

Svenska friskolors etableringsbeslut

Karin Edmark

Svenska friskolors etableringsbeslut^a

av

Karin Edmark^b

2018-10-12

Sammanfattning

Den här rapporten undersöker hur fristående lågstadieskolors etableringsbeslut samvarierar med mått på lokala karakteristiska, genom att studera friskolor som öppnade efter friskolereformen 1992. Resultaten visar att det är högre sannolikhet att friskolor öppnar i områden där en stor del av elevunderlaget har högutbildade föräldrar och där befolkningstätheten är hög. Det är vidare lägre sannolikhet att friskolor etableras i områden där en stor del av eleverna har svenskfödda föräldrar. Ovanstående resultat ges för en rad alternativa geografiska områdesdefinitioner, som används för att undersöka om de skattade sambanden påverkas av den geografiska aggregeringsnivån. Resultaten överensstämmer också med vad med vad tidigare liknande undersökningar funnit.

^a Den här rapporten är en sammanfattning av en mer omfattande version på engelska, se Edmark (2018). Jag är tacksam för kommentarer och synpunkter från Eirini Tatsi, Özge Öner, Helena Nilsson, Nikolay Angelov, Helena Holmlund, Caroline Hall, Mika Haapanen, Linuz Aggeborn, Mattias Öhman, Lina Hedman, Eva Andersson, Bo Malmberg och Pontus Hennerdal, från deltagare vid forskningsseminarier vid Uppsala Center for Labor Studies vid Uppsala universitet, Institutet för social forskning vid Stockholms universitet, Institutet för bostads- och urbanforskning vid Uppsala universitet, Högskolan i Jönköping, samt från deltagare vid 2018 års IIPF-Congress och 2018 års ERSA-kongress. Projektet har finansierats genom bidrag från Vetenskapsrådet genom forskningsprojekt 721-2008-5114 samt 721-2014-1783.

^b Karin Edmark, karin.edmark@sofi.su.se, Institutet för social forskning, Stockholms universitet.

Innehållsförteckning

1	Introduktion.....	3
2	Bakgrund.....	4
3	Datamaterial.....	6
4	Geografiska mått.....	8
5	Regressionsanalys och resultat.....	12
5.1	Regressionsanalys.....	12
5.2	Huvudsakliga resultat.....	13
5.3	Kompletterande resultat.....	18
6	Avslutande sammanfattning.....	19
	Referenser	20
	Appendix.....	21

1 Introduktion

Den här rapporten undersöker fristående lågstadieskolors¹ etableringsbeslut. Närmare bestämt undersöks vilka lokala faktorer som samvarierade med sannolikheten att en friskola öppnade i ett grannskap under åren efter friskolereformen 1992. Undersökningen genomförs med regressionsanalys, och använder ett stort datamaterial som innehåller detaljerad geografisk information om såväl skolor som elevunderlaget.

Friskolornas etableringsbeslut har undersökts i ett par tidigare svenska rapporter (se Angelov och Edmark, 2016, och Holmlund m.fl. 2014). Den här undersökningen skiljer sig från de tidigare på följande punkter: för det första används flera alternativa mått på geografiska områden, eller grannskap, än vad som gjorts tidigare.² Eftersom statistiska mått på lokala områdeskaraktistika tenderar att påverkas av vilken geografisk aggregeringsnivå som används, är det intressant att undersöka om de samband som skattas mellan skoletableringar och lokala faktorer i den här rapporten förändras när olika områdesdefinitioner används. För det andra undersöks etablering av fristående lågstadieskolor, medan tidigare rapporter fokuserat på alla grundskolor respektive högstadieskolor. För det tredje finns det vissa skillnader i vilka lokala faktorer som undersöks och hur de definieras.

Dessa skillnader till trots, är de huvudsakliga resultaten samstämmiga med tidigare rapporter: friskolor etableras oftare i områden där en stor del av elevunderlaget har högutbildade föräldrar, där befolkningstätheten är högre, samt där en lägre andel av elevunderlaget har svenskfödda föräldrar. Dessa samband är dessutom tydliga oavsett vilken geografisk områdesdefinition som används, och gäller oavsett om regressionsanalysen konstanthåller för andra lokala faktorer eller inte.

Ett annat mönster som framträder, även om det är något mer känsligt för förändringar i analysmetoden, är att friskolor mindre ofta etableras i kommuner med vänstermajoritet i kommunfullmäktige. Även detta resultat framträder oavsett om analysen konstanthåller för övriga lokala faktorer eller inte.

¹ Med detta avses alla skolor som har lågstadieårskurser. Genom att avgränsa undersökningen till just lågstadieskolor undviks risken att dubbelräkna skolor som i skolregistret räknas som flera skolenheter på grund av att olika stadier utgör olika skolenheter. Lågstadieskolor är dessutom intressanta att studera eftersom de troligen har en stark lokal koppling i meningen att elever i de lägre årskurserna ofta bor nära skolan de går i.

² Det ska noteras att också Angelov och Edmark (2016) använder några alternativa mått på lokala områden. Den här undersökningen bidrar dock med fler och andra typer av definitioner.

Det kan vidare påpekas att resultatet för vissa av de studerade variablerna påverkades ganska kraftigt av vilken nivå på geografisk aggregering som användes. Till exempel tyder resultaten för många av de områdesdefinitioner som används på ett positivt samband mellan den lokala inkomstspridningen och sannolikheten för friskoleetablering, men sambandet är inte alltid statistiskt säkerställt. För variabeln för elevernas medelbetyg varierar vidare det skattade sambandet så kraftigt mellan definitionerna att ingen slutsats kan dras. En generell lärdom av detta är att det kan vara lämpligt att undersöka hur användandet av olika geografisk aggregeringsnivå påverkar resultaten i undersökningar som använder variabler på grannskapsnivå.

I följande avsnitt ges först en bakgrund av friskolereformen och den efterföljande expansionen av friskolor, vilket följs av beskrivningar av datamaterialet och de geografiska områdesmått som används i analysen. Därefter presenteras rapportens metod och resultat, samt en avslutande sammanfattning.

2 Bakgrund

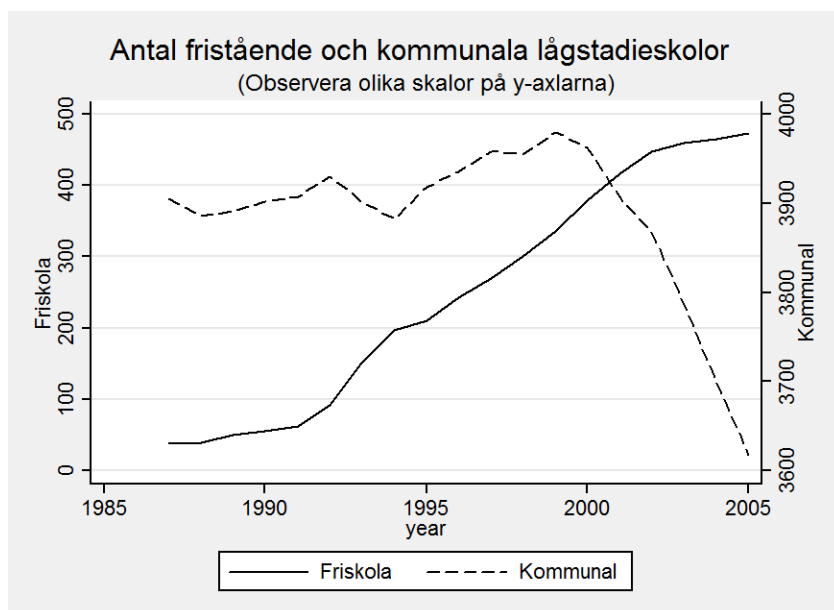
Friskolereformen 1992 innebär att det offentliga ekonomiska stödet till friskolorna kom att höjas kraftigt, och hamnade ungefär i nivå med ersättningen till de kommunala skolorna. Dessutom slopades den tidigare begränsningen att ersättning bara skulle utgå till friskolor som antingen erbjöd särskild pedagogik, var internatskolor eller utgjorde skolor för utländska elever. Samtidigt var de formella kraven för att få tillstånd att starta friskola tämligen låga under friskolereformens första årtionde³, och växande elevkullar innebär ett ökande elevunderlag.

Sammantaget var förutsättningarna för att starta friskolor således goda, och antalet friskolor växte kraftigt under åren efter reformen. Figur 1 nedan illustrerar den kraftiga expansionen av fristående lågstadieskolor under framför allt 1990-talet, och Figur 2 visar att andelen elever i friskolor ökade markant. Även antalet kommunala lågstadieskolor växte under 1990-talets andra halva, för att sedan sjunka från och med 00-talet, då även kohorterna av elever i lågstadieåldern minskade (se Figur 2).

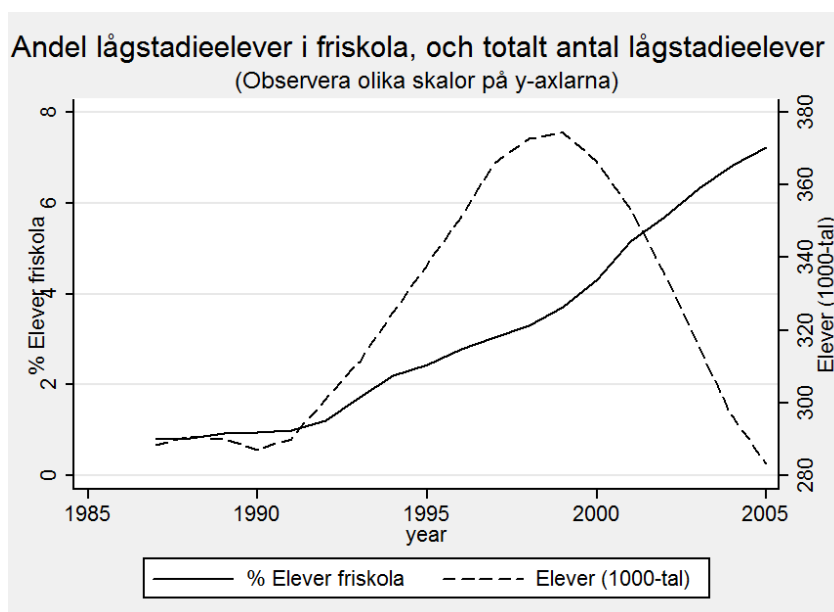
Den dramatiska ökningen av antalet friskolor innebär att det finns goda möjligheter att undersöka de fristående skolornas val av etableringsställe, helt enkelt eftersom det finns många etableringsbeslut att undersöka. I följande avsnitt beskrivs det datamaterial som används i den här undersökningen.

³ Se Angelov och Edmark (2016) för en beskrivning av ansökningsprocessen för att starta en friskola.

Figur 1 Antal fristående och kommunala lågstadieskolor



Figur 2 Andel elever i friskola och totalt antal elever i lågstadiet



3 Datamaterial

Den här rapporten drar nytta av friskoleexpansionen som illustrerades i avsnittet ovan genom att undersöka hur de friskoleetableringar som skedde under perioden 1992–2000 samvarierar med lokala områdeskaraktistika uppmätta år 1991, precis före friskolereformen. I den engelska versionen av rapporten visas även resultat för alternativa tidsperioder.

Undersökningen underlättas av att det för den här perioden finns tillgång till detaljerad geografisk information om skolor och elever, liksom flera variabler som kan tänkas vara relevanta för friskolornas etableringsbeslut. Undersökningen använder närmare bestämt registerbaserat datamaterial från Statistiska Centralbyrån (SCB) som innehåller geografisk information över skolor och elevers bostäder, i form av koordinatrutor på 100m*100m precisionsnivå. På skolnivå finns dessutom information över huruvida huvudmannen är fristående eller kommunal, samt om antal elever per årskurs. På elevnivå innehåller datamaterialet uppgifter om familjebakgrund i form av föräldrarnas födelseland och utbildningsnivå, samt hushållens disponibla inkomst. För eleverna finns vidare information om slutbetyget i årskurs 9. Flergenerationsregistret används för att koppla föräldrar till barn. Rapporten använder också information från SCB över fördelningen av partier i kommunfullmäktige, samt information från Skolverket över kommunernas resurstilldelning till skolorna. För en mer detaljerad beskrivning av datamaterialet, se den engelska versionen av den här rapporten, Edmark (2018).

På basis av den ovanstående informationen tas ett antal variabler fram som används för att skapa mått på lokala områdeskaraktistika. Variablerna har valts ut baserat på en enkel teoretisk modell för friskolors val av etableringsplats.⁴ Avsikten är att ge vägledning om vilka faktorer som varierar mellan olika etableringsställen och som kan tänkas påverka en friskolas val av etableringsplats. Modellen inkluderar variabler som påverkar en skolas ekonomi, såsom hur mycket ersättning skolan kan förvänta sig från elevernas hemkommuner, och om tillräckligt många elever förväntas vara intresserade av skolan för att fylla klassrummen. Det senare fångas upp genom variabler för områdets elevtäthet samt förekomsten, och kvaliteten, av andra, konkurrerande, skolor i området. Skolkvalitet är visserligen svårt att mäta, och i den här rapporten används de lokala elevernas medelbetyg som en proxy för detta.

Modellen inkluderar också variabler för elevernas familjebakgrund, som hushållets inkomster och föräldrarnas utbildnings- och utrikes bakgrund.

⁴ Den utförliga modellen presenteras i den engelska versionen av den här rapporten, Edmark (2018). Här anges bara en översiktlig sammanfattning.

Dessa variabler avser att fånga upp att skolornas huvudmän kan tänkas vara intresserade av att locka vissa grupper av elever till skolan, till exempel för att skolan har en viss profil. Det är också möjligt att elevernas familjebakgrund är relevant för huvudmän av andra skäl. Ett skäl kan vara en önskan att locka ”lätta elever”, t.ex. elever med välutbildade föräldrar, till skolan, ett annat att hjälpa elever från en mer utsatt social bakgrund. Det ska dock understrykas att den här rapporten inte kan fastslå exakt vilka bakomliggande motiv som ligger bakom de samband som uppmäts i regressionsanalysen.

Nedan anges definitionerna för de variabler som används i undersökningen:

- *Högutbildade föräldrar*: mäts som en dummy-variabel som antar värdet ett om minst en av elevens föräldrar har en avslutad eftergymnasial utbildning.
- *Svenskfödda föräldrar*: mäts som en dummy variabel som antar värdet ett om minst en av elevens föräldrar är född i Sverige.
- *Disponibel inkomst*: Definieras som disponibel inkomst på hushållsnivå. Om denna skiljer sig åt mellan föräldrarna, så används moderns mått på inkomst.
- *Ungefärlig variabel för förväntad skolpeng*: Som en proxy-variabel för skolpeng används den genomsnittliga ersättningen på kommunnivå per elev till kommunens egna skolor. Anledningen till att denna används, och inte den genomsnittliga ersättningen per elev i friskolor, är att flera kommuner saknade friskolor under den period som avses. Informationen för den här variabeln har hämtats från Skolverkets databas SIRIS. Eftersom uppgifter för år 1991 inte fanns tillgängliga, används för den här variabeln istället uppgifter för år 1992. Denna och de tre ovanstående variablerna mäts bland elever i 7–9-årsåldern.
- *Medelbetyg från 9ans slutbetyg*: Till skillnad från övriga elevvariabler mäts den här variabeln för äldre elever, när de går ut 9an. Anledningen är att tidigare mått på utbildningsresultat saknas för den tidsperiod som undersöks här. Den här variabeln är därmed ett relevant mått på de lokala lågstadiееlevernars utbildningsresultat bara i den mån detta samvarierar med de äldre elevernas.
- *Ungefärlig variabel för förväntade lokalkostnader*: Som mått på förväntade lokalkostnader används den genomsnittliga ersättningen för lokaler per elev som utgick till kommunala skolor. Datakällan är densamma som för måttet på skolpeng ovan, och måttet finns bara tillgängligt på kommunnivå och som tidigare från år 1992.

- *Vänstermajoritet*: Mäts med en dummyvariabel som antar värdet ett om Socialdemokraterna och Vänsterpartiet tillsammans har minst 50% av platserna i kommunfullmäktige.

4 Geografiska mått

Såsom påpekades i inledningen använder den här rapporten flera olika typer av geografiska områdesmått. För det första används två vanligt förekommande områdesmått, d.v.s. kommuner respektive SAMS-områden. SAMS-områden är ett mått på grannskap, framtaget av Statistiska Centralbyrån. För det andra kompletteras dessa mått med ett antal geografiska aggregeringar som skapats utifrån den teoretiska modellen för skoletableringar med hjälp av de geografiska koordinatrutorna för skolor och elevers bostadsadresser.

Varför används så många olika definitioner? I den geografiska litteraturen är det ett välkänt faktum att uppmätta områdeskaraktistika kan påverkas av den geografiska aggregeringsnivå som används.⁵ Genom att använda ett antal olika geografiska indelningar ges en indikation på hur känsliga resultaten för friskoleetableringar är för detta.

Ett annat skäl till att ta fram kompletterande mått till kommuner och SAMS, är att båda dessa kan kritiserar för bristande homogenitet. Kommunerna varierar som bekant markant i storlek och är i många fall troligen för stora för att utgöra relevanta enheter för friskolornas etableringsval. Även SAMS varierar kraftigt i storlek, och skillnaderna förklaras inte, eller åtminstone inte enbart, av att olika bostadsområden eller grannskap faktiskt kan vara olika stora. Enligt Amcoff (2012) är ett skäl till den bristande homogeniteten att det fanns stora skillnader mellan kommunerna i de underliggande områdesindelningar som användes för att skapa SAMS-måttet. Detta innebär till exempel att SAMS-områden i Stockholms kommun i regel är mycket större – både till yta och till befolkning sett – än i Göteborgs kommun.⁶ Om – och i så fall hur – den här aspekten av områdesindelningen påverkar resultaten av olika undersökningar som använder SAMS går det inte att göra någon generell utsago om. Sannolikt beror det på den specifika fråga som undersöks.

⁵ Fenomenet benämns i den geografiska litteraturen för "the Modifiable Areal Unit Problem". Se exempelvis Wong (2004), Hennerdal och Meinild Nielsen (2017), Östh m fl (2015), Mitra och Builung (2012), samt Andersson och Musterd (2010) för beskrivningar samt praktiska applikationer.

⁶ Det kan påpekas att en ny regional indelning från och med 2018 tagits fram av SCB, så kallade Demografiska Statistikområden, DeSo. DeSo avser att ersätta SAMS, och den bristande homogeniteten i SAMS beskrivs av SCB vara ett motiv till detta.

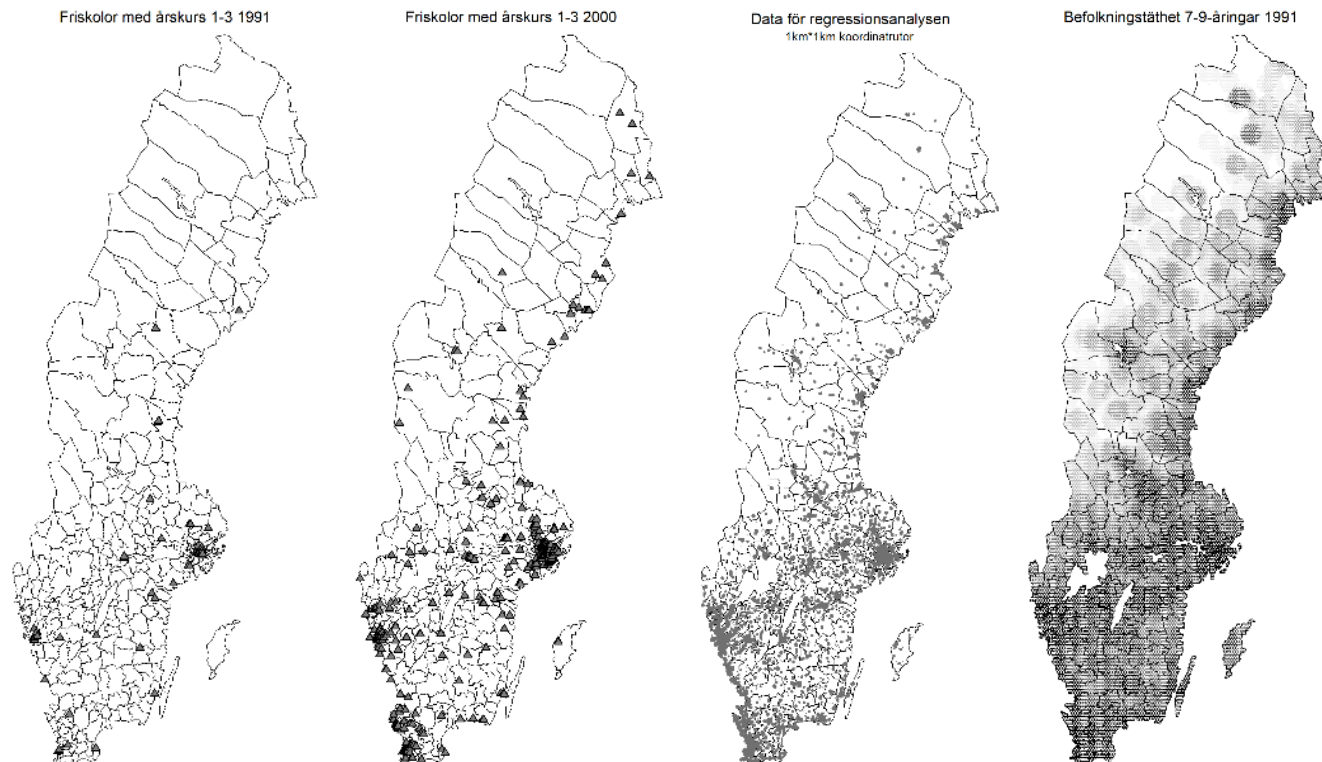
För att undersöka hur känsliga sambanden mellan friskoleetableringar och lokala faktorer är för förändringar av den geografiska aggregeringsnivån används därför, utöver SAMS och kommuner, även följande geografiska mått:

För det första mäts utfallsvariabeln ”etablering av en friskola” med en indikatorvariabel som anger om en friskola etableras inom geografiska rutor av storleken 1km*1km. Det vill säga, hela landet delas in i ett 1km*1km rutnät, och utfallsvariabeln indikerar om en friskola etableras i en sådan ruta.⁷ Det ska påpekas att stora delar av rutnätet täcker obefolkade naturområden som aldrig kommer att vara aktuella som etableringsställe. Dessa rutor tas bort, genom att analysen begränsas till rutor där det bor minst 30 elever inom en radie av 3 km.

Kartorna i Figur 3 nedan visar dels fördelningen av fristående lågstadieskolor över landet under 1991 respektive 2000, dels fördelningen av de rutor som inkluderas i analysen. Dessutom visas en karta över befolkningstätheten bland 7–9-åringar 1991 (mörkare färg indikerar en högre befolkningstäthet). Som kan ses i figuren omfattar rutorna i analysmaterialet mer befolkningstäta områden, och det är föga förvånande också där vi ser att friskolor etableras.

⁷ Som ett alternativ tas även resultat fram som använder ett mindre rutnät av 0,5km*0,5km-rutor. Resultatet för dessa ges i den engelska versionen av den här rapporten, se Edmark (2018).

Figur 3 Kartor över friskolor, de geografiska områden som ingår i det rutbaserade analysmaterialet, samt befolkningstäthet bland 7–9-åringar



Ovanstående beskrivna rutor utgör alltså enheterna för utfallsvariabeln för friskoleetableringarna. Till varje ruta – d.v.s. till varje möjligt etableringsställe – kopplas vidare ett antal mått för de områdesvisa faktorerna, eller högerledsvariablerna i regressionsanalysen.

För vissa av de lokala variablerna finns det bara en möjlig områdesindelning. Detta gäller för variabeln för vänstermajoritet i kommunfullmäktige, vilken naturligen bara varierar mellan kommuner. Det gäller också för proxyvariabeln för förväntade lokalkostnader, eftersom denna bara finns tillgänglig på kommunnivå. För de övriga variablerna är det emellertid möjligt att skapa olika typer av geografiska indelningar, och nedan förklaras hur detta görs.

För det första skapas följande alternativa områdesdefinitioner för de variabler som beskriver elevernas familjebakgrund (högutbildade respektive svenskfödda föräldrar, samt nivå och spridning i disponibel inkomst), samt för det ungefärliga måttet på vilken skolpeng eleverna kan förväntas medföra till skolan.⁸ Notera att variablerna definieras i relation till variabeln för friskoleetableringar, det vill säga de geografiska koordinatrutorna.

- i. Elever som bor inom 3 km från rutans mittpunkt.
- ii. Elever som bor inom 1,5 km från rutans mittpunkt.
- iii. De 50 elever som bor närmast rutans mittpunkt.
- iv. De 100 elever som bor närmast rutans mittpunkt.

De två första måtten bygger på antagandet att det är de elever som bor inom ett visst avstånd från skolan som utgör dess potentiella elevgrupp. De två andra baseras istället på tanken att det är de elever som bor närmast skolan som utgör dess potentiella elever, utan att sätta någon gräns för maximalt avstånd.⁹

Det ska understrykas att dessa mått ska ses som komplement, då de bygger på lite olika antaganden om lågstadielevs (och deras föräldrars) preferenser för skolor. Å ena sidan kan det vara så att de flesta bara är intresserade av skolor som ligger inom ett kort avstånd från hemmet, vilket stämmer in på de två första måtten. Å andra sidan kan det tänkas att vad som är ett acceptabelt avstånd till skolan varierar mellan olika delar av landet, t ex mellan landsort och stad. De två senare måtten, som baseras på de närmsta eleverna men inte anger något specifikt avståndsmått, kan då vara att föredra.

⁸ Detta är ett mått som avser att ge en uppskattning om vilken ersättning friskolan får från elevens hemkommun. För lokala områden som omfattar delar av flera kommuner blir det ett vägt medelvärde av kommunernas ersättningsnivåer.

⁹ Den här typen av mått har använts för att definiera grannskap i flera undersökningar, till exempel Östh m.fl. (2015) och Hennerdahl och Meinild Nielsen (2017). Dataprogrammet Equipop har utvecklats för att beräkna sådana mått, se <http://equipop.kultgeog.uu.se/>. De mått som används i den här undersökningen har dock beräknats genom att använda Mata-funktionen i programmet Stata.

För det andra används två alternativa mått på befolkningstäthet, vilka definieras på basis av gränsvärdena i mått i.–ii. ovan, d.v.s. antal elever som bor inom 3 km respektive 1,5 km från rutans mittpunkt.

För det tredje mäts förekomsten av redan existerande skolor med två alternativa mått, genom att räkna antal skolor inom 6 km respektive 3 km från rutans mittpunkt. Detta bygger vidare på antagandena för upptagningsområdesdefinitionerna i punkt i.–ii., d.v.s. om eleverna är intresserade av skolor inom 3 km (1,5 km) från hemmet, så kommer en nystartad skola komma att konkurrera om elevunderlaget med skolor inom maximalt det dubbla avståndet. För att ta hänsyn till att en nyetablerad friskola kanske ser kommunala och fristående skolor som olika typer av konkurrenter, skapas vidare två mått, ett för kommunala och ett före fristående skolor.¹⁰

Analysen kommer således att genomföras för ett antal alternativa definitioner på lokala områden, och hur dessa definieras skiljer sig dessutom åt mellan variablerna. I Tabell A1 i Appendix ges en översikt av och beskrivande statistik för variablerna. I det efterföljande avsnittet presenteras analysens metod och resultat.¹¹

5 Regressionsanalys och resultat

5.1 Regressionsanalys

Med hjälp av ovan beskrivna variabler genomförs en regressionsanalys över sambandet mellan respektive områdesvariabel och sannolikheten för friskoletablering i området.

Modellen skattas med en logistisk regression, vilket innebär att sambandet mellan indikatorvariabeln för friskoletablering i område g , y_g , och en matris av alla områdesspecifika variabler X_g , antas kunna beskrivas med följande ekvation¹²¹³:

$$P(y_{gm}|X_g) = \frac{e^{\alpha + X_g'\beta + \varepsilon_g}}{1 + e^{\alpha + X_g'\beta + \varepsilon_g}} \quad (1)$$

¹⁰ Det kan dock påpekas att alla lokala variabler mäts 1991, vilket innebär att det här måttet fångar upp det lilla antal friskolor som fanns före friskolereformen.

¹¹ För en tabell inklusive de variabler som används i olika känslighetsanalyser hänvisas till Tabell A1 respektive Table A1 i den engelska versionen av den här rapporten, Edmark (2018).

¹² En mer utförlig beskrivning av modellen finns tillgänglig i den engelska versionen av den här rapporten Edmark (2018).

¹³ För SAMS- och kommunnivå estimeras dessutom alternativa regressionsanalyser som använder en kontinuerlig utfallsvariabel för antal friskolor, istället för den binära indikatorvariabeln för huruvida minst en friskola etableras eller inte. För dessa används istället regressionsanalys baserat på poissonfördelningen.

Regressionsmodellen i ekvation (1) beskriver en analys där alla lokala faktorer inkluderas samtidigt i modellen, och där de skattade sambanden för respektive variabel således gäller givet att man konstanthåller för övriga faktorer. Detta påverkar främst variabler som är korrelerade, som den lokala inkomstnivån och utbildningsnivån. När de samtidigt tas med i modellen kommer till exempel skattningen för den lokala inkomstnivån att mäta sambandet med sannolikheten för friskoleetablering, givet, eller utöver, den lokala utbildningsnivån.

För att ge en mer fullständig beskrivning av sambanden mellan lokala faktorer och friskoleetableringar anges därför också i den här rapporten resultat från separata regressioner för varje variabel. I termer av exemplet ovan kommer den skattade parametern för den lokala inkomstnivån då att inkludera även samband som uppstår på grund av korrelationen med utbildningsnivån, det vill säga genom att områden med hög inkomstnivå även tenderar att ha en hög utbildningsnivå.

5.2 Huvudsakliga resultat

Resultaten av regressionsanalysen när alla variabler inkluderas simultant i enlighet med ekvation (1) visas i Tabell 1, medan resultaten av separata skattningar för respektive variabel anges i Tabell 2. Som utfallsvariabel används i kolumn 1–4 en dummyvariabel som antar värdet ett om minst en friskola etablerades under 1992–2000 mätt för de geografiska rutor på 1km*1km som beskrevs i avsnitt 4. Kolumn 5–8 visar resultatet för skattningar på SAMS-respektive kommunnivå. Förutom skattningarna för de binära utfallsvariablerna, som anges i kolumn 5 och 7, visas också resultaten för den kontinuerliga utfallsvariabeln antal friskoleetableringar i kolumn 6 och 8.¹⁴ Anledningen till att även dessa redovisas är att det i flera kommuner och vissa SAMS-områden etableras många friskolor, och det därmed är lämpligt med en kontinuerlig variabel som fångar upp detta.

De estimat som visas i tabellerna är i form av genomsnittliga elasticiteter. Det vill säga, de visar de genomsnittliga procentuella samvariationerna mellan utfallsvariabeln och de lokala faktorerna. En 1 procents högre andel av elever som har minst en förälder med eftergymnasial utbildning samvarierar till exempel enligt Tabell 1 med mellan 0,7 och 1,8 procents högre sannolikhet för friskoleetablering, beroende på vilken specifikation som avses. Ett än starkare samband, fast negativt, skattas för andelen elever med minst en svenskfödd förälder. En ökning på 1 procent i den variabeln

¹⁴ Tabell 1 och Tabell 2 visar standardfel som klustrats på kommun- eller SAMS-nivå. Den engelska versionen av rapporten redovisar även alternativa standardfel baserade på Conley (1999). Dessa skiljer sig inte på något avgörande vis från de resultat som redovisas här.

samvarierar i genomsnitt med mellan 2 och 4 procents *lägre* sannolikhet för friskoleetablering. Även befolkningstätheten bland barn i skolåldern samvarierar med friskoleetableringarna: en procents ökning i den här variabeln samvarierar i genomsnitt med knappt 0,3 procents högre sannolikhet för friskoleetablering.

Tabellerna tyder vidare på ett negativt samband mellan vänstermajoritet i kommunfullmäktige och sannolikheten för friskoleetablering. För den här variabeln är det mer intuitivt att uttrycka storleken på estimatet i termer av en 0–1-förändring än i termer av de elasticiteter som anges i tabellen. Uttryckt på detta sätt tyder resultaten på att områden som ligger i kommuner med vänstermajoritet i genomsnitt har en 0,6–1 procent lägre sannolikhet för friskoleetablering, jämfört med områden i kommuner utan vänstermajoritet.

De resultat som beskrevs ovan avser de variabler för vilka samstämmiga resultat ges av alla de alternativa geografiska specifikationerna i Tabell 1, i den mån att estimaten i alla (eller så gott som alla) specifikationer är statistiskt signifikant skilda från noll samt har samma tecken.

För vissa av variablerna skiljer sig resultaten dock ganska markant beroende på vilken geografisk indelning som används. Detta gäller bland annat för 9: ans medelbetyg, samt den lokala inkomstspridningen. Särskilt för medelbetyget varierar både storleken på parameterskattningen och dess signifikansnivå kraftigt mellan specifikationerna. För proxy-variabeln för förväntade lokalkostnader sticker vidare de SAMS-baserade skattningarna ut med relativt stora och statistiskt signifikanta positiva estimat, medan de andra specifikationerna ger betydligt mindre och icke statistiskt säkerställda elasticiteter.

Detta kan generellt sett ses som exempel på att den geografiska indelningen kan vara av betydelse, och specifikt som att de här estimaten ska tolkas med stor försiktighet. Eller med andra ord, eftersom det inte är självklart vilken eller vilka av de geografiska indelningarna som är att föredra i det här fallet, så bör skattningar som varierar mellan specifikationerna tolkas med försiktighet (eller inte alls).

Skattningarna i Tabell 1 gällde som tidigare påpekades en analys där alla variabler inkluderades samtidigt i modellen. Detta innebär alltså att de anger samvariationen mellan en viss variabel och sannolikheten för friskoleetablering, *givet* alla andra områdesspecifika variabler.

Resultaten av de alternativa separata regressionerna för en variabel i taget, som alltså *inte* konstanthåller för områdesvisa skillnader i de andra variablerna, visas i Tabell 2. Notera att varje rad i tabellen avser en separat regression.

Resultatet i Tabell 2 visar att de huvudsakliga resultaten står sig även i sådana separata skattningar: högutbildade föräldrar och befolkningstäthet bland barn i skolåldern samvarierar positivt med sannolikheten för friskoleetablering, medan svenskfödda föräldrar och vänstermajoritet i kommunfullmäktige är negativt korrelerade med densamma.

För andra variabler uppvisas emellertid stora skillnader jämfört med Tabell 1. Till exempel skattas nu en statistiskt säkerställd positiv association mellan ett områdes inkomstnivå och sannolikheten att en friskola startar där. Hur ska det förstås, när parameterestimaten för samma variabel i Tabell 1 i regel var statistiskt insignifikanta (och till tecknet negativa)? Visserligen etableras friskolor oftare i områden med en högre nivå på disponibel inkomst, men sannolikt fångar det positiva estimatet i Tabell 2 upp samvariationen med andra variabler som tenderar att vara korrelerade med inkomsten, såsom utbildningsnivå och befolkningstäthet. En annan variabel som skiljer sig starkt mellan tabellerna är 9ans medelbetyg: variabeln uppvisar ett relativt starkt positivt samband med friskoleetablering för de flesta specifikationerna när ingen hänsyn tas i analysen till andra skillnader mellan områdena, men uppvisar ett negativt eller inget samband *allt annat lika* – d.v.s. när de andra lokala faktorerna tas med i modellen.

Tabell 1 Genomsnittliga marginaleffekter i form av elasticiteter för friskoleetableringar 1992–2000. *Specifikation (1)-(4) baseras på 1km*1km-rutor, befolkningstäthet inom 3km, och skoltäthet inom 6 km. I specifikation (5)-(8) är alla mått aggregerade till SAMS- respektive kommunnivå*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Utfallsvariabel:</i>		<i>≥1 friskola, 1km*1km-rutnät</i>			<i>≥1 friskola</i>	<i>Antal friskolor</i>	<i>≥1 friskola</i>	<i>Antal friskolor</i>
<i>Specifikation elevbakgrund:</i>		<i>≤ 1,5km</i>	<i>50 närmsta</i>	<i>100 närmsta</i>	<i>SAMS</i>	<i>SAMS</i>	<i>Kommun</i>	<i>Kommun</i>
Medelbetyg åk 9	≤ 3km -3,184** (1,402)	-0,0217 (0,378)	-0,270 (1,624)	0,241 (2,598)	-2,771** (1,131)	-2,689** (1,146)	-8,303** (3,922)	-11,37*** (3,661)
Högutbildad förälder	1,779*** (0,182)	0,719*** (0,0645)	1,457*** (0,166)	1,531*** (0,203)	1,105*** (0,183)	1,035*** (0,177)	1,007* (0,537)	1,740*** (0,476)
Inkomst	-0,398 (1,036)	-0,230 (0,276)	-1,850* (0,953)	-1,064 (0,954)	-0,255 (0,591)	-0,205 (0,575)	3,586* (1,908)	0,798 (0,889)
Inkomstspridning	0,131** (0,0572)	0,0978** (0,0446)	0,140** (0,0686)	0,0851 (0,0557)	-0,0190 (0,0657)	0,00718 (0,0716)	-0,0771 (0,109)	0,0304 (0,0914)
Svenskfödd förälder	-4,171*** (0,877)	-2,182*** (0,339)	-3,032*** (0,950)	-3,834*** (0,843)	-2,459*** (0,650)	-1,973*** (0,656)	-3,084* (1,733)	-6,232*** (1,137)
Befolkning 7–9-år	0,275*** (0,0657)	0,297*** (0,0664)	0,256*** (0,0558)	0,256*** (0,0596)	0,284*** (0,0853)	0,201*** (0,0746)	0,422* (0,226)	0,298 (0,193)
Proxy skolpeng	-0,873 (1,737)	-0,608 (1,721)	-0,604 (1,793)	-0,640 (1,791)	-3,322** (1,504)	-2,978** (1,431)	0,255 (1,822)	-2,809* (1,573)
Vänstermajoritet	-0,194*** (0,0653)	-0,211*** (0,0610)	-0,227*** (0,0665)	-0,212*** (0,0656)	-0,233*** (0,0703)	-0,254*** (0,0715)	-0,210** (0,0898)	-0,272*** (0,0603)
Proxy lokalkostnader	-0,0661 (0,692)	0,107 (0,712)	0,0397 (0,713)	-0,0328 (0,687)	1,234** (0,546)	1,129** (0,563)	-0,208 (0,689)	0,613 (0,595)
Kommunala skolor	-0,0801 (0,0729)	-0,0269 (0,0664)	-0,0278 (0,0668)	-0,0481 (0,0706)	0,135* (0,0742)	0,125 (0,0765)	0,0315 (0,269)	0,205 (0,163)
Fristående skolor	-0,00733* (0,00445)	-0,00581 (0,00683)	-0,00754 (0,00484)	-0,00552 (0,00467)	-0,000234 (0,00400)	0,000668 (0,00410)	-2,76e-05 (0,0141)	-0,0613*** (0,0216)
Observationer	50 325	37 875	50 292	50 304	3 361	3 361	283	283
Estimator	Logit	Logit	Logit	Logit	Logit	Poisson	Logit	Poisson
Log-likelihood	-1806	-1771	-1822	-1820	-657,9	-789,2	-132,7	-331,1

Standardfel anges inom parentes, och är klustrade på kommunnivå i alla specifikationer utom kolumn 5–6, som istället är klustrade på SAMS-nivå. *** p<0,01, ** p<0,05, * p<0,1.

Tabell 2 Genomsnittliga marginaleffekter i form av elasticiteter för friskoleetableringar 1992–2000. Resultatet från separata skattningar för respektive högerledsvariabler. *Specifikation (1)-(4) baseras på 1km*1km-rutor, befolkningstäthet inom 3km, och skoltäthet inom 6km. I specifikation (5)-(8) är alla mått aggregerade till SAMS-respektive kommunnivå*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Utfallsvariabel:</i>		<i>≥1 friskola, 1km*1km-rutnät</i>			<i>≥1 friskola</i>	<i>Antal friskolor</i>	<i>≥1 friskola</i>	<i>Antal friskolor</i>
<i>Specifikation elevbakgrund:</i>	<i>≤ 3km</i>	<i>≤ 1,5km</i>	<i>50 närmsta</i>	<i>100 närmsta</i>	<i>SAMS</i>	<i>SAMS</i>	<i>Kommun</i>	<i>Kommun</i>
Medelbetyg åk 9	5,379*** (1,623)	1,259*** (0,382)	8,324*** (2,544)	13,81*** (3,484)	0,168 (0,888)	0,248 (0,956)	5,527** (2,525)	13,13*** (2,903)
Högutbildad förälder	2,028*** (0,232)	0,705*** (0,0904)	1,707*** (0,205)	1,945*** (0,223)	0,769*** (0,127)	0,754*** (0,136)	2,245*** (0,346)	2,100*** (0,307)
Inkomst	2,369*** (0,452)	0,108*** (0,0410)	1,336*** (0,278)	2,151*** (0,283)	0,645** (0,253)	0,711*** (0,248)	7,937*** (1,351)	3,262*** (0,676)
Inkomstspridning	0,215*** (0,0424)	0,121*** (0,0205)	0,0916*** (0,0205)	0,128*** (0,0266)	0,0413 (0,0284)	0,0541 (0,0336)	0,283* (0,163)	0,345*** (0,0976)
Svenskfödd förälder	-7,115*** (0,825)	-2,752*** (0,245)	-6,026*** (0,503)	-6,324*** (0,584)	-2,530*** (0,433)	-2,395*** (0,417)	-5,681*** (1,407)	-7,474*** (1,357)
Befolkning 7–9-år	0,298*** (0,0289)	0,298*** (0,0289)	0,298*** (0,0289)	0,298*** (0,0289)	0,492*** (0,0764)	0,386*** (0,0460)	0,623*** (0,0875)	0,248*** (0,0242)
Proxy skolpeng	3,145*** (1,131)	3,510*** (1,156)	3,123*** (1,142)	3,187*** (1,156)	0,646 (0,652)	0,614 (0,693)	-1,320* (0,701)	0,435 (0,894)
Vänstermajoritet	-0,377*** (0,0790)	-0,377*** (0,0790)	-0,377*** (0,0790)	-0,377*** (0,0790)	-0,328*** (0,0683)	-0,340*** (0,0694)	-0,228*** (0,0755)	-0,462*** (0,0745)
Proxy lokalkostnader	2,036*** (0,647)	2,036*** (0,647)	2,036*** (0,647)	2,036*** (0,647)	1,116*** (0,263)	1,097*** (0,282)	0,0631 (0,274)	1,116** (0,527)
Kommunala skolor	0,340*** (0,0195)	0,340*** (0,0195)	0,340*** (0,0195)	0,340*** (0,0195)	0,271*** (0,0547)	0,298*** (0,0507)	0,629*** (0,0911)	0,421*** (0,0239)
Fristående skolor	0,0494*** (0,00218)	0,0494*** (0,00218)	0,0494*** (0,00218)	0,0494*** (0,00218)	0,00592* (0,00349)	0,00636 (0,00399)	0,0367*** (0,00417)	0,0494*** (0,00455)
Observationer	50 388–50 682	42 681–50 682	50 388–50 682	50 682	3 391–3 485	3 391–3 485	284–286	284–286
Estimator	Logit	Logit	Logit	Logit	Logit	Poisson	Logit	Poisson
Log-likelihood	-1915 – -2333	-1915 – -2263	-1915 – -2327	-1915– -2323	-705,9 – -743,1	-849 – -891,4	-152,7 – -192,4	-470,6 – -710,6

Standardfel anges inom parentes, och är klustrade på kommunnivå i alla specifikationer utom kolumn 5–6, som istället är klustrade på SAMS-nivå. *** p<0,01, ** p<0,05, * p<0,1.

5.3 Kompletterande resultat

Utöver de analyser som presenterades i avsnittet ovan, har följande kompletterande analyser genomförts. För resultattabeller för dessa hänvisas till den engelska versionen av den här rapporten, Edmark (2018).

- Regressionsanalysen har genomförts för de respektive alternativa utfallsperioderna 1992–95 och 1992–2005. Dessutom genomfördes en analys för utfallsperioden 1996–2000 då de lokala variablerna uppmättes 1995 istället för 1991. De övergripande resultaten var likartade de som presenterats ovan. Den huvudsakliga skillnaden var att estimatet för vänstermajoritet i kommunfullmäktige inte längre var statistiskt säkerställt skilt från noll i alla specifikationer.
- Ett rutnät om 0,5km*0,5km användes för utfallsvariabeln friskoleetablering istället för rutnätet på 1km*1km. Detta påverkade inte rapportens huvudsakliga resultat.
- Befolkningstätheten bland barn i lågstadieåldern uppmättes inom en 1,5 km radie istället för inom 3 km radie, och antal existerande skolor uppmättes inom 3 km radie istället för inom 6 km radie. I dessa specifikationer var estimaten för den lokala inkomstspridningen oftast inte statistiskt säkerställt skilda från noll, medan estimaten för antal redan existerande kommunala skolor skattades som positiva och statistiskt säkerställt skilda från noll. I övrigt skilde sig inte resultaten nämnvärt från de i Tabell 1.
- Dummyvariabler för lokala arbetsmarknadsregioner, enligt SCB:s klassificering för år 1991, inkluderades i regressionsanalysen. Dessa skattningar resulterade i estimat för den lokala nivån på disponibel inkomst som var negativa och statistiskt signifikant skilda från noll, och estimat för inkomstspridningen som var positiva och statistiskt signifikanta, i de flesta av specifikationerna. Variabeln för vänstermajoritet i kommunfullmäktiga var inte statistiskt signifikant skild från noll, vilket troligen beror på att en stor del av variationen på kommunnivå ”äts upp” av arbetsmarknadsregion-variablerna.
- Dummyvariabler för kommuntyp, baserat på SKL:s kategorisering av kommuner, inkluderades i regressionsanalysen. Även detta resulterade ofta i statistiskt signifikanta estimat för variablerna för nivå och spridning i lokal disponibel inkomst. Samma resultat gavs när regressionsanalysen genomfördes med enbart observationer inom Stockholms län.

6 Avslutande sammanfattning

Sammanfattningsvis tyder rapportens resultat på att fristående lågstadieskolor som startade under perioden 1992–2000 oftare etablerades i områden med högre befolkningstäthet bland barn i skolåldern och där barnens föräldrar hade högre utbildningsnivå. Sannolikheten för friskoleetablering var vidare lägre i områden med en högre andel barn med svenskfödda föräldrar samt i områden i kommuner med vänstermajoritet i kommunfullmäktige. De här resultaten var stabila över olika alternativa geografiska områdesdefinitioner som användes i analysen, samt överensstämmer med vad tidigare rapporter på området funnit (Angelov och Edmark, 2016, samt Holmlund m.fl. 2014). Resultaten för dessa variabler framträdde dessutom både i separata regressioner för respektive variabler, och i analyser där alla variabler inkluderades samtidigt.

För andra variabler gjorde det emellertid stor skillnad vilken geografisk områdesdefinition som användes. Detta gäller bland annat för medelbetyget i 9:an samt för den lokala inkomstspridningen. En möjlig tolkning av detta är att den geografiska indelningen är potentiellt betydelsefull i den här typen av undersökningar. Det kan därför vara lämpligt att använda flera alternativa och flexibla mått, baserade på geografisk information på mikronivå. Detta torde bli enklare och mer vanligt i och med att tillgängligheten till både data och program för den här typen av skattningar ökar.

Referenser

- Amcoff, J. (2012). "Hur bra fungerar SAMS-områdena i studier av grannskapseffekter?", *Socialvetenskaplig tidskrift* nr 2.
- Andersson, R. och S. Musterd (2010). "What Scale Matters? Exploring the Relationship between Individuals' Social Position, Neighborhood Context and the Scale of Neighborhood", *Geografiska Annaler: Series B, Human Geography* 92(1): 23–43.
- Angelov, N. och K. Edmark (2016). "När skolan själv får välja – en ESO-rapport om friskolornas etableringsmönster", Rapport till expertgruppen för studier i offentlig ekonomi (ESO) 2016:3.
- Conley, T., G. (1999). "GMM estimation with cross sectional dependence", *Journal of Econometrics* 92(1):1–45.
- Edmark, K (2018). "Location choices of Swedish independent schools – How does allowing for private provision affect the geography of the education market?" IFAU Report 2018:16.
- Hennerdahl, P. och M. Meinild Nielsen (2017). "A Multiscalar Approach for Identifying Clusters and Segregation Patterns That Avoids the Modifiable Areal Unit Problem", *Annals of the American Association of Geographers* 107(3): 555–574.
- Holmlund, H., Häggblom, J., Lindahl, E., Martinson, S., Sjögren, A., Vikman, U. och B. Öckert (2014). "Decentralisering, skolval och fristående skolor: resultat och likvärdighet i svensk skola." IFAU Report 2014:25.
- Mitra, R. och R. N. Buliung (2012). "Built environment correlates of active school transportation: Neighborhood and the modifiable areal unit problem." *Journal of Transport Geography* 20(12): 51–61.
- Wong, D. (2004). "Comparing Traditional and Spatial Segregation Measures: A Spatial Scale Perspective", *Urban Geography* 25(1): 66–82.
- Östh J., Clark A. W., and B. Malmberg (2015). "Measuring the scale of segregation using k-nearest neighbor aggregates", *Geographical Analysis* 47(1): 34–49.

Appendix

Tabell A1 Beskrivande statistik i form av medelvärden, standardavvikelser (std) och antal observationer, för de regressionsvariabler som används i den här rapporten

Utfallsvariabler:		1km*1km rutor				Kommun	SAMS
Medelvärde:	Friskola (0–1)			0,0079		0,4161	0,0551
	varav Antal 1: or			400		119	192
	varav Antal 0: or			50 308		167	3293
Medelvärde:	Friskolor (antal)					1,3531	0,0666
Std:	Friskolor (antal)					3,4809	0,2995
Variabler för elevbakgrund:		≤ 3km	≤ 1,5km	50 närmsta	100 närmsta		
Medelvärde:	Medelbetyg åk9	3,19	3,19	3,20	3,20	3,19	3,20
Std:	Medelbetyg åk9	0,18	0,41	0,14	0,11	0,09	0,25
Medelvärde:	Högutb föräld	0,34	0,34	0,35	0,35	0,33	0,38
Std:	Högutb föräld	0,13	0,24	0,15	0,13	0,10	0,18
Medelvärde:	Inkomst	260 844	260 071	262 325	261 121	254 534	262 080
Std:	Inkomst	31 643	95 053	38 652	33 047	23 881	51 464
Medelvärde:	Inkomstspridning	105 280	90 660	97 822	104 790	121 401	105 213
Std:	Inkomstspridning	116 305	124 291	120 516	118 883	101 683	140 776
Medelvärde:	Svenskfödd föräld	0,93	0,93	0,93	0,93	0,92	0,90
Std:	Svenskfödd föräld	0,07	0,11	0,07	0,06	0,06	0,10
Medelvärde:	Proxy skolpeng	46 852	46 707	46 848	46 846	48 050	47 634
Std:	Proxy skolpeng	4555	4414	4557	4511	5866	4692
Variabler på kommunnivå:							
Medelvärde:	Vänstermajoritet			0,23		0,25	0,21
Std:	Vänstermajoritet			0,42		0,44	0,41
Medelvärde:	Proxy lokalkostn			11 029		11 347	11 527
Std:	Proxy lokalkostn			2592		2911	2743
Skoltäthet:				≤3km			
Medelvärde:	Befolkning 7–9 år			171		1003	65
Std:	Befolkning 7–9 år			299		1478	47
Befolkningstäthet:				≤6km			
Medelvärde:	Ant komm skolor			5,60		13,76	0,69
Std:	Ant komm skolor			8,86		14,22	0,71
Medelvärde:	Ant friskolor			0,15		0,21	0,01
Std:	Ant friskolor			0,93		1,20	0,11
Antal observationer:		≤3km	≤1,5km	50 närmsta	100 närmsta		
Medelvärde:	Antal 7–9-åringar ^a	170,52	43,20	50,00	100,00	1002,96	64,93
Std:	Antal 7–9-åringar	298,66	95,27	0,00	0,00	1478,16	47,26
Medelvärde:	Antal obs betyg ^b	61,45	15,59	50,00	99,97	360,47	22,33
Std:	Antal obs betyg	96,92	32,44	0,00	1,19	464,89	17,27
Antal:	Rutor/Kommuner/SAMS	50 708	50 708	50 708	50 708	286	3485 ^c

^a Antal observationer avser här antal observationer som användes i beräkningarna av de variabler för elevbakgrund som baseras på individer i lågstadieåldern.

^b Det kan noteras att det genomsnittliga antalet observationer för måttet baserat på de 100 närmsta eleverna är något lägre än 100 när det gäller medelbetygsvariabeln. Detta beror på att jag i beräkningarna av måttet begränsade elevunderlaget till de elever som bodde inom 100 km från respektive ruta, i syfte att underlätta beräkningarna i dataprogrammet. Det kan i något fall innebära färre än 100 elever.

^c Antal SAMS-områden som ingår i regressionsanalysen är lägre än det totala antalet SAMS-områden vilket uppgår till ca 9 000. Det beror på att bara SAMS-områden som innehåller minst 30 individer i 7–9-årsåldern, mått år 1991, ingår i analysen.

Institutet för arbetsmarknads- och utbildningspolitisk utvärdering (IFAU) är ett forskningsinstitut under Arbetsmarknadsdepartementet med placering i Uppsala.

IFAU ska främja, stödja och genom forskning genomföra uppföljningar och utvärderingar. Uppdraget omfattar effekter av arbetsmarknads- och utbildningspolitik, arbetsmarknadens funktionssätt och arbetsmarknadseffekter av socialförsäkringen.

I rapportserien presenteras såväl IFAU:s forskning som resultat av samarbeten med andra nationella och internationella forskningsorganisationer.

IFAU delar årligen ut bidrag till olika forskningsprojekt, vars resultat publiceras i rapportserien.

Rapporterna kan vara fristående eller publiceras tillsammans med ett Working paper.

Alla IFAU:s publikationer finns på www.ifau.se