

# Så används nedbrutna socioekonomiska indata i Sampers

Staffan Algers  
140403

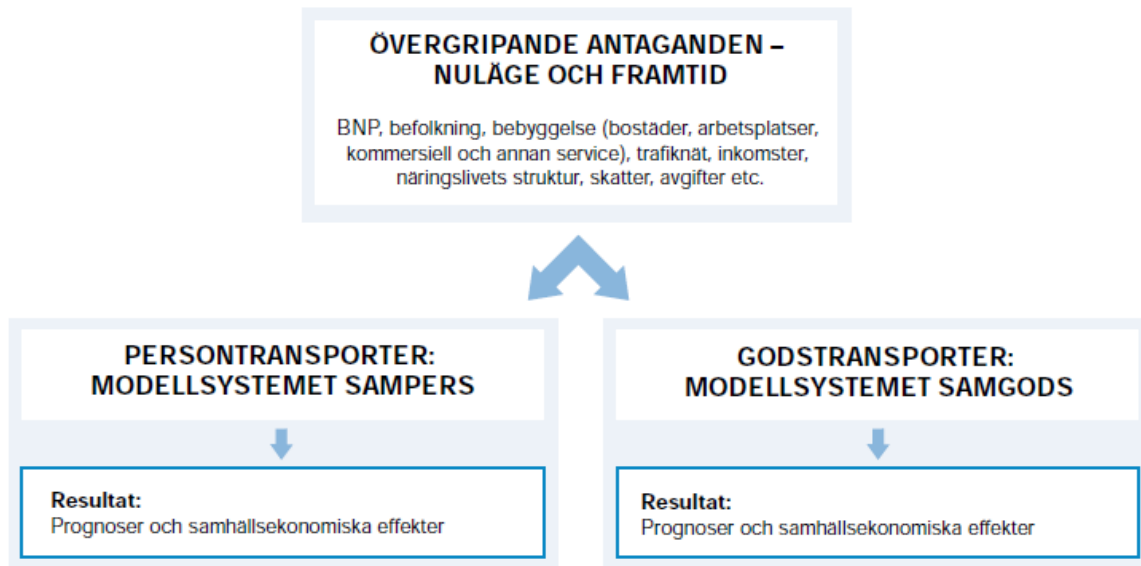
**TPmod AB**

## Inledning

Bakgrunden till denna PM är ett önskemål från Trafikverket att få en pedagogisk beskrivning av hur nedbrutna socioekonomiska indata används i Sampers<sup>1</sup>. Detta uppdrag har givits till Staffan Algiers (TPmod AB). Enligt uppdraget ska beskrivningen göras i form av en top-down ansats, där beskrivningen detaljeras efter hand. Uppdraget begränsas till hur socioekonomiska indata används i prognosmodellerna, varför Samkalk-delen inte berörs.

## Samperssystemet i stort

Samperssystemet används med Samgodssystemet för att beskriva effekterna av omvärldsförutsättningar och trafikpolitiska åtgärder på utnyttjandet av trafiksystemet i såväl fysiska (antal resor på olika detalj-nivåer) som ekonomiska termer (samhällsekonomi). Detta kan beskrivas i följande grova figur (hämtad från SIKa PM 2004:4)

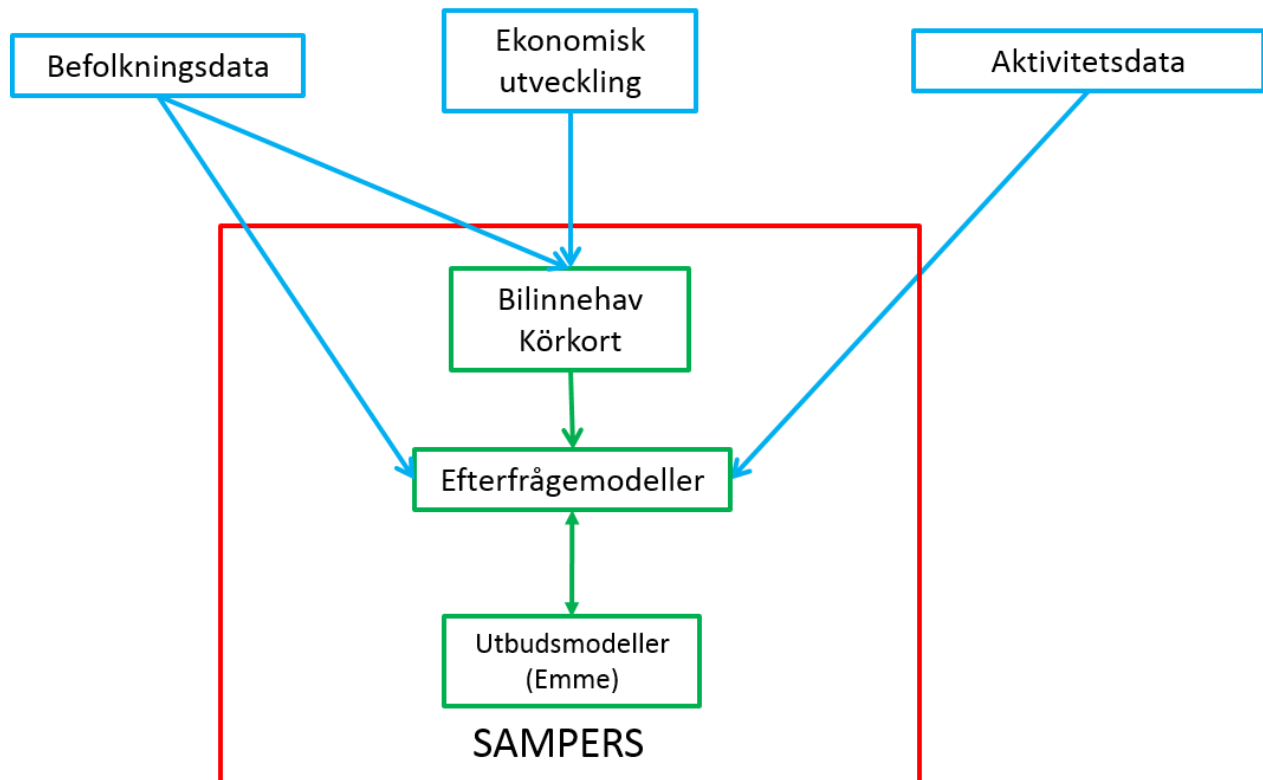


Som framgår av figuren är mycket av indata gemensamma för Sampers och Samgods. Här ska vi fokusera på Samperssystemet, och lämnar Samgods därhän. I denna beskrivning fokuseras på omvärldsdata, dvs data som påverkar resandet men som inte är en del av de trafikpolitiska åtgärder som ska utvärderas med Sampers. Detta är data över befolkningen (som genererar resandet) och aktiviteterna (som utgör målpunkter för resandet). Som ett första steg mot en mer detaljerad beskrivning redovisar vi i stort hur omvärldsdata interagerar med själva prognosmodellerna (fig 2).

I figur 2 visas de viktigaste indatakällorna tillsammans med de viktigaste delmodellerna i Sampers. Förutom befolknings- och aktivitetsdata har här också den ekonomiska utvecklingstakten tagits med, dels eftersom alla tre är relaterade till den regionalekonomiska modellen (STRAGO-rAps), dels eftersom

<sup>1</sup> Sampers version 2.9 som användes för framtagning av trafikprognoser i arbetet med åtgärdsplaneringen för 2014-2025. Texten är även giltig i fallet med beskrivningen av socioekonomiska data för de versioner av Sampers som används för Trafikverkets bastrafikprognos som ska börja gälla från 1/4-2014 och till att börja med ett år framåt. Denna översikt samt i den angivna referenser finns vid behov att tillgå via Trafikverket (dnr 2014/25100).

den har stor inverkan. Bilparksmodellen har här utelämnats, eftersom den inte mer än indirekt (via bilinnehavsmodellen) beror av befolknings- och aktivitetsdata. Vidare har trafiksystemindata utelämnats för att inte tynga framställningen i onödan. Blå färg indikerar data och databeroenden, och grön färg indikerar modeller och modellberoenden.



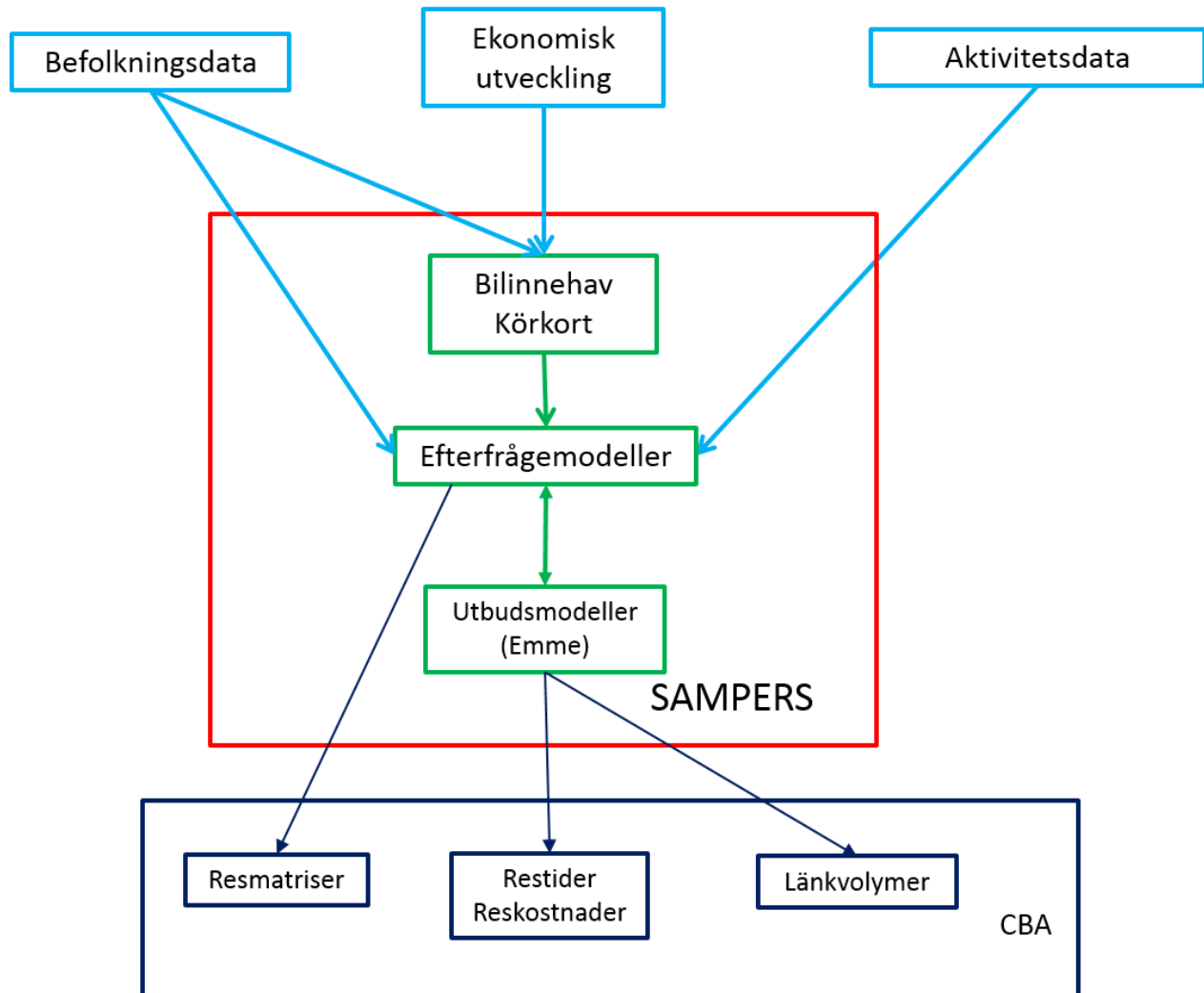
Figur 2 Schematisk bild av omvärldsindata till Sampers

Befolkningsvariabler är indata såväl i bilinnehavsmodellen som i själva efterfrågeberäkningen. Bilinnehavs- och körkortsmodellen kan sägas vara en modell för att beskriva befolkningsegenskaper som inte finns i de befolkningsprognoser som genereras till Sampers, utan som måste genereras med bilinnehavs- och körkortsmodellen för varje ny befolkningsprognos. Eftersom bilinnehavs- och körkortsmodellen är en del i Sampers beskrivs den explicit i figur 2. Hur ekonomiska data kommer in i Sampers har visats i tidigare PM<sup>2</sup> och behandlas inte vidare här.

Fig. 3 visar också de viktigaste resultaten från prognosmodellerna. Efterfrågemodellen beskriver individernas val av antal resor, färdssätt och målpunkt och resulterar i OD-matriser för respektive färdssätt (tabeller med antalet resor mellan de olika prognosområdena). Utbudsmodellen (Emme i figuren) beskriver trafikanternas ruttval i väg- och kollektivtrafiksystemet, och därmed trafiksystemets servicenivå i form av restider och reskostnader mellan de olika prognosområdena (beskrivs också i matrisform). Ett ytterligare resultat av ruttvalet är antalet resor med bil och kollektivtrafik på de olika

<sup>2</sup> Trafikverks-PM "120508 Sampdat rapport del 1 rev 120508" samt "120626 Sampdat rapport del 1 rev 120618"

väglänkarna och kollektivtrafiklinjerna. En viktig egenskap hos systemet beskrivs av den dubbelriktade pilen mellan efterfrågan och utbud. Den står för ett iterativt förfarande för att säkra konsistens mellan den servicenivå som är indata till efterfrågemodellen och den servicenivå som efterfrågan ger upphov till



Figur 3 Schematisk bild av omvärldsindata till Sampers samt prognosmodellresultat

För att tydliggöra kopplingarna på ett mer detaljerat sätt beskrivs befolkningsdata för sig och aktivitetsdata för sig.

### Befolkningsdata

Prognosberäkningen innebär egentligen att man för varje individ i varje prognosområde<sup>3</sup> beräknar sannolikheterna att välja de olika resalternativen (antal resor, färdstätt och målpunkt). Resefterfrågan fås sedan genom att summera sannolikheterna för alla individer i respektive prognosområde. Skulle alla

<sup>3</sup> Med prognosområde avses här de områden för vilka man får resultat från Sampers, vilket inte alltid är samma sak som de indataområden som man anger indata till Sampers (t.ex. befolkning).

individer i ett prognosområde vara lika skulle det endast krävas att man beräknade sannolikheterna för en individ och sedan multiplicerade dessa med antalet individer.

Efterfrågemodellerna som beskriver resebeteendet innehåller emellertid olika socioekonomiska variabler, varför man antingen måste göra beräkningar för flera olika typer av individer eller använda sig av genomsnittsvärden för de socioekonomiska variablerna. I det senare fallet gör man sig skyldig till ett aggregeringsfel som ger en viss bias varför detta bör undvikas. Fördelen med att använda genomsnittsvärden är att beräkningen går snabbare. Vid modellimplementeringen är det därför viktigt att göra en avvägning mellan aggregeringsbias och exekveringstid.

Idealt skulle man vilja ha befolkningen simultant nedbruten på alla de socioekonomiska variabler som förekommer i modellerna, eftersom dessa variabler ofta är korrelerade (rikare individer har oftare bil t.ex.). Det har dock inte varit möjligt. I Sampers används f.n. tabellerna SamsSys (ålder, kön och sysselsättning) och SamsInk (kön och inkomst) för att beskriva fördelningen på olika kombinationer av kategorier. För att beskriva fördelningen på bilinnehav och körkort används bilinnehavs- och körkortsmodellen.

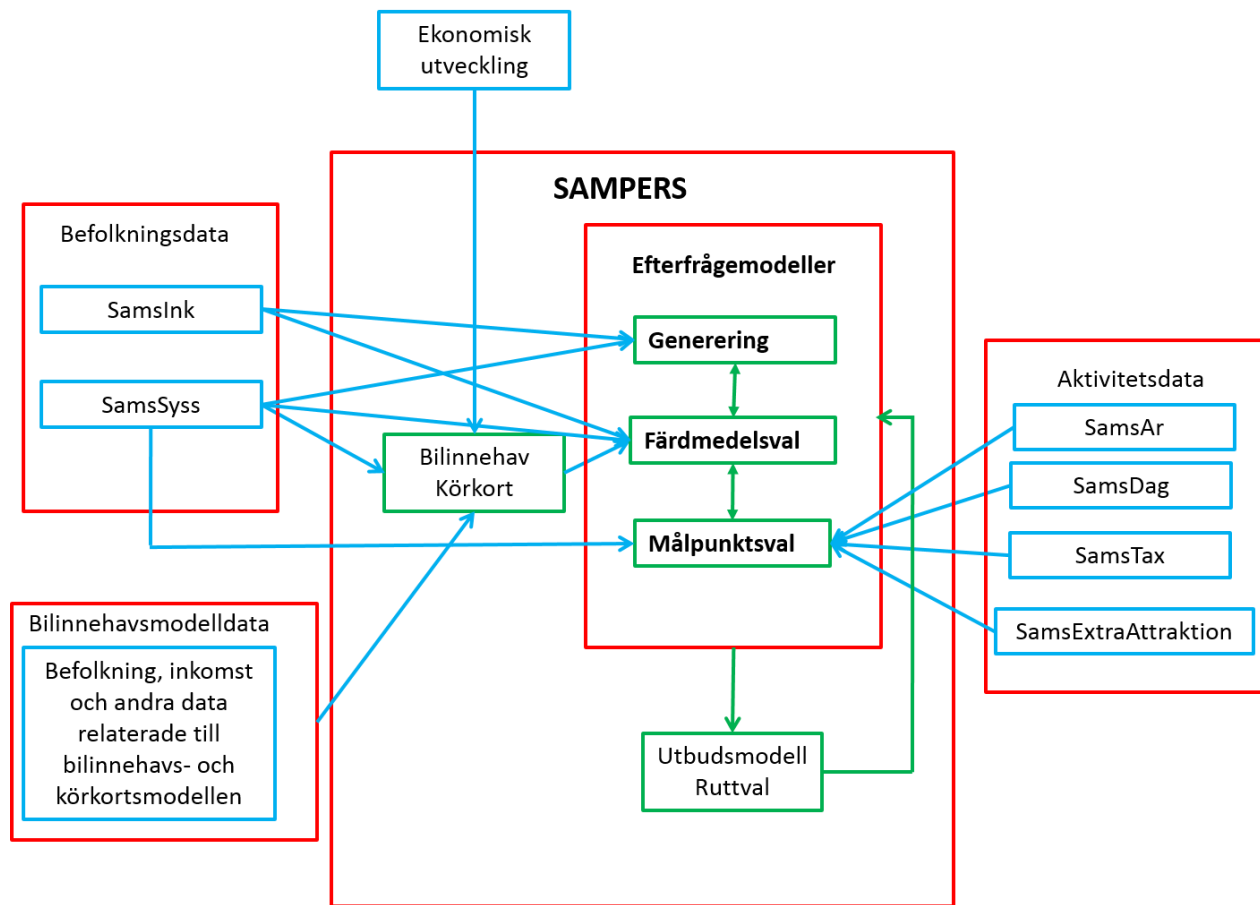
Fördelningarna på variablerna i SamsSys och SamsInk samt bil- och körkortsinnehav blir således oberoende, vilket kan ge en skev bild av verkligheten. Detta problem hanteras ibland genom att använda informationen i en resvaneundersökning som utgångspunkt och därefter för respektive prognosområde vikta om den mot prognoserade totalnivåer för de olika kategorierna. På detta sätt bevaras korrelationsstrukturen för de socioekonomiska variablerna.

Alla de aktuella variablerna har stor betydelse. Kön avspeglar beteendeskilnader mellan män och kvinnor. Dessa blir förhoppningsvis mindre i framtiden, men har idag påtaglig betydelse för flera av de olika resbesluten. Ålder påverkar också resandet i stor utsträckning, och den förväntade framtida åldersfördelningen skiljer sig påtagligt från den nuvarande. Sysselsättningen bestämmer arbetsresandet, och inkomsten påverkar känsligheten för kostnader och bilinnehavet.

Det är viktigt att upplösningen för de olika variablerna avspeglar inte bara dagens fördelning, utan också är rimlig för de prognoser som görs. Detta är f.n. ett problem för inkomstfördelningen, där fördelningen på högre inkomstklasser är för grov. Ett annat problem är att data till bilinnehavsmodellen måste genereras på annat sätt än från SamsInk- och SamsSysstabellerna.

I figur 4 visas de principiella kopplingarna mellan de olika Sams-tabellerna och Sampers delmodeller.

I verkligheten är bilden något mer komplicerad, eftersom bilinnehavsmodellen försörjs av data på en annan form än SamsSys- och SamsInk-tabellerna (bilinnehavsmodellen avser kommunnivå och använder data på kommunnivå). Detta ställer extra krav på användaren att vara observant på att använda konsistenta data för bilinnehavsmodellen och resten av Sampers.



Figur 4 Befolknings- och aktivitetsdata i Sampers

### Aktivitetsdata

Målpunkterna för resandet beskrivs av de aktiviteter som respektive målpunkt inrymmer. Det finns i princip två typer av sådana variabler – dels de som beskriver målpunktens storlek, och dels de som beskriver målpunktens kvalitet. I de nuvarande modellerna beskrivs storleken i huvudsak med antal anställda i olika näringsgrenar, och kvalitet främst med dummyvariabler för områdestyp (exempelvis vintersportanläggning). Tabellen SamsAr innehåller data om yta som tillsammans med data om antal boende och antal sysselsatta bildar täthetsmått som ingår i vissa nu befintliga modeller i Sampers. Tabellen SamsDag innehåller antalet anställda i olika näringsgrupper, medan tabell SamsExtraAttraktion främst innehåller kvalitetsvariabler. Tabellen SamsTax innehåller bl.a. fritidshusyta. Även befolkning kan beskriva ett områdes attraktivitet (för t.ex. besöksresor).

Beträffande var i processen olika data kommer in bör kanske först nämnas att efterfrågeberäkningen egentligen sker i ett steg. Det är mer relevant att tala om tre dimensioner än tre steg. Ordningen mellan dimensionerna är ett estimeringsresultat, och växlar mellan olika modeller (för färdmedels- och målpunktsdimensionerna). Oavsett var de olika variablerna beräkningsmässigt kommer in, så påverkar de alla tre dimensioner (genom modellernas logsumkopplingar).

Betydelsen av de olika variablerna varierar mellan de olika delmodellerna. Generellt styr de resandets fördelning i rummet, och är därmed centrala för beskrivningen av resandet. När det gäller exakt vilka data man behöver framöver är detta beroende av vilka modellresultat som den förestående omskattningen ger. Den nyligen genomförda utvecklingen av modeller för inköpsresor visar att ytterligare data (från HUI Research's detaljhandelsregister och verksamhetsställeregister) kan öka modellernas förklaringsgrad påtagligt. Samtidigt krävs att dessa mycket detaljerade data kan prognoseras på ett rimligt sätt när modellerna används. Man måste därför göra en avvägning mellan förklaringsgrad och förklaringsfaktorerna prognosbarhet för de tillämpningar som är aktuella.

## Nerbrutna socioekonomiska indata i vissa delmodeller

I detta avsnitt beskrivs mer konkret hur nerbrutna socioekonomiska indata kommer in i de delmodeller som pekats ut i uppdragsbeskrivningen. Dessa är dels de regionala arbets- och skolmodellerna, dels de långväga tjänste- och privatresorna.

### Arbetsresemodellen

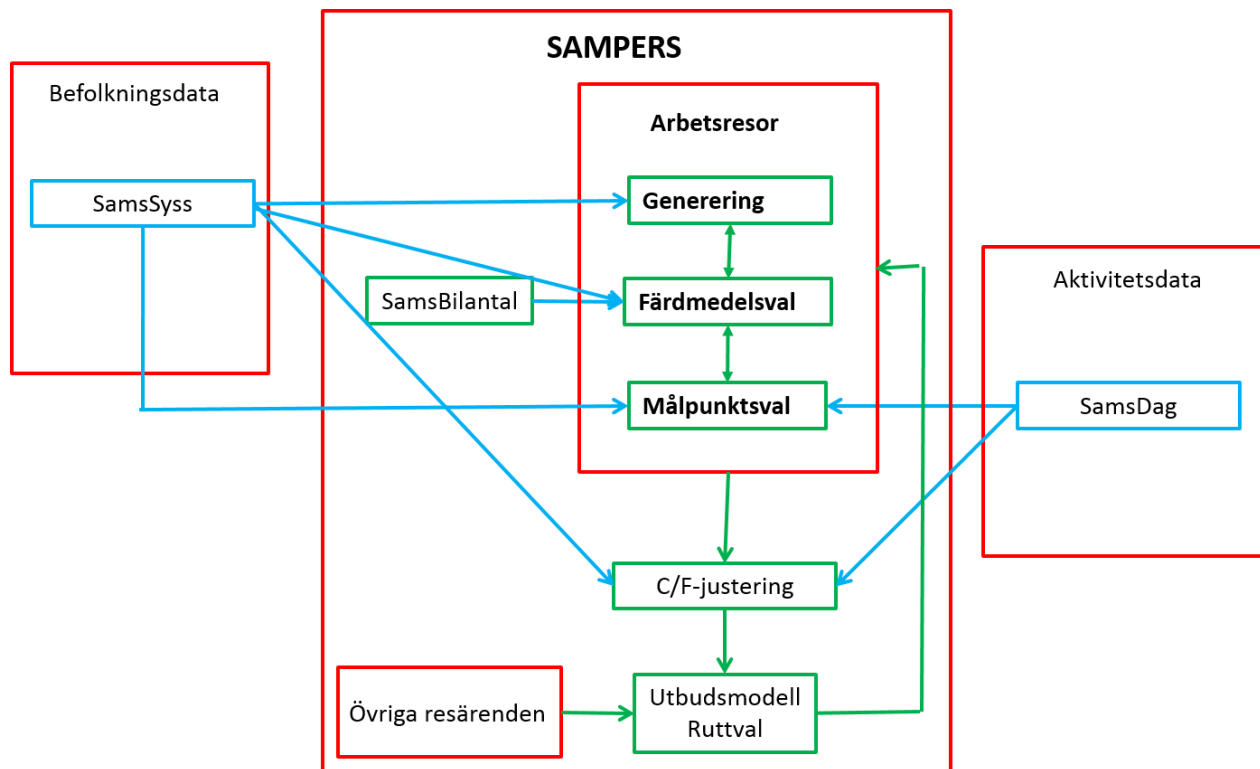
Alla data utnyttjas inte i alla delmodeller. När vi enbart ser på den regionala arbetsresemodellen blir bilden inte lika innehållsrik med avseende på indata. Vi förutsätter dessutom att bilinnehavsmodellen har körts, och nöjer oss med att representera resultatet med den resulterande tabellen SamsBilantal. I stället kompliceras bilden genom att vi lägger till den s.k. Cross/Fratarproceduren, som syftar till att skapa konsistens mellan arbetsresor och arbetsplatser på prognosområdesnivå.

Figur 5 visar hur kopplingarna mellan indata och modell ser ut för den regionala arbetsresemodellen. Först genererar efterfrågemodellen en fördelning av arbetsresorna på start- och målpunkt samt färdmedelsandelar för alla start-målrelationer. Alla sysselsatta reser inte till arbetet varje dag, varför antalet arbetsresor inte är detsamma som antalet arbetsplatser (antalet arbetsplatser är lika med antalet sysselsatta i SamsDag-tabellen). Antalet arbetsresor bestäms därför av modellen och inte av de absoluta talen i indata.

Modellen är vad man brukar kalla enkelt begränsad. Det innebär att det sammanlagda antalet resor till varje prognosområde inte nödvändigtvis motsvarar antalet arbetsplatser. Därför genomförs en s.k. matrisbalansering, i Sverige oftast kallad Cross/Fratarjustering. Den innebär att man förändrar den genom efterfrågemodellen erhållna arbetsresematrisen så att antalet arbetsresor från respektive till varje prognosområde motsvarar den sysselsatta befolkningen och antalet arbetsplatser. En mer utförlig beskrivning av proceduren finns i bilaga.

Befolkningsdata används förutom för att generera den sysselsatta befolkningen även för att generera andra socioekonomiska kategorier. I arbetsresemodellen förekommer flera socioekonomiska variabler, varav endast variabeln kön hämtas från SamsSys. Övriga variabler hämtas i form av medelvärden från den resvaneundersökning som modellen har skattats på. Dessa medelvärden är hårdkodade i Sampers, och förblir konstanta i prognossituationen (till dess att modellen omkodas med nya värden), medan SamsSys kan ändras/uppdateras.

När det gäller aktivitetsdata används endast det totala antalet arbetsplatser i varje prognosområde.



Figur 5 Sampers regionala arbetsresemodell

När modellen kalibrerats (efter estimering och implementering) har målsättningen hittills varit att den ska representera resvaneundersökningen så väl som möjligt (uppräknad till populationsnivå). Det är i princip möjligt att kalibrera modellen så att den återskapar information om resandet från någon annan källa. Data om befolkning och arbetsplatser är dock inga kalibreringsmålsättningar i sig, eftersom de inte avser resandet. Cross-Fratar-proceduren kan närmast ses som en restriktion på efterfrågemodellen som innebär att de i indata angivna markanvändningen blir mer styrande jämfört med efterfrågemodellen.

Kalibreringen sker genom att beräkna de tillägg till nyttofunktionerna (dvs de funktioner som innehåller restider och reskostnader etc.) som gör att överensstämmelsen mellan modell och kalibreringsmålsättning blir så god som möjligt. Kalibreringsmålsättningen definieras på en övergripande nivå för varje region, vanligen i form av reslängdsfördelningar och totaler för respektive färsätt. Kalibreringen ändrar valsannolikheterna, men inte de i modellskattningen erhållna parametrarna för olika påverkansfaktorer som exempelvis restider och reskostnader.

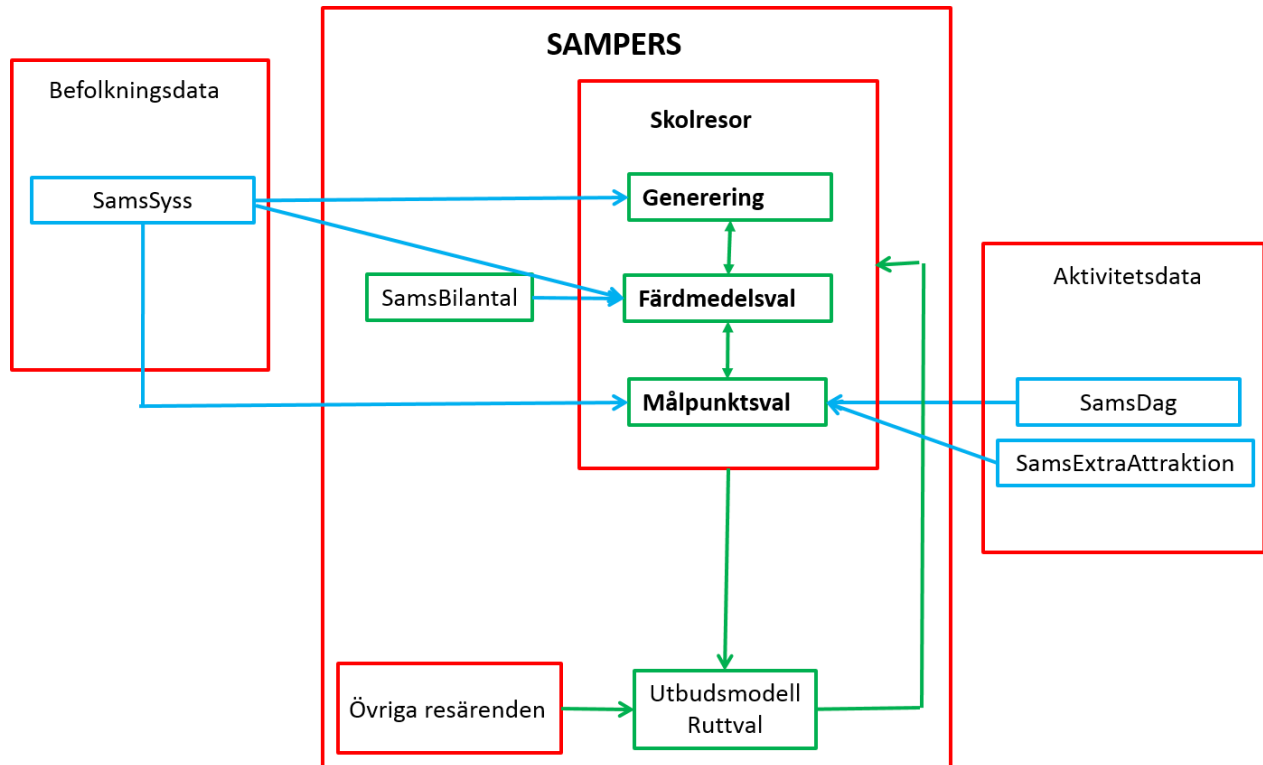
De matriser som läggs ut på näten innehåller inre bara de resor som motsvaras av resvaneundersökningen, utan också andra resor (kommersiell trafik etc.) i form av olika tilläggsmatriser i Sampers. Dessa totalmatriser kan också kalibreras, oftast med hjälp av flödesräkningar på vägnäten.

För utförligare dokumentation om arbetsresemodellen, se Sampers Teknisk rapport del 2A Regional Trips samt Utveckling av Sampers 2.1 Del 1 Estimeringen inkl tilläggstester, Transek rapport 2004:1.



## Skolresemodellen

Skolresemodellen spänner över hela den studerande befolkningen från grundskola till universitet, och avser således en ganska heterogen population. Det är därför inte konstigt att den innehåller det antal variabler som är kopplade till olika socioekonomiska grupper. Kopplingarna mellan modell och socioekonomiska data framgår av figur 6.



Figur 6 Sampers skolresemodell

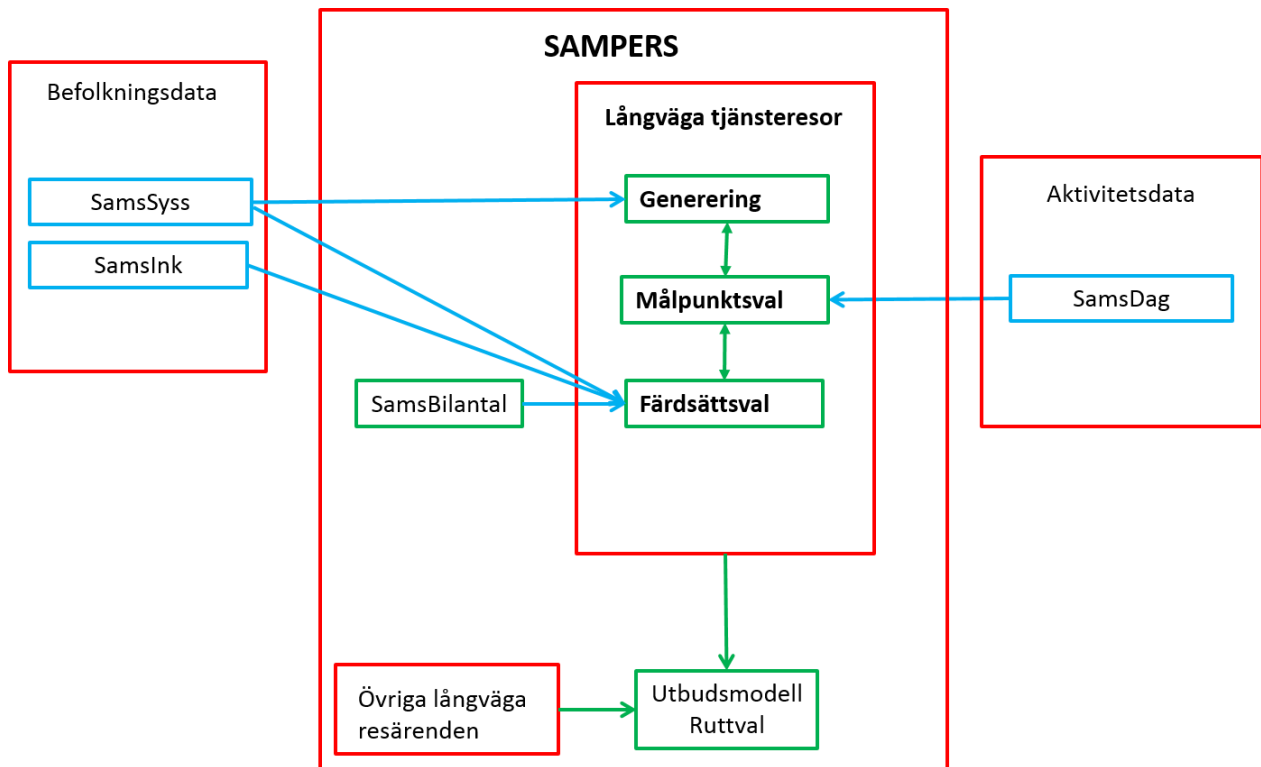
I SamsSys- samt SamsInk-tabellen används såväl kön, ålder, sysselsättning och inkomst och för att definiera olika populationssegment<sup>4</sup>. När det gäller aktivitetsfördelning används dagbefolkningen i utbildningssektorn (SamsDag), Totalbefolkningen (SamsSys) och antal studieplatser för stort respektive litet universitet (SamsExtraAttraktion).

För utförligare dokumentation om skolresemodellen, se Sampers Teknisk rapport del 2A Regional Trips samt Utveckling av Sampers 2.1 Del 1 Estimeringen inkl tilläggstester, Transek rapport 2004:1.

<sup>4</sup> Med populationssegment avses exempelvis "ej sysselsatta kvinnor i åldersgruppen 16-20 år".

## Modellen för långväga tjänsteresor

Kopplingarna mellan socioekonomiska indata och modellen för långväga tjänsteresor beskrivs i figur 7. Strukturen med avseende på de olika valdimensionerna innebär att färdmedelsvalet ligger under målpunktsvalet. Målpunktsvalet innehåller egentligen två nivåer, men det påverkar inte diskussionen här. Färdsättsvalet innebär också att ruttvalet för tåg hanteras i efterfrågemodellen, men inte heller det påverkar diskussionen här.



Figur 7 Sampers modell för långväga tjänsteresor

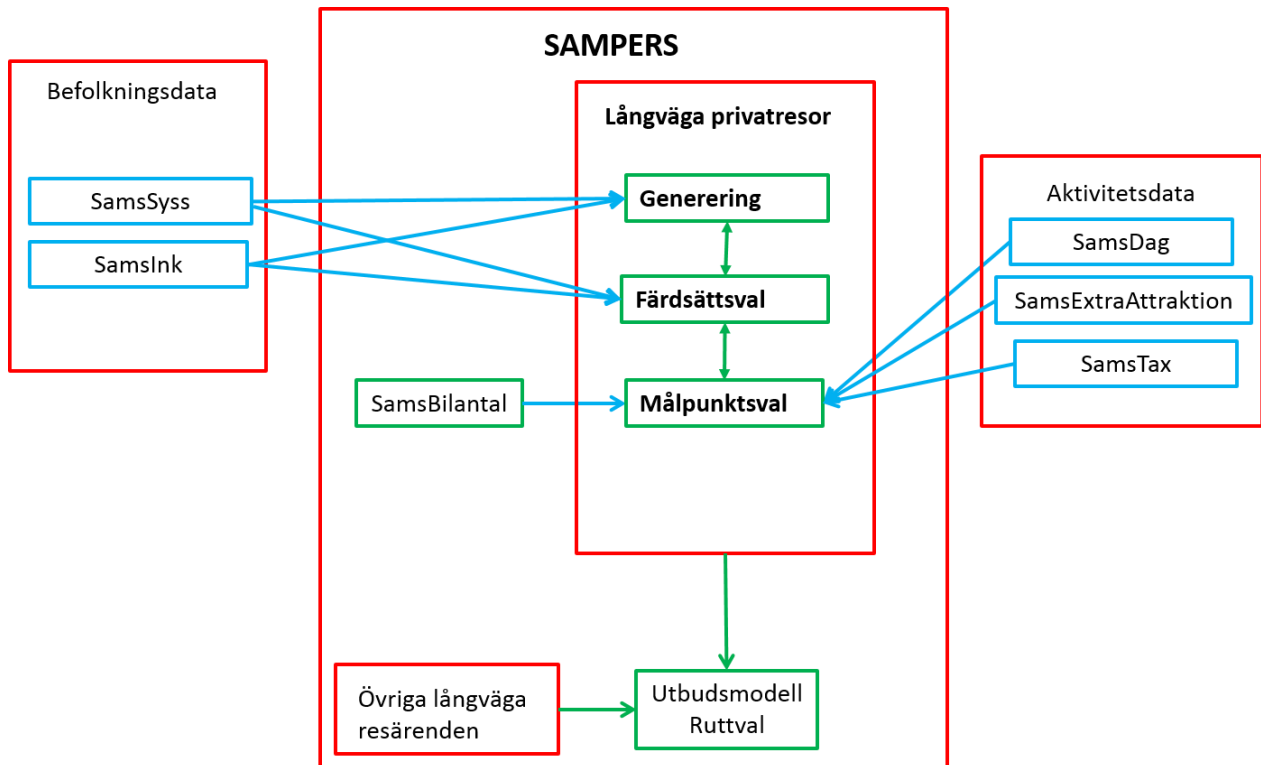
SamsSys används dels för att generera den sysselsatta befolkningen för genereringen, dels för att få fram kön och ålder som kommer in både i genererings- och i färdmedelsdimensionen. SamsInk används för att få en fördelning på inkomstkategorier. När det gäller attraktivitetsdata används den totala dagbefolkningen från SamsDag.

När det gäller modellkalibrering har modellen kalibrerats mot data från olika källor. Bil- och bussresandet har kalibrerats mot resvaneundersökningen, medan tågresandet har kalibrerats mot SJ's biljettstatistik (på grövre geografisk nivå) och flygresorna har kalibrerats mot luftfartsverkets statistik. Markanvändningsdata innehåller ingen information om det långväga resandet och är därför inte aktuellt som kalibreringsmål.

En mer detaljerad beskrivning av modellen finns i rapporten Höghastighetståg Del 2 – modellutveckling och känslighetsanalyser, WSP 10 feb 2012

## Modellen för långväga privatresor

Figur 7 visar kopplingen mellan socioekonomiska indata och modellen för långväga privatresor. SamsSys används för genereringen (totalbefolkningen åldersfördelad) men också för fördelning på kön för färdsetsval. SamsInk har använts för att inkomstfördela populationen vilket krävs i färdmedelsdelen. SamsDags näringsgrenar för kultur- och sportsektorn har använts för målpunktsbeskrivningen. SamsExtraAttraktion har använts för att beskriva utbud av vintersportfaciliteter (som dummyvariabel) mm. I modellen ingår också fritidshusyta, tagen från SamsTax-tabellen.



Figur 7 Sampers modell för långväga privatresor

Modellen har kalibrerats samtidigt med modellerna för långväga tjänste- och arbetsresor. På samma sätt som för tjänsteresorna kan inte markanvändningsvariablerna användas för att kalibrera modellerna för det långväga resandet.

En mer detaljerad beskrivning av modellen finns i rapporten Höghastighetståg Del 2 – modellutveckling och känslighetsanalyser, WSP 10 feb 2012. Den nya långväga modellen innehåller en separat modell för arbetsresor över 100 km enkel resa. Dessa ingick tidigare i privatresorna. Den regionala modellen innehåller visserligen ingen begränsning av reseavståndet, men antalet resor i denna avståndsklass är för litet för att kunna bli väl beskrivet i den regionala arbetsresomodellen. Arbetsresorna på längre avstånd är ofta av annan karaktär än vanlig pendling (vecko- och månadspendling).

## Bilaga: Cross-Fratar-proceduren i Sampers

Den s.k. Cross-Fratarproceduren i Samperssystemet finns inte beskriven i någon lättillgänglig dokumentation. Nedan följer en enkel beskrivning av proceduren

Arbetsresemodellen beskriver arbetsresandet som valet av att resa, färdmedel och destination. Valet är oberoende av andra individers val, vilket är ett mycket starkt antagande. Det finns inga garantier för att det inte blir fler tillresande än det finns arbetsplatser. Detta skulle ge orealistiska prognoser, varför en avstämning måste göras så att antalet tillresande sysselsatta stämmer överens med antalet arbetsplatser (korrigerade för att alla sysselsatta inte reser varje dag).

Korrigeringen sker genom en iterativ procedur. Följande steg genomförs:

- Först körs arbetsresemodellen, vilket ger
  - Antalet resor från område  $i$  till område  $j$  med färdväg  $m$ ,  $T_{ijm}$
  - Antalet resor totalt (summa över  $i$ ,  $j$  och  $m = T$ )
  - Antalet resor totalt till respektive område (summa över  $i$  och  $m = T_j$ )
  - Antalet resor totalt från respektive område (summa över  $j$  och  $m = T_i$ )
  - Antalet resor totalt i varje relation  $i$  till  $j$  (summa över  $m = T_{ij}$ )
  - Färdmedelsandelarna i varje relation  $i$  till  $j$  ( $FMV_{ij} = T_{ijm} / T_{ij}$ )
- Multiplicera antalet arbetsplatser i varje område  $S_j$  med kvoten mellan det totala antalet arbetsresor  $T$  och det totala antalet arbetsplatser  $S$ , ( $A_j = S_j * T/S$ ). Vi har därmed det önskade antalet avresande,  $T^*_j$ , och det önskade antalet tillresande,  $A^*_j$
- Justering av tillresande:
- Bilda för varje område kvoten mellan antalet modellberäknade tillresande och det önskade antalet tillresande ( $K'_j = T_j / A^*_j$ )
- Multiplicera alla resor från område  $i$  med kvoten mellan det modellberäknade antalet tillresande och det önskade antalet tillresande ( $(T'_{ij}) = K'_i * T_{ij}$ )
- Justering av avresande:
- Bilda för varje område kvoten mellan det nya antalet avresande ( $T'_i$ ) (summa  $T'_{ij}$  över  $j$ ) och det önskade antalet avresande ( $K''_i = T'_i / T^*_i$ )
- Multiplicera alla resor från område  $i$  med kvoten mellan det modellberäknade antalet avresande och det önskade antalet avresande ( $(T''_{ij}) = K''_i * T'_{ij}$ )
- Upprepa justeringarna av tillresande och avresande tills de justerade till- och avresande endast skiljer sig marginellt.
- Multiplicera därefter de så erhållna resorna  $T''_{ijm}$  med färdmedelsandelarna  $FMV_{ij}$ , vilket ger de önskade färdmedelsfördelningarna resmatriserna för arbetsresor