2005)], test scores [Guo and VanWey (1999), Haurin, Parcel, and Haurin (2002)], dropping out [Green and White (1997)], and graduating from school by a certain age [Aaronson (2000)]. Because our data are derived from the census files, we cannot make distinctions between the quality of the youth's school (e.g., school ranking), or the youth's academic performance within the school. Therefore, we adopt a measure similar to the one used in Conley and Glauber (2005) that compares the respondent's age with the highest schooling that he or she is currently enrolled in or has completed so far. The education system in Taiwan is similar to that of the United States, except that compulsory schooling is nine instead of 12 years. Therefore, from the age of six, children are required to take six years of elementary school and three years of junior high school. After finishing junior high school, those seeking additional education can go to senior high school (three years) and even higher after graduating from high school. Suppose a child of age 16 reports his or her highest

---

[8] According to the Taiwan Elderly Survey in the year 2000, 67.3% of adults aged over 65 coreside with their children.

schooling is junior high school. Then he or she either did not proceed to higher education or had been held back a grade in previous school years. By examining one's age and highest schooling, we can compare a child's educational attainments with those of peers in the same age cohort.

There are, however, two complications with this measure. First is that the cut-off birthday for school admission may result in some children starting school late.[9] For instance, a September-born child may be almost one year older than a child born the following August but they are in the same school grade. Because the census data only record an individual's age (in years) at the time of the census interview, we are unable to determine whether a child meets the full age requirement at the time of school enrollment. Thus, some 15-year-old children may already be in senior high school, while others are still in junior high school.[10] Second, there are two types of senior high schools (general versus vocational) and colleges (general versus junior) in Taiwan. Although the quality difference between various types of schools is small in some countries, the gap is large here because students are enrolled into schools based on their test scores on school entrance exams. Generally, general high schools are more difficult to enter, as are general colleges.[11]To resolve these difficulties, we first restrict the sample to youths of ages 16 and 17, and ages 19 and 20. Youths aged 15 and 18 are removed because their educational measures are harder to define. Next, we check if the respondent's reported schooling matches the highest schooling of his or her age. More specifically, we examine if youths of ages 16 or 17 attended general high schools (nonvocational), and whether youths of ages 19 or 20 attend general colleges (nonjunior). In the discussion that follows, we refer to the younger sample as the "teen" sample and the older sample as the "young adult" sample.

D. Description of Analysis Sample

We work with two analysis samples, both described in Table 2. To demonstrate the effect of our sample selection criteria, we continue to present sample statistics by age cohorts. In total, there are

---

[9] The cut-off birthday in Taiwan is similar to that of the United States: children must be six years old (full) by September 1st to be enrolled in the school.

[10] The 2000 Taiwan census is conducted at the end of that year. Therefore, roughly half of all 15-year-old children are in junior high school, while the rest are in senior high school.

[11] For instance, the minimum score for entering a public high school in Taipei in 2004 was 220 points, about 30 points higher than that of public vocational schools. Likewise, the minimum score for entering general college is considerably higher than that of junior college in Taiwan.

283,959 teens and 188,937 young adults because more young adults are removed during the selection process. In both samples, except for youths aged 20, we have more males than females, reflecting the special gender preference in Taiwan.[12] Because of the sibling's age restriction, a higher proportion of first-borns are observed in young adults than teens. No significant difference, however, is observed in the average number of siblings among different age cohorts.

The educational attainment of youths is listed in the first set of rows of Table 2. A little over half of teens were enrolled in general high schools at the time of the census; 35% were in vocational high schools, while the remaining teens were out of school. The variation in schooling among young adults is larger. About 40–50% of young adults continued schooling after high school (e.g., general or junior colleges), while another 40–50% chose to stop after general or vocational high schools. Only 5–10% of young adults stopped their education after compulsory schooling.

One concern with our educational measure is whether the cut-off birthday affects schooling. If that is the case, we should observe a large discrepancy in schooling between two consecutive ages. Table 2 provides some evidence regarding this concern. For teens, there are only limited schooling differences between ages. In fact, the proportion of those attending general high school for 17-year-old youths is actually lower than that of 16 year olds, showing that the cut-off is not a concern for teens.

The schooling comparison among young adults is a little bit complicated. Our data for a child's education show a rising trend of schooling between the two age cohorts. For instance, the proportion of youths attending general college increased from 17% to 25%, and attending junior college increased from 22% to 30%. Nevertheless, this observation seems unlikely to be because of the cut-off birthday because the number of young adults in each age cohort enrolled in general and junior colleges remains almost the same.[13] Instead, the increase in schooling reflects the fact that those who did not seek higher education left home for work. Because our sample removes youths that live alone, young adults

---

[12] The observation that there are more 20-year-old females than males is likely to reflect the fact that males are more likely to work away from home. As a result, the category of youths aged 20 that coreside with parents is dominated by females.

[13] The number of youths enrolled in general and junior colleges is 19,056 and 24,400 for youths of age 19, and 19,736 and 23,573 of age 20.

that live with their family at the age of 20 tended to enroll in higher education. In other words, the rising schooling trend is primarily because of our selection criteria, a point that we return to later.

Table 2 also reports variables describing the parental background of the youths, including age, education, and work status. The average parental age of young adults is two years older than that of teens, reflecting the age difference between teens and young adults. In both the teen and the young adult samples, mothers are less likely to have acquired higher levels of education than fathers, especially for colleges or above. Likewise, the difference in working status between fathers and mothers is quite large. Over 90% of fathers in both samples hold a full-time job, while only around 60% of mothers do. Nevertheless, in some families mothers shoulder more economic burden than fathers, with about 10% of the sample being female-headed households.

The Taiwan census data include a wide range of descriptions of housing environment, including floor space, number of rooms, age of building, tenure status, and the location from which the family last migrated. The floor space of the house is measured by square meter. On average, the typical respondent lives in a building 10 to 20 years old, with 3.5 rooms, and 130 square meters. To better account for overcrowdedness, we construct three dummies that compare the number of bedrooms in a house with the number of children in a family. Typically, parents share a bedroom, so the comparison is based on the remaining bedrooms (minus the parents' bedroom) and the number of children. A household is considered as having high crowdedness if some children share a room, medium crowdedness if every child has his or her own room, and low crowdedness if every child has more than one room. By this standard, more than 60% of respondents live in a house with medium crowdedness; the rest reside in households with limited private space. These rates remain almost unchanged with respect to the teen or young adult sample.

More than 90% of youths live in self-owned households, reflecting the high rate of owner-occupied houses in Taiwan. In most cases, the youths in the sample have been at the same residence for more than 10 years; less than 14% of youths moved into the current residence within the last five years, of which around 3% moved within the local vicinity (within the same village); the rest migrated from other villages.

E. Area Dummies and Family Heterogeneity

Before showing the estimation results, it is useful to first describe the area dummies, which aim to control for unobserved family heterogeneity. Because the census data record detailed address information, area dummies can be constructed from the highest level (county) to the lowest (lin). For instance, Taipei, the capital of Taiwan, consists of 12 towns, 435 villages, and 9741 lins. The average number of square kilometers of a town, village, and lin in Taipei are 22.6, 0.624, and 0.028, respectively.[14] Not surprisingly, as seen in Table 3, the sample number in an area drops sharply as the level of government jurisdiction moves from towns to lins. While there are, on average, 780 teens and 520 young adults in a town, each lin accommodates only 3.1 teens and 2.4 young adults. From the percentile distribution based on lin, at least half of lins have only one teen and one young adult at the time of the census. Despite this, there is still a great deal of variation in many other lins in the sample. This can be seen from the numbers in parentheses, showing that the number of teens or young adults at the first quartile, based on the whole sample, is 3 and 2, respectively.

If area dummies are good controls for family heterogeneity, we should observe that the extent of variation within a neighborhood declines when a smaller neighborhood is used. To demonstrate the relationship between family heterogeneity and area dummies, Table 4 shows "within" and "between" standard deviations (SD) of housing environment variables. Because these SDs may exhibit different patterns in cities and rural regions, we further separate our sample into two groups based on the number of residents in the town: large towns (more than 100,000 residents) and small towns (less than 100,000 residents). For the purpose of exposition, we only list these numbers at the village and lin level. Consistent with our expectation, "between" SD rises and becomes larger than "within" SD for the vast majority of housing variables as the neighborhood level moves from village to lin. Nevertheless, we do not find a clear difference in SD between large and small towns in the sample.

[14] The total size of Taipei is 271.8 square kilometers.

4. Empirical Analysis

A. Basic Specification

Results of our basic specification using the two analysis samples are presented in Table 5. Our basic specification estimates the linear probability model including all variables displayed in Table 2, except the father's age because father's age is highly correlated with mother's age. Because estimates from the teen and young adult samples are similar, we first discuss the similarities in the two results, and then discuss the differences.

As is typical for these types of regressions, our coefficients imply strong links between a child's education and parental schooling. The better the parental schooling, the more likely it is that youths will seek higher education. As discussed earlier, the findings could reflect parental preferences over a child's education or a child's generic abilities inherited from higher-educated parents or both. In addition, higher educational achievements are also positively associated with an increase in the mother's age and the father's employment status, but negatively related to the mother's employment status. There is no observed difference in the education of children raised in female- and male-headed households.

Two variables of a child's characteristics deserve special attention. First, our results indicate that the sibship size produces a small and negative marginally significant effect after controlling for all other factors. Our estimates suggest that adding one sibling reduces the chance of going to general high school by 0.3% and general college by 0.4%. Second, our estimates indicate that being the first born largely increases the chance of enrollment in general high school and college, by 6.6% and 3.2%, respectively.

Our results demonstrate a strong link between a child's education and housing variables. A positive correlation is observed between a child's education and the floor space. An increase of 100 square meters, for instance, is associated with an increase in the chance of enrollment in general high school and college by 1.5% and 1.2%, respectively. Likewise, children living in owner-occupied houses have a higher chance of getting into general high school or college, as are children living in newer houses. However, interpreting these results requires caution. It is possible that the results reflect the fact that parents that are more willing to invest in a house are likely to create positive benefits for their children's learning.[15] It is also possible that these coefficients may reflect our inability to control for household income. Perhaps new, larger, self-owned houses produce a positive effect simply because they are associated with a child's family's well-being. We will discuss this issue later.

Youths who have recently moved from other locations (migrated 3–5 years ago) are less likely to be enrolled in general high school. The greater the distance they moved, the larger the negative effect on a child's education. Because the housing effect usually takes time to materialize, this effect should be attributed to residential stability in an earlier period, i.e., at the time of junior high school. Residential stability is valuable to teens probably because they do not have to learn to adapt to a new social network (junior high schools are usually very close to where teens live).[16] Notice that the effect of residential stability is less evident among teens. This could be because general high school admissions are based on every young adult's test score. As a result, many young adults cannot benefit from the existing social network as they did at junior high school because they must attend distant general high schools.

One parameter of particular interest is the household crowdedness. Similar to Goux and Maurin (2005), our results also confirm the importance of private space on a child's education.[17] Nevertheless, its effect is more complex and possibly nonlinear. For instance, teens growing up in families of medium crowdedness are more likely to enroll in high school than those in high- or low-crowdedness houses. For young adults, those raised in medium crowdedness perform equally well as those in low-crowdedness houses. Notice that our estimation also controls for a house's floor space. Changing from

---

[15] Green and White (1997) explained why home ownership might positively influence children's cognitive and behavioral outcomes. First, there is a stronger investment incentive for owners compared with renters. Better physical home environments increase the probability of success of the children of owners. Second, compared with renters, there is higher self-esteem among owners, resulting in greater emotional support for the children. Finally, there is greater geographic stability creating a neighborhood network that is likely to promote a child's outcome.

[16] A longer tenure (or less mobility) often implies a more stable home and school (peer) environment, which helps to invest in building social capital that enhances a child's outcomes. Therefore, a longer tenure tends to lead to better outcomes for children. For details, see Coleman (1988).

[17] Because our estimation setting is different from that of Goux and Maurin (2005), a comparison may be inappropriate. However, we estimated an additional model using the average number of rooms per person as the proxy for household crowdedness. Results of that estimation are similar to that of Goux and Maurin (2005), where we found a significant and negative impact of household crowdedness on a child's education.

high- to low-crowdedness houses does not refer to an increase in floor space and number of rooms at the same time. Instead, the effect should be interpreted as increasing the private space, but reducing the shared space, in a household (e.g., smaller living room). From our estimates, it appears that there is an optimal mix of private and shared space that helps a child's schooling.

As stated earlier, there is a risk of bias generated from our sample selection. Most notable is the restriction on the eldest sibling's age and on coresiding with a mid-aged adult. If such a restriction induces a new bias into the estimation, we should observe differences in results that use only 17-year-old youths from the teen sample and only 19-year-old youths from the young adult sample. This can be seen from Table 5, where we list estimated results of youths from the age cohorts of 17 and 19, respectively. As demonstrated in the table, we observe only modest differences between results using the full sample and half of the sample. None of the estimated coefficients, however, changes its sign after restricting the sample, and the vast majority remain statistically significant. These results imply that our sample selection, at most, results in small biases in the estimation.

B. Effects of Area Dummies

One key concern regarding our findings is whether our results demonstrate the importance of housing variables or just the inability to control for unobserved family heterogeneity. For instance, strong associations between a child's education and housing variables found in the estimation could possibly be caused by failure to control for the household's income, one kind of unobserved family heterogeneity. To address this question properly, it is important to show some evidence that adding area dummies indeed mitigates the concern of unobserved family factors. Table 6 lists the estimated results using area dummies at the town, village, and lin levels, respectively. For the purpose of comparison, we also include results without controlling for neighborhood fixed effects. As indicated from this table, the total number of area dummies at each level is 0, 364, 7508, and 91,929, respectively, and a slightly smaller number for young adults. Given that the number of area dummies varies so much, it is not surprising that regressions controlling for different levels of neighborhood effects yield dramatically different estimates. For instance, the coefficient of sibship size in the teen sample changes from –0.016 when there are no area controls to –0.006 and –0.003 when controlling at the village and lin level, respectively. In fact, the Hausman test suggests that any two sets of estimates are statistically different.[18]If positive relationships between housing variables, especially floor space, ownership status, housing age, and child's education, are posited as another channel to display the income effect, we should anticipate the effect becoming smaller when looking across children residing in the same neighborhood. Families residing in the same neighborhood should have similar family assets or potential earnings. Throughout the table, however, estimates of housing variables continue to show significant effects on the youth's educational outcome, some of which become even larger after controlling for many more area dummies. While it is still possible that our results are biased because large variations exist within the same neighborhood, the results do not seem to suggest that our findings are driven by unmeasured household income.

Another way to examine the effect of area dummies is to compare our results with findings in previous studies accounting for unobserved endogeneity through IV methods or family dummies. Generally, these studies found the coefficient on sibship size changes from statistically significant in OLS estimation to insignificant in IV estimation [e.g., Angrist, Lavy, and Schlosser (2005), Black, Devereux, and Salvanes (2005)]. It is therefore interesting to see whether adding more area dummies generates a similar result. From the table, it is clear that the coefficient of sibship size diminishes when a finer level of area controls is included. At the level of the lin, the coefficient of sibship size for teens is only marginally significant at the 10% level Obviously, a finer area control reduces the effect of sibship size, a sign supporting the reduction of unobserved family heterogeneity.

C. IV Estimation

So far we have shown that estimates with area fixed effects exhibit a pattern similar to recent studies employing IV strategy. Nevertheless, it is still uncertain whether neighborhood dummies are good controls for unobserved family heterogeneity. A more convincing method is to compare area fixed effect results with IV results so that the extent of endogenous bias can be determined.

[18] The smallest chi-square value occurs when comparing results of village fixed effects with those of lin fixed effects. Even for these, the value is 31.82 for young adults and 114.56 for teens, strongly rejecting the null hypothesis of the Hausman Test that these two sets of estimates are indifferent.

Nevertheless, this is not easy because our regression includes, in addition to the number of siblings, a variety of variables characterizing a child's housing environment. Unless we are able to find an instrument for every housing variable, implementing a full-scale IV estimation is extremely difficult.

In light of this difficulty, we have decided to conduct IV estimation in a different way. We first use multiple births and preferences toward a mixed sibling-sex composition to construct the instrument of sibship size.[19] Through exogenous variations because of multiple births at the third-born and sibling-sex composition of the oldest three children, we can look at the effect of three or more births on the educational outcome of the first- and second-born child in families with at least three children. Neighborhood dummies are also included to aid family controls.

The estimates in the first two columns of Table 7 report the first-stage and IV results for teens and young adults in families of at least three children. Because our sample is reduced to less than half of its original size because of the restriction on the number of children, we control for the village instead of the lin fixed effects. All instruments are significant in the first stage. Family size goes up by 0.83–0.87 in response to multiple births at the third born. Likewise, the family size increases by 0.38–0.42 for families whose first three siblings are girls; this reflects Taiwanese parents' preference for boys over girls.

Controlling for the village fixed effects, IV results again show that the number of siblings has little effect on the child's education. Moreover, we do not observe clear differences in the coefficients of housing variables between regression results and IV results. The vast majority of housing variables still hold their original signs and magnitudes. To formally examine whether IV results differ from area fixed effects results, we reestimate the area fixed effects model using village dummies based on this new sample.[20] The Hausman Test shows the chi-square value for these two sets of estimates is 3.5 for the teen sample and 0.29 for the young adult sample; both fail to reject the null hypothesis that the IV results and regression results are statistically indifferent.

---

[19] Taiwan Census data only record the age of each family member. Therefore, multiple births are identified by checking whether two consecutive children share the same age. It is possible that our method overstates the number of multiple births for families whose age gap between two consecutive children is less than one year. Nevertheless, we believe the likelihood of a mother having two children in one year is limited.

[20] The Hausman test is conducted based on the 21 explanatory variables in the regressions. Coefficients of fixed dummies are not considered.

Results reported above account for the potential endogeneity between sibship size and child's education. However, our estimates can still be subject to biases if housing variables are endogenously determined based on the number of children (e.g., parents may decide to move to a bigger house once they have more children). Given we cannot find an instrument for every housing variable, we restrict our sample to those who have moved into their current residence one year before the second child was born. For these families, the chance that their housing variables are correlated with the exogenous variations in sibship size (e.g., multiple births or sex composition) should be considerably lower, and therefore should shed some light on the effect of housing variables. The remaining columns of Table 7 present the first-stage and IV estimates for this particular sample. Although the first-stage results continue to confirm the validity of our instruments, coefficients on many housing variables become insignificant after imposing the restriction, likely because of a much smaller sample imposed by the move-in year constraint. Nevertheless, the majority of housing variables still hold their sign, showing at least some evidence of their importance.

D. Gender Differences

When discussing housing variables, one often-raised question is whether gender differences exist. Do boys need a bigger house? Do girls have special needs for private space? To explore this possibility, Table 8 expands the estimation by allowing for gender interactions on three variables: first-born, floor space, and household crowdedness. As expected, first-born boys have a higher school enrollment than first-born girls; this is likely because boys in Taiwan's society are subject to more social pressure than girls.

Both household crowdedness and floor space exhibit some gender differences. In addition, those gender differences seem to change for different ages. The chance of school enrollment is higher for boys raised in households with larger floor space, but there is no observable gender difference in the teen sample. On the contrary, girls raised in medium-crowdedness households have a higher chance of getting into high school. However, the gender difference disappears in the young adult sample. It appears that different housing needs exist for boys and girls at different times in their lives.

5. Conclusions

Understanding factors that determine the children's educational attainments is an important research question in the social sciences. The answer is not only crucial for human capital formulation, a key driver of economic growth, but also essential for income distribution purposes because education is considered a driver for income mobility.[21] Among those components, housing environment provided by the parents is often considered of great relevance [Haveman and Wolfe (1995)]. While it is widely believed that a better housing environment stimulates a child's learning, there is limited evidence as to the causal link between housing environment and a child's schooling.

In this study, we seek to uncover the effect of housing environment on children's educational attainments. Differing from Goux and Maurin (2005), who use exogenous variations in the child's private space as instruments, we control for unobserved family heterogeneity through their residential choices. In general, families living within a close distance share similar parental preferences, household assets, and earning potential. In addition, children in the same neighborhood typically go to the same school. Using the Taiwan census files that include the unique address information of every household in the records, we compare the chance of general high school or general college enrollment for youths of the same age and in the same neighborhood. After controlling for area fixed effects using tens of thousands of area dummies, our results indicate the importance of housing variables in determining a child's schooling. The educational attainment of children is positively associated with increases in floor space, increases in residence stability, and the ownership status, but negatively related to increases in building age. The results are robust even accounting for the endogeneity between sibship size and child's education using IV estimation.

Several findings deserve special attention. First, a first-born child, particularly a boy, is more likely to perform well in school. While the finding may reflect the fact that parents, particularly those in Taiwan, tend to put more pressure on first-borns, our finding is consistent with Black, Devereux, and Salvanes (2005) who argue that birth order, not family size, matters for a child's outcome. A more

[21] According to Haveman and Wolfe (1995), the government's spending on children in terms of primary and secondary education in 1992 is estimated to be 235 billion, or roughly 4% of GDP in U. S. In Taiwan, the spending on compulsory education is a little less than 3 percent of GDP.

careful analysis that explores a full range of effects of birth order and possibly its interactions with housing variables may be necessary.

Second, our results are different from the findings of Goux and Maurin (2005) regarding the effect of a child's private space. Although our results also confirm the importance of household crowdedness, its effect appears to be nonlinear because the chance of school enrollment is higher for children raised in medium-crowdedness houses than those in low-crowdedness houses. Further investigations on the effect of household crowdedness may also be necessary to uncover the exact impact.

Finally, and most importantly, our identification uses the area fixed effects to control for unobserved family heterogeneity. While we have shown evidence supporting this approach, we caution readers that there might still be uncontrolled family factors, such as genetic differences or interactions between parents and children, in the estimation.

The main contribution of this paper is to provide causal evidence regarding the effect of housing environment on a child's education. Although many studies have attempted to establish the link between housing environment and children's educational achievement (e.g., ownership, residential stability), our paper appears to be the first that offers a complete picture of the effect of a wide range of housing variables. Our paper has demonstrated the importance of a few housing variables (e.g., tenure status and house floor space). Future studies could use our findings as the basis to consider more effective policy instruments to enhance children's educational attainments in designing housing policy.

References

Aaronson, D. (2000), "A Note on the Benefits of Homeownership," *Journal of Urban Economics*, 47, 356–369.

Angrist, J. D., Lavy, V and J. D. Schlosser (2005), "New Evidence on the Causal Link between the Quantity and Quality of Children," Mimeo.

Becker, G., and H. G. Lewis (1973), "On the Interaction between the Quantity and Quality of Children," *Journal of Political Economy*, 81(2), 279–288.

Becker, G., and N. Tomes (1976), "Child Endowments and the Quantity and Quality of Children," *Journal of Political Economy,* 84(4), 143–162.

Black, S., P. J. Devereux, and K. G. Salvanes (2005), "The More the Merrier? The Effect of Family Composition on Children's Education," *Quarterly Journal of Economics*, 120(2), 669–700

Blake, J. (1981), "Number of Siblings and Educational Mobility," *American Sociological Review*, 50, 84–94.

Blake, J. (1989), *Family Size and Achievement*, Berkeley: University of California Press.

Boehm, T., and A. Schlottmann (1999), "Does Home Ownership by Parents Have an Economic Impact on Their Children?" *Journal of Housing Economics*, 8, 217–232.

Caceres, J. (2004), "Impact of Family Size on Investment in Child Quality: Multiple Births as Natural Experiment," Mimeo, Department of Economics, University of Maryland.

Coleman, J. (1988), "Social Capital in the Creation of Human Capital," *American Journal of Sociology*, 94, 95–120.

Conley, D., and R. Glauber (2005), "Parental Educational Investment and Children's Academic Risk: Estimates of the Impact of Sibship Size and Birth Order from Exogenous Variation in Fertility," NBER Working Paper: 11302.

Goux, D., and E. Maurin (2005), "The Effect of Overcrowded Housing on Children's Performance at School," *Journal of Public Economics*, 89, 797–819.

Green, R., and M. White (1997), "Measuring the Benefits of Homeowning: Effects on Children," *Journal of Urban Economics*, 41, 441–461.

Guo, G., and L. K. VanWey (1999), "Sibship Size and Intellectual Development," *American Sociological Review*, 64, 169–187.

Lee, B. A., R. S. Oropesa, and J. W. Kanan (1994), "Neighborhood Context and Residential Mobility," *Demography*, 31(2), 249–270.

Haveman R., and B. Wolfe (1995), "The Determinants of Children's Attainments: A Review of Methods and Findings," *Journal of Economic Literature*, 33(4), 1829–1878.

Haurin, D. R., R. D. Dietz, and B. A. Weinberg (2003), "The Impact of Neighborhood Homeownership Rates: A Review of the Theoretical and Empirical Literature," *Journal of Housing Research*, 13(2), 119–151.

Haurin, D. R., T. Parcel, and R. J. Haurin (2002), “Does Homeownership Affect Child Outcomes?” *Real Estate Economics*, 30(4), 635–666.

Phillips, Meredith. (1999), “Comment: Sibship Size and Academic Achievement: What We Now Know and What We Still Need to Know,” *American Sociological Review*, 64 (2), 188–192.

Powell, B., and L. C. Steelman (1993), “The Educational Benefits of Being Spaced Out: Sibship Density and Educational Progress,” *American Sociological Review*, 58 (3), 367–381.

Schultz, T. P. (2005) “Effects of Fertility Decline on Family Well-Being: Evaluation of Population Programs,” Draft for MacArthur Foundation Consultation Meeting, Philadelphia, PA.

Table 1: Youths Used in the Analysis
(Table Entries are Number of Observations Meeting Selection Criteria)

Age	Total number of youths available in the census	...and live in households that have at least one adult (aged over 35)	…and have no other relatives (nuclear families)
15	336,040	315,248	219,341
16	358,437	333,616	235,237
17	371,046	344,711	246,760
18	387,969	338,183	224,750
19	398,667	334,469	218,989
20	387,682	292,182	175,669

Table 1: Cont.
(Table Entries are Number of Observations Meeting Selection Criteria)

Age	…and have valid father and mother info (exclude single parent)	…and the eldest sibling is less than or equal to 22 years old	…and moved into the current residence 3 years ago
15	158,577	149,821	132,404
16	170,769	158,145	140,616
17	178,860	160,429	143,343
18	160,706	137,039	122,611
19	155,629	122,284	109,924
20	123,481	87,234	79,013

Table 2: Summary Statistics of the Youth's Education, Family Background, and Housing

	Mean/Percent		Mean/Percent	
	Age (16)	Age (17)	Age (19)	Age (20)
Child's Education				
Junior High School or Less	8.72%	8.30%	6.87%	5.18%
Vocational High School	36.45%	40.35%	32.16%	25.16%
General High School	54.83%	50.89%	21.43%	14.86%
Junior College	0.00%	0.31%	22.20%	29.83%
General College or Above	0.00%	0.15%	17.34%	24.98%
Child's Characteristics				
Male	52.03%	52.21%	52.00%	40.65%
First Born	49.42%	54.19%	63.56%	70.07%
Number of Siblings (Self Included)	2.49	2.45	2.45	2.43
	(0.859)	(0.871)	(0.886)	(0.915)
Parental Background				
Female Economic Head	10.48%	10.80%	11.32%	11.50%
Mother's Age	41.86	42.50	43.61	44.36
	(3.398)	(3.390)	(3.246)	(3.158)
Mother's Education (0-6 years)	28.30%	31.72%	40.31%	41.19%
Mother's Education (6-9 years)	28.59%	27.70%	26.18%	24.12%
Mother's Education (9-12 years)	31.02%	29.42%	24.75%	24.92%
Mother's Education (12+ years)	12.09%	11.16%	8.77%	9.77%
Mother's Employment	60.03%	60.09%	58.67%	57.73%
Father's Age	44.81	45.48	46.64	47.36
	(3.454)	(3.407)	(3.212)	(3.087)
Father's Education (0-6 years)	23.12%	26.08%	34.17%	34.31%
Father's Education (6-9 years)	24.67%	23.62%	22.17%	20.31%
Father's Education (9-12 years)	30.75%	29.85%	27.34%	27.01%
Father's Education (12+ years)	21.45%	20.45%	16.32%	18.37%
Father's Employment	93.25%	92.74%	91.16%	90.97%
Housing Environment				
Rent	6.54%	6.36%	6.66%	6.02%
Space (Square Meter/100)	1.32	1.32	1.31	1.32
	(0.697)	(0.703)	(0.699)	(0.696)
Number of Rooms	3.52	3.53	3.53	3.53
	(1.268)	(1.265)	(1.279)	(1.260)
Low Crowdedness	29.56%	31.45%	31.70%	32.72%
Medium Crowdedness	36.44%	36.03%	35.77%	35.13%
High Crowdedness	34.00%	32.53%	32.53%	32.15%
Move from Vicinity*	3.20%	3.06%	2.90%	2.65%
Move from Distant Area*	10.01%	9.27%	9.20%	9.13%
Building Age (0-10 years)	22.57%	21.59%	19.71%	18.82%
Building Age (10-20 years)	40.99%	40.76%	40.04%	40.14%
Building Age (20+ years)	36.43%	37.65%	40.25%	41.04%
Number of Observations	140,616	143,343	109,924	79,013

Standard deviations are in parentheses.

*The youth is considered as a new mover if his/her current address differs from that of 5 years ago

Table 3: Observation Number at Various Levels of Government Jurisdiction[a]

	High School (Age 16/17)			College (Age 19/20)		
	Town	Village	Lin	Town	Village	Lin
Min	1	1	1	1	1	1
Max	8,708	372	95	6,228	244	42
Mean	780.11	37.82	3.09	519.06	25.44	2.40
	(1124.02)	(38.42)	(2.66)	(772.35)	(25.40)	(1.87)
25% Percentile	130	11	1	82	7	1
	(798)	(37)	(3)	(549)	(25)	(2)
50% Percentile	130	11	1	82	7	1
	(1803)	(67)	(4)	(1388)	(45)	(3)
75% Percentile	990	55	4	629	36	3
	(3239)	(103)	(7)	(2166)	(68)	(5)
# of jurisdictions	364	7,508	91,929	364	7,428	78707

[a]The percentile in parentheses reflects the observation number over the whole distribution

Table 4: "Within" and "Between" Standard Deviation of Housing Variables at the Level of Village and Lin

	High School (Age 16/17)				College (Age 19/20)			
	Large Towns[b]		Small Towns		Large Towns		Small Towns	
	Village	Lin	Village	Lin	Village	Lin	Village	Lin
Tenure Status	0.079	0.213	0.085	0.179	0.088	0.223	0.090	0.177
	(.26)	(.207)	(.204)	(.153)	(.261)	(.193)	(.195)	(.136)
Space (Square Meter/100)	0.348	0.571	0.389	0.684	0.366	0.585	0.413	0.695
	(.552)	(.4)	(.696)	(.492)	(.55)	(.372)	(.681)	(.438)
Number of Rooms	0.551	0.972	0.804	1.365	0.523	0.983	0.899	1.381
	(1.026)	(.769)	(1.349)	(.936)	(1.026)	(.72)	(1.335)	(.846)
Low Crowdedness	0.150	0.340	0.206	0.399	0.158	0.370	0.232	0.423
	(.429)	(.347)	(.457)	(.345)	(.435)	(.326)	(.459)	(.318)
Medium Crowdedness	0.121	0.351	0.176	0.371	0.137	0.381	0.200	0.391
	(.48)	(.398)	(.453)	(.353)	(.476)	(.366)	(.447)	(.319)
High Crowdedness	0.147	0.353	0.201	0.385	0.156	0.374	0.216	0.394
	(.457)	(.372)	(.451)	(.342)	(.456)	(.344)	(.44)	(.308)
Move from Vicinity	0.050	0.130	0.067	0.149	0.049	0.134	0.081	0.146
	(.169)	(.136)	(.173)	(.129)	(.16)	(.119)	(.161)	(.111)
Move from Distant Area	0.094	0.238	0.110	0.227	0.106	0.247	0.141	0.246
	(.297)	(.229)	(.267)	(.198)	(.284)	(.204)	(.265)	(.18)
Building Age (0-10 years)	0.179	0.348	0.185	0.370	0.176	0.346	0.196	0.371
	(.359)	(.213)	(.41)	(.269)	(.342)	(.19)	(.389)	(.235)
Building Age (10-20 years)	0.223	0.433	0.204	0.399	0.229	0.444	0.222	0.418
	(.45)	(.261)	(.447)	(.292)	(.445)	(.243)	(.444)	(.267)
Building Age (20+ years)	0.251	0.449	0.247	0.441	0.256	0.459	0.266	0.455
	(.421)	(.244)	(.449)	(.289)	(.428)	(.232)	(.451)	(.264)
# of observations	169,290	169,290	114,669	114,669	114,138	114,138	74,799	74,799

[a]The number in the parentheses is the "within" standard deviation of housing variables.

[b] A town is defined as large if its number of residents exceeds 100,000 residents and as small if less than 100,000 residents.

Table 5: Estimated Results of the Youth's Educational Achievements

	High School		College	
	Age (16/17)	Age(17)	Age(19/20)	Age (19)
Child's Characteristics				
Male	-0.006	-0.008	0.001	-0.026
	(0.002)***	(0.003)**	(0.002)	(0.003)***
First Born	0.066	0.073	0.032	0.017
	(0.002)***	(0.004)***	(0.003)***	(0.004)***
Number of Siblings (Self Included)	-0.003	-0.001	-0.004	-0.005
	(0.002)*	(0.003)	(0.002)**	(0.003)*
Parental Background				
Family Head	-0.001	-0.002	-0.005	0.000
	(0.004)	(0.006)	(0.004)	(0.006)
Mother's Age	0.017	0.026	0.04	0.018
	(0.006)***	(0.009)***	(0.007)***	(0.010)*
Mother's Age Square	0.000	0.000	0.000	0.000
	(0.000)**	(0.000)***	(0.000)***	(0.000)
Mother's Education (6-9 years)	0.034	0.033	0.018	0.014
	(0.003)***	(0.005)***	(0.003)***	(0.004)***
Mother's Education (9-12 years)	0.069	0.069	0.057	0.045
	(0.003)***	(0.005)***	(0.004)***	(0.005)***
Mother's Education (12+ years)	0.143	0.147	0.128	0.117
	(0.005)***	(0.008)***	(0.005)***	(0.008)***
Mother's Employment	-0.008	-0.007	0.000	-0.001
	(0.002)***	(0.004)*	(0.003)	(0.004)
Father's Education (6-9 years)	0.034	0.029	0.024	0.016
	(0.003)***	(0.005)***	(0.003)***	(0.005)***
Father's Education (9-12 years)	0.059	0.054	0.06	0.054
	(0.003)***	(0.005)***	(0.003)***	(0.005)***
Father's Education (12+ years)	0.149	0.144	0.153	0.143
	(0.004)***	(0.007)***	(0.005)***	(0.006)***
Father's Employment	0.024	0.02	0.027	0.027
	(0.004)***	(0.007)***	(0.004)***	(0.006)***
Housing Environment				
Rental Status	-0.031	-0.025	-0.058	-0.053
	(0.005)***	(0.008)***	(0.005)***	(0.007)***
Space (Square Meter/100)	0.015	0.009	0.012	0.012
	(0.002)***	(0.003)***	(0.002)***	(0.003)***
Low Crowdedness	0.005	0.007	0.014	0.009
	(0.004)	(0.006)	(0.004)**	(0.006)
Medium Crowdedness	0.011	0.014	0.013	0.011
	(0.003)**	(0.005)**	(0.003)**	(0.005)*
Move from Vicinity*	-0.024	-0.026	-0.023	-0.022
	(0.006)***	(0.011)**	(0.008)***	(0.010)**
Move from Distant Area*	-0.040	-0.041	0.001	-0.003
	(0.004)***	(0.007)***	(0.005)	(0.006)
Building Age (10-20 years)	-0.012	-0.006	0.002	-0.004
	(0.004)***	(0.006)	(0.005)	(0.006)
Building Age (20+ years)	-0.024	-0.023	-0.009	-0.015
	(0.004)***	(0.007)***	(0.005)*	(0.007)**
Number of Area Dummies (lin)	91929	71294	78707	62122
Number of Observations	283,959	143,343	188,937	79,013

* significant at 5%; ** significant at 1%; Standard deviations are in parentheses.

Table 6: Estimated Results of the Youth's Educational Achievements (Robustness Checks)[a]

	High School (Age 16/17)			College (Age 19/20)		
Level of Jurisdiction	Town	Village	Lin	Town	Village	Lin
Number of Dummies	364	7,508	91,929	364	7,428	78,707
Child's Characteristics						
First-Born	0.072	0.072	0.066	0.029	0.030	0.032
	(0.002)**	(0.002)**	(0.002)**	(0.002)**	(0.002)**	(0.003)**
Number of Siblings	-0.007	-0.005	-0.003	-0.004	-0.004	-0.004
	(0.001)**	(0.001)**	(0.002)	(0.001)**	(0.001)**	(0.002)*
House Characteristics						
Tenure Status	-0.027	-0.031	-0.031	-0.054	-0.058	-0.058
	(0.004)**	(0.004)**	(0.005)**	(0.004)**	(0.004)**	(0.005)**
Space (Square Meter/100)	0.018	0.017	0.015	0.015	0.014	0.012
	(0.001)**	(0.002)**	(0.002)**	(0.001)**	(0.002)**	(0.002)**
Low Crowdedness	-0.003	0.001	0.005	0.008	0.008	0.014
	(0.003)	(0.003)	(0.004)	(0.003)**	(0.003)**	(0.004)**
Medium Crowdedness	0.007	0.009	0.011	0.011	0.010	0.013
	(0.002)**	(0.002)**	(0.003)**	(0.002)**	(0.002)**	(0.003)**
Move from Vicinity	-0.031	-0.030	-0.024	-0.021	-0.018	-0.023
	(0.005)**	(0.005)**	(0.006)**	(0.005)**	(0.005)**	(0.008)**
Move from Distant Area	-0.040	-0.042	-0.040	0.001	0.003	0.001
	(0.003)**	(0.003)**	(0.004)**	(0.003)	(0.003)	(0.005)
Building Age (10-20 years)	-0.017	-0.017	-0.012	0.002	0.001	0.002
	(0.002)**	(0.003)**	(0.004)**	(0.003)	(0.003)	(0.005)
Building Age (20+ years)	-0.026	-0.027	-0.024	-0.007	-0.010	-0.009
	(0.003)**	(0.003)**	(0.004)**	(0.003)**	(0.003)**	(0.005)
Hausman Test						
vs Town Fixed Effect		447.60	295.42		201.99	75.27
vs Village Fixed Effect			114.56			31.82

** significant at 5%; * significant at 1%; standard deviations are in parentheses.

[a]The estimation controls for the mother's age, education and employment of mother and father, economic head, and child's gender.

Table 7: Results of IV Estimation Using Twin Births and Sex-Composition as Instrument[a]

Sample Selection	First and second child in families with at least 3 children		...and moved in before the 2nd child is born	
	High School (Age 16/17)	College (Age 19/20)	High School (Age 16/17)	College (Age 19/20)
First-stage Results				
Instruments				
First Two Boys	-0.048	-0.071	-0.059	-0.053
	(0.006)***	(0.007)***	(0.012)***	(0.018)***
First Two Girls	0.043	0.056	0.053	0.048
	(0.005)***	(0.005)***	(0.009)***	(0.013)***
First Three Boys	0.041	0.048	0.047	0.022
	(0.007)***	(0.009)***	(0.014)***	-0.022
First Three Girls	0.382	0.420	0.385	0.446
	(0.005)***	(0.006)***	(0.010)***	(0.014)***
Twins at the Third Birth	0.869	0.828	0.857	0.811
	(0.017)***	(0.021)***	(0.035)***	(0.057)***
Fixed Effect/ IV Results				
Child's Characteristics				
First Born	0.070	0.029	0.061	0.030
	(0.003)***	(0.003)***	(0.007)***	(0.007)***
Number of Sibling	0.000	-0.006	0.011	-0.021
	(0.010)	(0.008)	(0.021)	(0.017)
House Characteristics				
Rental Status	-0.024	-0.051	-0.008	-0.049
	(0.006)***	(0.005)***	(0.005)	(0.018)***
Space (Square Meter/100)	0.014	0.009	0.006	0.012
	(0.003)***	(0.002)***	(0.005)	(0.005)**
Low Crowdedness	0.002	0.003	-0.008	-0.019
	(0.006)	(0.005)	(0.011)	(0.010)*
Medium Crowdedness	0.019	0.013	0.010	0.006
	(0.005)***	(0.004)***	(0.010)	(0.009)
Move from Vicinity	-0.030	-0.024	-0.086	-0.057
	(0.009)***	(0.008)***	(0.025)***	(0.025)**
Move from Distant Area	-0.040	0.005	-0.108	0.026
	(0.006)***	(0.005)	(0.017)***	(0.015)*
Building Age (10-20)	-0.020	-0.001	-0.104	0.142
	(0.005)***	(0.004)	(0.078)	(0.135)
Building Age (20+)	-0.025	-0.009	-0.094	0.155
	(0.005)***	(0.004)**	(0.078)	(0.135)
Number of Area Dummies	7145	7145	6290	5578
Number of Observations	94457	94457	26885	17546

* significant at 10%; ** significant at 5%; *** significant at 1%; Standard deviations are in parentheses.

[a]The estimation controls for the mother's age, education and employment of the mother and father economic head, gender, as well as neighborhood fixed effects at village level.

Table 8: Results of the Youth's Educational Attainments (Gender Interactions)[a]

	High School Age (16/17)	College Age (19/20)
Child's Characteristics		
Male	-0.015	-0.017
	(0.006)**	(0.007)**
First Born*Boy	0.055	0.031
	(0.003)***	(0.005)***
First Born*Girl	0.042	0.023
	(0.004)***	(0.004)***
Number of Sibling	-0.01	-0.007
	(0.002)***	(0.002)***
Housing Environment		
Rental Status	-0.031	-0.058
	(0.005)***	(0.005)***
Space (Square Meter/100)*Boy	0.014	0.014
	(0.003)***	(0.003)***
Space (Square Meter/100)*Girl	0.014	0.009
	(0.003)***	(0.003)***
Low Crowdedness*Boy	0.002	0.018
	(0.005)	(0.005)***
Low Crowdedness*Girl	0.006	0.01
	(0.005)	(0.005)**
Medium Crowdedness*Boy	0.004	0.011
	(0.004)	(0.005)**
Medium Crowdedness*Girl	0.009	0.012
	(0.004)**	(0.004)***
Move from Vicinity*	-0.023	-0.023
	(0.006)***	(0.008)**
Move from Distant Area*	-0.039	0.001
	(0.004)***	(0.005)
Building Age (10-20)	-0.012	0.002
	(0.004)***	(0.005)
Building Age (20+)	-0.025	-0.009
	(0.004)***	(0.005)*
Number of Area Dummies	91929	78707
Number of Observations	283,959	188,937

* significant at 10%; ** significant at 5%; *** significant at 1%; Standard deviations are in parentheses.

[a]The estimation controls for neighborhood dummies at lin level