

RAPPORT

Syntetisk befolkning med hushållsinformation – som markanvändningsdata i transportmodellerna

Peter Almström

Svante Berglund (uppdragsledare)

Ulrika Isberg

INNEHÅLL

SAMMANFATTNING	3
INLEDNING	4
Bakgrund	4
Syfte och mål	5
Syntetisk befolkning	6
METOD	9
Prognosticera demografiskt driven hushållsbildning	9
Skapa en syntetisk befolkning	10
Arbetsgång	10
Implementation av metod	14
Syntetisering	14
Kodning av prognosresultat på agenter	18
RESULTAT OCH VALIDERING	19
Hushållsbildning	19
Syntetisk befolkning	23
AVSLUTANDE KOMMENTARER	27
BILAGA 1 SAMTLIGA HUSHÅLLSTYPER	29
BILAGA 2 ÅLDERSKLASSER	30
BILAGA 3 DETALJER FÖR DEN IMPLEMENTERADE METODEN	31
BILAGA 4 PROGRAMMETS INSTÄLLNINGAR I DETALJ	34

Sammanfattning

Dagens svenska transportmodeller bygger på individdata från skattning till tillämpning. Samtidigt är det väl känt att vissa beslut av stor betydelse för ett hushålls transporter styrs av dess sammansättning, inkomster och boende. Exempel på sådana beslut är bilinnehav och antal resor som utförs av individer som tillhör hushållet. Det finns därmed skäl att utveckla indata till våra modeller så att hushållsinformation finns med i modellernas tillämpningsdata.

I rapporten redogörs för hur man kan göra en syntetisk befolkning med hushållsinformation knuten till individer. Utgångspunkten är en demografiskt driven hushållsbildningsmodell. Med demografiskt driven avses att det är befolkningens sammansättning och preferenser som avgör antalet hushåll, inte bostadsmarknaden. I hushållsdimensionen hanteras hushållen med avseende på antal vuxna och antal personer <20 år i hushållet, med totalt 15 hushållstyper. Uppdelningen görs separat för småhus och flerfamiljshus. I ett andra steg skapas en syntetisk befolkning. Här utgör hushållstyp ett villkor så att varje zon fylls av individer som bygger upp den hushållsstruktur som beräknats i föregående steg. Andra fördelningar som metoden tar hänsyn till är efter kön, ålder, inkomst och antal förvärvsarbetande. Allt är implementerat i prototypprogramvara¹ för att göra arbetsgången repeterbar och spårbar.

De modeller och arbetssätt som tagits fram ger god överensstämmelse med data och en demonstrationsdatabas² har tagits fram med hushåll och befolkning för hela landet som motsvarar en SAMS-bas samt med information om individens hushållsstatus och hushållsinkomst.

Huvudsyftet med projektet var att tillfoga hushållsinformation till modellen och därigenom förbättra prognosernas precision. Den utveckling som gjorts har i första hand syftat till att förbättra de regionala modellerna men en möjligen större potential hos data ligger i framtida utveckling av storstadsmodellerna. De modeller för storstad som skisserats bygger på att man följer en individs resmönster under en period. Den här typen av data kan även användas för att förbättra långväga modeller där resorna ofta är betingade av hushållens sammansättning.

För processen att skapa data till trafikverkets modeller innebär tillvägagångssättet inte någon stor skillnad. Samma prognosdata som tidigare ska tas fram på samma zonindelning. Det tillkommer ett arbetsmoment med att skapa hushåll och agenter som beräknas ta ca 2 veckor inklusive grundläggande kvalitetskontroll. Samtliga arbetsmoment kan automatiseras med skript, delar av kvalitetskontrollen kräver dock manuell hantering.

Trafikverkets representant som beställare har varit Leonid Engelson och rapporten har granskats av Daniel Jonsson (KTH).

¹ Tillgänglig från författarna för granskning.

² Tillgänglig från författarna för granskning.

Inledning

Bakgrund

De transportmodeller vi använder idag har nått en viss mognad och fungerar på det hela taget bra. Det finns ett tydligt forskningsstöd för att de osäkerheter som ryms inom modellerna är kopplade till förutsättningar snarare än statistiska osäkerheter i parameterestimaten. Det finns således skäl att i forsknings- och utvecklingsverksamhet jobba mer med indata.

De data vi använder i dagens transportmodeller är baserade på information om individer. Det gäller i hela modellkedjan från skattning till tillämpning. Många av de beslut som fattas och har stor betydelse för transporterna är grundar sig emellertid på hushållens situation avseende storlek, sammansättning och hushållens ekonomi. Hushållets egenskaper i dessa avseenden styr långsiktiga beslut såsom boende och innehav av bil. De resvaneundersökningar som används som skattningsunderlag för transportmodeller är individbaserade i betydelsen att det är individers resmönster som undersöks men information om individens hushållssituation samlas in i form av bakgrundsdata. I samband med arbete med omskattning av Sampers³ blev det tydligt att förklaringen i modellerna kan förbättras om man använder de data som finns tillgängliga i RVU avseende individens hushållssituation. Det har dock inte varit möjligt att implementera modeller med de önskade variablerna till följd av avsaknad av markanvändningsdata (det vill säga, markanvändningsdata som idag är baserade på individer utan information om hushållen).

Genom att utveckla markanvändningsdata så att de även innehåller hushållsinformation, kan transportmodellerna i flera avseenden förbättras. Det gäller främst modellens egenskaper avseende innehav av bil, resegenerering och konkurrens inom hushållet avseende bil. Innehav av bil styrs i hög grad av hushållets inkomster snarare än enskilda individers inkomst och förekomst av barn i hushållet har stor betydelse för innehav av bil. Generering av vissa restyper som att skjutsa påverkas av hushållets sammansättning. En viktig faktor i valet av färdmedel är inte bara förekomst av bil i hushållet utan hur många andra som har körkort och konkurrerar om samma bil. I nuvarande version av Sampers behandlas bostadszonen som en bilpool vilket det vore önskvärt att justera till att bara gälla hushållet.

För att hantera de förutsättningar som påverkar behoven av transporter och de begränsningar som påverkar tillgången till alternativa färd sätt behöver vi hålla samman individen och dess hushåll. Ett sätt att göra detta är att skapa en syntetisk befolkning av *agenter*⁴ med information om agenten och det hushåll som agenten bor i. I många av dagens transportmodeller låter man befolkningen representeras av agenter, med eller utan hushållsinformation.

³ Projekt under 2016.

⁴ I samband med att syntetiska data skapas benämner vi observationerna agenter.

Det huvudsakliga skälet till att göra den här utvecklingen är att öka modellernas precision men vi får även andra egenskaper på köpet. När modellerna körs kommer vi att generera information för de syntetiska individerna som kan vara mycket användbart. Vi kommer exempelvis att ha tillgång till logsumma per individ som tillsammans med bakgrundsinformation kan användas till fördelningsanalyser utifrån inkomst, kön med mera.

Syfte och mål

Projektets mål är att anvisa en metod för att skapa en syntetisk befolkning där vi lägger till information om hushållet till individerna. Det gäller hushållsinkomst och hur familjen är sammansatt.

Målet är att knyta följande data till individerna:

- Vilken typ (storlek, bostadstyp) av hushåll bor individen i?
- Vilken är hushållets inkomst?
- Hur många barn finns i hushållet?

Notera att vi inte i ett första skede tänker oss en grundläggande datastruktur som är hushållsbaserad utan att endast tillföra hushållsinformation till individerna.

De data som används i pågående Sampersskattning visas i tabellen nedan. Bil i hushållet och bilkonkurrens är modellgenererade data medan sällskap är genererat från fördelningar som inte är individ- eller områdesspecifika. Med information om hushållets sammansättning kan vi tillföra information om:

- Förekomst av barn i hushållet
- Antal vuxna i hushållet

Tillsammans med modellgenererade data avseende körkort och bilar i hushållet kan vi förbättra informationen avseende bilkonkurrens.

	Gruppering	Besök	Skola grund	Skola gym	Skola övr	Skjutisa	SeBaVa	Övrigt	Sällaninik öp	Daigliginik öp	Tjänste	Arbete	Rekrea-tion
Inkomst	0-10	0		0	0	0	0	0	0	0	0	0	0
	10-150	1		0	1	1	1	1	0	1	0	1	1
	150-300	2		0	1	2	2	2	0	2	0	2	2
	300-	2		0	1	2	2	2	0	3	0	3	2
Boendetyt	Villa		W, BC	W, BC		PT						W	C, CP
	Fhus												W
	6-10 (Grund)	PT, CP, BC	BC, PT			W				BC, CP			
Alder	11-15 (Grund)	PT, CP, BC	CP			W				BC, CP			W
	16-17 (Gymm)							PT					W
	18-19 (Gymm)							PT					
	20-25 (Vux)							PT					
	26-64 (Vux)												
	65-74					CP							
Bil i hushäll	75-							C					
	>=1	C, CP	CP		C	C	C, CP	C, CP	C, CP	BC		C	C
	Kvinnu	C, CP, BC	PT, BC	W, BC	BC, PT	C	C, CP	C, CP	C	C, CP	C	BC, C	BC, C, CP
Sällskap	Ensam	BC	W				BC	BC		BC			BC
	>=3	C			C	C	C	C	C			C	C
Bil-konkurrens					C	C	C	C	C			C	C

Tabell 1. Indata i pågående Sampersomsättning. I tabellen avser siffrorna i raderna för inkomst antal klasser som tillämpas i modellen, bokstäverna för övriga rader avser vilket färdsets nyttofunktion som variabeln ingår i (C: bil, CP: bil passagerare, PT: kollektivtrafik, W: gång, BC: Cykel).

Syntetisk befolkning

Vad är en syntetisk befolkning? En syntetisk befolkning är en representation av data där vi har ett antal individer eller agenter som för varje zon, om man aggregerar dem,

summerar upp till zonbefolkningens statistiska egenskaper. Det finns inget behov av att knyta agenterna till några riktiga individer men agenterna ska vara sådana att egenskaperna är korrelerade på samma sätt som i en verklig befolkning. Att använda sig av en syntetisk befolkning möjliggör en ökad flexibilitet hos modellen och att egenskaper hos agenterna som är korrelerade kan behållas.

I nuvarande version av Sampers representeras befolkningen av flera tabeller med antal personer per zon fördelat på grupper. Varje tabell omfattar en eller flera variabler som beskriver befolkningen i sina dimensioner som kan vara exempelvis ålder och kön. Representerar man befolkningen med tabeller får man av praktiska skäl hålla sig inom ett begränsat antal dimensioner då tabellen annars ökar med en faktor som motsvaras av antal kategorier för varje dimension som ska beskrivas. För att data inte ska explodera i storlek delar man upp de i flera tabeller som hanterar ett begränsat antal dimensioner. Priset man betalar är att korrelationerna mellan olika egenskaper går förlorade i data. I tabellen nedan har vi en tabell för två zoner och antal män och kvinnor i zonerna.

Tabell 2. Exempel på den datarepresentation man använder för närvarande.

<i>ZonID</i>	<i>Ant_män</i>	<i>Ant_kv</i>
1	1	2
2	2	2

I en syntetisk befolkning frångår man tabellen som form att representera en befolkning. Innehållet i tabellen ovan kan representeras som:

Tabell 3. Exempel på agentbaserad datarepresentation.

<i>Löp_nr</i>	<i>Kön</i>	<i>Bo_Zon</i>
1	1	1
2	2	1
3	2	1
4	1	2
5	1	2
6	2	2
7	2	2

I exemplet blir det en längre tabell med en individuell representation, vill man tillfoga en egenskap hos befolkningen är det relativt enkelt genom att lägga till en variabel vilket gör att storleken hos data enbart ökar linjärt. I exemplet har vi bara expanderat tabellen för att illustrera datamodellen utan att det behöver utgöras av syntetiska individer. Att tillämpa en modell på de senare data utgör ingen tekniskt problem och måttliga förändringar av modellens programkod. Vill man lägga till en egenskap i data

räcker det med en kolumn till och tabellens relationer behöver inte justeras. Detaljer hur den syntetiska befolkningen skapas återkommer vi till i rapporten.

Varför inte en riktig befolkning? Man skulle rent tekniskt kunna ta registret över totalbefolkningen och köra modellen på. Förutom uppenbara integritetsproblem skulle det inte tillföra mycket av värde så länge som syntetiseringen blivit välgjord. Ett tyngre praktiskt skäl är emellertid att vi sällan studerar nulägen utan ofta en tidpunkt långt fram, 20-30 år, och ändå inte är hjälpta av register. Vi måste därför anvisa en metod där vi utgår från Trafikverkets etablerade data för framtidsscenarioer.

Metod

I det här kapitlet beskrivs metoden att skapa en syntetisk befolkning med hushållsinformation.

Projektet är en förstudie där olika angreppssätt för att besvara frågor om hushållsbildning, hushållsinkomst och syntetisk befolkning med hushållsinformation undersökts i syfte att hitta en fungerande metod.

Metoden att tillföra information om hushåll till markanvändningsdata utförs i två steg:

1. Prognosticera demografiskt driven hushållsbildning
2. Skapa en syntetisk befolkning

Metoden baseras på data som redan nu tas fram för markanvändningen i Sampers. Det avser befolkningens fördelning på ålder och kön, vilka är styrande för hushållsbildningen. Även bostadstyp (flerbostadshus respektive småhus) tas hänsyn till i den mån det är möjligt.

För att kunna skapa syntetisk befolkning med hushållsinformation som indata till Sampers är ett första steg att ta fram prognoser för hushållsbildningen från befolkningsframskrivningar. För nuläget finns numera hushållsstatistik via lägenhetsregistret. För att skapa prognoser för hushåll var utgångspunkten initialt att använda en befintlig regional modell se om det går att kalibrera om den så den passar för hela Sverige.

Prognosticera demografiskt driven hushållsbildning

Det finns flera operationella modeller för att prognostisera den demografiskt drivna hushållsbildningen. Med demografiskt driven avser vi att det är befolkningens sammansättning och preferenser som avgör antal hushåll, inte bostadsmarknaden. Det medför att vi antar att det byggs så många nya bostäder som behövs utifrån den demografiskt drivna hushållsbildningen. Det är inte analogt med hur det ser ut idag, speciellt inte i storstäderna. Men antagandet är i linje med hur långsiktiga scenarier brukar tas fram.

Ett exempel på modell är *hushållskvoter* som Boverket brukar använda sig av i sina analyser. Nackdelen med hushållskvoter är att de endast ger antal hushåll och inte hushållsstorlekar. Därmed är modellen inte tillämplig i detta projekt. Tillväxt- och regionplaneförvaltningen, SLL brukar i bostadsbehovsprojekt använda sig av en modell som ger både antal hushåll och hushållsstorlekar. Denna modell är en så kallad *hushållsfrekvensmodell* och finns implementerad för Stockholms län. Modellen användes till exempel i arbetet med den regionala utvecklingsplanen för Stockholm, RUF 2010. Denna senare modell skulle kunna estimeras för de olika länen i Sverige och användas för att ta fram hushållsprognoser för alla Sveriges län.

Valet föll till slut på att ta fram en ny hushållsfrekvensmodell som i mångt och mycket liknar SLL:s modell men som bara tar hänsyn till ålder, kön och bostadstyp. De data som modellen behöver är antal personer per län indelade efter ålder, kön och

bostadstyp. Dessa data finns som regel framtagna på länsnivå för Trafikverkets långsiktiga scenarier.

Modellen ger hushållsstorleken samt antal vuxna respektive barn (under 20 år) i hushållet. Hur många individer i hushållet som har körkort och hur många bilar som hushållet disponerar är också viktigt. Dessa är modellresultat från en bilinnehavs- och körkortsmo­dell men är beroende av hushållets sammansättning.

Skapa en syntetisk befolkning

Från befolkningsframskrivningen och hushållsprognosen på SAMS-områden ska sedan en syntetisk befolkning skapas. Det finns vedertagna algoritmer och även färdiga programvaror för detta. I korthet går de ut på att individer dras slumpvis för att skapa en population. Populationens sammansättning jämförs sedan med olika måltal. Därefter byts personer i populationen slumpvis ut tills dess att måltalen är uppfyllda med önskad precision. För individerna i den syntetiska befolkningen ingår egenskaper för hushållet som de ingår i.

Arbetsgång

I det här avsnittet beskrivs de olika stegen för att prognosticera den demografiskt drivna hushållsbildningen och skapa en syntetisk befolkning något mer i detalj. Implementationen har skett i skript som redovisas i bilagor. Till sist beskrivs vilka resultat metoden ger och hur dessa senare används i trafikmodellerna.

Metoden att tillföra information om hushåll till markanvändningsdata utförs, som beskrivs ovan, i två steg:

1. Prognosticera demografiskt driven hushållsbildning
2. Skapa en syntetisk befolkning

När vi här går in på detaljerna delas beräkningarna in i fem moment.

1. Skapa en aggregerad befolkning efter kön, ålder och bostadstyp per trafikzon
2. Bilda antal hushåll efter storlek för varje zon
3. Villkora de socioekonomiska egenskaper i zonerna utifrån befintliga markanvändningsdata.
4. Dra slumpvis individer som uppfyller de socioekonomiska egenskaperna.
5. Bryt ut individerna och låt de behålla hushållsinformation

De tre första stegen motsvarar att prognosticera den demografiskt drivna hushållsbildningen, medan steg fyra och fem skapar den syntetiska befolkningen. De fem stegen beskrivs nedan:

1) Skapa en befolkning efter kön, ålder och bostadstyp

Utgångspunkten är en delmängd av de demografiska tabeller som används i nuvarande version av Sampers dvs. ålder och kön. Dessa tas regelbundet fram för både basår och prognosår. Ett tabellskal visas i nedan.

I prognoserna på trafikzonsnivå tar man hänsyn till bostadsstockens sammansättning i form av bostadstyp och byggår (en delmängd av informationen i TYKO)⁵. Befolkningen är således konsistent med områdets karaktär och hushållsammansättning. Utifrån bostadsyta per bostadstyp (småhus och flerbostadshus) och byggår samt genomsnittlig boendtäthet per bostadstyp, byggår, ålder och kön kan befolkningen i varje SAMS-område delas upp på småhus och flerbostadshus. Eftersom befolkningen i varje zon behandlas som helt skilda i en grupp som bor i flerfamiljshus och en grupp som bor i villor delas grupperna upp i separata tabeller. Tabell 4 nedan finns för både flerfamiljshus och villor. Befolkningen stäms av per ålder och kön.

Tabell 4. Befolkning per zon, ålder, kön och bostadstyp som utgör indata till hushållsbildningsmodellen.

Trafikzon	BefSum	Män			Kvinnor		
		M_age_1	M_age_2	M_age_n	K_age_1	K_age_2	K_age_n
1	Antal	Antal	Antal	Antal	Antal	Antal	Antal
2	Antal	Antal	Antal	Antal	Antal	Antal	Antal
3	Antal	Antal	Antal	Antal	Antal	Antal	Antal
n	Antal	Antal	Antal	Antal	Antal	Antal	Antal

Befolkningen behöver inte vara indelad efter samma klasser som de som används i trafikmodellen utan den är i första hand avsedd för att utgöra indata till hushållsbildningsmodellen.

2) Bilda antal och storlek på hushåll

Hushållsbildningsmodellen bygger på statistik över hushållsbildningen per län från SCB. Statistiken ger sannolikheter för en person att ingå i hushåll av olika typ. Sannolikheterna varierar med ålder och kön på individen och hushållstyperna delas in i bostadstyp, storlek och antal personer under 20 år.

Hushållen delas upp med ledning av antal vuxna respektive barn:

Vuxna: 1, 2, 3, 4, 5+

⁵ TYKO (typkodsområden) är en klassning där fastigheter som har likartad befolkningsstruktur och befolkningsutveckling har förts samman till homogena områden. Klassificeringen görs utifrån bebyggelsetyp (småhus, bostadsrättslägenhet, hyreslägenhet, byggnadsår och värdenivå).

Barn (under 20 år): 0, 1, 2, 3, 4+; För transportmodellen spelar det ingen större roll⁶ hur många barn det är i hushållet, bara att det är barn.

Totalt 15 olika alternativ av hushåll för småhus respektive flerbostadshus (se Bilaga 1 för komplett lista).

Över tid har det visat sig att sannolikheterna att ingå i olika typer av hushåll är relativt stabila. Nulägets andelar kommer att användas, dessa bör beräknas baserat på statistik för de senaste 3-4 åren.

Centrala indata för såväl trafikefterfrågemodellen som hushållsbildningsmodellen är fördelningen på villor och flerbostadshus. Data i tabellen nedan görs för båda bostadstyperna.

Beräkningsgång i hushållsbildningsmodellen är att:

- a) Beräkna sannolikheten att tillhöra en viss hushållstyp (i) per åldersgrupp (a), kön (s) och bostadstyp (b).

$$P(i|a, s, b)$$

- b) Beräkna antal personer som tillhör hushåll (h) per typ efter ålder och kön: ${}^i h_i^{a,s,b} = P(i|a, s, b) \times n^{a,s,b}$, genom att multiplicera med befolkningen (n) per segment. Informationen i det här steget är tillräcklig för att skapa en syntetisk befolkning.

- c) Summera till antal personer som tillhör hushåll av typen i : ${}^i h_i^b = \sum_{a,s} {}^i h_i^{a,s,b}$, notera att vi har fortfarande inte antal hushåll utan bara personer med hushållsinformation⁷. Från dessa data skulle vi kunna syntetisera en population men med enkla medel går det att ta ett steg till.

- d) Beräkna antal hushåll: $h_i^b = \frac{{}^i h_i^b}{{}^i h_i^n}$, genom att dividera med antalet personer i hushållskategorin. Dvs. $h_i^n = 2$ för hushåll med två personer.

Beräkningen avser en zon så det hela får upprepas för samtliga zoner. Det förväntade resultatet ger en tabell som nedan med antal hushåll efter vuxna och barn per zon.

Tabell 5. Tabell med antal hushåll per zon efter typ där kolumnhuvudena avser: Hushåll_[antal personer]_[antal barn]. Observera att endast de första sju hushållstyperna (av 15) i tabellen visas.

Trafikzon	Hh10	Hh20	Hh21	Hh30	Hh31	Hh32	Hh40
1	antal	antal	antal	antal	antal	antal	antal
2	antal	antal	antal	antal	antal	antal	antal
3	antal	antal	antal	antal	antal	antal	antal
n	antal	antal	antal	antal	antal	antal	antal

⁶ Skulle dock förbättra beskrivningen av ärendet skjutsa och resor till service, barnomsorg och vård om vi hade information om 0-6 åringar.

⁷ Här är vi egentligen klara med vad vi har lovat att genomföra enligt ansökan, vi har hushållsinformation knutet till individen som motsvarar informationen i tabell 1 samt information om antalet barn.

I det här beräkningssteget har vi antal hushåll men inga egenskaper hos hushållen som inkomster eller en demografi inom hushållen. I tillvägagångssättet har vi inte slagit samman individerna i hushåll utan bara skapat antal hushåll. Från tabellen skapas en lista över samtliga hushåll i Sverige. De egenskaper som i detta steg kan tillskrivas hushållen är vilket SAMS-område de ligger i, vilken typ av hushåll det är och vilken bostadstyp det rör sig om.

3) Villkora de socioekonomiska egenskaper i zonerna

Sedan tidigare finns data/prognoser för inkomster, demografi och boendeform per zon. Den befolkning som skapas ska även fortsättningsvis representera samma demografiska sammansättning som de hittillsvarande tabellerna. Det innebär att agenterna ska bygga upp samma inkomstfördelning och samma demografi. De dimensioner som bygger upp det socioekonomiska variabelerna finns i tabellen nedan.

Tabell 6. Socioekonomiska egenskaper som ligger till grund för syntetiseringen.

<i>Egenskap</i>	<i>Antal klasser</i>	<i>Geografisk nivå</i>
Inkomst	5	Trafikzon
Ålder	8	Trafikzon
kön	2	Trafikzon
Bostadstyp	2	Trafikzon
Förvärsarbetande efter ålder och kön	5*2	Trafikzon

Det inkomstbegrepp vi vill använda är hushållsinkomst⁸. De data som hittills har tagits fram avser individinkomst.

Ta fram agenter som per zon uppfyller de socioekonomiska egenskaperna.

4) Fyll hushållen med agenter som uppfyller de socioekonomiska egenskaperna.

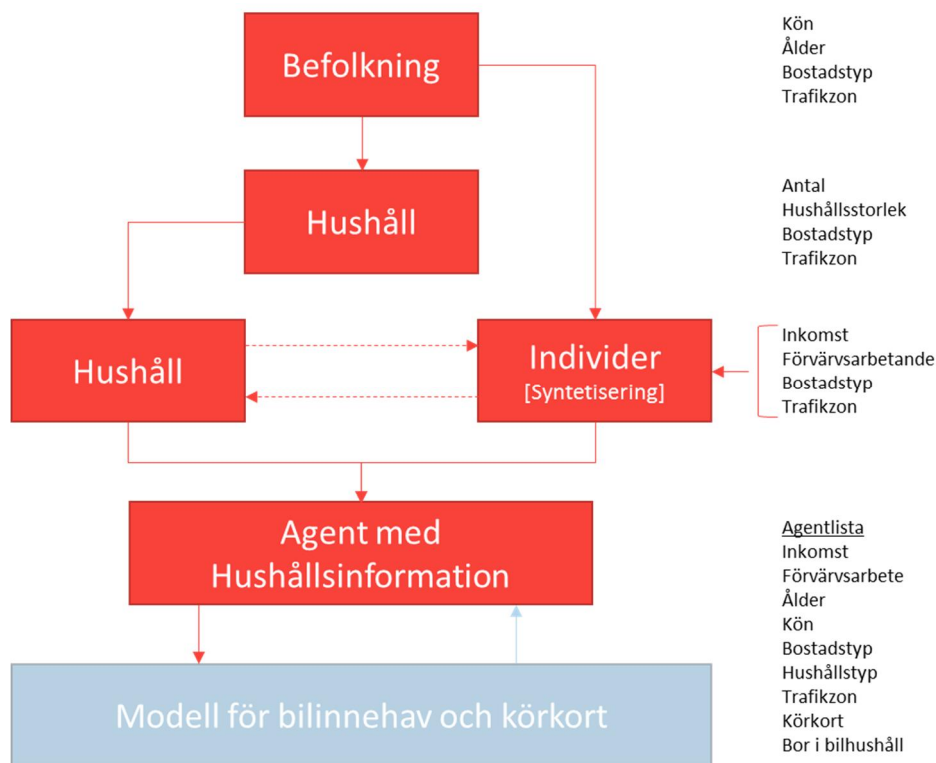
I det här steget sker en syntetisering av hushållen genom att dra hushåll till zonen.

5) Bryt ut agenterna och låt dom behålla HH-information?

Efter syntetiseringen finns agenter kopplade till hushållen och det är möjligt att koppla de syntetiska hushållens data till agenten.

I figur 1 nedan visas flöden av data och tillgänglig information i varje steg.

⁸ Se separat PM som togs fram i samband med omskattningen av Sampers: "Pengar i Sampers" som beskriver hur inkomster och kostnader hanteras i Sampers.



Figur 1. Arbetsflöde och resultat. Röda pilar avser data som skapas innan trafikmodellen körs medan den grå pilen är data som agenterna tilldelas i samband med modellkörningen.

Implementation av metod

Samtliga beräkningssteg är programmerade i makron så att vi kan göra om beräkningarna enkelt om vi ändrar något i hushållslistan. Makrona är av karaktären forskarhack och finns tillgängliga för granskning men bör utvecklas och kvalitetssäkras innan de används för produktion av data av andra än utvecklarna.

Metod och modeller implementeras i en prototypprogramvara (programspråk GISDK) inklusive inställningar för syntetiseringen. Program levereras för granskning, inte för produktion. I bilaga 3 redovisas makron, tabeller för indata och tabeller för utdata.

I likhet med arbetsgången ovan har vi två steg, ett som avser hushållsbildningen och ett som avser syntetiseringen av individerna.

Syntetisering

För syntetiseringen har TransCAD:s inbyggda metod använts. Den behöver två indatafiler. Den första filen är en individfil som innehåller ett brett spektrum av individer så att de kan representera totalpopulationen. Källan för individfilen är inte central men det är viktigt att den baseras på verkliga personer så att korrelerade egenskaper följer med. Vi har använt oss av data från resvaneundersökningen. Det

är viktigt att denna fil innehåller värden för alla dimensioner som avstämnings ska göra för mot randvärdena per SAMS-område. Filen måste också innehålla observationer för alla kategorier i alla dimensioner, annars går det inte att uppfylla alla randvillkor⁹.

Den andra filen är villkorsfilen som innehåller randvillkoren för alla kategorier för de respektive dimensionerna. Dimensionerna som används är hushållstyp, ålder, förvärvsarbete eller ej och inkomstklass. Dessa dimensioner delas upp i kategorier enligt:

- Antal personer per åldersklass (0-19, resp. 20-w år), hushållstyp och bostadstyp
- Antal personer per kön, åldersklass och bostadstyp
- Antal förvärvsarbete per kön, åldersklass och bostadstyp
- Antal personer per inkomstklass och kön

Det är för dessa kategorier i respektive dimension som randvillkoren per SAMS sätts.

Förvärvsstatus kommer från tabellen SAMSSys vilket innebär att förvärvsstatus för randvillkoren följer RAMS, dvs att personen har förvärvsarbetet minst 4 timmar under november.

För att minska minnesanvändningen och beräkningstiden syntetiseringsprocessen har klassindelningen för dimensionerna aggregerats något jämfört med SAMS-databasen och hushållsprognosen.

Inkomsterna har delats in i fem klasser (per kön):

1. Personer som är 0-15 år (som inte finns med i tabellen SAMSIK)
2. Personer 16 år och äldre med årlig inkomst 0 - 1 000 kr
3. Personer 16 år och äldre med årlig inkomst 1 000 - 160 000 kr
4. Personer 16 år och äldre med årlig inkomst 160 000 - 320 000 kr
5. Personer 16 år och äldre med årlig inkomst på över 320 000 kr

Befolkningen har delats in i följande åldersklasser (per kön per bostadstyp):

1. 0-6 år
2. 7-15 år
3. 16-19 år
4. 20-29 år
5. 30-49 år
6. 50-64 år
7. 65-79 år
8. 80-w år

Förvärvsarbete nattbefolkning har delats in i följande åldersklasser (per kön per bostadstyp, numrering som i föregående lista):

⁹ Det är också uppfyllt med undantag för den yngsta gruppen, 0-6 år, där vi hittat på individer samt för den äldsta gruppen som är trunkerad i ålder.

3. 16-19 år
4. 20-29 år
5. 30-49 år
6. 50-64 år
7. 65-w år
0. Ej förvärvsarbetande (totalt)

Hushållstyperna¹⁰ för syntetiseringen är (per bostadstyp per åldersklass (0-19 år, 20-w år):

10. 1 vuxen (20-w år), 0 barn (0-19 år)
11. 1 vuxen, 1 barn
12. 1 vuxen, 2+ barn
20. 2+, 0 barn
21. 2+, 1 barn
22. 2+, 2+ barn

Resultatet från syntetiseringen blir en lista med agenter som per SAMS uppfyller randvillkoren. För nuläget (2013) innebär det att listan innehåller cirka 9,6 miljoner agenter. I bilaga 4 redovisas programmets inställningar i detalj.

Resultatet från syntetiseringen redovisas i en stor tabell, där varje agent har en rad och kolumnerna ger information om ålder, kön, hur de bor (bostadstyp, hushållstyp), inkomst, om de är förvärvsarbetande m.m.. Se Tabell 7, s.17, för exempel på utdrag av resultat. I Tabell 8, s. 18, ges en förklaring till vad de olika kolumnvariablerna representerar.

¹⁰I listan avser första position antal vuxna och andra position antal barn.

Tabell 7. Tabell med en syntetisk befolkning.

[Zone ID]	[HH ID]	[UE_n]	[WEIGHT]	[UE_n]	[HH_n]	[HH_TYP]	[HH_TYP2]	[ABHH]	[KBA]	[KBAF]	[KI]	[HH_INK]	[Villa]	[HH_N_AFB]	[PO_SEX]	[PO_AGE]	[PO_FORV]	[INK_uppdal]
1140001	1	0.9975	143861010	36145	114	10	10	2110	114	114	13	120000	1	1	1	23	1	120000
1140001	2	0.9975	113561005	10874	114	10	10	2110	114	114	14	250000	1	1	1	21	1	250000
1140001	3	0.9975	113521003	10700	114	10	10	2110	114	114	14	300000	1	1	1	22	1	300000
1140001	4	0.9975	113521003	10700	114	10	10	2110	114	114	14	300000	1	1	1	22	1	300000
1140001	5	0.9975	141251010	30576	114	10	10	2110	114	114	15	360000	1	1	1	29	1	360000
1140001	6	0.9975	123721004	21715	114	10	10	2110	114	0	13	75639	1	0	1	23	0	75639
1140001	7	0.9975	130921003	24713	114	10	10	2110	114	0	14	220000	1	0	1	28	0	220000
1140001	8	0.9975	111341019	4340	114	10	10	2110	114	0	15	390000	1	0	1	27	0	390000
1140001	9	0.9975	114651054	14640	115	10	10	2110	115	115	14	260000	1	1	1	45	1	260000
1140001	10	0.9975	113611023	10967	115	10	10	2110	115	115	14	260000	1	1	1	32	1	260000
1140001	11	0.9975	112821017	8812	115	10	10	2110	115	115	15	425000	1	1	1	31	1	425000
1140001	12	0.9975	143251025	34888	115	10	10	2110	115	115	15	400000	1	1	1	39	1	400000
1140001	13	0.9975	135051004	27817	115	10	10	2110	115	0	13	160000	1	0	1	30	0	80000
1140001	14	0.9975	112741047	8646	115	10	10	2110	115	0	13	73000	1	0	1	46	0	7000
1140001	15	0.9975	114251046	13116	115	10	10	2110	115	0	14	232009	1	0	1	40	0	232009
1140001	16	0.9975	111421045	4581	115	10	10	2110	115	0	14	166200	1	0	1	40	0	165000
1140001	17	0.9975	142251053	32802	116	10	10	2110	116	116	14	300000	1	1	1	51	1	300000
1140001	18	0.9975	143541052	35517	116	10	10	2110	116	116	15	500000	1	1	1	62	1	500000
1140001	19	0.9975	141451059	31038	116	10	10	2110	116	116	15	550000	1	1	1	63	1	550000
1140001	20	0.9975	123431018	21422	116	10	10	2110	116	116	15	340000	1	1	1	50	1	340000
1140001	21	0.9975	114171057	12861	116	10	10	2110	116	0	13	120000	1	0	1	52	0	120000
1140001	22	0.9975	123521020	21514	116	10	10	2110	116	0	14	255437	1	0	1	50	0	255437

Tabell 8. Tabell med agenter.

<i>Variabel</i>	<i>Innehåll</i>
Zone ID	Samsområde
HH ID	Hushållsidentitet
WEIGHT	Vikt
HH_TYP	Hushållstyp från HH-modell
HH_TYP2	Hushållstyp syntetisering
ABHH	Åldersklass, bostadstyp och HH typ 2
KBA	Kön, bostadstyp och åldersklass (8 kl.)
KBAF	Förvärvsarbete per åldersklass (2 kl.), kön och bostadstyp
KI	Kön och inkomstklass
HH_INK	Hushållsinkomst
Villa	Bostadstyp 1 = småhus, 2 = FF-hus
HH_N_ARB	Antal förvärvsarbetande i Hushåll
P0_SEX	Kön, 1 = man, 2 = kvinna
P0_AGE	Ålder
P0_FORV	Förvärvsstatus 1 = förv, 0 = ej förv.
INK_uppdat	Agentinkomst

Kodning av prognosresultat på agenter

I Sampers modeller för bilinnehav och körkort kommer dessa variabler att koda på agenten.

- Körkort
- Bilinnehav
- Logsummer från efterfrågemodellen som används i modellen för generering och bilinnehav. Logsummerna kan sedan användas för att göra fördelningsanalyser.

Resultat och validering

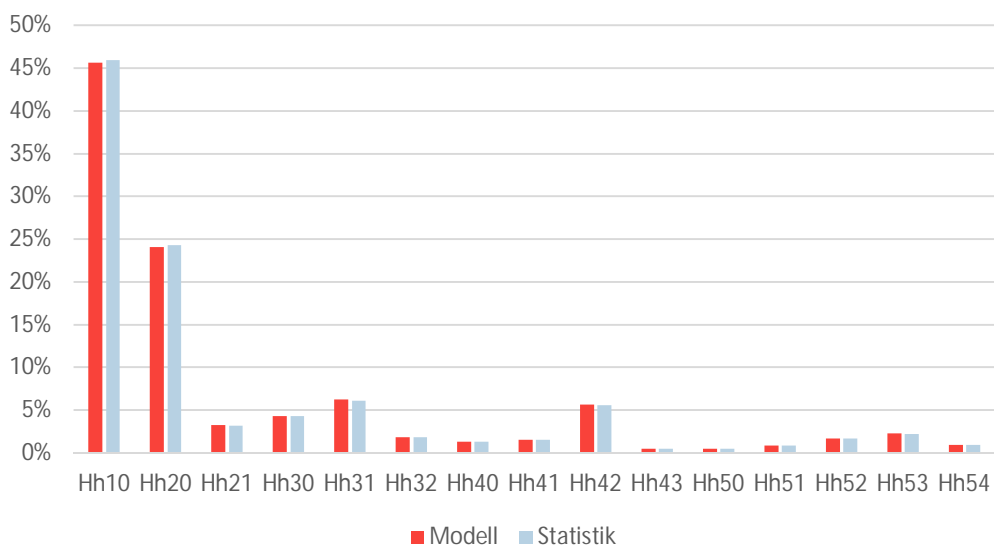
I det här kapitlet visas metodens resultat och en validering görs mot tidigare befintliga data (statistik som beskriver demografin).

Hushållsbildning

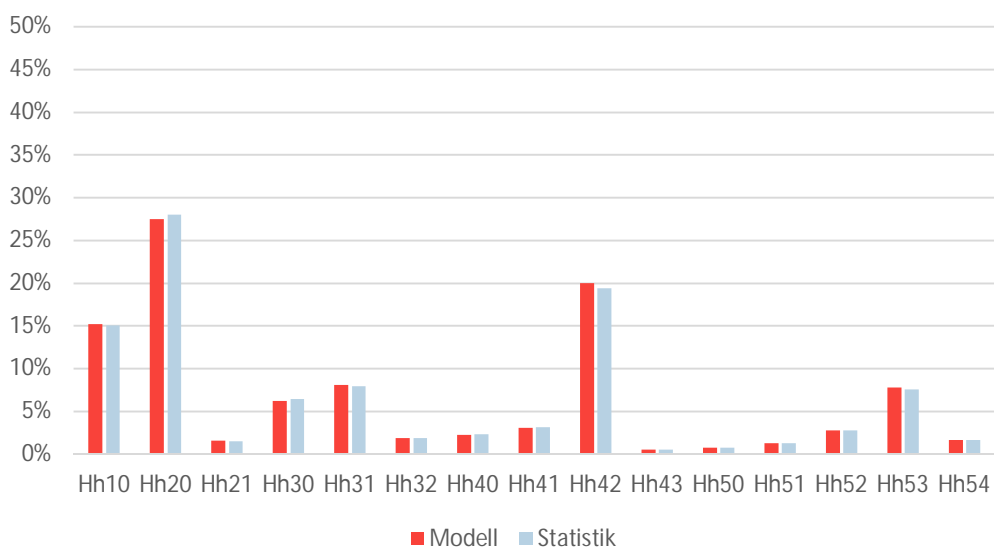
Beräkningar för hushållsbildning har gjorts för nuläget med den ovan beskrivna modellen för hushållsbildning. Data för befolkning och bostadsytor som beräkningarna utgår ifrån avser år 2013 (samma år som basprognosen från 2016 avser). Boendetätheter givet ålder, kön, bostadstyp och bostadens ålder är genomsnitt för perioden 2000-2010. Sannolikheterna för hushållsbildningsmodellen är för år 2016.

Figur 2 visar fördelning på hushållstyper i Stockholms län, modellberäknad och statistik. **Figur 3** visar motsvarande beräkningar och statistik för befolkningen fördelad på hushållstyper. Skillnaden mellan figurerna är att i den senare beaktar storleken på hushållen vid framtagande av fördelningen. Överlag är överensstämmelsen mellan beräknat utfall och statistik god. Det är inte förvånande eftersom hushålls-sannolikheterna som använts i beräkningarna bygger på samma statistik som redovisas i figurerna (dessa avser år 2016). Beräkningarna använder sig dock även av data och statistik som avser andra årtal och annan geografiskt uppdelning (så som befolkningen som avser år 2013, och parametrarna för boendetäthet som är riksgenomsnitt för perioden 2000-2010). Det är därför som överensstämmelsen mellan modell och statistik inte är helt perfekt.

Flerbostadshus: Hushåll fördelade på hushållstyp

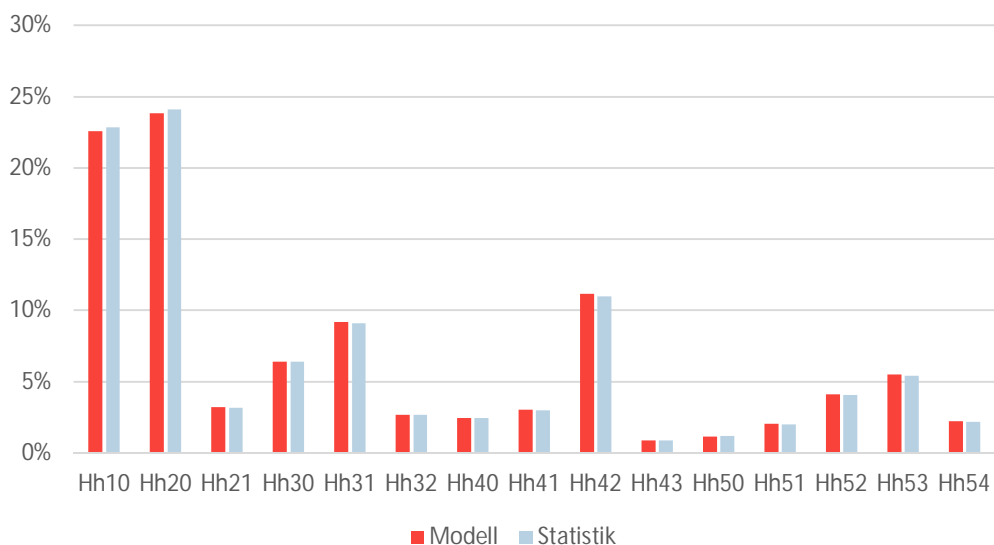


Småhus: Hushåll fördelade på hushållstyp

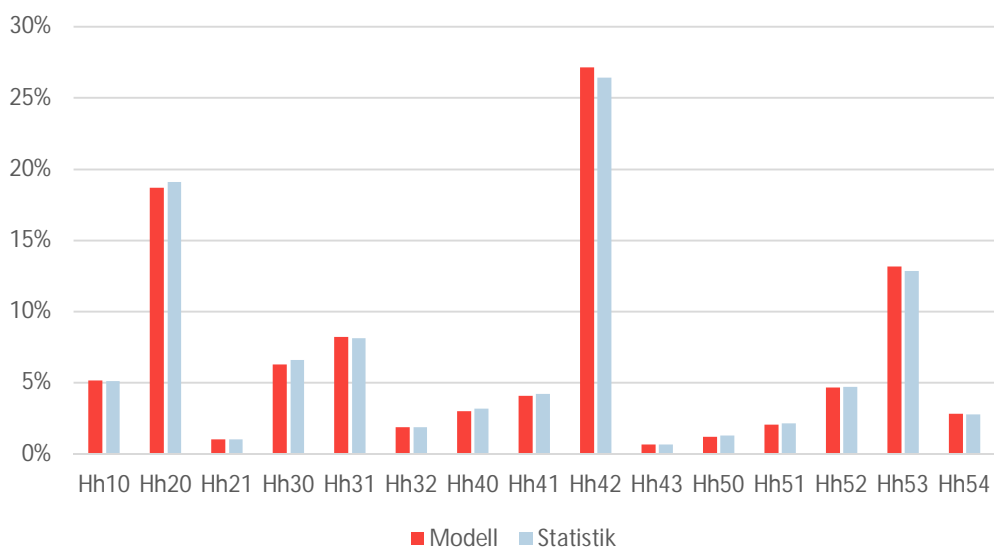


Figur 2 Hushåll fördelade på hushållstyper, jämförelse mellan modell och statistik, för Stockholms län. Flerbostadshus (överst) och småhus (underst). Hushållstypernas klasser följer modellen Hh[antal personer totalt][varav antal barn].

Flerbostadshus: Befolkningen fördelad på hushållstyp

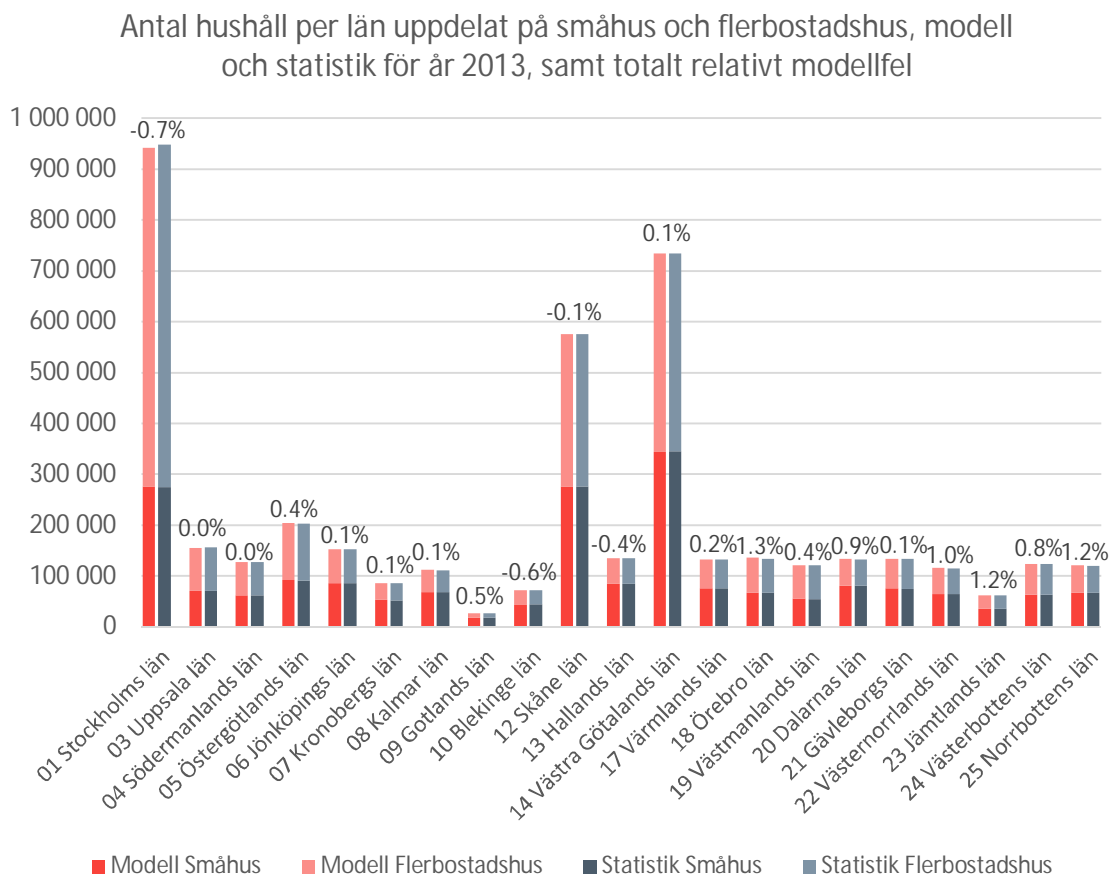


Småhus: Befolkningen fördelad på hushållstyp



Figur 3 Befolkningen fördelad på hushållstyper, jämförelse mellan modell och statistik, för Stockholms län. Flerbostadshus (överst) och småhus (underst). *Hushållstypernas klasser följer modellen Hh[antal personer totalt][varav antal barn].*

Figur 4 visar modellberäknat antal hushåll per län uppdelat på småhus och flerbostadshus för år 2013, samt statistik för detsamma. Siffrorna i figuren avser det relativa modellfelet totalt per län.



Figur 4 Antal hushåll per län uppdelat på småhus och flerbostadshus, modell och statistik för år 2013, samt totalt relativt modellfel

För år 2013 beräknar modellen för riket som helhet 2 300 för många hushåll jämfört med statistiken, vilket ger ett relativt modellfel på 0,05 %. Orsaken till att detta fel uppstår kan härledas till att sannolikheterna i hushållsmodellen är skattad på data för år 2016 och att uppdelningen av befolkningen på småhus och flerbostadshus i modellen görs utifrån bostadsarea och riksgenomsnitt för boendetätheter givet husens ålder. Användningen av riksgenomsnitt medför att uppdelningen av befolkningen på småhus och flerbostadshus per SAMS-område får små avvikelser jämfört med statistik. Dessutom bygger boendetätheterna på statistik för perioden 2000-2010, vilket inte är exakt identiskt med utfallet 2013.

Det finns kommuner som har klart större avvikelse mellan modellutfall och statistik än vad länen har. Det är förväntat eftersom läns-genomsnitt används för hushållsmodellen och det finns kommuner som skiljer sig åt mycket i de flesta län. Ett alternativ för att minska felen för enskilda kommuner kan vara att gå ifrån

länsgenomsnitt för hushållsmodellen och istället arbeta med kommungrupper, till exempel SKL:s kommungrupper. Att använda sannolikheter för hushållsbildning på kommunnivå bedömer vi inte vara ett alternativ eftersom många kommuner är för små för att det ska vara möjligt att få stabila sannolikheter.

Syntetisk befolkning

Även för den syntetiska befolkningen har beräkningar gjorts för ett nuläge (år 2013, som är basår i SAMS-databasen). Motsvarande beräkningar kan även göras för ett framtidsscenario (t.ex. år 2040 eller år 2060). I tre figurer på de kommande sidorna visas vad modellen har skattat jämfört med vad randvillkoren beskriver. Randvillkoren är hämtade från SAMS-databasen och beskriver:

- Antal personer per kön, åldersklass och bostadstyp
- Antal förvärvsarbetande per kön, åldersklass och bostadstyp
- Antal personer per inkomstklass och kön

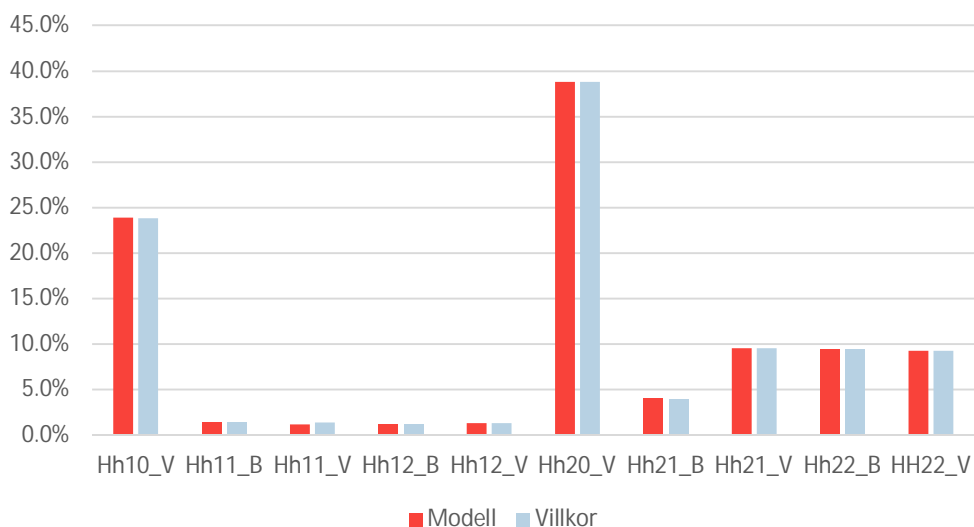
Vilket kombineras med den modellerade informationen om hushållsbildning på SAMS-områdesnivå (vilket beskrivs ovan under rubriken "hushållsbildning"):

- Antal personer per åldersklass (0-19, resp. 20-w år), hushållstyp och bostadstyp

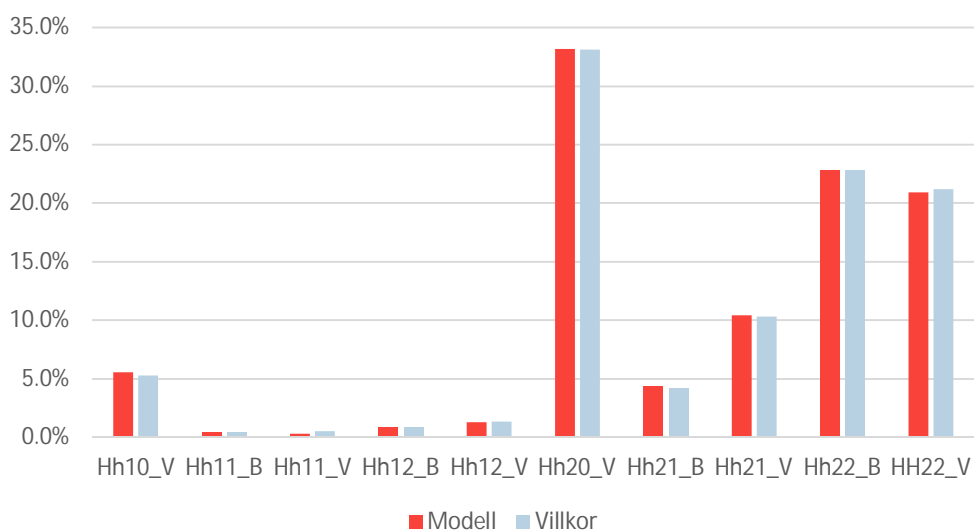
Figurer visas nedan vilka illustrerar befolkningen fördelad på hushållstyp, åldersklass respektive inkomstklass och kön för ett enskilt SAMS-område. Beräkningar har gjorts för samtliga SAMS-områden i hela Sverige, men kan av naturliga skäl inte redovisas i figurer i en rapport. Resultatet erhålls i en stor matris (se Tabell 7, s. 17).

Samtliga figurer visar en god överensstämmelse mellan modell och randvillkor. När data på enskilda SAMS-områden bryts ner på bostadstyp och t.ex. åldersklasser blir det en större relativ differens mellan modellerad syntetisk befolkning och vad randvillkoren visar. Det beror av att urvalsmängden i en sådan nedbrytning kan vara liten, och en liten absolut differens blir lätt en stor relativ differens.

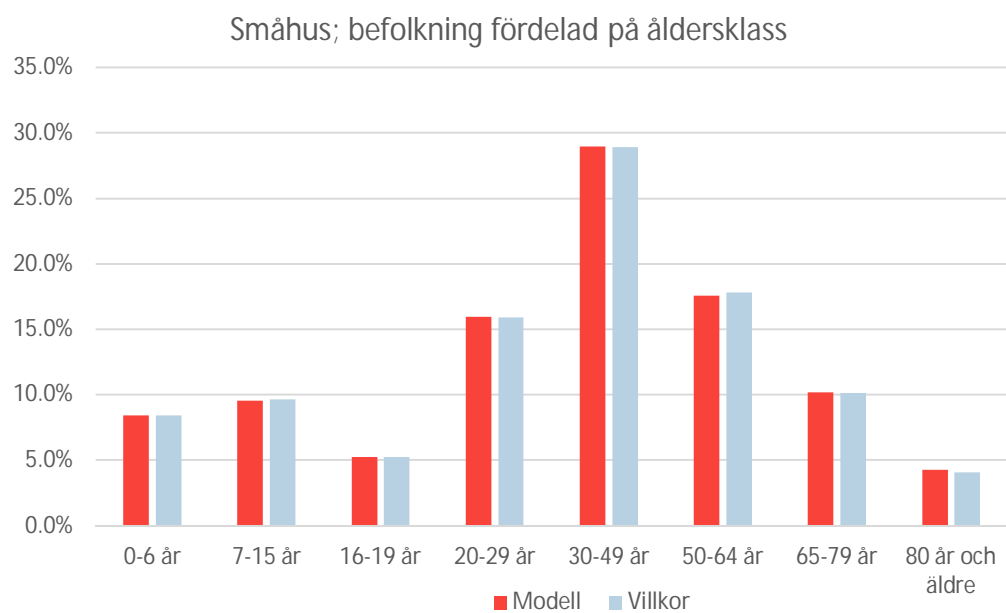
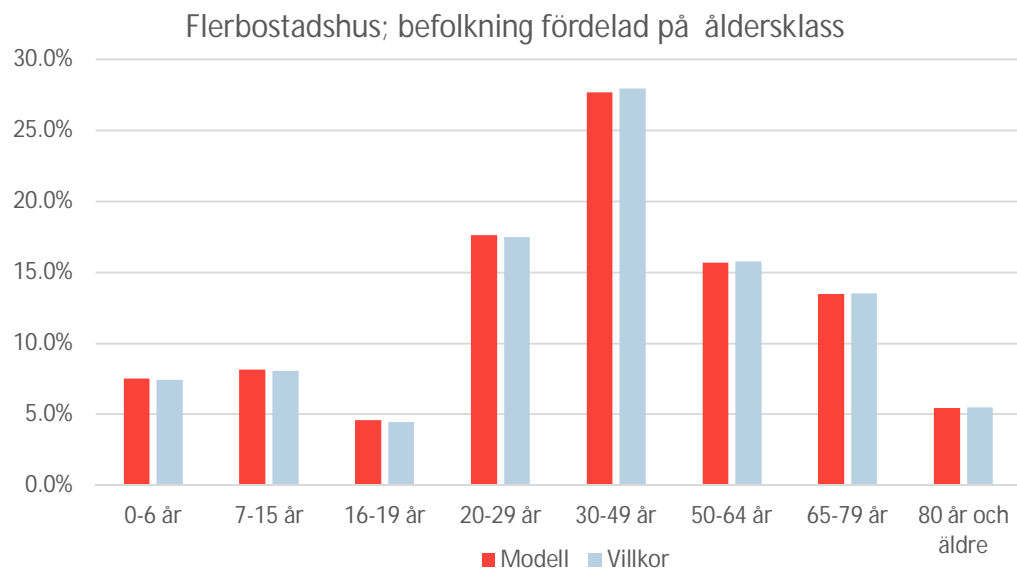
Flerbostadshus; befolkningen fördelad på hushållstyp



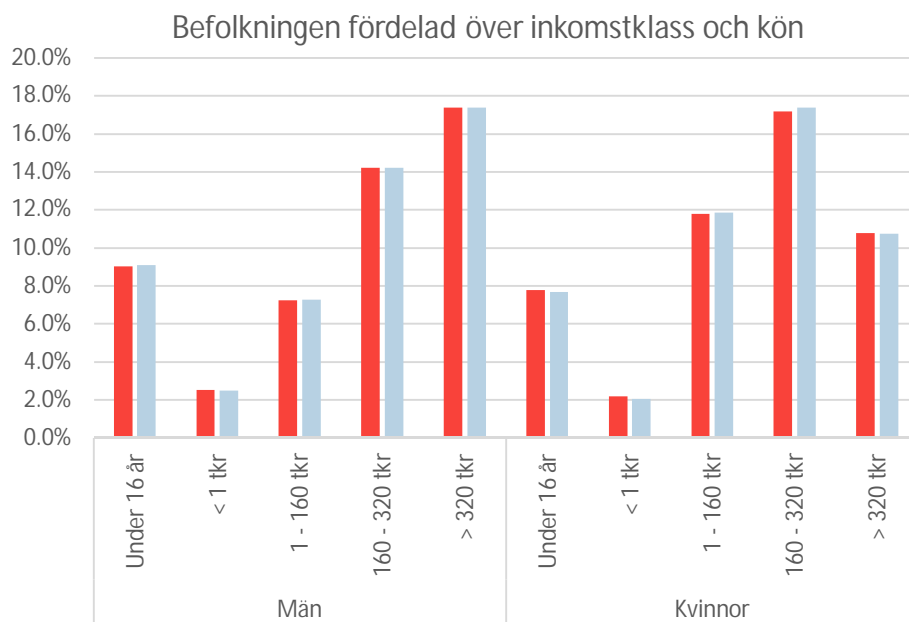
Småhus; befolkningen fördelad på hushållstyp



Figur 5 Befolkningen fördelad på hushållstyp och barn (< 20 år, markerade _B) respektive vuxna (>20 år, markerade _V), jämförelse mellan modell och villkor, för SAMS 114001. Flerbostadshus (överst) och småhus (underst). Hushållstypernas klasser följer modellen Hh[antal vuxna personer][antal barn].



Figur 6 Befolkningen fördelad på åldersklasser, jämförelse mellan modell och villkor, för SAMS 114001. Flerbostadshus (överst) och småhus (underst).



Figur 7 Befolkningen fördelad på inkomstklasser och kön, jämförelse mellan modell och villkor, för SAMS 114001. Flerbostadshus (överst) och småhus (underst).

Avslutande kommentarer

Arbetet med data syftar till att göra en kvalitativ förändring av indata till transportmodellerna där vi går från att tillämpa modellerna på aggregerade medelvärden för zoner till agenter med hushållsinformation. Initiativet till förändringen har sitt ursprung i observationen att hushållsinformation visade sig tillföra mycket för att förklara innehav av bil och resegenerering. Vi har också tidigare sett problem med att använda medelinkomster att beräkna bilkonkurrens på zonnivå. Medelvärden av zonegenskaper är inte optimalt för att förklara resebeteenden. Medelinkomsten i en zon kan mycket väl vara fullt tillräcklig för att ge ett högt bilinnehav medan det bakom medelvärdet döljer sig en fördelning av inkomster som kanske ger en fördelning på hushåll med höga inkomster med flerbilshushåll och hushåll med låga inkomster som normalt ger mycket lågt bilinnehav. Att arbeta med fördelningar ställer prognosmakaren inför både möjligheter och svårigheter. Normalt antar man oförändrade fördelningar i prognosen men möjligheten finns också att variera fördelningarna vilket tillför ett verktyg i analysen.

Kommer det att skilja i svårighetsgrad när man vill göra känslighetsanalyser med en syntetisk befolkning? Svaret kan variera. I nuvarande Sampers är det möjligt att justera realinkomstutvecklingen med en parameter och det skulle vara möjligt att göra samma sak här och bara skriva upp (eller ner) inkomsten per agent med en faktor. Problemet med partiella ändringar av enskilda ekonomiska variabler är emellertid att normalt är en högre realinkomstutveckling i ekonomin förknippad med andra förvärvsgrader och ökningen hör samman med att vissa agenter går från arbetslöshet till arbete och inte en högre ökningstakt av lön. I Nuvarande hantering saknas all form av konsistens eftersom inkomsterna kan ökas utan att det varken påverkar arbetskraftsdeltagande eller bilinnehav. Med en syntetisk befolkning ställs frågan mer på sin spets och vi får en direkt inkonsistens i agentlistan där inkomst och arbetskraftsdeltagande inte längre är korrelerat.

Vill man ändra något avseende befolkningen, som exempelvis inkomsterna, bör man göra en förnyad syntetisering mot villkor som har en intern konsistens dvs. där inkomster och arbetskraftsdeltagande är korrelerade på individnivå. Det innebär att med den ökade inkomsten kommer också en arbetsresa. För att lösa detta kan Trafikverket tillhandahålla alternativa agentlistor för de vanligaste alternativscenarierna man vill göra känslighetsanalyser för.

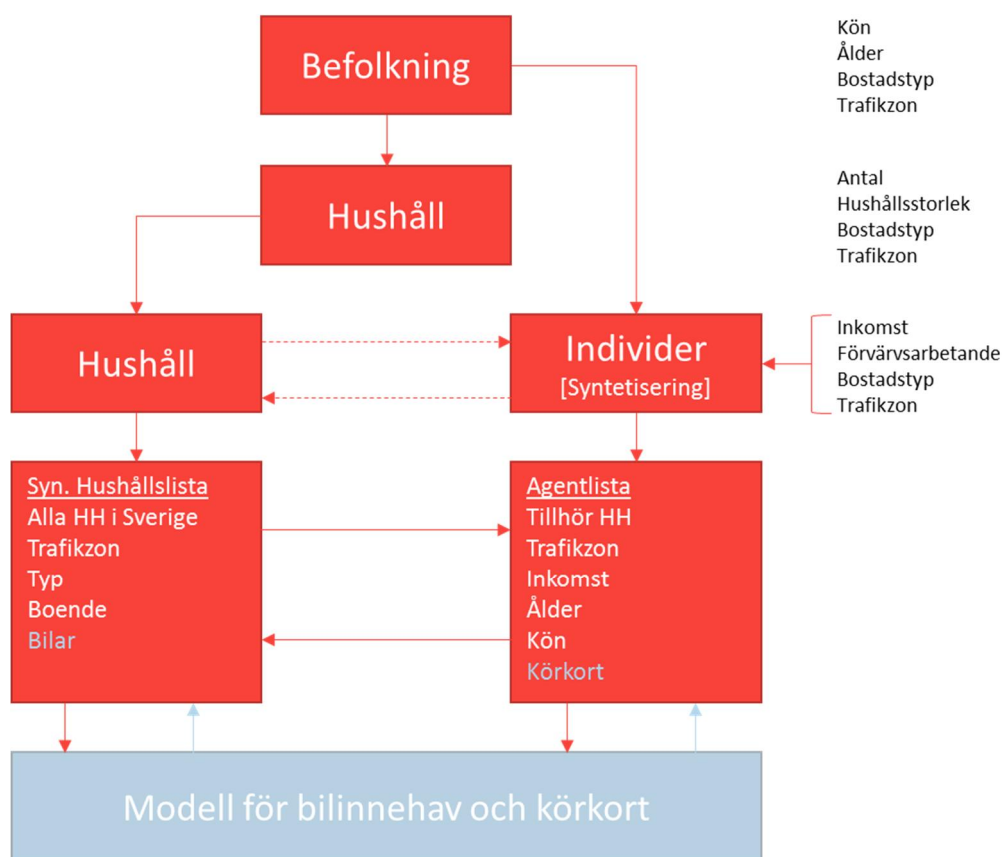
Vad innebär det här arbetssättet i tillkommande arbetsinsats? Inte någon dramatisk skillnad, det som görs redan nu för att ta fram SAMS-baser måste göras även i fortsättningen. SAMS-baserna är utgångspunkten för modellerna för hushållsbildning och som randvillkor för syntetiseringen. Det tillkommer naturligtvis arbetsmoment när data ska förfinas och vi beräknar att det som mest handlar om två veckor i kalendertid¹¹ inklusive kvalitetskontroll och syntetisering. Arbetsinsatsen i timmar beror på hur utvecklade program vi tar fram för att göra de steg som kommer efter framtagandet av SAMS-baserna. Under utvecklingsarbetet har vi gjort enkla program

¹¹ Avser en upplärd person.

som behöver utvecklas för att fungera som redskap i produktion av utomstående utförare.

Den datastruktur och det arbetssätt som presenterats är väl anpassat för att användas i exempelvis en storstadsmodell (IHOP) och kan även fungera bra i en långväga modell.

Slutligen en justering som är möjlig att göra i framtida arbete. Det gäller kopplingen mellan hushåll och agent som kan utvecklas. Nu har vi information om hur hushållet som agenten tillhör ser ut men vi har inte skapat hushåll med agenter i. Det är en utveckling som det går att jobba vidare med vilket skulle öppna möjligheter för ytterligare mer förfinade modeller. I figuren nedan visas en struktur av den typen (jämför med figur 1 ovan).



Figur 8. Tänkbar vidareutveckling.

Bilaga 1 Samtliga hushållstyper

Tabellen visar hushållstyper för hushållsprognosen och de aggregerade hushållstyper som använts för att skapa den syntetiska befolkningen

<i>Hushållstyp</i>		<i>Hushållstyp syntetisering</i>	
Hh10	1 pers, varav 0 pers 0-19	Hh10	1 vuxen, 0 barn
Hh20	2 pers, varav 0 pers 0-19	Hh20	2+ vuxna, 0 barn
Hh21	2 pers, varav 1 pers 0-19	Hh11	1 vuxen, 1 barn
Hh30	3 pers, varav 0 pers 0-19	Hh20	2+ vuxna, 0 barn
Hh31	3 pers, varav 1 pers 0-19	Hh21	2+ vuxna, 1 barn
Hh32	3 pers, varav 2 pers 0-19	Hh12	1 vuxen, 2+ barn
Hh40	4 pers, varav 0 pers 0-19	Hh20	2+ vuxna, 0 barn
Hh41	4 pers, varav 1 pers 0-19	Hh21	2+ vuxna, 1 barn
Hh42	4 pers, varav 2 pers 0-19	Hh22	2+ vuxna, 2+ barn
Hh43	4 pers, varav 3 pers 0-19	Hh12	1 vuxen, 2+ barn
Hh50	5+ pers, varav 0 pers 0-19	Hh20	2+ vuxna, 0 barn
Hh51	5+ pers, varav 1 pers 0-19	Hh21	2+ vuxna, 1 barn
Hh52	5+ pers, varav 2 pers 0-19	Hh22	2+ vuxna, 2+ barn
Hh53	5+ pers, varav 3 pers 0-19	Hh22	2+ vuxna, 2+ barn
Hh54	5+ pers, varav 4 pers 0-19	Hh12	1 vuxen, 2+ barn

Förutom dessa finns i statistiken även hushållstypen övriga. Dessa bortses från i beräkningarna. Cirka 1,5 % av Sverige befolkning 2016 bodde i ett "övriga-hushåll".

Bilaga 2 Åldersklasser

Tabellen visar åldersklasserna i SAMS-databasen och de aggregerade åldersklasser som använts för att skapa den syntetiska befolkningen.

[Åldersklasser]	[Åldersklass syntetisering]
0_6	0_6
7_12	7_15
13_15	7_15
16_17	16_19
18_19	16_19
20_24	20_29
25_29	20_29
30_34	30_49
35_39	30_49
40_44	30_49
45_49	30_49
50_54	50_64
55_59	50_64
60_64	50_64
65_69	65_79
70_74	65_79
75_79	65_79
80_84	80_w
85_w	80_w

Bilaga 3 Detaljer för den implementerade metoden

Enligt notationen avser bostadstyp 1 småhus och bostadstyp 2 flerbostadshus. Hushållen är namngivna enligt formen [HhNn], där Hh = hushåll, N = total antal personer i HH, n = antal barn i HH

Hushållsbildning

Makro: BEFBOST

Program som tar en befolkning per sams, ålder och kön och delar upp den per bostadstyp (småhus, flerbostadshus).

Tabeller IN

Aldersklasser.bin: åldersklasser enligt Sampers

Samssyss2013.bin: Befolkning och förvärsarbetande nattbefolkning per SAMS, ålder och kön

smh_area_2015.bin: Småhusyta (m2) per SAMS efter bostädernas ålder (18 klasser) och SAMS

flbh_area_2015.bin: Yta (m2) flerbostadshus per SAMS efter bostädernas ålder (18 klasser)

Botat.bin: Boendetäthet per ålder, kön bostadstyp och bostadens ålder, se tabell nedan för ytterligare detaljer.

Tabell 9. Innehåll i botat.bin

Variabelnamn	Taltyp	Längd	Decimaler	Beskrivning
bostalder	Integer (4 bytes)	11	0	Bostadens byggår i 5-årsklasser
bosttyp	Integer (4 bytes)	11	0	Småhus = 1, Flerbostadshus = 2
kon	Integer (4 bytes)	11	0	Man = 1, Kvinna = 2
alder	Integer (4 bytes)	11	0	SCB-åldersklasser enligt tabell
botat	Real (8 bytes)	15	4	Pers 100m2/bostadsyta

Tabeller UT

bef_smh.bin: befolkning per kön, ålder och SAMS i småhus

bef_flbh.bin: befolkning per kön, ålder och SAMS i flerbostadshus

Dessa tabeller är indata till macro ANTHUSHALL

forv_smh.bin: förvärsarbetande nattbefolkning per kön, ålder och SAMS i småhus

forv_flbh.bin: förvärsarbetande nattbefolkning per kön, ålder och SAMS i flerbostadshus

Makro: ANTHUSHALL

Skapar antal hushåll per typ och SAMS i småhus respektive flerbostadshus.

Tabeller IN

Aldersklasser.bin: åldersklasser enligt Sampers

age_scbage.bin: åldersklasser för hushållssannolikheterna (0-19, 20-29, 30-49, 50-64, 65-w)

bef_smh.bin: befolkning per kön, ålder och SAMS i småhus

bef_flbh.bin: befolkning per kön, ålder och SAMS i flerbostadshus

Hh_slk: sannolikhet att tillhöra en viss hushållstyp, beräknas per ålder och kön för småhus och flerbostadshus

Tabeller UT

Hh_smh.bin: antal hushåll per SAMS i småhus efter hushållstyp

Hh_flbh.bin: antal hushåll per SAMS i flerbostadshus efter hushållstyp

Dessa tabeller är indata till macro HUSHALLSLISTA

Makro: HUSHALLSLISTA

Skapar en hushållslista med alla beräknade hushåll i riket.

Tabeller IN

Hh_smh.bin: antal hushåll per SAMS i småhus efter hushållstyp

Hh_flbh.bin: antal hushåll per SAMS i flerbostadshus efter hushållstyp

Tabeller UT

Hh_list.bin: lista över alla beräknade hushåll i riket, listar SAMS, bostadstyp, hushållstyp och plats för agenterna som tillhör hushållet.

Syntetisk befolkning

Beräkningssteget att skapa randvillkor per SAMS-område för den syntetiska befolkningen är programmerad i ett makro så att vi kan göra om beräkningarna enkelt om vi ändrar något i metoden. Makrot är av karaktären forskarhack och finns tillgängliga för granskning men bör utvecklas och kvalitetssäkras innan det används för produktion av data av andra än utvecklarna.

Metod och modeller implementeras i en prototypprogramvara (programspråk GISDK) inklusive inställningar för syntetiseringen. Program levereras för granskning inte för produktion.

Makro: VILLKOR

Skapar en tabell med de villkor som en syntetisk befolkning ska uppfylla.

Tabeller IN

bef_smh.bin: befolkning per kön, ålder och SAMS i småhus

bef_flbh.bin: befolkning per kön, ålder och SAMS i flerbostadshus

forv_smh.bin: förvärvsarbetande nattbefolkning per kön, ålder och SAMS i småhus

forv_flbh.bin: förvärvsarbetande nattbefolkning per kön, ålder och SAMS i flerbostadshus

Hh_smh.bin: antal hushåll per SAMS i småhus efter hushållstyp

Hh_flbh.bin: antal hushåll per SAMS i flerbostadshus efter hushållstyp

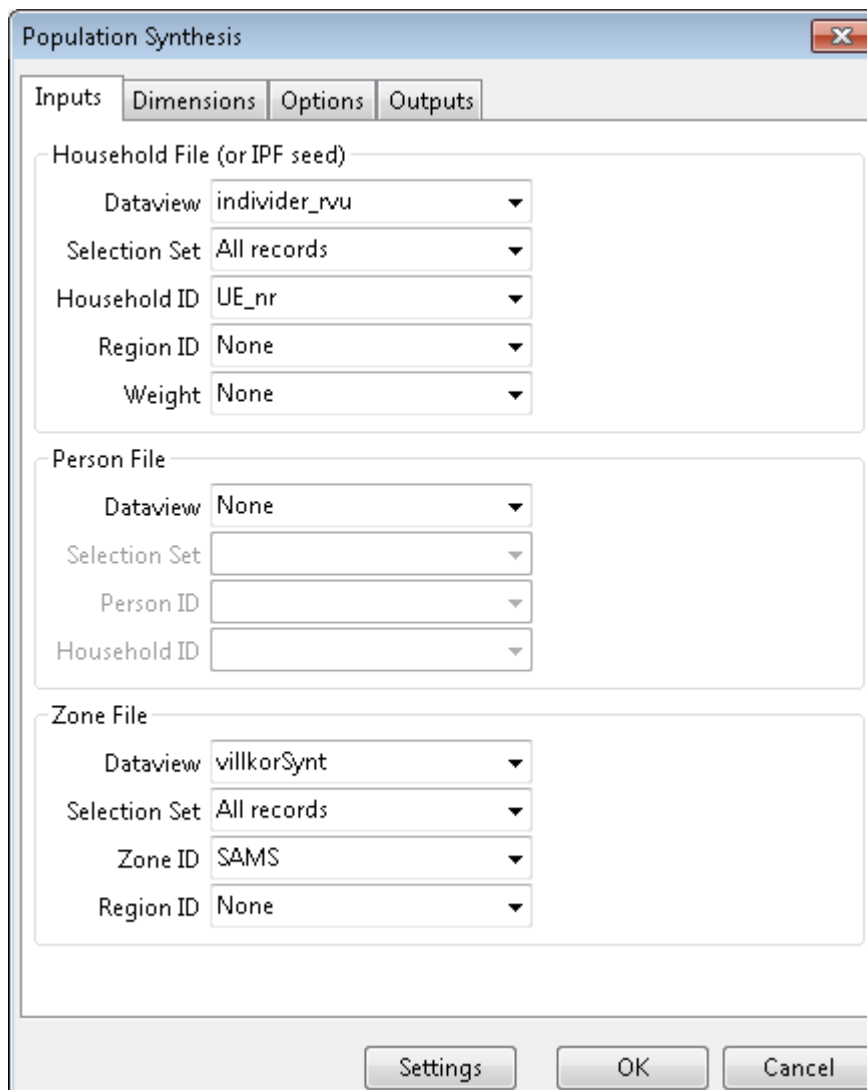
samsink2013_3a.bin: Befolkning 16-w år per SAMS och inkomstklass

Tabeller UT

villkorSynt.bin: tabell med samtliga randvillkor som används vid syntetiseringen per SAMS.

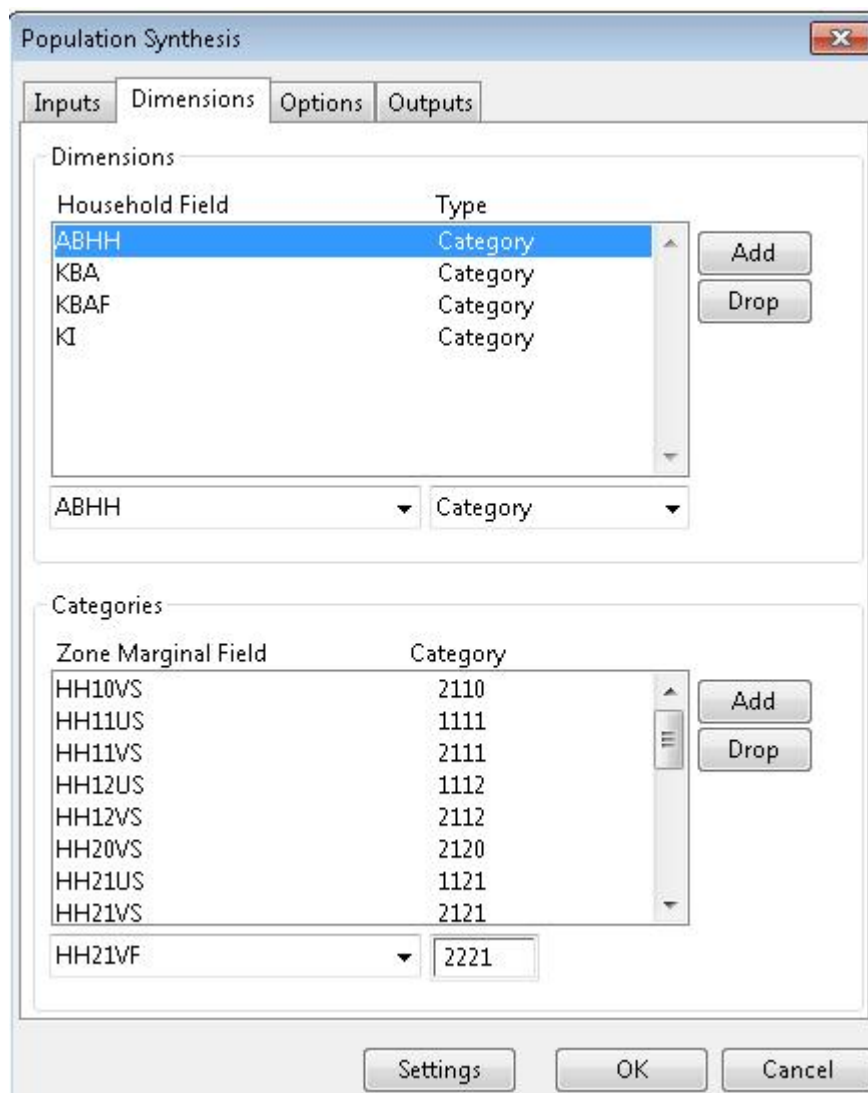
Bilaga 4 Programmens inställningar i detalj

Algoritmen för syntetiseringen nås via menysystemet i TransCAD Standard version under *Planning-Planning Utilities-Population Synthesis* vilket öppnar dialogrutan nedan. I fälten för *Household field (or IPF seed)* specificeras listan för individer från resvaneundersökningen och vilken kolumn i filen som är den unika individnyckeln. Algoritmen är egentligen tänkt för att skapa syntetiska hushåll men genom att lägga in individer istället för hushåll får vi en syntetisk befolkning.



Section	Field	Value
Household File (or IPF seed)	Dataview	individer_rvu
	Selection Set	All records
	Household ID	UE_nr
	Region ID	None
	Weight	None
Person File	Dataview	None
	Selection Set	
	Person ID	
	Household ID	
Zone File	Dataview	villkorSynt
	Selection Set	All records
	Zone ID	SAMS
	Region ID	None

Under fliken *Dimensions* ansätts vilka dimensioner och kategorier med randvillkor per SAMS som ska användas för syntetiseringen. Här används fyra dimensioner: antal personer per hushållstyp, ålder och bostadstyp (ABHH), antal personer per åldersklass, kön och bostadstyp (KBA), antal förvärvsarbetande per åldersklass, kön och bostadstyp (KBAF), antal personer per inkomstklass och kön (KI).



The screenshot shows the 'Population Synthesis' software interface. It has four tabs: 'Inputs', 'Dimensions', 'Options', and 'Outputs'. The 'Dimensions' tab is active.

Dimensions Panel:

Household Field	Type
ABHH	Category
KBA	Category
KBAF	Category
KI	Category

Buttons: Add, Drop

Selected: ABHH (Household Field), Category (Type)

Categories Panel:

Zone Marginal Field	Category
HH10VS	2110
HH11US	1111
HH11VS	2111
HH12US	1112
HH12VS	2112
HH20VS	2120
HH21US	1121
HH21VS	2121

Buttons: Add, Drop

Selected: HH21VF (Zone Marginal Field), 2221 (Category)

Bottom buttons: Settings, OK, Cancel

Population Synthesis

Inputs Dimensions Options Outputs

Dimensions

Household Field	Type
ABHH	Category
KBA	Category
KBAF	Category
KI	Category

Add Drop

KBA Category

Categories

Zone Marginal Field	Category
BefM30_49S	115
BefM50_64S	116
BefM65_79S	117
BefM80_wS	118
BefKv0_6S	211
BefKv7_15S	212
BefKv16_19S	213
BefKv20_29S	214

Add Drop

BefKv30_49F 225

Settings OK Cancel

Population Synthesis

Inputs Dimensions Options Outputs

Dimensions

Household Field	Type
ABHH	Category
KBA	Category
KBAF	Category
KI	Category

Add Drop

KBAF Category

Categories

Zone Marginal Field	Category
ForvM16_19S	113
ForvM20_29S	114
ForvM30_49S	115
ForvM50_64S	116
ForvM65_wS	117
ForvKv16_19S	213
ForvKv20_29S	214
ForvKv30_49S	215

Add Drop

ForvM16_19S 113

Settings OK Cancel

Population Synthesis

Inputs Dimensions Options Outputs

Dimensions

Household Field	Type
ABHH	Category
KBA	Category
KBAF	Category
KI	Category

Add Drop

KI Category

Categories

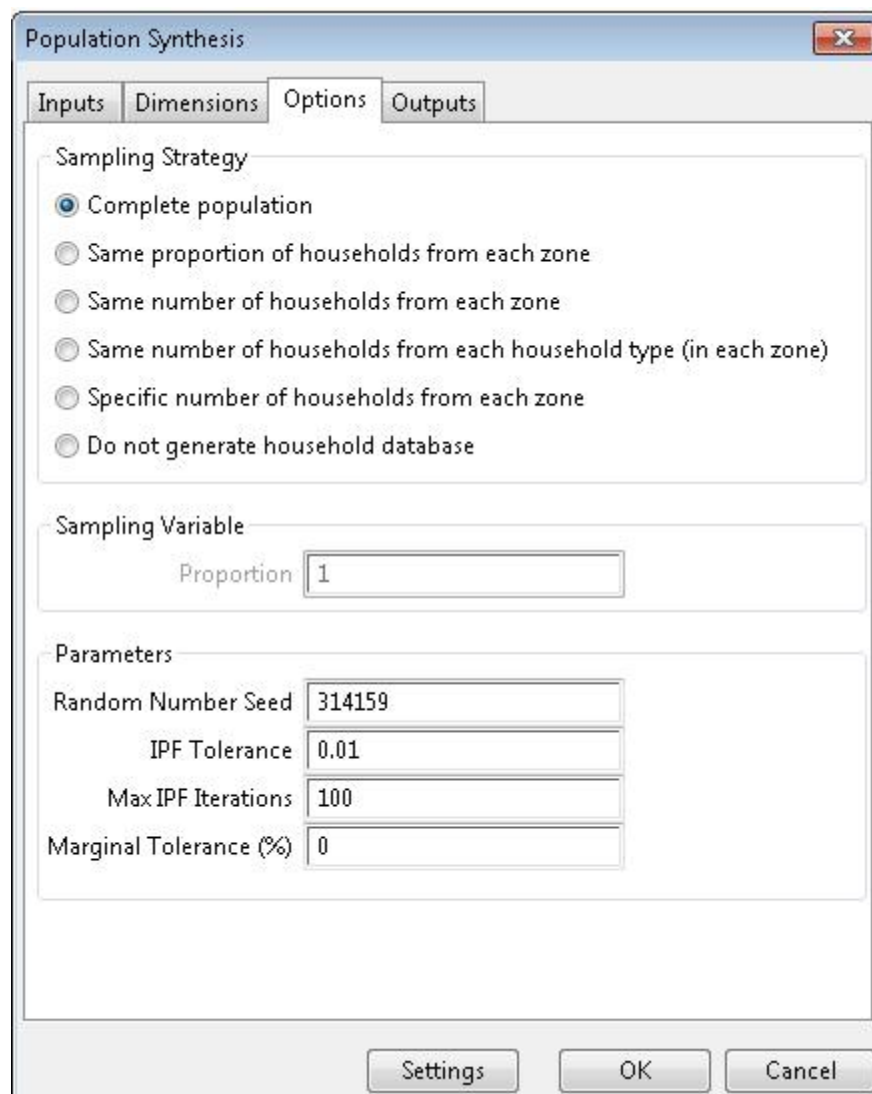
Zone Marginal Field	Category
InkM0	11
InkM1	12
InkM2	13
InkM3	14
InkM4	15
InkK0	21
InkK1	22
InkK2	23

Add Drop

InkM0 11

Settings OK Cancel

Under fliken *Options* specificeras samplingsstrategin, slumpfelsfröet och toleranser.



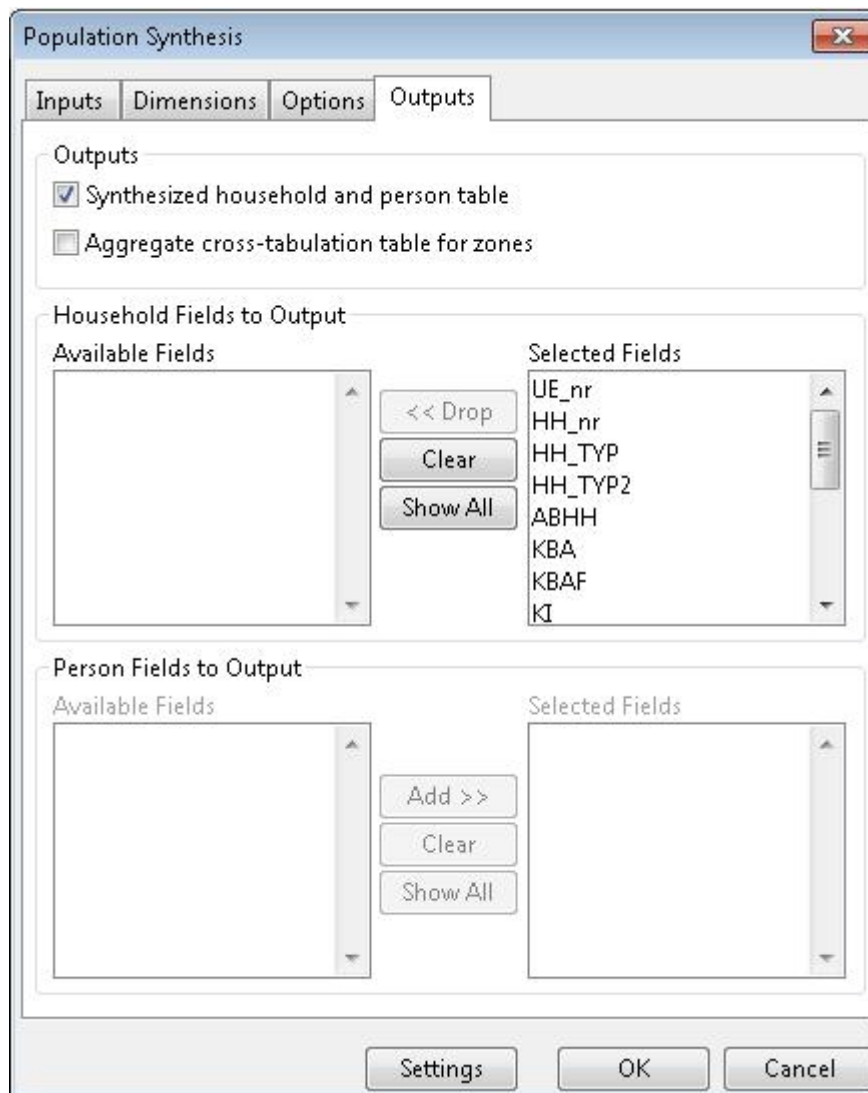
The screenshot shows the 'Population Synthesis' dialog box with the 'Options' tab selected. The dialog has four tabs: 'Inputs', 'Dimensions', 'Options', and 'Outputs'. The 'Options' tab contains the following settings:

- Sampling Strategy:** A group of radio buttons with 'Complete population' selected. Other options include 'Same proportion of households from each zone', 'Same number of households from each zone', 'Same number of households from each household type (in each zone)', 'Specific number of households from each zone', and 'Do not generate household database'.
- Sampling Variable:** A text box labeled 'Proportion' with the value '1'.
- Parameters:** A table of input fields:

Random Number Seed	314159
IPF Tolerance	0.01
Max IPF Iterations	100
Marginal Tolerance (%)	0

At the bottom of the dialog are three buttons: 'Settings', 'OK', and 'Cancel'.

Under fliken *Outputs* specificeras vilka fält ur individlistan som ska flyttas med till listan med de syntetiska individerna. Aggregerad korstabell rekommenderas inte eftersom den vid testerna blivit för stor och medfört att RAM-minnet tagit slut när TransCAD försöker öppna tabellen.



The screenshot shows the 'Population Synthesis' dialog box with the 'Outputs' tab selected. The 'Outputs' section has two checkboxes: 'Synthesized household and person table' (checked) and 'Aggregate cross-tabulation table for zones' (unchecked). Below this are two sections for field selection: 'Household Fields to Output' and 'Person Fields to Output'. The 'Household Fields to Output' section has an 'Available Fields' list (empty), a 'Selected Fields' list containing 'UE_nr', 'HH_nr', 'HH_TYP', 'HH_TYP2', 'ABHH', 'KBA', 'KBAF', and 'KI', and buttons for '<< Drop', 'Clear', and 'Show All'. The 'Person Fields to Output' section has an 'Available Fields' list (empty), an empty 'Selected Fields' list, and buttons for 'Add >>', 'Clear', and 'Show All'. At the bottom are 'Settings', 'OK', and 'Cancel' buttons.