

Folkbokföringsfelet

En metodrapport om skattning av
övertäckning, undertäckning och folkbokförda
på fel adress

Folkbokföringsfelet En metodrapport om skattning av övertäckning, undertäckning och folkbokförda på fel adress

Producent SCB BV/REG och PMU/MIÖ
701 89 Örebro
010-479 40 00

Diarienummer 244198/8925943

Förfrågningar Fredrik Carlsson
010-479 63 78
Fredrik.Carlsson@scb.se

Rikard Gard
010-479 67 59
Rikard.Gard@scb.se

Henrik von Hofsten
010-479 68 59
Henrik.vonhofsten@scb.se

Karin Wegfors
010-479 64 98
Karin.Wegfors@scb.se

Stina Åsling Rønning
010-479 62 07
Stina.Ronning@scb.se

Förord

Registret över totalbefolkningen (RTB) består av individer folkbokförda i Sverige. I registret ingår det individer som är folkbokförda, men som inte borde vara det, vilket ger upphov till övertäckning i registret. Det finns också individer som inte är folkbokförda, men som borde vara det, vilket utgör undertäckning i registret. Slutligen finns det även individer som bor i Sverige men som inte är folkbokförda på den adress där individen borde vara folkbokförd om folkbokföringslagen efterlevts.

Skatteverket arbetar med att utveckla en metod för att på ett systematiskt sätt kunna följa upp kvaliteten i folkbokföringsregistret. Ett sätt att åstadkomma det är att med kvantitativa metoder uppskatta folkbokföringsfelet där folkbokföringsfelet är en avvikelse mellan vad en individ borde ha för uppgifter och vad samma individ faktiskt har för uppgifter.

Denna rapport har arbetats fram på uppdrag av Skatteverket och utgör en underlagsrapport till Skatteverkets rapport "Kvalitetsuppföljning i folkbokföringsregistret", september 2018. De resultat som presenteras i rapporten är endast preliminära och kan komma att ändras.

SCB i september 2018

Henrik von Hofsten

Innehåll

Förord	2
1. Inledning.....	5
1.1 Termer och begrepp	5
1.2 Register och datakällor.....	6
2. Modell - Övertäckning	8
2.1 Alternativ 1	8
2.1.1 Delmängden misstänkt övertäckning.....	8
2.1.2 Indikatorer.....	10
2.1.3 Potentiella indikatorer	13
2.1.4 Förkastade indikatorer.....	13
2.1.5 Osäkerhetskällor.....	13
2.2 Alternativ 2	15
2.2.1 Delmängden kontrollgrupp.....	15
2.2.2 Modellspecifikation.....	16
2.2.3 Osäkerhetskällor.....	17
2.3 Jämförelse av resultat mellan alternativ 1 och alternativ 2.....	17
Tabell 1.....	17
3. Modell - Undertäckning	18
3.1 Modellen	18
3.2 Ingående komponenter	19
3.2.1 Sena aviseringar – α	19
3.2.2 Högskolestuderande – β	20
3.2.3 Korttidsmigranter – γ	20
3.2.4 Inkomst i Sverige men inte folkbokförda - μ	20
3.3 Potentiell komponent	21

Samordningsnummer i Fordonsregistret	21
4. Modell för skattning av felaktig folkbokföringsadress	22
4.1 Modellen	22
4.2 Ingående komponenter	23
4.2.1 Impulser – α	23
4.2.2 Studenter - β	24
4.2.3 Stora övriga hushåll – γ	25
4.2.4 Individer med långa avstånd till arbetet– μ	26
4.2.5 Liten boendeyta – θ	27
4.3 Potentiella komponenter.....	27
4.3.1 Pass- och ID-uppgifter	27
4.3.2 Körkortsuppgifter	28
4.3.3 Studenter som bor och studerar i samma kommun	28
4.4 Förkastade komponenter	28
4.4.1 Bilägaren och besiktning av bilen	28
4.5 Modellens osäkerhet	28
4.6 Skattning	29
Tabell 2.....	29
4.7 Jämförelse och validering	30
5. Modell – Mätfel av andra egenskaper	31
5.1 Urvalsundersökning för att skatta mätfelet.....	32
Referenser	33
Bilaga.....	34

1. Inledning

Skatteverket arbetar med att utveckla en metod för att på ett systematiskt sätt kunna följa upp kvaliteten i folkbokföringsregistret. Ett sätt att åstadkomma det är att med kvantitativa metoder uppskatta folkbokföringsfelet där folkbokföringsfelet är en avvikelse mellan vad en individ borde ha för uppgifter och vad samma individ faktiskt har för uppgifter.

Som ett led i arbetet har SCB kontaktats för att på uppdrag av Skatteverket genomföra en registeranalys för att utveckla en metod för att uppskatta folkbokföringsfelet i folkbokföringsregistret via SCB:s register, då utifrån fyra kvalitetsaspekter:

- Övertäckning
- Undertäckning
- Mätfel – Felaktig folkbokföringsadress
- Mätfel – Andra typer av egenskaper där fel förekommer

Metodarbetet som presenteras i den här rapporten har utgått från en befintlig modell som SCB använder för att skatta övertäckningen i RTB, registret över totalbefolkningen (SCB, 2015). Modellen för undertäckningen bygger även den på en tidigare modell medan modellen för att skatta antal med felaktigt folkbokföringsadress har utvecklats inom ramen för det här arbetet. De skattningar som presenteras i den här rapporten är endast preliminära och kan komma att ändras.

Eftersom rampopulation är densamma för folkbokföringsregistret och registret över totalbefolkningen och målpopulationen är densamma för de båda registren så bedöms samma modell kunna användas för att uppfylla Skatteverkets behov av att uppskatta folkbokföringsfelet i folkbokföringsregistret. Givet att rätt underlag finns tillgängligt kan beskrivna metoder kontinuerligt uppdateras utifrån SCB:s register.

1.1 Termer och begrepp

För att undvika missförstånd är vissa centrala termer och begrepp definierade under den här rubriken.

Övertäckning – En täckningsbrist där rampopulationen innehåller fler individer än målpopulationen. I RTB finns samtliga individer som var folkbokförda i Sverige ett visst år. En del av dessa individer bor egentligen någon annanstans och borde inte vara folkbokförda i Sverige.

Undertäckning – En täckningsbrist där rampopulationen innehåller färre individer än målpopulationen. Vissa individer som bor i Sverige och som inte är folkbokförda borde egentligen vara det enligt folkbokföringslagen.

Mätfel – En avvikelse mellan det efterfrågade sanna värdet och det värde som faktiskt observerats. Exempel på mätfel är folkbokförd på annan adress än där man enligt lag borde vara folkbokförd om alla regler och lagar efterlevts. Andra exempel kan vara egenskaper som kön, ålder eller civilstånd där det efterfrågade sanna värdet skiljer sig mot det observerade.

Egenskapsfel – Synonymt med mätfel. En term som används hos Skatteverket.

1.2 Register och datakällor

Inom ramen för det här arbetet har flera olika register och datakällor identifierats som behövs för att ta fram skattingarna på folkbokföringsfelet och samtliga presenteras nedan.

Registret över totalbefolkningen (RTB) – Det statistiska register som SCB ansvarar för som utgår från folkbokföringsregistret men som har bearbetats för att användas till produktion av statistik.

Inkomst- och taxeringsregistret (IoT) - Registret ligger till grund för den officiella inkomst- och skattestatistiken. Registret innehåller uppgifter från Skatteverket (taxeringsuppgifter, kontrolluppgifter från arbetsgivare, kontrolluppgifter från pensions- och försäkringsutbetalare, kontrolluppgifter över räntor och utdelningar samt standardiserade räkenskapsutdrag), Försäkringskassan, Centrala studiestödsnämnden, Statens tjänstepensionsverk, Pensionsmyndigheten, Försvarmakten, Skolverket och Socialstyrelsen.

Universitets- och högskoleregistret – Registret ligger till grund för officiell högskolestatistik på grundnivå, avancerad nivå och forskarnivå. Uppgifterna samlas in från Sveriges högskolor och universitet.

Databasen för statistiska integrationsvariabler (STATIV) – Statistikdatabasens syfte är att utgöra underlag för beskrivning och analys avseende integration, segregation, jämställdhet och migration. Innehåller uppgifter från flera andra individregister som produceras av SCB samt data från Migrationsverket Arbetsförmedlingen, Försäkringskassan och Socialstyrelsen.

Lägenhetsregistret - Registret används för den årliga bostadsstatistiken och för den löpande hushålls- och boendestatistiken. Uppgifterna kommer från Lantmäteriet som är den myndighet som förvaltar och ansvarar för registret.

Fordonsregistret – Registret baseras på Transportstyrelsens vägtrafikregister över samtliga fordon som är registrerade i Sverige.

Elever i grundskolan – Registret innehåller samtliga elever inskrivna i grundskola, förskoleklass eller fritidshem per den 15 oktober.

Registerbaserad aktivitetsstatistik (RAKS) – Uppgifterna beskriver en individs totala försörjning och anknytning till arbetsmarknaden.

Kontrolluppgiftsbaserad lönesummestatistik (LSUM) – Registret visar lönesummor, preliminär A-skatt och vilka sociala ersättningar som betalas ut. Uppgifterna baseras på kontrolluppgifter som hämtas från Skatteverket.

Registret för migrations- och asylstatistik (MOA) – Registret innehåller uppgifter om beviljade uppehållstillstånd för utomnordiska immigranter. Uppgifterna hämtas från Migrationsverket.

Den gränsregionala statistiken – Uppgifterna visar alla personer som haft sin huvudsakliga löneinkomst i annat nordiskt land under ett givet år.

Skatteverkets impulsdata i folkbokföringsdatabasen – Data som visar iakttagelser från andra myndigheter och privatpersoner om indikationer på fel i folkbokföringen.

Utöver de registren ovan som används har två andra register identifierats som kan vara av intresse vid vidareutveckling av metoden:

Polisens pass- och ID-register

Transportstyrelsens körkortsregister

2. Modell - Övertäckning

Övertäckning i RTB definieras som de individer som finns i RTB men som inte skulle finnas där om folkbokföringslagen följdes fullt ut. Den vanligaste orsaken till övertäckning är att individer utvandrar utan att anmäla det till Skatteverket.

I ett arbete som SCB genomförde 2015 tog man fram en metod som skattar övertäckningen i RTB. Metoden som togs fram består av en modell som använder indikatorer för att avgöra om en individ bör tillhöra övertäckningen eller RTB. I dialog med Skatteverket har SCB tittat på hur den nuvarande modellen skulle kunna utvecklas.

För mer information om modellen som togs fram 2015 hänvisas läsaren till rapporten "Övertäckning i registret över totalbefolkningen" (SCB, 2015).

2.1 Alternativ 1

Alternativ 1 är en mindre vidareutveckling av 2015 års modell där sex indikatorer har tillkommit och definitionen av gruppen misstänkt övertäckning justerats för att inkludera individer som enbart haft garantipension.

2.1.1 Delmängden misstänkt övertäckning

Misstänkt övertäckning är en delmängd av RTB som består av de individer som finns med i RTB men som inte lämnat ett avtryck i något av SCB:s centrala statistikregister RTB:s händelseregister, STATIV eller IoT det senaste året. Vilket är detsamma som att säga att individen anses tillhöra RTB om den lämnat ett avtryck i något av registren.

RTB

Om en individ lämnat avtryck i något av RTB:s händelseregister anses den tillhöra RTB:

- Födda
- Immigration
- Civilståndsändringar
- Medborgarskapsbyte
- Inrikes flyttningar
- Emigration
- Döda

Det finns några undantag till avtryck i RTB:

- En person som ändrar civilstånd till änka/änkling/efterlevande partner, eftersom civilståndet sätts av Skatteverket när personens make/maka/partner dör
- När en person folkbokförs på Utan känd hemvist. Detta sker av Skatteverket då personens adress och vistelseort är okänd, trots en noggrann utredning
- Personer med sekretessmarkering, bl.a. eftersom dessa personer ofta är kvarskrivna på den gamla adressen och gör färre avtryck än personer utan sekretessmarkering. Skatteverket gör i regel en omprövning av sekretessen varje år

STATIV

Om en individ har ett värde för någon av följande variabler anses den tillhöra RTB:

- Sysselsättningsstatus
- Studiedeltagande

- Inskripen på Arbetsförmedlingen
- Antal dagar inskripen på Arbetsförmedlingen under året
- Deltagande i svenska för invandrare (SFI)
- Studerande på gymnasiet
- Avgången från gymnasiet under året
- Grund för bosättning som pensionär/tillräckliga medel

Ett undantag är om en individ har invandrat som pensionär, alternativt har tillräckliga ekonomiska medel för att försörja sig själv, så antas individen inte tillhöra misstänkt övertäckning.

IoT

Om en individ har ett värde för någon av följande variabler, förutom enbart garantipension, anses den tillhöra RTB:

- Studiemedel och studiehjälp (lån och bidrag) för familj
- Ekonomiskt bistånd för familj
- Introduktionsersättning till flyktingar för familj
- Arbetsinkomst för familj. Beräknas som summan av lön, företagarkomst, sjukpenning, föräldrapenning och ersättning i samband med militärtjänstgöring
- Sammanräknad förvärvsinkomst för familj. Beräknas som summan av inkomst av tjänst och inkomst av näringsverksamhet
- Enbart garantipension. Den här variabeln fanns inte med i 2015 års övertäckningsmodell och är den enda skillnaden i hur misstänkt övertäckning klassificeras.

Utvidgning av misstänkt övertäckning med enbart garantipension

Gruppen misstänkt övertäckning har utvidgats med individer som enbart har garantipension jämfört med övertäckningsmodellen som togs fram 2015. Variabeln är härledd genom garantipensionen och inkomst av tjänst. Om båda dessa belopp är lika stora och skilda från noll innebär det att man enbart fått garantipension.

Antal individer i misstänkt övertäckning

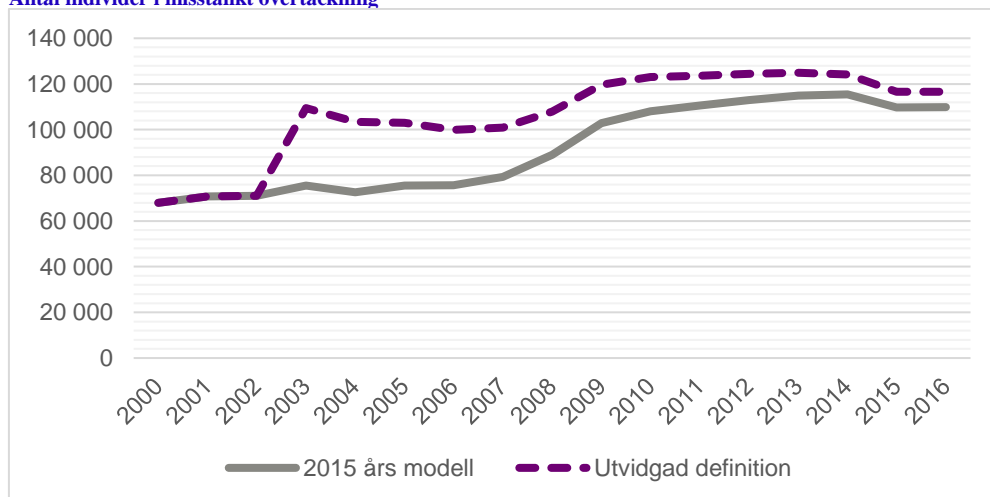


Diagram 1 - En jämförelse av antal individer i misstänkt övertäckning för modellen som togs fram 2015 gentemot modellen med utvidgad definition

Diagram 1 visar att för de tre första åren (2000 – 2002) fanns inga uppgifter om garantipension varför den skattade övertäckningen blir identiska de åren. Över tid

minskar skillnaden mellan 2015 års modell och modellen med utvidgad definition av misstänkt övertäckning. Det beror på att det är färre personer som har enbart garantipension idag än förr.

2.1.2 Indikatorer

Om vi har en population $U = \{1, \dots, i, \dots, N\}$, exempelvis RTB, innebär det att vi har N antal individer. Vi kan definiera $U_{m\ddot{o}} = \{i: i \in U \text{ och } i \notin A\}$ som **misstänkt övertäckning** där $i = 1, \dots, i, \dots, n$. Mängden A utgörs av de individer som lämnat ett avtryck i något av SCB:s centrala statistikregister RTB, IoT eller STATIV det senaste året.

Den delmängd $U_{m\ddot{o}}$ vi skapat antar vi går att förbättra ytterligare för att skatta övertäckningen. Vi skapar K indikatorer där K_2 är antalet indikatorer för individer som borde tillhöra övertäckningen och K_1 antalet indikatorer för individer som borde tillhöra RTB. Vi definierar \hat{I}_{ki} som en indikatorvariabel där $k = 1, \dots, k, \dots, K$ är ordnad efter K_2 sen K_1 för alla i som tillhör delmängden **misstänkt övertäckning**. Varje indikator är skapad med sin egen algoritm som använder sig av en eller flera variabler från statistiska register. En modell kan skapas med flera indikatorer; den modell som valts har totalt 33 indikatorer där 19 indikerar övertäckning och 14 indikerar RTB. Se bilaga 1.

Varje indikator tilldelas en vikt w_k där högre vikt innebär att indikatorn får större genomslag vid sammanräkningen. I alternativ 1 kan w_k ta värdet 1, 2 eller 3.¹ Slutligen summeras vikterna över alla indikatorer där en individ anses tillhöra övertäckningen ifall summan av vikterna är större för indikatorerna för övertäckning än för att tillhöra RTB, annars anses individen tillhöra RTB. Det här innebär att övertäckningen bestäms som en funktion av indikatorerna för delmängden $U_{m\ddot{o}}$ med följande formel:

$$R_i = \begin{cases} 1 & \text{om } \sum_{k=1}^{K_1} \hat{I}_{ki} w_k < \sum_{k=K_1+1}^K \hat{I}_{ki} w_k \text{ eller } \hat{I}_{ki} = 0 \forall k \\ 0 & \text{om } \sum_{k=1}^{K_1} \hat{I}_{ki} w_k \geq \sum_{k=K_1+1}^K \hat{I}_{ki} w_k \end{cases}$$

Det totala antalet individer som tillhör övertäckningen beräknas sedan genom $\sum_{i=1}^n R_i$.

Nya indikatorer

I 2015 års modell fanns 27 indikatorer. I modellen med utvidgad definition föreslås 33 indikatorer vilket innebär att sex nya indikatorer tillkommer. Samtliga indikatorer har testats utifrån enskilda variabler eller i kombination med andra och nedan följer en lista på de nya indikatorerna som bedömdes tillföra ny information:

- **Enbart garantipension samtidigt som inga skattefria bidrag/förmåner har erhållits**, indikerar att en individ tillhör övertäckningen. Man kan få garantipension utan att varaktigt bo i Sverige. Att endast se till garantipension är därför inte en indikator på att man tillhör RTB. Skulle man däremot erhålla skattefria bidrag/förmåner är ett rimligt antagande att man bor varaktigt i Sverige. I samband med enbart garantipension är det vanligt att personerna även får bostadstillägg. Pensionsmyndigheten gör en viss kontroll i samband med att personer får bostadstillägg. Det förhindrar att personer som tillhör övertäckningen får bostadstillägg. Antalet individer med enbart garantipension minskar stadigt från

¹ Vikterna har bestämts på förhand med hjälp av befintlig kunskap och antaganden från de som arbetat med modellen och presenteras i listan över indikatorerna som finns i bilagan.

34 000 ner till 7 500 för tidsperioden i diagrammet nedanför. Regeln granskar bara innevarande år och det finns data från och med år 2003.

- **Automatiska nolltaxerare.** Har man inte lämnat in en deklaration ingår man i gruppen automatiska nolltaxerare vilket är en indikation på att man tillhör övertäckningen. Den är dock ganska svag då man mycket väl kan bo varaktigt i Sverige utan att lämna in deklarationen. En grupp som normalt inte behöver lämna in deklarationer är barn. Därför har vi lagt till att man måste vara över 25 år. I princip alla som får träff på denna indikator finns i den gamla misstänkta övertäckningen. Det rör sig om mellan 30 000 till 58 000 individer, se diagram 2. Regeln granskar bara innevarande år och det finns data från och med år 2000.
- **Utgående uppehållstillstånd.** Alla individer med ett uppehållstillstånd som gått ut för minst ett år sedan betraktar vi som stor risk för övertäckning. Det handlar om mellan 2 500 till 11 000 individer, se diagram 2. Regeln granskar föregående år och data finns från och med år 2010.
- **Inskrivna vid grundskolan.** Elever som fanns inskrivna vid grundskolan i oktober året efter betraktar vi som en indikator på att man tillhör RTB. Man skulle kunna vända på det och säga att om man inte är inskriven vid grundskolan är det en indikation på att tillhöra övertäckningen, men det gäller bara barn i grundskoleålder. Antalet barn som finns i misstänkt övertäckning och är inskrivna vid grundskolan är mellan 7 000 till 9 000, se diagram 2. Regeln granskar bara innevarande år och det finns data från och med år 2007.
- **Anknytning till arbetsmarknaden.** Anknytning till arbetsmarknaden innebär att man har någon av följande status: helårsanställd, nyanställd, avgången, delårsanställd, kombinatör eller företagare vilket är en indikator att tillhöra RTB. Antalet individer bland gruppen misstänkt övertäckning som har anknytning till arbetsmarknaden är mellan 500 till 1 000, se diagram 2. Regeln granskar bara innevarande år och det finns data från och med år 2003.
- **Särskild postadress till utlandet.** Utländsk adress anmäld till Skatteverket, indikerar att personen tillhör övertäckningen. En person som vill ha sin post till utlandet bor troligen inte i Sverige. Antalet individer med särskild postadress till utlandet varierar mellan 350 till 700 per år, se diagram 2. Regeln granskar innevarande år och det finns data från och med år 2000.

Antal individer efter de nya indikatorerna

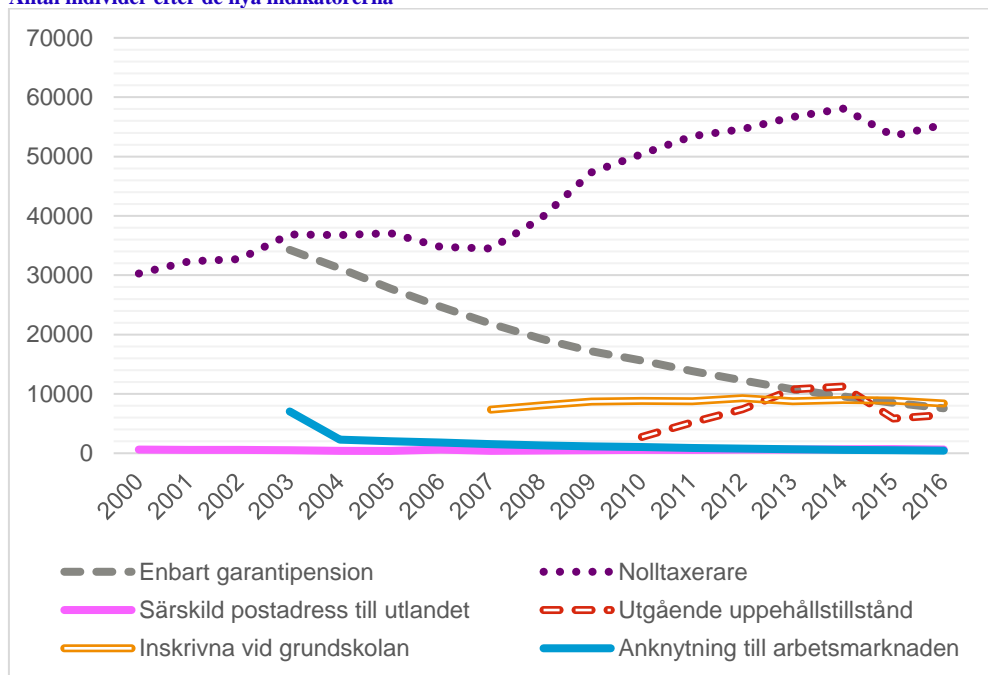


Diagram 2 - antal individer som fått träff på de nya indikatorerna, oavsett om man fått träff på tidigare indikatorer eller inte

En av målsättningarna med de nya indikatorerna har varit att den del av misstänkt övertäckning som inte får några träffar alls ska bli mindre vilket borde leda till mer träffsäkra och stabila skattningar.

Av de nya indikatorerna är det framförallt *nolltaxerare* som ger träff för de som inte haft träff på någon indikator tidigare. Sedan är det inskrivna vid grundskolan som matchar mot de som inte fått träffar på andra indikatorer tidigare. De övriga nya indikatorerna leder inte till att fler individer får träff men kan användas vid sammanvägningen.

Antal individer som tidigare inte fått träff på en indikator

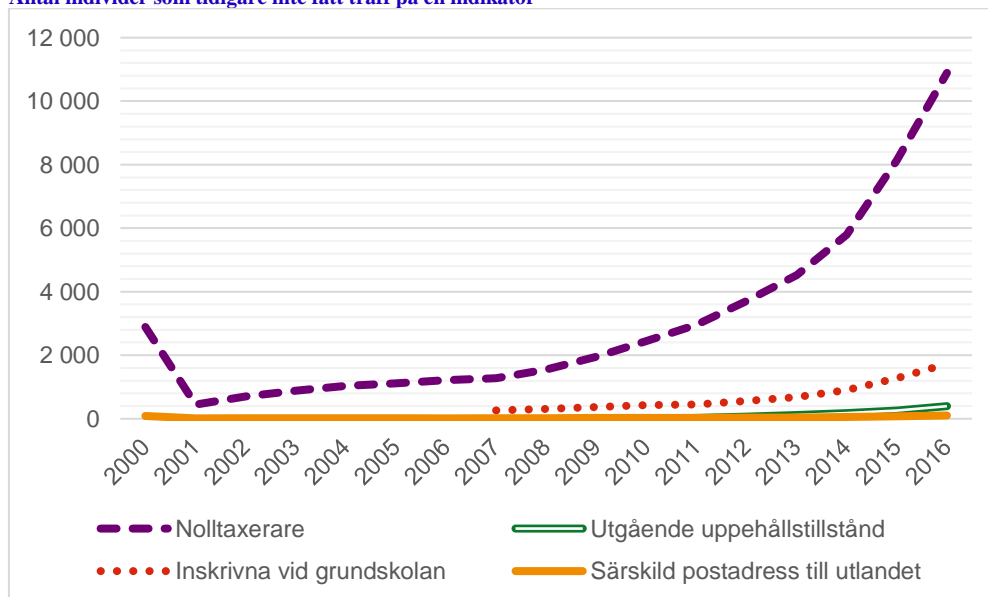


Diagram 3 - antal individer som tidigare inte fått någon träff på någon indikator men som fått träff på de nya indikatorerna

2.1.3 Potentiella indikatorer

Pass- och ID-uppgifter

Polisen har register över vilka individer som har svenskt pass och ID. Denna uppgift används av andra statistikbyråer i Europa för att göra en så kallad ”Signs of life analysis” vilket är detsamma som indikatorer för om man tillhör målpopulationen eller inte. Uppgifter om pass och ID skulle kunna följas över tid och om pass eller ID förnyas i Sverige, återkommande genom åren, är det en stark indikator på att individerna ska tillhöra RTB.

Pass- och ID-uppgifter har inte tagits med i modellen eftersom SCB i nuläget inte har tillgång till uppgifterna. Inom ramen för uppdraget har det kartlagts huruvida det ur ett juridiskt perspektiv är möjligt att hantera dessa uppgifter. Bedömningen är att det är möjligt att uppgifterna kan användas.

Körkort

På samma sätt som Pass- och ID-uppgifter kan användas skulle körkortsuppgifter kunna användas. Det är i nuläget oklart vilka uppgifter som finns tillgängliga. På motsvarande sätt som för Pass- och ID-uppgifterna har SCB inte tillgång till detta.

2.1.4 Förkastade indikatorer

Två indikatorer som arbetades fram men som förkastades var etableringsgrad och antal arbetade månader.

Etableringsgrad

Etableringsgrad visar i vilken grad man har anknytning till arbetsmarknaden utifrån individens förvärvsinkomst. Förvärvsinkomsten ska vara över ett visst gränsvärde, som är olika beroende på kön, ålder och utbildningsnivå, för att man ska vara etablerad.

Etableringsgrad gav inga träffar före 2008 och fr.o.m. 2008 ger indikatorn träff på exakt samma individer som får träff på anknytning till arbetsmarknaden.

Antal månader man arbetat

Antal månader en individ arbetat hämtas från månadsmarkeringen på kontrolluppgiften från Skatteverket och gav inga träffar alls bland misstänkt övertäckning. Det fanns en förhoppning att månadsmarkeringen skulle kunna användas till att identifiera individer som varit här en kortare stund men det visade sig inte stämma.

2.1.5 Osäkerhetskällor

Statistik är förenat med osäkerhet. Tre osäkerhetskällor bakom skattningen av övertäckningen som behöver nämnas handlar om *klassificeringen* av gruppen misstänkt övertäckning, *tilldelningen* av vikterna för indikatorerna samt *ingen träff* för någon indikator.

Klassificering av gruppen misstänkt övertäckning

Misstänkt övertäckning klassificeras utifrån tre olika register med flera olika variabler och tillhörande regler. Det är inte säkert att misstänkt övertäckning täcker samtliga som faktiskt är övertäckning vilket innebär att det finns en osäkerhet. Det här beror helt

enkelt på att variablerna som används kanske inte fångar upp samtliga som faktiskt är övertäckning och det är väldigt svårt att bedöma osäkerheten.

Tilldelning av vikter

Hur tilldelningen av vikter går till är av betydelse för hur skattningen blir i slutändan. Men det är inte helt klart hur man ska avgöra vilka indikatorer som är starkast och vilka som är svagast. Det är även oklart om en viss kombination av indikatorer är mer eller mindre viktig än en enstaka indikator. Det här innebär en osäkerhet för den här modellen. Eftersom tilldelningen av vikterna bestämdes subjektivt, utan en statistisk modell, finns det en risk att de valdes på ett sätt som kan ge en skev skattning.

Ingen träff på någon indikator

Av de som ingår i gruppen misstänkt övertäckning finns det vissa individer som inte får träff på någon indikator d.v.s. inget som indikerar om de tillhör RTB eller övertäckningen. Det är en svagt uppåtgående trend mellan 2001 och 2014 som ökar de två sista åren 2015 och 2016, se diagram 3. Det är framför allt två regler som använder data framåt i tiden som inte får några träffar de två sista åren. Reglerna är inaktivitet till aktivitet på samma fastighetsadress samt inaktivitet följt av emigration. Något val måste göras i det slutliga steget vilket är att modellen markerar de utan träff som att de tillhör RTB. Det här är ett osäkert antagande eftersom det enligt nuvarande metod inte går att avgöra om de borde tillhöra RTB eller övertäckningen. I diagram 4 syns det att indikatorer för att tillhöra RTB minskar för 2015 och 2016 vilket är en följd av att vissa indikatorer bygger på data för år som inte finns tillgängligt. Den syns även att indikatorerna för att tillhöra RTB minskar mer än indikatorerna för att tillhöra övertäckningen, vilket är en anledning att klassificera de som saknar träff för indikatorer som att de tillhör RTB. Det ger dessutom jämnare skattningar mellan åren vilket kan vara önskvärt beroende på vad man vill använda skattningen av övertäckningen till. För 2016 innebär det att ungefär 25 000 individer som tillhör misstänkt övertäckning och inte får någon indikatorträff, klassificeras som tillhörande RTB.

Antal individer med Indikatorer för och emot att tillhöra RTB

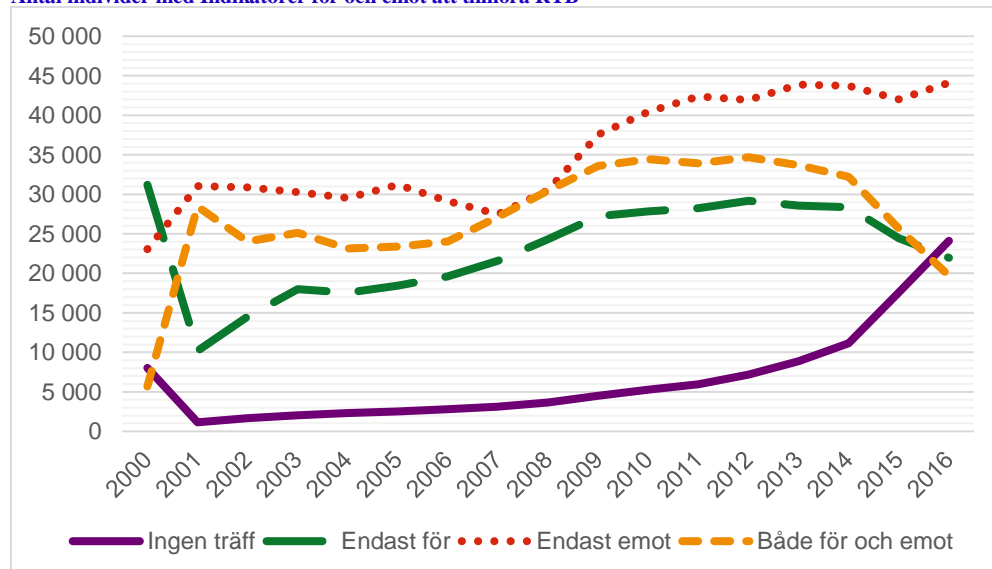


Diagram 4 - antal individer och vilken typ av indikator de fått träff på

I diagrammet ovan syns att antalet individer som inte får träff på någon indikator varierar mellan drygt 1 000 och drygt 24 000. Mellan 20 000 och 35 000 individer fick träffar på indikatorer både för och emot. Av de individer som har indikatorer både för och emot att

tillhöra RTB är det mellan 1 000 och 6 000 individer varje år som har en stark indikator som är motsatsen till vad modellen säger, se diagram 5. Det finns fler individer som har en stark indikator för övertäckning när modellen visar på att man tillhör RTB än vice versa. Det tyder på att modellen favoriserar att tillhöra RTB när vikten är densamma för alla indikatorer.

Antal individer med en stark indikator som går emot modellens resultat

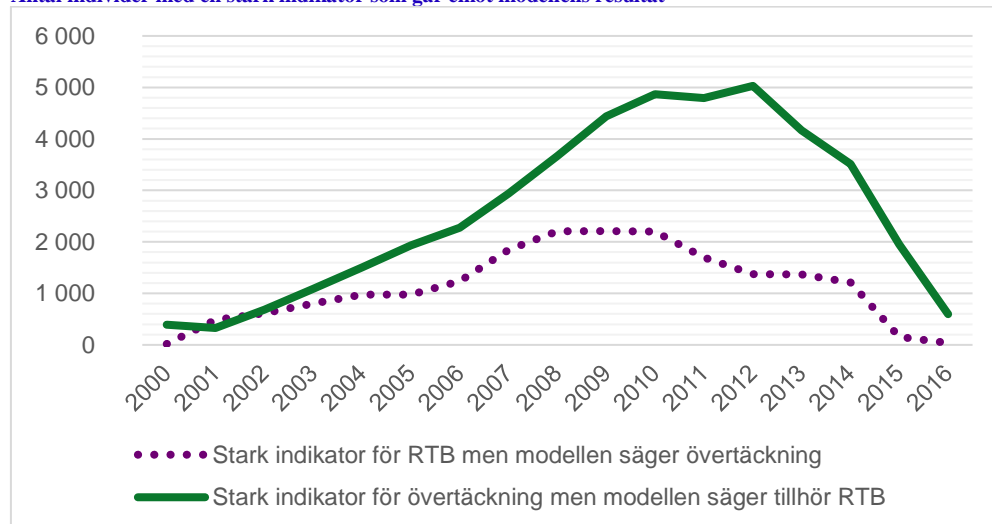


Diagram 5 - antal individer med en stark indikator som visar på motsatt riktning gentemot modellen

2.2 Alternativ 2

Det är önskvärt att frågå den subjektiva tilldelningen av vikter som görs i alternativ 1 och istället använda vedertagen statistisk metod för att beräkna vikterna. I det här avsnittet presenteras ett annat alternativ för skattning av övertäckningen som går ut på att identifiera en kontrollgrupp (som är så nära facit vi kommer) för att med logistisk regression ta fram nya vikter till indikatorerna.

2.2.1 Delmängden kontrollgrupp

Om det går att identifiera en delmängd av misstänkt övertäckning för en viss tidpunkt, $U_{m\ddot{o},t}$, som ett facit är det möjligt att använda en statistisk modell på den delmängden för beräkning av vikterna som sedan används för den större gruppen misstänkt övertäckning för tidpunkten övertäckningen ska skattas. Utmaningen är att identifiera en delmängd som kan fungera som ett facit, vilket vi väljer att kalla kontrollgrupp. Vi definierar $U_{Kontroll,t} = \{i: i \in U_{m\ddot{o},t} \text{ och } i \in A_{Kontroll,t}\}$ vilket innebär att reglerna som avgör vilka som hamnar i $A_{Kontroll,t}$ är avgörande för skattningen av övertäckningen. Det är antaganden som ligger till grund för reglerna vilket innebär att modellens tillförlitlighet till stor del vilar på dessa antaganden.

En kontrollgrupp har skapats med år 2010 som startpunkt där varje individ har följts 6 år framåt i tiden t.o.m. år 2016. Individer som dyker upp i händelseregister som antyder att den inte har tillhört RTB den perioden, vilket är om individen dör eller emigrerar, klassificeras som övertäckning. Samma register och variabler som används till att klassificera misstänkt övertäckning används här. Alla variabler, förutom död och emigrerad, som används till att klassificera misstänkt övertäckning används även till att klassificera individen som att den tillhör RTB.² Det finns individer som inte dyker upp i

² Antagandet här är att den status som blir känd inom 6 år är den status som individen haft hela tiden sedan 2010. Vi skapar en kontrollgrupp men osäkerhet kvarstår om individen har klassificerats korrekt eller inte.

något händelseregister; dessa klassificeras som okänd status. I rapporten från 2015 undersökte man hur lång tid det tog för individer att registreras i något händelseregister och man upptäckte att 90 % av alla individer har registrerats på något sätt 6 år senare (SCB, 2015). Det här är anledningen till att en individ följs 6 år framåt i tiden.

Anledningen till att år 2010 väljs som startpunkt är att det är den närmsta tidpunkten bakåt i tiden där flertalet individer kommer att ha registrerats fram t.o.m. år 2016. Det är önskvärt att använda statistik så nära inpå som möjligt eftersom det blir mer relevant eftersom resultatet blir mer aktuellt.

Det kan vara bra att poängtera att man kan välja andra eller flera startpunkter och även ett längre tidsintervall än 6 år för att skapa kontrollgruppen. Det kan vara bra att betona att vikterna i den här rapporten beräknas utifrån en kontrollgrupp från 2010 års misstänkta övertäckning och används sedan för samtliga år av misstänkt övertäckning.

Misstänkt övertäckning bestod år 2010 av 123 000 individer där ungefär 59 000 klassificerades att tillhöra RTB, 36 000 klassificerades tillhöra övertäckningen och 28 000 klassificerades med okänd status.

Antal individer bland misstänkt övertäckning med senare känd status

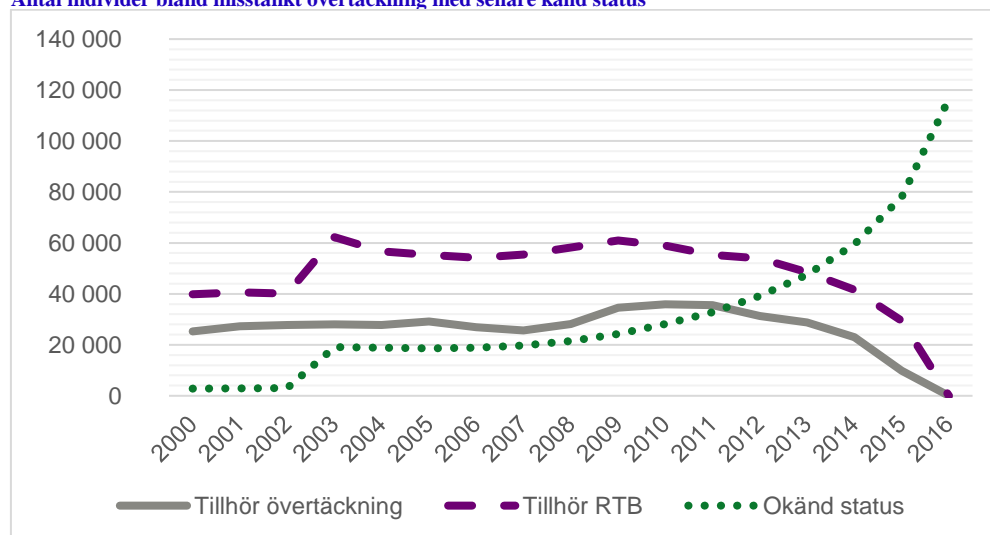


Diagram 6 - antal individer bland misstänkt övertäckning med senare känd status. År 2005 ska tolkas som att det fanns runt 105 000 individer i gruppen misstänkt övertäckning där cirka 20 000 inte fått någon känd status år 2016

2.2.2 Modellspecifikation

För att beräkna vikterna används en statistisk modell som hanterar den binära responsvariabeln om man tillhör övertäckning eller ej. Modellen som föreslås är därför en logistisk regression med samma indikatorer som till alternativ 1 som förklarande variabler. Maximum likelihood-metoden används för skattningen av modellens parametrar vilket blir de vikter som används till modellen som presenteras i avsnitt 2.1.2. Indikatorer.

Den modellspecifikation som används i den här rapporten består av 30 indikatorvariabler vilka finns med i hjälpvektorn som förklaringsvariabler³. Den logistiska regressionen utgår sedan utifrån 2010 års kontrollgrupp för skattningen av parametrarna. Modellens skattning av övertäckningen redovisas i nästa avsnitt.

2.2.3 Osäkerhetskällor

Alternativ 2 har samma osäkerhetskällor som alternativ 1. Skillnaden ligger i sättet som vikterna tilldelas. För de individer som bildar kontrollgruppen ligger ett antagande om att vi har kunskap om det sanna värdet vilket vi i själva verket inte har. Men om antagandet bakom kontrollgruppen håller är det här sättet att tilldela vikter inte lika osäkert och alternativ 2 skulle vara att föredra gentemot alternativ 1 eftersom man skulle förlita sig på en statistisk modell.

2.3 Jämförelse av resultat mellan alternativ 1 och alternativ 2

Skattningen av övertäckningen skiljer sig mellan de båda modellerna men om skillnaden är stor eller liten beror helt på hur skattningen för övertäckningen ska användas av statistikanvändare. Nedanför presenteras en tabell över de båda alternativen.

Tabell 1

En jämförelse av skattningen för övertäckningen mellan alternativ 1 (modellen med subjektivt tilldelade vikter) och alternativ 2 (En kontrollgrupp som skapats utifrån 2010 års misstänkt övertäckning där vikterna beräknas med en statistisk modell) och den misstänkta övertäckningen

År	Alternativ 1	Alternativ 2	Misstänkt övertäckning
2010	63 500	56 700	123 000
2011	66 000	60 600	123 600
2012	66 800	61 600	124 500
2013	69 600	65 700	124 900
2014	70 400	68 500	124 100
2015	72 400	68 000	116 500
2016	78 500	84 200	116 500

³ Indikatorer med information från högskoleregistret har utgått då vi saknade tillgång till registret vid tidpunkten för regressionsanalysen. Dessutom har indikatorn inaktivitet följt av emigration utgått då den korrelerar nästan helt med kontrollgruppen.

3. Modell - Undertäckning

Undertäckning består av individer som inte är folkbokförda i Sverige men som *borde* vara det, om folkbokföringslagen efterlevts. Att hitta dessa individer är svårt och skulle bli kostsamt om man använde gränskontrollanter eller besöksintervjuare. Det är betydligt billigare att utreda undertäckningen via register. Men att skatta undertäckningen via register är svårt eftersom dessa individer gör få eller inga avtryck i register.

Vi har därför valt att använda en modell som är framtagen i syfte att skatta en total befolkningssiffra för Sverige (Bengtsson, 2015). Där utgjorde modellen för undertäckning en delkomponent vid beräkningen av den totala befolkningssiffran.

Modellen som beskrivs i den rapporten är nästan densamma som den som tillämpas här, bortsett från att asylsökande och papperslösa då ingick i delkomponenten för undertäckningen. Det beror på att målpopulationen för skattningen av den totala befolkningssiffran skiljer sig från RTB:s och folkbokföringsregistrets målpopulation. I RTB och folkbokföringsregistret är målpopulationen de individer som borde vara folkbokförda ifall regler och lagar efterlevs. För den totala befolkningssiffran är målpopulationen de individer som normalt har sin dygnsvila i landet bortsett från tillfällig frånvaro för rekreation, semester, besök hos vänner och släktingar, affärsresor, läkarbehandling eller religiösa pilgrimsfärder.

3.1 Modellen

Modellen är en deterministisk komponentmodell där varje komponent modelleras för sig för att sedan summeras.

Modellen kan skrivas på formeln:

$$Y = \alpha + \beta + \gamma + \mu$$

Där

Y = Antal individer som ingår i gruppen undertäckning

α = Antal individer med sena aviseringar

β = Antalet utländska högskolestuderande som bor och pluggar i Sverige men inte är folkbokförda här

γ = Antalet korttidsmigranter

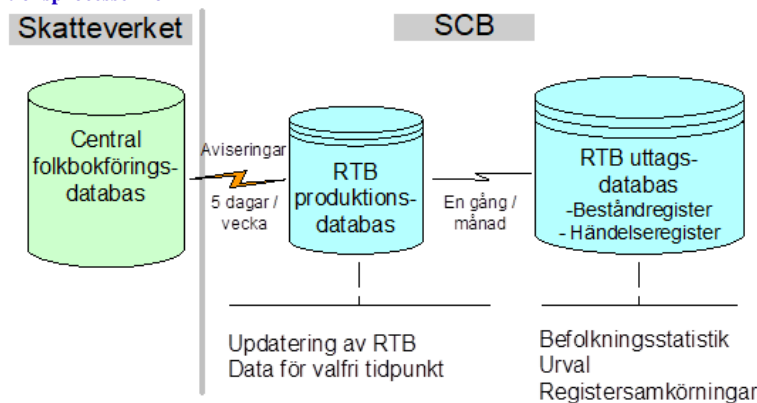
μ = Antalet individer med inkomst i Sverige men ej folkbokförda

3.2 Ingående komponenter

Nedan beskrivs de ingående komponenterna översiktligt. För mer information se SCB:s interna rapport om befolkningssiffran som nämns ovan.

3.2.1 Sena aviseringar – α

Produktionsprocessen för RTB



Figur 1

Bilden ovan visar hur uppgifterna från Skatteverket, så kallade aviseringar, uppdaterar RTB. Aviseringarna läggs in i RTB:s produktionsdatabas som uppdateras 5 dagar per vecka. En gång i månaden tas informationen ur produktionsdatabasen för att skapa statistikregister som används för t.ex. statistikproduktion, urvalsdragningar och mikrodatautlämnande. När statistikregistren är framställda är de "lästa" och uppdateras inte senare med nya uppgifter, bl.a. eftersom den officiella befolkningsstatistiken baseras på dessa register. En avvägning mellan kvalitetsaspekterna aktualitet och täckning har gjort att man avvaktar en månad innan uttaget ur produktionsdatabasen görs. Ett exempel för att illustrera detta: Uttaget för statistikregistren som avser befolkningen 31/12 2017 gjordes 31/1 2018. Då har de allra flesta aviseringarna som avser 31/12 2017 eller tidigare inkommit till SCB.

Det finns dock aviseringar som inkommer **efter** 31/1 2018 som skulle påverka befolkningen 31/12 2017 om man väntat en längre tid än en månad. Till undertäckningen i RTB räknas sena aviseringar rörande invandring och födselar. T.ex. en födelse som ägt rum 21 december 2017 men rapporteras av Skatteverket 3 februari 2018. Denna födelse leder till en undertäckning i statistikregistret som avser 31/12 2017 och det är de årsvisa statistikregistren som är utgångspunkten för alla beräkningar.

Antalet för denna skattning år 2016 var 1260 individer, 122 födda och 1138 immigranter.

3.2.2 Högskolestuderande – β

En del av undertäckningen består av ej folkbokförda studenter, som studerar i Sverige mer än ett år. Dessa studenter borde enligt folkbokföringslagen vara folkbokförda i Sverige.

I högskoleregistret registreras alla högskolestudenter oavsett om de är folkbokförda eller inte. Genom att kontrollera vilka studenter som är inskrivna längre än ett år hittas de studenter som borde vara folkbokförda.

Undertäckningen bland ej folkbokförda högskolestuderande avseende år t skattas enligt följande. För åren $t-1$ och bakåt till 2001 beräknas andelen ej folkbokförda studenter som är inskrivna tre terminer i rad av totalt antal studenter som inte är folkbokförda. För att beräkna andelen år t där information saknas för tre terminer framåt, används ett genomsnitt av de tidigare framräknade andelarna. Den genomsnittliga andelen sett över tid multipliceras sedan med de ej folkbokförda studenterna för år t för att få fram en aktuell skattning.

Vilket kan skrivas på formeln:

$$U_t^{student} = \bar{x}_{t*} \times S_t$$

där $U_t^{student}$ står för studenter som utgör undertäckning vid tidpunkt t , \bar{x}_{t*} är den genomsnittliga andelen studenter som studerat tre terminer på rad för tidsperioden $t*$ vilket är perioden år $t-1$ tillbaka till år 2001 och S_t är antalet studenter som inte är folkbokförda i Sverige vid tidpunkt t .

Antalet för denna skattning år 2016 var 2775 individer.

3.2.3 Korttidsmigranter – γ

Ett av kraven för att bli avregistrerad som utvandrad är att man har för avsikt att vara bosatt utomlands under minst ett år. Individer som av någon anledning återinvandrar till Sverige inom ett år från utvandringen kallas korttidsmigranter. Det finns en överenskommelse mellan de nordiska länderna som innebär att man inte kan flytta från ett nordiskt land till ett annat och bli folkbokförd utan att samtidigt bli avregistrerad i det land varifrån man flyttar. Reglerna för folkbokföring skiljer sig dock åt mellan länderna. För att bli folkbokförd i exempelvis Norge räcker det med att man har för avsikt att stanna minst sex månader. Förutom skillnader i regelverk kan intentionen från början vara att man ska stanna mer än tolv månader men att man trots det återvänder efter kortare tid än så. Av ovan nämnda skäl kan man diskutera om samtliga korttidsmigranter bör betraktas som undertäckning eller inte. Det finns en osäkerhet kring huruvida man överskattar undertäckningen något för korttidsmigranterna. Men bedömningen är att överskattningen i sådana fall är marginell.

I modellen för undertäckning skattas antalet korttidsmigranter enligt följande. Datum för individens (senaste) utvandring ett år jämförs med individens datum för (första) invandring året därpå. Om differensen mellan datumen understiger 366 dagar, är individen en korttidsmigrant.

Antalet för denna skattning år 2016 var 2430 individer.

3.2.4 Inkomst i Sverige men inte folkbokförda - μ

I Inkomst- och taxeringsregistret (IoT) förekommer det individer med samordningsnummer. Ett samordningsnummer fungerar som en identitet för individer

som inte är eller har varit folkbokförda, t.ex. om man arbetat en kortare tid i Sverige. De som ingår i IoT är dels den grupp som valt att deklarerat med en vanlig kontrollblankett (KU10) eller Kontrolluppgift – Sjöinkomst (KU16) från Skatteverket. I IoT saknas dock de som skattar med så kallad artistskatt, en vanlig konstruktion för exempelvis många av de idrottare som ingår i svenska elitlag.

Vidare ingår även uppgifter från bland annat Försäkringskassan och CSN för individer med samordningsnummer.

Antalet individer med inkomst i Sverige men ej folkbokförd skattas genom att summera alla i IoT med samordningsnummer ett år som förekom året före och med disponibel inkomst över ett prisbasbelopp. Dessa anses som ett bidrag till undertäckningen. Valet av ett prisbasbelopp motiveras med att den gränsen används i flera andra sammanhang, exempelvis som gräns för inkomst då man inte har någon rimlig möjlighet att försörja sig.

Antalet för denna skattning år 2016 är 4810 individer.

3.3 Potentiell komponent

Nedan följer en komponent som skulle kunna ingå i modellen men där mer arbete behövs för att klargöra om komponenten ska vara med eller inte.

Samordningsnummer i Fordonsregistret

Fordonsregistret innehåller de bilar som finns registrerade i Sverige där individer med samordningsnummer ingår. Det fanns 24 733 individer med samordningsnummer som ägde ett fordon 31/12 2016.

Skattningen utgår ifrån de individer som ägde samma bil i mer än ett år och som var i trafik under samma tid. Endast personbilar övervägs eftersom lastbilar ägs i samband med företagande vilket gör det svårt att avgöra om ägaren är densamma som kör lastbilen. Det fanns 5 291 individer med dessa kriterier 31/12 2016 och av dem var 725 individer dubletter som fanns med i IoT så de borde plockas bort ifall den här skattningen används.

Resonemanget är att individer som äger en bil i Sverige som varit i trafik men saknar personnummer egentligen borde vara folkbokförda här eftersom de kört bil i mer än ett år. Att bilen är i trafik innebär att bilen måste vara försäkrad. Det är då troligt att ägaren befinner sig i Sverige, annars skulle inte bilen behöva försäkras.

Vid diskussion med de som förvaltar Fordonsregistret framkom det att det kan vara svårt att använda registret för att bedöma kvalitén i RTB eftersom den som står registrerad på en bil inte nödvändigtvis är densamma som kör bilen. Av den anledningen behövs mer diskussion kring registret och förståelse för vilka individer man kan vara säkra på är förare till bilarna som finns registrerade i Sverige.

4. Modell för skattning av felaktig folkbokföringsadress

Mätfelet definieras som avvikelsen mellan den *faktiskt observerade* folkbokföringsadressen och där individen *borde* vara folkbokförd om folkbokföringslagen efterlevts, där själva mätningen är den adress som observerats som finns i registret. Till vår kännedom finns det i dagsläget inga registerbaserade modellansatser för att skatta antalet individer med felaktig folkbokföringsadress vilket innebär att en ny metod och modell måste utvecklas. I det här kapitlet presenterar vi ett förslag på en modell som skattar antalet individer folkbokförda på fel adress.

4.1 Modellen

Modellen som vi utvecklat är en deterministisk komponentmodell som liknar modellen som används vid skattning av undertäckningen. Varje komponent modelleras var för sig och summeras i slutändan ihop. Varje komponent kan skattas med flera mindre delmodeller där ett genomsnitt av dessa delmodeller beräknas för att få fram den enskilda komponentens skattning. Den här metoden har hämtat inspiration från Robert T Clemen (1989) som går igenom litteratur där man kombinerat resultatet från flera prognoser och sedan beräknat medelvärdet och funnit att det kan ge bättre skattningar, vilket är anledningen till att vi valt att använda medelvärdet av flera olika delmodeller för varje komponent.

Modellen kan skrivas på formeln:

$$Y = \frac{\sum_{i=1}^{N_1} \alpha_i}{N_1} + \frac{\sum_{i=1}^{N_2} \beta_i}{N_2} + \frac{\sum_{i=1}^{N_3} \gamma_i}{N_3} + \frac{\sum_{i=1}^{N_4} \mu_i}{N_4} + \frac{\sum_{i=1}^{N_5} \theta_i}{N_5}$$

Där

Y = Skattat antal individer med fel adress

N_j = Antal modeller för den j :te komponenten

α_i = Den i :te modellen för att skatta antal impulser som leder till en åtgärd om ändring av folkbokföringsadress

β_i = Den i :te modellen för att skatta antal studenter med fel adress

γ_i = Den i :te modellen för att skatta antal individer folkbokförda i ett övrigt hushåll med 7 eller fler individer

μ_i = Den i :te modellen för att skatta antal individer som arbetspendlar orimligt långa sträckor

θ_i = Den i :te modellen för att skatta antal individer som bor på en orimligt liten boendeyta

Komponenterna skattas så att de blir ömsesidigt uteslutande. Det görs genom att först beräkna impulserna, där individer som även ingår i de andra komponenterna kan förekomma. Sedan beräknas antalet studenter med fel adress där man selekterar bort de som även ingick i impulserna, vilket innebär att studenterna även kan ingå i de tre övriga

komponenterna. Beräkningarna görs i den ordning som finns i formeln ovan. Den sista komponenten, liten boendeyta, selekterar bort samtliga andra individer som förekommer i de fyra föregående komponenterna.

4.2 Ingående komponenter

I det här kapitlet beskrivs varje komponent mer ingående. Om skattningen för en komponent baseras på medelvärdet från flera mindre modeller så beskrivs de olika modellerna.

Modellens träffsäkerhet vilar på de antaganden som ligger till grund för respektive komponents modell och hur pass rimliga de är. Det innebär att om ett antagande inte håller finns det risk att skattningen blir dålig.

Framtagningen av de gränser som används till de ingående komponenterna är subjektiv. Gränserna har tagits fram genom analys av percentilerna samtidigt som vi resonerat kring variabelns sakinnehåll utifrån syftet med det vi vill skatta. En ren registerstudie ger inte utrymme för en rent objektiv ansats, utan gränsdragning genom sunt förnuft behövs för att kunna bygga modellen. Det är möjligt att göra mer objektiva gränsdragningar med hjälp av andelar som beräknas utifrån urvalsundersökningar. Registerstudien skulle alltså kunna kompletteras med resultat, i form av vikter eller andelar, från ett mindre urval där man kontaktat urvalspersonerna och intervjuat dem.

4.2.1 Impulser – α

Om någon obehörig har folkbokfört sig hemma hos någon annan och den som egentligen bor där meddelar Skatteverket detta, så inkommer en impuls till Skatteverket. Impulser kan även komma från andra myndigheter i Sverige som hör av sig till Skatteverket om de misstänker att något är fel i folkbokföringen. Vi har valt att använda de impulser som leder till åtgärd om ändrad folkbokföringsadress vid skattningen av antal individer som är felaktigt folkbokförda eftersom de klassificerats med felaktig folkbokföringsadress hos Skatteverket.

Antal impulser år 2017

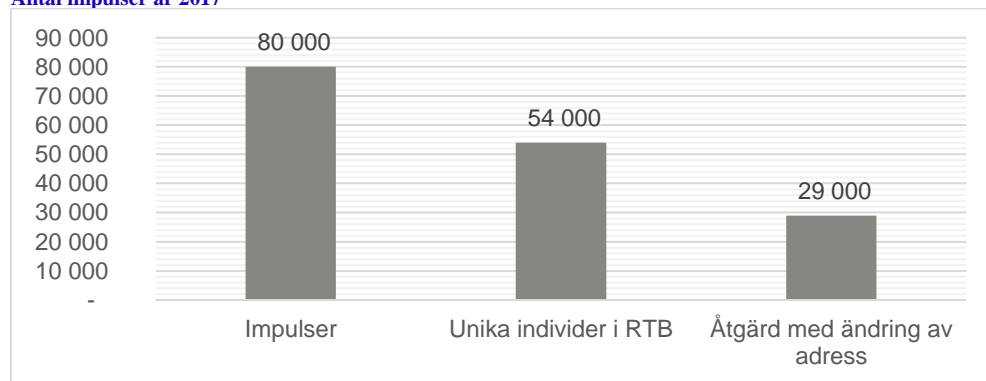


Diagram 7 - antal impulser år 2017 samkörda med RTB 2016 med nedbrytningar av impulserna i mindre grupper där de mindre grupperna är en delmängd av den förgående större gruppen

Det var 80 000 impulser som inkom till Skatteverket med ett ärende om flyttning år 2017. Impulserna utgjordes av 54 000 unika individer som även fanns med i RTB år 2016. Av de impulserna som kom in så ledde 29 000 till åtgärd om ändrad adress. I modellen för impulser antar vi att dessa ändringar inte hade gjorts om det inte varit för impulserna och att det av den anledningen bör ses som ett folkbokföringsfel. Skattningen för impulser som felaktig adress är alltså 29 000 och det finns endast förslag på en enkel modell.

En utmaning med impulserna är att koda vilka individer som har lett till åtgärd om felaktig folkbokföringsadress eftersom det är en textvariabel där handläggare hos Skatteverket kan ha fyllt i på olika sätt. Det är inte givet att de 29 000 individer som kodats till åtgärd utgör ett folkbokföringsfel. Vilket kan bero på kodningen eller själva antagandet (att en impuls leder till åtgärd är inte nödvändigtvis ett fel utan skulle kunna bero på en fördröjning inom folkbokföringen som hinner uppfattas som ett fel). Här skulle det vara bra att stärka kunskapen kring kvaliteten i datamaterialet genom fortsatt dialog med Skatteverket kring impulsdata.

4.2.2 Studenter - β

En del studenter folkbokför sig inte på den lägenhet där de faktiskt bor. Omfattningen är däremot inte tillräckligt utredd. Med hjälp av SCB:s register kan man skatta antalet studenter som är felaktigt folkbokförda genom att utgå ifrån högskoleregistret och samköra med uppgifter från RTB.

Eftersom RTB är en stock med referenstiden sista december för varje år finns ett datum för när man är folkbokförd och det finns uppgifter om individens folkbokföringskommun. Den uppgiften kan samköras med högskoleregistret där uppgifter om studier med studieort finns för varje termin. Även uppgifter om man läser på distans finns som då går att selekteras bort.

Antal studenter som bor och studerar i olika kommuner 2016

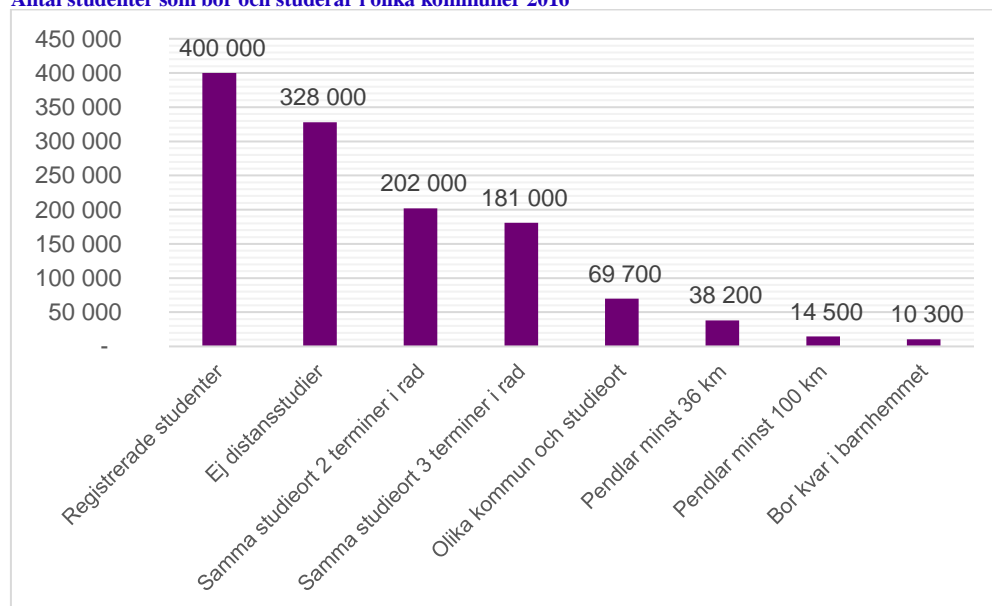


Diagram 8 - antal studerande nedbrutet i mindre grupper där de mindre grupperna är en delmängd av den förgående större gruppen

På det här sättet kan man se hur många studenter det finns som har studerat minst tre terminer i följd i samma kommun men är folkbokförd i en annan kommun mellan den andra och tredje terminen. Det går inte att anta att alla dessa studenter är felaktigt folkbokförda eftersom det kan vara rimligt att bo i en kommun men studera i en annan, givet att pendlingsavståndet inte är alldeles för långt. Men det går att samköra uppgifter om avstånd mellan kommunernas mittpunkter för att välja ut de individer som anses ha ett alldeles för långt pendlingsavstånd. Exempelvis antas att en student inte pendlar mer än 100 km mellan boende och studieort. Till sist kan de studenter som är barn i ett hushåll eller bor kvar hos sina föräldrar selekteras.

Det här sättet att bryta ned en större grupp till mindre och mindre delar genom vissa regler är ett sätt att identifiera antalet studenter som är folkbokförda på fel adress. För varje nedbrytning antas att nedbrytningen blir mer och mer säker på att det är en felaktig folkbokföring. Samtidigt förloras information eftersom vissa studenter som selekterats bort längst vägen fortfarande kan ha fel adress. För att hantera det används medelvärdet av de grupper som kommer lägre i hierarkin.

Om det bedöms att man är ganska säker på att de 10 300 studenterna som är folkbokförda hos sina föräldrar med minst 100 km till sin studieort är felaktigt folkbokförda blir det minimigränsen. Det är rimligt att anta att flera av de som pendlar mindre än 100 km även är felaktigt folkbokförda. Den 3:e kvartilen visar att 25 procent av alla studenter har som mest 36 km att pendla till studieorten vilket bedöms vara ytterligare en rimlig gräns eftersom det är svårt att tro att 25 procent av alla studenter pendlar 36 km. Det kan även vara så att individer som pendlar mindre än 36 km är felaktigt folkbokförda vilket varför att den delgruppen även väljs med. Det innebär fyra olika modeller, om än väldigt enkla, som ligger till grund för medelvärdesskattningen på 33 200 individer som beräknas vara felaktigt folkbokförda.

4.2.3 Stora övriga hushåll – γ

Vid en analys av impulsdata syntes ett tydligt samband mellan impulserna och hushållets storlek och om man tillhörde ett övrigt hushåll där övrigt hushåll definieras som en övrig kategori till ensamstående hushåll och sambohushåll. I gruppen övriga hushåll ingår bl.a. syskonhushåll, flergenerationshushåll eller inneboende. Även tidningar har skrivit om folkbokföringsfelet med koppling till hushållets storlek och hushållstyp där man gjort sina egna modellberäkningar av folkbokföringsfelet utifrån hushållsstorleken och relations sambanden inom hushållet vilket varit till inspiration för den här komponenten (Hem och hyra, 2018).

Uppgifter om hushåll hämtas från RTB:s registerbaserade hushållsstatistik. Där finns uppgifter om vilken lägenhet man är skriven på, hur många individer som är skrivna på lägenheten samt hushållstyp som delas in i sammanboende, ensamboende och övriga hushåll, alla med eller utan barn.

Antal individer i stora övriga hushåll 2016

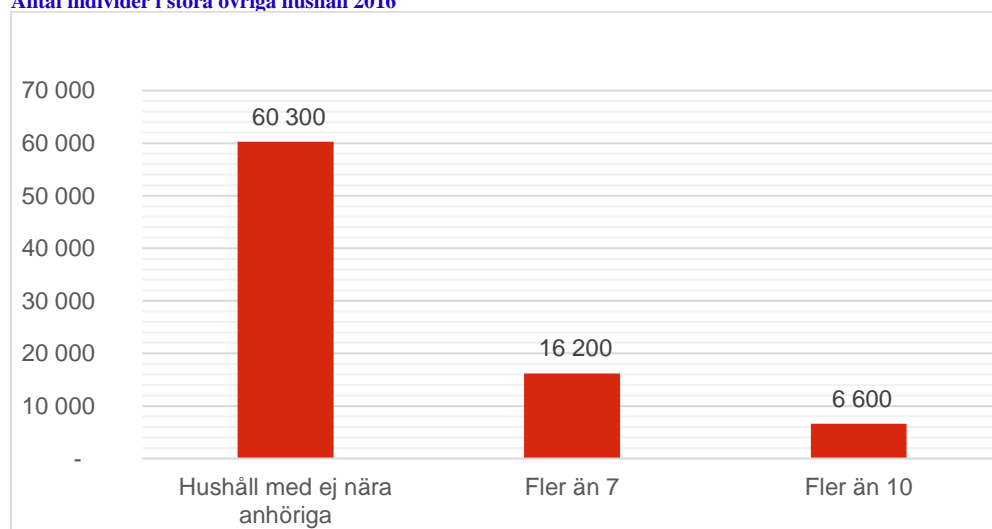


Diagram 9 - antal individer nedbrutet i mindre grupper där de mindre grupperna är en delmängd av den föregående större gruppen

Det finns drygt 1,2 miljoner individer som bor i övriga hushåll. Av dessa bor 60 300 i ett hushåll där det finns minst tre individer utan nära anhöriga i hushållet. Det finns anledning att tro att en del av dessa individer utgör ett folkbokföringsfel vilket är varför den gruppen tas med som en modell. Att samtliga dessa har ett folkbokföringsfel verkar dock inte rimligt, även om det är möjligt. Därför har gruppen brutits ned tre gånger med större och större hushåll. Det innebär att den här komponenten består av tre enklare modeller där medelvärdet blir 27 700.

4.2.4 Individer med långa avstånd till arbetet– μ

I STATIV finns uppgifter om vilket arbetsställe en individ har ifall denne arbetar. Om man arbetar i en kommun men är folkbokförd i en annan är det möjligt att man är felaktigt folkbokförd eftersom det kan vara så att man har sin egentliga hemvist i den kommun man arbetar. Genom att mäta avståndet mellan folkbokföringskommunen och arbetsställekommunens mittpunkter har vi en variabel som går att selektera på för att skatta antalet felaktigt folkbokförda. Det innebär antaganden om att det är orimligt att pendla över ett visst avstånd samt att variabeln arbetsställe är korrekt, vilket inte behöver vara fallet eftersom arbetsställe är en registervariabel som även den kan innehålla mätfel. Vilket även gäller för studenterna, stora övriga hushållen och liten boendeyta.

Vi antar att endast de personer som inte är gifta, inte är registrerade partners eller de som inte lever i samboförhållanden utgör den grupp som står för felaktig folkbokföring. Vilket exempelvis innebär att om en person är gift och pendlar över 1 000 km antas det vara rimligt.

Antal individer med långa avstånd mellan adress och arbetet 2016

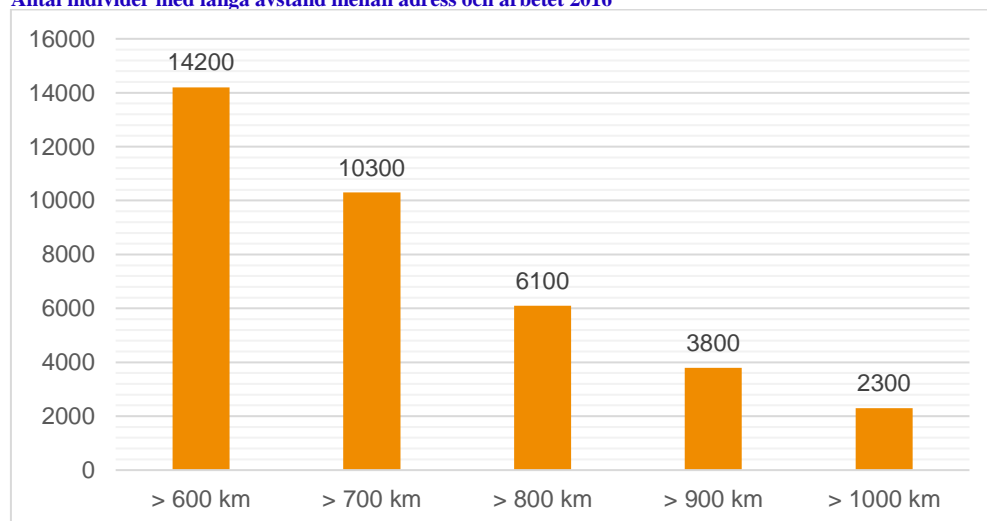


Diagram 10 - antal individer, som inte är sambo eller gifta, med långa avstånd till arbetet, nedbrutet i mindre grupper där de mindre grupperna är en delmängd av den förgående större gruppen

Att samtliga 14 200 individer med mer än 600 km till arbetsorten skulle ha felaktig adress är inte rimligt eftersom det är möjligt att kunna veckopendla det avståndet. Men några av dessa tror vi har felaktig adress. Vi bryter ned gruppen med 100 km per gång, varje gång ner tills 1000 km där vi antar att merparten arbetspendlare har felaktig adress. Om vi räknar ut medelvärdet för de fem olika (enkla) delmodellerna, med subjektivt valda gränser, får vi 7 300 som är felaktigt folkbokförda.

4.2.5 Liten boendeyta – 0

Vissa hushåll bor mer trångt än andra och vad som är en rimlig boendeyta beror på flera faktorer. Vårt resonemang kring kopplingen mellan boendeyta och folkbokföringsfelet utgår ifrån att det finns en gräns för hur trångt man kan tänkas bo.

I den registerbaserade hushållsstatistiken från RTB finns ett ID-nummer för varje lägenhet. Det kan man samköra med lägenhetsregistret och kombinera uppgifter om boarea personer per lägenhet för att beräkna boarea per individ.

Antal individer med liten boendeyta i övriga hushåll 2016

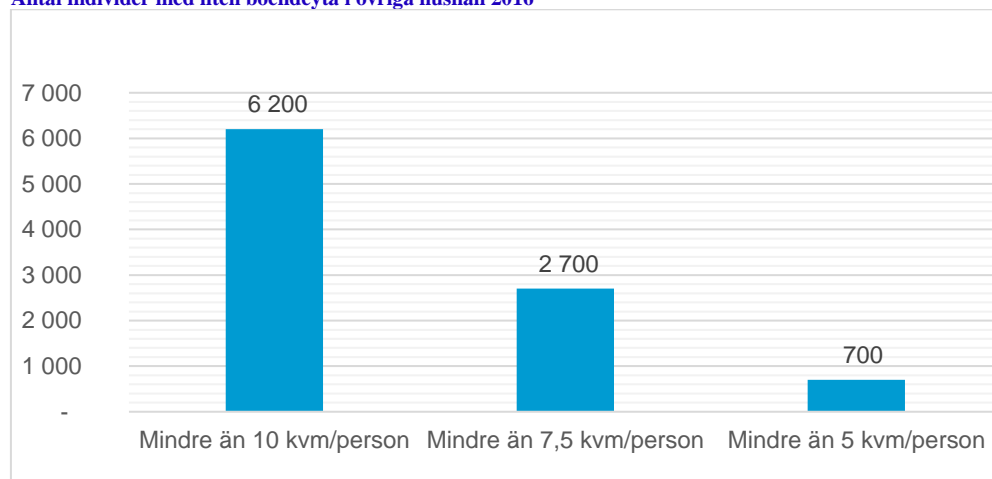


Diagram 11 - antal individer med liten boendeyta som bor i ett övrigt hushåll nedbrutet i mindre grupper där de mindre grupperna är en delmängd av den förgående större gruppen

Resonemanget gällande liten boendeyta och valet i vilka grupper som medelvärdet ska beräknas utifrån är detsamma som för komponenten stora övriga hushåll. Det finns tre grupper som medelvärdet beräknats utifrån vilket ger en skattning på 3 200 individer.

4.3 Potentiella komponenter

I det här kapitlet går vi igenom några komponenter som skulle kunna användas vid skattningen av antalet folkbokförda på fel adress men som behöver utredas mer ifall de ska ingå i modellen. Anledning till varför varje komponent inte ingått står under respektive del.

4.3.1 Pass- och ID-uppgifter

Polisen har register över vilka individer som har svenskt pass och ID. I registret finns det uppgifter om vart man ansökte om pass eller ID och vart någonstans det lämnades ut. Dessa uppgifter skulle kunna användas för att skatta var man egentligen bor; om man antar att de flesta som beställer och hämtar ut sitt pass på samma plats bor där. Om en individ beställt och hämtat ut pass från samma ort de senaste 10-15 åren är det en stark indikation på att den faktiskt bor där. Dessa uppgifter fungerar för att identifiera individer som bor i en kommun men är folkbokförda i en annan men det går inte att använda för att identifiera felaktigt folkbokförda inom den kommun man är skriven.

Pass- och ID-uppgifter har inte tagits med i modellen eftersom SCB i nuläget inte har tillgång till uppgifterna. Inom ramen för uppdraget har det kartlagts huruvida det ur ett juridiskt perspektiv är möjligt att hantera dessa uppgifter. Bedömningen är att det är möjligt att uppgifterna kan användas.

4.3.2 Körkortsuppgifter

På samma sätt som Pass- och ID-uppgifter kan användas skulle körkortsuppgifter kunna användas. Det är i nuläget oklart vilka uppgifter som finns tillgängliga. På motsvarande sätt som för Pass- och ID-uppgifterna har SCB inte tillgång till detta.

4.3.3 Studenter som bor och studerar i samma kommun

Skattningarna för studenterna gäller endast studenter som bor i en kommun men studerar i en annan. Det innebär att man missar de studenter som bor och studerar i samma kommun som är folkbokförda på fel adress. För att skatta det skulle man kunna specificera en modell genom en logistisk regression med gruppen *olika kommun och studieort* som population, *bor kvar i hemmet* som responsvariabel med bland annat *studielån* och *inkomst* som förklaringsvariabler. En sådan modell togs fram men förklaringsgraden blev låg och resultatet blev orimligt eftersom andelen studenter som bor i samma kommun rimligtvis inte har samma andel felaktigt folkbokförda som studenter som bor och studerar i olika kommuner, vilket blev utfallet. Av dessa två anledningar är komponenten placerad som en potentiell komponent. Men man behöver hitta en vektor med hjälpinformation som ger högre förklaringsgrad.

4.4 Förkastade komponenter

En komponent ingick i slutändan inte i modellen men dokumenteras i den här rapporten inför framtida arbeten med modellen.

4.4.1 Bilägaren och besiktning av bilen

Det fanns en idé att använda sig av vägtrafikregistret (fordonsregistret) och samköra med RTB och se var besiktningen gjordes. Om en bil besiktas i en kommun flera år i rad men ägaren är folkbokförd i en annan som ligger långt ifrån så skulle det indikera att individen är felaktigt folkbokförd.

Det är osäkert att använda fordonsregistret för att utvärdera kvaliteten i RTB bl.a. eftersom ägaren av bilen inte nödvändigtvis är samma individ som kör den. En individ kan dessutom äga flera bilar, i olika kommuner, och det finns inget krav på var bilen ska besiktas. På grund av detta valde vi att inte gå vidare med uppgifter från registret över fordon.

4.5 Modellens osäkerhet

Det här kapitlet innehåller en kombination av kvalitativt och kvantitativt resonemang för att belysa modellens osäkerhet.

Statistik är förenat med osäkerhet. I urvalsundersökningar redovisar man en punktskattning tillsammans med ett osäkerhetsintervall med en konfidensgrad om 95 % med vilket vi säger att osäkerhetsintervallet kommer täcka det sanna värdet i 95 % av fallen. Tekniskt kan osäkerhetsintervall presenteras antingen som i någon mening objektiva konfidensintervall eller som subjektiva bedömningsintervall.

Eftersom detta är en registerundersökning och ingen urvalsundersökning är sannolikhetsfördelningen för modellen okänd. Det innebär att vi inte kan ta fram ett objektiva konfidensintervall. Men vi kan däremot ta fram ett subjektivt bedömningsintervall genom att för varje komponent välja ut max- och min-modellerna vilket ger oss ett sätt att konstruera ett osäkerhetsintervall. Ett sådant osäkerhetsintervall är svårtolkat men givet att våra antaganden bakom max- och min-modellerna för varje komponent stämmer kan vi vara säkra på att antalet individer folkbokförda på fel adress ligger någonstans i det intervallet. Men vi kan inte kvantifiera hur säkra vi är eftersom vi inte känner till den underliggande sannolikhetsfördelningen. Osäkerhetsintervallets giltighet vilar alltså på huruvida modellens antaganden håller eller inte. Antaganden som

att **vi kan vara säkra på att** det minsta antalet felaktigt folkbokförda studenter är 10 300 eller **vi tror att** det endast finns en skattning för impulser som leder till åtgärd.⁴

Osäkerhetsintervallets gränser kan skrivas på formeln:

$$Y_{Min} = \min\{\alpha_i\} + \min\{\beta_i\} + \min\{\gamma_i\} + \min\{\mu_i\} + \min\{\theta_i\}$$

$$Y_{Max} = \max\{\alpha_i\} + \max\{\beta_i\} + \max\{\gamma_i\} + \max\{\mu_i\} + \max\{\theta_i\}$$

Själva skattningen, Y, ligger någonstans mellan dessa värden och inte nödvändigtvis i mitten; max och min är inte symmetriska kring punktskattningen.

4.6 Skattning

Modellens skattning för RTB 2016 ger att cirka 100 000 individer har felaktig folkbokföringsadress, men den skattningen är förenad med osäkerhet. Ser man till min- och maxvärden ger modellen värden som varierar mellan 49 000 – 179 000. Tolkningen är att vi är ganska säkra på att antalet personer som är felaktigt folkbokförda ligger någonstans i det intervallet. Men vi kan inte kvantifiera hur säkra vi är eftersom det är svårt att tilldela modellen en sannolikhetsfördelning. Osäkerhetsintervallet är dessutom inte symmetriskt kring punktskattningen.

Tabell 2

Sammanställning av värdena som används till modellen

Delmodell	Impuls som leder till åtgärd	Studenter ej samma kommun	Stora övriga hushåll	Långa avstånd till arbetet	Liten boendeyta
1	29 000	10 300	6 600	2 300	700
2		14 500	16 200	3 800	2 700
3		38 200	60 300	6 100	6 200
4		69 700		10 300	
5				14 200	

I den här tabellen syns samtliga ingående komponenter och de olika delmodellerna för varje komponent. Det är impulserna som leder till åtgärd, studenterna som är folkbokförda och studerar i olika kommuner samt de stora övriga hushållen som har störst inverkan på skattningen. Vilket även syns i diagrammet nedanför.

⁴ Ett bedömningsintervalls giltighet vilar på utvärderingar, utredningar, analyser, ämnes- och registerkunskaper som gett erfarenhet inom området.

Min-, medel- och maxskattning

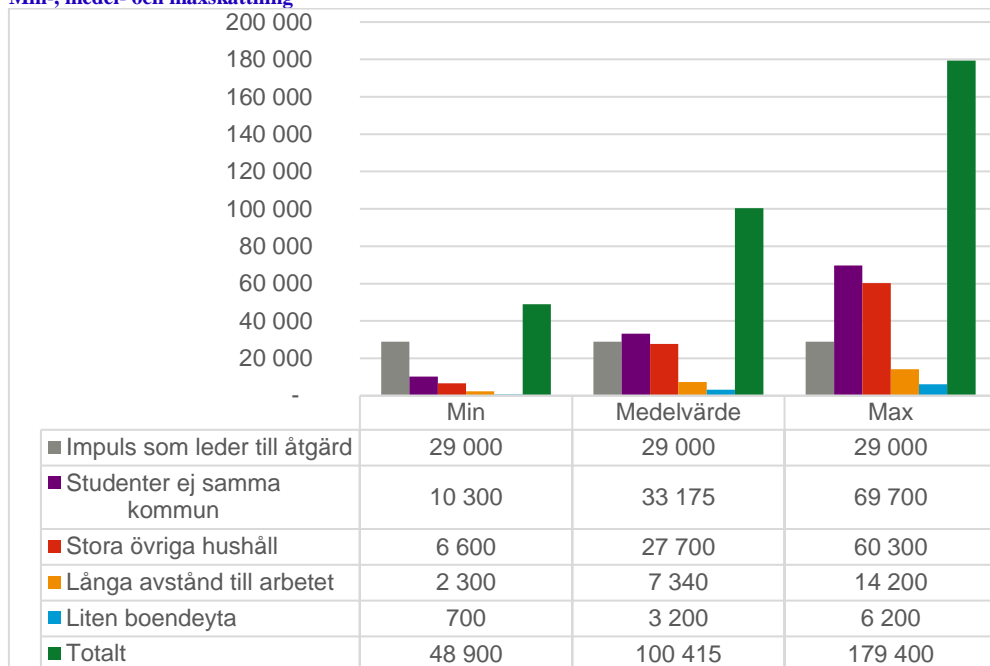


Diagram 12 – De ingående komponenternas min-, max- och medelvärden sammanställda för varje enskild komponent samt totalen

4.7 Jämförelse och validering

Det går att jämföra resultatet från den här modellen med resultatet från andra modeller. Skatteverket tog fram en skattning på ett osäkerhetsintervall till mellan 45 000 – 128 000 (Skatteverket, 2013) för en treårsperiod (2010 – 2012). Om vi delar intervallets gränsvärden med tre får vi en ungefärlig bild av osäkerhetsintervallet för ett år vilket blir 15 000 – 42 700. Den här skattningen är låg i förhållande till vår skattning och det går inte att validera vår modell genom jämförelse mot Skatteverkets. Skatteverkets metod baseras på impulsdata medan det endast är en komponent i vår modell vilket är en förklaring till den låga skattningen som Skatteverket tagit fram.

I en evalveringsstudie från SCB (2014) skickade man ut en enkät till 15 000 individer där man ställde frågor kring adressen och bostaden med huvudsyftet att evalvera hushållets egenskaper. De som arbetade med enkäten ansåg i efterhand att den var svår att använda till att evalvera folkbokföringsfelet för adressen eftersom alternativ två för första frågan blev för vagt.

Om man gör en rak uppräkningsstudie av svaren från den undersökningen samt kodar de som saknar värde som felaktig adress och nästan rätt adress (alternativ två) som rätt adress får man en skattning (för felaktig adress) på 151 000 ± 33 000 individer vilket betyder ett osäkerhetsintervall mellan 118 000 – 184 000, med 95 % konfidensgrad. Svartsbenägenheten är däremot troligen kopplad till observationsvariabeln och eftersom ingen bortfallsanalys med kalibrering gjorts innebär det troligtvis att punktskattningen borde vara högre än 151 000. Vi misstänker alltså en bias i skattningen. Dessutom finns det en tveksamhet till om målvariabeln stämmer väl överens med observationsvariabeln d.v.s. om frågan var tillräckligt välformulerad för att mäta antalet personer folkbokförda på fel adress. Detta medför en viss osäkerhet till skattningen.

Även om tillförlitligheten i skattningen är osäker så ger den här urvalsundersökningen något att förhålla sig till som ändå säger att vi är i närheten (beroende på hur man ser

det) av vad som kan vara den sanna siffran för antalet personer folkbokförda på fel adress.

Om resonemanget kring urvalsundersökningens skevhet i skattningen stämmer (att skattningen borde vara högre än den visar) innebär det att registerstudien underskattar antalet felaktigt folkbokförda. En trolig förklaring till det är att det är svårt att identifiera fel i register genom att enbart använda sig av register och det behövs flera metoder för att få en bra uppfattning. Gränsdragningarna för vad som sedan skattas som felaktig adress har gjorts utifrån deskriptiv analys och resonemang i grupp vilket innebär subjektivt valda gränser vilket även det kan vara en anledning till den misstänkta underskattningen.

Modellen som baseras på register bedöms vara mer osäker än den från evalveringsstudien eftersom registermodellen baseras på antaganden och subjektiva gränsdragningar medan evalveringsstudien är en objektiv urvalsundersökning där bortfallet är den största källan till osäkerhet. Båda modellens osäkerhetsintervall överlappar däremot med varandra vilket innebär att det inte går att dra några slutsatser om själva skattningen eftersom båda modeller kan ha rätt.

Det är möjligt att använda registermodellens komponenter vid design av en ny evalveringsstudie vilket borde leda till lägre osäkerhet samtidigt som resultatet (andelsskattningar för olika grupper) från den evalveringsstudien kan användas för att fortsätta arbetet med registermodellen.

5. Modell – Mätfel av andra egenskaper

Utöver oövertäckningsfel, undertäckningsfel och felaktiga folkbokföringsadresser kan det även finnas mätfel i de övriga uppgifterna som finns om individer i folkbokföringen. Till dessa uppgifter hör t.ex. ålder, kön, civilstånd, medborgarskap och födelseland. Någon metod för att mäta förekomsten av denna form av fel har inte utvecklats inom ramen för

detta projekt. Nedan resoneras i korthet kring hur en mätning av omfattningen av dessa fel skulle kunna ske.

5.1 Urvalsundersökning för att skatta mätfelet

Det som diskuterats är, precis vad som nämns i avsnittet ovanför om jämförelse och validering, att genomföra en urvalsundersökning för ett stort urval individer och med hjälp av telefon, webb- och pappersenkäter ta kontakt med dem för att se var de bor, var de är folkbokförda samtidigt som man frågar om andra egenskaper från RTB stämmer. Vi bedömer att det är den bästa metoden för att ta reda på mätfelet för andra egenskaper samtidigt som ett underlag om felaktig adress samlas in som kan användas som benchmark gentemot den registermodellen som genomförts och även användas för vidareutveckling av registermodellen.

Fördelen med urvalsundersökningen är att man kan fråga om vad som helst och vara, någorlunda, säker på att svaren stämmer med det man frågar om. Det ger dessutom ett objektiva osäkerhetsintervall med grunden i vetenskaplig teori. Utöver det går det att designa undersökningen för att täcka flera redovisningsgrupper ifall andra typer av analyser ska göras och det går att använda lärdomar från registermodellen i designen av undersökningen.

Nackdelen med urvalsundersökningen är att svarsbenägenheten förmodligen är starkt korrelerad med målvariabeln; vi hör av oss till individens folkbokföringsadress för att höra med dem om de bor där. För att hantera denna problematik bör man designa undersökningen väldigt noggrant och fokusera kontaktstrategierna mot webb/telefon/besöksintervju.

Referenser

Bengssten, Tor (2015), "Resultatrapport Total befolkningssiffra", SCB BV/REG internrapport 2015-08-20

Clemen, Robert T. (1989), "Combining forecasts: A review and annotated bibliography", International Journal of Forecasting 5 (1989) 559-583

Hem och Hyra (2018), "Allt fler fuskar med folkbokföringen", <https://www.hemhyra.se/nyheter/allt-fler-fuskar-med-felaktiga-adresser> (Hämtad 2018-07-23)

Skatteverket (2013), "En analys om folkbokföringsfelet, dess kostnader och konsekvenser", rapport, 2013-04-03

SCB (2014), "Evalvering av Census 2011", SCB PCA/MIH internrapport 2014-02-10

SCB (2015), "Övertäckning i Registret över totalbefolkningen – en registerstudie", Befolkning och välfärd 2015:1

Bilaga

Den här bilagan innehåller en beskrivning av samtliga indikatorer som använts till övertäckningen. Det finns två typer av indikatorer, antingen indikeras att individen tillhör RTB eller att individen tillhör övertäckningen. Nedan listas samtliga indikatorer med respektive vikt.

1. **Gränspendling (vikt hög)**, indikerar att objektet tillhör RTB-populationen. Om inaktivitet sammanfaller med registrerad gränspendling finns det starka skäl att tro att personen tillhör RTB-populationen. Registervariabeln för gränspendling är en mycket tillförlitlig variabel. Denna regel granskar bara det aktuella året och fungerar därför lika bra för alla undersökta år. En brist är att det endast finns data för gränspendling till Norge och Danmark.
2. **Högskolestudier (vikt hög)**, indikerar att objektet tillhör RTB-populationen. Om inaktivitet sammanfaller med högskolestudier finns det skäl att tro att personen tillhör RTB-populationen. För att avgöra om högskolestudier pågår måste personen blivit registrerad för mer än 10 högskolepoäng. I praktiken motsvarar den en halv termin med heltidsstudier. Regeln granskar bara det aktuella året och fungerar därför lika bra för alla undersökta år.
3. **Inaktivitet följt av död (vikt normal)**, indikerar att personen tillhör RTB-populationen. Om en person markeras som död har det vanligen skett i Sverige och det är inte otroligt att en tids sjukdom före döden har medfört att dessa personer syns mindre i registren. Därför finns det skäl att tro att personen tillhör RTB-populationen. Regeln granskar efterföljande år, vilket gör att regeln kan missa inaktiva som ännu inte har dött.
4. **Återkommande aktiviteter i RTB utan aktivitet i STATIV och IoT (vikt normal)**, indikerar att personen tillhör RTB-populationen. En person som inte gör avtryck i STATIV eller IoT, men har återkommande RTB-händelser som gör personen aktiv dessa år, kan antas tillhör RTB-populationen även de mellanliggande åren. Personen har bevisat att den tillhör RTB-populationen utan att det syns i STATIV eller IoT och eftersom RTB-händelser är av karaktären sällanhändelser är det troligt att personen tillhör RTB-populationen även mellanliggande år. Den valda regeln kräver data från minst 6 år varav minst 3 ska ha en RTB-händelse.
5. **Inaktiv med aktiv RTB-familjemedlem (vikt låg)**, indikerar att personen tillhör RTB-populationen. En RTB-familjemedlem är en person som har samma familje-ID som den undersökta personen. Om en person som är inaktiv har någon RTB-familjemedlem som är aktiv har personen ändå en stark koppling till att själv tillhöra RTB-populationen. Regeln granskar bara det aktuella året och fungerar därför lika bra för alla undersökta år. Familje-ID är personnumret för den äldste personen i en RTB-familj.
6. **Inaktiv registrerad på församlingen skriven (vikt låg)**, indikerar att personen tillhör RTB-populationen. På församlingen skriven tyder på att myndigheten vet att personen åtminstone har funnits på församlingen, men inte haft någon adress. Personer som registreras på församlingen skriven kan mycket väl vara sådana personer som har små eller inga aktiviteter i de aktuella registren. Det är därför troligt att personen finns kvar och att personen tillhör RTB-populationen. Regeln granskar bara det aktuella året och fungerar därför lika bra för alla undersökta år.
7. **Från inaktivitet till aktivitet på samma adress (vikt normal)**, indikerar att personen tillhör RTB-populationen. En person som varit inaktiv och blir aktiv på samma adress som den tidigare haft. Det är rimligt att tro att personen har funnits på adressen men inte gjort avtryck i registren. Det får därför fungera som en indikation på att personen tillhör RTB-populationen. Regeln granskar efterföljande år, vilket gör att man riskerar att missa inaktiva som ännu inte blivit aktiva på samma adress.
8. **Utländska medborgare som har immigrerat efter 60 års ålder (vikt normal)**, indikerar att personen tillhör RTB-populationen. En person som immigrerar efter 60 års ålder antas ha tjänat ihop till pension i något annat land. Det gör att personen troligtvis inte gör avtryck i IoT eller STATIV och eftersom RTB-händelser inte inträffar så ofta kan personen se inaktiv ut fast den tillhör RTB-populationen. Regeln granskar bara det aktuella året och fungerar därför lika bra för alla undersökta år.

9. **Personer med tillräckliga medel (vikt normal)**, indikerar att personen tillhör RTB-populationen. En person som har tillräckliga medel som grund för bosättning vid immigration kan antas inte behöver jobba och därför är det troligt att en sådan person inte kommer att göra avtryck i registren. Regeln granskar bara det aktuella året och fungerar därför lika bra för alla undersökta år.
10. **Inaktiv med ränteinkomster (vikt hög)**, indikerar att personen tillhör RTB-populationen. En person med ränteinkomster som är två prisbasbelopp eller större antas klara sig utan andra inkomster. Dessutom bör en person med höga ränteinkomster ha en ganska stark anknytning till Sverige. När det finns möjlighet till försörjning och anknytning till Sverige är det en indikation på att personen tillhör RTB-populationen. Regeln granskar bara det aktuella året och fungerar därför lika bra för alla undersökta år.
11. **Familjeinkomst som är större än ett prisbasbelopp (vikt normal)**, indikerar att personen tillhör RTB-populationen. Om en inaktiv person har en RTB-familjemedlem med en inkomst över ett prisbasbelopp räknas det som en indikator på att personen tillhör RTB-populationen. Någon i RTB-familjen har en registrerad inkomst så det finns i alla fall någon försörjning. Regeln granskar bara det aktuella året och fungerar därför lika bra för alla undersökta år.
12. **Familjeinkomst som är större än 0 (vikt låg)**, indikerar att personen tillhör RTB-populationen. Detta är en svag indikator men om det finns någon i RTB-familjen som har en inkomst i Sverige får det fungera som en indikator för att personen tillhör populationen. Regeln granskar bara det aktuella året och fungerar därför lika bra för alla undersökta år.
13. **Inaktiv följt av emigration (vikt hög)**, indikerar att personen tillhör övertäckningen. Om en person är inaktiv för att ett senare år bli markerad som emigrerad är det troligt att personen var emigrerad i redan i samband med inaktiviteten. Det tar en tid att kontrollera att personen verkligen har emigrerat innan registrering sker. Regeln granskar efterföljande år vilket gör att regeln kan missa inaktiva som ännu inte blivit registrerade som emigranter.
14. **Högskolestudier följt av inaktivitet för utländska medborgare (vikt normal)**, indikerar att personen tillhör övertäckningen. En utländsk student som kommer till Sverige för att studera ska folkbokföra sig här om studierna är längre än ett år. Om studenten sedan återvänder hem och inte anmäler emigration efter avslutade studier blir denne inaktiv. Regeln granskar föregående år, vilket gör att regeln kan missa inaktiva som studerat före år 2000.
15. **Studier som grund för bosättning och inaktiv (vikt normal)**, indikerar att personen tillhör övertäckningen. En person som har studier som grund för bosättning och inte finns med i högskoleregistret har troligen avslutat sina studier. Om personen då också är inaktiv är det en indikation på att studierna är klara och personen har lämnat Sverige. Regeln granskar bara det aktuella året och fungerar därför lika bra för alla undersökta år.
16. **Immigration följt av inaktivitet samt ingen familj i Sverige (vikt låg)**, indikerar att personen tillhör övertäckningen. En person som nyligen immigrerat och som inte syns i något register och inte har någon känd familj i Sverige. Normalt syns nyligen immigrerade personer i ett eller flera register. Under dessa omständigheter är troligt att personen har lämnat Sverige. Regeln granskar föregående år vilket gör att regeln kan missa inaktiva som immigrerat före år 2000.
17. **Immigration följt av aktivitet i IoT följt av inaktivitet (vikt hög)**, indikerar att personen tillhör övertäckningen. En person som immigrerar för att arbeta i Sverige blir aktiv i IoT. Om sedan person lämnar Sverige utan att anmäla emigration kommer den att bli inaktiv. En person som kommit hit för att arbeta under en kort period kan mycket väl sakna kunskap eller intresse att själv anmäla emigration. Regeln granskar två föregående år, vilket gör att regeln kan missa inaktiva som immigrerat före år 2000.
18. **Aktivitet i IoT följt av inaktivitet följt av ny aktivitet i IoT på ny adress (vikt normal)**, indikerar att personen tillhör övertäckningen. En person som är aktiv i IoT (jobbar) för att sedan bli inaktiv och aktiv igen fast på en ny adress är en indikation på att personen har varit borta under en period och när denne kommer tillbaka syns det på att den har en ny adress. Detta mönster tyder på att personen varit borta något år utan att ha anmält emigration och när den kommer tillbaka anmäler den en ny adress. Regeln granskar föregående år och efterföljande år, vilket gör att regeln kan missa inaktiva som har aktivitet före eller efter de undersökta åren.

19. **Inaktiv utan känd hemvist (vikt normal)**, indikerar att personen tillhör övertäckningen. En person som saknar känd hemvist och samtidigt är inaktiv är en klar indikation på att personen tillhör övertäckningen. Regeln granskar bara det aktuella året och fungerar därför lika bra för alla undersökta år.
20. **Sen avisering och inaktiv (vikt hög)**, indikerar att personen tillhör övertäckningen. En person som försvinner från RTB utan markering om emigration eller död tyder på en sen avisering av någon av händelserna. Då är det troligt att det har gällt hela den inaktiva perioden och det bör röra sig om övertäckning. Regeln granskar efterföljande år, vilket gör att regeln kan missa inaktivitet om personen inte har förvunnit före 2012.
21. **Inaktivitet och inkomst som är 0 efter att ha varit mer än 0 (vikt låg)**, indikerar att personen tillhör övertäckningen. En person som blir inaktiv samtidigt som inkomsten blir noll efter att ha haft inkomst tidigare, är en indikation på att personen troligen lämnat Sverige. Regeln granskar föregående år, vilket gör att regeln kan missa inaktiva som haft en inkomst före år 2000.
22. **Utländsk medborgare som blir ensam i sin familj (vikt låg)**, indikerar att personen tillhör övertäckningen. En utrikes född person eller utländsk medborgare som får förändrade familjeförhållanden och blir ensam i familjen och samtidigt blir inaktiv, är en indikation på att person har lämnat Sverige. Regeln granskar föregående år, vilket gör att regeln kan missa inaktiva som haft andra familjeförhållanden före år 2000.
23. **En yngre familjemedlem finns kvar på samma adress (vikt låg)**, indikerar att personen tillhör övertäckningen. Om en person som tillhör ett familje-ID och detta byts till den aktuella personens familje-ID betyder det att de äldre personerna i familjen har flyttat. Om en familj splittras och en yngre person blir kvar på samma adress samtidigt som den blir inaktiv är det en indikation på övertäckning. Regeln granskar föregående år, vilket gör att regeln kan missa inaktiva som haft andra familje-ID före år 2000.
24. **Inaktiv med arbete som grund för bosättning (vikt normal)**, indikerar att personen tillhör övertäckningen. En person som haft arbete som grund för bosättning och blir inaktiv bör kunna ses som en indikation på övertäckning. Regeln granskar bara det aktuella året och fungerar därför lika bra för alla undersökta år.
25. **Inaktiv efter död (vikt hög)**, indikerar att personen tillhör övertäckningen. En person som är död bör inte ingå i RTB-populationen. Den här regeln ska inte få några träffar, men om så är fallet bör det vara frågan om övertäckning. Regeln granskar föregående år vilket gör att regeln kan missa inaktiva som har markerats döda före år 2000.
26. **Inaktiv före immigration (vikt hög)**, indikerar att personen tillhör övertäckningen. En person bör inte kunna vara inaktiv innan den har immigrerat. Det är därför en indikation på övertäckning. Regeln granskar efterföljande år, vilket gör att regeln kan missa inaktiva om personen inte har markerats med immigration före år 2012.
27. **Inaktivitet efter emigration (vikt hög)**, indikerar att personen tillhör övertäckningen. En person bör varken kunna bli inaktiv eller aktiv efter det att den har emigrerat. Om så ändå är fallet tyder det på att något är fel. Är person då markerad som inaktiv är det en indikation på övertäckning. Regeln granskar föregående år, vilket gör att regeln kan missa inaktiva som har markerats emigrerade före år 2000.
28. **Endast garantipension (vikt hög)**, indikerar att personen tillhör övertäckningen. Även personer som inte bor i Sverige kan få garantipension. Därför är inte garantipension i sig en indikator på att personen tillhör RTB-populationen. Om personen inte har någon annan inkomst eller bidrag är det snarare en indikator på att personen tillhör övertäckningen. Regeln granskar bara innevarande år och fungerar därför lika bara för alla undersökta år efter 2003.
29. **Automatisk nolltaxerare över 25 år (vikt låg)**, indikerar att personen tillhör övertäckningen. Personer som med automatik blir nolltaxerare har inte lämnat in någon deklaration och därför finns heller inga tecken på att personen finns i Sverige. Barn och ungdomar som går i skolan och inte har några inkomster är normalt automatiska nolltaxerare utan att tillhöra övertäckningen. Är personen däremot över 25 år bör den ha någon inkomst eller annan ersättning (bidrag).

Regeln granskar bara innevarande år och fungerar därför lika bra för alla undersökta år.

30. **Utgående uppehållstillstånd (vikt hög)**, indikerar att personen tillhör övertäckningen. En person som inte har förnyat sitt uppehållstillstånd utan låtit det gå ut under föregående år finns troligen inte länge kvar i Sverige. Regeln granskar föregående år, vilket gör att den kan missa inaktiva som har utgående uppehållstillstånd före 2010.
31. **Inskrivna vid grundskolan (vikt normal)**, indikerar att personen tillhör RTB-populationen. Finns en person inskriven vid grundskolan den 15 oktober året innan aktuellt år kan man anta det troligt att personen finns kvar även aktuellt år. Kontrollen av föregående år ligger inbyggt i variabeln. Därför granskar regeln innevarande år och fungerar bra för alla år efter 2007.
32. **Anknytning till arbetsmarknaden (vikt låg)**, indikerar att personen tillhör RTB-populationen. Alla personer med någon anknytning till arbetsmarknaden, arbetar, har påbörjat arbete och har slutat arbetat, innebär att det någon gång under innevarande år varit i Sverige och jobbat. Regeln granskar innevarande år och fungerar bra på alla år från och med 2003.
33. **Särskild postadress till utlandet (vikt normal)**, indikerar att personen tillhör övertäckningen. En person som vill ha sin post från skatteverket till utlandet bor troligen inte i Sverige. Regeln granskar innevarande år och fungerar bra på alla år