



**Statistiska Institutionen
Lunds Universitet**

**STA104:1
HT 2008
Magisteruppsats, 15 poäng**

Fordonsbestånd och nyregistreringar

**- en statistisk analys av regionala skillnader och
strukturella faktorer**

Författare:
Henrik Svanström
Samuel Roos

Handledare:
Panagiotis Mantalos

Abstract

The official statistics on communication and transports in Sweden is administered by *Statens Institut för kommunikationsanalys* (SIKA). SIKA's commission includes production of statistics on the Swedish vehicle fleet. These statistics provide important information to financial institutes, ministries, industry and mass media among others. Furthermore, the vehicle statistics include measures of comparison between different counties and municipalities. These regional measures suggest large differences concerning the number of passenger cars per inhabitant and registered passenger cars per inhabitant in different regions. SIKA has had a number of inquiries seeking an explanation to these differences.

The objectives of this study are; **i)** to examine regional differences in the number of passenger cars per inhabitant **ii)** to investigate and present factors that influence the number of passenger cars per inhabitant and to compose an adequate statistical model **iii)** to examine regional differences in the number of newly registered passenger cars per year and inhabitant **iv)** to investigate and present factors that influence the number of newly registered passenger cars per year and inhabitant and to compose an adequate statistical model.

The study makes use of a wide spectrum of demographic-, economic-, social- and infrastructural explaining variables. Most of the variables are acquired from *Sveriges kommuner och landsting's* classification of municipalities.

The study shows that the main differences in passenger cars per inhabitant exist between major cities and more sparsely populated areas. The major cities show a significantly lower number of passenger cars per inhabitant than other areas of the country. A polynomial regression model suggests that a higher proportion of inhabitants in the age intervals 35 – 44 and 55 – 64 and a higher proportion of detached housing have a positive effect on the number of passenger cars per inhabitant. The availability of public transport has a negative effect on the number of passenger cars per inhabitant.

Furthermore, the study shows that suburban municipalities have a higher number of newly registered passenger cars per year and inhabitant than other municipalities. Sparsely populated municipalities prove to have a lower number of newly registered passenger cars than other municipalities. A two-sided panel data model with fixed effects shows that mean income, employment, proportion of commuters, the proportion of inhabitants in the age interval 15 – 44 and the number of passenger cars per inhabitant are positively associated with the number of newly registered passenger cars per year and inhabitant.

Keywords: Regression analysis, Longitudinal data

1. Inledning

Statens institut för kommunikationsanalys (SIKA) är en statlig myndighet under Näringsdepartementet, med ansvar för övergripande analys av transport- och kommunikationssystemen. Detta verksamhetsområde innefattar analys av bantrafik, kollektivtrafik, kommunikationsmönster, luftfart, post, tele, sjöfart och vägtrafik. SIKA tillhandahåller, utvecklar och publicerar den officiella statistiken för respektive område.

Som del av framtagningen av officiell vägtrafikstatistik ingår att utifrån vägtrafikregistret föra statistik över den svenska fordonsparken. Detta innefattar att redovisa statistik över det befintliga fordonsbeståndet samt nyregistreringar av fordon. Fordonsstatistiken fyller en viktig funktion för bl.a. finansiella institut, departement, branschorganisationer, massmedier och andra som följer transportsektorns och bilbranschens utveckling ur olika perspektiv.

Den officiella fordonsstatistiken redovisas med en regional upplösning på läns- och kommunnivå. Statistiken inkluderar bl.a. jämförelsemått mellan olika regioner, såsom fordonstäthet, d.v.s fordonsbestånd per invånare, samt antal nyregistreringar per invånare. Detta föranleder att SIKA har fått ett stort antal förfrågningar från bl.a. län och kommuner som söker en förklaring till de stora regionala skillnader som förekommer i statistiken.

Denna uppsats skrivs på ämnesförslag från SIKA, med avsikt att statistiskt belysa och förklara den regionala variation som förekommer i fordonsstatistiken. Temat för undersökningen har givits av SIKA, medan tillvägagångssätt och specifik avgränsning har utarbetats självständigt.

Uppsatsens syfte är att inledningsvis undersöka och beskriva regionala skillnader beträffande fordonsbestånd per invånare med avseende på registrerade personbilar. Uppsatsen avser att därefter belysa vilka faktorer som påverkar personbiltäthet på nationell och regional nivå. Det innebär att ta reda på och kontrollera samband mellan fordonstäthet och tillämpliga förklarande variabler. Avsikten är att utifrån dessa variabler skapa en adekvat förklarande modell.

Syftet är att därefter undersöka och redogöra för regionala skillnader beträffande antalet nyregistrerade fordon. Vidare så är avsikten att undersöka vilka faktorer som inverkar på antalet nyregistreringar på nationell och regional nivå och att utifrån dessa skapa en förklarande modell.

Både när det gäller fordonsbestånd och nyregistreringar innefattas samtliga registrerade personbilar som ägs av fysiska personer eller personliga företag.

Syftet leder till en frågeställning enligt:

- Vilka regionala skillnader förekommer beträffande fordonstäthet med avseende på personbilar?
- Vilka regionala skillnader förekommer beträffande antalet nyregistreringar med avseende på personbilar?
- Vilka faktorer inverkar på antalet nyregistreringar samt personbiltäthet på nationell och regional nivå?

Fordon som ägs av juridiska personer inkluderas inte i undersökningen. Det kan antas att väsentligt skilda faktorer inverkar på juridiska - respektive fysiska personers anskaffande och innehav av fordon. Vidare så registreras fordon ägda av juridiska personer inte nödvändigtvis

i den kommun där de faktiskt används. En allmän förklaringsmodell, som inkluderar fordon oavsett ägare, riskerar därmed att vara inadekvat.

Uppsatsen disponeras i tre avsnitt. Den första delen består av kapitel 1 och kapitel 2 som redogör för bakgrund, syfte och avgränsning respektive datamaterial och bearbetning av data. Därefter följer den andra delen som redogör för undersökningen med avseende på regionala skillnader i fordonsbestånd per invånare. Kapitel 3 beskriver de metoder som används, kapitel 4 redogör för tillvägagångssätt och kapitel 5 skildrar arbetet med att ta fram en förklarande modell. Slutligen redovisas undersökningens resultat och slutsatser i kapitel 6. Den tredje delen redovisar undersökningen av regionala skillnader i antal nyregistreringar. Kapitel 7 beskriver metod, kapitel 8 redogör för tillvägagångssätt och kapitel 9 redovisar arbetet med modellbestämning. Resultat och slutsatser redovisas i kapitel 10.

2. Datamaterial

Uppgifter om nyregistreringar, antalet personbilar i trafik samt personbilstäthet har hämtats från SIKAs databas över fordonsstatistik. SIKAs fordonsstatistik baseras på uppgifter om samtliga registreringspliktiga fordon i Vägverkets vägtrafikregister. Vissa mindre fel kan förekomma i datamaterialet, tex. som följd av kodningsfel i vägtrafikregistret.

Fordonsstatistiken redogör för beståndet av fordon, antalet nyregistreringar, vilken kategori som äger fordonen samt biltäthet på regional nivå. I undersökningen gällande fordonsbestånd per invånare inkluderas endast fordon som var i trafik vid årsskiftet 2006/2007. Avställda fordon inkluderas därmed inte. Som personbil räknas *bil huvudsakligen inrättad för befordran av personer, dock högst förare och åtta passagerare*.

När det gäller antal nyregistreringar sträcker sig uppgifterna över tidsperioden 1997-2006. Under denna tid har Knivsta och Nykvarn tillkommit som nya kommuner. Observationer tidigare än kommunernas tillkomst har inte ersatts.

Datamaterial som ligger till grund för undersökningens förklarande variabler kommer huvudsakligen från SCB:s statistikdatabas. Utöver detta har uppgifter hämtats från Arbetsmarknadsverket och SIKA. I samtliga fall används senast tillgänglig data. Tidpunkten för uppgifterna varierar därför mellan 2005 och 2007.

3. Metod

Syftet med den inledande undersökningen är att belysa förhållandet mellan den beroende variabeln *fordonstäthet* och ett antal förklarande variabler. För att åstadkomma detta används en *multipl regressionsmodell*. Avsnittet nedan avser att översiktligt redogöra för de metoder som används i undersökningen.

Utformningen av regressionsmodellen syftar till att välja ut de förklarande variabler som maximerar andelen förklarad variation. De specifika modellkriterier och test som används i undersökningen redovisas i appendix.

3.1 Regressionsmodellen

Antag att en variabel Y är beroende av k förklarande variabler X_1, X_2, \dots, X_k . Förhållandet mellan dessa kan då beskrivas med en *regressionsmodell*. Modellen kan skrivas:

$$(3-1) \quad Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik} + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, k$$

där:

Y_i är den i :te observationen på den beroende variabeln

$\beta_0, \beta_1, \dots, \beta_k$ är okända parametrar

X_{i1}, \dots, X_{ik} är det i :te observerade värdet på den j :te variabeln

ε_i är en slumpmässig felterm med $E(\varepsilon_i) = 0$ och $Var(\varepsilon_i) = \sigma^2$ (konstant). ε_i och

ε_j är okorrelerade så att $Cov(\varepsilon_i, \varepsilon_j) = 0$ för alla $i, j; i \neq j$.

Det i :te observerade värdet på Y är därmed summan av **1**) den konstanta termen $\beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik}$ och **2**) den slumpmässiga feltermen ε_i . Eftersom $E(\varepsilon_i) = 0$, så följer att:

$$\begin{aligned} E(Y_i) &= E(\beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik} + \varepsilon_i) = \\ &\beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik} + E(\varepsilon_i) = \\ &\beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik} \end{aligned}$$

Det observerade värdet Y_i , med motsvarande värden X_{i1}, \dots, X_{ik} , är därmed en slumpvariabel med medelvärdet:

$$E(Y_i) = \beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik}$$

Regressionsmodellen kan i matrisform skrivas;

$$\begin{array}{ccccccc} \mathbf{Y} & = & \mathbf{X} & \mathbf{\beta} & + & \mathbf{\varepsilon} & \\ (n \times 1) & & (n \times (k+1)) & ((k+1) \times 1) & & (n \times 1) & \end{array}$$

Regressionsmodellens variation delas in i:

$$SS_T = \mathbf{y}'\mathbf{y} - \frac{\left(\sum_{i=1}^n y_i\right)^2}{n}, \quad SS_E = \mathbf{y}'\mathbf{y} - \mathbf{b}'\mathbf{X}'\mathbf{y} \quad \text{och} \quad SS_R = \mathbf{b}'\mathbf{X}'\mathbf{y} - \frac{\left(\sum_{i=1}^n y_i\right)^2}{n}.$$

3.1.1 Minsta-kvadratmetoden

För att skatta de okända parametervärdena $\beta_0, \beta_1, \dots, \beta_k$ används *minsta-kvadratmetoden*. Låt b vara en skattning av β . *Minsta-kvadratmetoden* innebär att det värde b som minimerar residualkvadratsumman

$$S(b) = \sum_{i=1}^n (y_i - b_0 - b_1 x_{i1} - \dots - b_k x_{ik})^2 = (\mathbf{y} - \mathbf{Xb})'(\mathbf{y} - \mathbf{Xb})$$

används som skattning av β . *MK-estimatorn* är konsistent om $p \lim N^{-1} \mathbf{X}'\mathbf{u} = 0$.

3.1.2 Val av förklarande variabler

För att undersöka om en eller flera regressionskoefficienter $\beta_0, \beta_1, \dots, \beta_k$ är signifikant skild från 0, används *t-test*, *F-test* eller *partiellt F-test*.

För att pröva en specifik regressionskoefficient, $H_0 : \beta_j = 0$, används *t-test*:

$$t_0 = \frac{b_j}{\sqrt{\hat{\sigma}^2 C_{jj}}}$$

där t_0 är $t_{(n-1)}$ under H_0 . $\hat{\sigma}^2 = \frac{SS_E}{n-k}$ och C_{jj} är det diagonalelement i $(\mathbf{X}'\mathbf{X})^{-1}$ som korresponderar med b_j .

Partiellt F-test mäter ökningen i MS_R då en eller flera ytterligare variabler X_j inkluderas i modellen. Antag att $H_0 : \beta_{j+1} = \beta_{j+2} \dots = \beta_k = 0$. Testvärdet ges då av:

$$F = \frac{MS_R(X_{j+1}, \dots, X_k | X_1, \dots, X_j)}{MSE}$$

där H_0 är $F \sim F_{(k-j, n-k)}$.

3.1.3 Residualgenskaper

I regressionsmodellen (3-1) antas ε_i vara oberoende normalfördelade med $E(\varepsilon_i) = 0$ och $Var(\varepsilon_i) = \sigma^2$ (konstant). Ett villkor för en adekvat regressionsmodell är att modellens residualer e_i har samma egenskaper som ε_i . För att undersöka om dessa villkor är uppfyllda används grafisk analys av residualplottar samt formella test som beskrivs i appendix.

4. Tillvägagångssätt

Mot bakgrund av att liknande studier saknas sedan tidigare så genomförs undersökningen utan kännedom om en potentiell eller förmodad standardmodell. Detta innebär att analysen av möjliga samband inleds förutsättningslöst och utan hänsyn till en särskild teori. Det finns i och med det ett stort antal variabler som kan tänkas ha betydelse i undersökningen, samtidigt som inga faktorer är givna att inkluderas på förhand. Avsikten är därför att till en början reducera antalet tänkbara förklarande variabler och samtidigt få ett inledande begrepp om vilka faktorer som kan antas ha betydelse för *antalet personbilar per invånare*.

För att åstadkomma detta genomförs inledningsvis *stegvis regression* med *biltäthet* som beroende variabel. Stegvis regression är en modellvalsprocedur som inkluderar eller exkluderar förklarande variabler enligt förutbestämda kriterier; *entry α* och *exit α* . Det första steget är att inkludera samtliga variabler. Steg två innebär att variabler vars pvärde vid partiellt F-test överstiger *exit α* , tas ur modellen. Då inga ytterliggare variabler kan tas bort prövas tidigare borttagna variabler mot *entry α* . När slutligen inga oberoende variabler kan läggas till eller tas bort från modellen, avslutas den stegvisa regressionen (Draper & Smith, 1981).

Vid *stegvis regression* är det viktigt att vara medveten om att ett stort antal oberoende hypotesprövningar innebär att den valda signifikansnivån α i realiteten blir betydligt högre. Den faktiska signifikansnivån kan beräknas med $1 - (1 - \alpha)^p$ där p anger antalet oberoende prövningar (Draper & Smith, 1981).

I undersökningen av *fordonstäthet* fungerar *stegvis regression* som utgångspunkt för framtagning av en specifik regressionsmodell. Med detta tillvägagångssätt kommer slutsatser dragna direkt från datamaterialet att skapa en ram för fortsatt analys. *Stegvis regression* rekommenderas i detta syfte av bl.a. Draper & Smith (1981) och Rawlings, Pantula och Dickey (1998).

De variabler som inledningsvis används i *stegvis regression* är i många avseenden identiska med de som används i Sveriges Kommuner och Landstings kommungruppsindelning. Kommungruppsindelningen klassificerar Sveriges kommuner i nio grupper efter strukturella egenskaper med avseende på befolkning, näringsliv och arbetsmarknad. Utöver de faktorer som ingår i kommungruppsindelningen inkluderas även åldersstruktur, regional förekomst av kollektivtrafik samt ekonomiska och sociala variabler. Definitioner och beskrivning av kommungrupperna redovisas i bilaga 1.

Översiktligt så kan de valda variablerna delas in enligt följande;

- *Befolkning* variabler som avser att redogöra för befolkning, befolkningstäthet samt åldersmässig fördelning inom respektive region. Den åldersmässiga indelningen görs i 10årsklasser samt åldersgruppen 65.

- *Arbetsmarknad* – variabler som beskriver arbetskraftens regionala indelning efter näringsgren. Uppdelningen av arbetskraft efter näringsgren görs i tio grupper som bygger på EU:s näringsgrensstandard SNI 2002.

- *Transport och trafik* – variabler som beskriver förekomsten av kollektivtrafik och pendling inom de olika regionerna. Eftersom det saknas uppgift om *utbud av kollektivtrafik i km per invånare* på kommunnivå, används motsvarande uppgift på länsnivå som proxyvariabel.

- *Ekonomiska och sociala variabler* – variabler som avser att återspegla ekonomisk styrka inom respektive region. De mått som används är *inkomst* och *arbetslöshet*.

Sammanlagt 26 förklarande variabler ingår i undersökningen. Beskrivning av samtliga variabler finns i bilaga 2.

Tabell 4.1. Beskrivande statistik för undersökningens samtliga variabler.

Variabel	Medel.	Std.	Min	Max	Q1	Median	Q3
BILPER1000	412,1	41,4	213,4	487,1	395,9	419,3	438,4
Befolkning							
INV	31425	60605	2541	782885	10006	15236	32507
INVKM2	129,0	433,8	0,2	4163,0	12,6	26,4	72,7
ALDER1	122,6	10,1	100,4	179,6	116,5	121,6	127,7
ALDER2	100,8	20,6	60,1	217,2	88,2	96,6	106,9
ALDER3	135,6	14,7	92,3	180,1	127,7	133,5	141,2
ALDER4	131,5	7,1	105,1	155,6	127,1	131,0	135,7
ALDER5	144,8	13,0	109,4	184,1	136,4	145,4	153,6
ALDER6	146,9	55,8	71,2	297,9	95,8	130,7	201,2
MDLALDER	42,4	2,3	36,2	47,8	40,9	42,6	44,0
SMAHUS	0,62	0,15	0,03	0,93	0,53	0,65	0,74
Arbetsmarknad							
SEKTOR1	17,9	11,5	0,0	57,0	9,0	15,5	26,0
SEKTOR2	112,1	68,9	4,0	458,0	61,5	98,5	147,3
SEKTOR3	4,9	5,5	0,0	50,0	2,0	4,0	6,0
SEKTOR4	32,5	10,1	12,0	97,0	26,0	32,0	37,0
SEKTOR5	80,4	38,1	33,0	415,0	57,8	72,0	92,0
SEKTOR6	31,7	14,5	14,0	134,0	23,0	28,0	36,0
SEKTOR7	40,6	30,1	10,0	268,0	23,0	33,0	46,0
SEKTOR8	22,3	15,2	1,0	101,0	12,0	17,0	27,0
SEKTOR9	56,4	11,2	35,0	125,0	50,0	55,0	61,0
SEKTOR10	85,3	28,9	28,0	254,0	66,0	82,0	99,0
Transport / trafik							
KOLLTRKM	67,9	22,3	36,0	125,0	53,0	63,0	74,0
LANDAREAL	1415	2461	8,72	19371	346	673	1305
UTPENDL	197,0	118,4	23,4	519,7	102,2	169,0	264,7
Ekon. och soc.							
INK	197,6	23,7	167,1	366,5	183,9	192,3	202,8
ARBLOS	3,55	1,19	1,10	7,30	2,70	3,40	4,30

5. Analys

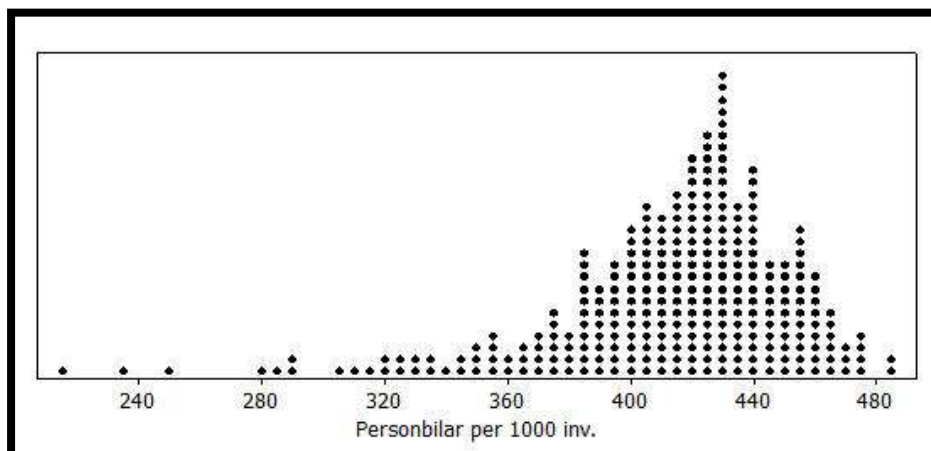
5.1 Regionala skillnader i fordonstäthet

Den nationella fordonstätheten med avseende på fordon ägda av fysiska personer och personliga företag uppgick vid årsskiftet 2006-2007 till 418 personbilar/1000 inv. enligt SIKA:s fordonsstatistik. Detta kan jämföras med motsvarande siffror för landets län som vid samma tidpunkt varierade mellan 281 och 449 personbilar/1000 inv. Det föreligger med andra ord stora skillnader mellan landets regioner.

SIKA:s fordonsstatistik visar att fordonstätheten är som lägst i storstadslänen. Stockholms län ligger långt under riksgenomsnittet med en fordonstäthet på 281 personbilar/1000 inv. Västra Götalands- och Skåne län ligger något högre med 367 respektive 378 personbilar/1000 inv. Högst fordonstäthet noteras i Gotlands län med 449 personbilar/1000 inv. I övrigt kan hög fordonstäthet observeras i framförallt landets norra och centrala län.

Med en regional upplösning på kommunnivå framträder skillnaderna mer tydligt. Den regionala variationen kan som tidigare indikerats i stor utsträckning tillskrivas storstadskommunerna och kringliggande storstadsområden. Statistiken visar att fordonstätheten i Stockholmsområdet och framförallt Stockholms stad, med 218 personbilar/1000 inv. samt Solna- och Sundbybergs kommun avviker kraftigt från det nationella genomsnittet. Förutom dessa kan ett litet antal kommuner i Skåne, Storstockholm samt Göteborg betraktas som låga extremvärden.

Diagram 5.1. Fördelningen för personbilstäthet efter kommun vid årsskiftet 2006-2007.



För att undersöka om personbilstäthet skiljer sig beroende på kommungruppering används ensidig *variansanalystabell*. Resultatet indikerar tydliga skillnader. Fordonstätheten visar sig vara signifikant lägre i kommungrupperna *storstad*, *förortskommun* och *större stad*. Jämförelser mellan övriga kommungrupper visar endast enstaka avvikelser. Variansanalystabellen redovisas i bilaga 3.

Resultatet indikerar att variationen i personbilstäthet generellt överensstämmer med regionala skillnader i befolkningstäthet. Storstadsområden och områden med hög befolkningskoncentration uppvisar i allmänhet låg personbilstäthet. När det gäller landets övriga regioner och kommuner är det emellertid svårt att urskilja allmänna skillnader. Skillnaderna i fordonstäthet är jämförelsevis små och signifikant avvikande observationer saknas.

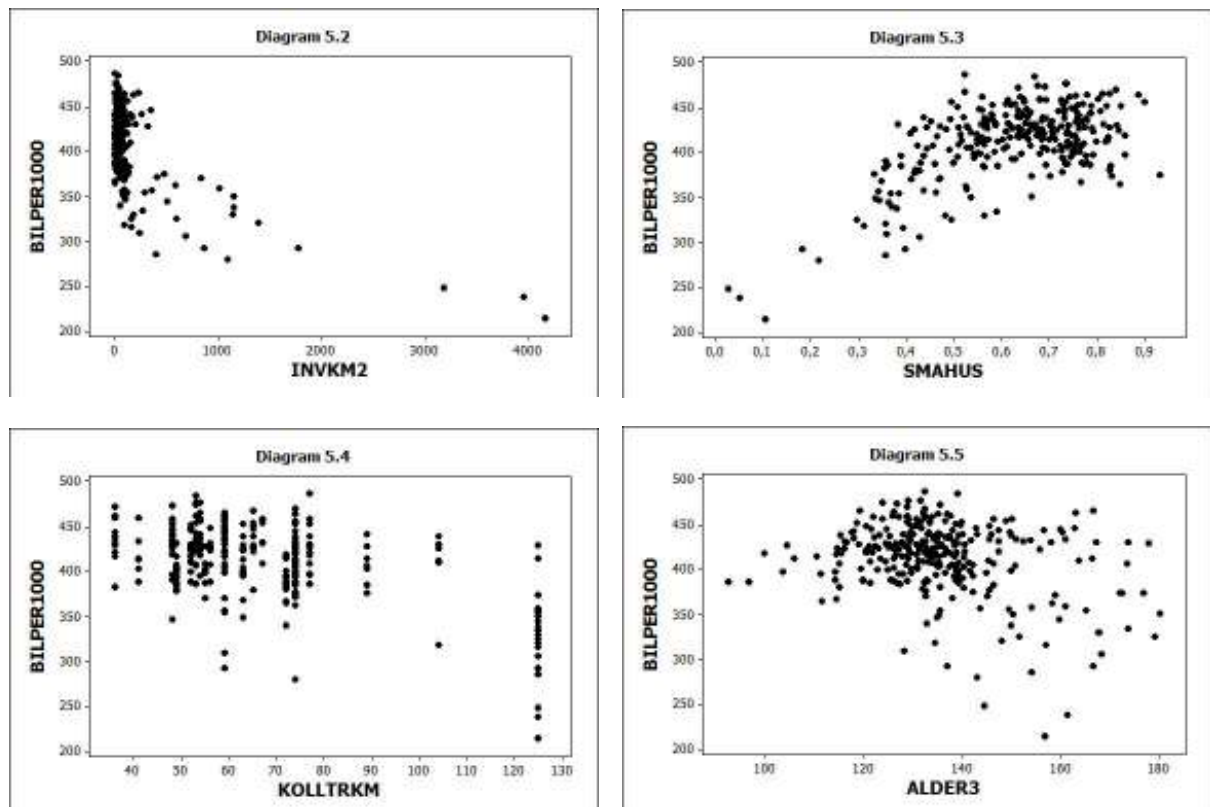
5.2 Modellbestämning

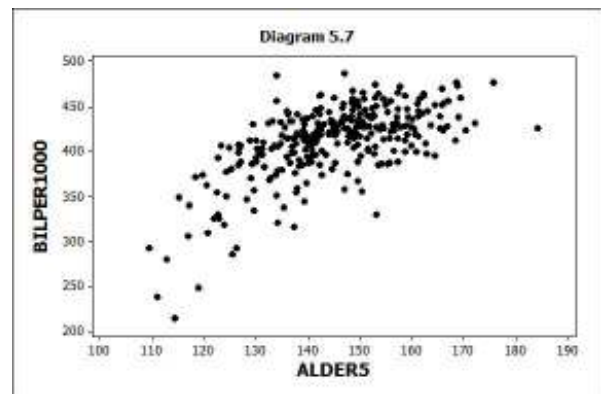
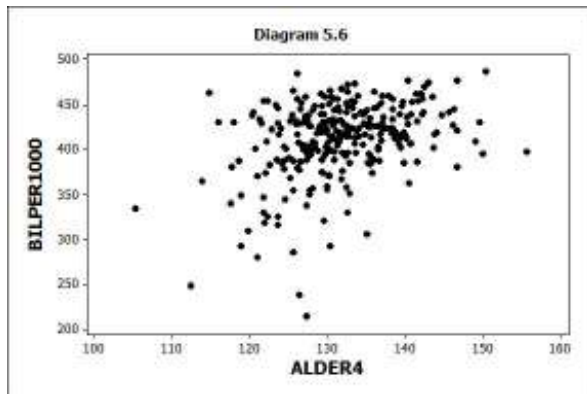
Inledningsvis analyseras korrelationsmatrisen för den beroende och samtliga oberoende variabler. Undersökningen visar att samtliga förklarande variabler har signifikant korrelation med *BILPER1000*. Framförallt variablerna *INVKM2*, *KOLLTRKM* och *ALDER2* visar mycket hög korrelation med den beroende variabeln. I övrigt förekommer viss korrelation mellan ett antal förklarande variabler, dock inte i sådan utsträckning att variabler behöver exkluderas.

Undersökningen fortsätter därefter med *stegvis regression*, där samtliga variabler inkluderas i en inledande modell (bilaga 4). För att reducera effekterna av masssignifikans används låga $\alpha = 0,001$ *to-enter* och $\alpha = 0,05$ *to-remove*. Regressionen resulterar i en modell med de förklarande variablerna *INVKM2*, *MDLALDER*, *SMAHUS*, *KOLLTRKM*, *SEKTOR1*, *ALDER2*, *ALDER3*, *ALDER4* samt *ALDER5*. Modellen ger signifikanta parametrar och relativt hög förklaringsgrad, $\bar{R}^2 = 75,6$. Dock så visar VIF-värdena tecken på kraftig multikollinearitet. Utöver detta visar även *lack-of-fit*-test att modellen är inadekvat anpassad. Vi fortsätter därför med att successivt ta bort de variabler som visar tydligast tecken på multikollinearitet.

En ny reducerad modell, utan signifikant multikollinearitet, innehåller variablerna *INVKM2*, *SMAHUS*, *KOLLTRKM*, *ALDER3*, *ALDER4*, *ALDER5*. Modellen ger som tidigare signifikanta parametrar och god förklaringsgrad, dock så kvarstår problemen med dålig anpassning. Sambanden mellan *BILPER1000* och vissa förklarande variabler karakteriseras av kurvatur. Diagrammen 5.2-5.7 illustrerar icke-linjära samband mellan den beroende- och förklarande variabler.

Diagram 5.2-5.7. BILPER1000 mot variablerna INVKM2, SMAHUS, KOLLTRKM, ALDER3, ALDER4, ALDER5.





Nästa åtgärd blir därför att undersöka en polynom regressionsmodell. Nya variabler i form av andragradspolynom och interaktionstermer skapas successivt för de faktorer som karakteriseras av icke-linjärt samband med *BILPER1000*. Anpassningen till polynom regression leder till nya modeller där *INVKM2* och *ALDER4* blir icke-signifikanta och därmed exkluderas. För återstående variabler prövas tredjegrads polynom, vilket emellertid inte nämnvärt förbättrar modellens förklaringsgrad. Andragradspolynom förefaller därför tillräckligt för att förklara variationen i *BILPER1000*.

Undersökningen leder oss till en slutlig modell enligt:

$$BILPER1000 = 324 + 0,762 ALDER3 + 1,45 ALDER5 - 0,432 KOLLTRKM + 32,6 SMAHUS - 0,0245 (ALDER5)^2 + 0,0121 (KOLLTRKM)^2 - 229,04 (SMAHUS)^2 + \varepsilon$$

5.3 Utvärdering av den valda modellen

Den slutliga modellen visar god **förklaringsgrad**, $\bar{R}^2 = 74,6$. Samtliga parameterskattningar är signifikanta och har förväntat tecken.

Tabell 5.1. Skattade parametervärden för den valda modellen.

n = 290

Variabel	Parameter-skattning	SE Koeff.	t-värde	Pr (T>t)	VIF
Intercept	324,070000	15,4400	20,99	0,0000	-
ALDER3	0,762300	0,1158	6,58	0,0000	1,9
ALDER5*	1,452800	0,1316	11,04	0,0000	1,9
ALDER5*2	-0,024523	0,0062	-3,98	0,0000	1,3
KOLLTRKM*	-0,432450	0,0833	-5,19	0,0000	2,3
KOLLTRKM*2	-0,012059	0,0023	-5,25	0,0000	2,9
SMAHUS*	32,610000	10,7500	3,03	0,0030	1,8
SMAHUS*2	-229,040000	40,5700	-5,65	0,0000	1,7

Variabler markerade * har medelvärdescentrerats för att eliminera multikollinearitet.

Variabler markerade *2 avser andragradstermer.

Modellen undersöks med avseende på residualgenskaper. Plottar över residualer e mot skattade värden \hat{y} , samt residualer mot förklarande variabler X_j redovisas i diagram 5.7-5.14. Diagram 5.7 indikerar att residualvariansen är oberoende av nivån på skattade värden. Inte heller plottarna över e mot förklarande variabler X_j visar tecken på beroende; residualernas variation tycks i samtliga fall vara konstant och utan underliggande struktur. Slutligen så

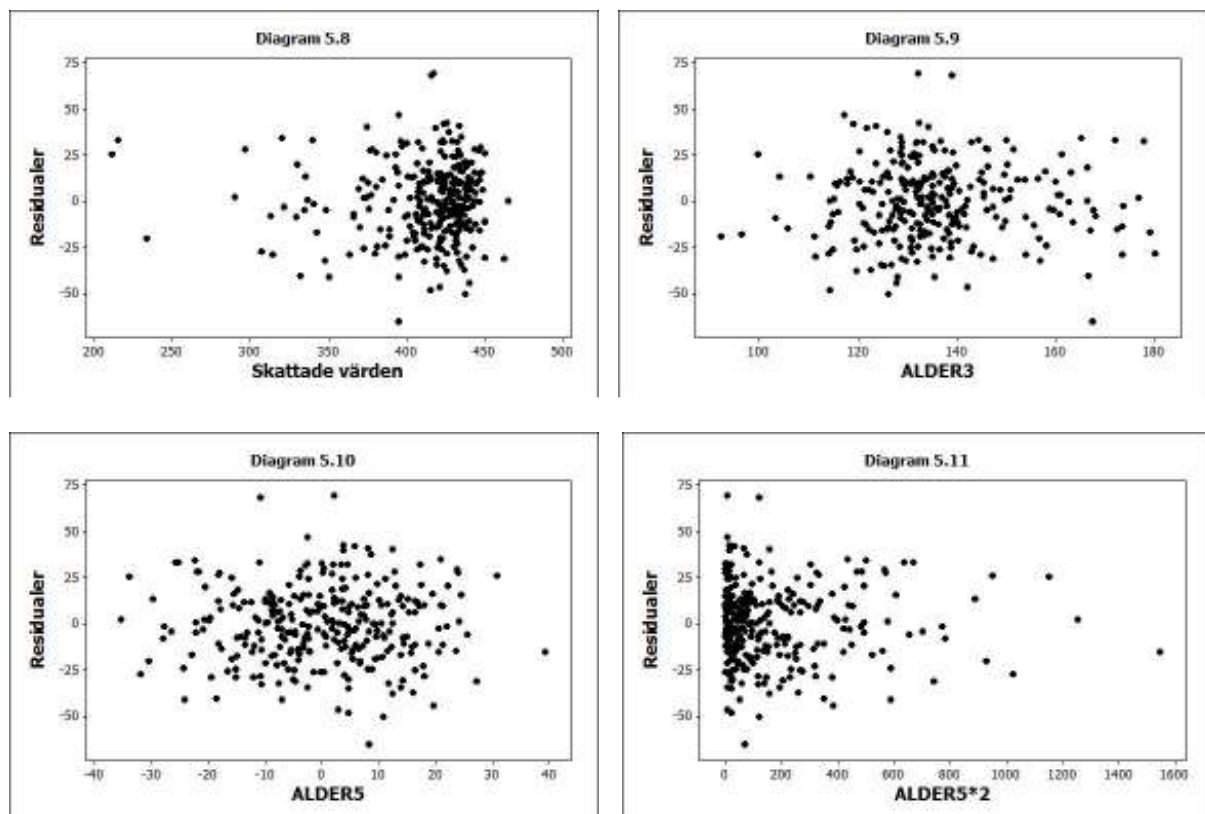
kontrolleras **heteroskedasticitet** med *Whites test*. Testet visar sig icke-signifikant och styrker indikationerna på att modellen inte lider av heteroskedasticitet.

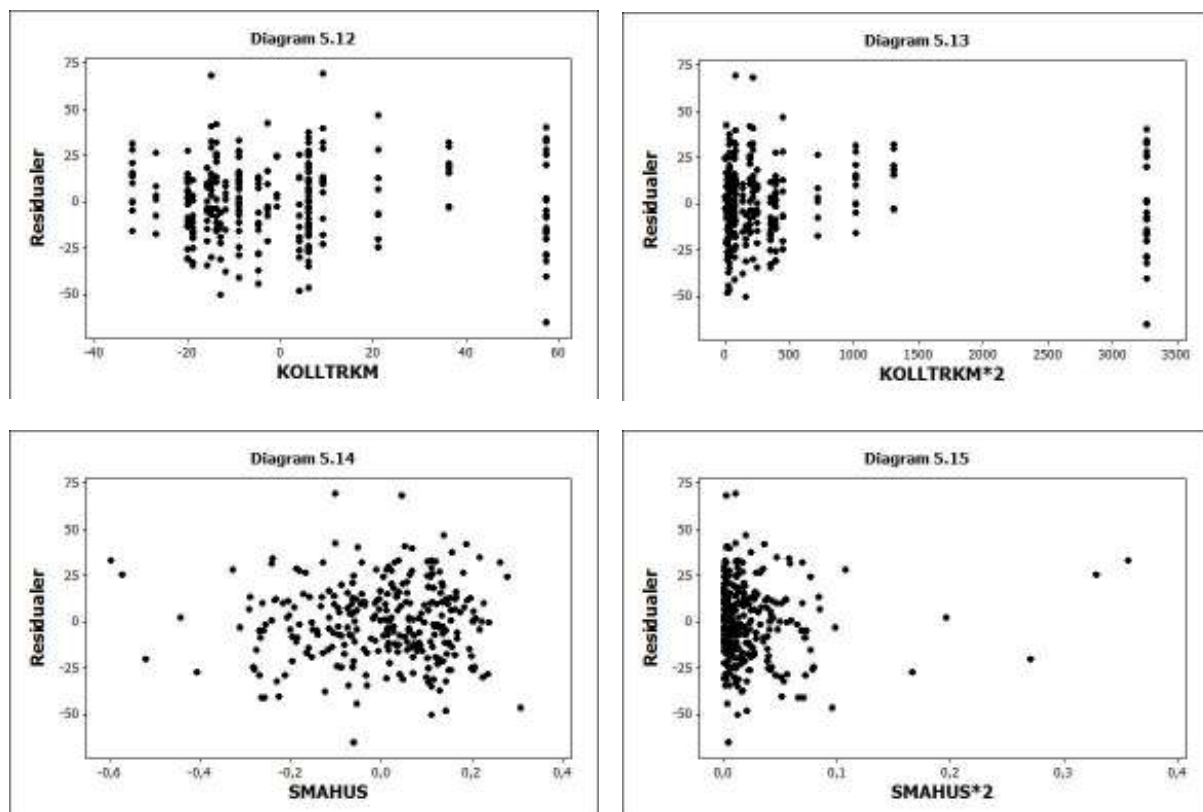
Residualerna kontrolleras även med avseende på **inflytelserika värden**. En undersökning med *Cooks D*, visar att de tre storstadskommunerna samt Solna, Sundbyberg och Valdemarsvik har stort inflytande på skattningen av β . Detta är inte oväntat, då samtliga dessa observationer, förutom Valdemarsvik, är tydliga extremvärden med avseende på fordonstäthet. Inflytandet kan såtillvida uppfattas som snarare följderna av naturlig variation, än anpassningsproblem i modellen. För att ytterligare undersöka hur modellen påverkas av **inflytelserika värden** kontrolleras R^2_{PRED} . Värdet $R^2_{PRED} = 73,5$ ligger nära $\bar{R}^2 = 74,6$ och tyder på att modellens prediktiva kapacitet är god (Montgomery, 1976). Sammantaget så tycks inte **inflytelserika värden** indikera fel i modellen.

Sammanfattningsvis kan modellen betraktas som adekvat. Modellen har god förklaringsgrad och residualerna är oberoende och lika fördelade. Vidare så har tidigare problem med anpassningsgraden försvunnit. Test för **lack-of-fit** ger $p > 0,1$ och indikerar därmed god anpassning. Viss **multikollinearitet** förekommer, dock inte i sådan grad att det kan betraktas som ett problem. Medelvärde för *VIF* är 1,97, med ett högsta värde på 2,9.

Modellens fullständiga diagnostik redovisas i bilaga 5.

Diagram 5.8-5.15. Residualer mot skattade y-värden och samtliga förklarande variabler.





6. Resultat

Regressionsanalysens resultat indikerar att regional åldersstruktur har signifikant betydelse för *antalet personbilar per invånare*. Den inledande korrelationsmatrisen visar att samtliga åldersgrupper förutom 65- har signifikant korrelation med *antalet personbilar per invånare*. Undersökningen visar vidare att kommuner med hög befolkningsandel i åldersintervallen 45-54 och 55-64 tenderar att ha ett större antal personbilar per invånare. Övriga åldersgrupper uppvisar ett negativt samband med den beroende variabeln. I den slutliga modellen inkluderas endast åldersgrupperna 35-44 och 55-64, båda med positiv inverkan på *antalet personbilar per invånare*.

Vidare så kan konstateras att *utbud av kollektivtrafik i km per invånare* har en signifikant negativ inverkan på *antalet personbilar per invånare*. Detta framstår som ett väntat resultat, då en väl utbyggd kollektivtrafik rimligtvis bör minska behovet av personbilar. Samtidigt finns en mycket hög positiv korrelation mellan *utbud av kollektivtrafik* och de faktorer som kännetecknar områden med hög befolkningstäthet. Det är därför troligt att effekterna av *kollektivtrafik i km per invånare* i viss mån svarar för effekterna av andra utelämnade faktorer, såsom exempelvis inkomst, arbetsmarknadsstruktur och tätortsgrad.

Modellen visar vidare att *andelen småhus av det totala bostadsbeståndet* har en signifikant positiv effekt på *antalet personbilar per invånare*. Detta överensstämmer med tidigare indikationer på att hög befolkningstäthet sammanfaller med låg personbiltäthet. En stor andel småhus kan i sammanhanget generellt betraktas som en indikator på kommuner med låg befolkningstäthet.

Sammanfattningsvis förefaller undersökningens resultat att bekräfta ett samband mellan befolkningstäthet och personbiltäthet. För att återgå till kommungruppsindelningen så är åldersstrukturen i *storstäder* och *större städer* generellt olik den i övriga kommungrupper. Andelen invånare i de övre åldersintervallen är allmänt högre i mindre tätbefolkade

kommuner. Åldersstruktur förefaller därför att till viss del förklara den väsentligt lägre fordonstätheten i *storstäder* och *större städer*. Tillsammans med *utbud av kollektivtrafik* och *andelen småhus* pekar detta på en regional variation i *fordonstäthet* som i stor utsträckning kan relateras till strukturella skillnader mellan mer tätbefolkade regioner och övriga landet.

Trots att modellen innehåller ett begränsat antal variabler kan det naturligtvis antas att betydligt fler faktorer inverkar i olika grad. Vidare, som en följd av korrelation mellan förklarande variabler kan effekten av en enskild faktor inte isoleras utan måste uppfattas i relation till övriga faktorer. De valda variablerna utgör emellertid tillsammans den modell som mest adekvat beskriver variationen i antalet personbilar per invånare.

7. Metod

I undersökningen av *regionala skillnader i antal nyregistreringar* används *paneldata*; data som varierar över både individer och tid. Detta avsnitt avser att översiktligt redogöra för de metoder som används.

7.1 Modeller för paneldata

En allmän regressionsmodell för paneldata kan skrivas

$$(7-1) \quad y_{it} = x_{it}'\beta_{it} + z_i'\alpha + \varepsilon_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T$$

där

y_{it} är det observerade värdet på den beroende variabeln;

x_{it} är observerade värden på K oberoende variabler;

$z_i'\alpha$ är *heterogenitet* eller *individspecifik effekt*, där z_i anger en grupp- eller individspecifik egenskap som antas vara konstant över tid;

ε_{it} är en slumpmässig felterm; och

i indexerar individ i tvärsnittsdata och t indexerar tid.

En utökad *tvåvägsmodell* inkluderar parametern γ_t , som anger tidsspecifik effekt.

Beroende på olika antaganden om *heterogenitet*, kan specifika varianter av (7-1) formuleras på en rad sätt. Vidare kan parametrarna α och β skattas med olika metoder beroende på modell. Nedan redovisas de modeller och metoder som tas upp i undersökningen.

Poolad regression: Om z_i endast innehåller en konstant term, kan modellen beräknas som en *vanlig linjär regressionsmodell*. Interceptet α och lutningen β skattas då konsistent och effektivt med *minsta-kvadratmetoden* av

$$y_{it} = \alpha + x_{it}'\beta + \varepsilon_{it}$$

Modell med fixa effekter (FE): Om z_i är oobserverad, men korrelerad med x_{it} , är *MK-estimatoren inkonsistent* och lider av *utelämnad-variabel bias*. Av hänsyn till *individuell heterogenitet*, används istället en modell med *fixa effekter* i form av

$$(7-2) \quad y_{it} = x_{it}'\beta + \alpha_i + \varepsilon_{it}.$$

Den individspecifika effekten $\alpha_i = z_i'\alpha$, $i = 1, \dots, N$ mäter *tidsinvariant observerad heterogenitet* och ger FE-modellen ett skattningsbart villkorligt medelvärde.

Vid FE-modell kan β skattas konsistent med *inom- eller första differens-estimatoren*. *Inom-estimatoren* beräknas med MK-skattning av

$$y_{it} - \bar{y}_i = (x_{it} - \bar{x}_i)'\beta + (\varepsilon_{it} - \bar{\varepsilon}_i), \quad i = 1, \dots, N, \quad t = 1, \dots, T$$

Inom-estimatoren är *effektiv* vid FE-modell.

Modell med slumpmässiga effekter (RE): Om *oobserverad individuell heterogenitet* antas vara *okorrelerad* med x_{it} , kan modellen formuleras

$$(7-3) \quad \begin{aligned} y_{it} &= x'_{it}\beta + E[z'_i\alpha] + \{z'_i - E[z'_i\alpha]\} + \varepsilon_{it} \\ &= x'_{it}\beta + \alpha + u_i + \varepsilon_{it}. \end{aligned}$$

RE-modellen implicerar att u_i är en individspecifik slumpterm, som är snarlikt ε_{it} , men *identisk* för varje tidsperiod t .

Vid RE-modell kan β skattas konsistent med samtliga ovanstående metoder, samt *mellan-* och RE-estimatorn. RE-estimatorn beräknas med MK-skattning av

$$\begin{aligned} y_{it} - \hat{\lambda}\bar{y}_i &= (1 - \hat{\lambda})\mu + (x_{it} - \hat{\lambda}\bar{x}_i)' \beta + v_{it} \\ \text{där } v_{it} &= (1 - \hat{\lambda})\alpha_i + (\varepsilon_{it} - \hat{\lambda}\varepsilon_i) \text{ är asymptotiskt iid och } \hat{\lambda} \text{ är konsistent för} \\ \lambda &= 1 - \frac{\sigma_\varepsilon}{\sqrt{\sigma_\varepsilon^2 + T\sigma_\alpha^2}}. \end{aligned}$$

RE-estimatorn är *effektiv* vid RE-modell.

Tabell 7.1. Konsistens vid olika estimatorer och modeller. Figur hämtad från Cameron & Trivedi (2006).

Estimator av β	Modell		
	Poolad	RE	FE
Poolad MK	Konsistent	Konsistent	Inkonsistent
FE	Konsistent	Konsistent	Konsistent
Första differens	Konsistent	Konsistent	Konsistent
RE	Konsistent	Konsistent	Inkonsistent

7.2 Test för olika modeller

I undersökningen används två metoder för pröva *fixa-* respektive *slumpmässiga effekter*. För att testa *fixa effekter*; låt β_f vara en n -dimensionell vektor av *fixa effekt-parametrar*.

Nollhypotesen, $\beta_f = 0$, är $F_{(n, M-K)}$ och testvärdet ges av

$$\hat{\beta}_f \hat{S}_f^{-1} \hat{\beta}_f / n, \text{ där } \hat{S}_f \text{ är den skattade kovariansmatrisen för } \beta_f.$$

För att pröva nollhypotesen *slumpmässiga effekter* mot alternativet *fixa effekter*, används *Hausman-test* (Greene, 1993). Antag två estimatorer; $\hat{\beta}_a$ och $\hat{\beta}_b$. Under nollhypotesen är båda konsistenta, men bara $\hat{\beta}_a$ är asymptotiskt effektiv. Under mothypotesen är bara $\hat{\beta}_b$ konsistent.

Testvärdet m är χ^2_k , där k är antalet dimensioner i $\hat{\beta}_a$ och $\hat{\beta}_b$ och

$$m = (\hat{\beta}_b - \hat{\beta}_a)' (\hat{S}_b - \hat{S}_a) (\hat{\beta}_b - \hat{\beta}_a), \text{ där } \hat{S}_b \text{ och } \hat{S}_a \text{ är konsistenta estimatorer av kovariansmatriserna för } \hat{\beta}_b \text{ och } \hat{\beta}_a.$$

8. Tillvägagångssätt

Avsikten med undersökningen är att belysa vilka faktorer som påverkar *antalet nyregistrerade personbilar ägda av fysiska personer eller personliga företag*. För att åstadkomma detta utformas en *regressionsmodell för paneldata*. Som beroende variabel används *antal nyregistreringar per 1000 inv efter kommun*, med årsvisa observationer för tidsperioden 1997-2006. Den beroende variabeln består av totalt 2891 observationer.

En fördel med modeller för *paneldata*, i jämförelse med *tvärsnittsdata*, är ökad precision i skattningarna. Genom att datamaterialet sträcker sig över tid, ökar antalet observationer. Detta kan utnyttjas, antingen genom att observationer över tid för en individ ”*poolas*”, eller genom modeller som tar hänsyn till *observerad individuell heterogenitet*.

Med detta tillvägagångssätt kan *antal nyregistreringar* prövas mot olika regressorer, samtidigt som modellen kontrollerar för eventuella kommun- och tidsspecifika effekter. Vidare så skapas möjlighet att undersöka enskilda kommuners beteende över tid. Genom att undersökningen inkluderar observerade värden för en tioårsperiod, uppnås hög precision i skattningarna.

De förklarande variabler som används, avser att återspegla ett spektrum av samhällsområden. Översiktligt kan de valda variablerna delas in enligt följande:

- *Befolkning* – variabeln *INVKM2* beskriver befolkningstäthet för respektive kommun. *ALDER1-ALDER6* redogör för den åldersmässiga fördelningen i 10-årsklasser samt åldersgruppen 65-.
- *Transport och trafik* – variabeln *UTPENDL* beskriver *antal utpendlare per 1000 inv.* för respektive kommun. Variabeln *BEST* redogör för *kommunalt fordonsbestånd per 1000 inv.* Med denna variabel kontrolleras sambandet mellan befintligt fordonsbestånd och nya fordonsregistreringar.
- *Ekonomiska och sociala variabler* – variabler som avser att återspegla ekonomisk styrka och ekonomiska förändringar över tid. Tidsserien för *antal nyregistrerade personbilar* är en vanlig konjunkturindikator i ekonomiska sammanhang och används ofta som ett allmänt mått på hushållens och företagens köpkraft och framtidstro. Variabler som beskriver ekonomiska förändringar över tid, kan därför antas ha central betydelse i undersökningen. De mått som används är *medelinkomst(INK)* och *antal förvärvsarbetande per 1000 invånare(ARB)*.

Samtliga förklarande variabler består av observationer för tidsperioden 1997-2006. Variablerna redovisas i bilaga 6.

9. Regionala skillnader i antal nyregistrerade fordon

SIKA:s fordonsstatistik för 2006 visar att antalet nyregistreringar uppgick till 13,6 personbilar per 1000 inv., sett över hela landet. Detta är en marginell minskning sedan året innan och 2004. Sett över en tioårsperiod har antalet nyregistreringar ökat något från genomsnittet 12,6 för 1997. Tidsserien för de senaste 10 åren visar en topp under 1999 och 2000, för att därefter avta och stabiliseras.

Flest nyregistreringar per inv. under 2006 noterades i kommunerna Vellinge, Nykvarn och Svedala, samtliga med fler än 23 nyregistrerade personbilar per 1000 inv. Minst antal nyregistreringar observerades i Dorotea och Sorsele, med 3,1 respektive 3,6 per 1000 inv.

För perioden 1997-2006 noteras högst genomsnittligt antal nyregisteringar i Vellinge, Strömstad och Olofström, samt ytterliggare 9 kommuner; alla med ett medelvärde på över 20 per 1000 inv. Vellinge är vidare fördelningens enda återkommande extremvärde. För samma tidsperiod observeras de lägsta medelvärdena i Sorsele, Åsele och Vilhelmina.

Tabell 9.1. Beskrivande statistik för antal nyregisteringar per 1000 inv. efter kommun 1997-2006.

Variabel	Medel	Std.	Min	Max	Q1	Median	Q3
NYREG97	12,6	3,4	5,6	32,4	10,5	12,7	14,6
NYREG98	13,8	3,5	6,4	28,3	11,5	13,7	16,2
NYREG99	17,1	4,3	3,7	33,0	14,1	17,1	19,6
NYREG00	18,0	4,6	5,6	34,6	15,3	18,0	21,3
NYREG01	12,9	3,6	3,4	26,4	10,4	13,0	15,6
NYREG02	13,2	3,7	2,6	25,7	10,6	13,5	15,8
NYREG03	14,1	4,1	3,8	29,8	11,4	14,7	16,9
NYREG04	14,5	4,1	3,0	27,3	11,6	14,9	17,4
NYREG05	14,0	4,0	3,9	27,0	11,4	14,5	16,7
NYREG06	13,6	3,9	3,1	28,4	11,0	13,8	16,0

Fördelningen för 2006, uppdelat efter kommungrupp, visar att gruppen *förortskommun* har högst medelvärde med 16,8 nyregisteringar per 1000 inv. Lägst medelvärde noteras för *glesbygdskommuner* med 8,2 per 1000 inv. I övrigt förekommer inga markanta avvikelser. Fördelningen är i princip oförändrad under hela tidsperioden 1997-2006 (bilaga 7).

För att undersöka om antalet nyregisteringar skiljer sig beroende på kommungruppering, används *ensidig variansanalystabell*. Resultatet visar att medelvärdet för gruppen *glesbygdskommuner* är signifikant lägre i jämförelse med övriga kommungrupper. Gruppen *förortskommuner* har högre medelvärde än flertalet grupper. Förutom dessa förekommer inga signifikanta skillnader. Vidare så är utvecklingen över tid snarlik för samtliga kommungrupper.

10. Analys

10.1 Modellbestämning

Inledningvis prövas en *poolad regressionsmodell* som innehåller samtliga förklarande variabler. Resultatet visar att variablerna *ALDER3*, *ALDER5*, *ALDER6* och *ARB* är icke-signifikanta, med $\alpha = 0,01$. Förklaringsgraden är måttlig, $\bar{R}^2 = 0,288$. Vidare så visar modellen tecken på kraftig multikollinearitet och dålig anpassning. Efter att ha reducerat modellen stegvis, erhålls en slutlig *poolad regressionsmodell* där de förklarande variablerna *INK*, *BEST*, *ALDER1*, *ALDER2*, *ALDER5* och *ALDER6* ingår. Förklaringsgraden är i princip oförändrad jämfört med tidigare, och modellen visar endast svaga tecken på multikollinearitet. Dock så visar *Whites test* att modellen lider av heteroskedasticitet. Diagnostik för modellen samt skattade parametrar redovisas i bilaga 8.

Efter detta undersöks en *tvåvägsmodell med slumpmässiga effekter*, med samtliga förklarande variabler. Resultatet visar att alla variabler, förutom *INVKM2* och *ALDER4-ALDER6* är signifikanta. Nollhypotesen för *Hausman-testet* förkastas emellertid och skattning med

slumpmässiga effekter och *poolad-MK estimator* visar sig därmed inkonsistent (bilaga 9). Undersökningen fortsätter därför med att pröva en modell med *fixa effekter*.

En *tvåvägsmodell med fixa effekter* ger samma resultat som tidigare med avseende på skattade parametrar; *INVKM2* samt *ALDER4-ALDER6* visar sig icke-signifikanta (bilaga 10). För att pröva H_0 : *inga fixa effekter* används F-test, som ger $p < 0,0001$.

Efter stegvis eliminering av icke-signifikanta variabler, ges en slutlig modell med *fixa individ- och tidsspecifika effekter* av:

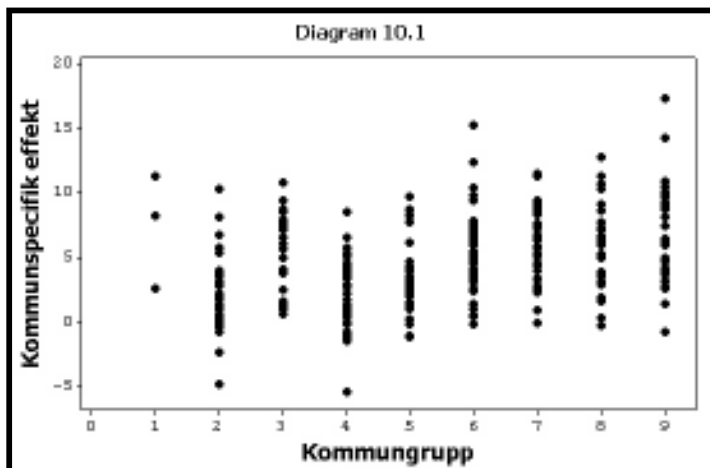
$$NYREG = 0,021881 INCOME + 0,019776 ARB + 0,03691 BEST + 0,014524 UTPENDL + 0,038 ALDER1 + 0,053666 ALDER2 + 0,057101 ALDER3 + \alpha_i + \gamma_t + \varepsilon_{it}$$

Modellen ger en förklaringsgrad på $R^2_{within} = 0,5833$. Värdet \bar{R}^2 hamnar på 0,8846 och inkluderar då även *individ- och tidsspecifika effekter* i beräkningen av förklarad variation.

Modellens *kommunspezifika effekter* visar stor spridning. Högst positiv effekt uppvisar kommunerna Strömstad och Olofström, med 17,4 och 15,3 jämfört med referenskommunen Upplands-Väsby. Kommunerna Knivsta och Sundbyberg har mest negativ effekt med -5,4 respektive -4,8. En plott över kommunspezifisk effekt mot kommungrupp redovisas i diagram 10.1. Diagrammet indikerar inga tydliga skillnader; inflytandet av kommunspezifisk effekt tycks inte vara beroende av kommungruppsstillhörighet.

När det gäller *tidsspezifisk effekt* har koefficienten för samtliga år positivt inflytande på *antal nyregistreringar*. Koefficienterna för åren 2001 och 2002 ligger emellertid nära 0. Störst tidsspezifisk effekt uppvisar åren 1999 och 2000 med ett positivt inflytande på 3,14 respektive 3,35, jämfört med referensåret 1997.

Diagram 10.1. Kommunspezifisk mot kommungruppsstillhörighet.



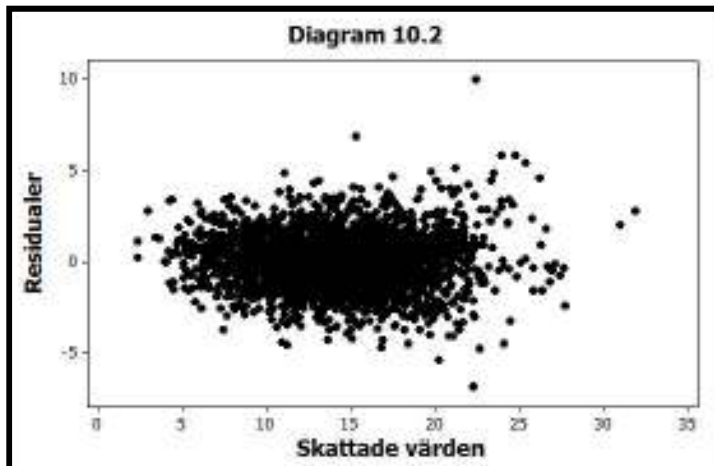
10.2 Modellutvärdering

Ett vanligt problem i *paneldatamodeller* är att residualerna lider av **1) seriell korrelation** (dvs. korrelation över t för givet i) och **2) heteroskedasticitet**. Vid skattning med *OLS-estimatorn* leder dessa problem till kraftig *underskattning av vanliga standardfel*, och som följd *överskattning av de skattade parametrarnas t-värden*. För giltig inferens krävs korrigeringsmetoder. För *korta paneler* kan problemen åtgärdas med *Whites korrektion* och *heteroskedasticitets-robusta standardfel*. För *längre paneler* krävs *panelrobusta standardfel*, som justerar för både

heteroskedasticitet och seriell korrelation. Individspecifika effekter kan till viss del, dock inte helt, reducera problemen med *seriell korrelation* (Cameron & Trivedi, 2007).

Modellen undersöks inledningsvis med avseende på *heteroskedasticitet*. Diagram 10.2 visar *modellens residualer mot skattade värden*. Diagrammet visar tydligt på icke-konstant varians; spridningen i residualerna ökar för stora värden på \hat{y} . Modellen tycks därmed lida av *heteroskedasticitet*. Detta bekräftas av ett modifierat Wald-test för *heteroskedasticitet*, som med H_0 : konstant varians över i , ger $p < 0,0001$ (Baum, 2001).

Diagram 10.2. Modellens residualer mot skattade värden.



Seriell korrelation undersöks genom *Wooldridges test för autokorrelation*, med H_0 : ingen autokorrelation av första ordningen (Wooldridge 2002). Testet visar sig kraftigt signifikant och ger $p < 0,0001$.

Resultaten visar att modellen lider av både *heteroskedasticitet* och *seriell korrelation*. Korrektur av standardfelen är därmed nödvändig. Tabell 10.1 visar *ojusterade standardfel*, *standardfel som justerats för heteroskedasticitet* samt *standardfel som justerats för både heteroskedasticitet och seriell korrelation*. De *ojusterade-* och *heteroskedasticitets-robusta standardfelen* visar sig relativt lika, medan de *panelrobusta standardfelen* ger betydligt större värden. Detta reflekteras i de skattade koefficienternas t-värden, där variablerna *UTPENDL* och *INK* har $p > 0,01$ respektive $p > 0,1$, med *panelrobusta standardfel*.

Sammantaget är det tydligt att slutsatser bör dras efter *panelrobust korrigering* av modellen. Underskattning av standardfelen innebär att variabeln *INK* felaktigt förefaller signifikant. Även signifikansen för variabeln *UTPENDL* visar sig något överskattad vid användning av ojusterade standardfel.

I övrigt förefaller modellen adekvat. Samtliga koefficienter har förväntat tecken och förklaringsgraden på $R^2_{within} = 0,5833$ respektive $\bar{R}^2 = 0,8846$, är relativt god. Diagnostik för modellen redovisas i bilaga 11.

Tabell 10.1. Skattade koefficienter och vanliga-, robusta- samt panelrobusta standardfel.

n=2891

Variabel	Koeff. skattn.	Vanliga SE	t	Pr(t>T)	Robusta SE	T	Pr(t>T)	Panel SE	t	Pr(t>T)
INK	0,02188	0,00911	2,40	0,0164	0,00979	2,24	0,025	0,01450	1,51	0,1312
ARB	0,01978	0,00383	5,17	<,0001	0,00402	4,92	<,0001	0,00554	3,57	0,0004
BEST	0,03691	0,00408	9,06	<,0001	0,00617	5,98	<,0001	0,01210	3,05	0,0023
UTPENDL	0,01452	0,00422	3,44	0,0006	0,00464	3,13	0,002	0,00610	2,38	0,0174
ALDER1	0,03800	0,01010	3,77	0,0002	0,01037	3,66	0,0003	0,01304	2,91	0,0036
ALDER2	0,05367	0,00915	5,86	<,0001	0,01041	5,16	<,0001	0,01613	3,33	0,0009
ALDER3	0,05710	0,00884	6,46	<,0001	0,01017	5,61	<,0001	0,01641	3,48	0,0005

11. Resultat

Undersökningens resultat visar att regional ålderstruktur har signifikant betydelse för antalet nyregisterade fordon. Resultaten visar att kommuner med en hög andel invånare i åldersintervallet 15-44 tenderar att ha ett större antal nyregistreringar.

När det gäller ekonomiska variabler antyder de inledande resultaten ett positivt samband mellan nyregistreringar och genomsnittlig förvärvsinkomst. Efter justering av resultaten kan något klart samband emellertid inte påvisas. Undersökningen visar däremot ett tydligt samband mellan den beroende variabeln och andelen förvärvsarbetande. Sammantaget så förefaller det tydligt att antalet nyregistreringar, inte oväntat, kan kopplas till ekonomisk styrka. Detta återspeglas även i modellens tidsspecifika effekter, som under högkonjunkturen 1999-2000 ger påfallande höga värden.

Även då det gäller transport- och trafikvariabler förekommer samband med den beroende variabeln. Undersökningen visar på att befintligt fordonsbestånd har ett klart positivt inflytande på antalet nyregisterade fordon. Detta kan tolkas som att ett stort fordonsbestånd genererar ett större antal nyregistreringar. Ett mer osäkert samband förekommer mellan andelen utpendlare och antalet nyregistreringar.

Sammanfattningsvis tycks regionala skillnader i antalet nyregisterade fordon överensstämma med strukturella skillnader med avseende på befolkningsstruktur, ekonomisk styrka och geografiskt läge. Som tidigare framgått av variansanalystabellen i kapitel 9 förekommer skillnader framförallt mellan förorts-, glesbygds- och övriga kommuner. Förortskommuner, dvs. kommuner där mer än 50 procent av befolkningen pendlar till framförallt storstäder, har i genomsnitt större andel invånare i de lägre åldersintervallen, samt god ekonomisk styrka. På motsvarande sätt har glesbygdskommuner i allmänhet en äldre befolkning och lägre genomsnittlig inkomst. Inledande resultat, som visar på ett större antal nyregistreringar i förortskommuner och ett mindre antal i glesbygdskommuner, förefaller såtillvida överensstämma med den framtagna modellen.

Det är i sammanhanget intressant att notera att undersökningen med avseende på fordonstäthet visade signifikant lägre fordonstäthet i förortskommuner, respektive högre fordonstäthet i glesbygdskommuner. En rimlig förklaring är att god ekonomisk styrka innebär kontinuerlig uppdatering av fordonsbeståndet.

Litteratur

Baum, C.F. (2001) "The Stata Journal: Residual Diagnostics for Cross-section Time Series Regression Models", (Nummer 1, Sida 101-104, 2001)

Cameron, A.C., Trivedi, P.K. (2005) "Microeconometrics, methods and applications", (New York: Cambridge University Press, 2005)

Draper, N.R.; Smith, H. (1966): "Applied regression analysis", (New York : Wiley, cop., 2. upplaga 1981)

Greene, W.H. (1993) "Econometric Analysis", (New Jersey: Prentice Hall, inc, 3. upplaga 1997)

Montgomery, D.C. (1976) "Design and Analysis of Experiments", (New Jersey: Wiley, cop, 6. upplaga 2006)

Rawlings, J.O., Pantula, S.G., Dickey, D.A. (1989) "Applied Regression Analysis", (New York: Springer-Verlag, inc, 2. upplaga 1998)

Wooldridge, J.M. (2002) "Econometric Analysis of Cross Section and Panel Data", (Cambridge, Mass: MIT Press, cop, 2002)

APPENDIX

Variansanalystabell

Antag de observerade värdena $X_{\ell 1}, X_{\ell 2}, \dots, X_{\ell n}$ över en faktor ℓ med g olika grupper.

Nollhypotesen *lika medelvärden över faktorn* ℓ formuleras $H_0 : \mu_1 = \mu_2 = \dots = \mu_g$.

Grupp-specifika medelvärden μ_ℓ kan betraktas som summan av det gemensamma medelvärdet μ och grupp-specifik effekt τ_ℓ . Detta ger parametreringen $\mu_\ell = \mu + \tau_\ell$ och nollhypotesen omformuleras till $H_0 : \tau_1 = \tau_2 = \dots = \tau_g$. Det observerade värdet $X_{\ell j}$ uttrycks:

$$X_{\ell j} = \mu + \tau_\ell + e_{\ell j}$$

För att pröva nollhypotesen *lika medelvärden över* ℓ dekomponeras den totala variationen SS_{obs} i en variansanalystabell med komponenterna SS_{tr} (*inomgruppsvarians*) och SS_{res} (*residualvarians*). Om andelen *inomgruppsvarians* är stor, förkastas nollhypotesen. Prövningen görs med *F-test*:

$$F = \frac{SS_{tr} / (g - 1)}{SS_{res} / (\sum_{\ell=1}^g n_\ell - g)} > F_{g-1, \sum n_\ell - g}(\alpha)$$

VIF

För att undersöka om en modell lider av multikollinearitet används VIF-test. Testet låter den oberoende variabel som ska testas vara responsvariabel för övriga oberoende variabler. För detta test fås ett R^2 värde som används i VIF-testet:

$$VIF_i = \frac{1}{1 - R_i^2}, i = 1, 2, \dots, k \text{ där } k \text{ är antalet oberoende variabler i regressionsmodellen.}$$

Om $VIF = 1$ innebär det att de oberoende variablerna kan anses som helt oberoende av varandra. $VIF > 1$ indikerar viss multikollinearitet och $VIF > 5$ tyder på en dåligt anpassad modell.

R^2

Styrkan i sambandet mellan Y och X_1, X_2, \dots, X_k mäts med:

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T}$$

R^2 mäter andelen av den totala variationen i Y som förklaras av sambandet med X_1, X_2, \dots, X_k . Eftersom R^2 alltid ökar då fler förklarande variabler inkluderas i modellen, används ofta det justerade värdet:

$$\bar{R}^2 = 1 - \frac{SS_E(n-k)}{SS_T / (n-1)} = 1 - \left(\frac{n-1}{n-k} \right) (1 - R^2)$$

En ytterliggare variant, R_{PRED}^2 , används för att bedöma modellens prediktiva förmåga. Detta värde beräknas med

$$R_{PRED}^2 = 1 - \frac{PRESS}{SS_T}$$

där

$$PRESS = \sum_{i=1}^n e_{(i)}^2 = \sum [y_i - \hat{y}_{(i)}]^2 .$$

Lack-of-fit

För att pröva om en modell är adekvat anpassad används *F-test* för *lack-of-fit*. Kvadratsumman för modellens residual delas in i två delar:

$$SS_E = SS_{PE} + SS_{LOF}$$

där

SS_{PE} är kvadratsumman som beror av *rena fel*, och SS_{LOF} är kvadratsumman som beror av *lack-of-fit*. För en modell med ett totalt antal observationer på $n = \sum_{i=1}^m n_i$ och med p estimerade parametrar, ges F-värdet på testet för *lack of fit* genom:

$$F_0 = \frac{SS_{LOF}/(m-p)}{SS_{PE}/(n-m)}$$

där m är antalet nivåer av regressorn X_j , $j = 1, 2, \dots, k$. Om F_0 överstiger tabellvärdet för en given sannolikhet lider modellen av *lack-of-fit*. Modellen är då inte tillräcklig för att beskriva observerad data.

Cooks D

För att undersöka huruvida extremvärden påverkar parameterskattningarna beräknas värdena på *Cooks D* för samtliga observationer. En observation med *Cooks D-värde* > 1 betraktas normalt som ett extremvärde. Testet för *Cooks D* beräknas med en konstant p och skrivs enligt följande:

$$D_i = \frac{r_i^2}{p} \frac{h_{ii}}{(1-h_{ii})}, \quad i = 1, 2, \dots, n$$

där

h_{ii} = diagonalelementet i som ligger i $n \times n$ matrisen H . Där H skapar en vektor skattade värden av en vektor observerade värden.

$$\hat{y} = Hy$$

$$r_i = \frac{e_i}{\sqrt{\hat{\sigma}^2(1-h_{ii})}}, \quad i = 1, 2, \dots, n$$

Hausman-test

Test som baseras på jämförelser mellan två estimatorer kallas Hausman-test. Antag två estimatorer $\hat{\theta}$ och $\tilde{\theta}$ och en testsituation där:

$$H_0 : p \lim(\hat{\theta} - \tilde{\theta}) = 0$$

$$H_a : p \lim(\hat{\theta} - \tilde{\theta}) \neq 0$$

Under nollhypotesen är båda estimatorerna konsistenta och $\tilde{\theta}$ är effektiv. Under mothypotesen är endast $\hat{\theta}$ konsistent. Testvärdet ges av:

$$H = (\hat{\theta} - \tilde{\theta})'(N^{-1}\hat{V}_H)^{-1}(\hat{\theta} - \tilde{\theta})$$

där \hat{V}_H är den skattade variansmatrisen. Nollhypotesen förkastas då $H > \chi^2_\alpha(q)$.

Vid *fixa-effekter* är *inom-estimatorn* $\hat{\beta}_w$ konsistent. RE-estimatorn $\hat{\beta}_{RE}$ är då inkonsistent. Vid *slumpmässiga effekter* är $\hat{\beta}_w$ konsistent medan $\hat{\beta}_{RE}$ är konsistent och effektiv.

Modifierat Wald-test för heteroskedasticitet

Modifierat Wald-test för gruppvis heteroskedasticitet utgår från nollhypotesen $\sigma_i^2 = \sigma^2$ för $i = 1, 2, \dots, N_g$, där N_g betecknar antalet grupper. Låt $\hat{\sigma}_i^2 = T_i^{-1} \sum_{t=1}^{T_i} e_{it}^2$ vara en skattning av den *i*:te gruppens residualvarians. Definiera därefter:

$$V_i = T_i^{-1} (T_i - 1)^{-1} \sum_{t=1}^{T_i} (e_{it}^2 - \hat{\sigma}_i^2)^2$$

som är den skattade variansen av σ_i^2 . Testvärdet för det *modifierade Wald-testet* ges av:

$$W = \sum_{i=1}^{N_g} \frac{(\hat{\sigma}_i^2 - \hat{\sigma}^2)^2}{V_i}$$

som är $\chi^2_{N_g}$ under nollhypotesen.

Bilaga 1. Sveriges Kommuner och Landstings kommungruppsindelning.

1. Storstäder (3 kommuner)

Kommun med en folkmängd som överstiger 200 000 invånare.

2. Förortskommuner (38 kommuner) Kommun där mer än 50 procent av nattbefolkningen pendlar till arbetet i någon annan kommun. Det vanligaste utpendlingsmålet skall vara någon av storstäderna.

3. Större städer (27 kommuner)

Kommun med 50 000– 200 000 invånare samt en tätortsgrad överstigande 70 procent.

4. Pendlingskommuner (41 kommuner)

Kommun där mer än 40 procent av nattbefolkningen pendlar till arbetet i någon annan kommun.

5. Glesbygdskommuner (39 kommuner)

Kommun med mindre än 7 invånare per kvadratkilometer och mindre än 20 000 invånare.

6. Varuproducerande kommuner (40 kommuner)

Kommun med mer än 40 procent av nattbefolkningen mellan 16 och 64 år, anställda inom varutillverkning och industriell verksamhet. (SNI92)

7. Övriga kommuner, över 25 000 inv. (34 kommuner)

Kommun som inte hör till någon av tidigare grupper och har mer än 25 000 invånare.

8. Övriga kommuner, 12 500-25 000 inv. (37 kommuner)

Kommun som inte hör till någon av tidigare grupper och har 12 500-25 000 invånare.

9. Övriga kommuner, mindre än 12 500 inv. (31 kommuner)

Kommun som inte hör till någon av tidigare grupper och har mindre än 12 500 invånare.

Källa: Sveriges Kommuner och Landsting.

Bilaga 2. Undersökningens samtliga förklarande variabler.

BILPER1000 – antal registrerade personbilar ägda av fysiska personer eller personliga företag per 1000 invånare (årsskiftet 2006-2007). Källa: SIKA.

INV – antal invånare (2006). Källa: SCB.

INVKM2 – antal invånare per km² (2005). Källa SCB.

LANDAREAL – landareal (2005). Källa: SCB.

MDLALDER – medelålder (2005). Källa: SCB.

SMAHUS – andel småhus av det totala bostadsbeståndet (2006). Källa: SCB.

UTPENDL – antal utpendlare per 1000 invånare 16 år och äldre (2005). Källa: SCB.

KOLLTRKM – utbud av kollektivtrafik i km per invånare efter län (2006). Källa: SIKA.

ARBLOS – andel arbetslösa i procent (januari 2007). Källa: AMV.

INK – sammanräknad genomsnittlig förvärvsinkomst för invånare 16 år och äldre (årsskiftet 2005-2006). Källa: SCB.

SEKTOR1 – antal anställda inom sektorn *jordbruk, skogsbruk, jakt och fiske* per 1000 invånare 16 år och äldre (2005). Källa: SCB.

SEKTOR2 – antal anställda inom sektorn *utvinning av mineral, tillverkningsindustri* per 1000 invånare 16 år och äldre (2005). Källa: SCB.

SEKTOR3 – antal anställda inom sektorn *energi och vattenförsörjning, avfall* per 1000 invånare 16 år och äldre (2005). Källa: SCB.

SEKTOR4 – antal anställda inom sektorn *byggindustri* per 1000 invånare 16 år och äldre (2005). Källa: SCB.

SEKTOR5 – antal anställda inom sektorn *handel; transport, magasinering, kommunikation* per 1000 invånare 16 år och äldre (2005). Källa: SCB.

SEKTOR6 – antal anställda inom sektorn *personliga och kulturella tjänster* per 1000 invånare 16 år och äldre (2005). Källa: SCB.

SEKTOR7 – antal anställda inom sektorn *kreditinstitut, fastighetsförvaltning, företagstjänster* per 1000 invånare 16 år och äldre (2005). Källa: SCB.

SEKTOR8 – antal anställda inom sektorn *civila myndigheter, försvar* per 1000 invånare 16 år och äldre (2005). Källa: SCB.

SEKTOR9 – antal anställda inom sektorn *forskning och utveckling, utbildning* per 1000 invånare 16 år och äldre (2005). Källa: SCB.

SEKTOR10 – antal anställda inom sektorn *hälso- och sjukvård, socialtjänst* per 1000 invånare 16 år och äldre (2005). Källa: SCB.

ALDER1 – antal personer i åldersgruppen 15-24 per 1000 invånare (2006). Källa: SCB.

ALDER2 – antal personer i åldersgruppen 25-34 per 1000 invånare (2006). Källa: SCB.

ALDER3 – antal personer i åldersgruppen 35-44 per 1000 invånare (2006). Källa: SCB.

ALDER4 – antal personer i åldersgruppen 45-54 per 1000 invånare (2006). Källa: SCB.

ALDER5 – antal personer i åldersgruppen 55-64 per 1000 invånare (2006). Källa: SCB.

ALDER6 – antal personer i åldersgruppen 65- per 1000 invånare (2006). Källa: SCB.

Bilaga 3

Beskrivande statistik över NYREG efter kommungrupp.

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1,00	30	14,4633	3,84667	,70230	13,0270	15,8997	9,30	24,50
2,00	380	17,2347	4,04931	,20773	16,8263	17,6432	9,00	34,60
3,00	270	14,7330	3,45647	,21035	14,3188	15,1471	7,30	27,20
4,00	402	15,6868	3,86212	,19262	15,3081	16,0655	6,10	30,60
5,00	390	9,0656	3,07975	,15595	8,7590	9,3723	2,60	22,10
6,00	399	14,6188	3,31842	,16613	14,2922	14,9454	7,70	24,60
7,00	340	15,0574	4,25796	,23092	14,6031	15,5116	5,10	32,40
8,00	370	14,4738	3,60239	,18728	14,1055	14,8421	5,80	28,10
9,00	310	14,3877	3,18280	,18077	14,0320	14,7434	5,90	22,90
Total	2891	14,3793	4,28200	,07964	14,2232	14,5355	2,60	34,60

Variansanalystabell

NYREG

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	15013,746	8	1876,718	142,424	,000
Within Groups	37975,977	2882	13,177		
Total	52989,723	2890			

Multiple Comparisons

Dependent Variable: NYREG

Dunnett T3

(I) KG	(J) KG	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1,00	2,00	-2,77140(*)	,73238	,020	-5,2940	-,2488
	3,00	-,26963	,73313	1,000	-2,7941	2,2549
	4,00	-1,22348	,72824	,947	-3,7362	1,2892
	5,00	5,39769(*)	,71941	,000	2,9059	7,8895
	6,00	-,15546	,72168	1,000	-2,6526	2,3417
	7,00	-,59402	,73929	1,000	-3,1334	1,9454
	8,00	-,01045	,72684	1,000	-2,5198	2,4989
	9,00	,07559	,72519	1,000	-2,4299	2,5811
	2,00	1,00	2,77140(*)	,73238	,020	,2488
3,00		2,50177(*)	,29563	,000	1,5548	3,4488
4,00		1,54792(*)	,28329	,000	,6412	2,4547
5,00		8,16910(*)	,25975	,000	7,3374	9,0008
6,00		2,61594(*)	,26599	,000	1,7644	3,4675
7,00		2,17738(*)	,31060	,000	1,1829	3,1719

	8,00	2,76095(*)	,27968	,000	1,8656	3,6563
	9,00	2,84699(*)	,27537	,000	1,9652	3,7288
3,00	1,00	,26963	,73313	1,000	-2,2549	2,7941
	2,00	-2,50177(*)	,29563	,000	-3,4488	-1,5548
	4,00	-,95385(*)	,28522	,031	-1,8676	-,0401
	5,00	5,66732(*)	,26186	,000	4,8279	6,5068
	6,00	,11417	,26804	1,000	-,7449	,9732
	7,00	-,32439	,31237	1,000	-1,3251	,6764
	8,00	,25918	,28164	1,000	-,6432	1,1616
	9,00	,34522	,27736	1,000	-,5438	1,2342
4,00	1,00	1,22348	,72824	,947	-1,2892	3,7362
	2,00	-1,54792(*)	,28329	,000	-2,4547	-,6412
	3,00	,95385(*)	,28522	,031	,0401	1,8676
	5,00	6,62117(*)	,24784	,000	5,8279	7,4145
	6,00	1,06802(*)	,25437	,001	,2539	1,8822
	7,00	,62946	,30071	,735	-,3334	1,5924
	8,00	1,21303(*)	,26866	,000	,3531	2,0730
	9,00	1,29907(*)	,26416	,000	,4533	2,1449
5,00	1,00	-5,39769(*)	,71941	,000	-7,8895	-2,9059
	2,00	-8,16910(*)	,25975	,000	-9,0008	-7,3374
	3,00	-5,66732(*)	,26186	,000	-6,5068	-4,8279
	4,00	-6,62117(*)	,24784	,000	-7,4145	-5,8279
	6,00	-5,55316(*)	,22786	,000	-6,2824	-4,8239
	7,00	-5,99171(*)	,27865	,000	-6,8844	-5,0990
	8,00	-5,40814(*)	,24371	,000	-6,1884	-4,6279
	9,00	-5,32210(*)	,23874	,000	-6,0867	-4,5575
6,00	1,00	,15546	,72168	1,000	-2,3417	2,6526
	2,00	-2,61594(*)	,26599	,000	-3,4675	-1,7644
	3,00	-,11417	,26804	1,000	-,9732	,7449
	4,00	-1,06802(*)	,25437	,001	-1,8822	-,2539
	5,00	5,55316(*)	,22786	,000	4,8239	6,2824
	7,00	-,43856	,28447	,991	-1,3498	,4726
	8,00	,14501	,25034	1,000	-,6564	,9464
	9,00	,23106	,24551	1,000	-,5552	1,0173
7,00	1,00	,59402	,73929	1,000	-1,9454	3,1334
	2,00	-2,17738(*)	,31060	,000	-3,1719	-1,1829
	3,00	,32439	,31237	1,000	-,6764	1,3251
	4,00	-,62946	,30071	,735	-1,5924	,3334
	5,00	5,99171(*)	,27865	,000	5,0990	6,8844
	6,00	,43856	,28447	,991	-,4726	1,3498
	8,00	,58357	,29732	,838	-,3686	1,5357
	9,00	,66961	,29326	,559	-,2698	1,6090
8,00	1,00	,01045	,72684	1,000	-2,4989	2,5198
	2,00	-2,76095(*)	,27968	,000	-3,6563	-1,8656
	3,00	-,25918	,28164	1,000	-1,1616	,6432
	4,00	-1,21303(*)	,26866	,000	-2,0730	-,3531
	5,00	5,40814(*)	,24371	,000	4,6279	6,1884
	6,00	-,14501	,25034	1,000	-,9464	,6564
	7,00	-,58357	,29732	,838	-1,5357	,3686
	9,00	,08604	,26029	1,000	-,7475	,9196
9,00	1,00	-,07559	,72519	1,000	-2,5811	2,4299
	2,00	-2,84699(*)	,27537	,000	-3,7288	-1,9652

3,00	-,34522	,27736	1,000	-1,2342	,5438
4,00	-1,29907(*)	,26416	,000	-2,1449	-,4533
5,00	5,32210(*)	,23874	,000	4,5575	6,0867
6,00	-,23106	,24551	1,000	-1,0173	,5552
7,00	-,66961	,29326	,559	-1,6090	,2698
8,00	-,08604	,26029	1,000	-,9196	,7475

* The mean difference is significant at the .05 level.

Bilaga 4

Predictor	Coef	SE Coef	T	P	VIF
Constant	-563,1	112,1	-5,02	0,000	
INVKM2	-0,036005	0,004121	-8,74	0,000	2,2
MDLALDER	10,676	1,580	6,76	0,000	9,5
SMAHUS	84,65	14,34	5,90	0,000	3,3
KOLLTRKM	-0,63363	0,06533	-9,70	0,000	1,5
SEKTOR1	-0,5249	0,1661	-3,16	0,002	2,5
ALDER2	0,4840	0,1500	3,23	0,001	6,6
ALDER3	1,6787	0,2105	7,97	0,000	6,6
ALDER4	0,9236	0,2075	4,45	0,000	1,5
ALDER5	0,8918	0,1454	6,13	0,000	2,5

S = 20,4262 R-Sq = 76,4% R-Sq(adj) = 75,6%

PRESS = 126177 R-Sq(pred) = 74,49%

Lack of fit test

Possible curvature in variable INVKM2 (P-Value = 0,020)

Possible interaction in variable INVKM2 (P-Value = 0,088)

Possible curvature in variable MDLALDER (P-Value = 0,001)

Possible curvature in variable SMAHUS (P-Value = 0,000)

Possible interaction in variable SMAHUS (P-Value = 0,020)

Possible curvature in variable SEKTOR1 (P-Value = 0,000)

Possible interaction in variable SEKTOR1 (P-Value = 0,002)

Possible curvature in variable ALDER2 (P-Value = 0,009)

Possible interaction in variable ALDER2 (P-Value = 0,034)

Possible curvature in variable ALDER3 (P-Value = 0,091)

Possible interaction in variable ALDER3 (P-Value = 0,035)

Possible curvature in variable ALDER5 (P-Value = 0,010)

Possible lack of fit at outer X-values (P-Value = 0,000)

Overall lack of fit test is significant at P = 0,000

Bilaga 5

Predictor	Coef	SE Coef	T	P	VIF
Constant	324,07	15,44	20,99	0,000	
ALDER3	0,7623	0,1158	6,58	0,000	1,9
ALDER5*	1,4528	0,1316	11,04	0,000	1,9
ALDER5*2	-0,024523	0,006166	-3,98	0,000	1,3
KOLLTRKM*	-0,43245	0,08334	-5,19	0,000	2,3
KOLLTRKM*2	-0,012059	0,002295	-5,25	0,000	2,9
SMAHUS*	32,61	10,75	3,03	0,003	1,8
SMAHUS*2	-229,04	40,57	-5,65	0,000	1,7

S = 20,8393 R-Sq = 75,2% R-Sq(adj) = 74,6%

PRESS = 131106 R-Sq(pred) = 73,49%

No evidence of lack of fit (P >= 0,1).

**White's test for Ho: homoskedasticity
against Ha: unrestricted heteroskedasticity**

chi2(32) = 35.51
Prob > chi2 = 0.3062

Cameron & Trivedi's decomposition of IM-test

Source	chi2	DF	P
Heteroskedasticity	35.51	32	0.3062
Skewness	5.36	7	0.6163
Kurtosis	0.64	1	0.4234
Total	41.51	40	0.4047

Bilaga 6. Undersökningens förklarande variabler.

NYREG – antal nyregistrerade personbilar ägda av fysiska personer eller personliga företag per 1000 invånare (1997–2006). Källa: SIKA.

BEST – fordonsbestånd per 1000 invånare (1997–2006). Källa: SIKA.

INVKM2 – antal invånare per km² (1997-2006). Källa SCB.

INK – genomsnittlig förvärvsinkomst (1997–2006). Källa: SCB.

ARB – antal förvärvsarbetande per 1000 invånare (1997–2006). Källa: SCB.

UTPENDL – antal utpendlare per 1000 invånare 16 år och äldre (1997–2006). Källa: SCB.

ALDER1 – antal personer i åldersgruppen 15-24 per 1000 invånare (1997–2006). Källa: SCB.

ALDER2 – antal personer i åldersgruppen 25-34 per 1000 invånare (1997–2006). Källa: SCB.

ALDER3 – antal personer i åldersgruppen 35-44 per 1000 invånare (1997–2006). Källa: SCB.

ALDER4 – antal personer i åldersgruppen 45-54 per 1000 invånare (1997–2006). Källa: SCB.

ALDER5 – antal personer i åldersgruppen 55-64 per 1000 invånare (1997–2006). Källa: SCB.

ALDER6 – antal personer i åldersgruppen 65- per 1000 invånare (1997–2006). Källa: SCB.

Bilaga 7.

Diagram 1. Genomsnittligt antal nyregistreringar efter kommungrupp 1997 – 2006.

År	Kommungrupp								
	1	2	3	4	5	6	7	8	9
1997	12,2	14,5	13	13,9	8,6	13	13,5	13,5	12
1998	13,6	16,7	14,2	14,8	9,7	13,9	14,4	14,4	13,5
1999	16,9	20,6	17,6	18,3	11,7	17,1	18	18	16,4
2000	19,1	22,1	18,5	19,5	11,6	18	18,9	18,9	18,2
2001	13,7	16,2	13,2	14	7,9	13	13,2	13,2	12,9
2002	13,3	16,1	13,4	14,3	7,8	13,7	13,7	13,7	13,4
2003	13,9	16,6	14,4	15,5	8,3	14,4	15,2	15,2	14,6
2004	14,4	17,1	14,7	15,9	8,5	14,8	15,4	15,4	14,9
2005	13,6	16,5	14,3	15,5	8,5	14,2	14,4	14,4	14,4
2006	13,9	16	14,1	15,1	8,2	14,3	13,9	13,9	13,6

Bilaga 8.

Source	SS	df	MS	
Model	14635.5164	6	2439.25273	Number of obs = 2891
Residual	38354.2067	2884	13.2989621	F(6, 2884) = 183.42
Total	52989.7231	2890	18.3355443	Prob > F = 0.0000
				R-squared = 0.2762
				Adj R-squared = 0.2747
				Root MSE = 3.6468

nyreg	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
ink	.0284142	.0028508	9.97	0.000	.0228244 .034004
best	.0308349	.0022685	13.59	0.000	.0263868 .0352829
alder1	-.076922	.0065433	-11.76	0.000	-.089752 -.064092
alder2	.0586904	.0056644	10.36	0.000	.0475838 .0697971
alder5	-.0150728	.0053636	-2.81	0.005	-.0255897 -.004556
alder6	-.0373908	.0028881	-12.95	0.000	-.0430537 -.0317279
_cons	7.997337	2.158288	3.71	0.000	3.765393 12.22928

Variable	VIF	1/VIF
alder2	3.05	0.327970
alder6	2.50	0.399617
best	1.86	0.538199
ink	1.62	0.615818
alder5	1.45	0.688149
alder1	1.21	0.827436
Mean VIF	1.95	

White's test for H₀: homoskedasticity
against H_a: unrestricted heteroskedasticity

chi2(27) = 106.72
Prob > chi2 = 0.0000

Cameron & Trivedi's decomposition of IM-test

Source	chi2	df	p
Heteroskedasticity	106.72	27	0.0000
Skewness	62.66	6	0.0000
Kurtosis	5.95	1	0.0147
Total	175.32	34	0.0000

Bilaga 9.

```

Random-effects GLS regression           Number of obs   =   2891
Group variable: id                     Number of groups =   290

R-sq:  within = 0.5823                 Obs per group:  min =    4
        between = 0.3267                avg   =   10.0
        overall = 0.3987                max   =   10

Random effects u_i ~ Gaussian          wald chi2(20)   =   3724.50
corr(u_i, X) = 0 (assumed)            Prob > chi2     =   0.0000

```

nyreg	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
ink	.0248355	.0072963	3.40	0.001	.010535	.039136
arb	.0145508	.0034693	4.19	0.000	.007751	.0213505
best	.0351937	.003481	10.11	0.000	.028371	.0420163
invkm2	-.0001319	.0005462	-0.24	0.809	-.0012025	.0009388
pendlare	.005422	.0019765	2.74	0.006	.001548	.0092959
alder1	.0335351	.0094561	3.55	0.000	.0150015	.0520688
alder2	.0466283	.0102013	4.57	0.000	.0266342	.0666225
alder3	.03203	.011481	2.79	0.005	.0095277	.0545324
alder4	-.0192202	.0090457	-2.12	0.034	-.0369495	-.0014909
alder5	.0005661	.0030867	0.18	0.854	-.0054837	.0066159
alder6	-.0008497	.0081361	-0.10	0.917	-.0167963	.0150969
t2	.5971295	.1464697	4.08	0.000	.3100542	.8842048
t3	3.298986	.1778206	18.55	0.000	2.950464	3.647508
t4	3.593145	.2295167	15.66	0.000	3.1433	4.042989
t5	-1.675997	.2689623	-6.23	0.000	-2.203154	-1.148841
t6	-1.652863	.3163934	-5.22	0.000	-2.272983	-1.032744
t7	-.8095249	.3658616	-2.21	0.027	-1.5266	-.0924494
t8	-.7747794	.4112341	-1.88	0.060	-1.580783	.0312247
t9	-1.510936	.4570382	-3.31	0.001	-2.406714	-.6151573
t10	-2.311199	.5112124	-4.52	0.000	-3.313157	-1.309241
_cons	-23.86463	5.865504	-4.07	0.000	-35.36081	-12.36846
sigma_u	2.8868949					
sigma_e	1.5392489					
rho	.77864263	(fraction of variance due to u_i)				

Hausman-test för slumpmässiga- mot fixa effekter:

```

(b)      (B)
fixed    random

```

b = consistent under Ho and Ha; obtained from xtreg
B = inconsistent under Ha, efficient under Ho; obtained from xtreg

Test: Ho: difference in coefficients not systematic

```

chi2(19) = (b-B)'[(V_b-V_B)^(-1)](b-B)
          = 41.25
Prob>chi2 = 0.0022

```

Bilaga 10.

Fixed-effects (within) regression
 Group variable: id

Number of obs = 2891
 Number of groups = 290

R-sq: within = 0.5845
 between = 0.2811
 overall = 0.3406

Obs per group: min = 4
 avg = 10.0
 max = 10

corr(u_i, xb) = -0.4147

F(20,2581) = 181.53
 Prob > F = 0.0000

nyreg	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ink	.0198081	.0096265	2.06	0.040	.0009317	.0386845
arb	.0190771	.0040009	4.77	0.000	.0112318	.0269223
best	.0363235	.004133	8.79	0.000	.0282192	.0444278
invkm2	-.0028331	.0032302	-0.88	0.381	-.0091673	.003501
utpend1	.0143304	.0043047	3.33	0.001	.0058894	.0227713
alder2	.0453551	.011696	3.88	0.000	.0224205	.0682897
alder1	.0407185	.0107735	3.78	0.000	.0195929	.0618441
alder3	.03879	.0124469	3.12	0.002	.0143831	.063197
alder4	-.0225634	.0095809	-2.36	0.019	-.0413504	-.0037763
alder5	-.0012256	.0031124	-0.39	0.694	-.0073287	.0048774
alder6	-.008679	.0091537	-0.95	0.343	-.0266284	.0092704
t2	.4922133	.1624115	3.03	0.002	.1737433	.8106834
t3	3.17281	.2185762	14.52	0.000	2.744208	3.601413
t4	3.389481	.2969695	11.41	0.000	2.807158	3.971803
t5	-1.876498	.3529817	-5.32	0.000	-2.568653	-1.184342
t6	-1.887387	.418121	-4.51	0.000	-2.707273	-1.0675
t7	-.9634962	.4758484	-2.02	0.043	-1.89658	-.030413
t8	-.980613	.533014	-1.84	0.066	-2.025791	.0645654
t9	-1.710099	.5900188	-2.90	0.004	-2.867057	-.5531403
t10	-2.544307	.6575801	-3.87	0.000	-3.833745	-1.254869
_cons	-26.42716	6.426634	-4.11	0.000	-39.02904	-13.82528
sigma_u	3.505211					
sigma_e	1.5392489					
rho	.83833781	(fraction of variance due to u_i)				

F test that all u_i=0: F(289, 2581) = 35.43 Prob > F = 0.0000

Bilaga 11.

Fixed-effects (within) regression
 Group variable: id

Number of obs = 2891
 Number of groups = 290

R-sq: within = 0.5833
 between = 0.3100
 overall = 0.3620

Obs per group: min = 4
 avg = 10.0
 max = 10

corr(u_i, xb) = -0.4290

F(16,289) = 155.12
 Prob > F = 0.0000

(Std. Err. adjusted for 290 clusters in id)

nyreg	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
ink	.0218813	.0144974	1.51	0.132	-.0066525	.0504152
arb	.0197765	.005547	3.57	0.000	.0088588	.0306941
best	.03691	.0121053	3.05	0.003	.0130842	.0607357
utpendl	.0145239	.0061	2.38	0.018	.0025177	.02653
alder1	.0380001	.0130402	2.91	0.004	.0123344	.0636659
alder2	.0536661	.0161273	3.33	0.001	.0219243	.0854079
alder3	.0571013	.0164061	3.48	0.001	.0248108	.0893919
t2	.4645969	.1867666	2.49	0.013	.0970016	.8321922
t3	3.139735	.3312506	9.48	0.000	2.487765	3.791705
t4	3.359195	.4996667	6.72	0.000	2.375747	4.342642
t5	-1.879362	.5631617	-3.34	0.001	-2.98778	-.7709433
t6	-1.861106	.6524408	-2.85	0.005	-3.145244	-.576968
t7	-.8856605	.759967	-1.17	0.245	-2.381432	.6101114
t8	-.8757934	.8398609	-1.04	0.298	-2.528813	.7772262
t9	-1.58013	.9102625	-1.74	0.084	-3.371715	.2114543
t10	-2.411554	1.009023	-2.39	0.017	-4.397519	-.4255879
_cons	-35.87834	6.872123	-5.22	0.000	-49.4041	-22.35258
sigma_u	3.4602544					
sigma_e	1.5401953					
rho	.83463864	(fraction of variance due to u_i)				

wooldridge test for autocorrelation in panel data

H0: no first-order autocorrelation
 F(1,289) = 383.590
 Prob > F = 0.0000

Modified wald test for groupwise heteroskedasticity in fixed effect regression model

H0: $\sigma(i)^2 = \sigma^2$ for all i

chi2 (290) = 11431.35
 Prob>chi2 = 0.0000