Location choices of Swedish independent schools

Karin Edmark[1] ![ORCID]

Abstract

This paper studies the location decisions of the Swedish privately provided independent schools. It makes use of the great expansion of such schools following the 1992 independent school reform, to study the local determinants of independent school entry. The analysis thus provides evidence on the location choices made by private agents in a mixed market setting. The modifiable areal unit problem is addressed by employing a set of alternative measures for local school markets, constructed from highly detailed geographical data on schools and students. The results suggest that independent schools were more likely to choose locations with a larger share of students with high-educated parents; a higher student population density; and a lower share of students with Swedish-born parents. There is also some evidence that independent schools were less likely to locate in municipalities with a left-wing political majority.

1 Introduction

This paper studies the location decisions of the privately provided and voucher-funded schools[1] that emerged in the wake of the Swedish independent school reform of 1992. In particular, it tests what characteristics of the local school market were correlated with independent primary level[2] school entry in 1992–2000, with

---

[1] These are in the present paper denoted independent schools, as this term has been used in previous studies of the Swedish voucher schools, see e.g. Böhlmark and Lindahl (2015). The Swedish term is "fristående skolor".

[2] I choose to focus on the primary schools, as they are likely to have a strong connection to the local neighbourhood in the sense that young students often attend a school near home. Restricting the analysis to these schools also avoids some data difficulties, see the working paper version of this paper for details.

✉ Karin Edmark
   karin.edmark@sofi.su.se

1   The Swedish Institute for Social Research (SOFI), IFN, IFAU, CESifo, UCLS, Stockholm
    University, 106 91, Stockholm, Sweden

respect to aspects such as the family background of the local student population, the local political majority, and the quality and density of the existing neighbourhood schools. The 1992 reform introduced practically free entry, including for for-profit companies, and thus introduced strong market incentives into the education sector. As it gave rise to the entry of a large number of independent schools, it provides an excellent opportunity to study the location decisions of private providers in a mixed market setting.

This paper makes two main contributions: (i) It provides descriptive evidence on the geographical pattern of a dramatic independent school expansion, by analysing what local characteristics were correlated with a higher likelihood for independent school location.[3] (ii) It carefully assesses the sensitivity of the results to the modifiable areal unit problem (MAUP),[4] i.e. it investigates if the regression estimates are sensitive to the type of spatial aggregation employed. This is done by making use of highly detailed geographical information to construct a set of alternative definitions of local school markets, and by investigating if the estimated parameters are sensitive to gradual changes in the local market definition, thus providing an informal test of the MAUP.[5]

The MAUP is, I would argue, in general not sufficiently addressed in the literature outside the domain of economic geography. For example, previous studies on school locations in general tend to make use of the spatial units that are readily available, such as school districts, zip codes, or census tracts in the US education literature, or SAMS areas or municipalities in the Swedish context, without much discussion of whether or not this is the appropriate level of aggregation.[6] With the increased accessibility of geo-coded data, as well as sufficiently powerful software,[7] there is scope to take this issue more seriously, for example by using tailor-made local units, and this applies to all studies using spatially aggregated data.

The main results of this study suggest that the likelihood for independent school entry was correlated with the local student population density as well as with the local student family background. In particular, the independent school entry probability was higher in locations where students were of high-educated family

---

[3] It shall be emphasized that the results are descriptive in character; in particular I do not claim to provide an exhaustive model of the school location patterns—something that would require a more thorough structural model of the school market. Furthermore, some of the included variables can only be included by rough proxies and shall thus be interpreted with extra caution. This will be commented in more detail in Sects. 3 and 5.

[4] See Wong (2009) for an overview of the MAUP.

[5] Fotheringham and Wong (1991) point out the potential relevance of the MAUP in multivariate statistical analysis, and Wong (2009) suggests that using different aggregation scales and schemes is a practically viable approach to deal with the MAUP that should be available to most researchers.

[6] An overview of the previous literature on school locations can be found in the working paper version of this paper, Edmark (2018). For previous research articles, see Bifulco and Buerger (2015), Glomm et al. (2005), Burdick-Wills et al. (2013), Henig and MacDonald (2002), Koller and Welsch (2017), Downes and Greenstein (1996), Barrow (2006), Ferreyra and Kosenok (2018), Mehta (2017), Rincke (2007). The Swedish independent school locations have also been studied by two Swedish reports: Angelov and Edmark (2016) and Holmlund et al. (2014).

[7] GIS-based software facilitates the analysis but is not necessary—all the analysis in this paper was carried out using Stata.

background and was lower in locations, where a large share of students had at least one Swedish-born parent. There was also some indication of a lower likelihood for independent school entry in municipalities with a local left-wing political majority, although this result was not robust to changes in the outcome time period.

The above-mentioned results were robust to the various alternative and flexible definitions of local school markets. For other variables, such as the local income dispersion, average GPA, and voucher level, the definition of the local market, however, had a substantial impact on the results. This underlines the importance of accounting for the MAUP in studies using spatially aggregated variables.

The remaining sections of this paper are organized as follows: Sect. 2 gives a brief overview of the Swedish independent school reform, Sect. 3 provides information on the data variables, and Sect. 4 describes the MAUP and the spatial aggregation measures. Section 5 presents the results, and Sect. 6 concludes.

2 The Swedish independent school reform[8]

Prior to the independent school reform of 1992, the vast majority of Swedish children were educated in a municipal school. Schooling could also take place in an independent school, but these were limited to schools that were either of alternative pedagogical profile, boarding schools, or schools for foreign nationals.

In the fall of 1991, a tight parliamentary election brought a right-wing coalition government to power, and by July the following year it had implemented an extensive reform of the independent school system. The reform introduced a voucher based funding system for the independent schools and abolished the restriction of independent school status to the above listed types. The vouchers were to be paid out by the students' home municipalities, and at a level basically on par with the funding provided to the municipality-operated schools.[9]

For the period under study in this paper, the application process for start-up independent schools was handled by The Swedish National Agency for Education. According to the regulation, approval should be granted if the provider was deemed competent to provide education according to the goals and (since 1997) the value system of the Swedish education system, and had a credible economic plan.

The 1992 reform meant a significant improvement of the conditions for the independent schools and led to an immediate rise in the number of independent schools, as illustrated in Fig. 1.[10]

The rapid expansion of the independent schools is also visible in the maps of Fig. 2, which show the locations of all grade 1–3 independent schools in the last

---

[8] For a more detailed overview of the independent school reform, see the working paper version of this paper.

[9] The exact regulation for the voucher levels has varied over the years, but the general rule for the earlier years was that it should equal the municipalities own per student school funding with a possible deduction of 15–25%, see the working paper version of this paper for more details.

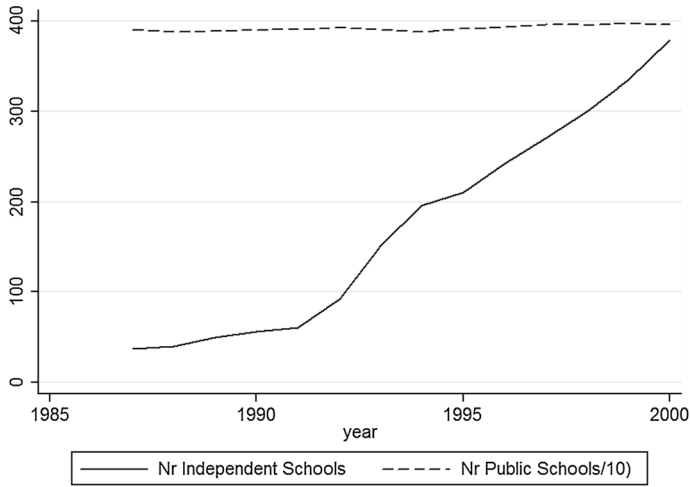[10] More detailed information can be found in the working paper version of this paper.

Fig. 1 Number of independent and public schools offering grades 1–3 in 1988–2000. Note that the number of public schools has been divided by ten
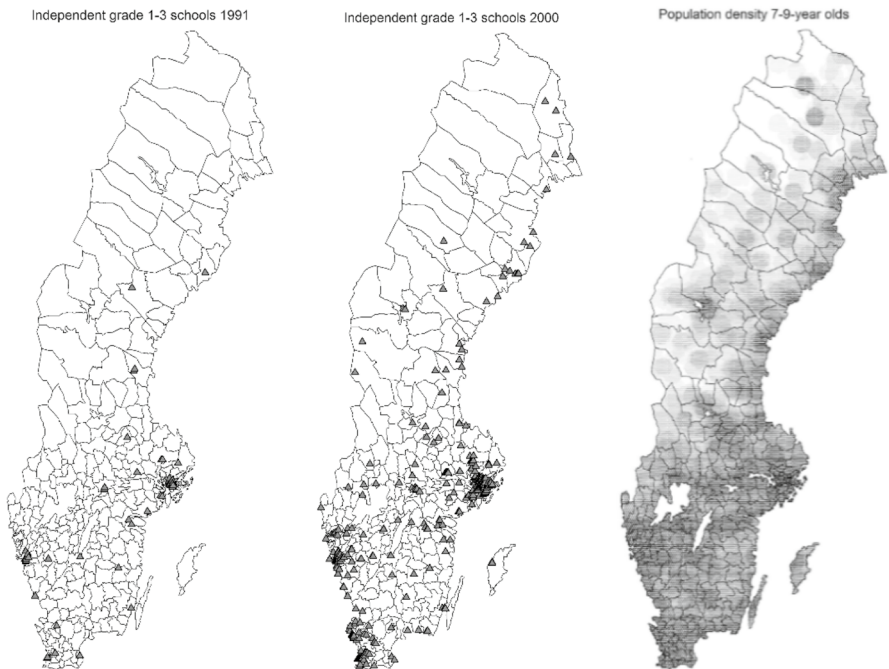

Fig. 2 Maps over the location of independent schools offering grades 1–3 in 1991 and 2000, and the population density among 7–9-year olds in 1991

pre-reform year of 1991, as well as in year 2000. It also shows the population density among school-age children in 1991. As indicated by the maps, the independent schools opened up in many parts of the country, but tend to cluster in the more densely populated areas.

3 Data variables of the empirical model

Before we move on to the issue of the spatial definition of local entities, this section describes the data variables of the analysis of the location choices of the independent schools that entered the education market after the 1992 independent school reform.[11] All right-hand side regression variables will be measured in 1991,[12,13] the year before the reform, and the empirical model will test how well they predict the subsequent location choices made by schools opening up in 1992–2000.[14]

3.1 Right-hand side regression variables

The aim is to include variables that capture a broad range of location specific characteristics that potentially influence the location choices of independent school providers.[15] To this end, I first construct a set of variables for students' family background, namely: (i) the share of students with at least one parent with post-secondary education; (ii) the share of students with at least one Swedish-born parent; and (iii) the average and standard deviation of the disposable household income, all measured among the local population aged 7–9. These variables are motivated on the grounds that school providers may prefer to locate near students with a certain socio-economic profile in the hope of attract them to the school. On the one hand, a provider with a social mission to bring good education to disadvantaged areas may opt for neighbourhoods, where students in general come from a low-education family background. On the other hand, another provider may prefer students from a high-education or high-income background, in order to earn a reputation as an elite type

---

[11] Detailed information on the data sources and variable definitions is available in the working paper version of this paper, see Edmark (2018). The datasets generated during and/or analysed during the current study are based on register data from Statistics Sweden. Due to the regulations of Statistics Sweden, the data cannot be distributed by the author.

[12] By measuring local level variables just before the reform, we furthermore reduce the risk for endogeneity bias due to potential reverse causality. Reverse causality may in this context arise if local characteristics are affected by the actual or expected entry of independent schools, for example through effects on the demographic composition.

[13] The exceptions are the below described proxy variables for voucher levels and costs for facilities, which are, due to the lack of earlier data, measured in 1992.

[14] Measuring the explanatory variables in 1991 eliminates the risk that the results are biased due to endogenous school market characteristics. Such endogeneity can arise if actual or anticipated entry of an independent school affects the characteristics of the school market, for example if the plans for opening an independent school in a neighbourhood affects moving patterns.

[15] The working paper version of this paper includes a theoretical model which guided the choices of variables, see Edmark (2018). The model was based on the assumption that independent schools choose location based on how it affects expected revenues and costs.

school, or if such students are perceived as being associated with lower costs. These student background variables may also be related to the demand side, if independent schooling is more popular among students of some family background. It shall be underlined that the results of the empirical model will not enable us to distinguish between these potential underlying channels—several of them may be at play simultaneously—but will rather show the aggregate picture.

Second, the model will include variables for the local student population density, measured as the number of students aged 7–9 since this corresponds to the lower primary school age,[16] and for local the school density, defined as the number of public and independent schools, respectively, in the local market.[17] These variables are included as they are likely to capture the local demand for an entering school: more students mean a higher demand in general, and even more so if the number of existing schools in the neighbourhood is low. The local demand may furthermore be higher if the quality of the existing schools is low. In order incorporate this notion in the model, I add as a proxy variable for education quality, the grade point average (GPA) among local students at age 16, which is the earliest age at which any education attainment information is provided in the national registers for the studied period. This is naturally a crude measurement for lower primary school quality, both as it also reflects the quality of higher education, but it is the information that is available for the period under study.[18] GPA is furthermore a limited indicator of quality if we consider that schools may differ in their grading standards, but is on the other hand an indicator that is visible to both families and school providers.

Third, I add a political variable in the form of an indicator for left-wing local political majority, defined as the Social Democrats and the Left Party jointly having at least 50% of the seats of the municipal council. The council decides on issues such as school choice policy, public transport, and construction permits, and can thus impact the day-to-day operation of the schools. The left-wing parties have from the start been more sceptical to independent schools than the right-wing parties, and the hypothesis is thus that the independent schools may therefore seek to avoid left-wing municipalities.

Finally, the model includes variables for local costs for facilities and wages, which are the main cost items for schools, as well as a measure of the per student voucher a schools can expect to receive from the municipalities. Due to a lack of more precise information,[19] these variables will be approximated for using the

---

[16] Schools with lower primary education often also offer grades for students up until age 12, and sometimes even until age 15. I, however, choose to base the variables on students age 7–9, as this is the age group that is common to all the schools in the regression sample.

[17] It can be noted, however, that the overall number of existing independent schools will be very small, as these will in the baseline regression be measured in 1991.

[18] It can also be pointed out that the variable is unadjusted for student socio-economic background. However, as the econometric model includes variables for parental education, household income, and foreign background, it will in effect hold constant for these socio-economic background variables.

[19] For some years and municipalities, there is survey-based information on student vouchers, see Angelov and Edmark (2016) for more information, but the information is too limited to be useful for this analysis.

following variables: The expected costs for facilities will be measured using the per student cost for school premises in the municipality-operated schools, and labour market region dummy variables will be added to the regression in a robustness analysis to account for the fact that teacher wages may vary regionally.[20] The expected per student vouchers are measured as the per student expenditure in the municipalities' own schools. This is motivated by the fact that the vouchers shall, according to the regulation, approximately follow the resources provided to the municipal schools. While these measures are included to give a more comprehensive model, it shall be underlined that they are likely measured with error, and their estimates shall thus be interpreted with extra caution.

3.2 Outcome variable

The outcome variable of the regression model is measured as a binary indicator of independent school entry. In the baseline specification it takes the value one if at least one independent school opened in a local school market at any point in time between 1992, the first year of the independent school reform, and year 2000.[21,22] This means that I will study the location decisions of the independent school start-ups during the first 9 years of the independent school reform. Results for alternative time periods are reported in a sensitivity analysis in Sect. 5.3.

The binary definition of the outcome variable does not take into account the intensity of the outcome variable, i.e. whether one or more independent schools choose a specific location. As will become clear in the following section, I will for most of the regression analysis define the geographical units of the analysis based on very small geographical entities, such that there will rarely be more than one independent school per unit. However, for the regressions using the larger SAMS and municipalities as spatial units, I will complement the binary outcome with a continuous outcome variable in the form of the number of start-up independent schools.

---

[20] The Labour market region indicators are generated by Statistics Sweden based on local commuting patterns. As the analysis will show, the main results will remain unaltered when these dummy variables are included; however, they will induce a substantial share of the geographical units of analysis to be dropped due to multicollinearity since several of the Labour market regions are quite small and perfectly predict the outcome variable. Therefore, I have chosen to include this specification only as a robustness test, and not as the main specification, which means that local teacher wage proxy is omitted from the baseline regression equation.

[21] When generating the outcome variable, I exclude all cases where an entering independent school shares geographical coordinate information with a school (independent or public) that existed already in 1991. This is done as I cannot rule out that these schools are transformations or extensions of previously existing schools, rather than new start-ups.

[22] This means that all start-ups between year 1992 and year 2000 are included in the analysis, including also schools that did not remain in business until year 2000. This is feasible as it is the schools' location choices—not their success rate—that is analysed.

4 Spatial definitions of potential school locations and school market characteristics

In contrast to most of the previous literature on this subject, this study will *not* rely on pre-existing administrative geographical units (school districts, zip codes, etc.). The reason is that such areas—which were not generated for the particular research question at hand—may not constitute feasible measures of local characteristics in the context of schools' location choices. What is the relevant unit of measurement may additionally vary across the regression variables. This section first provides a brief overview of the modifiable areal unit problem in relation to school locations, and then turns to the spatial definitions that are used in the analysis.

Before doing so, it shall be acknowledged that this is not the first school location paper to deviate from using (only) administrative neighbourhood measures.[23] Compared to the previous studies, however, this paper goes a step further by using a wider set of alternative measures of geographical school markets that are tailor-made for the analysis.

4.1 The modifiable areal unit problem in the context of school locations

As was mentioned in the introduction, the general message of the MAUP is that the level and shape of the spatial aggregation may affect the results that are obtained from the analysis. This may seem self-evident, but the MAUP is useful in defining and analysing the concept in more detail.[24] In particular, the problem is divided into two parts: the scale and zoning problems. The first relates to, as the name suggests, how the scale of spatial unit affects measurement, while the zoning problem has to do with its shape.

The problem can be illustrated in the context of school locations, for example by considering the US literature, which has often used school districts as the spatial unit of aggregation.[25] Relating to the MAUP, the root of the problem is that these districts were not created in order to be used in studies of school locations, but are rather local government units. This means that if we are lucky, they may incidentally coincide with what is the appropriate level of aggregation for school location decisions. But it may well be the case that they are too large—or too small—(*the scale problem*), or of the wrong shape (*the zoning problem*), to be useful. It could for example be that a school that chooses to locate near the border of a school district does so with the aim to serve students in parts of all districts near that border, instead of serving only the students in the district of location. If so, using the school district as unit of analysis means measuring the relevant variables with error. Or, it could be the case that one (large) school district comprises a number of (smaller)

---

[23] Burdick-Wills et al. (2013), Ferreyra and Kosenok (2018) Glomm et al. (2005), Downes and Greenstein (1996), Angelov and Edmark (2016) and Koller and Welsch (2017) either use measures other than administrative units, or take into account also the neighbouring units.

[24] Wong (2009) explains and exemplifies the MAUP.

[25] See Edmark (2018) for a review of the literature.

potential school locations, in which case the level of aggregation is too high and may mask underlying patterns.[26] The issue is furthermore complicated by the fact that what is the relevant unit of measurement may vary across the regression variables, such that different levels of aggregation are relevant for different variables.

These concerns are naturally not limited to school districts, but potentially apply to any spatial measure that was not tailor-made for the particular analysis at hand. In the Swedish context, one commonly used local spatial unit is the Small Area Market Statistics (SAMS), generated by Statistics Sweden.[27] It is also relatively common to base regional analyses on the municipalities, i.e. the lowest level of local government.[28] Both of these measures, however, suffer from drawbacks that can be related to the MAUP. As noted by Amcoff (2012), differences in the underlying local data that were used to construct the SAMS units in the 1990s have led to large, and apparently arbitrary, size differences in the SAMS areas in different parts of the country. A striking example is that the SAMS areas are significantly smaller in central Gothenburg (Sweden's second largest city) than in central Stockholm (the largest city).[29] The municipalities are also very different in size, ranging from the smallest municipality, Bjurholm, with a population of 2450, to the largest, Stockholm, with over 960,000 inhabitants.[30] While this does not rule out that SAMS or municipalities are sometimes the appropriate spatial measures, it seems plausible that for many research problems these sizeable differences in scale could be problematic.

4.2 Spatial definition of the outcome variable

In an effort to deal with the MAUP, and starting out with the question of how to best measure the school location variable, I note that in principle, every "spot on the map"—every coordinate point—constitutes a potential location for an entering independent school. That said, letting each coordinate point constitute a separate location unit is not a feasible option for the regression analysis, as it would result in low statistical power due to too many (too small) location spots.[31] I therefore aggregate the

---

[26] As pointed out by Wong (2009), increasing the scale tends to reduce the variation between spatial units, by "smoothing out" the variation in the data over a larger area.

[27] Böhlmark et al. (2016) use SAMS in a study on school choices and segregation, and Holmlund et al. (2014) and Angelov and Edmark (2016), use SAMS in studies of school location patterns. See also Amcoff (2012) for an extensive overview of references to literature using SAMS areas in neighbourhood analyses.

[28] See Böhlmark and Lindahl (2015) for an example in the education related literature.

[29] Starting from 2018, the SAMS are replaced with a new local are unit, called DeSo, see https://www.scb.se/hitta-statistik/artiklar/2018/demografiska-statistikomraden-en-ny-regional-indelning-under-kommuner/. The differences in SAMS across the country, dating back to the different methods that were used to generate SAMS, are in this webpage mentioned as a drawback with the SAMS.

[30] The figures refer to 2018, see Statistics Sweden: https://www.scb.se/hitta-statistik/statistik-efter-amne/befolkning/befolkningens-sammansattning/befolkningsstatistik/pong/tabell-och-diagram/helarsstatistik-kommun-lan-och-riket/folkmangd-i-riket-lan-och-kommuner-sista-december-och-befolkningsforandringar/.

[31] The reason is that the location spots that an entering school is in practice choosing among are likely to be much larger in size than geographical coordinate points. Defining school markets based on all geographical coordinate points would hence give the statistical analysis too low power. Intuitively, this fol-

Fig. 3 Maps over the regression sample grid cells for the 1 km×1 km grid, the location of independent schools offering grades 1–3 in 2000, and the population density of 7–9-year olds in 1991

coordinate points to a larger grid, consisting of 1 km×1 km squares.[32] The choice of exactly 1 km×1 km sized grid cells is arbitrary, and as an alternative, I also provide results from using a smaller 0.5 km×0.5 km sized grid, thus changing the scale of the spatial units in the terminology of the MAUP.

The generated grid cells constitute the locations that can be chosen by the entering independent school. The outcome variable of the regression analysis will thus be defined as a dummy variable which takes the value one if at least one independent school has started up in the grid cell during the period of study, and zero otherwise.[33] Grid cells located in low-populated areas, defined as having fewer than 30 students residing within a 3 km radius, will, however, be excluded from the regression sample, as they are unlikely to be considered by entering schools. The resulting regression sample of grid cells for the 1 km×1 km specification is shown in the left-hand side map in Fig. 3. As indicated by the middle and right-hand side map (which

---

Footnote 31 (continued)

lows from the fact that spots that are very close to a chosen location would count as not chosen locations (zeroes in the binary outcome variable) even though they would in fact belong to the category of spots that were chosen by an entering independent school (ones in the binary outcome variable).

[32] This was done using the command spmap in Stata.

[33] The outcome variable is generated based on geographical information on schools in the form of 100 m×100 m squares for the geographical coordinates of the school addresses.

are copied from Fig. 2), the regression sample grid cells cover the more populated areas of Sweden and vast majority of the actual independent school location choices.

In addition to these grid-based specifications, I will, as previously commented, also provide results from using SAMS and municipality as unit of analysis for the sake of comparison.

4.3 Spatial definition of the explanatory variables

For the explanatory regression variables, the spatial aggregation will vary depending on what is suitable for the specific variable, and depending on the restrictions posed by the data at hand.

First, the regression variables for student background, i.e. parental education level, immigrant background, disposable income and GPA, and for the voucher level,[34] shall be generated such that they reflect the characteristics of the pool of students that an entering school can be expected to attract if it chooses a certain location. Taking this into account, I define the following four alternative spatial measures:

(i)   Students residing within a 3-km radius from a grid cell midpoint.
(ii) Students residing within 1.5-km radius from a grid cell midpoint.
(iii) The 50 students residing nearest the grid cell midpoint, and
(iv) The 100 students residing nearest the grid cell midpoint.[35]

The two first measures are based on the notion that primary school students are likely to prefer schools located near their home and use fixed cut-offs in order to measure proximity.[36] The precise cut-offs are chosen ad hoc, but they reflect the assumption that primary level students are likely to prefer schools relatively near home. Whereas I lack data on the school of attendance for students of primary school age, the median distance to the school of attendance for students in the final grade of compulsory school, who are typically 16 years of age, was 1.6 km in 2000, and the average distance was 7.2 km. This suggests that, even among students much older than the 7–9-year olds that are of relevance to this study, most attend a school in the near vicinity. This link is likely to be even stronger for younger students.

Alternatives (iii) and (iv) instead take into account the fact that what is viewed as an acceptable travel distance is likely to differ across regions, for example depending on access to public transport. They thus assume that it is the students residing

---

[34] The level of the voucher depends on the student's home municipality, and the voucher level thus depends on the home municipality of the students attending the school. An alternative would be to simply measure the voucher level in the municipality of location, but this would ignore the fact that an independent school that locates near a municipality border may very well attract students from the bordering municipality.

[35] In order to speed up the calculations of the nearest neighbour measures, which were calculated using Stata's Mata program, I imposed the restriction that only students residing within 100 km from the grid cell midpoint were included.

[36] All distance-based variables are based on data on coordinate pairs for residential addresses, in the form of 100 m × 100 m squares. The distances are measured as the crow flies. Actual travel distances for different modes of transport could not be computed due to lack of access to the necessary software.

nearest the school that are likely to be more interested in the school, without explicitly taking into account the travel distance.[37]

Second, the variable for population density will be measured using the same cut-off values as in the two first student-based measures above (note that the nearest-student type alternatives (iii) and (iv) are not useful for this variable), since this variable is relevant for the entering school to the extent that it affects the number of students who are potentially interested in attending it. That is, population density is in the regression analysis defined as the number of age 7–9 individuals residing within 3 km from a grid cell midpoint, or alternatively, within 1.5 km from a grid cell midpoint.

Third, the variables for school density will be measured using double the cut-off distances for the population density measure, namely 6 km and 3 km from a grid cell midpoint, respectively. This is based on the notion that if schools are assumed to attract students within a given radius (3 km and 1.5 km above), then they are expected to compete with schools within twice that distance.

For the variable municipal political majority, finally, the relevant geographical area is naturally the municipality, so each observation—whether based on grid cell, SAMS, or municipality—will for this variable be assigned the value of political majority corresponding to its municipality. Furthermore, the proxy variable for costs for school premises is also measured at the level of the municipality, although in this case due to lack of more detailed information.

In addition to the above tailor-made spatial measures, I will also show results for two commonly used measures of local areas for Sweden, SAMS, and municipalities. In those regressions, both the explanatory and the outcome variables are measured at the SAMS/municipality level.

The analysis will thus be carried out on a set of alternative regression variables, based on slightly different spatial assumptions. The aim is, as stated above, to evaluate if the results change when the different measures are used, thereby indicating if the MAUP is present in the current setting.

Table 1 summarizes descriptive statistics for the alternative variables based on the 1 km × 1 km grid cells, and for the SAMS and municipality level measures. (Descriptive statistics for the variables when using 0.5 km × 0.5 km grid cells are available in the working paper version of the article, see Edmark 2018.) The table uses the following abbreviations for the variable names: *dS2000* denotes the binary outcome variable defined based on the period 1992–2000, and *NrS2000* denotes the continuous outcome measures (for number of start-up schools) for the same periods. *GPA* denotes the grade point average, *High educ parent* denotes the variable for parental education level, *Fam disp inc* is the family disposable income measure, and *Std Fam disp inc* gives the standard deviation for the same measure. *Sw parent* denotes the variable for Swedish-born parent, *School-age pop* is the density among the school-age population, and *Voucher proxy* denotes the proxy variable for the

---

[37] "K nearest neighbours" measures have been used to define neighbourhoods in several studies, for example Östh et al. (2015) and Hennerdahl and Meinild Nielsen (2017). The program Equipop has been developed to calculate measures of K nearest neighbours, see http://equipop.kultgeog.uu.se/. The measures used in this study were, however, calculated using Stata (Mata).

Table 1 Descriptive statistics for all generated regression variables for the 1 km×1 km grid, and for the municipality and SAMS-level specifications

Outcome variables		1 km×1 km grid cell location units				Municipality	SAMS
Mean of	dS2000	0.0079				0.4161	0.0551
dS2000	Nr of ones	400				119	192
dS2000	Nr of zeroes	50308				167	3293
Mean of	nrS2000					1.3531	0.0666
Sd of	nrS2000					3.4809	0.2995
Student background variables		≤3 km	≤1.5 km	50 nearest	100 nearest		
Mean of	GPA9	3.19	3.19	3.20	3.20	3.19	3.20
Sd of	GPA9	0.18	0.41	0.14	0.11	0.09	0.25
Mean of	High educ parent	0.34	0.34	0.35	0.35	0.33	0.38
Sd of	High educ parent	0.13	0.24	0.15	0.13	0.10	0.18
Mean of	Fam disp inc	260844	260071	262325	261121	254534	262080
Sd of	Fam disp inc	31643	95053	38652	33047	23881	51464
Mean of	Std Fam disp inc	105280	90660	97822	104790	121401	105213
Sd of	Std Fam disp inc	116305	124291	120516	118883	101683	140776
Mean of	Sw parent	0.93	0.93	0.93	0.93	0.92	0.90
Sd of	Sw parent	0.07	0.11	0.07	0.06	0.06	0.10
Mean of	Voucher proxy	46852	46707	46848	46846	48050	47634
Sd of	Voucher proxy	4555	4414	4557	4511	5866	4692
Municipality level variables							
Mean of	Left	0.23				0.25	0.21
Sd of	Left	0.42				0.44	0.41
Mean of	Costs premises	11029				11347	11527
Sd of	Costs premises	2592				2911	2743
School and population density		≤3 km	≤1.5 km				
Mean of	School-age Pop	171	43			1003	65
Sd of	School-age Pop	299	95			1478	47
		≤6 km	≤3 km				
Mean of	Nr mun schools	5.60	1.87			13.76	0.69
Sd of	Nr mun schools	8.86	2.84			14.22	0.71
Mean of	Nr priv schools	0.15	0.04			0.21	0.01
Sd of	Nr priv schools	0.93	0.35			1.20	0.11
Number of observations		≤3 km	≤1.5 km	50 nearest	100 nearest		
Mean of	Nr of obs age 7–9[a]	170.52	43.20	50.00	100.00	1002.96	64.93
Sd of	Nr of obs age 7–9	298.66	95.27	0.00	0.00	1478.16	47.26
Mean of	Nr of obs GPA9[b]	61.45	15.59	50.00	99.97	360.47	22.33
Sd of	Nr of obs GPA9	96.92	32.44	0.00	1.19	464.89	17.27
Nr of	Grid cells/ Municip/SAMS	50708	50708	50708	50708	286	3485[c]

Table 1 (continued)

ᵃThe number of observations per unit refers to the number of observations used for the calculations of the student background variables, in the alternative specifications

ᵇIt can be noted that the average number of observations for the 100 nearest students is slightly lower than 100. This is due to the occurrence of missing observations, and to the fact that the calculations for the nearest neighbour specifications was limited to students residing within 100 km from the grid cell, in order to facilitate the computation

ᶜNote that only SAMS areas with at least 30 age 7–9 students, measured in 1991, are included in the regression data. This explains why the number of included SAMS areas falls far short of the total number of approximately 9000 SAMS areas

voucher level. *Left* denotes the local political majority variable, and *Costs premises* gives the proxy variable for costs for premises. Finally, *Mun sch density* and *Indep sch density* denote the density variables for municipal and independent schools, respectively.

5 Regression model and results

5.1 Regression model and estimation issues

The likelihood that an independent school chooses location g in municipality m, $P(y_{gm})$, is modelled as a logistic function of the matrix of local characteristics X_{gm}, which implies the following regression equation[38]:

$$P(y_{gm}|X_{gm}) = \frac{\exp\left(\alpha + X'_{gm}\beta + \varepsilon_{gm}\right)}{1 + \exp\left(\alpha + X'_{gm}\beta + \varepsilon_{gm}\right)}. \tag{1}$$

As previously mentioned, the outcome variable is based on independent school start-ups in 1992–2000, and the explanatory variables are measured in 1991 in order to avoid endogeneity bias. The downside of measuring local characteristics in 1991, however, is that they will be imperfect measures for the local characteristics in later years, and likely more so the further we move in time from the starting year. The correlation between the variables over time is, however, quite strong, which suggests that this is not a major issue.[39,40]

---

[38] The working paper version of this paper includes additional results from penalized maximum likelihood estimations, and standard errors following Conley (1999). The results from these alternative specifications are very similar to the results that are presented here.

[39] Table A2 in the working paper version of this article (Edmark 2018) shows the correlation between the explanatory variables measured in 1991 and 1995.

[40] One can also note that the potential measurement problem is likely to be smaller for the shorter alternative outcome period of 1992–1995, than for the longer outcome period 1992–2000. Comparing the results from these specifications hence gives a hint on whether or not measurement error affects the results. It is thus reassuring that the results are generally very stable across specifications for different time periods. See Table A5 in the working paper version of this article (Edmark 2018).

5.2 Main results

Table 2 shows the regression results for the outcome period 1992–2000 for the set of alternative local school market specifications. The results in Table 2 are presented in the form of elasticities, in order to facilitate comparison between the specifications. Note that all columns show the results of logit estimations for the binary outcome variable for independent school locations, except for columns (6) and (8) which show the results from Poisson regression models using the continuous outcome variables measuring the number of independent schools opening up in SAMS (column 6) and municipalities (column 8). These estimations are shown as a complement to the binary specifications for the SAMS and municipality regressions, and are in particular relevant for the municipalities, which often have several independent schools.

The results in Table 2 suggest that the probability for independent school entry is higher in locations where a larger share of the student population has high-educated parents, and lower where a larger share of the student population has Swedish-born parents. It is furthermore positively correlated with the local student population density and is negatively correlated with a left-wing municipal political majority. In terms of estimate sizes, the estimated elasticities are the largest for the share of students with Swedish-born parents: a 1% increase in this variable is estimated to be correlated with a 2–4% lower probability of independent school location, depending on the specification. The second largest elasticity, 0.7–1.8%, is estimated for the share of students with high-educated parents.

For the variable left-wing political majority, it is more intuitive to express the estimate size in terms of a 0–1 change than in terms of the elasticities given in the table. Expressed in this manner, the results suggest that locations in municipalities with a left-wing political majority are on average 0.6–1% less likely than others to experience independent school entry.[41]

The above results are consistent for all of the alternative school market definitions, except for a statistically insignificant estimate on population density in the municipality level regression in column 8. Some of the other estimated elasticities, however, differ markedly depending on the level of spatial aggregation. For example, the elasticities in columns (1)–(3) of Table 2 suggest that a 1% higher standard deviation in household income is correlated with an approximately 0.1% increase in the likelihood that an independent school opens up. However, this relation turns insignificant when measured among the 100 nearest students in column (4), and is insignificant, and sometimes even changes sign, in the SAMS and municipality level specifications in columns (5)–(8).

In this case, which of the specifications shall we trust? As described in Sect. 4.3, the alternative grid-based spatial measures were derived from slightly different measures on students' willingness to travel to school, and the resulting impact on competition between schools. However, our knowledge on students' actual willingness to travel to school is limited, and so we cannot distinguish, based on the results, what is the more correct specification. What we can say, however, from the MAUP,

---

[41] Estimates are available upon request.

Table 2 Average marginal effects: elasticities, outcome period 1992–2000. Population density ≤3 km, school density ≤6 km. 1 km×1 km grid

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Outcome variable	dS2000	dS2000	dS2000	dS2000	dS2000	nrS2000	dS2000	nrS2000
Specification student variables	≤3 km	≤1.5 km	50 nearest	100 nearest	SAMS	SAMS	Municip	Municip
GPA	−3.184**	−0.0217	−0.270	0.241	−2.771**	−2.689**	−8.303**	−11.37***
	(1.402)	(0.378)	(1.624)	(2.598)	(1.131)	(1.146)	(3.922)	(3.661)
High educ parent	1.779***	0.719***	1.457***	1.531***	1.105***	1.035***	1.007*	1.740***
	(0.182)	(0.0645)	(0.166)	(0.203)	(0.183)	(0.177)	(0.537)	(0.476)
Fam disp inc	−0.398	−0.230	−1.850*	−1.064	−0.255	−0.205	3.586*	0.798
	(1.036)	(0.276)	(0.953)	(0.954)	(0.591)	(0.575)	(1.908)	(0.889)
Std Fam disp inc	0.131**	0.0978**	0.140**	0.0851	−0.0190	0.00718	−0.0771	0.0304
	(0.0572)	(0.0446)	(0.0686)	(0.0557)	(0.0657)	(0.0716)	(0.109)	(0.0914)
Sw parent	−4.171***	−2.182***	−3.032***	−3.834***	−2.459***	−1.973***	−3.084*	−6.232***
	(0.877)	(0.339)	(0.950)	(0.843)	(0.650)	(0.656)	(1.733)	(1.137)
School-age Pop	0.275***	0.297***	0.256***	0.256***	0.284***	0.201***	0.422*	0.298
	(0.0657)	(0.0664)	(0.0558)	(0.0596)	(0.0853)	(0.0746)	(0.226)	(0.193)
Voucher proxy	−0.873	−0.608	−0.604	−0.640	−3.322**	−2.978**	0.255	−2.809*
	(1.737)	(1.721)	(1.793)	(1.791)	(1.504)	(1.431)	(1.822)	(1.573)
Left	−0.194***	−0.211***	−0.227***	−0.212***	−0.233***	−0.254***	−0.210**	−0.272***
	(0.0653)	(0.0610)	(0.0665)	(0.0656)	(0.0703)	(0.0715)	(0.0898)	(0.0603)
Costs premises	−0.0661	0.107	0.0397	−0.0328	1.234**	1.129**	−0.208	0.613
	(0.692)	(0.712)	(0.713)	(0.687)	(0.546)	(0.563)	(0.689)	(0.595)
Mun sch density	−0.0801	−0.0269	−0.0278	−0.0481	0.135*	0.125	0.0315	0.205
	(0.0729)	(0.0664)	(0.0668)	(0.0706)	(0.0742)	(0.0765)	(0.269)	(0.163)
Indep sch density	−0.00733*	−0.00581	−0.00754	−0.00552	−0.000234	0.000668	−2.76e−05	−0.0613***
	(0.00445)	(0.00683)	(0.00484)	(0.00467)	(0.00400)	(0.00410)	(0.0141)	(0.0216)
Observations	50,325	37,875	50,292	50,304	3361	3361	283	283

Table 2 (continued)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Outcome variable	dS2000	dS2000	dS2000	dS2000	dS2000	nrS2000	dS2000	nrS2000
Specification student variables	≤ 3 km	≤ 1.5 km	50 nearest	100 nearest	SAMS	SAMS	Municip	Municip
Specification	No added dummies	No added dummies	No added dummies	No added dummies	No added dummies	No added dummies	No added dummies	No added dummies
Estimator	Logit	Logit	Logit	Logit	Logit	Poisson	Logit	Poisson
Log-likelihood	−1806	−1771	−1822	−1820	−657.9	−789.2	−132.7	−331.1

Standard errors are clustered on municipality in parentheses, except for the SAMS-level regressions, where standard errors are clustered on the SAMS level

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

is that using too large a scale in general tends to reduce the variance in the data, by smoothing out local extreme values. This in turn makes it harder to identify underlying associations that occur at a smaller scale. It is possible that the above-mentioned insignificant estimates for the standard deviation in household income in columns (4)–(8), i.e. the specifications using the larger of the "nearest students" measures, SAMS and municipalities, are a result of using too large a scale.

The local student grade point average (GPA) is furthermore estimated to be strongly and statistically significantly correlated with the likelihood for independent school location for the specifications using a 3-km cut-off in column (1) and the SAMS-based specifications in columns (5)–(6), and even more so when using the municipality level in columns (7)–(8). The elasticity is, however, small and statistically insignificant for the other specifications in columns (2)–(4). The estimate for the proxy variable for costs for premises also varies a lot across the specifications. While this variable is always measured at the municipality level, the alternative specifications still differ in the level at which school entry is measured. In this case, the SAMS-level specifications in columns (5)–(6) stand out by yielding statistically significant and positive elasticities, i.e. of opposite direction of what we expected theoretically. The estimates are, however, insignificant and of much smaller magnitude, and sometimes of negative sign, for the other specifications. I deem it likely that the large and positive estimates for the SAMS-level specifications are the result of some omitted variable which correlates with the independent school entry variable at the SAMS level. In general, the variation in the estimates for GPA and costs for premises across the specifications is in itself an indication of the MAUP—it exemplifies how the level of aggregation can impact the estimated results. However, just as for the estimate on the standard deviation in household income above, we cannot deduct from the results which one of the specifications is more correct.

5.3 Additional robustness tests

In addition to the above estimations, the independent school location choice was estimated after making the following alterations to the baseline specification of Table 2[42]:

- Labour market region dummies were added to the regressions, in order to account for the fact that teacher wage levels may differ between regions. For the grid cell-based specifications, this had little impact on the overall results. The most striking differences were that the estimated positive elasticity for the standard deviation of household income came out as statistically significant in all grid cell-based specifications, and the positive estimate for the elasticity for the average disposable income level was statistically significantly different from zero in all specifications. The dummy variable for left-wing political majority was no longer statistically significant. This is, however, not surprising given that

---

[42] Results are available in the working paper version of this article, see Edmark (2018).

the Labour market regions are likely to capture a lot of the municipality level variation. The SAMS-level estimates were overall more sensitive to the inclusion of Labour market region dummies than the grid cell-based estimations[43]: the elasticities for the share of Swedish-born parents and the cost for premises were no longer statistically significantly different from zero. In addition, a negative and statistically significant elasticity was estimated for the average disposable income.

- Estimating the regression for alternative outcome years (1992–1995, and 1992–2005, respectively), and estimating the regression for outcome years 1996–2000 and measuring explanatory variables in 1995. The overall pattern of results was similar to the baseline across specifications. The main difference was that the left-wing majority in the municipal council was not always statistically significantly different from zero.
- Using a smaller (500 m×500 m) grid (instead of the baseline 1 km×1 km size grid) to measure potential location points. This yielded results that were very similar to the baseline specification.
- Defining population density as number of students residing within 1.5 km from the grid cell (instead of the 3 km in the baseline specification) and measuring school density within 3 km (instead of 6 km). Most estimates were unaltered by this; however, the standard deviation of income, which was statistically significant in most specifications in the baseline case, was now insignificant in most of the specifications. Municipality school density, which was previously insignificant, was on the contrary now positive and statistically significant in all four specifications.
- Adding dummy variables for municipality type, based on the categories defined by the Swedish Association of Local Authorities and Regions,[44] and estimating the regression only for the municipalities within the Stockholm County. The results were overall very similar to the baseline specification. The more prominent difference was that the estimated elasticities for the average and standard deviation in local disposable income were more often statistically significantly different from zero than in the baseline specifications.

6 Concluding discussion

The overall results of this study suggest that the likelihood for independent school entry during the studied period was higher in locations where a larger share of school-age children had high-educated parents and where the local population

---

[43] This is not surprising, as the number of SAMS is much lower than the number of grid cells, meaning that the Labour market region dummies capture relatively more of the variation among observations for the SAMS specification.

[44] These categories are: large cities, suburban, municipalities with many commuters, large towns, manufacturing towns, other municipalities with a population of more than 25,000, other municipalities with a population of 12,500–25,000, other municipalities with a population below 12,500, and rural municipalities.

density among school-age children was higher. The likelihood was furthermore lower in locations with a larger share of school-age children with Swedish-born parents. These results were stable across the alternative spatial specifications and the alternative regression models that were estimated.

These results are well in line with previous results on the USA and Sweden. Many of the US studies suggest that the likelihood for charter or private school location is positively correlated with the local adult education level and with a higher level of dispersion in terms of ethnicity or higher shares of students with foreign background. The previous Swedish reports, Angelov and Edmark (2016) and Holmlund et al. (2014), also suggest a higher likelihood for independent school entry in locations with a larger share of students from high-educated or foreign family background, and in more densely populated areas.

The results also showed that a left-wing political majority in the local council was negatively correlated with independent school entry in the baseline specification, although this correlation was not always statistically significant when alternative outcome periods were used.

The independent schools furthermore tended to locate in areas with a higher income dispersion, although this result was not stable across all spatial specifications. Other variables were even more sensitive to variations in the spatial aggregation measures. This held for example for the average GPA among local students, for the level of local household income, and for the proxy variables for local voucher levels and estimated costs for facilities. Although the analysis was not a formal test of which spatial aggregation was the best, the results indicate that the spatial unit of analysis can have significant impact on the estimated results, thus highlighting the potential importance of the MAUP.

References

Amcoff J (2012) Hur bra fungerar SAMS-områdena i studier av grannskapseffekter? Socialvetenskaplig tidskrift nr 2

Angelov N, Edmark K (2016) När skolan själv får välja – en ESO-rapport om friskolornas etableringsmönster. Rapport till expertgruppen för studier i offentlig ekonomi (ESO) 2016:3

Barrow L (2006) Private school location and neighbourhood characteristics. Econ Educ Rev 25:633–645

Bifulco R, Buerger C (2015) The influence of finance and accountability policies on location of New York State Charter Schools. J Educ Finance 40(3):193–221

Böhlmark A, Lindahl M (2015) Independent schools and long-run educational outcomes: evidence from Sweden's large-scale voucher reform. Economica 82(327):508–551

Böhlmark A, Holmlund H, Lindahl M (2016) Parental choice, neighbourhood segregation or cream skimming? An analysis of school segregation after a generalized choice reform. J Popul Econ 29(4):1155–1190. https://doi.org/10.1007/s00148-016-0595-y

Burdick-Wills J, Keels M, Schuble T (2013) Closing and openings schools: the association between neighbourhood characteristics and the location of new educational opportunities in a large urban district. J Urban Aff 35(1):59–80

Conley TG (1999) GMM estimation with cross sectional dependence. J Econom 92(1):1–45

Downes T, Greenstein S (1996) Understanding the supply decisions of nonprofits: modelling the location of private schools. Rand J Econ 27(2):365–390

Edmark K (2018) Location choices of Swedish independent schools—how does allowing for private provision affect the geography of the education market? Working Paper No 11/2018, Swedish Institute for Social Research (SOFI), Stockholm University

Ferreyra M, Kosenok G (2018) Charter school entry and school choice: the case of Washington, D.C. J Public Econ 159:160–182

Fotheringham AS, Wong D (1991) The modifiable areal unit problem in multivariate statistical analysis. Environ Plan A Econ Space 23(7):1025–1044

Glomm G, Harris D, Lo T (2005) Charter school location. Econ Educ Rev 24:451–457

Henig JR, MacDonald JA (2002) Location decisions of charter schools: probing the market metaphor. Soc Sci Q 83(4):962–980

Hennerdahl P, Meinild Nielsen M (2017) A multiscalar approach for identifying clusters and segregation patterns that avoids the modifiable areal unit problem. Ann Am Assoc Geograph 107(3):555–574

Holmlund H, Häggblom J, Lindahl E, Martinson S, Sjögren A, Vikman U, Öckert B (2014) Decentralisering, skolval och fristående skolor: resultat och likvärdighet i svensk skola. IFAU Report 2014:25

Koller K, Welsch DM (2017) Location decisions of charter schools: an examination of Michigan. Educ Econ 25(2):158–182

Mehta N (2017) Competition in public school districts: charter school entry, student sorting, and school input determination. Int Econ Rev 58(4):1089–1116

Östh J, Clark AW, Malmberg B (2015) Measuring the scale of segregation using k-nearest neighbor aggregates. Geograph Anal 47(1):34–49

Rincke J (2007) Policy diffusion in space and time: the case of charter schools in California School Districts. Reg Sci Urban Econ 37:526–541

Wong D (2009) The modifiable areal unit problem (MAUP). In: Fotheringham AS, Rogerson PA (eds) The SAGE handbook of spatial analysis. SAGE Publications Ltd, Thousand Oaks. https://doi.org/10.4135/9780857020130.n7