



Restaurangkluster

-En studie av områdesfaktorer för gynnsam restaurangetablering

Elias Myrén

Abstract:

Location choice can be considered a key competition element in multiple industries. In the restaurant business, location can be crucial for restaurants to stay competitive and viable over time. This study aims to investigate what kind of factors contribute to making an area more suitable for restaurant localization. Initially by looking at demographic composition in different areas in Sweden, investigating which areas restaurants tend to cluster to. Thereafter, the study investigates if this tendency to cluster to specific areas can provide an additional advantage for the individual restaurant through cumulative attraction. These two questions are analysed through regression analysis. The regressions use restaurant density, demographic statistics, and data on restaurant revenue to make estimations regarding what demographic and restaurant composition to look for in an area when opening restaurants or planning new neighbourhoods. The result of the study found significant positive correlation between restaurant density and variables such as median income and population density. Different household compositions were also found to be significantly correlated to restaurant density, although the direction of correlation varied with household size. A high restaurant density was later found to be positively correlated to restaurant revenue. This study therefore suggests that the area's demographic composition should be closely considered when examining the suitability of an area for restaurant localization. Although revenue might not provide the full picture of a restaurant's success, clusters of restaurants should not be immediately dismissed in fear of competition when researching suitable location for similar businesses.

Keywords: restaurant localization, cumulative attraction, restaurant cluster, monopolistic competition, demographic area statistics, restaurant industry.

Kandidatuppsats Nationalekonomi, 15hp

Vårtermin 2020

Handledare: Anders Boman

Institutionen för nationalekonomi med statistik

Handelshögskolan vid Göteborgs universitet

Sammanfattning:

Val av lokalisering kan anses som en viktig del i konkurrensbeteendet inom många industrier. I restaurangbranschen kan läget vara avgörande för att en verksamhet ska vara konkurrens- och livskraftig. Denna studie ämnar till att undersöka vilka faktorer som bidrar till att skapa områden som anses lämpliga för restaurangverksamhet. Inledningsvis genom att studera den demografiska sammansättningen i olika områden i Sverige för att undersöka vilken typ av områden restauranger tenderar att placera sig i. Därefter undersöker studien om denna klustertendens kan vara en framgångsfaktor i sig själv för den enskilda restaurangen genom *cumulative attraction*. Dessa två frågor analyseras genom regressionsanalys. Regressionerna använder demografiska områdesdata, restaurangtäthet samt omsättningsdata för respektive restaurang för att uppskatta vilken demografisk profil samt existerande restaurangutbud som kan anses attraktivt när nya restauranger väljer läge eller när nya områden ska detaljplaneras. Resultatet visade på signifikant positivt samband mellan restaurangtäthet och variabler så som medianinkomst och invånartäthet. Hushållssammansättning sågs också ha signifikant samband med restaurangtätheten i ett område, dock kan riktningen för sambanden variera beroende på hushållsstorlek. Hög restaurangtäthet sågs senare också positivt korrelera med restaurangers omsättning. Denna studie pekar därmed på att ett områdes demografiska sammansättning bör beaktas när olika områden utvärderas inför restaurangetablering. Trots att omsättningen inte kan ge hela bilden av en verksamhets framgång menar denna studie också på att områden med täta restaurangkluster inte bör avfärdas av rädsla för hög konkurrens, denna typ av områden kan likväl vara lämpliga när olika lägesalternativ utvärderas för restaurangetablering.

Nyckelord: restauranglokalisering, cumulative attraction, restaurangkluster, monopolistisk konkurrens, demografiska statistikområden, restaurangmarknaden.

Innehåll

| | | |
|----------|--|-----------|
| 1 | Inledning | 1 |
| 1.1 | Bakgrund | 1 |
| 1.2 | Problemformulering..... | 1 |
| 1.3 | Syfte & frågeställningar..... | 2 |
| 2 | Teoretisk referensram & tidigare forskning | 3 |
| 2.1 | Konkurrensteori & vinstmaximering..... | 3 |
| 2.2 | Tidigare studier..... | 5 |
| 3 | Metod & Data | 8 |
| 3.1 | Data..... | 8 |
| 3.2 | Metod..... | 10 |
| 3.3 | Avväganden för statistisk analys | 14 |
| 3.4 | Etiska aspekter & god forskningssed..... | 14 |
| 4 | Resultat..... | 16 |
| 4.1 | Områdesvariabler för restaurangverksamhet..... | 16 |
| 4.2 | Värdet av kluster..... | 19 |
| 5 | Diskussion | 21 |
| 5.1 | Regression 1 & 2 | 21 |
| 5.2 | Regression 3 & 4 | 22 |
| 5.3 | Övriga iakttagelser..... | 23 |
| 6 | Studiens begränsningar & framtida forskning | 24 |
| 6.1 | Bolagsformer för klusteranalys | 24 |
| 6.2 | Val av beroendevariabel | 24 |
| 6.3 | Demografiska statistikområden | 24 |
| 6.4 | Förslag till vidare forskning | 25 |
| 7 | Slutsats..... | 26 |
| 8 | Källförteckning..... | 27 |
| 9 | APPENDIX A. | 29 |

1 Inledning

I detta kapitel presenteras bakgrunden samt studiens problemområde. Därefter presenteras studiens syfte och frågeställningar.

1.1 Bakgrund

Den enskilda restaurangens framgång eller misslyckande har studerats i en mängd olika former och i en mängd olika städer. Enligt en seglivad myt klarar 90% av alla nystartade restauranger inte ens ett år utan går i konkurs eller byter ägare. Detta har dock visats inte riktigt stämma överens med verkligheten. Studier på den amerikanska restaurangmarknaden menar istället att ungefär 26% av alla nyöppnade restauranger inte klarar sig igenom det första året. Inom en treårsperiod kan dock så mycket som 59,74% av alla nyöppnade restauranger ha bytt ägare. (Parsa, Self, Njite, & King, 2005). Även om närmare 60% ägarbyte inom loppet av tre år kan tyckas högt är det långt ifrån myten om att majoriteten inte klarar sig igenom det första året.

Under kommande arbete kommer termen *restaurang* innefatta all verksamhet som i dagligt tal brukar benämnas som restaurang, bar eller kafé. Företag så som cateringverksamhet eller centralkök inkluderas inte i studien och ingår därför ej i denna term. Vad som definierar om en restaurang är lyckad eller ej är inte heller helt självklart. Rimligtvis kan olika lönsamhetsmått användas för att utvärdera detta. På grund av bristande tillgång till alla tillämpliga lönsamhetsmått fokuserar denna studie på restaurangers omsättning, konsekvenser av detta diskuteras i avsnitt 6.2 i *Studiens begränsningar* samt i avsnitt 5.2 i *Diskussion*.

Så vad som faktiskt krävs för att lyckas är en fråga som också studerats och diskuteras tidigare. Diverse demografiska områdesfaktorer, restaurangägarens förmåga och karaktärsdrag, förmåga att differentiera sig från konkurrensen har nämnts som möjliga framgångsfaktorer. Restaurangens läge anses av många också vara en av de avgörande faktorerna för verksamhetens vara eller icke-vara (Self, Feree Jones, & Botieff, 2015). Vad som menas med ett bra läge kan i sin tur troligtvis definieras av en mängd olika variabler, demografiska förutsättningar kan vara en av dem.

Vilka variabler som avgör den enskilda restaurangens val av placering kan också variera beroende på vilken typ av restaurangverksamhet som bedrivs. Faktum är dock att restauranger har en tendens att etableras i närhet av varandra och skapa kluster av snarlika verksamheter. Möjligtvis för att tillsammans kunna generera fler besökare till området, eller på grund av annan uppenbar fördel det valda området har som gör det särskilt attraktivt för den typen av verksamhet (Parsa et al., 2005).

Samtidigt som klustereffekten kan anses öka attraktionskraften för området bör det också innebära en ökad konkurrens restaurangerna emellan om många liknande verksamheter verkar inom ett litet geografiskt område. Detta går till viss del emot klassiskt ekonomisk teori om hur företag bör bete sig för att maximera vinst. Just klustereffekten har dock studerats och påvisats ha en positiv inverkan på restaurangverksamhetens omsättning även i Sverige. Under 2016 genomfördes en studie över restauranger i Göteborg där resultatet pekade på att det fanns en klar fördel att ha nära till verksamheter inom samma näringsgren (Evidens, 2016).

1.2 Problemformulering

Vad som egentligen krävs för att lyckas med en restaurangetablering beror troligtvis på en mängd olika faktorer och varierar rimligtvis med typ av restaurang. Som tidigare nämnts finns också en viss motsättning mellan de empiriska studier som gjorts och den klassiska ekonomiska teorin gällande konkurrens. Detta borde inte vara något ovanligt i sig i och med

att ekonomiska teorier endast är modeller och kan inte rimligtvis förväntas passa samtliga marknader i alla länder. Det gör dock att det blir intressant att undersöka först varför kluster skapas. Därefter undersöka om det finns en tydlig effekt av att ligga inuti ett restaurangkluster för den enskilda verksamheten. Kan det påvisas att kluster i sig själv är en framgångsfaktor och har en positiv inverkan på den enskilda restaurangens omsättning? Eller är klustertendensen som syns i restaurangbranschen endast en liten del i restaurangers val av lokalisering på marknaden. Genom att utöka undersökt område från Göteborg i tidigare studie (Evidens 2016), till att undersöka Sverige i sin helhet kan studien förhoppningsvis ytterligare öka förståelsen för de områdesfaktorer som på ett brett plan påverkar antalet restauranger i området.

1.3 Syfte & frågeställningar

Syftet med studien är att inom Sverige undersöka om det går att identifiera klara områdesfaktorer som generellt bidrar till en hög andel restauranger inom området. Därefter undersöka om en ansamling av restauranger har någon ekonomisk nytta för den enskilda restaurangen.

- Vilka demografiska områdesfaktorer bidrar till att en hög andel restaurangverksamheter etableras i ett område?
- Hur stor effekt har kluster av liknande verksamheter (kaféer, barer eller restauranger) på en enskild restaurangs omsättning, har klustret ett värde i sig?

2 Teoretisk referensram & tidigare forskning

Följande avsnitt går igenom delar av grundläggande ekonomisk teori som kan vara aktuella oavsett bransch- eller företagstyp. Därefter presenteras delar av den tidigare forskning som genomförts på restaurangmarkanden med fokus på lokaliseringsstudier.

2.1 Konkurrensteori & vinstmaximering

2.1.1 Konkurrens i restaurangbranschen

För att beskriva hur ett företag teoretiskt sett hittar sin vinstmaximerande produktionsnivå under olika typer av konkurrens delar klassiska mikroekonomiska teorier inledningsvis in olika varu- eller tjänstemarknader i någon av följande marknadsformer: perfekt konkurrens, monopol, monopolistisk konkurrens och oligopol (Greenlaw & Shapiro, 2017).

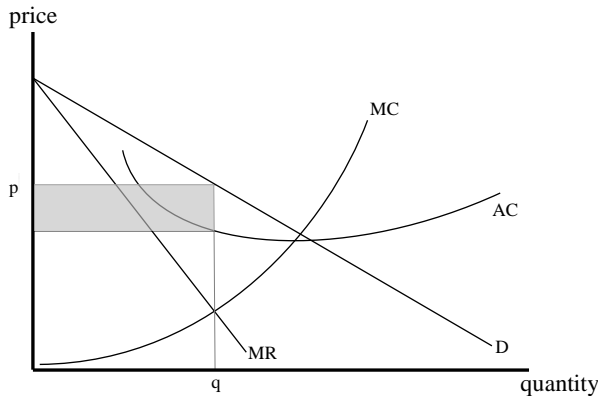
Restaurangbranschen är en av de branscher som emellanåt används som exempel för att beskriva en marknad där *monopolistisk konkurrens* råder. Denna marknadsform kännetecknas i likhet med perfekt konkurrens av låga inträdesbarriärer och många mindre till mellanstora aktörer där ingen aktör har en dominerande position på marknaden. Till skillnad från en perfekt konkurrensmarknad är dock inte varorna eller tjänsterna som företagen på marknaden bjuder ut perfekta substitut, det som produceras av varje företag är på något sätt annorlunda jämfört sina konkurrenter. Ur detta hänseende liknar det mer en monopolmarknad då varje företag kan sägas ha monopol på sin specifika vara eller tjänst. Med det faktum att varor eller tjänster inte är perfekta substitut kommer också en viss möjlighet för företagen att själva sätta sina priser, vilket i sin tur ger möjligheter till större ekonomisk vinst. På en marknad med låga inträdesbarriärer kommer sådan möjlighet dock innebära att nya producenter tar sig in på marknaden för att ta del av överskottet. Eventuellt genom att erbjuda närliggande substitut till redan existerande varor eller tjänster. Så det som gör det till en konkurrensmarknad istället för monopolmarknad är att den består av flera differentierade företag som söker ungefär samma kundgrupp (Greenlaw & Shapiro, 2017).

2.1.2 Vinstmaximering på monopolistisk konkurrensmarknad

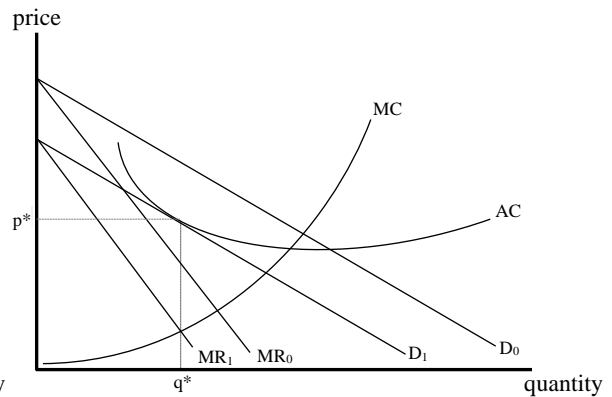
Möjlighet att i viss mån sätta självständiga priser innebär för företagen på en monopolistisk konkurrensmarknad att det på kort sikt finns möjligheter att göra större ekonomisk vinst än vad som kan anses vara normalläge. På lång sikt innebär dock det enkla inträdet till marknaden att andra kommer se möjligheten till ekonomisk vinst och etablera sig på samma marknad. Med ytterligare konkurrens kommer efterfrågekurvan för det enskilda företaget skifta till vänster och på lång sikt försvinner möjligheten till större ekonomisk vinst. Det vill säga en ökad konkurrens minskar vinsten.

Figur 1 visar hur ett företags utbudna kvantitet på en monopolistisk konkurrensmarknad rör sig från ett kortsiktigt läge i grafen till vänster där det gråa området utgör den extravinst som anses överskrida normalvinst och motiverar nya inträden på marknaden. Till det långsiktiga läget i grafen till höger där möjligheten till vinst medfört att nya aktörer tar sig in på marknaden vilket förskjuter det enskilda företaget efterfråge- samt marginalintäktskurva åt vänster och vinstmöjligheten som fanns kortsiktigt har tagits upp av konkurrenter.

Short-run Equilibrium



Long-run Equilibrium



Figur 1: Omarbetade grafer från: Rittenberg, L. Tregarthen, Timothy D. Principles of Economics, (CC BY-NC-SA 3.0)

| | |
|---|-------------------------------------|
| MR = marginalintäkt | MC = marginalkostnad |
| D = efterfråga | AC = genomsnittskostnad (per enhet) |
| p^* = långsiktigt jämvikts/vinstmaxpris | p = kortsiktigt vinstmax-pris |
| q^* = långsiktig jämvikts/vinstmaxkvantitet | q = kortsiktig vinstmax-kvantitet |

2.1.3 Preferens för variation

En monopolistisk konkurrensmarknad betyder i teorin också att respektive företag inte producerar vid sin lägsta kostnad per enhet, och kommer inte frivilligt öka sin produktion trots att det kan sänka produktionskostnaden per enhet om det genomförs. Detta av den enkla anledningen att det skulle innebära att marginalintäkten skulle hamna under marginalkostnaden. Detta ses jämfört perfekt konkurrens som ineffektivt resursutnyttjande i ekonomisk teori. Perfekt konkurrens är dock endast en modell och inte nödvändigtvis något att sträva efter. Differentieringen som uppkommer på den monopolistiska konkurrensmarknaden kan ha ett värde i sig. Konsumenter får fler alternativ vilket innebär större möjlighet för den enskilda konsumenten att hitta en vara eller tjänst som bäst matchar de egna preferenserna (Rittenberg & Tregarthen, u.å).

2.1.4 Negativ skalavkastning

För restaurangbranschen betyder detta att en större produktion eller verksamhet inte nödvändigtvis resulterar i ett bättre resultat för den enskilda restaurangen. Det varierar dock inom branschen och beror tills viss del på restaurangtyp. Snabbmatsrestauranger kan relativt länge utöka sin verksamhet och fortfarande anses effektiva medan majoriteten av fine-dining-restaurangerna är relativt små operationer. Detta då det vid ett besök till en finare restaurang förväntas större individuell uppmärksamhet till kunden. I många fall finns också en inneboende förväntan hos konsumenten att menyn ska vara nyskapande och variera över årstider eller efter råvarutillgång. I en större organisation kan denna förmåga till konstant förändring och innovation bli lidande varför det finns få riktigt stora gourmetrestauranger (Gwartney, Stroup, Sobel & Macpherson, 2009). Rimligtvis kan detta vara en bidragande faktor till att restaurangbranschen kännetecknas av flertalet mindre företag såväl som större snabbmatskedjor.

2.2 Tidigare studier

2.2.1 Områdesvariabler för restaurangetablering

Vad det faktiskt betyder att ha nära till sin marknad är inte helt självklart. I en tid med ökad mobilitet behöver inte det omedelbara närområdet vara avgörande för att en restaurang ska hitta tillräcklig kundbas.

I en studie av Taipei ansågs olika typer av transportmöjligheter väga tungt och spela en avgörande roll för olika restaurangverksamheters möjligheter till god omsättning. Med transport menas i detta fall dels möjlighet till olika sorters kollektivtrafik, dels andra faktorer så som antalet parkeringsplatser. Transportmöjligheter gäller inte enbart kunderna utan även för restaurangens och minimering av dess framtida transportkostnader. Författarna framhåller dock vikten av att vilka faktorer som väger tyngst kan variera beroende på land eller till och med stad (Tzeng, Teng, Chen, & Opricovic, 2002). Detta påpekande från författarna om stor variation i framgångsfaktorer motiverar dessutom studerat område i detta arbete ytterligare.

Tidigare studier pekar dock åt något olika håll. Somliga menar att demografiska faktorer så som befolkningstäthet, invånares ålder, medelinkomst, invånares utbildningsnivå, boendeform (hyresrätt kontra någon form av ägande) har en positiv korrelation med antalet restauranger som etableras i ett område (Yang, Roehl, & Huang, 2017). Detta är något som företag, inte minst franchiserestauranger, också lägger vikt vid. I en arbetsprocess från en fastighetskonsult för att utvärdera mest lämplig placering av nästa franchiserestaurang, är första steget att identifiera områden som matchar de demografiska profilerna som har visats ge bra resultat för restauranger i andra områden (Fisher, 1997).

Andra studier menar att det inte nödvändigtvis är socioekonomiska faktorer i det specifika området en restaurang ligger i som är avgörande. En studie från Atlanta pekar ut det faktum att vissa restauranger tenderar till att placera sig vid genomfartsleder där många människor passerar. Ett sådant läge gör att kunder kan komma från vitt skilda områden och varje restaurang har ett stort potentiellt marknadsområde. Detta kan då vara en större förklaringsfaktor än den demografiska situationen i närområdet (Pilsbury, 1987). Samtidigt hittar samma studie att en viss typ av restaurang ofta dyker upp i samma typ av område, som exempel uppges kinesiska restauranger vara överrepresenterade i konservativa arbetarklassområden jämfört övriga stadsdelar.

Denna bild av variation beroende på restaurangtyp bekräftas av andra studier. Inkomst ses i andra studier generellt ha en positiv korrelation med antalet restauranger eller serveringar. Det kan dock se annorlunda ut, i en studie från 2002 fann man tre gånger fler alkoholservingar i de fattigaste områdena jämfört de rikaste (Morland, Wing, Diez Roux & Poole, 2002).

2.2.2 Klustertendensen på restaurangmarknaden

Flertalet studier som undersökt hur företag placerar sig på en marknad refererar också till något som kallas *Principle of minimum differentiation*. Denna princip tar avstamp i studier publicerade 1929 av Harold Hotelling och behandlar hur företag teoretiskt sett väljer att differentiera sig gentemot sina konkurrenter. Bland annat genom val av geografisk placering.

Hotelling (1929) pekade först på det faktum att företagens marknad av möjliga köpare sällan är en stationär punkt utan är utspridda över en yta eller i sin enklaste form längs en linje (bygata). Om mer än ett företag existerar inom samma bransch kommer de dock inte placera sig långt ifrån varandra och försöka hitta egna geografiska segment av marknaden. De kommer istället vilja maximera sin kundkrets också genom att överta konkurrenters kunder. Detta leder i sin tur till en tendens till ansamling av företag till en plats där åtkomsten till marknaden anses optimal för bägge dessa ändamål, vilket leder till minimal geografisk differentiering. Principen behandlar också företags tendenser att inte heller i varu- eller tjänsteutbud vilja differentiera sig för mycket från sina konkurrenter, också för att öka marknadsandelar genom övertagande av kunder från konkurrenter.

Den beskrivna ansamlingen av liknande företag anses dock ha ett värde i sig och har gett upphov till ytterligare en flitigt använd princip, nämligen *Principle of Cumulative Attraction*. Principen säger att ansamlingen av företag till en plats kommer öka den generella attraktiviteten för den platsen, vilket gör att kunderna får större benägenhet att söka sig dit. Genom att ligga nära sina konkurrenter och skapa kluster kommer företag därför ha större omsättning än om företagen varit spridda över geografien (Litz & Rajaguru, 2008; Nelson, 1958).

I en empirisk undersökning av faktorer som bidrar till en lyckad restaurangetablering undersöktes också just tendensen till kluster. Studien tog fasta på och undersökte en mängd olika stadskvaliteter och hur de skapade värde för olika intressenter i en stad. En av dessa intressenter var restaurangverksamheter. Överensstämmande med tidigare stycke visade undersökningen av Evidens (2016) att restaurangverksamheter i stor grad drog nytta av att ha liknande verksamheter inom närområdet. Studien visade att omsättningen för restauranger, kaféer eller barer var större om det fanns liknande verksamheter inom en 100 meters radie. Mer samlad konkurrens verkade här skapa bättre förutsättningar till ekonomisk framgång (Evidens, 2016). Vilket är i linje med det som presenterats av Nelson redan 1958.

Detta har även diskuterats i något nyare studier som menar att om företag är tillräckligt differentierade jämfört övriga konkurrenter kommer det också bidra till minimal geografisk spridning då samtliga företag kommer placera sig där de har minst avstånd till marknaden. Konkurrens genom prissättning är annars en stark centrifugalkraft (Fujita & Thisse, 1996). Då restaurangbranschen som anses vara en marknad där varje företag i viss mån har möjlighet att fritt sätta sina priser, i och med att de befinner sig på en monopolistisk konkurrensmarknad, är denna centrifugalkraft möjligtvis svagare inom restaurangbranschen.

Slår man samman Fujita och Thisse (1996), Nelson (1958) och Hotelling (1928) finns det därför en bild av att företag till stor del drar nytta av att ligga nära varandra förutsatt att de har något differentierade produkter. Differentieringen kommer dock endast vara så stor att ett alltför intensivt priskrig undviks, för stor skillnad till övriga företags utbud skulle medföra en risk att tappa marknadsandelar till konkurrenter.

För såväl studien som undersökte den positiva effekten av restaurangkluster (Evidens, 2016) som undersökningar av demografiska områdesvariabler (Yang et al., 2017; Pilsbury, 1987) är det viktigt att komma ihåg att det vid nästan alla statistiska analyser är svårt att säkert säga något om kausalitet. Det vill säga vilken av variablerna som faktiskt påverkar den andra (Esaiasson, Gilljam, Oscarsson, Wängnerud & Towns, 2017). Man kan därför tänka sig att det likväl kan vara ett överflöd av restauranger eller andra restaurangverksamheter som bidrar till att attrahera en viss typ av människor eller kundgrupp till ett område, istället för att områdesinvånarna är de som attraherar restaurangverksamheterna.

3 Metod & Data

I följande avsnitt presenteras den metod som använts för att undersöka studiens frågeställningar. Denna del inleds med presentation av använd data. Detta då viss insikt i, och kunskap om begrepp specifika för den data som används kan underlätta förståelsen av metodpresentationen.

3.1 Data

3.1.1 Datakällor

Data som används för att genomföra regressionsanalyser och besvara frågeställningen kommer främst från två olika källor: *Statistiska centralbyrån* (SCB) samt *Retriever business*.

3.1.2 Statistiska centralbyrån

Från SCB inhämtas demografisk- och socioekonomiska data som senare kommer utgöra några av de oberoende variablerna. Denna data kommer på en geografisk indelningsnivå som kallas *Demografiska statistikområden* (DeSo). Denna typ av områden är en geografisk indelning av Sverige i 5985 mindre områden och för respektive område finns statistik över bland annat:

- Invånarantal i området och antal av respektive kön
- Medelinkomst, samt andel inv. med inkomst över respektive under riksmedian
- Hushållssammansättning, så som antal samboende par eller ensamhushåll
- Antal registrerade fordon
- Antal invånare efter utbildningsnivå: grundskola, gymnasial, eftergymnasial minst tre år och så vidare

Utöver dessa variabler finns ytterligare statistik av demografisk karaktär. SCB:s inhämtningsdatum för DeSo-data varierar mellan 2017 och 2019 (SCB, 2020).

All tillgänglig statistik kommer dock inte användas i regressionerna som oberoende variabler. De variabler som väljs ut att ingå i den inledande regressionen baseras på de demografiska och socioekonomiska variabler som användes i Yang et al, studie från 2017 av antal restauranger i amerikanska postnummerområden. Dock stämmer den data som finns att tillgå för den här studien i Sverige inte exakt överens med den censusdata som använts i studien i USA. Den tidigare undersökningen använde också en annan regressionsmodell vilket gör att somliga variabler var anpassade till den modellen. På grund av dess två omständigheter sker ett visst bortfall av oberoende variabler jämfört studien från 2017. Se ”metod” för regressionspecification.

Viktigt att poängtera är att majoriteten av de oberoende variablerna anges som andelar av totalen inom respektive område. Exempelvis är det inte antalet män som finns i ett område som används, utan hur stor andel av befolkningen i området som är män, uttryckt som ett procenttal mellan 0 och 100.

3.1.3 Retriever business - Företagsdata

Från *Retriever Business* (u.å) exporterades besöksadress till företag inom restaurang samt barverksamhet, i Retrievers klassifikation av restaurangverksamhet inkluderas även kaféer. Verksamhet som exkluderas är sådana som cateringverksamhet, storkök, hotellverksamhet. Inkluderade bolagsformer är: Handelsbolag, kommanditbolag, enskild firma och aktiebolag.

För aktiebolag hämtas också data för verksamhetens omsättning. Data inhämtas för hela Sverige. Totalt hämtas 24 627 datapunkter där respektive punkt består av en restaurang. Initialt bortfall skedde på grund av att besöksadress inte fanns angiven eller att geokodningsverktyget inte kunde lokalisera adressen, slutgiltigt användes 20 768 restaurangpunkter.

Först och främst användes den registrerade besöksadressen för respektive verksamhet för att placera ut varje enskild restaurang, bar eller kafé på en kartbild. Detta för att kunna mäta restaurangtätheten inom respektive DeSo-område. Detta beräknas genom antalet restauranger inom området delat med områdets area i kvadratkilometer och uttrycks som *Restauranger/km²*. Geografisk utplacering är också grundläggande för att kunna mäta hur tätt klustret av andra liknande verksamheter är runt varje individuell verksamhet. Det vill säga hur många andra restauranger som återfinns närområdet, här definierat som inom 100 meter.

För att genomföra regression angående eventuell positiv effekt av att befinna sig i ett kluster behövs ett lämpligt lönsamhetsmått. För denna studie användes företagets omsättning som mått på verksamhetens framgång. Det är möjligt att andra lönsamhetsmått kan anses mer lämpliga, sådana mått var dock vid upprepade tillfällen svåra att nå via den använda databasen. En möjlig kandidat till beroendevariabeln hade också kunnat vara omsättning/anställd. Då en genomgång av företagsdata visade på att många företag kan ha hög omsättning men noll registrerade anställda ansågs det ha för hög risk att bli missvisande.

Tyvärr fanns företags årsredovisningar inte heller tillgängliga på samma sätt för samtliga bolagsformer varför denna data endast hämtades för aktiebolag. I beräkandet av klustertätheten användes dock företag av alla bolagsformer även om det endast är aktiebolagen som sedan utgjorde de slutgiltiga observationerna i regressionen. Efter bortfall innebär detta att 9733 aktiebolag inkluderades i denna regression.

Sammanfattningsvis betyder detta att vad som utgjorde en observation såg olika ut beroende på undersökt frågeställning:

Regression 1 och 2: Varje DeSo-område utgjorde en observation, för varje område finns statistik över tidigare nämnda demografiska- och socioekonomiska sammansättning, områdesarea samt antal serveringsverksamheter. De två sistnämnda användes för att beräkna restaurangtätheten.

Regression 3 och 4: Varje aktiebolag utgjorde en observation. För varje bolag finns statistik över demografiska- och socioekonomiska sammansättningen i det direkta närområdet, baserat på vilket DeSo-område verksamheten ligger i. För varje observation finns också statistik över hur många liknande verksamheter som finns inom 100 meter. Därtill information om omsättningen för respektive verksamhet.

3.1.4 Databehandling / Bearbetning

Nedan beskrivs i korthet de steg som sker i programvaran ArcMap för att beräkna restaurangtäthet såväl som klustertäthet. Restaurangtätheten uttrycks som *Restauranger/km²* och beräknas för respektive DeSo-område. Klustertäthet anges ett heltal genom antal nådda restauranger inom 100 meter, detta beräknas för respektive restaurang.

Utifrån den besöksadress som angetts i data hämtad från Retriever placerades alla restauranger ut på en kartbild. Utplaceringen skedde automatiskt genom ett verktyg för geokodning från programvarupaketet ArcGIS.

För regressionerna till den första frågeställningen behöver restaurangtätheten beräknas. Detta genom att när samtliga restauranger är geokodade läggs de över en kartbild innehållande tidigare beskrivna DeSo-områden, statistiken för DeSo-områdena samt information om respektive restaurang återfinns i tillhörande tabell. Därefter kombineras de två kartlagren (restaurangverksamhet + DeSo-områden) baserat på geografisk position och ett nytt kartlager skapas. Tabellen till detta nya kartlager innehåller utöver all demografisk statistik för respektive område nu även en kolumn över antalet restaurangverksamheter som återfinns inom varje område. I och med denna sammanslagning kan man nu räkna ut den ungefärliga tätheten av restaurangverksamheter för respektive DeSo-område.

För regressionerna till den andra frågeställningen behöver klustertätheten beräknas. Detta genom att skapa ett buffertområde runt respektive restaurangpunkt med en radie av 100 meter. Därefter används ArcMaps funktion för att automatiskt räkna antalet restauranger inom respektive buffertområde, framräknat antal tilldelas utgångspunkten (resp. restaurangverksamhet).

3.2 Metod

3.2.1 Översiktlig beskrivning och motivering av studiens metod

Studiens frågeställningar besvarades genom multipel regressionsanalys av *Ordinary-Least-Squares* (OLS) typ. Kort kan sägas att denna typ av regression är en linjär modell i sitt grundutförande. När regressionslinjen eller framräknad modell tas fram genom OLS läggs vikten på att minimera summan av modellens *Residualer*. Med residual, också kallar felterm, menas avståndet mellan det av regressionen uppskattade värdet och det i verkligheten observerade värdet. Genom att fokus läggs på minimering av residualerna kan dessa senare användas vid utvärdering av hur regressionen presterar. Detta kommer bli viktigt senare när olika modellspecifikationer ska jämföras. Begrepp som också kommer användas flitigt för att beskriva residualerna är *homoskedasticitet* och *heteroskedasticitet*. Med homoskedasticitet menas att residualerna har ungefär samma varians för alla uppskattade värden. Motsatsen till detta är heteroskedasticitet och betyder då att variansen för feltermen varierar eller är olika stor längs regressionslinjen. (Hair, Black, Babin, & Anderson, 2014). Accepterad signifikansnivå för studien var $p < 0,05$. Regressioner utfördes i två steg. Första steget var att undersöka vad som drar restaurangverksamheter till vissa platser eller områden. Därefter ytterligare regressioner för att undersöka värdet av denna klustertendens. De statistiska analyserna utfördes uteslutande i programvaran *STATA*.

De inledande regressionerna var av mer undersökande karaktär där fokus var att i så stor utsträckning som möjligt, med den data som fanns tillgänglig, besvara frågeställningen om vilka områdesfaktorer som leder att serveringsverksamheter etableras på en viss plats. Med andra ord: vilka variabler bidrar till att kluster av restauranger skapas. I dessa regressioner utgjordes den beroende variabeln av restaurangtätheten respektive log-transformering av restaurangtätheten i respektive DeSo-område. De oberoende variablerna utgjordes av den data som beskriver den demografiska och till viss del socioekonomiska sammansättningen i området. Det vill säga de oberoende variablerna var i majoriteten av fallen uttryckt som

andelar av områdesbefolkningen. Genom demografiska data uttryckt som andelar istället för antal togs hänsyn till att DeSo-områden är av olika storlek och innehåller olika antal invånare.

Metoden för genomförd studie och framförallt den inledande regressionen följer i stora drag den metod som används av Yang et al., (2017). Detta då den beroende variabeln var densamma: restaurangtäthet. De förklarande demografiska och socioekonomiska variablerna som användes följer också den tidigare studien så gott datatillgången tillät. Studien genomförd 2017 (Yang et al.,) använde sig av data indelad efter postnummerområden i USA, storleken för sådana områden kan variera mycket men sett till antal invånare innehåller dessa mellan 1500 och 8000 personer (United States Census Bureau, 2020). För denna studie användes också en geografisk indelning dock med något mindre omfattande områden, 654 till 4243 personer per område (SCB, 2020). Vilket medförde att studien kunde genomföras på en mer detaljerad nivå och med det förhoppningsvis bättre svara på hur mycket faktorer i det omedelbara närområdet spelar in på restaurangtätheten.

Därefter genomfördes två regressioner med syfte att besvara den andra frågeställningen: huruvida tendensen att placera sig nära andra liknande verksamheter, här kallat kluster, i sig själv har en positiv effekt på omsättningen. Detta genom att den beroende variabeln i dessa regressioner var omsättning. De förklarande variablerna var samma som i tidigare regressioner plus variabeln *RestaurangerInom100m* vilken mäter antalet liknande verksamheter inom 100 meter från respektive restaurang. Att mäta klustertäthet med detta avstånd baseras på den tidigare studien av Göteborgs kommun (Evidens, 2016), där man med detta avstånd fann signifikant positivt samband mellan restaurangverksamheters omsättning och högt klusterindex. Vilket också är anledningen till användandet av omsättning som beroende variabel.

3.2.2 Regression av restaurangtäthet

Innan regressionerna inleddes genomfördes vissa statistiska undersökningar av datamaterialet för att bestämma lämplig regressionsspecifikation. Värt att poängtera återigen är att för den inledande regressionen utgjorde varje DeSo-område en observation. Efter bortfall på grund av brister i data användes 5983 DeSo områden i de första regressionerna.

Ett histogram av den inledningsvis tilltänkta beroende variabeln, *Restauranger/km²*, för den inledande regressionen visade på en fördelning med skevhet åt höger där ett fåtal observationer särskilt stod ut. För att undvika att regressionen påverkas av oupptäckta direkta mätfel kontrolleras för detta enligt process hämtad från *UCLA Stata support* ("Regression diagnostics", 2020). Detta krävde dock en första regression och därmed bestämmande av de oberoende variablerna.

På grund av skillnader i datatillgång jämfört studien av Yang et al., (2017) fanns flera möjliga kombinationer av förklaringsvariabler att inkludera i de slutgiltiga regressionerna. För att undvika problem med multikollinearitet kontrollerades för parvis korrelation innan de oberoende variablerna bestämdes. Korrelationer över 0,7 alternativt under -0,7 medförde att en av de variabler som uppvisade sådan korrelation togs bort, dessa gränsvärden baseras på resonemang från Hair et al., (2014). Därefter repeterades kontrollen på nytt tills korrelationerna mellan samtliga oberoende variabler antog värden i spannet -0,7 till 0,7 (Se APPENDIX: A, Tabell 5). Detta förfarande ledde fram till att följande demografiska och socioekonomiska variabler utgjorde de slutgiltiga oberoende variablerna i regressioner 1 och 2 (beteckning i regressionstabeller inom parentes):

- *Andel av områdets invånare som var män (%Män)*
- *Andel av områdets invånare som bor i hyresrätt (%Hyrratt)*
- *Andel av områdets invånare som bor i bostadsrätt (%Borätt)*
- *Årlig medianinkomst i området, angett i tusentals kronor (Median_Ink)*
- *Andel invånare i ålder 25 – 64 år med högskoleutbildning alternativt annan eftergymnasial utbildning om 3 år eller längre (%HögskoleUtb)*
- *Andel hushåll i området som bestod av samboende par med barn (%SamBarn)*
- *Andel hushåll i området som bestod av samboende utan barn (%SamUbarn)*
- *Andel hushåll i området som bestod av ensamstående vuxen med barn (%EnsBarn)*
- *Befolkningstäthet (Invkm2)*

Efter bestämmande av oberoende variabler fortsatte arbetsgången för att undersöka extremvärden. I denna process genomfördes en första regression med ovan nämnda förklaringsvariabler. Därefter undersöktes de tio observationer, med andra ord DeSo-områden, utpekade av arbetsgången som extremvärden, detta baserat främst på observationernas *Residual* samt *Leverage*-värden. *Residual* är som tidigare nämnt ett mått för att beskriva avståndet mellan det av regressionsekvationen uppskattade värdet och det verkligt uppmätta värdet. *Leverage* är i sin tur ett mått på hur långt från sitt eget medelvärde en observation av en oberoende variabel befinner sig. Observationer med högt *leverage* för vissa variabler kan tydligt påverka vilka värden regressionskoefficienterna uppskattas till (Hair et al., 2014).

I två av de utpekade DeSo-områdena upptäcktes orimligt stort antal serveringsverksamheter sett till området placering och storlek. Dessa observationer exkluderades från analysen. Undersökningen av övriga områden utpekade som extremvärden hittades ingen indikation på att det skulle finnas ett genomgående geokodningsfel. De flesta observationer utpekade som extremvärden återfanns i stadskärnor där hög restaurangtäthet kan anses rimlig. Bägge borttagna observationer stod ut främst genom sitt residualvärde. För de observationer som hade särskilt högt *leverage* kunde ingen uppenbar felmätning upptäckas varför ingen av dessa observationer exkluderas.

Undersökning och exkludering av extremvärden ledde till att totalt 5983 Observationer/DeSo-områden används i den första regressionen.

Efter exkludering av förmodade mätfel var fördelningen för *Restauranger/km2* fortfarande tydligt skev åt höger. Förekommande sätt att hantera variabler med denna fördelning är att transformera om variabeln i logaritmisk form (Hair et al, 2014). I och med denna möjlighet till omskrivning av beroende variabeln genomförs två regressioner för att undersöka vad som bidrar till restaurangetablering. En där den beroende variabeln är *Restauranger/km2*. Därefter en regression där beroende variabeln är den naturliga logaritmen av *Serveringsverksamheter/km2*, här namngiven: *logRestauranger /km2*. Beslutet att transformera om beroendevariabeln har också grund i tidigare studie (Yang et al., 2017) där en log-linjär statistisk metod används, dock inte av exakt samma typ som för denna studie.

Problem kring logtransformering uppstod i och med förekomsten av områden som innehåller noll restauranger, naturliga logaritmen av noll är inte möjligt. Detta genererade istället 1053 ”missing values”. Två kontrollerande regressioner genomfördes där nollområden sattes till 1 respektive 0,1 innan transformering. Resultatet av dess vissas i Tabell 7 och 8 i appendix.

Jämförelse mellan dessa regressioner och den regression där nollområden uteslutits helt visade på endast små skillnader. Exkludering av sådana områden ansågs därför godtagbar. Det är dock värt att ha i åtanke inför analys av resultaten.

De slutgiltiga ekvationerna för att undersöka områdesfaktorer som bidrar till kluster av restaurangverksamheter syns i ekvation 1 & 2.

3.2.3 Regression av klusterverkan

Efter dessa regressioner togs nästa modell fram som syftade till att undersöka eventuellt värde av att ligga i ett kluster av liknande verksamheter. Variablerna från tidigare regressioner inkluderades här för att utgöra kontrollvariabler mot den variabel som var av störst intresse: *RestaurangerInom100m*. Som beroende variabel sattes som tidigare diskuterat *Omsättning*. Mätfelet som identifierats och tagits bort tidigare berodde på felaktigt utplacerade serveringsverksamheter. I och med att respektive verksamhet utgör en observation i följande regressioner antogs inför kommande regressioner att de som var felaktigt utplacerade redan exkluderats i tidigare stadiet varför ingen ny kontroll av extremvärden utfördes. Det fanns heller ingen möjlighet att bedöma rimligheten för extremvärden i den beroende variabeln, *Omsättning*, i detta fall då statistiken var hämtad direkt från bolagens årsredovisningar.

Histogram av beroendevariabeln visade dock även här på höger-skev fördelning. Förfarandet med logaritmering av variabeln upprepades därför också för denna regression. Två regressioner genomfördes därmed för att undersöka eventuell effekt av kluster. Se ekvation 3 & 4.

3.2.4 Slutgiltiga regressionsekvationer

Regression1 - Ekv.1:

$$\text{Restauranger}/\text{km}^2 = \beta_0 + \beta_1 \% \text{Män} + \beta_2 \% \text{Hyrratt} + \beta_3 \% \text{Borätt} + \beta_4 \% \text{Median_Ink} + \beta_5 \% \text{HögskoleUtb} + \beta_6 \% \text{SamBarn} + \beta_7 \% \text{SamUbarn} + \beta_8 \% \text{EnsBarn} + \beta_9 \% \text{Invkm2} + \varepsilon$$

Regression2 - Ekv.2:

$$\log \text{Restauranger}/\text{km}^2 = \beta_0 + \beta_1 \% \text{Män} + \beta_2 \% \text{Hyrratt} + \beta_3 \% \text{Borätt} + \beta_4 \% \text{Median_Ink} + \beta_5 \% \text{HögskoleUtb} + \beta_6 \% \text{SamBarn} + \beta_7 \% \text{SamUbarn} + \beta_8 \% \text{EnsBarn} + \beta_9 \% \text{Invkm2} + \varepsilon$$

Regression3 - Ekv.3:

$$\text{Omsättning} = \beta_0 + \beta_1 \text{RestaurangerInom100m} + \beta_2 \% \text{Män} + \beta_3 \% \text{Hyrratt} + \beta_4 \% \text{Borätt} + \beta_5 \% \text{Median_Ink} + \beta_6 \% \text{HögskoleUtb} + \beta_7 \% \text{SamBarn} + \beta_8 \% \text{SamUbarn} + \beta_9 \% \text{EnsBarn} + \beta_{10} \% \text{Invkm2} + \varepsilon$$

Regression4 - Ekv.4:

$$\log \text{Omsättning} = \beta_0 + \beta_1 \text{RestaurangerInom100m} + \beta_2 \% \text{Män} + \beta_3 \% \text{Hyrratt} + \beta_4 \% \text{Borätt} + \beta_5 \% \text{Median_Ink} + \beta_6 \% \text{HögskoleUtb} + \beta_7 \% \text{SamBarn} + \beta_8 \% \text{SamUbarn} + \beta_9 \% \text{EnsBarn} + \beta_{10} \% \text{Invkm2} + \varepsilon$$

β_0 = Konstant i regressionskvationen, den punkt där regressionslinjen korsar Y-axeln

β_i = Oberoende variabels koefficient

ε = Felterm, den del av variation i beroendevariabeln som inte förklaras av övriga inkluderade variabler

(Hair et al., 2014)

3.3 Avväganden för statistisk analys

Den regressionsmetod (OLS) som används i studien är en relativt grundläggande statistisk modell som ofta är den första som avhandlas i böcker om statistisk analys och ekonometri. Den har använts av vissa studier i referenslistan i olika former och med olika transformeringar av variabler. Dock är det inte den modell som nödvändigtvis är mest lämplig sett till datamaterialet. Den i denna studie frekvent citerade undersökningen av liknande frågeställning i USA (Yang et al., 2017) använde en regressionsmodell kallad negativ-binomial-regression. Sett till likheterna med denna studie är det möjligt att en sådan regression hade gett mer tillförlitliga resultat. Avvägandet gjordes dock att den begränsade tidsramen för denna studie inte gav tillräckligt utrymme för en korrekt, trovärdig tillämpning av en ny och okänd regressionsmetod.

Exakt vilket samband som finns mellan de beroende och oberoende variablerna skulle också kunna undersökas mer i detalj innan regressionskvation bestämdes. De övervägande linjära eller log-linjära förhållanden som antogs gälla här grundas till stor del på studier i andra länder. Det skulle inte vara särskilt förvånande om korrelationer och typ av samband varierar beroende på kultur och skillnader i infrastruktur världen över. Även här möjliggör dock likheter i variabelspecifikation mellan studier också viss jämförelse, vilket kan påvisa om de finns uppenbara skillnader mellan länder och världsdelar.

3.4 Etiska aspekter & god forskningssed

Vetenskapsrådets publikation angående god forskningssed (2017) diskuterar vikten av försiktighet vid exkludering av data. Detta togs i beaktning på så sätt att endast de observationer som tydligt kunde sägas bero på mätfel eller fel i utplaceringen av restauranger exkluderas. I arbetsprocessen för att identifiera extremvärden fanns flera ytterligare observationer som enligt processen flaggades som extremvärden. Att vårdslöst ta bort dessa hade dock medfört risker då det inte fanns tydliga bevis för att dessa värden berodde på direkta felaktigheter i inhämtning eller utplacering, i brist på tydliga bevis kunde inget sakligt argument göras för att ta bort dessa värden. Om observationer tas bort utan att ett klart gränsvärde eller uppenbart argument kan urskiljas blir studien svår att replikera.

Vetenskapsrådet (2017) pekar också på vikten av att samtligt material, metod men också eventuella experimentella undersökningar är tydligt presenterade för att ge andra en möjlighet att testa reproducerbarheten. Av dessa anledningar inkluderas även de regressioner som senare genom analys av residualer anses lida av heteroskedasticitet. Detta dels för att göra jämförelse med andra regressioner enklare och därmed öppna upp för kritik av vald metod, dels för att motivera variabeltransformeringar och på ett transparent sätt visa på studiens arbetsgång.

Vidare menar Vetenskapsrådet att en diskussion av studiens begränsningar och hur dessa påverkar slutsatser är grundläggande för att uppnå någon som helst kvalitet i det presenterade resultatet. Av denna anledning finns det dels en diskussion av vald statistiska analysmetod, se avsnitt 3.3. Utöver detta finns också en utförligare diskussion kring några av studiens övriga begränsningar, denna placeras efter diskussionen av resultatet för studien. Denna diskussion tar upp hur vissa val i studien och datamaterial påverkar resultatet samt i vilken mån detta begränsar studiens generaliserbarhet.

4 Resultat

Nedan presenteras resultatet av studien och dess regressioner. Produkten av genomförda regressioner är till största del tabeller som presenterar variablernas koefficienter och eventuell signifikans. Beskrivande text finns till respektive tabell, informationen i texten och tabellerna är till övervägande del densamma.

Inför tolkning av regressionstabeller:

- Enheten för de variabler som används i andelsform är procent, vilket här innebär att en enhets förändring i oberoende variabel betyder plus/minus en procentenhet.
- I de regressioner där beroendevariabeln är i logaritmiska form blir tolkningen av koefficienterna i procentform för samtliga typer av oberoende variabler. Det vill säga koefficienterna anger procentuell förändring i Y (beroendevariabeln) vid en enhets förändring i X (oberoende variabel)

4.1 Områdesvariabler för restaurangverksamhet

Följande två tabeller med tillhörande förklaringar är resultatet av de regressioner som syftar till att besvara den första frågeställningen angående vad som drar restauranger till vissa områden och bidrar till skapandet av restaurangkluster.

Tabell 1: OLS-regression med robusta standardfel, serveringsverksamhet/km² som beroende variabel

| Restkm ² | Coef. | St.Err. | t-value | p-value | [95% Conf | Interval] | Sig |
|---------------------|---------|---------|---------|---------|-----------|-----------|-----|
| %Män | 0.375 | 0.200 | 1.87 | 0.061 | -0.018 | 0.767 | * |
| %Hyrrätt | 0.074 | 0.028 | 2.63 | 0.009 | 0.019 | 0.130 | *** |
| %Borätt | -0.016 | 0.025 | -0.65 | 0.514 | -0.066 | 0.033 | |
| Median_Ink_tkr | 0.127 | 0.014 | 9.39 | 0.000 | 0.101 | 0.154 | *** |
| %HögskoleUtb. | -0.142 | 0.043 | -3.33 | 0.001 | -0.226 | -0.058 | *** |
| %SamBarn | -0.408 | 0.050 | -8.21 | 0.000 | -0.506 | -0.311 | *** |
| %SamUbarn | 0.197 | 0.058 | 3.39 | 0.001 | 0.083 | 0.312 | *** |
| %EnsBarn | -0.752 | 0.137 | -5.49 | 0.000 | -1.021 | -0.484 | *** |
| Invkm ² | 0.003 | 0.000 | 14.33 | 0.000 | 0.003 | 0.004 | *** |
| Constant | -44.702 | 12.089 | -3.70 | 0.000 | -68.401 | -21.003 | *** |

| | | | |
|--------------------|-----------|----------------------|-----------|
| Mean dependent var | 9.154 | SD dependent var | 30.873 |
| R-squared | 0.470 | Number of obs | 5983.000 |
| F-test | 114.932 | Prob > F | 0.000 |
| Akaike crit. (AIC) | 54241.486 | Bayesian crit. (BIC) | 54308.453 |

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Tabell 1 visar resultatet av den inledande regressionen där restaurangverksamheter/km² i sin ursprungsform används som beroendevariabel. Av de signifikanta andelsvariablerna är det *Andel i hyresrätt* och *Andel samboende utan barn* som har positiv koefficient. En procentenhets ökning i hyresrättsvariabeln indikerar enligt regressionen en ökning av antalet restaurangverksamheter/km² om +0,074/km². En procentenhets ökning i samboende-utan-barnvariabeln hör enligt regressionen samman med en ökning av antalet restaurangverksamheter/km² om +0,197/km². Positiva koefficienter samt signifikanta är också variablerna Median_ink(tkr) och inv/km². En ökning av medianinkomsten i ett område med 1000 kr/år korrelerar enligt regressionen med en ökning av antalet restaurangverksamheter/km² med +0,127/km². En ökning av antalet inv/km² med en person har i regressionen en korrelation med ökning av antalet restaurangverksamheter/km² om +0,003/km².

Signifikanta variabler med negativ koefficient är för denna regression: *Andel högskoleutbildade*, en procentenhets ökning i denna variabel innebar en minskning av antalet restauranger/km² om -0,142/km². *Andel samboende m. barn* en procentenhets ökning i denna variabel innebar en minskning av antalet restauranger/km² om -0,408/km². *Ensamstående m. barn* en procentenhets ökning i denna variabel innebar en minskning av antalet restauranger/km² om -0,752/km².

Denna regression är dock genomförd på data som inte är normalfördelad utan snedvriden åt höger där beroendevariabeln antar ett maxvärde som ligger långt ifrån sitt medelvärde (Se APPENDIX: A, Tabell 9). Regressionen visar också efter grafisk visualisering av dess residualer tydliga tecken på heteroskedasticitet och icke-normalfördelade residualer (se APPENDIX: A, figur 2 & 3). Detta medför att modellen inte förväntas göra konsekvent likvärdiga uppskattningar av beroendevariabeln. Av denna anledning presenteras denna tabell främst för att tydliggöra jämförelser med regression2 som syns i Tabell 2.

Tabell 2: OLS-regression med robusta standardfel, naturliga logaritmen av restaurangverksamhet/km² som beroende variabel. Nollområden exkluderade.

| logRestkm2 | Coef. | St.Err. | t-value | p-value | [95% Conf | Interval] | Sig |
|--------------------|--------|-----------|----------------------|---------|-----------|-----------|-----|
| %Män | -0.214 | 0.011 | -19.28 | 0.000 | -0.235 | -0.192 | *** |
| %Hyrrätt | 0.058 | 0.002 | 36.18 | 0.000 | 0.055 | 0.062 | *** |
| %Boratt | 0.046 | 0.001 | 33.38 | 0.000 | 0.044 | 0.049 | *** |
| Median_Ink_tkr | 0.006 | 0.001 | 9.91 | 0.000 | 0.004 | 0.007 | *** |
| %HögskoleUtb. | 0.009 | 0.002 | 3.84 | 0.000 | 0.004 | 0.013 | *** |
| %SamBarn | 0.054 | 0.004 | 14.44 | 0.000 | 0.046 | 0.061 | *** |
| %SamUbarn | -0.059 | 0.005 | -12.35 | 0.000 | -0.069 | -0.050 | *** |
| %EnsBarn | -0.112 | 0.010 | -10.84 | 0.000 | -0.132 | -0.092 | *** |
| Invkm2 | 0.000 | 0.000 | 19.33 | 0.000 | 0.000 | 0.000 | *** |
| Constant | 6.970 | 0.635 | 10.98 | 0.000 | 5.725 | 8.215 | *** |
| Mean dependent var | | -0.065 | SD dependent var | | | 2.708 | |
| R-squared | | 0.766 | Number of obs | | | 4930.000 | |
| F-test | | 1639.719 | Prob > F | | | 0.000 | |
| Akaike crit. (AIC) | | 16680.617 | Bayesian crit. (BIC) | | | 16745.648 | |

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Tabell 2 visar resultatet av regressionen där restaurangverksamheter/km² har transformerats i logaritmisk form och utgör den beroende variabeln. Samtliga variabler visar här på ett signifikant samband på med beroendevariabeln. Bland andelsvariablerna ses *Andelen män* samt *Ensamstående vuxen m. barn* som de variabler med störst koefficienter. En procentenhets ökning av andelen män i ett område innebär enligt denna regression en minskning av restauranger/km² med -21,4%/km². Likaså innebär en procentenhets ökning av andelen ensamstående m. barn i ett område en minskning av restaurangverksamheter/km² med -11,2%/km².

Andelsvariablerna med positiv koefficient är följande: *Andel i hyresrätt* vilken har koefficient som tolkas som att en procentenhets ökning i denna variabel motsvarar ökning av restaurangtätheten om +5,8%/km². *Andel i bostadsrätt* vilken har koefficient som tolkas som att en procentenhets ökning i denna variabel motsvarar ökning av restaurangtätheten om +4,6% /km². *Andel med högskoleutbildning* vilken har koefficient som tolkas som att en procentenhets ökning i denna variabel motsvarar ökning av restaurangtätheten om +0,9%/km². *Andel samboende par m. barn* har koefficient som tolkas som att en

procentenhets ökning i denna variabel motsvarar ökning av restaurangtätheten om +5,4%/km². *Andelen samboende utan barn* är signifikant men har negativ koefficient och en procentenhets ökning av denna variabel innebär därför minskning av restaurangtätheten om -5,9%/km²

För de variabler som inte utgörs av andelar av områdesbefolkningen, *Medianinkomst* och *Invånartäthet*, har bägge två positiva koefficienter. Om medianinkomsten i ett område ökar med 1000 kronor per år innebär det en ökning av beroendevariabeln om +0,6%/km² enligt denna regression. För inv/km² syns att variabeln är signifikant men koefficienten är så pass liten att den inte visas. Detta är endast ett resultat av antalet redovisade decimaler i tabellen. En ökning av invånarantalet med en person per kvadratkilometer har rimligtvis mycket liten effekt. Ytterligare kontroll i Stata visar att koefficienten för inv/km² är cirka 0,000073. För att sätta det i paritet med måttet för medianinkomst kan det uttryckas som att en ökning med 1000 inv/km² enligt regressionen hänger samman med en ökning av antalet restaurangverksamheter/km² om +7,3%/km².

Det resultat som kan vara anmärkningsvärt jämfört regression1 är dock de oberoende variabler vars koefficientriktning eller signifikans förändrats. För variablerna *Andel i bostadsrätt* samt *Andel män* syns inget signifikant samband med beroendevariabeln i Tabell 1, i Tabell 2 visas däremot bägge vara signifikanta. *Andel med högskoleutbildning* är signifikant i bägge modeller men negativ koefficient i första modellen och positiv i den andra. Likaså *Andel samboende par m. barn* och *Samboende par utan barn* som är signifikanta i bägge regressioner men har omvända tecken i Tabell 1 jämfört Tabell 2.

Grafisk visualisering av residualerna för regression2 visar tecken på heteroskedasticitet även för denna modell, dock ser residualerna ut att ligga närmare en normalfördelning (se APPENDIX: A, figur 4 & 5).

4.2 Värdet av kluster

Följande två tabeller med tillhörande förklaringar är resultatet av de regressioner som syftar till att besvara den andra frågeställningen huruvida kluster har ett värde i sig för den enskilda restaurangen.

Tabell 3: OLS-regression med robusta standardfel, restaurangverksamhetens omsättning som beroende variabel

| Omsättning(tkr) | Coef. | St.Err. | t-value | p-value | [95% Conf | Interval] | Sig |
|--------------------|------------|------------|----------------------|---------|------------|------------|-----|
| AntalRest100m | 120.823 | 125.418 | 0.96 | 0.335 | -125.024 | 366.669 | |
| %Män | 790.600 | 229.216 | 3.45 | 0.001 | 341.288 | 1239.912 | *** |
| %Hyrrätt | 29.388 | 14.718 | 2.00 | 0.046 | 0.538 | 58.239 | ** |
| %Borätt | 30.657 | 24.934 | 1.23 | 0.219 | -18.219 | 79.533 | |
| Median_Ink_tkr | 25.102 | 15.997 | 1.57 | 0.117 | -6.257 | 56.460 | |
| %HögskoleUtb. | 68.287 | 45.688 | 1.50 | 0.135 | -21.270 | 157.844 | |
| %SamBarn | -120.206 | 47.950 | -2.51 | 0.012 | -214.198 | -26.214 | ** |
| %SamUbarn | -91.578 | 82.240 | -1.11 | 0.266 | -252.786 | 69.631 | |
| %EnsBarn | -17.099 | 166.391 | -0.10 | 0.918 | -343.260 | 309.062 | |
| Invkm2 | -0.191 | 0.040 | -4.76 | 0.000 | -0.269 | -0.112 | *** |
| Constant | -38600.000 | 10477.017 | -3.69 | 0.000 | -59100.000 | -18100.000 | *** |
| Mean dependent var | | 7007.316 | SD dependent var | | | 28477.707 | |
| R-squared | | 0.011 | Number of obs | | | 9709.000 | |
| F-test | | 12.021 | Prob > F | | | 0.000 | |
| Akaike crit. (AIC) | | 226637.747 | Bayesian crit. (BIC) | | | 226716.735 | |

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Tabell 3 visar resultatet av regressionen där respektive restaurangs omsättning används som beroende variabel. Variabeln av störst intresse är den som mäter klustertätheten:

AntalRest100m. För denna variabel syns inget signifikant samband med beroendevariabeln.

Signifikant samband med *Omsättning* har variablerna: *Andelen män*, *Andel i hyresrätt*, *Andel samboende m. barn* samt *Invånartätheten/km2*.

En procentenhets ökning av andelen män i ett område betyder i denna regression en ökning av omsättningen med 790 060 kronor/år. En procentenhets ökning av andelen boende i hyresrätt betyder i sin tur en ökning av omsättningen med 29 000 kronor/år. För andelen samboende par med barn medför en procentenhets ökning en minskning av omsättningen med 120 206 kronor/år. Likaså tyder ökning av antalet invånare med en person per kvadratkilometer en minskning av omsättningen med -191 kr/år.

Återigen syns dock mönstret med heteroskedasticitet och icke-normalfördelade residualer likt regression 1 (se APPENDIX: A, figur 6 & 7).

Tabell 4: OLS-regression med robusta standardfel, naturliga logaritmen av restaurangverksamhetens omsättning som beroende variabel

| logOmsättning | Coef. | St.Err. | t-value | p-value | [95% Conf | Interval] | Sig |
|----------------|--------|---------|---------|---------|-----------|-----------|-----|
| AntalRest100m | 0.052 | 0.007 | 7.34 | 0.000 | 0.038 | 0.066 | *** |
| %Män | 0.019 | 0.008 | 2.46 | 0.014 | 0.004 | 0.034 | ** |
| %Hyrrätt | 0.006 | 0.001 | 5.19 | 0.000 | 0.004 | 0.008 | *** |
| %Borätt | 0.002 | 0.001 | 1.37 | 0.171 | -0.001 | 0.004 | |
| Median_Ink_tkr | 0.003 | 0.000 | 5.95 | 0.000 | 0.002 | 0.004 | *** |
| %HögskoleUtb. | 0.004 | 0.002 | 2.13 | 0.033 | 0.000 | 0.008 | ** |
| %SamBarn | -0.017 | 0.003 | -5.02 | 0.000 | -0.024 | -0.011 | *** |
| %SamUbarn | -0.004 | 0.004 | -0.89 | 0.372 | -0.011 | 0.004 | |
| %EnsBarn | -0.026 | 0.010 | -2.63 | 0.009 | -0.045 | -0.007 | *** |
| Invkm2 | -0.000 | 0.000 | -5.72 | 0.000 | 0.000 | 0.000 | *** |
| Constant | 6.332 | 0.432 | 14.64 | 0.000 | 5.485 | 7.180 | *** |

| | | | |
|--------------------|-----------|----------------------|-----------|
| Mean dependent var | 7.911 | SD dependent var | 1.597 |
| R-squared | 0.046 | Number of obs | 9709.000 |
| F-test | 43.692 | Prob > F | 0.000 |
| Akaike crit. (AIC) | 36214.296 | Bayesian crit. (BIC) | 36293.285 |

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Tabell 4 visar resultatet av regressionen där naturliga logaritmen av respektive restaurangs omsättning används som beroende variabel. Variabeln av störst intresse är den som mäter klustertätheten: *AntalRest100m*. För denna variabel syns här signifikant positivt samband med beroendevariabeln. Enligt denna regression innebär en ökning med en ytterligare serveringsverksamhet inom 100 meter, en ökning av restaurangers omsättning med +5,2%.

Därtill är variablerna *Andelen män*, *Andel i hyresrätt*, *Medianinkomst* samt *Andelen högskoleutbildade invånare* alla signifikanta med positiva koefficienter. En procentenhets ökning av andelen män i ett område motsvaras av en ökning av omsättningen med +1,9%. En procentenhets ökning i andel boende i hyresrätt motsvaras i sin tur av en +0,6% ökning av omsättningen. För medianinkomsten innebär en ökning av områdets medianinkomst med 1000 kr/år en ökning av restaurangers omsättning med +0,3%. En procentenhets ökning av andelen högskoleutbildade motsvarar i sin tur en ökning av omsättningen med +0,04%.

Variablerna *Andelen samboende m. barn*, *Invånare/km2* samt *Andel ensamstående m. barn* är alla tre signifikanta med negativa koefficienter. En procentenhets ökning av andelen samboende med barn innebär enligt denna regression minskning av omsättningen med -1,7%, samma förändring i andelen ensamstående med barn innebär en minskning av omsättningen med -2,6%. Koefficienten för *Invånare/km2* är så pass liten att den inte visas i tabellen men är i verkligheten -0,0000148 vilket medför att en ökning av antalet invånare/km2 med en person motsvaras av en minskning av omsättningen med -0,00148%.

Histogram av residualerna för regression4 visar tecken på att närma sig en normalfördelning jämfört regression3, dock inte så att de kan sägas vara helt normalfördelade. Scatterplot av residualerna visar svaga tecken på en ellipsform runt referenslinjen (se APPENDIX: A, figur 8 & 9).

5 Diskussion

I detta kapitel diskuteras studiens resultat och kopplas tillbaka till tidigare forskning. Kapitlet behandlar även studiens begränsningar och förslag till framtida forskning.

5.1 Regression 1 & 2

Residualdiagram för regression1 (APPENDIX: A, figur 3) tyder på att denna modell skattar bättre för låga värden på den beroende variabeln i och med relativt små residualler i början av referenslinjen. För större värden ökar dock både residualerna och dess spridning markant. Residualvariansen är således inte densamma för samtliga värden i ekvationen vilket här tolkas som ett tydligt tecken på heteroskedasticitet. Även om residualerna inte kan anses homoskedastiska för regression2 vid inspektion av dess residualdiagram syns inte en lika tydlig form för residualspridningen (se APPENDIX: A, figur 5). I och med att dessa residualler dessutom verkar vara normalfördelade är analysen att regression2 bör ge en bättre skattning sett över hela datamängden.

Lägg till detta att ingen hänsyn tas till varken typ av- eller storleken på verksamheten. Regression2 som tycks täcka in ett större spann av observationer, i och med mindre tydlig form på residualvariansen, bör då generera koefficienter med större generaliserbarhet än koefficienterna från Regression1. Det var också det som var målet med studien, att hitta faktorer som på ett generellt plan bidrar till etablering av serveringsverksamheter. I och med denna ansats att hitta variabler som i generell mening bidrar till ett stort antal restauranger kan studien möjligtvis användas som ett första steg när exempelvis nya verksamheter ska bestämma etableringsplats. Alternativt som ett inlägg i en planprocess av ett större område för att ge vägledning till hur mycket mark som bör vigas till verksamheter av restaurangtyp jämfört den förväntade demografiska sammansättningen i området. Detta för att nå en lämplig balans mellan invånare och olika typer verksamheter.

Av ovanstående resonemang antas variabler och koefficienter framtagna i regression2 ge en mer generell bild och därmed något trovärdigare resultat i att hitta områdesvariabler som på övergripande nivå bidrar till att restaurangkluster skapas.

Jämfört studien från USA (Yang et al.,2017) stämmer det resultat som fås från regression2 i denna studie relativt bra överens med de variabler som visats signifikanta i tidigare studie. Andelen män i närområdet har negativ koefficient i bägge studierna. Samtidigt har medianinkomst, invånartäthet, andel i hyresrätt och andelen högskoleutbildade alla positiva koefficienter precis som föregående studie. Den mer tvetydiga skillnaden från tidigare undersökning gäller de variabler som handlar om de enskilda hushållens sammansättning. Yang et al., (2017) påvisade att ju större hushåll i området desto mindre var tätheten av restaurangverksamheter. För undersökningen genomförd här syns negativ koefficient för variabeln andel ensamstående med barn samt andelen samboende utan barn. För variabeln andel samboende par m. barn ses dock en positiv koefficient vilket kan vara förvånande givet att denna variabel rimligtvis borde innehålla de största hushållen. Då ingen uppenbar förklaring kan utläsas kan denna skillnad troligtvis bero av faktorer som inte lyckats specificeras i regressionsekvationen. En teori, som dock inte hittats klara bevis för i denna studie, är om genomsnittligt antal barn i varje hushåll varierar mellan länderna. Detta skulle då kunna förklara skillnader i koefficientriktning. Antalet barn skulle också i olika grad kunna korrelera med hushållsinkomst. Om exempelvis fattiga familjer i USA tenderar att ha fler barn än fattiga familjer i Sverige kan variabeln för hushållsstorlek vara en uppskattning på

hushållets disponibla inkomst snarare än att familjestorleken skulle vara avgörande för restaurangtätheten. Alternativt att på vilket sätt hushållen är spridda i geografin varierar mellan länderna. Om det i USA finns en tydligare gräns mellan områden med barnfamiljer kontra mindre hushåll kan detta möjligtvis ge tydligare resultat för variabeln som mäter hushållsstorlek och dess inverkan.

Koefficienten för *Andelen män* kan i resultatet sticka ut som hög i och med att uppskattningen gör att en procentenhets förändring i denna variabel har +21,4% som motsvarande förändring i den beroende variabeln. Vid närmare kontroll av den beskrivande statistiken för de olika oberoende variablerna (se APPENDIX: A, Tabell 9) syns dock att medelvärdet för denna variabel ligger runt 50% och med en relativt liten standardavvikelse. Detta pekar på att fördelningen mellan män och kvinnor är relativt jämn över DeSo-områdena. En procentenhet upp eller ner kan då anses vara en relativt stor förändring, vilket då kan förklara den stora koefficienten. Detta resonemang bör även vara tillämpligt för variabeln med näst störst koefficient: *Ensamstående med barn* vilken också ses ha relativt låg standardavvikelse.

5.2 Regression 3 & 4

För undersökningen av om det finns ett värde för serveringsverksamheter av att befinna sig ett kluster av liknande verksamheter fanns viss skillnad mellan de två regressionerna och resultatet är därför inte helt entydigt. Störst vikt läggs dock vid resultatet från regression 4. Detta baseras likt föregående regressionsjämförelse på skillnader i heteroskedasticitet och normalfördelning av residualer. Regression 4 verkar enligt samma resonemang som för regression 1 och 2 ge något bättre skattningar över en större datamängd.

Enligt tabell 4 verkar det finnas en fördel rent omsättningsmässigt för restauranger att befinna sig nära många andra verksamheter av samma slag, även när andra områdesfaktorer kontrolleras för. Detta bekräftar resultatet från tidigare studie av Göteborgs kommun (Evidens, 2016). Även om omsättning inte är det perfekta måttet för huruvida en restaurang är lyckad eller inte bör god omsättning ge bra möjligheter till god lönsamhet. Därmed kan det diskuteras om resultatet för denna studie delvis går emot den klassiska konkurrens- och vinstmaximeringsteori som togs upp inledningsvis. Där visades att inträde av fler konkurrenter bör förskjuta företagens efterfrågekurva åt vänster och över tid eliminera möjligheter till onormala vinster, sådan förskjutning av efterfrågekurvan bör ha negativ effekt på omsättningen. Resultatet i denna studie pekar dock i riktningen att nyttan som högre klustertäthet har på omsättningen skulle vara större än de nackdelar som kommer med ökad konkurrens. Det kan rimligtvis finnas flera skäl till fler företag inte etableras på marknaden även om möjlighet till god omsättning eller vinst finns. Ren lokalbrist kan vara en faktor i de områden som redan har en hög andel restaurangverksamheter.

Det kan också vara så att det finns variabler som inte specificerats i denna studie som egentligen är den övervägande anledningen till att restaurangverksamheter etableras på viss plats. Goda transportmöjligheter är en tänkbar sådan som lyfts tidigare. Om värdet av att ligga nära exempelvis kollektivtrafiknoder är långt mycket större än den negativa effekten av ökad konkurrens kommer kluster skapas runt sådana punkter oavsett. Skulle detta vara fallet är det inte säkert att kluster har ett värde i sig eller att det har ett betydligt mindre värde än vad som framkommit i denna studie.

Värt att påpeka är dock att stor omsättningen inte nödvändigtvis behöver betyda stor lönsamhet. Att ingen skillnad görs på typ av verksamhet samt det faktum att endast aktiebolag

inkluderas i denna del av analysen kan rimligtvis också påverka resultatet av regressionen. Intuitivt kan man tänka sig att en verksamhet som har en hög andel spontana gäster skulle dra större nytta av kluster som ökar den generella attraktiviteten för närområdet, jämfört en restaurang som endast serverar bokade besök som planeras långt i förväg. En uppdelning efter verksamhetstyp hade därför varit önskvärd.

5.3 Övriga iakttagelser

Jämför man resultaten i tabell 2 och tabell 4 finns det en del koefficienter som ger en något otydlig bild av undersökt ämne. I regression² rörande vilka områdesfaktorer som bidrar till en ökad täthet av serveringsverksamheter är koefficienten för *andelen män* negativ. I regression⁴ som undersöker omsättningen för verksamheter är koefficienten för *andelen män* istället positiv. Omvänt förhållande gäller dessutom variabeln *andel samboende m. barn*. I den första delen av studien sågs denna variabel ha en positiv koefficient och bidra till ökad restaurangtäthet. I tabell⁴ framkom istället en negativ koefficient som tyder på minskad omsättning för serveringsverksamheter när andelen sådana hushåll i närområdet ökar.

Detta kan ses som paradoxalt då de flesta restaurangverksamheter borde vara intresserade av att maximera sin omsättning och då inte placera sig i områden där förutsättningarna för detta är sämre. Detta kan återigen vara ett resultat av att ingen skillnad på typ av verksamhet görs. Det är inte en omöjlig tanke att områden med en hög andel barnfamiljer skulle ha en högre andel mindre omfattande verksamheter så som kaféer eller mindre restauranger jämfört exempelvis stadskärnor. Sådant förhållande skulle då resultera i en negativ koefficient för andel barnfamiljer relativt omsättningen men skulle snarare vara en konsekvens av verksamhetens praktiska storlek snarare än att just barnfamiljer skulle vara en särskilt dålig kundgrupp.

6 Studiens begränsningar & framtida forskning

6.1 Bolagsformer för klusteranalys

Liksom många studier begränsas även denna uppenbart av datatillgången, och tillhörande brister i det geokodningsverktyg som används för att placera ut restauranger. Att data angående företagets omsättning inte finns tillgänglig på samma sätt för samtliga bolagsformer medför problem med huruvida studiens resultat är generaliserbara för alla typer av restaurangverksamheter. Ingen djupare undersökning har gjorts angående hur bolagsform varierar med typ av restaurangverksamhet. Intuitivt kan man dock tänka sig att enklare bolagsformer som enskild firma eller handelsbolag i högre grad kan attrahera något mindre omfattande verksamheter. Detta skulle innebära att man konsekvent går miste om en viss typ av observation, exempelvis med mindre omsättning, vilket om så är fallet gör koefficienterna för variabeln *RestaurangerInom100m* missvisande. För att få klusterberäkningarna så korrekta som möjligt har dock samtliga bolagsformer inkluderats i kartbilden när beräkningarna genomförs, först därefter tas bolagsformer utan omsättningsdata bort. Genom detta får de 9733 aktiebolag som ändå används ett så rättvisande klustervärde som möjligt. Att endast inkludera aktiebolag kan dock tänkas ha vissa positiva konsekvenser. Genom att enklare företagsformer inte är med exkluderas rimligtvis en del bolag som inte bedriver stadigvarande verksamhet. Snabb överblick av data visade exempelvis på enskilda firmor som troligtvis endast tjänar till att föreningar eller evenemang ska kunna driva sporadisk serveringsverksamhet.

6.2 Val av beroendevariabel

För frågeställning ett där restaurangtätheten undersöks görs ingen skillnad mellan olika typer av restauranger. Rimligtvis kan olika demografiska faktorer spela olika stor roll beroende på vilkens sorts restaurang som bedrivs. Då den data som inhämtats inte innehöll detaljerad information kring restaurangtyp kunde sådan indelning inte göras även om det varit önskvärt för att få bättre bild av variationen inom restaurangbranschen. Att ingen sådan indelning genomförs genererar därmed koefficienter som får ses som mycket generella kring vad som är bra förutsättningar för restaurangverksamhet.

För frågeställning två används aktiebolags omsättning som ett mått på lönsamhet eller ekonomisk framgång, detta kan tänkas bli missvisande när ingen skillnad görs mellan typer av restaurangverksamhet. Större restauranger eller snabbmatsrestauranger som betjänar stora mängder kunder varje dag har troligtvis stor omsättning. Är personalkostnader och andra omkostnader i samma storleksklass kan detta betyda att den faktiska lönsamheten för restaurangen inte nödvändigtvis är bättre än för en mindre verksamhet. Omsättning används trots detta, dels på grund av upprepade svårigheter att nå den databas i Retriever som övriga lönsamhetsmått hittades i. Det möjliggör också viss jämförelse mot tidigare studie som också använt omsättning som beroende variabel. Då övriga förklarande variabler i tidigare studie inte helt klargjorts för bör dock denna jämförelse hållas till på generell nivå. En möjlig kandidat till beroendevariabeln hade också kunnat vara omsättning/anställd. Då en genomgång av företagsdata visade på att många företag kan ha hög omsättning men noll registrerade anställda ansågs det ha för hög risk att bli missvisande.

6.3 Demografiska statistikområden

Användandet av DeSo-områden som statistikunderlag medför också begränsningar. Dels då de endast innehåller statistik av demografisk och socioekonomisk karaktär. Detta medför att

vissa variabler som påvisats viktiga i andra studier så som god transporttillgång (Tzeng et al. 2002) inte inkluderas och kontrolleras för. Själva områdesindelningen i sig kan rimligtvis också påverka uppskattade koefficienter. Såväl invånartätheten som tätheten av restaurangverksamheter beräknas som områdesantal /områdesarea. Detta medför att tätheter och andra variabler som beaktas i många fall endast beräknas för det omedelbara närområdet, framförallt i storstäder där DeSo-områden kan ha liten area. En restaurang, bar- eller kafés kundområde är i många fall mycket större än det enskilda kvarteret det ligger i, varför det i viss mån hade varit bättre om närliggande områden och dess data också kunnat viktas in i regressionerna. Det som denna mindre indelning dock medför är att det kan ge mer precis vägledning till vilken typ av område nyetablering av restaurangverksamhet ska leta efter. Hade större områden använts så som kommuner eller stadsdelsnämnder hade en nyetablering kunnat ha väldigt varierande miljöer inom området att välja mellan.

6.4 Förslag till vidare forskning

Vidare studier kring ämnet skulle kunna fokusera på andra möjliga förklaringsvariabler än endast de demografiska. Som tidigare nämnt är exempelvis närhet till effektiv och enkel transport något som andra studier pekat på som förklaringsfaktor för en restaurangs framgång. Transport skulle också vara extra intressant ur den synvinkeln att hur människor väljer att transportera sig kan variera mycket beroende på vilket land man befinner sig i. I vissa länder är kollektivtrafik avgörande för att möta transportbehovet medan det i andra länder har mindre vikt.

En djupare studie av samma demografiska förklaringsvariabler som för denna studie men med en indelning av restaurangverksamheter utefter vilken typ av restaurang som bedrivs skulle också öka förståelsen avsevärt för hur nyetablering av restauranger kan göras så effektivt som möjligt. Detta då det rimligtvis finns en stor variation i vilken typ av kunder som en viss typ av restaurang anser sig ha som målgrupp. En sådan uppdelning var inte möjlig i denna studie på grund av bristen på så pass detaljerad information i den data som hämtades.

Vidareutveckling av denna studie som bygger på användning av geografiska informationssystem (GIS) skulle kunna vara att istället för användning av DeSo-områden använda statistik på individnivå, denna data är till viss del redan insamlad av SCB men är inte fritt tillgänglig på samma sätt som DeSo-statistiken. Om data på individnivå kan användas i kombination med exempelvis nätverksanalyser av trafiknätet i ett GIS-program skulle man kunna skapa en mer precis bild av hur stort respektive restaurangs potentiella kundområde faktiskt är. Kan man lägga till en analys av fotgängartrafiken i en stad skulle man ytterligare kunna öka förståelsen för effektiv restaurangplacering. Möjligtvis hela vägen ner till hur restauranger bör placera sig i själva gatunätet.

7 Slutsats

Frågeställningen kring vilka områdesfaktorer som skapar bäst förutsättningar för etablering av serveringsverksamhet och som i hög grad korrelerar med hög restaurangtäthet har besvarats till största del i tabell 2. Trots att ingen skillnad görs mellan typer av restaurangverksamheter kan signifikanta resultat fås fram. De följer dessutom till stora delar resultat från tidigare studier. Resultaten till frågeställning ett visade att i denna studie kunde ökade värden på variabler så som medianinkomst i området och invånartäthet bidra till positiva förutsättningar för etablering av restaurangverksamhet. Både andel boende i hyresrätt samt andel boende i bostadsrätt sågs också ha positiv korrelation med antalet serveringsverksamheter i området. Därtill sågs hög andel ensamstående m. barn samt hög andel män vara de två faktorerna som tydligast som korrelerar med en låg restaurangtäthet.

Genom att studien genomförts på nationell nivå för Sverige kan studien till stora delar bekräfta tidigare resultat från studier genomförda i USA för demografiska variabler, samt resultat i Göteborg för klustervariabeln. Då studien genomförts på något finare detaljnivå jämfört referensstudie från USA bör dessa resultat kunna ge ytterligare vägledning i utvärdering av områdets lämplighet för etablering av restaurangverksamhet. På ett större plan också vägledning vid framtagande av större detaljplaner för att få en lämplig balans mellan boende och olika typer av verksamheter.

Frågeställningen kring huruvida klustertendensen som syns inom restaurangbranschen har ett värde för den enskilda serveringsverksamheten besvarades till stora delar i tabell 4. Det generella svaret utifrån de analyser som genomförts inom ramen för denna studie är att täta restaurangkluster ser ut att innehålla restauranger med större omsättning. Med förutsättningen att hög omsättning är positivt verkar en hög klustertäthet därför bidra positivt till de verksamheter som befinner sig i klustret. Dock behövs en djupare undersökning om detta skiljer sig mellan olika typer av verksamheter inom branschen. En liknande analys men med olika lönsamhetsmått hade också varit lämplig för att göra resultaten mindre känsliga för den enskilda verksamhetens storlek och omfattning, och kunde då på ett bättre sätt undersökt ekonomiska framgång.

8 Källförteckning

- Esaiasson, P., Gilljam, M., Oscarsson, H. Wängnerud, L. & Towns, A. (2017). *Metodpraktikan: Konsten att studera samhälle, individ och marknad*. Stockholm: Wolters Kluwer
- Evidens & Spacescape. (2016). VÄRDESKAPANDE STADSUTVECKLING. Hämtad: 2020-05-02 http://framtiden.se/wp-content/uploads/2017/02/Vardeskapande-stadsutvecklingRapport-161206_mail1.pdf
- Fisher, Douglas P. (1997) Location, Location, Location: Ensuring a Franchisee's Success, *Hospitality Review: Vol. 15 : Iss. 1 , Article 4*. Hämtad 2020-05-10 <https://digitalcommons.fiu.edu/hospitalityreview/vol15/iss1/4>
- Fujita, M., & Thisse, J. (1996). Economics of Agglomeration. *Journal of The Japanese and International Economies*, 10(4), 339-378. DOI: 10.1017/CBO9780511805660
- Greenlaw S. A. & Shapiro, D. (2017). *Principles of Economics 2e*, Houston, Texas: OpenStax. Hämtad 2020-05-01 <https://openstax.org/books/principles-economics-2e/pages/1-introduction>
- Gwartney, J. D., Stroup, R. L., Sobel, R. S., & Macpherson, D. A. (2009). *Economics: private and public choice*. 12th ed., macro-first version, South-Western Cengage Learning
- Hair, J., Black, W., Babin, B., & Anderson, R. (2014). *Multivariate data analysis* (Seventh edition, Pearson new international ed.)
- Hotelling, H. (1929). Stability in Competition. *The Economic Journal*, 39(153), 41-57. DOI:10.2307/2224214
- Litz, R.A. & Rajaguru, G. (2008). Does small store location matter? A test of three classic theories of retail location, *Journal of Small Business & Entrepreneurship*, Vol. 21 No. 4, pp. 477-92. DOI: 10.1080/08276331.2008.10593436
- Morland, K., Wing, S., Diez Roux, A., & Poole, C. (2002). Neighborhood characteristics associated with the location of food stores and food service places. *American Journal of Preventive Medicine*, 22(1), 23-29.
- Nelson, R. (1958). *The selection of retail locations*. New York: F.W. Dodge.
- Parsa, H. G., Self, J.T., Njite, D., & King, T. (2005). Why Restaurants Fail. *Cornell Hotel and Restaurant Administration Quarterly*, 46(3), 304–322. DOI: [10.1177/0010880405275598](https://doi.org/10.1177/0010880405275598)
- Pillsbury, R. (1987). FROM HAMBURGER ALLEY TO HEDGEROSE HEIGHTS: TOWARD A MODEL OF RESTAURANT LOCATION DYNAMICS, *The Professional Geographer*, 39:3, 326-344, DOI: 10.1111/j.0033-0124.1987.00326.x
- Prayag, G., Landré, M., & Ryan, C. (2012). Restaurant location in Hamilton, New Zealand: Clustering patterns from 1996 to 2008. *International Journal of Contemporary Hospitality Management*, 24(3), 430-450. DOI: 10.1108/09596111211217897

- Retriever Business. (u.å.). *Företagsdata: Restaurang- och barverksamhet*. Retriever. Hämtad 2020-05-10: <http://web.retriever-info.com.ezproxy.ub.gu.se/services/businessinfo/displaySelectedCompanies>
- Rittenberg, L. & Tregarthen, Timothy D. (u. å) *Principles of Economics*. Hämtad 2020-05-04 https://ocw.mit.edu/ans7870/14/14.01SC/MIT14_01SCF11_rtext.pdf
- Self, J.T., Feree Jones, M., & Botieff, M. (2015) Where Restaurants Fail: A Longitudinal Study of Micro Locations, *Journal of Foodservice Business Research*, 18:4, 328-340, DOI: 10.1080/15378020.2015.1068670
- Statistiska centralbyrån. (2020). *Öppna geodata för DeSO – Demografiska statistikområden*. Hämtad 2020-01-05: <https://www.scb.se/vara-tjanster/oppna-data/oppna-geodata/deso--demografiska-statistikomraden/>
- Thise, J-F. (2019). Urban, Rural, and Regional Economics: *Economics of Agglomeration*, DOI: 10.1093/acrefore/9780190625979.013.152
- Tzeng, G., Teng, M., Chen, J., & Opricovic, S. (2002). Multicriteria selection for a restaurant location in Taipei. *International Journal of Hospitality Management*, 21(2), 171-187. DOI: 10.1016/S0278-4319(02)00005-1
- UCLA: Statistical Consulting Group. (2020). REGRESSION WITH STAT – REGRESSION DIAGNOSTICS. Hämtad 2020-05-10: <https://stats.idre.ucla.edu/stata/webbooks/reg/chapter2/stata-webbooksregressionwith-statachapter-2-regression-diagnostics/>
- United States Census Bureau. (2020). *Census 2000: Geographic Terms and Concepts* Hämtad 2020-05-15: <https://www2.census.gov/geo/pdfs/reference/glossry2.pdf>
- Vetenskapsrådet. (2017). God forskningsd (Reviderad utgåva ed.)
- Yang, Y., Roehl, W., & Huang, J. (2017). Understanding and projecting the restaurantscape: The influence of neighborhood sociodemographic characteristics on restaurant location. *International Journal of Hospitality Management*, 67, 33-45. <https://doi.org/10.1016/j.ijhm.2017.07.005>.

9 APPENDIX A.

Tabell 5: Korrelationstabell oberoende variabler regression 1 & 2

| Variables | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|--------------------|--------|--------|--------|--------|--------|--------|--------|-------|-------|
| (1) %Män | 1.000 | | | | | | | | |
| (2) %Hyrrätt | -0.192 | 1.000 | | | | | | | |
| (3) %Borätt | -0.218 | -0.122 | 1.000 | | | | | | |
| (4) Median_Ink_tkr | 0.034 | -0.437 | 0.368 | 1.000 | | | | | |
| (5) %HögskoleUtb | -0.058 | -0.121 | 0.564 | 0.655 | 1.000 | | | | |
| (6) %SamBarn | 0.255 | -0.628 | -0.327 | 0.437 | 0.057 | 1.000 | | | |
| (7) %SamUbarn | 0.036 | -0.490 | -0.195 | 0.209 | -0.128 | 0.244 | 1.000 | | |
| (8) %EnsBarn | -0.064 | 0.213 | -0.017 | -0.105 | -0.167 | 0.125 | -0.464 | 1.000 | |
| (9) Invkm2 | -0.204 | 0.155 | 0.493 | 0.265 | 0.452 | -0.214 | -0.261 | 0.112 | 1.000 |

Tabell 6: Korrelationstabell oberoende variabler regression 3 & 4

| Variables | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
|--------------------|--------|--------|--------|--------|--------|--------|--------|--------|-------|-------|
| (1) AntalRest100m | 1.000 | | | | | | | | | |
| (2) %Män | -0.037 | 1.000 | | | | | | | | |
| (3) %Hyrrätt | 0.184 | -0.191 | 1.000 | | | | | | | |
| (4) %Borätt | 0.224 | -0.217 | -0.122 | 1.000 | | | | | | |
| (5) Median_Ink_tkr | 0.079 | 0.034 | -0.438 | 0.368 | 1.000 | | | | | |
| (6) %HögskoleUtb. | 0.215 | -0.058 | -0.120 | 0.564 | 0.653 | 1.000 | | | | |
| (7) %SamBarn | -0.219 | 0.253 | -0.628 | -0.327 | 0.438 | 0.056 | 1.000 | | | |
| (8) %SamUbarn | -0.164 | 0.034 | -0.491 | -0.195 | 0.210 | -0.129 | 0.246 | 1.000 | | |
| (9) %EnsBarn | -0.017 | -0.065 | 0.213 | -0.017 | -0.105 | -0.167 | 0.125 | -0.463 | 1.000 | |
| (10) Invkm2 | 0.214 | -0.196 | 0.158 | 0.489 | 0.258 | 0.450 | -0.216 | -0.266 | 0.109 | 1.000 |

Tabell 7: Logaritm av serveringsverksamhet/km2 (nollområden satta till =1)

| logRestkm2 | Coef. | St.Err. | t-value | p-value | [95% Conf | Interval] | Sig |
|--------------------|-----------|---------|----------------------|---------|-----------|-----------|-----|
| %Män | -0.208 | 0.010 | -20.02 | 0.000 | -0.228 | -0.187 | *** |
| %Hyrrätt | 0.058 | 0.001 | 38.64 | 0.000 | 0.055 | 0.061 | *** |
| %Borätt | 0.046 | 0.001 | 36.62 | 0.000 | 0.044 | 0.049 | *** |
| Median_Ink_tkr | 0.005 | 0.001 | 10.46 | 0.000 | 0.004 | 0.006 | *** |
| %HögskoleUtb. | 0.010 | 0.002 | 4.88 | 0.000 | 0.006 | 0.014 | *** |
| %SamBarn | 0.050 | 0.003 | 15.13 | 0.000 | 0.044 | 0.057 | *** |
| %SamUbarn | -0.058 | 0.004 | -13.08 | 0.000 | -0.066 | -0.049 | *** |
| %EnsBarn | -0.104 | 0.009 | -11.05 | 0.000 | -0.122 | -0.085 | *** |
| Invkm2 | 0.000 | 0.000 | 19.64 | 0.000 | 0.000 | 0.000 | *** |
| Constant | 6.531 | 0.598 | 10.92 | 0.000 | 5.358 | 7.703 | *** |
| Mean dependent var | -0.324 | | SD dependent var | | 2.727 | | |
| R-squared | 0.763 | | Number of obs | | 5983.000 | | |
| F-test | 1991.708 | | Prob > F | | 0.000 | | |
| Akaike crit. (AIC) | 20391.397 | | Bayesian crit. (BIC) | | 20458.364 | | |

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Tabell 8: Logaritm av serveringsverksamhet/km2 (nollområden satta till = 0,1)

| logRestkm2 | Coef. | St.Err. | t-value | p-value | [95% Conf | Interval] | Sig |
|----------------|--------|---------|---------|---------|-----------|-----------|-----|
| %Män | -0.229 | 0.013 | -18.06 | 0.000 | -0.253 | -0.204 | *** |
| %Hyrrätt | 0.057 | 0.002 | 30.84 | 0.000 | 0.053 | 0.060 | *** |
| %Borätt | 0.045 | 0.002 | 27.56 | 0.000 | 0.041 | 0.048 | *** |
| Median_Ink_tkr | 0.007 | 0.001 | 10.94 | 0.000 | 0.006 | 0.008 | *** |
| %HögskoleUtb. | 0.007 | 0.003 | 2.43 | 0.015 | 0.001 | 0.012 | ** |
| %SamBarn | 0.034 | 0.004 | 8.00 | 0.000 | 0.026 | 0.042 | *** |
| %SamUbarn | -0.076 | 0.006 | -13.28 | 0.000 | -0.087 | -0.065 | *** |
| %EnsBarn | -0.124 | 0.013 | -9.88 | 0.000 | -0.148 | -0.099 | *** |
| Invkm2 | 0.000 | 0.000 | 15.38 | 0.000 | 0.000 | 0.000 | *** |
| Constant | 7.826 | 0.751 | 10.42 | 0.000 | 6.354 | 9.298 | *** |

| | | | |
|--------------------|-----------|----------------------|-----------|
| Mean dependent var | -0.729 | SD dependent var | 3.031 |
| R-squared | 0.673 | Number of obs | 5983.000 |
| F-test | 1330.108 | Prob > F | 0.000 |
| Akaike crit. (AIC) | 23577.873 | Bayesian crit. (BIC) | 23644.840 |

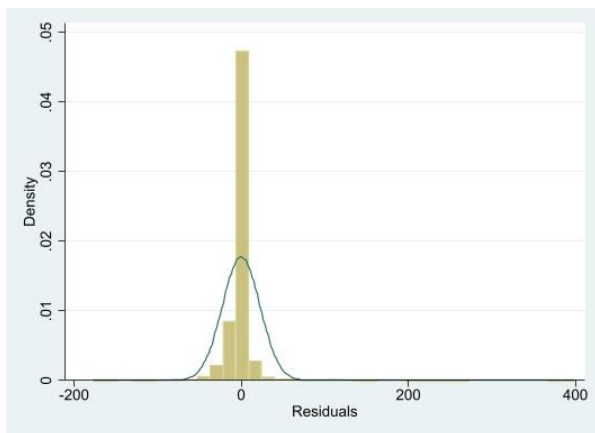
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Tabell 9: Beskrivande statistik regression 1 & 2

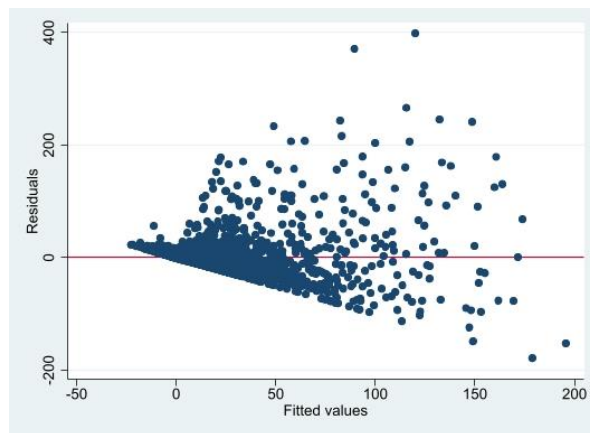
| Variable | Obs | Mean | Std.Dev. | Min | Max |
|---------------|------|---------|----------|--------|----------|
| Restkm2 | 5983 | 9.154 | 30.873 | 0 | 518.45 |
| logRestkm2 | 4930 | -.065 | 2.708 | -9.121 | 6.251 |
| %Män | 5983 | 50.296 | 2.13 | 38.083 | 70.442 |
| %Hyrrätt | 5983 | 28.379 | 27.818 | 0 | 100 |
| %Borätt | 5983 | 18.482 | 24.235 | 0 | 100 |
| Median_InkTkr | 5983 | 285.447 | 61.542 | 3.012 | 562.797 |
| % HögskoleUtb | 5983 | 26.795 | 13.473 | 4.481 | 79.769 |
| %SamBarn | 5983 | 23.958 | 11.065 | 0 | 74.403 |
| %SamUbarn | 5983 | 24.826 | 7.318 | 3.747 | 48.808 |
| %EnsBarn | 5983 | 7.062 | 2.629 | 0 | 21.242 |
| Invkm2 | 5983 | 3280.12 | 6075.139 | .069 | 57390.42 |

Tabell 10: Beskrivande statistik regression 3 & 4

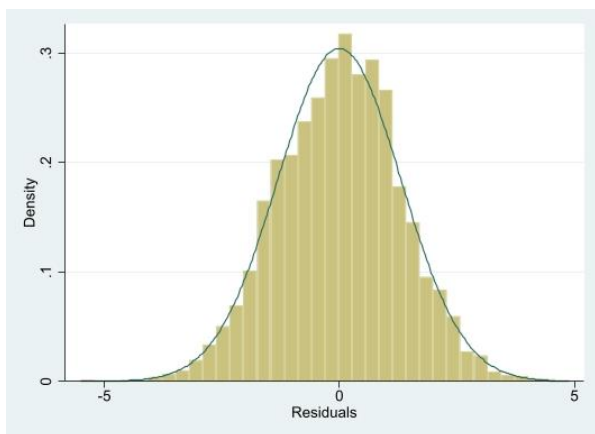
| Variable | Obs | Mean | Std.Dev. | Min | Max |
|---------------|------|----------|----------|--------|----------|
| Omsättning | 9733 | 6999.481 | 28443.76 | 1 | 1360000 |
| logOmsättning | 9733 | 7.91 | 1.596 | 0 | 14.121 |
| AntalRest100m | 9733 | 1.376 | 2.266 | -1 | 18 |
| %Män | 9733 | 49.791 | 2.263 | 38.083 | 70.035 |
| %Hyrrätt | 9733 | 38.225 | 26.724 | 0 | 100 |
| %Borätt | 9733 | 28.169 | 25.401 | 0 | 100 |
| Median_InkTkr | 9733 | 295.168 | 60.769 | 23.436 | 562.797 |
| % HögskoleUtb | 9733 | 31.763 | 13.874 | 4.481 | 74.205 |
| %SamBarn | 9733 | 18.773 | 10.398 | .35 | 68.839 |
| %SamUbarn | 9733 | 23.296 | 6.09 | 5.851 | 48.808 |
| %EnsBarn | 9733 | 6.691 | 2.239 | 0 | 20.745 |
| Invkm2 | 9733 | 5119.684 | 8042.733 | .069 | 56289.02 |



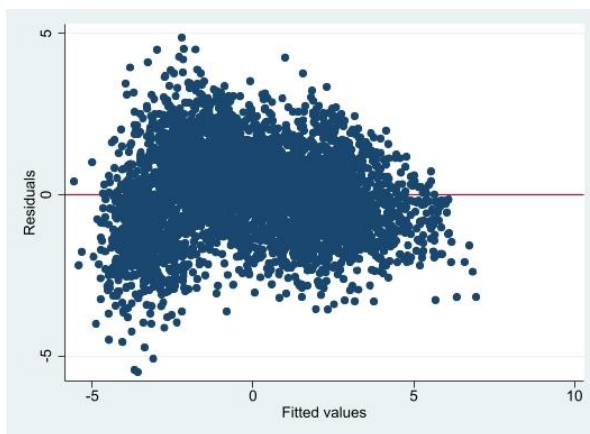
Figur 2: Histogram residualer regression1
Normalfördelningskurva inlagd i blått.



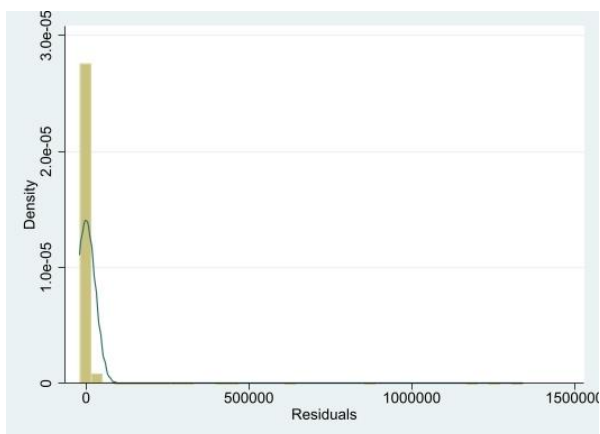
Figur 3: Scatterplot residualer regression1.
Röd referenslinje anger residual/skattningsfel = 0



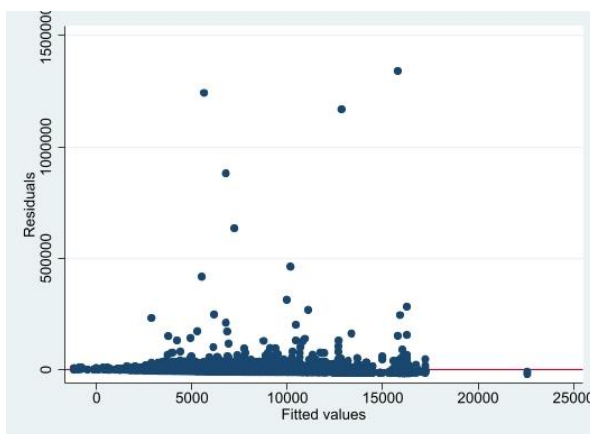
Figur 4: Histogram residualer regression2
Normalfördelningskurva inlagd i blått.



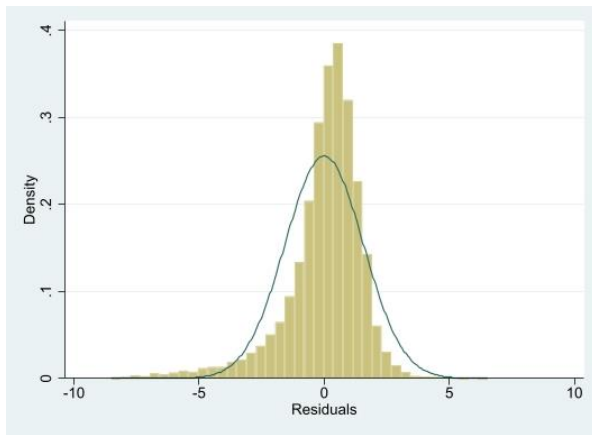
Figur 5: Scatterplot residualer regression2
Röd referenslinje anger residual/skattningsfel = 0



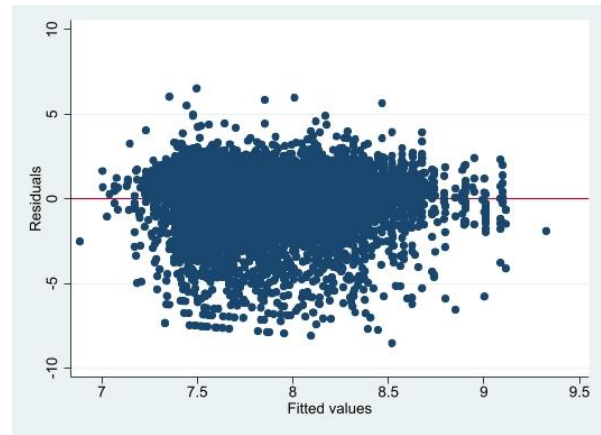
Figur 6: Histogram residualer regression3
Normalfördelningskurva inlagd i blått.



Figur 7: Scatterplot residualer regression3
Röd referenslinje anger residual/skattningsfel = 0



Figur 8: Histogram residualer regression4
Normalfördelningskurva inlagd i blått.



Figur 9: Scatterplot residualer regression4
Röd referenslinje anger residual/skattningsfel = 0