CPB Netherlands Bureau for Economic Policy Analysis

The urban rural-education gap: do cities indeed make us smarter?

Despite extensive research into the urban-rural wage gap, the urban-rural education gap has received far less attention. This paper investigates in the context of the Netherlands whether children growing up in urban and rural communities make different educational choices, conditional on observed cognitive ability and a wide range of family characteristics.

The paper finds an elasticity of university attendance w.r.t. population density of 0.07, which is robust across a variety of specifications. Furthermore, the results cannot be explained by sorting on unobserved variables under the assumptions of Oster (2017).

CPB Discussion Paper

Raoul van Maarseveen

April 2020

The urban rural-education gap: do cities indeed make us smarter? [*1]

Raoul van Maarseveen[†]

April 16, 2020

Abstract:

Despite the existence of a large urban-rural education gap in many countries, little attention has been paid whether cities enjoy a comparative advantage in the production of human capital. Using Dutch administrative data, this paper finds that conditional on family characteristics and cognitive ability, children who grow up in urban regions consistently attain higher levels of human capital compared to children in rural regions. The elasticity of university attendance w.r.t. population density is 0.07, which is robust across a wide variety of specifications. Hence, the paper highlights an alternative channel to explain the rise of the city.

JEL codes: I20, J24, R10

[*] I would like to express my gratitude to Niklas Bengtsson, Cristina Bratu, Matz Dahlberg, Peter Fredriksson, Edward Glaeser, Lawrence Katz, Matthew Lindquist, Björn Öckert, Albert Saiz, Anna Thoresson, Anna Tompsett, Santiago Truffa and Dinand Webbink, as well as the participants to the Urban Economics Association conference 2019 and the VU Amsterdam/Uppsala University seminars for their excellent comments and suggestions. Furthermore, I would like to thank CPB Netherlands Bureau for Economic Policy Analysis for making the data available for this research, as well as their excellent suggestions. All results in this paper are based on calculations by the author using non-public microdata from Statistics Netherlands. The code used to generate the results is available upon request.

[1] The central research question of this paper was first mentioned in an earlier CPB publication (Van Maarseveen et al., 2017) in the Appendix “Suggesties voor vervolgonderzoek”, which argued that very little was known about regional differences in educational outcomes, particularly in Europe, and called for more future research on this topic. I developed the general idea for this paper during my time at CPB, which subsequently turned into a research project and the first chapter of my doctoral thesis at Uppsala University.

[†] Uppsala University, contact: *raoul.vanmaarseveen@nek.uu.se*

I Introduction

The resurgence of the city has received widespread attention from both academics and policy makers in the recent decades. Explanations typically highlight the success of cities at attracting well educated and skilled individuals from elsewhere, both by offering higher wages (see Combes et al., 2008) and superior consumption amenities (see Glaeser et al., 2001). Furthermore, cities are seen to have a comparative advantage in utilizing human capital in the production of goods and services, due to the sharing, matching and learning mechanisms described in Duranton and Puga (2004) and Rosenthal and Strange (2004).

However, the literature to date has taken a relatively static view of the relationship between cities and human capital. Most of the focus has been on the way in which cities attract and efficiently employ *existing* stocks of human capital. On the other hand, very little attention has been paid to whether cities also have a comparative advantage in *creating* human capital. This is surprising, as there are good reasons why agglomeration economies may also operate in the production of human capital in the education sector. Hence, this paper investigates whether growing up in an urban environment affects human capital investment decisions of children and young adults.

Despite the existence of a substantial urban-rural education gap in the majority of countries[2], it remains unclear to what extend this reflects a comparative advantage of cities in educating their population and to what extend it reflects the spatial sorting of households. To answer this question, I make use of the particular institutional setting in one country, The Netherlands. The Dutch educational system requires students to make conscious decisions about their level of human capital investment at various ages. At the end of both primary school and secondary school, students select into fairly rigorous schooling tracks which are clearly distinct in the level of human capital accumulation which they provide. Crucially, all students participate in high-stakes nationwide tests of academic ability at the end of primary school and secondary school. These test-scores are highly predictive of future academic outcomes, available for the entire population and thus provide a high-quality measures of the students' cognitive and academic ability at the same moment as when students make key educational decisions.

The empirical strategy leverages the large number of observed household and individual level characteristics, as well as the measures of cognitive ability, to control for heterogeneity between

[2] See Appendix A for the urban-rural education gap for a wide variety of countries based on the Demographic and Health Survey (DHS) and available census data from IPUMS international.

children who grow up in cities and rural areas.[3] Given very flexible controls for household characteristics and the observed measures of cognitive ability, do we see that children who grow up in rural communities make structurally different choices with regards to their educational investment compared to children who grow up in urban communities? In particular the availability of pre-determined measures of cognitive ability diminishes the concern that differences in academic potential drives differences in educational choices between urban and rural regions.

The analyses reveal substantial differences in the human capital decisions between children who grow up in urban areas and rural areas. Conditional on family background and cognitive ability, a one log-point increase in population density[4] is associated with a 1.68 percentage point increase in the likelihood that a child enrolls in an academic secondary school (VWO in Dutch), from a base of 23%. Similarly, amongst the group of high school students who obtained all prerequisites to enroll in the majority of university studies, a one log-point increase in population density is associated with a 0.8 percentage point increase in the likelihood that a child attends university, from a base of 84%. Thus, children who grow up in more urban environments are significantly more likely to select into the schooling tracks that provide higher levels of human capital accumulation. Taken together, the results imply that, conditional on family characteristics and observed academic ability, a one log-point increase in population density is associated with a 1.4 percentage point increase in the probability that a child will attend university from a base of 20%, which implies an elasticity of university attendance with respect to population density of 0.07. Similar to Combes et al. (2008) in the case of the urban wage premium, I find an elasticity about twice as large when individual- and family characteristics are not controlled for due to the spatial sorting of households.

To assess whether differences in unobserved characteristics between urban and rural students might be responsible for differences in human capital investment decisions, I employ the methodology of Oster (2017), using the selection on observed variables as a guide for the selection on unobserved variables. Under the assumptions of Altonji et al. (2005) and Oster (2017), that selection on

[3] A different empirical approach that could be applied is the methodology of Chetty and Hendren (2018), who identify the effect of US counties on educational outcomes by exploiting differences in the age of children when households move to generate differential exposure effects to counties. However, their identifying assumption, namely that the age of children at the time of the move is uncorrelated with their potential outcomes, does not hold in the Dutch context (see appendix D). Both the number of movers and the observed academic ability measured at the age of 11/12 negatively correlates with age of move for children older than 11/12, which indicates that this identifying assumption does not hold in the Dutch setting.

[4] Which equals 1.15 standard deviations.

unobserved characteristics is weakly less than selection on observed characteristics, omitted variable bias cannot account for the results obtained in this paper. In addition, the coefficients are similar across subgroups and cohorts and are not driven by functional form assumptions. Finally, the results change very little when municipality or provincial fixed effects are included or when using the historical densities of 1840 as instrument for modern densities to account for the potential endogeneity of urban settlements.

The findings of this paper contribute to three strands of literature. First, the paper contributes to the urban economics literature by suggesting an alternative channel to explain the resurgence of the city. The findings indicate that cities may not only have a comparative advantage in attracting and utilizing existing stocks of human capital as noted in existing literature, but may also play a key role in the creation of human capital. The findings that human capital accumulation is higher amongst children in cities also nicely complements the recent work by De la Roca and Puga (2017), who find that workers in cities acquire human capital at a faster rate compared to workers in rural areas. Hence, the paper provides an alternative channel to explain the recent rise of the city, distinct from the traditional "production city" and "consumer city" explanations. In addition, the findings suggest that studies that examine the costs of policies that limit access to cities, such as greenbelts in the UK, zoning laws in the US or Hukou system in China, may underestimate the costs of such policies, as they typically focus on the costs of denying workers access to productive places (see for instance Hsieh and Moretti (2019)).

Second, the results have implications for regional growth and the persistence of regional differences. As widely documented in the literature, the existing stocks of human capital in rural regions are typically much lower compared to urban regions. The findings of this study suggest that rural regions are in addition at a disadvantage when it comes to the expansion of their human capital stock, as children in rural areas invest less in their human capital formation, even when they have the same cognitive ability and family background as children in urban areas (which typically is not the case). This is particularly important due to the fact that human capital is the most important predictor of future economic growth in regions (Gennaioli et al, 2012), thus providing a mechanism that might explain the slow observed convergence in incomes between regions (Gennaioli et al, 2014).

Finally, the paper contributes to a rapidly growing literature on the spatial (in)-equality of opportunity, following Chetty et al. (2014a; 2018) in the US,, Alesina et al. (2019) in Africa and

Deutscher (2020) in Australia. This literature has clearly shown that the places in which children grow up matter for opportunities in life. In this paper I focus on one specific aspect of places, namely population density, and examine the relationship with educational attainment. The results indicate that children who grow up in rural communities do not appear to enjoy or take the same educational opportunities as children in urban communities, even in a country such as The Netherlands where the distance between urban centers and rural places is fairly limited from an international perspective. The findings of this paper thus suggest that the urban-rural differences in educational outcomes found across globe may not just be a result of spatial sorting of households, but may also reflect a lack of opportunities or awareness of opportunities for children residing in rural regions.

Finally, a limitation of this study is that it cannot attribute the effect of density on educational attainment to individual mechanisms. Based on the existing literature, there are three main channels through which density is likely to affect educational outcomes. First, the returns to education are higher in cities as agglomeration forces mainly complement the productivity of high-skilled workers (Baum-Snow et al., 2019; Autor, 2019), which raises the incentive to invest in human capital in urban areas. Secondly, the costs of obtaining education are lower in cities, both due to the smaller distance to educational institutes (Frenette, 2006), as well as the possibility that the more diverse school choice in cities improves the match between students their needs and interests and the educational institutes. Third, it might be costlier for youth to acquire information about future educational possibilities in rural areas, due to the absence of university outreach programs as well as the absence of strong network linkages to higher educational institutes in rural communities (Hoxby and Avery, 2012). However, a detailed causal investigation of individual mechanisms in the fashion of De la Roca and Puga (2017) and Dauth et al. (2018) in the case of the urban wage premium is beyond the scope of this paper.

The paper proceeds as following. Section 2 reviews the existing theoretical and empirical literature which links density to educational outcomes. Section 3 provides an overview of the Dutch context and data. Section 4 discusses the methodology and identification strategy. Section 5 presents the results and robustness analyses. Section 6 discusses the implications and concludes.

II Related literature

Based on the existing literature, there are good reasons why we might expect children with similar capabilities to make different educational choices depending on whether they live in an urban or rural community. This section will discuss the theoretical and empirical support for three of these mechanisms, as well as review the empirical literature linking density and educational attainment.

II.I Higher returns to education

A key determinant of educational choice in any model of educational investment are the expected returns to schooling. Models of educational investment, such as the seminal Ben-Porath model, typically assume that students invest in education until the point where the discounted increase in future wages is equal to the opportunity costs and direct costs of obtaining education, providing a tight link between perceived returns to education and human capital investments. Since agglomeration forces raise productivity mainly for the highly educated and skilled, while having a relatively small effect on productivity of the low skilled, the returns to education are on average higher in cities (Gould, 2007; Combes and Gobillon, 2015, Baum-Snow et al., 2018). This difference has become more pronounced over the last 20 years, during which the urban wage premium for non-college educated workers essentially disappeared in the US (Autor, 2019). Hence, in the absence of perfect mobility one would expect the higher returns to education in cities, ceteris paribus, to lead to higher educational investment by children growing up in urban places.

A key question though is whether children and young adults correctly infer the returns to education from their environment and if they respond accordingly. The way in which people form beliefs with regards to the returns to education is still largely unknown, but the experiment by Jensen (2010) suggests that these beliefs are not necessarily deeply held, and that beliefs are updated when new information is presented. Similarly, Jensen (2012) and Oster and Steinberg (2013) use respectively openings of call centers in India as shock to the local returns to education and find increased school enrollment. Adukai et al. (2020) find that educational attainment increases when villages in rural India are connected by roads to cities, and that this increase in educational investment increases with the returns to education in the newly connected cities.

II.II Costs of schooling

A second channels through which population density could affect educational outcomes is through a reduction in the perceived costs of attending schools, both due to a reduction in transportation costs to schools, as well as the possibility of a better match between students and educational institutes due to the larger school choice in cities. At the level of the primary and secondary school, the higher density of schools in urban areas means that children on average will spend less time and financial resources on commuting to a school of a given level and quality. In the case of tertiary education, commuting to college or university may become simply unfeasible for children in rural areas, hence necessitating a costly move to attend further education. As mentioned earlier, most models of educational choice predict that students invest in their education until the point where the marginal returns of education owing to higher future wages are offset by the direct costs and opportunity costs of education. As such, the higher commuting or moving costs that students in rural communities face may well lead to a reduction in educational investment. The empirical literature finds supporting evidence for the importance of commuting costs on educational decisions, particularly in the case of tertiary education. Card (1993) uses distance to college as instrument for educational attainment and finds a significant effect in the first stage. Frenette (2006) finds in Canada that children whose families live more than 80 kilometers from the nearest university have a 40% lower probability to enroll at university compared to individuals who grow up within 40 kilometers of a university. However, both studies include only a very limited set of control variables, and as such, it is not clear to what degree the results are driven by spatial sorting of households.

Furthermore, the higher density of educational institutions in cities may also allow for better matching between schools and students, as long as students and schools are heterogeneous on some dimension. Burgess et al. (2019) show in the UK that the majority of students do not list the nearest secondary school as their first preference, indicating that students do not perceive schools as a homogenous good. Such school-child match specific component can be based on academic preferences of the child, such as the level of instruction (see for instance Bau., 2019) or the focus on certain specialization tracks, or can be based on more personal preferences of the child, such as the religious orientation of a school (Cohen-Zada and Sanders, 2008). Typically, models of school choice assume that an outside option is available with a utility of zero, which depending on the context might consist of staying home, working or choosing a lower level of schooling. As long as students have a large range of schools to choose from, then the student will be likely to find a school which yields a higher utility then the outside option. However, the smaller the number of schools available, the more likely that unfavorable draws of the school-student match quality will result in the student choosing the outside option, which may reduce human capital investment in rural areas.

II.III Information and network effects

Third, density may affect human capital decisions through the availability of information on future schooling prospects or due to network effects. Hoxby and Avery (2012) find in the US that high performing students who apply to non-selective colleges are disproportionally located in rural areas, as the lower density of high performing students in rural areas makes it less profitable for colleges to engage in outreach campaigns. In addition, Hoxby and Avery (2012) suggest that the lower SES-composition of households typically found in rural areas may limits the exposure to alumni of various educational institutes, as well as reduced the expertise of local study counselors on how to advise high-performing students. Hence, limited information about future educational possibilities in rural areas provides a third potential channel through which density would be expected to affect educational investment.

II.IV Empirical evidence

Despite good reasons to expect children and young adults in urban areas to invest more in their human capital, the empirical evidence on this subject is very limited. Knight and Shi (1996) show large differences in years of educational attendance in China between urban and rural regions, although it remains unclear to what degree this is driven by household heterogeneity. Katz & Goldin (2008, p222) instead report a negative relationship between town size and high school graduation rates in the early 20th century in the US, which they suggest might be the result of the higher opportunity costs of education in cities. However, a key challenge in this literature has been dealing with the spatial sorting by households and unobserved heterogeneity between children. Separating spatial selection and area effects is particularly difficult in the case of educational outcomes, as panel data on the level of the individual is typically unavailable. As such, the methods developed to estimate the urban wage premium, which rely heavily on the inclusion of individual fixed effects, cannot easily be applied in the case of educational outcomes.

As a way to overcome the spatial sorting of households, some studies have relied on quasi-experimental variation to quantify the effect of exposure to neighborhood on educational outcomes. Chetty et al. (2016) use the MTO-program to study neighborhood effects in general, which provided a randomly selected group of families the opportunity to move to a better neighborhood with less poverty and higher incomes. They find that children who moved to better neighborhoods before the age of 13 had significantly higher schooling outcomes compared to those who did not receive the opportunity to move. However, such experiments are typically difficult to scale beyond the size of the individual city. In a more general approach, Chetty et al. (2018) use the difference in the age of children when they move between regions to identify neighborhood effects. They find a negative relationship between population density and upward mobility as measured by the income at age 26, but it remains unclear to what extend differences in human capital accumulation may have played a role.

III Context and data

In this paper, I utilize the Dutch context to investigate the relationship between population density and educational outcomes, which is particular well suited for this question. Tracking in the Dutch educational system starts relatively early at the age of eleven to twelve, which means that meaningful decisions on human capital investment are taken from an early age onwards. Crucially, the Netherlands institutes high-stakes national tests of academic ability shortly before children make their key educational decisions, namely the end of primary school and the end of secondary school. Hence, I observe detailed measures of cognitive and academic ability at the time when students make their educational decisions, which remains typically unobserved in most other settings. Furthermore, a key outcome in the education literature is college/university attendance. Earlier studies have shown that not only attending college or university matters for later life outcomes, but that also that the quality of the institution matters (Hoekstra, 2009). As the Dutch universities are relatively homogenous in terms of quality[5], I can use university enrollment as a more or less consistent outcome measure without being overly concerned about the precise institution which students attend. A drawback of the Netherlands is that it is relatively urbanized by international standards, thus limiting the variation in density. However, earlier studies in the

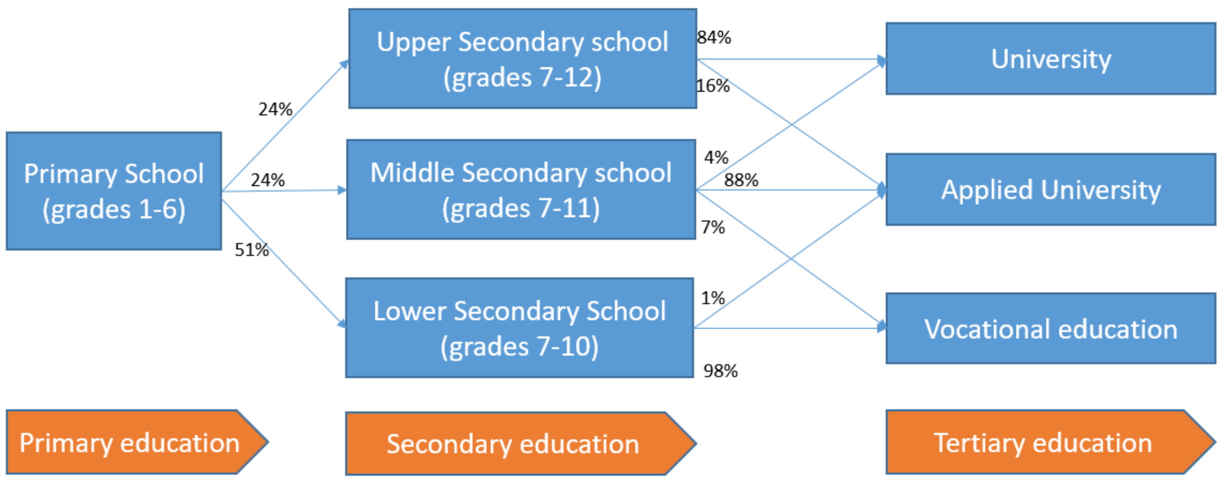
[5] For instance, all 13 Dutch universities rank between places 59 and 195 on the 2018 Times higher education ranking.

Netherlands on the urban wage premium found results in line with other countries (Groot et al., 2014; Verstraten et al., 2018).

III.I Context

Figure II provides an overview of the Dutch school system, including the flows between school types for the cohort born in 1996. Compulsory education starts at age 6, when all students are enrolled into primary school. Primary school continues for six years, at the end of which students participate in a national test which measures their ability in reading comprehension, mathematics and vocabulary, as well as their cognitive and studying ability.[6]

Figure II: tracking through the Dutch education system for the 1996 cohort.



Note: Figure based on the cohort born in 1996. The secondary school enrollment percentages reflects enrollment four years after completing primary school. The tertiary education percentages reflect enrollment within three years after finishing the highest secondary school degree. The transition from secondary school to tertiary school is shown for the highest obtained secondary school degree of each individual. For instance, students who both obtained a middle secondary school degree and an upper secondary school degree are counted towards the upper secondary school degree group. There are some streams within the secondary schooling system (for instance, students dropping from upper secondary school to middle secondary school), which are not displayed here. In addition, some students continue with applied university after obtaining a vocational education degree or with university after obtaining an applied university degree. These streams are not displayed here either, as this usually happens outside of the 3-year window.

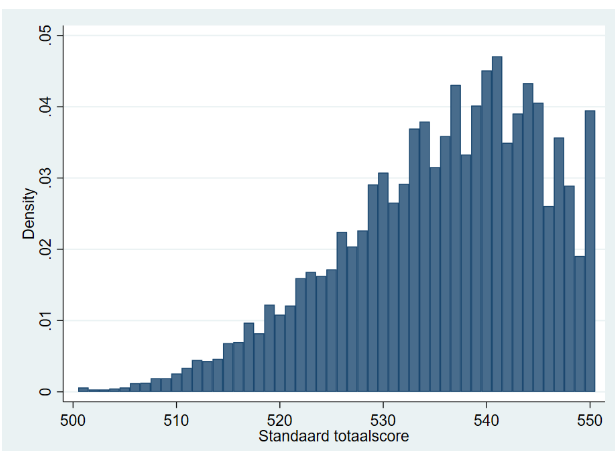
At the end of primary school tracking begins and students can enroll in three different levels of secondary school: upper, middle and lower secondary school. The levels differ both in length of study, the difficulty of the material and the access to tertiary education which a degree grants.

[6] This test consists of 200 multiple choice questions (100 on reading comprehension and vocabulary, 60 on Mathematics and 40 on studying and cognitive ability). The tests are centrally graded by the Citogroup agency, which also develops the test.

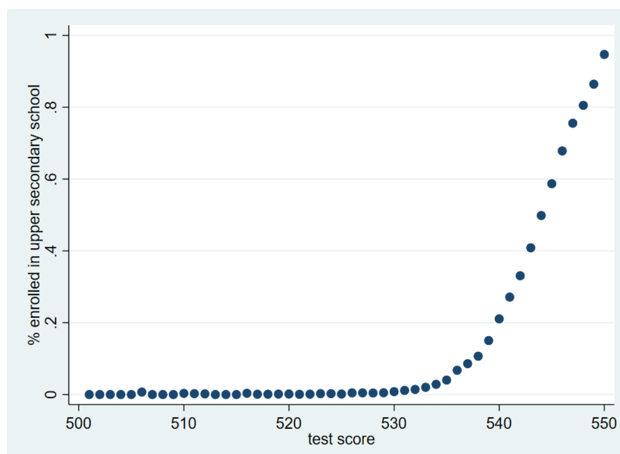
Students are free to apply to any of the three levels of secondary school after finishing primary school, but secondary schools can choose whether to accept students. During the admission decision, secondary schools primarily rely on the scores on the standardized test at the end of the primary school, as well as the recommendation of the primary school teacher. In the years 2003-2015, the score on the national standardized test was considered the leading admission criteria for secondary schools.[7]

Figure III shows the distribution of the test score for the cohort born in 1996 (panel a) as well as the probability that students with a given test-score enrolls in an upper-secondary school (panel b). The end of primary school test is highly predictive of whether a student enrolls in upper secondary school. Students who score less than 535 (which is about the median test score) are unlikely to be enrolled in an upper secondary school four years later, whereas amongst the group of students who obtain the maximum score of 550[8], 95% is enrolled at an upper secondary school four years later.[9] Figure B3 in the appendix shows how these percentages differ between urban and rural regions, providing a first indication of the differences in educational decisions.

Figure III: Frequency and probability of enrolling in upper secondary school by end-of-primary school test score



(a)



(b)

Note: Panel a shows the frequency for each of the 50 potential scores (501-550) that students could receive at the end-of-primary school test for the cohort born in 1996. Panel b displays the percentage of students that are enrolled at an upper secondary school four years after taking the test on the y-axis by the on the end-of-primary school test score.

[7] At the end of 2015, the system was reformed to make the primary school teacher's recommendation binding and secondary schools were no longer allowed to use the end-of-primary school test score. Politicians and teachers felt that the test had become the only selection criteria on which secondary schools evaluated children, which they argued put undue pressure on the children to perform at one specific moment in time.

[8] The maximum score of 550 is obtained by 5.7% of the students in the 1996 cohort

[9] The other 5 percent enrolled in a middle secondary school.

Once students enroll in a secondary school, they continue for another 4-6 years depending on the selected type of secondary school. At the end of secondary school all students take a national examination which determines if they are granted a degree. This end of secondary school test is specific for their level of secondary school.[10] Once students obtain a secondary school degree, they continue to the tertiary education.[11] The tracking system is designed such that certain levels of secondary school feed into a specific level of tertiary education. As can be seen from figure III, 84% of the students who obtain a degree of an upper secondary school are registered at a university three years after graduation. Similarly, 88% of the students who obtain a degree from a middle secondary school and 98% of the students completing the lower secondary school are respectively registered at an applied university[12] and a vocational education three years after graduation.[13] Hence, the choice that students make at the end of primary school matters greatly for their future educational prospects.

Tertiary education in the Netherlands differs from many other European countries in the sense that very few university studies have a binding constraint on the number of students they accept.[14] Students who have obtained the relevant qualifications can typically register for their study of choice at the university of their choice, without additional entry conditions. Tuition fees are around 1,800 euro per year for universities and applied universities and around 1,000 for vocational education, with governmental student loans available to meet these costs if needed. Primary schools and secondary schools are free of tuition. Primary schools and secondary schools are financed

[10] All students have a basic set of subjects (such as Dutch, English and Math A), as well as a set of subjects related to their chosen subject specialization (health, humanities, natural sciences or social sciences) in which they complete the national examination. The material covered in the exam as well as the difficulty varies depending on the type of secondary school.

[11] Upon obtaining a middle or lower secondary school degree, students with excellent grades have the opportunity to take the final two years of the next secondary school level. When discussing degree completion or conditioning on degree, these students are always classified according to the highest degree they obtained.

[12] The applied university is similar to the Fachhochschule in the German system. Applied university are more focused on practical skills compared to universities and part of the degree requirement typically involves various long term internships at companies. For illustrational purposes, individuals with a university/applied university/vocational education on average had an annual income of respectively 50.00/36.000/25.000 euro between 2007 and 2009 (CBS, 2011).

[13] These figures only count students who remained in the Netherlands after completing their secondary school. Around 2% of Dutch students register for a bachelor- or master degree abroad at some point during their education (Department of Education, 2016). As such, this small group of missing students is unlikely to be a concern for this study. As robustness check in the main analysis, I exclude all students from our analysis who are at any Dutch educational institution at least 2 out of the three years between the ages 19 and 21.

[14] The most important exceptions are medicine, veterinary medicine and dentistry, as these studies typically are expensive to offer compared to other studies.

directly by the national government, with the financing based on the number and type of students enrolled.

III.II Data

The data used in the analysis are primarily for the cohorts born between 1994 and 1998.[15] I restrict the sample to individuals born in the Netherlands and for whom both parents can be identified. For these individuals, I observe the place of residence between the ages 3 and 20, the enrollment status in all types of secondary and tertiary education and the results on the national tests at the end of primary school[16] and secondary school. In addition, a large numbers of family characteristics are available, including parental income, education, country of birth and year of birth. The summary statistics for all variables are provided in Table I.

Parental income is defined similarly as in Chetty & Hendren (2018) and consists of the sum of income for both parents when the child is between ages 14 and 18, divided by the number of parent-years with non-missing income. Some parents have a negative income, top-coded income or are missing income for more than 5 parent-years. For the parents with negative income and parents with fewer than 5 parent-year income observations, income is set to zero and a dummy is included for both groups in each regression. For top-coded incomes a dummy is also included. The robustness checks reveal that the results are not sensitive to exclusion of these (small) groups, which together contain about 1.5% of the observations. As the coverage of the income tax data may have improved slightly over the years, I allow the coefficient of parental income as well as the coefficients of the three dummy variables to vary across cohorts.

For each parent, I observe 13 possible levels of education.[17] I do not observe education for all parents, particularly for older parents who completed schooling before some of the national

[15] Some analyses can be carried out using a larger sample. For consistency, results in the main text are based on the cohorts 1994-1998 unless specified otherwise. The main tables also report results for a larger sample as robustness test whenever possible

[16] A small number of students is exempted from making the test due to disabilities and not all schools have agreed to make the test-scores available to Statistics Netherlands. Nonetheless, I observe the test-score for the large majority of the population (70%). The coefficients on the urbanization measure of columns (1) and (2) of table II and III are virtually unchanged (within 10% in all cases) when estimating it either on the full sample of individuals born between 1994-1998 or the 70% sample for whom I observe the test-score. Hence, different selection between rural and urban regions into which schools agreed to make the test scores available to Statistics Netherlands is not driving the results.

[17] These groups are kindergarten, primary school, some secondary education (low, middle or high), secondary education degree (low, middle or high), some university or applied university, applied university bachelor, university bachelor, university master or applied university master, and doctoral degree. Vocational education is not listed separately, but included in the three levels of secondary education together with the various levels of secondary school.

education registers started. Discussions with Statistics Netherlands reveal that these parents are more likely to be low educated, as highly educated parents are more likely to be included by at least one of the various educational registers. In the baseline result, missing education is included as a 14th education type. However, the results are robust to the exclusion of this group, as is shown in the main tables. To opt for an as flexible approach as possible when controlling for parental education, I include dummies for each of the 196 possible parental education combinations. Similarly, parental country of birth is grouped into "The Netherlands", "Europe" and "Non-European" for each parent, and 9 dummies are included to account for each possible parental combination. Finally, the age of the oldest parent at the time of the birth is added as additional control variable.

The next step is to construct an index of urbanization. For each individual, the zip code in which he or she resided between the years 1995-2018 is observed. The size of the zip codes is relatively small at 8km^2. I define urbanization as the log of the number of people living within 10 kilometers of the centroid of the zip code in which an individual resides (see appendix B for the details of the procedure).[18] The threshold of 10 kilometers is selected based on the fact that the majority of Dutch children travel to school by bicycle, where 10 kilometers appears a reasonable upper bound for their reach. Furthermore, the choice of a 10 kilometer radius is in line with recent studies of agglomeration economies on wages (De la Roca and Puga, 2017; Verstraten et al., 2018). Nonetheless, the number of individuals that lives within respectively 5, 10 and 20 kilometers is highly correlated[19], and as such, the exact distance cut-off has little influence on the results. As Figure IV shows, densities are the highest in the urbanized Western part of the Netherlands and are relatively low in the North and East.

[18] To avoid potential reverse causality problems, I calculate the density based on the spatial distribution of the population in 1995. As the robustness analysis reveals, the results remain virtually unchanged when I use the spatial distribution of 1840 as instrumental variable.

[19] For instance, the correlation between density within 10 kilometers and the density within 5 kilometers and the density within 20 kilometers is 0.89 in both cases.

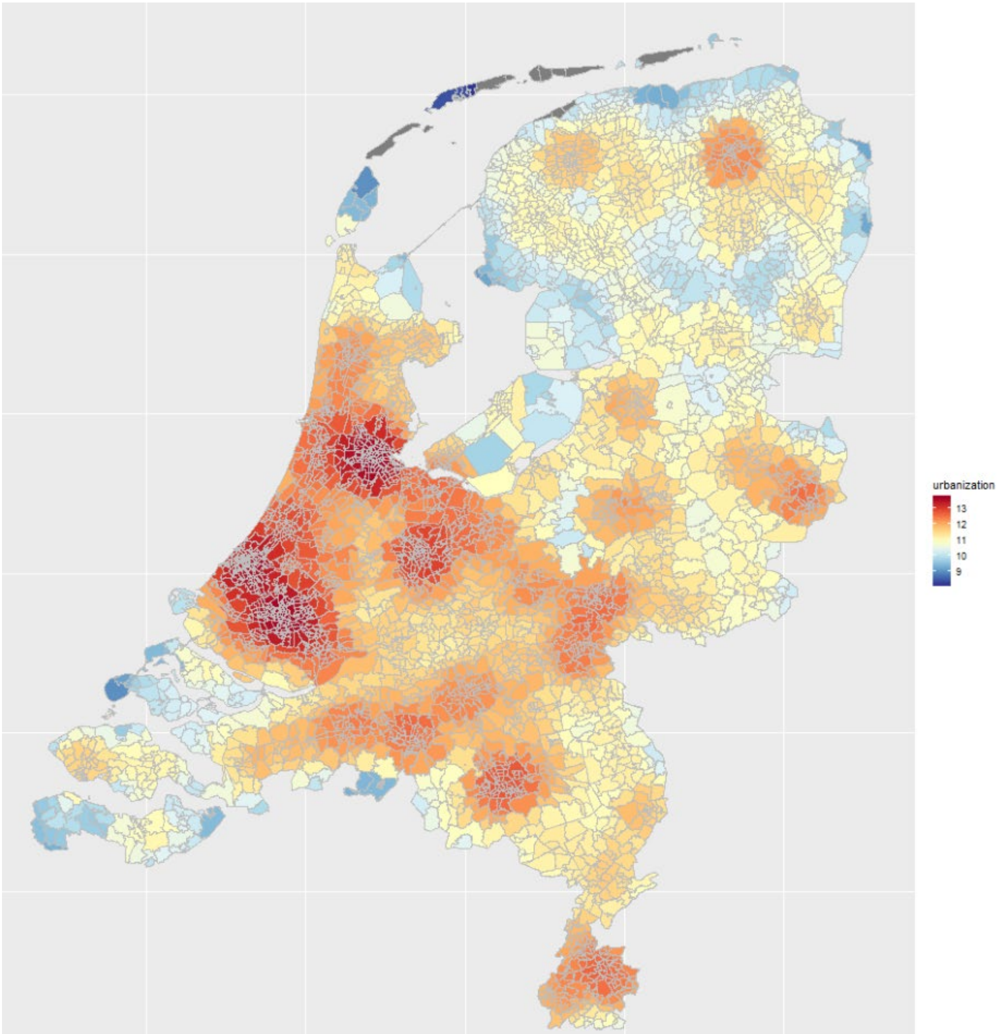
Table I: Summary statistics for individuals born in the Netherlands between the years 1994-1998.

Child characteristics	N	Mean	Sd. Dev.	p1	p99
Urbanization measured at age 11	631815	12.07	0.87	9.99	13.64
Year of birth	631815	1995.99	1.42	1994	1998
Parental characteristics					
log parental income	631890	10.36	0.83	8.82	11.68
Insufficient income data (dummy)	631890	0.001	0.03	0.00	0.00
Negative household income (dummy)	631890	0.003	0.05	0.00	0.00
Top coded incomes (dummy)	631890	0.011	0.10	0.00	1.00
Age of oldest parent at time of birth	631890	33.46	4.97	23.00	48.00
Country of birth mother (categorical)	631890	1.22	0.58	1.00	3.00
Country of birth father (categorical)	631890	1.23	0.59	1.00	3.00
Education level mother (categorical)	361625	9.49	3.45	2.00	14.00
Education level father (categorical)	340446	9.10	3.55	2.00	14.00
Secondary school enrollment and graduation variables					
End of primary school test score	631890	535.43	9.66	510.00	550.00
Upper secondary school enrollment	631890	0.23	0.42	0.00	1.00
Middle secondary school enrollment	631890	0.24	0.43	0.00	1.00
Lower secondary school enrollment	631890	0.50	0.50	0.00	1.00
Upper secondary school graduation	631890	0.19	0.39	0.00	1.00
Middle secondary school graduation[20]	631890	0.25	0.43	0.00	1.00
Lower secondary school graduation	631890	0.45	0.50	0.00	1.00
GPA upper secondary school degree	121106	6.63	0.69	5.50	8.50
Specialization track upper secondary school degree (categorical variable)	121106	4.47	2.65	1.00	10.00
Tertiary education enrollment (within 3 years of high school graduation, conditional on graduating)					
University enrollment	400305	0.20	0.40	0.00	1.00
Applied University enrollment	400305	0.29	0.46	0.00	1.00
Vocational education enrollment	400305	0.50	0.50	0.00	1.00

Note: The table shows the summary statistics for the individuals born between 1994 and 1998 in the Netherlands. Due to the privacy-sensitive nature of the microdata, it is not possible to report minima or maxima, hence the 1st and 99th percentiles values are displayed instead. As explained in the main text, education is not available for all parents. Secondary school enrollment is measured 4 years after completing primary school. The GPA and specialization track of the upper secondary school degree are only observed for the students who graduated from upper secondary school. Tertiary education enrollment is based on cohorts born in 1994-1996, as for the cohorts born in 1997/1998 not enough time has passed for all students to measure. Tertiary education enrollment is defined as the highest enrollment within 3 years of high-school graduation.

[20] The graduation rate of middle secondary school is higher than enrollment due to the fact that about 10% of the children enrolled in an upper secondary school drop down to a middle secondary school during their final three years in secondary school (see also section V).

Figure IV: Urbanization measure per zip code



Note: Map displays the log of number of individuals within 10 kilometers, based on the 1995 population distribution. For graphical purposes, areas with an urbanization grade less than 9 are grey (containing 230 out of the 631.890 observations in the baseline sample). Appendix B provides details on the construction of the urbanization measure.

IV Empirical approach

The Dutch educational system provides two key decisions moments which can be exploited to analyze the impact of urbanization on educational outcomes. The first decision moment is at the end of primary school, when children select one of the three levels of secondary school. At this point, I observe the cognitive ability of the child as measured by the score on the national test, as well as a wide range of family characteristics. This allows me to study whether, conditional on observed academic ability and parental background, children who grow up in urban areas make different educational choices compared to children who grow up in rural environments. The estimating equation is displayed in equation (1), where the family characteristics contain the variables discussed in section III.II and dummies are included for all possible primary school test scores. The educational outcome measure is whether a child attends the highest level of secondary school (VWO in Dutch) 4 years after graduation from primary school.[21] The log-linear relationship between density and educational outcomes is well supported by the data, as will be shown in section V.II.

$$P(attends\ upper\ secondary\ school)_{it} = \beta_1 * Child\ characteristics_i + \beta_2 * Family\ characterstics_i + \beta_3 * f(primary\ school\ test\ score_i) + \beta_4 * Urbanization_i + \gamma_t + \varepsilon_{it} \quad (1)$$

The second decision moment is at the end of secondary school. Here I focus on the group of students who obtained a degree of an upper secondary school, which provides access to the large majority of university-subject combinations without further conditions as described in section III.II. Hence, equation (2) estimates the effect of density on the probability of attending an academic university with three years after graduating from high school[22], conditional having obtained a degree from an upper secondary school, high school GPA and family characteristics.[23]

$$P(attends\ university|upper\ secondary\ school\ degree)_{it} = \beta_1 * Child\ characteristics_i + \beta_2 * Family\ characteristics_i + \beta_3 * f(GPA\ high\ school_i) + \beta_4 * Urbanization_i + \gamma_t + \varepsilon_{it} \quad (2)$$

Under the assumption that the covariance between ε_{it} and the urbanization measure is zero, equations (1) and (2) will correctly identify the effect of growing up in an urban environment on educational outcomes. One concern is that parents move to places best fitted to realize the potential outcomes of their children, in which case the estimates of β_4 would be biased. However, I find no evidence of this in the Dutch setting. For the cohorts born between 1994 and 1998, only 5.4% of the families with children between the ages 6 and 17 make a substantial move (more than 20 kilometers), thus limiting the degree to which families respond to the realized potential of their children. In addition, within the group of children who move, there is no correlation between the

[21] The reason for analyzing the enrollment four years after completing primary school is that some secondary schools offer joint upper- and middle secondary school classes for the first one or two years of secondary school. To allow in addition for students that get ill or have to repeat a year, I analyze school enrollment in the fourth year after completing primary school.

[22] The reason for allowing a three year lag is that some students opt for a pause between completing secondary school and starting tertiary education.

[23] While it would also be possible to analyze where graduates of other secondary school levels end up, in practice this is more difficult as the alternative education is not quite clear. The majority of students with an upper secondary school degree end up at either a university or an applied university, whereas graduates of the middle secondary school end up at a far more diverse set of tertiary educational institutes (see figure II).

change in density and the observed academic ability of the children.[24] Nonetheless, I show that the results of equation (1) and (2) are robust to estimating it on either the full-sample or on the group of children who do not move across municipalities between the ages 6-17.

A second concern might be that parental characteristics vary between urban and rural places in ways which are not fully captured by the control variables. For instance, it might be that parents in urban areas are more ambitious than parents in rural areas, and that this is not fully captured by the differences in wages. Separating spatial sorting from area effects has been a key challenge in the urban literature and solutions to this have typically relied heavily on individual-fixed effects (see for instance Combes et al., 2008 and De la Roca and Puga, 2017). However, such approaches are typically not possible in the case of educational outcomes, due to the lack of consistent panel data on national educational outcomes of children.[25][26] Section V.II explores this potential threat to identification, using the methodology of Altonji et al. (2005) and Oster (2017) to assess the potential importance of sorting on unobserved characteristics.

V Results

Section V.I presents the baseline results of equations (1) and (2). Section V.II discusses the robustness of the baseline results to the functional form assumption, inclusion of region fixed effects, the endogeneity of the urbanization measure and sorting on unobserved characteristics.

V.I Baseline results

Table II presents the results of equation (1), analyzing how the decision of children to enroll in an upper secondary school depends on population density. Column (1) shows the regression when only the urbanization measure is included, whereas column (2) adds child- and family

[24] Result not included in the paper but available on request

[25] Such approach is difficult to implement for educational outcomes, as most countries lack repeated outcomes for children on a national level, which would allow fixed-effect estimations similar to Combes et al. (2008) with grades instead of wages. However, such individual-fixed effect approaches have been successfully applied within schools in the economics of education literature to identify effects of school level variables, such as teacher quality (see for instance Chetty et al., 2014b).

[26] An alternative way to estimate the effect of urbanization on educational outcomes would be to utilize the identification strategy of Chetty & Hendren (2018) and use differences in age between children who move across regions for identification of neighborhood effects. However, appendix D shows that the identifying assumption of Chetty & Hendren (2018), that the age of children at the time of the move is uncorrelated with potential outcomes, does not hold in the Dutch context.

characteristics and column (3) adds dummies for the end of primary school test scores. The coefficient of urbanization is statistically significant in all specifications at the 1% level. In economic terms, the baseline result of column (3) indicates that an increase in density by one log-point (which coincidentally is close to one standard deviation, see table I) is associated with a 1.7 percentage point increase in the probability that a child attends upper secondary school. Given that the mean percentage of children attending the upper secondary school is about 23%, this increase is substantial. When comparing the individual columns, the coefficient drops somewhat when family-characteristics are added, which is largely driven by the differences in parental education between urban and rural regions. The coefficient declines further when the test scores are added, which might reflect a lower initial academic ability of children in rural regions, or alternatively might be the effect of overcontrolling to the degree that growing up in an urban environment may also affect end-of-primary school test scores.

Columns 4-7 test the robustness of the results by excluding various subgroups. Column (4) excludes all children who are not enrolled in one of the three secondary school levels, as these may be enrolled in special needs schools or schools across the border with Germany/Belgium. Column (5) excludes children who moved between municipalities during their school-going age, as their families may have sorted themselves into places based on the potential outcomes of their children. Column (6) changes the sample and estimates the effect instead on the cohorts born between 1999 and 2002, for whom information is also available. Finally, column (7) reduces the sample to the individuals for whom the exact education level of both parents is known. The coefficient is very stable across the various subgroups, suggesting that these groups are not driving the results.

One concern might be that that children in urban environments are enrolled into classes which are too difficult for their level of ability, resulting in higher drop-out rates in urban areas.[27] Around 21% of the students who are enrolled in an upper secondary school four years after finishing primary school eventually do not obtain an upper secondary school degree, which is far from

[27] Alternatively, one may also worry that children in rural areas obtain a degree from a middle secondary school, then continue with some years of applied university before switching to university. In this case, children from rural areas would simply follow a different track to end up at university. I do find some evidence for this, as children who obtain a middle secondary school degree are slightly more likely to enroll in university in the seven years after graduating from middle secondary school if they grow up in a rural area rather than an urban area. However, the increased usage of this alternative route in rural areas is far too small to compensate for the negative effects described in table II and IV. Overall, a one log point increase in density reduces the probability that a student enrolls for university by 0.6% amongst the group of middle secondary school graduates (25% of the sample, see table I). Hence, taking this path to university into account reduces the overall effect of density on university enrollment by 0.25*0.6 = 0.15%. Hence, this alternative route can compensate for about 10% of the total effect of density on university enrollment of 1.4%.

negligible.[28] To control for the possibility that differences in dropout rates between urban and rural areas are driving the results, table III instead analyzes the probability that a student obtains a degree from an upper secondary school. The results are very similar to table II, indicating that a misallocation of students at the end of primary school in urban communities is unlikely to drive the results. Furthermore, Table C1 in the appendix shows the results when directly analyzing drop-out rates, by estimating the probability that a student obtains a degree from an upper secondary school, conditional on being enrolled in an upper secondary school four years after finishing primary school. The results indicate that children in urban areas are actually slightly less likely to drop out from an upper secondary school, again showing that differential dropout rates between urban and rural areas are not driving the results of table II.

The second key educational decision moment is after children obtain an upper secondary school degree, when they have to decide on the level of tertiary education. Table IV shows the estimates for equation (2), analyzing the effect of urbanization on the probability that a student enrolls at an university within 3 years of obtaining an upper secondary school degree, which provides access to the large majority of university-subject combinations. Column (1) again only includes the urbanization measure, whereas column (2) adds student and family characteristics and column (3) adds controls for the specialization track and the GPA on the national examination at the end of upper secondary school[29]. As I condition on the national end of high school examination scores (instead of end of primary school test scores), it is possible to use a larger sample and to also include the cohorts born between 1989 and 1993.[30]

[28] The majority (62%) of these students instead obtains a degree from a middle secondary school. The drop-out figure of 21% is in line with the statistics reported for this period by Statistics Netherlands (2019).

[29] The GPA is based on the national examinations at the end of high school. In the final two years of secondary school, students can enroll in one of the possible 4 tracks (humanities, social sciences, biology or natural sciences). The track determines the subjects in which the students take their final examinations.

[30] For the students born in 1994-1998, it is possible to include both the secondary school test scoreg as well as end of primary school test score. However, adding the end of primary test score results offers little explanatory power over the high school GPA and leaves the coefficient virtually unchanged. Hence, the loss of this variable due is more than outweigh by the higher precision obtained due to the larger sample.

Table II: effect of urbanization on probability of enrolling in upper secondary school

	Baseline Estimates			Sensitivity analysis				Area fixed effects		IV
	Ind. level controls	Ind. + Fam. controls	Ind. + Fam controls + test-score	Children observed in any education	Movers Excluded	Different cohorts (1999-2002)	Full parental education	Municipality-Fixed effects	Provincial -fixed effects	IV-estimates
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Urbanization at age 11	0.0342*** (0.0067)	0.0260*** (0.0037)	0.0167*** (0.0017)	0.0169*** (0.0018)	0.0180*** (0.0019)	0.0167*** (0.0020)	0.0141*** (0.0015)	0.0123** (0.0045)	0.0196*** (0.0020)	0.0178*** (0.00064)
Individual controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Test score-dummies	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.005	0.13	0.50	0.50	0.50	0.52	0.53	0.51	0.50	0.50
No. of obs.	631.731	631.731	631.731	624.008	554.826	503.200	242.156	631.731	631.731	628.396

Note: All results apart from column (6) based on individuals born between 1994 and 1998. Dependent variable is whether a child is enrolled in an upper secondary school 4 years after completing primary school. Individual controls include a gender-dummy and cohort-dummies. Family controls include dummies for parental education combinations (168 dummies), parental nationality combinations (9 dummies), year-of-birth of oldest parent (40 dummies), family income (interacted with cohort fixed effects), low-income dummy (interacted with cohort-fixed effects), top-coded income dummy (interacted with cohort-fixed effects) and dummy for insufficient income data (interacted with cohort fixed effects). For a detailed explanation of the construction of the urbanization measure or the family controls, see section III. Column 4 excludes children who are not observed in any school 4 years after completing primary school. Column 5 excludes all children who moved municipalities during school going age. Column 6 instead estimates the model on the cohorts born between the years 1999 and 2002. Column 7 removes all parents for whom uncertainty exists about the education level of one or both parents. Column 8 adds municipality-fixed effects for the municipality in which children live at age 11 (430 dummies), whereas column 9 adds province fixed-effect for the province in which children live at age 11 (12 dummies). Column (10) instruments for modern densities (based on 1995 population distribution, see Appendix B) with historical densities. For all columns apart from column 10 the standard errors are clustered on the municipality level. Column 1-9 are estimates by OLS, column 10 by 2SLS. First stage results of column 10 are reported in Appendix C3.

Table III: effect of urbanization on probability of obtaining a degree in upper secondary school,

	Baseline Estimates			Sensitivity Analysis			Area-fixed Effects		IV
	Ind. level controls	Ind. + Fam. controls	Ind. + Fam controls + test-score	Children observed in any education	Movers Excluded	Full parental education	Municipality-Fixed effects	Provincial-fixed effects	IV-estimates
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Urbanization at age 11	0.0296***	0.0226***	0.0146***	0.0148***	0.0156***	0.0130***	0.0113**	0.0153***	0.0158***
	(0.0060)	(0.0030)	(0.0013)	(0.0014)	(0.0014)	(0.0013)	(0.0039)	(0.0016)	(0.00062)
Individual controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Test score-dummies	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.004	0.12	0.44	0.45	0.45	0.47	0.45	0.44	0.44
No. of obs.	631.731	631.731	631.731	624.008	554.826	242.156	631.731	631.731	628.396

Note: All results based on individuals born between 1994 and 1998. Dependent variable is whether a child graduates from an upper secondary school. Individual controls include a gender-dummy and cohort-dummies. Family controls include dummies for parental education combinations (168 dummies), parental nationality combinations (9 dummies), year-of-birth of oldest parent (40 dummies), family income (interacted with cohort fixed effects), low-income dummy (interacted with cohort-fixed effects), top-coded income dummy (interacted with cohort-fixed effects) and dummy for insufficient income data (interacted with cohort fixed effects). For a detailed explanation of the construction of the urbanization measure or the family controls, see section III. Column 4 excludes children who are not observed in any school 4 years after completing primary school. Column 5 excludes all children who moved municipalities during school going age. Column 6 removes all parents for whom uncertainty exists about the education level of one or both parents. Column 7 adds municipality-fixed effects for the municipality in which children live at age 11 (430 dummies), whereas column 8 adds province fixed-effect for the province in which children live at age 11 (12 dummies). Column (9) instruments for modern densities (based on 1995 population distribution, see Appendix B) with historical densities. For all columns apart from column 9 the standard errors are clustered on the municipality level. Column 1-8 are estimates by OLS, column 9 by 2SLS. First stage results of column 9 are reported in Appendix C3.

Table IV: effect of urbanization on probability of enrolling at university/applied university, conditional on having an upper secondary school degree.

	Baseline Estimates			Sensitivity Analysis				Area-fixed Effects		IV
	Ind. level controls	Ind. + Fam. controls	Ind. + Fam controls + GPA	Children observed in any education	Movers Excluded	Cohorts of table II (1994-1998)	Full parental education	Municipali ty-Fixed effects	Provincial-fixed effects	IV-estimates
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Urbanization at age 11	0.0232***	0.0083***	0.0080***	0.089***	0.0084***	0.0087***	0.0072***	0.0098*	0.0135***	0.0071***
	(0.0022)	(0.0016)	(0.0017)	(0.0017)	(0.0018)	(0.0019)	(0.0018)	(0.0047)	(0.0022)	(0.00011)
Individual controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
GPA and specialization track	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.003	0.03	0.06	0.07	0.07	0.06	0.06	0.07	0.07	0.06
No. of obs.	289.109	289.109	289.109	281.179	253.430	139.559	110.921	289.109	289.109	287.763

Note: All results apart from column (6) based on individuals born between 1989 and 1998. Dependent variable is whether a child attends university within 3 years of graduating from upper secondary school. Individual controls include a gender-dummy and cohort-dummies. Family controls include dummies for parental education combinations (168 dummies), parental nationality combinations (9 dummies), year-of-birth of oldest parent (40 dummies), family income (interacted with cohort fixed effects), low-income dummy (interacted with cohort-fixed effects), top-coded income dummy (interacted with cohort-fixed effects) and dummy for insufficient income data (interacted with cohort fixed effects). For a detailed explanation of the construction of the urbanization measure or the family controls, see section III. Column 4 excludes children who are not enrolled in any form of tertiary education three years after graduating. Column 5 excludes all children who moved municipalities during school going age. Column 6 limits the sample to the cohorts born between 1994-1998, in line with the baseline sample of table II. Column 7 removes all parents for whom uncertainty exists about the education level of one or both parents. Column 8 adds municipality-fixed effects for the municipality in which children live at age 11 (430 dummies), whereas column 9 adds province fixed-effect for the province in which children live at age 11 (12 dummies). Column 10 instruments for modern densities (based on 1995 population distribution, see Appendix B) with historical densities. Columns 1-9 are estimated by OLS, column 10 by 2SLS. First stage results of column 10 are reported in Appendix C3. For all columns apart from column 10 the standard errors are clustered on the municipality level.

The results in table IV indicate that children who grow up in urban areas, conditional on their academic ability and family characteristics, are more likely to enroll at a university. The preferred specification in column (3) indicates that children who grow up in an area where log density is one-point higher (about one standard deviation) are 0.8 percentage points more likely to enroll in university, from a base of 84%. The coefficient remains very similar when excluding the small group of children who are not observed in any tertiary education (column 4), excluding children who moved between municipalities during their childhood (column 5), limiting the sample to the cohorts born between 1994 and 1998 (column 6) or limiting the sample to the children for whom no uncertainty exists over parental education (column 7).

The next question that arises is which alternatives these children select when they decide not to attend university. One alternative might be to instead enroll at the lower rated applied universities. As figure II shows, about 16% of the students who obtained an upper secondary school degree decide to enroll at an applied university, despite having the necessary qualifications to start at a university. Table C2 in the appendix shows the effect of density on the probability that a student enrolls at an applied university, conditional having obtained a degree from an upper secondary school. The coefficient is significant and negative, indicating that students growing up in rural communities are more likely to instead enroll at an applied university. The coefficients are nearly identical to the coefficients in table IV, suggesting that diversion of rural students into the lower rated applied universities fully explains the lower enrollment rates at universities found in table IV.

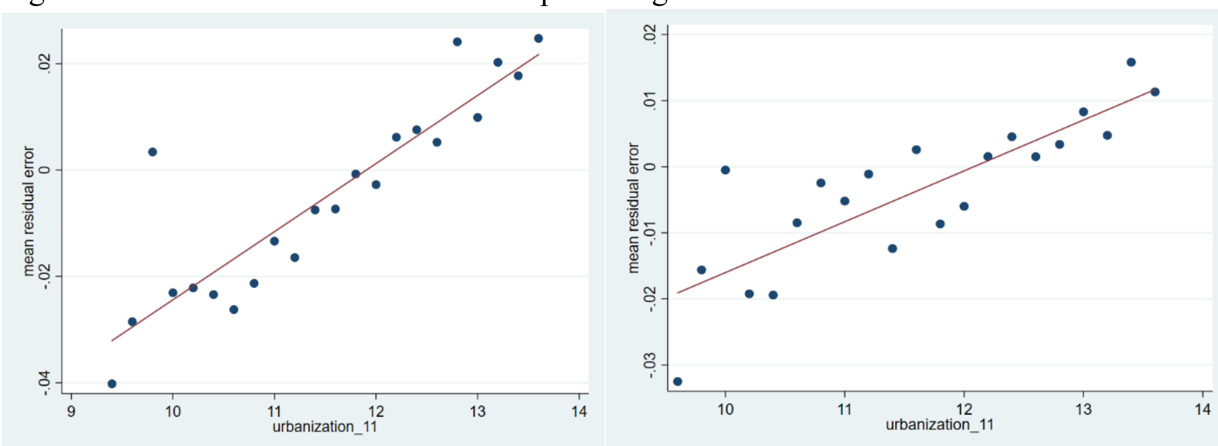
V.II Robustness

The results in section V.I indicate that population density may play a role in the educational decisions of children. Naturally, some concerns arise over the interpretation and the robustness of this result. This section will discuss four potential threats to identification: the log-linear functional form assumption, the influence of more general regional differences, the endogeneity of population density and sorting on unobserved characteristics.

1. *Functional form assumption*

All estimations so far have been based on the assumption of a log-linear relationship between population density and educational outcomes, which may or may not accurately represent the true functional form. Should the functional form be wrongly specified, it may bias the coefficients. In order to test this possibility, figure V plots the residuals of the baseline estimations (columns 3 in table II and table IV) without the urbanization measure against the urbanization measure itself. The residuals are averaged over 0.2 log point intervals. As figure V shows, the log-linear functional form assumption seems to fit the data quite well and hence is unlikely to bias the results. Furthermore, the residuals reveal that the results do not depend on any specific part of the density distribution: the log-linear relationship seems to describe the actual relationship well throughout the observed density distribution.

Figure V: Residuals of baseline estimates plotted against urbanization measure



Note: Average residuals of columns (3) of table II (left) and table IV (right) without the urbanization measure plotted against the urbanization measure. Each dot corresponds to the average residual in a 0.2 log-point bin. Bins containing fewer than 500 observations are not depicted.

2. *Influence of broad regional differences*

Second, the map of the urbanization measure displayed in figure IV shows that urbanization is highest in the West and relatively low in the North and East of the country. As a result, the urbanization measure may at least partially reflect broader (economic or cultural) regional differences within the country, rather than urbanization per se. To test this possibility, columns (8) and (9) in tables II - IV add respectively provincial and municipality fixed effects. The coefficient of the urbanization measure is in this case identified on the variation in population density within provinces and municipalities respectively. However, the coefficients changes relatively little and remain statistically significant in all cases, even though the standard errors increase substantially as most of the variation in the urbanization measure is discarded. Hence, the results are not driven by broad regional differences.

3. *Endogeneity of Urbanization*

Third, one might be concerned that population density itself is endogenous. Factors that attract population to certain areas and hence contribute to city formation may also directly affect educational outcomes. Furthermore, reverse causality can play a role if individuals migrate to areas with good schools or favorable schooling policies. Combes et al. (2008) highlight the potential of such contemporary factors to bias estimations in the case of the urban wage premium and use historical densities as instrumental variable. To alleviate concerns that the results in this paper are driven by the endogeneity of the urbanization measure, I follow Combes et al. (2008) and instrument current population densities with the population density measured based on the 1840 Dutch census.[31][32] The details of this procedure and a map of the population densities of 1840 are contained in appendix C3. Even though it is not a perfect historical instrument, as 4 of the 13 Dutch universities predate 1840, it should substantially reduce any bias rising from the endogeneity of the urbanization measure. The first stage is significant as shown in Appendix C3, which indicates that the instrument is relevant. The result of the 2SLS estimations are provided in column (10) in tables II-IV. In all three cases, the coefficient hardly changes when using the 2SLS estimator. Hence, the endogeneity of urbanization is not driving the results.

4. *Sorting on unobservables*

Finally, a key concern for many urban and regional studies is the spatial sorting of households on unobserved variables. Despite observing some of the key factors of importance for educational decisions, such as parental education and cognitive ability of the child, some factors such as ambition and non-cognitive skills remain unobserved. One method to assess the importance of such sorting on unobserved variables is provided by Altonji et al. (2005) and Oster (2017). Their procedure relies on the assumption that the sorting on observed variables is informative of sorting on unobserved variables, and in particular that the sorting on unobserved characteristics is weakly less than the sorting on observed characteristics. As such, the degree to which the coefficient of interest and the R^2 change when control variables are added can provide an indication to the potential importance of omitted variable bias.

Altonji et al. (2005) and Oster (2017) argue that sorting on unobserved variables in most cases is less severe than sorting on observable characteristics for two reasons. First of all, Altonji et al.

[31] I would like to express my gratitude to Paul Verstraten (CPB Netherlands Bureau for Economics Policy Analysis) for providing the shape files containing the boundaries of the 1840 municipalities as well as the digitalized 1840 census.

[32] A small number of zip-codes cannot be matched to the densities of 1840, as they are located on land that has been reclaimed from the sea after 1840. Nonetheless, the instrument is available for 99.6% of the observations.

(2005) argue that in the most extreme case, researchers observe a random set of variables since they typically have no direct influence over the data collection in surveys and administrative data. In such cases, the observed variables are a random subset, and as such, sorting on observed and unobserved variables should be equally important in generating bias. However, Altonji et al. (2005) argue that researchers typically direct their effort towards obtaining and including the control variables which are seen as being most likely to generate omitted variable bias. Hence, the included control variables in all likelihood contribute more to the omitted variable bias compared to the unobserved variables. For this reason, both Altonji et al. (2005) and Oster (2017) argue that a reasonable upper bound for the degree of selection on unobserved variables is the selection on observed variables. Furthermore, Altonji et al. (2005) argue that when there is a lag between observing the explanatory variables and the outcome measures, any idiosyncratic shocks that occurs between observing the explanatory variables and the outcome measure cannot bias effect of the predetermined explanatory variable. Hence, to the extent that (unobserved) idiosyncratic shocks are present, they provide an additional reason why sorting on unobserved variables may be less important than sorting on observed variables in generating omitted variable bias.

A key decision when applying the methodology of Altonji et al. (2005) and Oster (2017) is the choice for the upper bound on R^2_{max}, i.e. how much of the remaining variance the model variables which are unobserved by the researcher would explain. Oster (2017) argues that an upper bound of 1 on the R^2 is too restrictive, as measurement error and the true idiosyncratic error term contained in most models would prevent the researcher from reaching an R^2 of 1, even in cases where all relevant variables are observed. Instead, based on a simulation exercise, Oster (2017) suggests using an R^2 1.3 times larger than the R^2 of the regression with full controls, or to find a reasonable upper bound based on earlier research. For instance, in case of child outcomes, Oster (2017) argues that sibling correlations may provide a reasonable upper bound for the explanatory power of environmental and family characteristics. Given the decision with respect to the R^2_{max}, it is then possible to rescale the coefficient to account for the sorting on unobserved variables, using the simplified estimator provided in Oster (2017)[33]:

[33] The reported coefficients in table V are based on the general estimator as reported in section 3.2 of Oster (2017), which allows for a more general covariance structure of the unobserved variable with the observed explanatory and outcomes variables. However, the adjusted coefficients are virtually identical when using the simplified and general estimator in this study.

$$\beta^* = \tilde{\beta} - [\dot{\beta} - \tilde{\beta}] \frac{R_{max} - \tilde{R}}{\tilde{R} - \dot{R}} \quad (3)$$

where $\dot{R}$ and $\dot{\beta}$ are obtained from the short regression of density on educational outcomes reported in columns (1) of table II and IV and $\tilde{R}$ and $\tilde{\beta}$ are obtained from the full regression model specified in columns (3) of table II and IV.

Table V shows the adjusted coefficients of table II and IV when applying the correction for unobserved variables of Oster (2017), under assumption that selection on observed variables is equal to selection on unobserved variables. The upper half of table V displays the original coefficients and R^2's taken from tables II and IV, which are inserted into equation (3). The second half of table V provides the adjusted coefficients obtained from equation (3) under two different sets of R^2_{max}. In the case of the choice for secondary school level (table II), the coefficient remains statistically significant and economically relevant when adjusting for the omitted variable bias, both when using the R^2 based on sibling correlations in the data as well as Oster's suggested value of $1.3 * \tilde{R}$. The second column of table V adjusts the coefficients based on university enrollment (table IV). In this case, the sibling correlations are hard to interpret as they are only available for the small subset of siblings where both children graduate from an upper secondary school. When the coefficient is adjusted for omitted variable bias using Oster's suggestion of $1.3 * \tilde{R}$, the coefficient diminishes substantially, but remains positive. Taken together, the results in table V indicate that under the assumption that selection on unobserved variables is weakly less than selection on observed variables, omitted variable bias cannot explain the majority of the effect of population density on educational outcomes.

Table V: Importance of selection on unobserved variables

Original coefficients taken from table II/IV	Table II	Table IV
R^2 of short regression (column 1 in table II/IV): ($\dot{R}$)	0.005	0.003
Coefficient of short regression (column 1 in table II/IV): ($\dot{\beta}$)	0.0342***	0.0232***
R^2 of long regression (column 3 in table II/IV): ($\tilde{R}$)	0.502	0.064
Coefficient of long regression (column 3 in table II/IV): ($\tilde{\beta}$)	0.0167***	0.0080***

Coefficient adjusted for unobserved variables following Oster (2017)	Table II	Table IV
R_{max} obtained from sibling regressions	0.54	-
Adjusted coefficient (β^*) using R_{max} based on sibling regressions:	0.0153***	-
R_{max} as $1.3 * \tilde{R}$	0.65	0.083
Adjusted Coefficient (β^*) using R_{max} of $1.3 * \tilde{R}$	0.0106***	0.0028

Note: Adjustment for unobserved variables based on Oster (2017). The R^2 based on sibling regression comes from a regression of the model in table II, column III on the subset of siblings, with the secondary educational choice of the sibling added as control variable. To provide a conservative inference, the standard errors of table II/IV have been used to calculate significance.

VI Discussion and conclusion

The results in this paper show evidence that population density affects educational investment decisions of children in the context of The Netherlands. Conditional on observed ability and parental background, children who grow up in urban areas consistently choose to invest more in their education compared to children who grow up in more rural environments. This result is robust across various specifications, subgroups and spatial scope, and cannot be accounted for by the endogeneity of the urbanization measure or sorting on unobserved variables under the assumptions of Oster (2017). Taken together, the results imply that conditional on family characteristics and academic ability, a one log-point increase in density is associated with a 1.4 percentage point increase in the probability that a child attends university, from a base of 20%, which implies an elasticity of 0.07.[34] Translated into standard deviations, it implies that a one standard deviation increase in population density is associated with a 1.25 percentage point increase in the probability that a child attends university.

It is surprising that the potential for agglomeration economies to affect human capital decisions of children and young adults has not received more attention in the literature, particularly given the potential long-term effects on regional growth. Taken together, the results in this paper contribute to the literature in three ways. First of all, the paper shows that the observed urban-rural education gaps might reflect more than just spatial sorting by households. Children who grow up in cities consistently select higher levels of human capital accumulation, conditional on household characteristics and cognitive ability. As such, the paper extends the findings of De la Roca and Puga (2017) by showing that density does not only matter for human capital accumulation during working life, but that effects are already visible prior to labour market entry. Second, the results highlight an important channel that may inhibit convergence between rural and urban regions in the long run. Rural regions are not only endowed with lower initial levels of human capital, but in

[34] 1.23% of this increase is due to the increase 1.46% in probability that a child graduates from upper secondary school as seen in table III, taking into account that on average 84% continues with university (see Figure 1). A further 0.15% of the increase is due to the 0.8% increase in university attendance amongst the children who complete upper secondary school, which contains about 19% of the population (see table 1).

addition seem to be at a disadvantage in the production of human capital. This is noteworthy as the human capital stock appear the main driver of economic growth on a regional level (Gennaioli et al. 2014). Finally, the paper contributes to a rapidly growing literature highlighting the importance of the place in which children grow up on later life outcomes. Specifically, I show that children in rural communities do not seem to enjoy or take the same educational opportunities as children who grow up in urban communities, even conditioning on cognitive ability and household characteristics. Hence, the differences in educational attainment between urban and rural communities observed in a wide range of countries may reflect more than just the spatial sorting of households.

Finally, one question which remains is what mechanisms drive the increased human capital accumulation of children in more urban environments. Section II.I and II.II indicated that based on existing literature, one would expect the agglomeration mechanisms to play a role in the early-life human capital decisions, either through an increase in the returns to education in cities or by increasing the availability of schooling and by reducing the commuting or moving costs to attend further education. Furthermore, there might also be a role for network effects as highlighted in section II.III. However, a detailed exposition of a specific mechanism in the spirit of De La Roca and Puga (2017) or Dauth et al. (2018) is beyond the scope of this paper and will be left for future research.

VIII References

Adukia, A., Asher, S., & Novosad, P. (2020). Educational investment responses to economic opportunity: evidence from Indian road construction. American Economic Journal: Applied Economics, 12(1), 348-76.

Alesina, A., Hohmann, S., Michalopoulos, S., & Papaioannou, E. (2019). Intergenerational Mobility in Africa (No. w25534). National Bureau of Economic Research.

Altonji, J. G., Elder, T. E., & Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. Journal of political economy, 113(1), 151-184.

Andersson, F., Burgess, S., & Lane, J. I. (2007). Cities, matching and the productivity gains of agglomeration. Journal of Urban Economics, 61(1), 112-128.

Autor, D. (2019). Work of the past. Work of the future. American Economic Association: Papers and Proceeding.

Bau, N. 2019. 'Estimating an Equilibrium Model of Horizontal Competition in Education'. London, Centre for Economic Policy Research

Baum-Snow, N., Freedman, M., & Pavan, R. (2018). Why has urban inequality increased? American Economic Journal: Applied Economics, 10(4), 1-42.

Bryan, G., & Morten, M. (2019). The aggregate productivity effects of internal migration: Evidence from indonesia. Journal of Political Economy, 127(5), 2229-2268.

Burgess, S., Greaves, E., & Vignoles, A. (2019). School choice in England: evidence from national administrative data. Oxford Review of Education, 1-21.

Card, D. (1993). Using geographic variation in college proximity to estimate the return to schooling (No. w4483). National Bureau of Economic Research.

Chetty, R., Hendren, N., Kline, P., & Saez, E. (2014a). Where is the land of opportunity? The geography of intergenerational mobility in the United States. The Quarterly Journal of Economics, 129(4), 1553-1623.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014b). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. American Economic Review, 104(9), 2593-2632.

Chetty, R., & Hendren, N. (2018). The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects. The Quarterly Journal of Economics, 133(3), 1107-1162.

Chetty, R., Hendren, N., & Katz, L. F. (2016). The effects of exposure to better neighborhoods on children: New evidence from the Moving to Opportunity experiment. American Economic Review, 106(4), 855-902.

Cohen-Zada, D., & Sander, W. (2008). Religion, religiosity and private school choice: Implications for estimating the effectiveness of private schools. Journal of Urban Economics, 64(1), 85-100.

Combes, P. P., Duranton, G., & Gobillon, L. (2008). Spatial wage disparities: Sorting matters! Journal of urban economics, 63(2), 723-742.

Combes, P. P., & Gobillon, L. (2015). The empirics of agglomeration economies. In Handbook of regional and urban economics (Vol. 5, pp. 247-348). Elsevier.

Dauth, W., Findeisen, S., Moretti, E., & Suedekum, J. (2018). Matching in cities (No. w25227). National Bureau of Economic Research.

D'Costa, S., & Overman, H. G. (2014). The urban wage growth premium: Sorting or learning?. Regional Science and Urban Economics, 48, 168-179.

Deutscher, N. (2020). Place, Peers, and the Teenage Years: Long-Run Neighborhood Effects in Australia. American Economic Journal: Applied Economics, 12(2), 220-49.

Duranton, G., & Puga, D. (2004). Micro-foundations of urban agglomeration economies. In Handbook of regional and urban economics (Vol. 4, pp. 2063-2117). Elsevier.

Frenette, M. (2006). Too far to go on? Distance to school and university participation. Education Economics, 14(1), 31-58.

Gennaioli, N., La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2012). Human capital and regional development. The Quarterly journal of economics, 128(1), 105-164.

Gennaioli, N., La Porta, R., De Silanes, F. L., & Shleifer, A. (2014). Growth in regions. Journal of Economic growth, 19(3), 259-309.

Glaeser, E. L., Kolko, J., & Saiz, A. (2001). Consumer city. Journal of economic geography, 1(1), 27-50.

Gould, E. D. (2007). Cities, workers, and wages: A structural analysis of the urban wage premium. The Review of Economic Studies, 74(2), 477-506.

Groot, S. P., de Groot, H. L., & Smit, M. J. (2014). Regional wage differences in the Netherlands: Micro evidence on agglomeration externalities. Journal of Regional Science, 54(3), 503-523.

Hoekstra, M. (2009). The effect of attending the flagship state university on earnings: A discontinuity-based approach. The Review of Economics and Statistics, 91(4), 717-724.

Hoxby, C. M., & Avery, C. (2012). The missing" one-offs": The hidden supply of high-achieving, low income students (No. w18586). National Bureau of Economic Research.

Hoxby, C., & Turner, S. (2013). Expanding college opportunities for high-achieving, low income students. Stanford Institute for Economic Policy Research Discussion Paper, (12-014).

Hsieh, C. T., & Moretti, E. (2019). Housing constraints and spatial misallocation. American Economic Journal: Macroeconomics, 11(2), 1-39.

Jensen, R. (2010). The (perceived) returns to education and the demand for schooling. The Quarterly Journal of Economics, 125(2), 515-548.

Jensen, R. (2012). Do labor market opportunities affect young women's work and family decisions? Experimental evidence from India. The Quarterly Journal of Economics, 127(2), 753-792.

Katz, L. F., & Goldin, C. D. (2008). The race between education and technology. Cambridge: Harvard university press.

Kearney, M. S., & Levine, P. B. (2014). Income inequality, social mobility, and the decision to drop out of high school (No. w20195). National Bureau of Economic Research.

Knight, J., & Shi, L. (1996). Educational attainment and the rural-urban divide in China. Oxford Bulletin of Economics and Statistics, 58(1), 83-117.

Maarseveen, R. van, van der Pol, T., & Zwaneveld, P. (2017). Literatuurstudie ‘Effectiviteit van place-based beleid’. CPB Netherlands Bureau for Economic Policy Analysis, The Hague

Maarseveen, R. van, (2019). Forget about the kids: using the age of children when moving to identify neighborhood effects in Europe. Working paper available soon.

Manski, C. F. (1993). Adolescent econometricians: How do youth infer the returns to schooling?. In Studies of supply and demand in higher education (pp. 43-60). University of Chicago Press.

Minnesota Population Center. (2019). Integrated Public Use Microdata Series, International: Version 7.2 [dataset]. Minneapolis, MN: IPUMS. https://doi.org/10.18128/D020.V7.2

Mion, G., & Naticchioni, P. (2009). The spatial sorting and matching of skills and firms. Canadian Journal of Economics/Revue canadienne d'économique, 42(1), 28-55.

Oster, E. (2017). Unobservable selection and coefficient stability: Theory and evidence. Journal of Business & Economic Statistics, 1-18.

Oster, E., & Steinberg, B. M. (2013). Do IT service centers promote school enrollment? Evidence from India. Journal of Development Economics, 104, 123-135.

Roca, J. De La, & Puga, D. (2017). Learning by working in big cities. The Review of Economic Studies, 84(1), 106-142.

Rosenthal, S. S., & Strange, W. C. (2004). Evidence on the nature and sources of agglomeration economies. In Handbook of regional and urban economics (Vol. 4, pp. 2119-2171). Elsevier.

Shapiro, J. M. (2006). Smart cities: quality of life, productivity, and the growth effects of human capital. The review of economics and statistics, 88(2), 324-335.

Statistics Netherlands. (2011). Studeren loont. Inkomens van afgestudeerden in het mbo, hbo en wetenschappelijk onderwijs. Online publication.

Statistics Netherlands. (2019). Statline. Table available here.

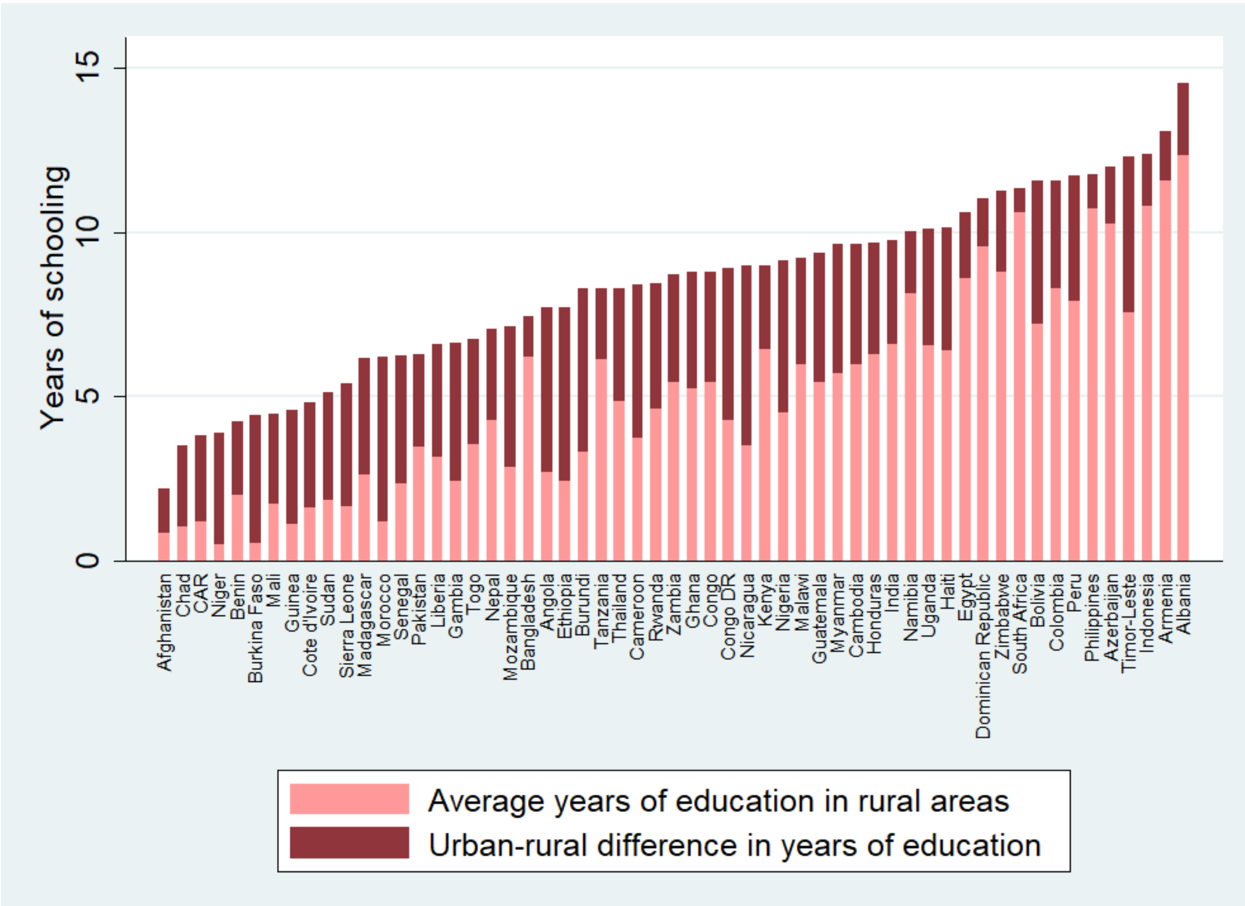
Verstraten, P., Verweij, G., & Zwaneveld, P. J. (2019). Complexities in the spatial scope of agglomeration economies. Journal of Regional Science, 59(1), 29-55.

Young, A. (2013). Inequality, the urban-rural gap, and migration. The Quarterly Journal of Economics, 128(4), 1727-1785.

IX Appendix

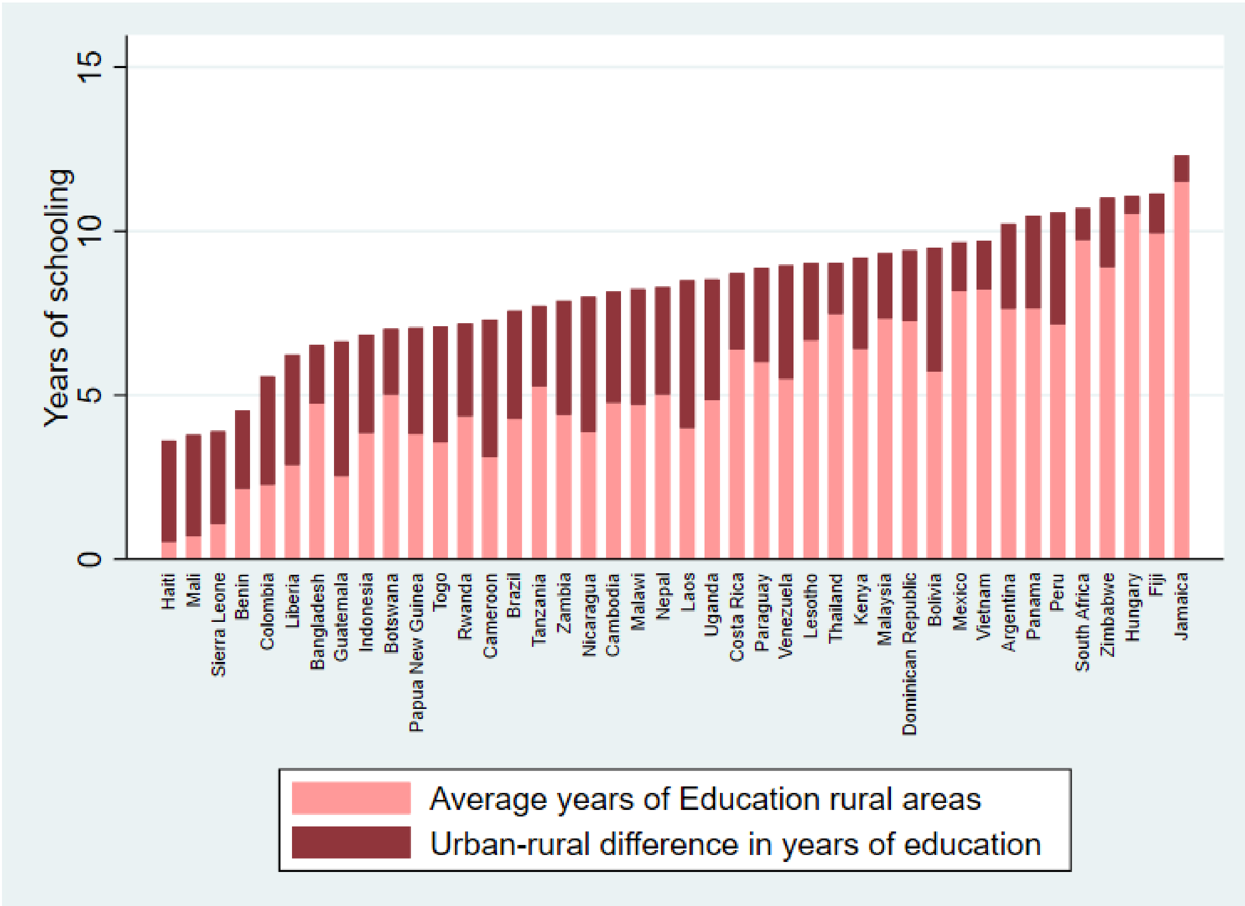
A: Additional measures of the Urban-rural educational gap

Figure A1: Urban-rural education gap based on 25-year-old males in DHS-data



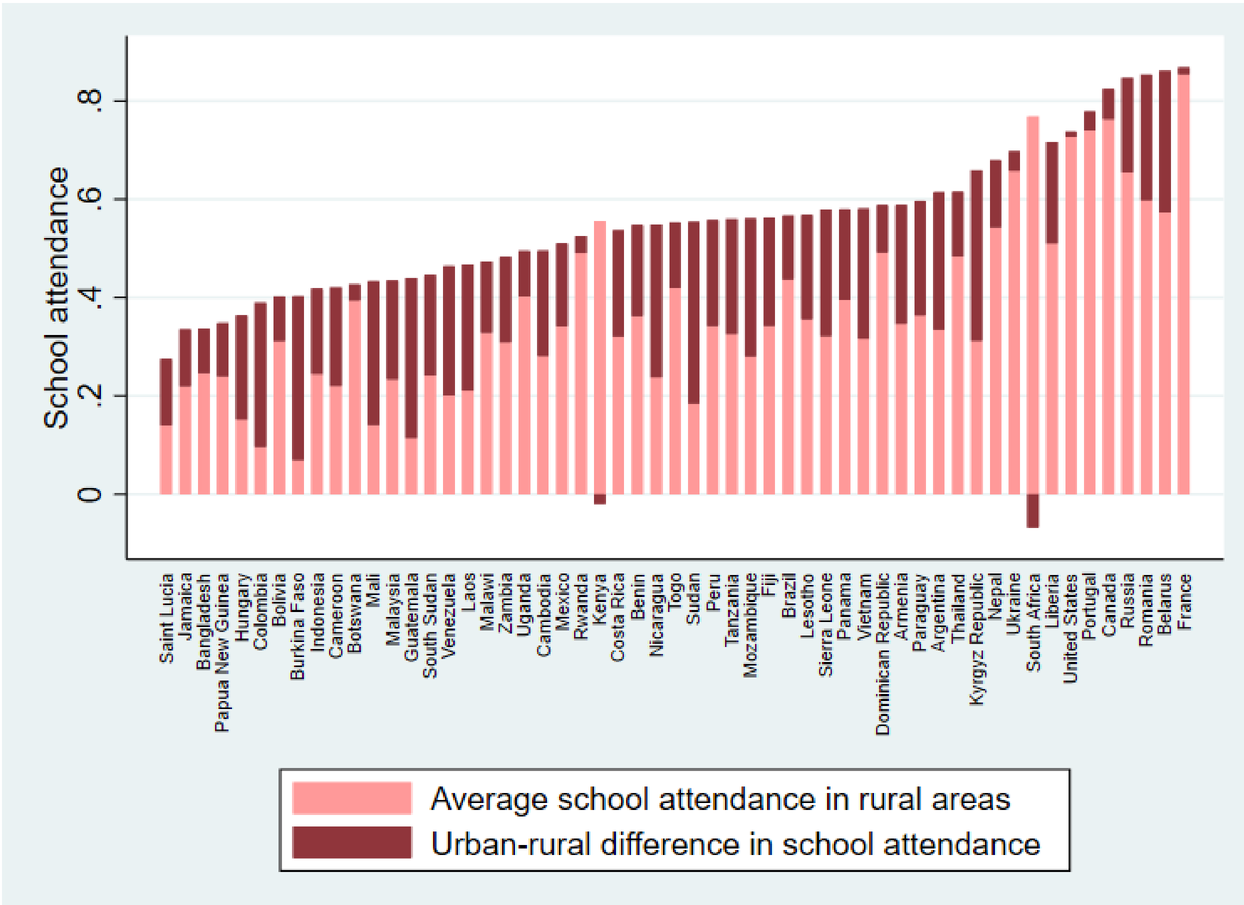
Note: Figure based on 57 countries included in the Demographic and Health Survey (DHS). Reported are the average years of schooling in urban and rural areas for 25-year-old males. The urban-rural definition follows the definition of DHS, which is based on the urban-rural definition of the country in question. Countries with fewer than 100 18-year old males for either the rural or urban region of the country have been excluded from the original sample of 75 countries.

Figure A2: Urban-rural education gap based on 25-year-old adults in IPUMS International-data



Note: Figure based on the latest census for the 42 countries included in the IPUMS International data made available by Minnesota Population Center (2019) which include an urban/rural definition in their census (which excludes most developed countries). Reported are the average years of schooling in urban and rural areas for 18-year-old males. The urban-rural definition is based on the definition of the statistical office of the country in question.

Figure A3: Urban-rural school attendance gap based on 18-year-old males in the IPUMS International data



Note: Figure based on the latest census for 55 countries included in the IPUMS International data which include an urban/rural definition in their census (which excludes most developed countries). The y-axis shows the percentage of children that indicate that they were attending school at the time of the census in urban and rural areas for 18-year-old males. The urban-rural definition is based on the definition of the statistical office of the country in question.

Acknowledgement for use of IPUMS data: The author wishes to acknowledge the statistical offices that provided the underlying data making this research possible: National Institute of Statistics and Censuses, Argentina; National Statistical Service, Armenia; Bangladesh Bureau of Statistics, Bangladesh; Ministry of Statistics and Analysis, Belarus; National Institute for Statistics and Economic Analysis, Benin; National Institute of Statistics, Bolivia; Central Statistics Office, Botswana; Institute of Geography and Statistics, Brazil; National Institute of Statistics and Demography, Burkina Faso; National Institute of Statistics, Cambodia; Central Bureau of Census and Population Studies, Cameroon; Statistics Canada, Canada; National Bureau of Statistics, China; National Administrative Department of Statistics, Colombia; National Institute of Statistics and Censuses, Costa Rica; National Statistics Office, Dominican Republic; National Institute of Statistics and Censuses, Ecuador; Central Agency for Public Mobilization and Statistics, Egypt; Bureau of Statistics, Fiji; National Institute of Statistics and Economic Studies, France; Federal Statistical Office, Germany; National Institute of Statistics, Guatemala; Institute of Statistics and Informatics, Haiti; Central Statistical Office, Hungary; BPS Statistics Indonesia, Indonesia; Central Statistics Office, Ireland; Central Bureau of Statistics, Israel; National Institute of Statistics, Italy; Statistical Institute, Jamaica; National Bureau of Statistics, Kenya; National Statistical Committee, Kyrgyz Republic; Statistics Bureau, Laos; Bureau of Statistics, Lesotho; Institute of Statistics and Geo-Information Systems, Liberia; National Statistical Office, Malawi;

Department of Statistics, Malaysia; National Directorate of Statistics and Informatics, Mali; National Institute of Statistics, Geography, and Informatics, Mexico; National Institute of Statistics, Mozambique; Central Bureau of Statistics, Nepal; National Institute of Information Development, Nicaragua; Statistics Division, Pakistan; Central Bureau of Statistics, Palestine; Census and Statistics Directorate, Panama; National Statistical Office, Papua New Guinea; General Directorate of Statistics, Surveys, and Censuses, Paraguay; National Institute of Statistics and Informatics, Peru; Central Statistics Office, Poland; National Institute of Statistics, Portugal; National Institute of Statistics, Romania; Federal State Statistics Service, Russia; National Institute of Statistics, Rwanda; Government Statistics Department, Saint Lucia; Statistics Sierra Leone, Sierra Leone; Statistical Office of the Republic of Slovenia, Slovenia; Statistics South Africa, South Africa; National Bureau of Statistics, South Sudan; Central Bureau of Statistics, Sudan; Bureau of Statistics, Tanzania; National Statistical Office, Thailand; National Institute of Statistics, Togo; Bureau of Statistics, Uganda; State Committee of Statistics, Ukraine; Bureau of the Census, United States; National Institute of Statistics, Venezuela; General Statistics Office, Vietnam; Central Statistics Office, Zambia; and National Statistics Agency, Zimbabwe.

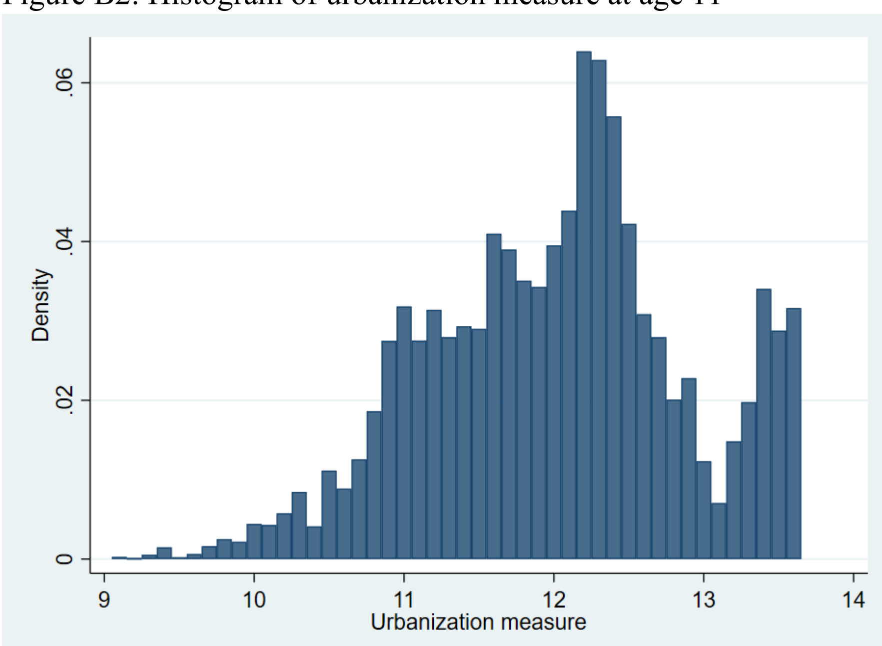
B Construction of urbanization measure

The administrative data includes information on place of residence from January 1995 onwards. Hence, I start by calculating the number of individuals registered per zip code on the 5th of January 1995. As next step, I determine the centroid of each zip code using GIS software. The map of the Netherlands with all the centroids is displayed in figure A1. For each centroid, I determine which other zip codes lie within a 10 kilometer radius of the centroids and add up the population of these zip codes. For zip codes that lie partially in the 10km radius, I multiply the share of the zip code area covered by the 10km radius with the population of that particular zip code.[35][36] The average zip code has 43 other zip-codes within a 10 kilometer radius and even the first percentile of zip codes has 8 other zip-codes within a 10 kilometer radius. The correlation between the number of individuals living within 5, 10 and 20 kilometers is fairly high at 0.89. Finally, I take the log of the number of individuals within the 10 kilometer radius as urbanization measure. Figure IV in the main text shows the resulting map of the urbanization measure. In addition, Figure B.2 shows a histogram of the urbanization measure.

[35] This means that I implicitly assume that the population is spread equally across zip codes. However, as zip codes are fairly small (average of 8 km2) compared to the 10km radius (314 km2) radius, this assumption has little effect on the relative differences in urbanization between regions.
[36] For some very isolated zip codes (for instance, on the islands in the North) there are no other zip codes within 10km, and hence the population with 10km is simply the zip codes own population.

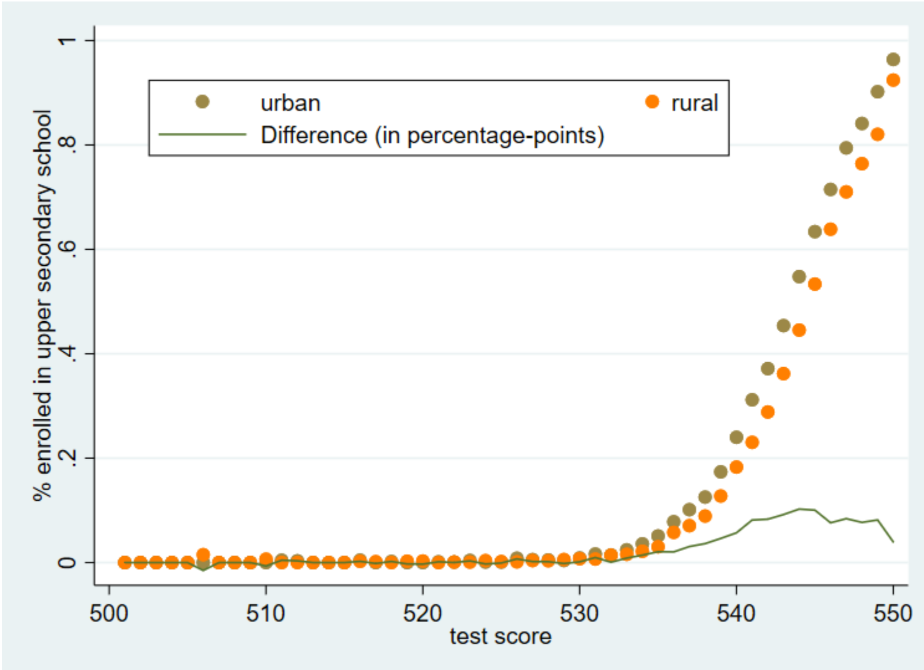
Figure B1: Map of Dutch zip codes and centroids.

Figure B2: Histogram of urbanization measure at age 11



Note: Histogram of urbanization measure for the cohorts born between 1994 and 1998. Areas with an urbanization score below 9 (containing 230 out of the 631.815 observations in the baseline sample) not displayed here.

Figure B3: probability of enrolling in upper secondary school by urbanization status and test score



Note: Note: Figure based on observations born in 1996, with the sample split between those who grew up in a place with less than the median density (rural) and those who grew up in a place with above-median density (urban). The y-axis displays the percentage of children enrolled in an upper secondary school 4 years after finishing primary school, the x-axis the score on the national end of primary school test. The green line show the difference in upper secondary school enrollment between urban and rural places in percentage points

C Additional robustness analyses

C.1 Probability of upper secondary school graduation, conditional on being registered at an upper secondary school

Table C1: effect of urbanization on probability of obtaining a degree in upper secondary school, conditional on enrollment in upper secondary school

	Baseline Estimates			Sensitivity Analysis		Area-fixed Effects		IV
	Ind. level controls	Ind. + Fam. controls	Ind. + Fam controls + test-score	Movers Excluded	Full parental education	Municipality-Fixed effects	Provincial-fixed effects	IV-estimates
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Urbanization at age 11	0.0075	0.0060***	0.0068**	0.0069**	0.0062*	0.0071	-0.0005	0.0071***
	(0.0041)	(0.0022)	(0.0021)	(0.0022)	(0.0025)	(0.0059)	(0.0028)	(0.0018)
Individual controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Test score-dummies	No	No	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.000	0.03	0.08	0.08	0.09	0.08	0.08	0.08
No. of obs.	146.879	146.879	146.879	130.634	60.945	146.879	146.879	146.248

Note: All results based on individuals born between 1994 and 1998. Dependent variable is whether a child graduates from an upper secondary school, conditional on being enrolled at the third grade of upper secondary school. Individual controls include a gender-dummy and cohort-dummies. Family controls include dummies for parental education combinations (168 dummies), parental nationality combinations (9 dummies), year-of-birth of oldest parent (40 dummies), family income (interacted with cohort fixed effects), low-income dummy (interacted with cohort-fixed effects), top-coded income dummy (interacted with cohort-fixed effects) and dummy for insufficient income data (interacted with cohort fixed effects). For a detailed explanation of the construction of the urbanization measure or the family controls, see section III. Column 4 excludes all children who moved municipalities during school going age. Column 5 removes all parents for whom uncertainty exists about the education level of one or both parents. Column 6 adds municipality-fixed effects for the municipality in which children live at age 11 (430 dummies), whereas column 7 adds province fixed-effect for the province in which children live at age 11 (12 dummies). Column (8) instruments for modern densities (based on 1995 population distribution, see Appendix B) with historical densities. For all columns apart from column 8 the standard errors are clustered on the municipality level. Column 1-7 are estimates by OLS, column 8 by 2SLS. First stage results of column 8 are reported in Appendix C3.

C.2 Applied university enrollment

Table C2: Effect of urbanization on applied university attendance

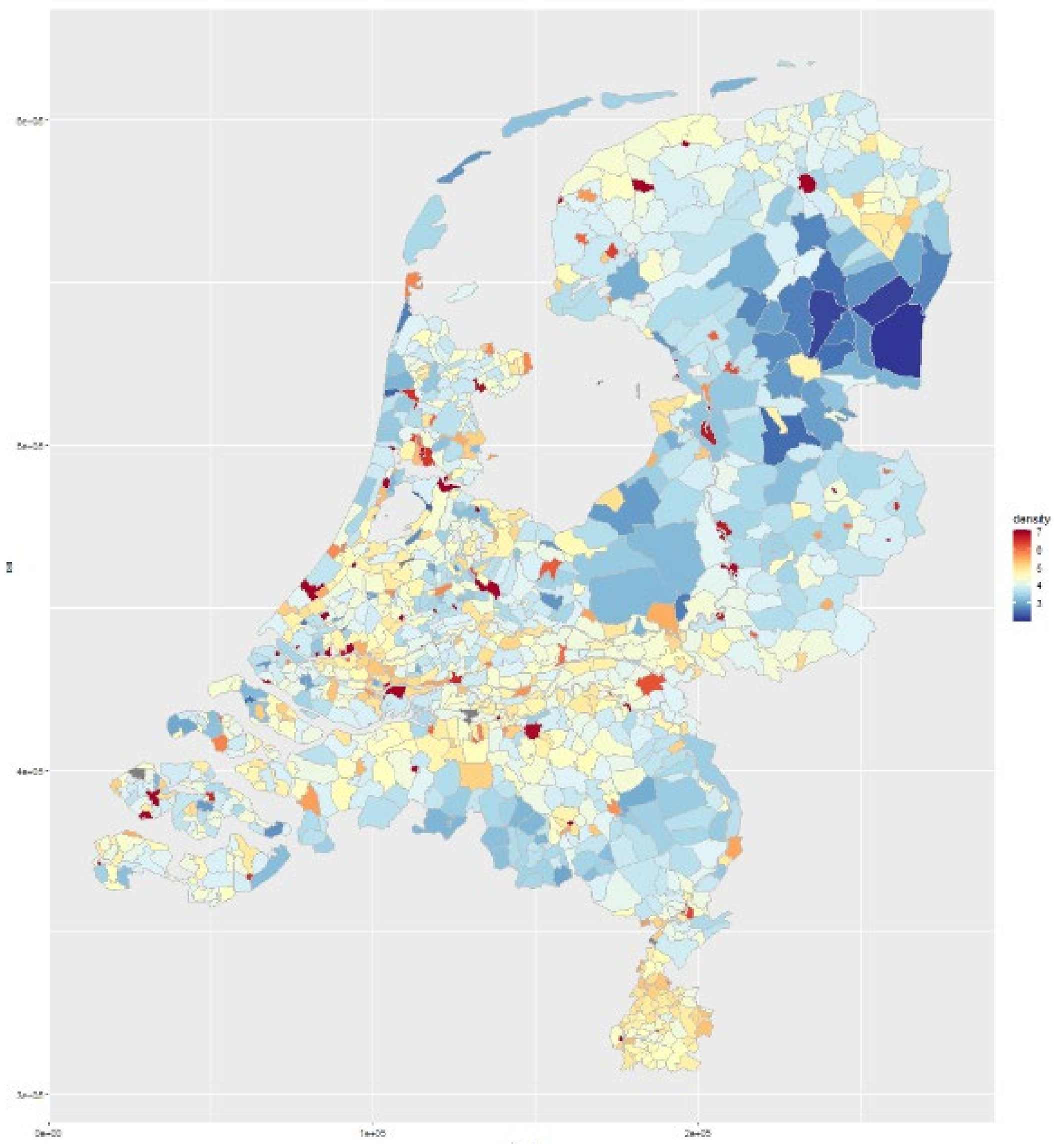
	Baseline Estimates			Sensitivity Analysis				Area-fixed Effects		IV
	Ind. level controls	Ind. + Fam. controls	Ind. + Fam controls + GPA	Children observed in any education	Movers Excluded	Cohorts of table II (1994-1998)	Full parental education	Municipality-Fixed effects	Provincial-fixed effects	IV-estimates
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Urbanization at age 11	-0.0233*** (0.0023)	-0.0094*** (0.0016)	-0.0089*** (0.0016)	-0.0090*** (0.0017)	-0.0095*** (0.0017)	-0.0101*** (0.0017)	-0.0081*** (0.0017)	-0.0090* (0.0041)	-0.0129*** (0.0018)	-0.0089*** (0.00098)
Individual controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
GPA and specialization track	No	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
R^2	0.00	0.03	0.07	0.07	0.07	0.06	0.07	0.07	0.07	0.07
No. of obs.	289.109	289.109	289.109	281.179	253.430	139.559	110.921	289.109	289.109	287.763

Note: All results apart from column (6) based on individuals born between 1989 and 1998. Dependent variable is whether a child attends applied university as highest education within 3 years of graduating from upper secondary school. Individual controls include a gender-dummy and cohort-dummies. Family controls include dummies for parental education combinations (168 dummies), parental nationality combinations (9 dummies), year-of-birth of oldest parent (40 dummies), family income (interacted with cohort fixed effects), low-income dummy (interacted with cohort-fixed effects), top-coded income dummy (interacted with cohort-fixed effects) and dummy for insufficient income data (interacted with cohort fixed effects). For a detailed explanation of the construction of the urbanization measure or the family controls, see section III. Column 4 excludes children who are not enrolled in any form of tertiary education three years after graduating. Column 5 excludes all children who moved municipalities during school going age. Column 6 limits the sample to the cohorts born between 1994-1998, in line with the baseline sample of table II. Column 7 removes all parents for whom uncertainty exists about the education level of one or both parents. Column 8 adds municipality-fixed effects for the municipality in which children live at age 11 (430 dummies), whereas column 9 adds province fixed-effect for the province in which children live at age 11 (12 dummies). Column 10 instruments for modern densities (based on 1995 population distribution, see Appendix B) with historical densities. Columns 1-9 are estimated by OLS, column 10 by 2SLS. First stage results of column 10 are reported in Appendix C3. For all columns apart from column 10 the standard errors are clustered on the municipality level.

C.3 Instrumental Variable (IV) first stage

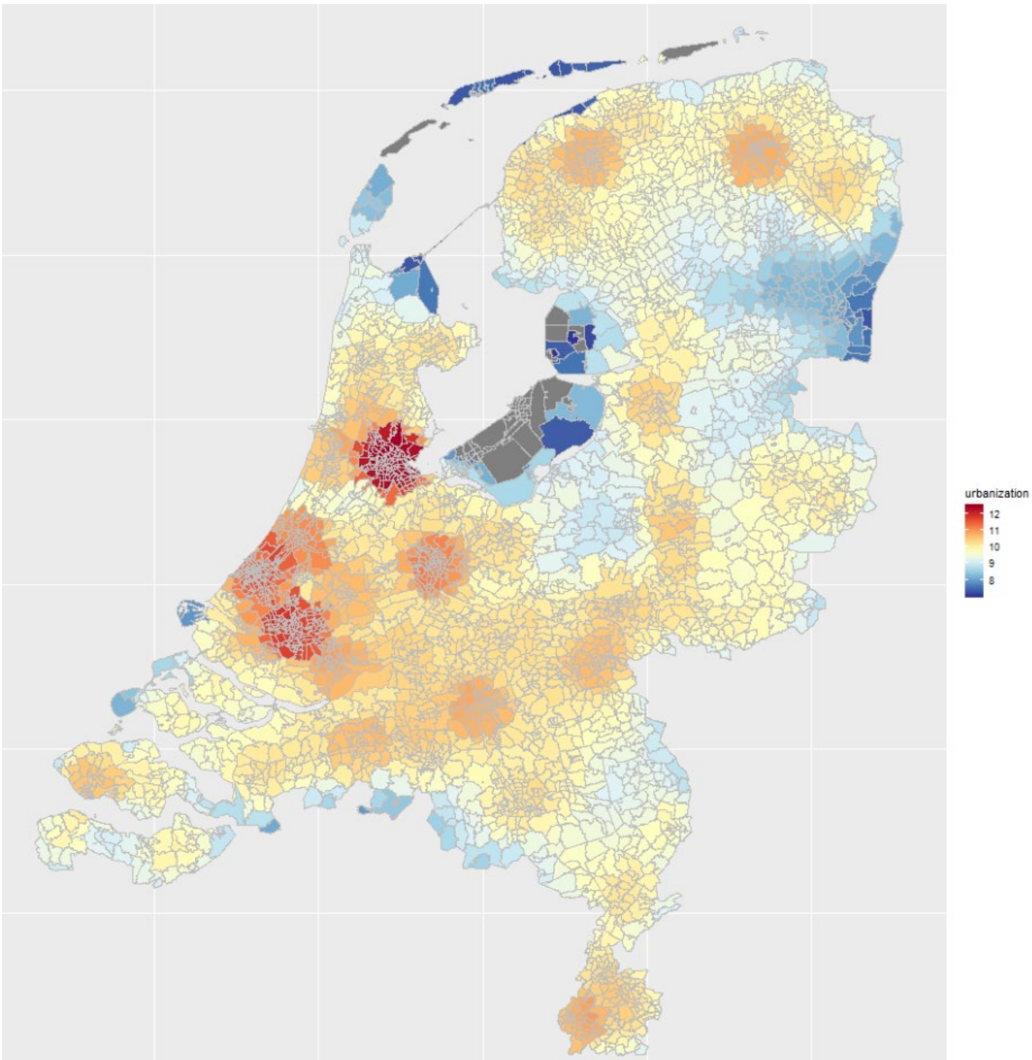
Figure C3.1 displays the population densities of Dutch municipalities in 1840. The number of municipalities (1340) is relatively large compared the modern day number of municipalities (400). The map below is used as input to construct the IV-measure, namely the number of individuals in 1840 living within 10 kilometers of all current zip codes using the procedure outlined in Appendix B. Notice that I make the implicit assumption that the population in 1840 was spread homogenous across the municipalities, as I only have the 1840 population statistics at the municipality level.

Figure C3.1: Population density per municipality in 1840.



Note: When calculating the densities based on the 1840 population map, I implicitly assumed that population is spread homogenously within municipalities. Notice that The Netherlands contained a substantial larger amount of inland water compared to the contemporary Netherlands, due to the land reclamation programs in the 19th and 20th century

Figure C3.2: Urbanization measure based on 1840 population distribution



Note: The map displays the urbanization measure for each zip code (log of number of people living within 10km), based the 1840 population distribution. See appendix B for the construction of the measure and Figure C3.1 for the 1840 population distribution. The grey areas have an urbanization measure below 8.The grey mass in the center of the Netherlands consists of land that has been reclaimed since the 1840's, and hence had very few to no individuals living with 10 kilometers in 1840, as can be also seen on figure C3.1.

Table C3: First stage of regressing urbanization based on the 1840 densities on contemporary urbanization measure (based on 1995 densities).

	First stage
Urbanization based on 1840 densities	0.635***
	(0.0045)
F-statistic	2349.94
R^2	0.58
No. of obs.	628.396

Note: First stage of 2SLS regression of baseline urbanization measure (based on the 1995 population distribution) on the historical urbanization measure (based on the 1840 population distribution).

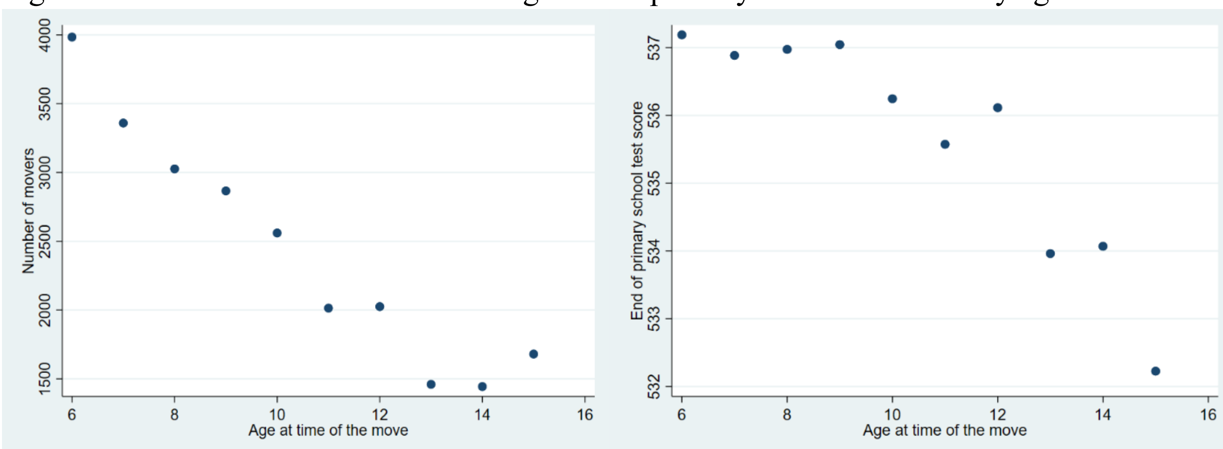
D Application of Chetty & Hendren (2018) identification strategy

One way to achieve causal identification and to deal with endogeneity concerns is to employ the methodology of Chetty & Hendren (2018), using families who move between regions for identification. The key identifying assumption is that within the groups of families moving between pairs of regions, the age of children at the time of the move is orthogonal to potential outcomes of children. As such, identification can be achieved by comparing the differences in outcomes for children who moved at different ages between regions. Appendix D.1 develops a formal framework to achieve identification under the assumptions of Chetty & Hendren (2018).

In line with Chetty & Hendren (2018), I analyze the subgroup of individuals who move more than a certain distance exactly once (100 miles in Chetty & Hendren (2018)). Crucially, the Dutch data allows for testing of their identifying assumption to some degree, as relatively early measures of cognitive ability are available. Unfortunately, the data reveals that the Chetty & Hendren (2018) assumption does not appear to hold well in the Dutch context. Figure D1 shows the number of families who move once a distance of at least 50 kilometers by age of the children, as well as the average end of primary school test score. As can be observed, both the number of movers (panel a) as well as the average test score (panel b) fall sharply as the age of the child at the time of the move increases. Particularly problematic is the decline in test score for the children who move between ages 12-16, as all these children took the end of primary school test before their move. As such, there appears strong evidence of a relationship between potential outcomes and the age of children at the time of the move in the Dutch context.

Table D1 in Appendix D2 shows that the results are very similar if I instead use families who move at least 20 kilometers once or who move at least 100 kilometers once. The decline in observed academic performance is substantial. Children who move at age 14/15 have an end of primary test score between a third and half a standard deviation lower than children who move before the age of 12. As such, the identification assumption of Chetty & Hendren (2018), that the age at the time of the move is uncorrelated with the potential outcomes, does not appear to hold in the Dutch context. The most likely explanation is that mobility levels in the Netherlands (as well as in Europe in general) are much lower than in the US, and hence the movers are a far more selective subgroup. Appendix D2 provides some more detail.

Figure D1: Number of movers and average end of primary school test score by age of move



(a) (b)

Note: Figure II (Left) indicates the number of families that moves by age of child for the cohorts born between 1994-1998. Figure III presents the median end-of-primary school test score for the children who do move. In line with Chetty & Hendren (2018), the figures are based on the subset of children who move once at least a certain amount of distance (100 miles in Chetty & Hendren, 2018). As the Netherlands is smaller, I instead limit the sample to the children who move once at least 50 kilometers between ages 6-16. The results are very similar when using a 20 or 100 minimum moving distance instead (see table D1 in the appendix).

D.1 Estimation model

The results in the main text show that individuals who grow up in rural communities invest less in education compared to their counterparts who grow up in cities, even conditional on observed academic ability and a wide range of parental characteristics. Nonetheless, it might be the case that families or children differ between rural and urban communities on some unobservable dimensions that are relevant for the educational decision of children.

In order to test the robustness of the results to such sorting on unobserved variables, I attempt to employ a second estimation strategy using the approach of Chetty & Hendren (2018). Their key identifying assumption is that the families that move between any pair of regions are a (highly) selective subgroup, but that within the group of movers, the age of the child at the time of the move is uncorrelated with the potential outcomes of the children. Chetty & Hendren (2018) show in the

US context that this assumption seems to hold up quite well and their results are highly robust to different variations on the estimation strategy.

Under the identifying assumption of Chetty & Hendren (2018), I can consistently estimate the effect of urbanization on educational outcomes. Equation D1 below is a slightly modified version of equation (6) of Chetty & Hendren (2018).[37]

$$P(attends\ university)_i = \sum_{s=1989}^{1996} I(s_i = s)(\alpha_s^1 + \alpha_s^2 * urbanization_o) + \sum_{m=6}^{16} I(m_i = m)\,\varphi_m + \sum_{m=6}^{16} b_m * I(m_i = m) * \Delta urbanization_{od} + \varepsilon_i \qquad \text{(D1)}$$

The first term includes a cohort-fixed effect as well as a cohort-specific coefficient for the degree of urbanization in the place of origin. The second term captures the effect of moving itself, which includes the disruption caused by the move which might be age-dependent. The third term crucially captures the effect of moving to a more or less urban area relative to the origin. The estimates of b_m capture two effects: sorting into more urban areas and the effect of exposure to an urban environment. Formally

$$b_m = \beta_m + \delta_m \qquad \text{(D2)}$$

Where β_m is the effect of spending 17 – m years in an environment that has a log-point higher density and δ_m reflects sorting into the city ($\frac{Cov\ (\Delta urbanization_i, \varepsilon_i)}{Var(\Delta urbanization_i)}$) at various ages. The key assumption of Chetty & Hendren (2018) is that the age at which families move is orthogonal to the potential outcomes of children, which implies that $\delta_m = \delta$ for all ages m. As a result, one can obtain the effect of one more year of exposure to an area one log-point denser as $\gamma_m = b_{m+1} - b_m$. In essence, the estimation strategy uses differences in exposure to more/less urban environments for children to identify the causal effects of places. As the effect of living in a city may depend on a child's age, the estimation strategy allows γ_m to vary with age.

D.2 Identification assumption in the Dutch/European context

How sensible this identifying assumption is in the Dutch (or European) context is unclear a priori and needs to be tested. In general, the levels of mobility are substantially lower in Europe compared to the US (both within countries as well as across countries), which means that the group of movers might be more selective. Fortunately, the data allows for the testing of the identifying assumption

[37] The modifications reflect that we are not interested in the convergence of movers to the population average of the destination, but instead are interested in a specific characteristic of a receiving region, namely urbanization.

to some degree. In particular, I can analyze the group of children who move at ages 12 or older, which is after children finish their primary school, and analyze whether observed characteristics vary with the age at the time of the move.

Table D1 provides the numbers of movers, as well as the mean end-of-primary school test score for the groups of children who move exactly once a distance of at least 20, 50 or 100 kilometers.[38][39] The results show a steep decline in both the number of movers, as well as the average and median test score. This decline is substantial: the average test-score of children who move at the age of 15 is about 40% of a standard deviation lower compared to children who move at the age of 12 at all three minimum moving distances. This decline is highly significant, and it hence appears that the age at the time of the move is related to the potential outcomes of children in the Dutch context. Furthermore, the number of children who move at a given age drops rapidly with the age of moving, which suggests that parents do seem to take the age of their children into account when they move. Figure D1 shows the same results graphically for the families that move at least 50 kilometers. Hence, the Chetty & Hendren (2018) method is thus not suitable as identification strategy in this particular context.

Table D1: number of observations and end-of-primary school test score by age of move

	Moved > 20km		Moved > 50km		Moved > 100km	
	Number of movers	Mean testscore	Number of movers	Mean testscore	Number of movers	Mean testscore
Moved aged 6	6710	536.6	3984	537.2	1913	537.2
Moved aged 7	5807	536.6	3359	536.9	1532	536.9
Moved aged 8	5188	536.5	3026	537.0	1364	536.8
Moved aged 9	4791	536.5	2866	537.0	1342	536.8
Moved aged 10	4183	536.3	2560	536.2	1129	536.1
Moved aged 11	3488	535.4	2014	535.6	917	535.7
Moved aged 12	3574	535.7	2025	536.1	970	536.1
Moved aged 13	2605	533.6	1460	534.0	650	533.7
Moved aged 14	2591	533.1	1444	534.1	635	534.1
Moved aged 15	3167	531.9	1680	532.2	697	531.9

Note: All figures based on cohorts born between 1994 and 1998. Moved age 6 means a child moved between his 6th and 7th birthday. Figures refer to the number of movers and test-scores for the groups of children who move exactly once at least 20/50/100 km between ages 6-16. For reference, the standard deviation on the individual level on the test is 9.7 (see table 1).

[38] The use of individuals who move a single time is in line with Chetty & Hendren (2018). The difference is that Chetty & Hendren (2018) uses individuals who move exactly once a distance of more than a 100 miles, whereas here I use a lower bound of 20/50/100 kilometers to reflect the smaller size of The Netherlands.

[39] A small group of individuals (0.23% of the sample) moves across municipalities more than 3 times between ages 6-16. In line with Chetty & Hendren (2018), I exclude these from the sample.