



Kombinerade mobilnätsdata och enkätdata för att beskriva resmönster – metod-PM

PM: 2025:8

Datum: 2025-04-28

Trafikanalys

Adress: Rosenlundsgatan 54 118 63 Stockholm

Telefon: 010 414 42 00

E-post: trafikanalys@trafa.se

Webbadress: www.trafa.se

Omslagsbild: Adobe Stock

Ansvarig utgivare: Mattias Viklund

Datum: 2025-04-28

Förord

Trafikanalys beskriver svenskarnas dagliga resvanor genom en årlig resvaneundersökning som baseras på enkätsvar.

Trafikanalys genomförde under 2023 en analys av hur mobilnätsdata kan användas för att beskriva våra resvanor. Vi kom då fram till att det är möjligt att bryta ner mobilnätsdata på månad och inom län för att beskriva resande. Samtidigt såg vi brister i att använda mobilnätsdata för att beskriva reserelationer. Nästa steg var att hitta lämpliga metoder för att kombinera mobilnätsdata med den befintliga resvaneundersökningen.

I denna PM redovisas Trafikanalys analys och metodarbete med att kombinera enkätdata från den årliga resvaneundersökningen med mobilnätsdata för att beskriva resmönster.

Björn Tano har varit projektledare, Andreas Holmström och Filippa Egnér har ingått i projektgruppen. Avdelningschefer Sofie Orrling och Andreas Tapani har ingått i styrgruppen.

Stockholm i april 2025

Sofie Orrling

Avdelningschef

Innehåll

1	Sammanfattning	5
2	Inledning	7
2.1	Syfte och mål	7
2.2	Upplägg i PM	7
3	Data och förutsättningar	9
3.1	Data	9
3.2	Förutsättningar	11
4	Metod	13
4.1	Linjär prediktion	13
4.2	Transfer learning	14
4.3	Linjär prediktion tillsammans med Transfer learning	17
4.4	Sammanfattning metod	17
5	Dataanalyser och tester	19
5.1	Korrelation och testning mellan RVU och Mobilnätsdata, brutet på månad	19
5.2	Korrelation och testning mellan RVU och Mobilnätsdata, brutet på län	26
5.3	Korrelation mellan RVU och övrig statistik	32
6	Diskussion och slutsatser	35
7	Publicering	37
8	Arbetet framåt	39
9	Källförteckning	40
	Tabellbilaga	41

1 Sammanfattning

Trafikanalys beskriver svenskarnas dagliga resande med hjälp av en väl utarbetad enkätundersökning, den nationella resvaneundersökningen *Resvanor i Sverige*, även kallad RVU¹. Samtidigt som svarsfrekvenser i enkätundersökningar sjunker ökar tillgången till alternativa datakällor. Det är därför viktigt att arbeta i linje med den utvecklingen. I detta projekt har vi implementerat metoder för att kombinera mobilnätsdata med RVU, för undersökningsåren 2020–2023. Syftet har varit att kombinera de två källorna mobilnätsdata och RVU på bästa sätt, avseende statistikens kvalitet. I denna PM beskriver vi arbetet bakom denna implementering. I en tidigare analys² kom vi fram till att resorna i mobilnätsdata kan brytas ner på månadsnivå samt på länsnivå (resor inom län). Baserat på dessa två nivåer har vi utvecklat och testat metoder för att identifiera den mest lämpliga lösningen för att beskriva resvolymen.

Flera olika alternativ av datakällor och kombinationer av dessa finns för att beskriva resvolymen:

- Mobilnätsdata
- Data från RVU
- En kombination av båda källorna
- Andra datakällor

Vår utgångspunkt var att använda metod c), en kombination av mobilnätsdata och RVU. I denna PM beskriver vi hur vi analyserat och testat olika metoder för att hitta den mest lämpliga lösningen. Resultatet visade att en robust blandad modell – som kombinerar linjär blandad modell med Transfer learning – var mest lämpad för att skatta antal resor per månad med de två datakällorna. På länsnivå visade sig en Transfer learning metod vara den mest lämpade metoden för att skatta resvolymen med de två datakällorna. Önskvärt är att använda en robust blandad modell som vi ser att vi kan göra när vi bryter på månad. Men på länsnivå ser vi outliers som påverkar en linjär modell väldigt mycket, så kallade influencers. Då kan det vara riskabelt att blanda in linjär prediktion. Transfer learning är däremot mer motståndskraftigt mot influencers än vad en linjär modell är.

Resultaten från detta projekt ingår inte i officiell statistik³ och publiceras på Trafikanalys webbsida: www.trafa.se/transportmonster/RVU-Sverige/kombinerade-mobilnatsdata-och-enkatdata-beskriver-resmonster-15129/. Publikationen innehåller både interaktiva tabeller och figurer.

¹ Trafikanalys, "Resvanor, Sverige"

² Trafikanalys, "PM 2023:6 Hur väl kan mobilnätsdata beskriva våra resvanor?"

³ SCB, "Sveriges officiella statistik"

2 Inledning

Kunskap om hur svenskarna reser utgör ett viktigt kunskapsunderlag för att kunna utforma ett framtida transportsystem som på ett hållbart sätt kan möta befintliga och kommande samhällsutmaningar. Det finns därför ett behov av att ta fram metoder för att fånga våra resmönster på ett så kostnadseffektivt sätt som möjligt utifrån de bästa källorna som är tillgängliga.

Traditionellt sett har resvaneundersökningar använts för detta syfte, men i och med att de flesta bär på en mobiltelefon har mobilnätsdata, det vill säga information om vilka mobilmaster vi är uppkopplade mot, föreslagits som källa för att komplettera dagens resvaneundersökningar. Dagens metoder för resvaneundersökningar har dessutom vissa utmaningar avseende problem med täckning, kostnader och uppgiftslämnarvärda. Det finns även ett behov av att fånga fler och mer detaljerade reserelationer, vilket mobilnätsdata kan bidra med.

Under 2023 analyserade⁴ vi om och hur Telias mobilnätsdata skulle gå att använda för att beskriva svenskarnas resmönster. Slutsatserna var dels att vi såg vissa problem med tillförlitligheten i Telias data, dels att vi såg en potential i att använda mobilnätsdata. För mobildata saknas dock information om ärendet, färdstätt och vem som genomför resorna som är en viktig del i resvaneundersökningarna.

2.1 Syfte och mål

Som ett nästa steg föreslog vi att undersöka hur mobilnätsdata skulle kunna kombineras med den nationella resvaneundersökningen RVU⁵. Trafikanalys har utifrån slutsatserna från det projektet därför genomfört ett arbete för att undersöka hur en sådan implementering skulle kunna genomföras. I denna PM beskriver vi slutsatserna från detta implementeringsarbete i syftet att utveckla de metoder vi har för att fånga in resvanor.

Målet var att komma fram till robusta metoder som kan användas för att dels beskriva resmönster, dels kan komma att användas i andra sammanhang då mobilnätsdata finns tillgänglig, utöver enkätdata. Ett större mål är med andra ord att hänga med i utvecklingen och använda mobilnätsdata i kombination med enkätdata även i andra sammanhang, men det innebär inte att vi kan kasta traditionella enkätundersökningar i papperskorgen.

I bästa fall kan mobilnätsdata användas i kombination med enkätdata till en början, i framtiden kan det eventuellt bli aktuellt att gå över helt till mobilnätsdata men där är vi inte nu. Fokus i detta projekt har varit att komma fram till, samt redogöra för, robusta metoder som kan användas för att kombinera mobilnätsdata med resvaneundersökningens enkätdata.

2.2 Upplägg i PM

I denna PM beskriver vi metoder för kombinerande av de två datakällorna mobilnätsdata och RVU. Vi beskriver datakällornas storheter avseende antal resor samt korrelationen mellan datakällorna i kapitel 3, därefter metoderna i kapitel 4.

⁴ Trafikanalys, "PM 2023:6 Hur väl kan mobilnätsdata beskriva våra resvanor?"

⁵ Trafikanalys, "Resvanor, Sverige"

I kapitel 5 redogör vi för analyser och metodval utifrån dessa, för resor per månad, per inom län samt per färdssätt och syfte.

Slutligen har vi sammanfattande slutsatser och en kortare diskussion i kapitel 6, vi redogör för ny publicering i kapitel 7 och skriver om möjligheter framöver i kapitel 8.

3 Data och förutsättningar

I Trafikanalys förra projekt drog vi slutsatsen att det fanns en god korrelation mellan Mobilnätdata och RVU för åren 2019–2021, men att det var stor nivåskillnad på antal resor totalt sett. Sedan dess har Telia genomfört metodförändringar i hur de räknar upp antal resor. Se svar från Telia⁶ angående varför deras dataunderlag förändrats:

"Den nya versionen är stabilare än den tidigare och påverkas inte lika mycket av slumpmässiga fluktuationer i signalkvalitet. Den innehåller också förbättringar i vår möjlighet att extrapolera från våra användare till befolkning, en utveckling som skett i samarbete med SCB. Ytterligare har vi utökat dataunderlaget med företagsabonnemang från och med 2020. Över lag är ett av resultaten att siffrorna är lägre än de var i vår tidigare version, V1.

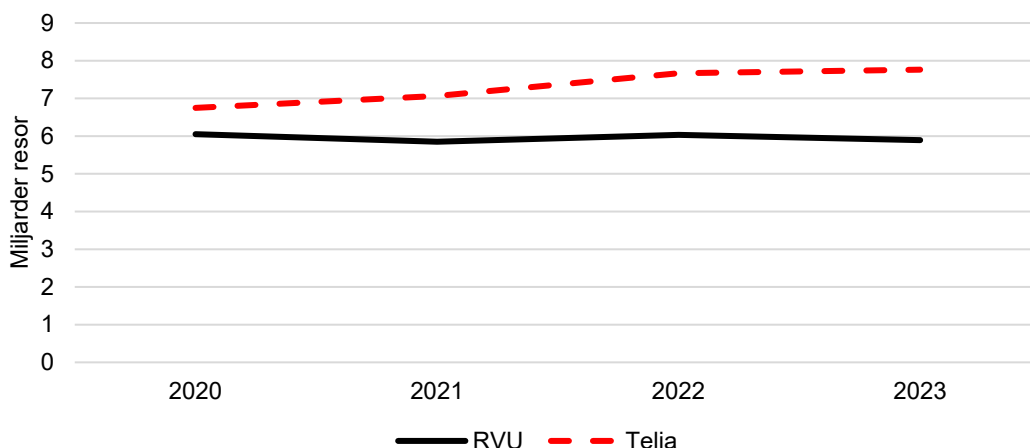
Datan är jämförbar från år till år, men år 2019 hade vi inte med företagsabonnemang vilket kan innebära att antalet resor är något underskattade."

Med anledning av att Telia har justerat sin beräkningsalgoritm har nivåskillnaderna mellan de två datakällorna har minskat. Det totala antalet resor är lägre i den nya versionen av Telias data jämfört med den tidigare versionen. Det är svårt att tolka resultaten från Mobilnätdata år 2019. Vi har därmed valt att exkludera 2019 års data i och använder endast data för åren 2020–2023 i analyserna i denna PM.

Telias mobilnätdata innehåller en variabel som beskriver signalkvalitet, och Telia bedömde att den bästa uppskattningen av det totala antalet resor erhålls genom att använda hela datamaterialet, det vill säga oavsett signalkvalitet. I denna PM har vi inkluderat samtliga signaler från Telias data, i enlighet med Telias rekommendationer från vår tidigare analys.

3.1 Data

Se Figur 1 för en sammanställning av antal resor per källa. För åren 2020–2023 skiljer sig utvecklingen mellan RVU och mobilnätdata åt. Medan antal resor enligt RVU har legat på en relativt stabil nivå, har antalet enligt mobilnätdata ökat. Eftersom utvecklingen visar olika riktningar för några år ger det en korrelation mellan de två källorna som är negativ, med ett värde på $-0,20$.



Figur 1. Antal delresor enligt RVU och antal resor enligt Telias mobilnätdata i miljarder, åren 2020–2023.

⁶ Svar i mejl till Trafikanalys den 9 september 2024.

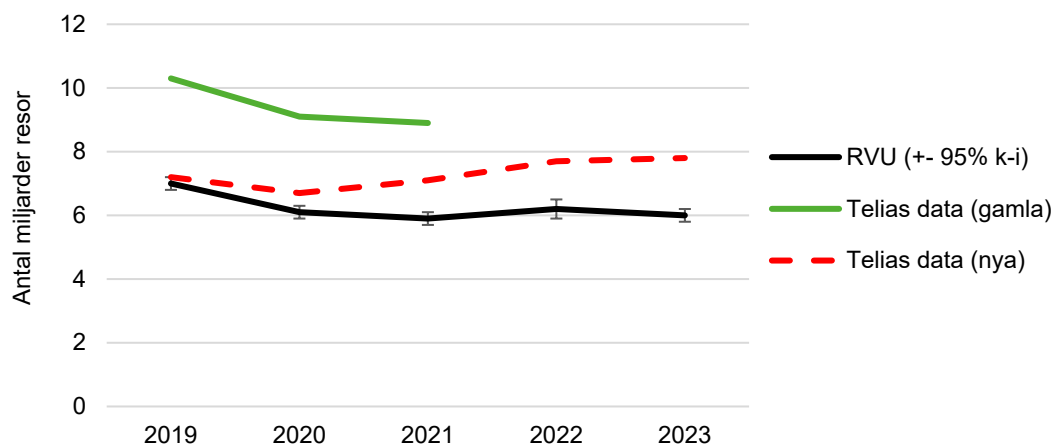
När vi skriver om resor i RVU avses alltid delresor. Definitionen av delresor i RVU ligger närmast definitionen av resor i Telias data vid en jämförelse mellan de två källorna. Mer information om detta finns beskrivet i föregående PM⁷.

I Tabell 1

samt i det efterföljande diagrammet i Figur 2 redovisas antal resor enligt båda källorna. Analysen i föregående PM byggde på Telias gamla data. För jämförelsens skull inkluderas därför även Telias data från den tidigare analysen, innan revideringen i deras siffror. I samtliga analyser i denna PM används Telias nya data.

Tabell 1. Antal resor i miljarder i de båda källorna, åren 2019–2023.

År	RVU (+/- 95% k-i)	Telias data (gamla)	Telias data (nya)
2019	7,0 (0,2)	10,3	7,2
2020	6,1 (0,2)	9,1	6,7
2021	5,9 (0,2)	8,9	7,1
2022	6,2 (0,3)		7,7
2023	6,0 (0,2)		7,8



Figur 2. Antal delresor enligt RVU med konfidensintervall och antal resor enligt Telia (enligt nya och gamla metoden) i miljarder, åren 2019–2023.

Anmärkning: 2019 är inte jämförbart med övriga år eftersom Telia bytt metod.

Telias data har, som beskrivits ovan, genomgått förändringar sedan den förra analysen. Trots detta ser vi att mobilnätdata fortfarande uppvisar liknande brister som vi identifierade för åren 2019–2021 i vår tidigare analys. Vid nedbrytning är det endast antal resor inom län samt per månad och år som kan redovisas med Telias data för att ge tillförlitliga skattningar.

Utöver detta kan vissa resor inom och mellan vissa kommuner i storstadsområdena redovisas, men i detta projekt har vi valt att fokusera på antal resor per inom län samt per månad.

I fortsättningen i detta projekt benämner vi Telias mobilnätdata som enbart Mobilnätdata, förutom i diagram där det står Telias data.

⁷ Trafikanalys, "PM 2023:6 Hur väl kan mobilnätdata beskriva våra resvanor?"

3.2 Förutsättningar

Nedan sammanfattas förutsättningarna för våra två datakällor och den önskade synergien mellan dem.

Vi vill veta hur många nationella resor som görs av befolkningen i Sverige, en given tidsperiod. De yngsta faller bort ur intressepopulationen, som gör att målpopulationen är de personer som är 6–84 år i RVU respektive 6 år och äldre i Mobilnätsdata. Det sökta antalet resor, som vi definierar som Y , blir därför

$$Y = \sum_{\{k \in U\}} \sum_{t=1}^T y_{kt},$$

där U är populationen (6–84 år i RVU, 6 år och äldre i Mobilnätsdata), T är given tidsperiod och y_{kt} är antal resor per person k och tidsintervall t .

För att skatta antal resor en given tidsperiod från RVU-data har vi estimatoren $\hat{Y}$, som kan beskrivas enligt:

$$\hat{Y} = \sum_{k \in U} \sum_{t=1}^T \delta_{kt} w_{kt} y_{kt},$$

där:

- $\delta_{kt} = 1$ om personen k undersöks tiden t , annars $= 0$
- w_{kt} = vikten för person k vid tiden t
- y_{kt} = Antal resor per person k och tiden t

För att skatta antal resor en given tidsperiod från Mobilnätsdata har vi antal mobiltelefoner en given tidsperiod, enligt:

$$X_t = \sum_{d \in D} x_{dt},$$

där:

- D = Populationen av mobiltelefoner
- x_{dt} = mobiltelefon x vid tiden t

Registrering av dessa mobiltelefoners signaler räknas upp till befolkningen av Telia. Skattning av totalt antal resor en given tidsperiod från Mobilnätsdata kan då beskrivas enligt följande:

$$X_t^U = \sum_{k \in U} \alpha_{kt} y_{kt} w_{kt},$$

där:

- α = antal upptäckta mobiltelefonsignaler
- y_{kt} = Antal resor per person k och tiden t
- w_{kt} = Telias uppräknings till befolkningen

Optimalt är att skattningarna från RVU och Mobilnätsdata per år överensstämmer med varandra och med det verkliga antalet resor, uttryckt som:

$$\hat{Y} = \sum_{k \in U} \sum_{t=1}^T \delta_{kt} w_{kt} y_{kt} = X_t^U = \sum_{k \in U} \alpha_{kt} y_{kt} w_{kt} = Y$$

En sådan överensstämmelse skulle ge tyngd åt RVU och vi skulle inte behöva använda Mobilnätsdata mer i denna analys. Det är förstås en utopi att dessa två källor skattar exakt samma antal resor, därav behöver vi vidare analysera hur vi bäst kombinerar de två källorna.

Viktigt att notera är att RVU frågar individer direkt om deras resor och platser, medan Mobilnätsdata bygger på signaler från mobiltelefoner till mobilmaster. Båda källorna innehåller osäkerheter, data kan registreras och tolkas felaktigt i båda källorna.

I avsnitten 0 och 5.2 beskriver vi hur vi testar och modellerar skattningarna med olika metoder, nedbrutet på månad respektive län. Läs mer om metoderna i kapitel 4.

4 Metod

I föregående projekt analyserade vi datamaterialen för att undersöka hur Mobilnätsdata kan användas tillsammans med RVU. Vi kom fram till att Mobilnätsdata kan vara lämplig att använda för att redovisa antal resor på månadsbasis samt för antal resor inom län på årsbasis. Vi vet samtidigt att vi inte kan använda endast Mobilnätsdata för att ersätta RVU-data, eftersom vi både ser oklarheter och har dålig insyn i hur Mobilnätsdata data tas fram. Vår plan i denna PM var därför att komma fram till en lämplig metod för att kombinera de två datakällorna, med lämplig avser vi kvalitetsaspekten och robustheten.

Innan en kombination av de två datakällorna kunde genomföras behövdes dock ytterligare analyser. Dessa analyser fokuserade på den möjliga redovisningen av antal resor nedbrutet på tid och geografi samt hur implementeringen skulle göras.

Vi ville även testa lämpligheten att skatta antal resor per färdstätt och syfte med hjälp av den data vi har i RVU för dessa två variabler.

I avsnitt 0 har vi testat om Mobilnätsdata korrelerar med RVU-data vid nedbrytning på tid (månad) för att kunna analysera vilken implementeringsmetod som är mest lämplig. I avsnitt 5.2 har vi gjort motsvarande analys för nedbrytning på geografi (län). I avsnitt 5.3 har vi jämfört RVU med andra datakällor brutet på färdstätt. Nedan beskrivs de metoder vi använt för att kombinera och analysera datakällorna.

4.1 Linjär prediktion

En metod för att kombinera de två datakällorna och därmed skatta antal resor, är att använda sig av linjär prediktion. Dels har vi vår RVU-data med urvalsfel, dels har vi Mobilnätsdata.

En linjär blandad modell kan då uttryckas som:

$$\hat{Y}_t = \beta_0 + \beta_1 X_t + v_t + e_t$$

där $\hat{Y}_t$ i vårt fall är skattade antal resor i RVU en viss månad eller ett visst län ett år, β_0 och β_1 är regressionskoefficienter enligt β_0 = interceptet och β_1 = lutningskoefficienten, X_t = Antal resor i Mobilnätsdata en viss månad eller ett visst län ett år, t = månad eller län, v = slumpfelet och e = urvalsfelet.

För att konstruera den bästa linjära prediktionsbaserade (LP) skattningen kan vi, enligt Zhang⁸, använda oss av Fay och Herriots modell⁹ för "small area estimation", för att skatta grupperna (månad eller län). Fay och Herriot använder sig av en variant av en James-Stein¹⁰ estimator, för att minska variansen genom att krympa individuella skattningar mot modell-skattningen.

Målet är att ta fram fördelningen över månad respektive län, i form av andelar per månad respektive län. Då kan vi använda oss av formeln:

$$\hat{Y}_t^{LP} = \hat{Y}_{\hat{p}_t^{LP}},$$

⁸ Zhang, "Disaggregation of trips using MNO data".

⁹ Fay, R.E. and Herriot, R.A, "Estimates of income for small places: An Application of James-Stein Procedures to Census Data".

¹⁰ James and Stein, Berkeley Symp. on Math. Statist. and Prob., "Estimation with Quadratic Loss".

där

$$\hat{p}_t^{LP} = \frac{\hat{\gamma}_t \hat{Y}_t + (1 - \hat{\gamma}_t)(\hat{\beta}_0 + \hat{\beta}_1 X_t)}{\sum_{l=1}^T \hat{\gamma}_l \hat{Y}_l + (1 - \hat{\gamma}_l)(\hat{\beta}_0 + \hat{\beta}_1 X_l)},$$

där vikten γ_t beror på hur stor osäkerhet som finns i modellen respektive urvalet:

$$\gamma_t = \frac{\sigma_v^2}{\sigma_v^2 + V(e_t)},$$

där $\sigma_v^2 = V(u_t)$.

Sammanfattning:

- Om urvalsfelet $V(e_t)$ är stort $\rightarrow$ Vi litar mer på modellskattningen.
- Om modellfelet σ_v^2 är stort $\rightarrow$ Vi litar mer på skattning från RVU-data, $\hat{Y}_t$.

4.2 Transfer learning

Om det finns outliers som påverkar en linjär modell avsevärt, så kallade influencers, så är det en indikation på att prediktion baserat på modellen är riskabel. I dessa fall kan metoden Transfer learning vara ett bättre alternativ för skattningarna, än linjär prediktion.

Transfer learning är en metod inom maskinlärning och statistik där en modell som har tränats på en uppgift används som utgångspunkt för att förbättra prestandan på en annan, relaterad uppgift. I stället för att träna en modell från grunden för varje ny uppgift, använder Transfer learning den kunskap som redan har inlärts från en annan uppgift för att förbättra resultaten.

Transfer learning-metoden i vårt projekt ger oss en estimator som liknar en James Stein-estimator¹¹. Denna metod syftar till att minimera det totala medelkvadratfelet (på engelska förkortat MSE, som står för Mean Squared Error) över alla estimatorer. I vårt fall handlar det om månadsestimatorer respektive länssestimatorer för antal resor. För att uppnå detta tillåter metoden en liten grad av bias för varje estimator, men kan reducera variansen av alternativa icke-biased estimatorer. I detta projekt innebär Transfer learning att vi använder RVU:s estimatorer som icke-biased men de kan få relativt stora varianser på grund av den begränsade urvalsstorleken.

Transfer learning med en James Stein-estimator hjälper oss att vikta RVU och Mobilnätsdata med en krympningsfaktor på RVU-data. Vi vill göra en optimal avvägning hur mycket vikt vi ska lägga på RVU:s estimatorer jämfört med alternativen, som är biased men med negligerbara varianser och i detta fall härledda från Mobilnätsdata, utan att direkt anta ett linjärt samband mellan de två källorna (vilket görs vid Linjär prediktion). Vilken vikt vi ska lägga på RVU, och därmed även på Mobilnätsdata, beror på hur precisa skattningarna från RVU-data är samt hur de två källorna korrelerar med varandra.

För att exemplifiera metoden i vårt projekt: I stället för att använda lika stor vikt vid de två källorna för att skatta Stockholms läns antal resor ett år, vore det bättre att ta hänsyn till optimala skattningar av de andra länens resor. Detta är särskilt fördelaktigt om målet är att reducera totala MSE för länssestimatorerna.

Nedan följer metodiken med Transfer learning ur en matematisk vinkel. Metoden¹² kombinerar en traditionell skattning, RVU:s skattning $\hat{Y}_t$, med en proxyvariabel X_t/X , Mobilnätsdata i

¹¹ James and Stein, Berkeley Symp. on Math. Statist. and Prob., "Estimation with Quadratic Loss".

¹² Zhang, "Disaggregation of trips using MNO data"

vårt fall, genom att använda en viktningsparameter $\gamma \geq 0$. Denna parameter styr hur mycket vikt man ger till respektive datakälla. Märk skillnaden mot vikten γ_t i $\hat{p}_t^{LP}$ där det var olika vikter per månad eller län men här är det samma totala vikt över alla månader eller län.

- Om proxyvariabeln X_t/X exakt motsvarar ett sant förhållande Y_t/Y , så ger Transfer learning en perfekt skattning utan urvalsfel, jämfört med Linjär prediktion där urvalsfel alltid förekommer.
- Om X_t/X och Y_t/Y är helt okorrelerade, ger Transfer learning ingen förbättring.

Nu vill vi optimera viktningen av de båda källorna vid en kombination av dessa, med hjälp av Transfer learning. Vi beskriver metoden i tre delar:

Del 1)

Om vi endast har RVU-data och ska skatta $p_t = Y_t/Y$ så vill vi minimera den negativa log-likelihooden $-\sum_{t=1}^T \hat{Y}_t \log p_t$, vi vill med andra ord maximera sannolikheten för skattningen. Lösningen, enligt Zhang, blir då

$$-\sum_{t=1}^T \hat{Y}_t \log p_t + \lambda \sum_{t=1}^T (p_t - 1)$$

under kravet att $\sum p_t = 1$, där λ = lagrangemultiplikator för att tillgodose kravet. Lösningen för skattningen blir då $\hat{p}_t = \hat{Y}_t/\hat{Y}$.

Del 2)

Nu vill vi justera sannolikheterna beskrivet med proxydata, det vill säga Mobilnätsdata i vårt fall. Zhang beskriver hur vi kan införa Transfer learning, där proxydata $q_t = X_t/X$ vägs in med en viktningsparameter γ , vilket innebär att man lägger till ett så kallat straff i optimeringsfunktionen för att avvika från q_t . Det nya optimeringsproblemet kombinerar:

- Likelihood från RVU-skattningen $\hat{Y}_t$

och

- En divergens (Kullback-Leibler) mellan p_t och q_t .

Det betyder att man minimerar $-\sum_{t=1}^T \hat{Y}_t \log p_t$ genom:

$$-\sum_{t=1}^T \hat{Y}_t \log p_t + \gamma \sum_{t=1}^T X_t (\log q_t - \log p_t)$$

med villkoret att $\sum p_t = 1$.

Med Transfer learning¹³ kan man då visa följande:

$$\hat{p}_t = \psi(\gamma) \cdot \frac{\hat{Y}_t}{\hat{Y}} + (1 - \psi(\gamma)) \cdot \frac{X_t}{X},$$

där

$$\psi(\gamma) = \frac{\hat{Y}/X}{\gamma + \hat{Y}/X}$$

¹³ Zhang, "Disaggregation of trips using MNO data"

Sammanfattning av del 2:

- När viktningsparametern $\gamma = 0 \rightarrow$ Vi använder endast RVU:s skattning $\hat{Y}_t$.
- När viktningsparametern $\gamma \rightarrow \infty \rightarrow$ Vi går helt över till proxydata (Mobilnätsdata) X_t/X .

Del 3)

I tredje steget beskriver vi hur man, enligt Zhang, väljer den optimala viktningsfaktorn $\psi(\gamma)$ som styr balansen mellan:

- $\hat{p}_t = \hat{Y}_t/\hat{Y}$ (RVU-data)

och

- $q_t = X_t/X$ (Mobilnätsdata)

Man tittar på det förväntade MSE mellan skattningen $\hat{p}_t$ och det sanna p_t , där

$u_t = q_t - p_t$ mäter hur mycket Mobilnätsdata avviker från den sanna andelen, Den sanna andelen p_t är ju dock inte känt i praktiken.

Formeln för det förväntade MSE är:

$$E\left(\sum_{t=1}^T (\hat{p}_t - p_t)^2\right) = \psi^2 \sum V(\hat{p}_t) + (1 - \psi)^2 \sum u_t^2$$

Eftersom p_t inte är känt i praktiken, ersätts ψ med en estimerad version $\hat{\psi}$, baserat på observerade data. För att minimera det förväntade MSE får man, enligt Zhang¹⁴, att:

$$\hat{\psi} = \frac{\hat{t}_u}{\hat{t}_u + \hat{t}_\varepsilon},$$

där:

- $\hat{t}_u = \frac{1}{T} \sum (\hat{p}_t - q_t)^2 - \hat{V}(\hat{p}_t)$ (som ersätter $\frac{1}{T} \sum_{t=1}^T u_t^2$ eftersom p_t ej är känt)
- $\hat{t}_\varepsilon = \frac{1}{T} \sum \hat{V}(\hat{p}_t)$

Den här uppskattningen av minsta MSE används sedan för att beräkna en förbättrad TL-skattning $\tilde{p}_t^{TL}$.

Sammanfattning av del 3:

- Den optimala viktningen mellan de två datakällorna (RVU och Mobilnätsdata) bestäms dels av $\hat{t}_u$, det vill säga hur väl Mobilnätsdata överensstämmer med RVU-data, dels av $\hat{t}_\varepsilon$, det vill säga hur osäker skattningen $\hat{Y}_t$ från RVU är.
- Ju mer källorna överensstämmer, desto lägre blir $\hat{\psi}$, och då blir det enligt formeln

$$\hat{p}_t = \psi(\gamma) \cdot \frac{\hat{Y}_t}{\hat{Y}} + (1 - \psi(\gamma)) \cdot \frac{X_t}{X}$$

mer vikt åt Mobilnätsdata.

- Ju säkrare skattning av RVU-data, $\hat{Y}_t$, desto högre blir $\hat{\psi}$ och då blir det mer vikt åt RVU.

¹⁴ Zhang, "Disaggregation of trips using MNO data"

4.3 Linjär prediktion tillsammans med Transfer learning

För att få en än mer robust metod än Transfer learning kan vi med fördel, om inte någon influencer påverkar modellen mycket, sammanslå Linjär prediktion och Transfer learning i en modell. En modell för att använda både Linjär prediktion och Transfer learning och som metod¹⁵ kan enligt Zhang skrivas som:

$$\hat{p}_t^{MX} = \omega \hat{\mu}_t + (1 - \omega) q_t,$$

där $\hat{\mu}_t = \hat{p}_t - q_t$ och vikten ω beror på hur mycket skattningarna för andel per månad respektive län från RVU ($\hat{p}_t$) och Mobilnätdata (q_t) avviker från varandra:

$$\omega = \frac{\sum_t (\hat{p}_t - q_t)^2}{\sum_t (\hat{p}_t - \mu_t/Y)^2 + \sum_t (\hat{p}_t - q_t)^2}$$

4.4 Sammanfattning metod

Utöver modellerna vi testat att beskriva antal resor med respektive källa var för sig.

Sammanfattningsvis har vi testat följande fem olika metoder för att beskriva antal resor per månad samt antal resor per inom län:

- Enbart RVU-data ($\hat{Y}$)
- Enbart Mobilnätdata (X)
- RVU-data kombinerat med Mobilnätdata med hjälp av skattningar i linjär blandad modell (LP)
- RVU-data kombinerat med Mobilnätdata med hjälp av Transfer learning-metodik (TL)
- RVU-data kombinerat med Mobilnätdata med hjälp av en kombination av LP+TL (MX-estimator)

Det vore önskvärt att ha tillgång till data från flera, helst alla, mobiloperatörer för att få en mer heltäckande bild av antal resor. Det skulle möjliggöra korsanalyser och göra det enklare att avgöra om förändringar eller diskrepanser beror på förändringar i befolkningens faktiska resande eller på skillnader mellan datakällor. Dock har vi tillgång till skattade resor i Trafikanalys övriga produkter, exempelvis *Regional linjetrafik* och *Körsträckor*¹⁶, som vi kan använda för att analysera och förstå eventuella skillnader mellan datakällor.

I avsnitt 5.3 har vi analyserat resultaten från dessa produkter i jämförelse med RVU-data, för att se om RVU:s andelar är lämpliga att använda för att skatta antal resor per färdväg och syfte med resan. Metoden för att kombinera RVU-data och Mobilnätdata, om RVU:s andelar visade sig hålla tillräcklig kvalitet, var med hjälp av rak uppräknings för andelarna i RVU. Det är med andra ord en simpel metod vi använt för att skatta antal resor för färdväg och syfte inom ramen för detta projekt.

Våra analyser och sedermera metoder genererade fördelningen över antal resor per månad, per inom län samt per färdväg och syfte, årligen. Antal resor har därefter viktats till Mobilnätdata totala antal resor per år.

¹⁵ Zhang, "Disaggregation of trips using MNO data"

¹⁶ Trafikanalys, "Regional linjetrafik och Körsträckor"

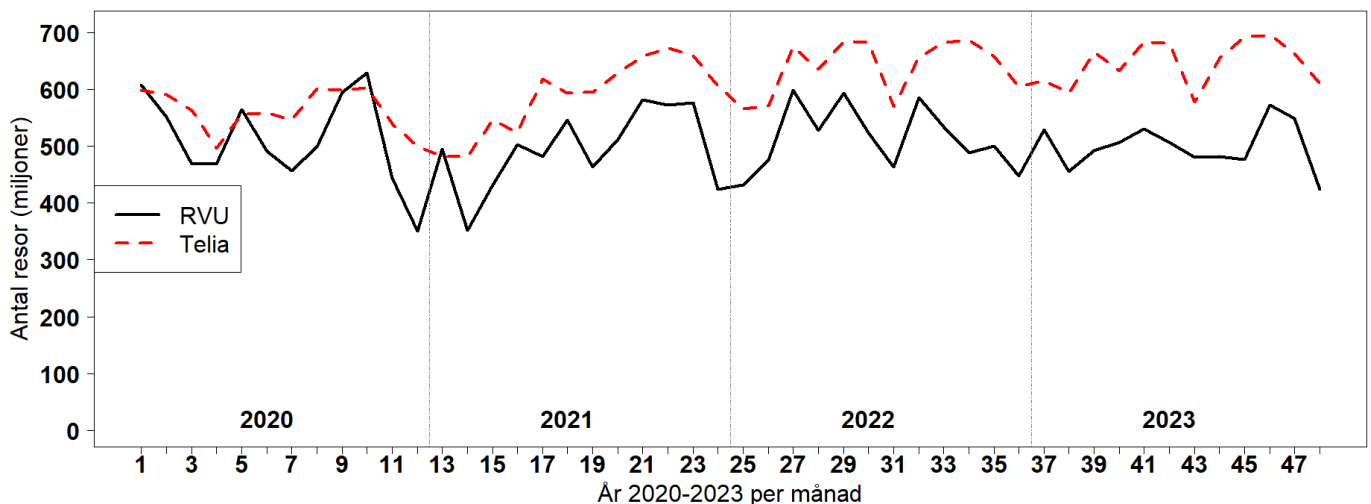
5 Dataanalyser och tester

I detta kapitel beskriver vi hur våra analyser ger oss information om vilken metod som är mest lämplig att använda för att kombinera de två datakällorna brutet på månad i avsnitt 5.1 respektive brutet på län i avsnitt 5.2. Respektive avsnitt börjar med redogörelse för hur väl datakällorna korrelerar med varandra, för att sedan visa på hur väl de olika källorna/metoderna passar för att beskriva antal resor.

I avsnitt 5.3 beskriver vi hur analyser av annan statistik tillsammans med RVU ger oss information vilken metod som är mest lämplig att använda för att kombinera de två datakällorna brutet på färdstätt och syfte med resan.

5.1 Korrelation och testning mellan RVU och Mobilnätsdata, brutet på månad

I Figur 3 visas antal resor enligt de två källorna och täcker månaderna för åren 2020–2023. Mobilnätsdata representeras av en streckad röd linje och RVU-data visas med en heldragen svart linje. Diagrammet visar att de två datakällorna följer varandra relativt väl när de bryts ner på månad, vilket bekräftar en fortsatt god korrelation mellan dem.



Figur 3. Antal resor per månad och år för de två källorna, åren 2020–2023. 1=januari 2020, 48=december 2023.

För att kunna bestämma den bästa möjliga metoden för att kombinera de två datakällorna, brutet på månad, har vi testat hur väl Mobilnätsdata och RVU-data förhåller sig till varandra per år¹⁷. Samtliga tester i detta avsnitt har utförts med programvaran R.

Vi sätter Y som RVU-data och X som Mobilnätsdata. RVU:s månadsestimatorer för antal delresor har approximerade normalfördelningar och estimerade varianser. RVU:s månadsestimatorer behandlas som icke-biased. Mobilnätsdatas antal resor behandlas som konstanter, med negligerbara varianser men med bias. För hypotesprövningarna har vi använt

¹⁷ Zhang, "Disaggregation of trips using MNO data".

oss av χ^2 -test (Chi²-test). Då kan nollhypotesen med tillhörande hypotesprövningar uttryckas enligt följande:

- a) $H_0: X_t = Y_t$, som med ord blir H_0 : Mobilnätsdatas antal resor är lika med RVU:s antal delresor över alla 12 månader ett givet år, där vi antar att RVU:s antal delresor motsvarar det verkliga antalet.
- b) $H_0: X_t \propto Y_t \Leftrightarrow \frac{X_t}{X} = \frac{Y_t}{Y}$, som med ord blir H_0 : Mobilnätsdatas antal resor är proportionell mot RVU:s antal delresor över alla 12 månader ett givet år, där vi antar att RVU:s fördelning över resor speglar den verkliga fördelningen.

Med andra ord testar vi om det är statistiskt bevisat att Mobilnätsdata skiljer sig från RVU-data per månad och år, eller om skillnaderna faller inom ramen för urvalsosäkerheten i RVU. Vi önskar få höga p-värden eftersom det innebär att källorna inte skiljer sig från varandra avseende antal resor och fördelningen av resor per månad. Det önskade resultatet är att vi inte ska kunna förkasta nollhypoteserna. Alternativa hypotesen, H_1 , är att det finns en signifikant skillnad mellan datakällorna avseende antal resor respektive fördelningen av resor per månad.

Resultaten visar att år 2020 inte är tillfredsställande, då vi förkastar nollhypoteserna på samtliga testade signifikansnivåer (α -nivåer). År 2021 uppvisar till viss del tillfredsställande resultat med p-värde större än 0,1 för både nollhypotes a och b. Åren 2022 och 2023 visar positiva resultat med höga p-värden som indikerar att vi inte kan förkasta nollhypoteserna, vilket är önskvärt. Detta innebär att för åren 2022 och 2023 kan vi inte påvisa någon statistisk skillnad för antal resor samt fördelningen av antal resor per månad mellan de två datakällorna. Se p-värden i Tabell 2.

Tabell 2. p-värden för hypotestest a) Mobilnätsdatas antal resor = RVU:s antal delresor över alla 12 månader ett givet år, samt för hypotestest b) Mobilnätsdatas antal resor är proportionell mot RVU:s antal delresor över alla 12 månader ett givet år.

År	Nollhypotes a)	Nollhypotes b)
2020	5,08 x 10 ⁻⁶	1,12 x 10 ⁻⁶
2021	0,12	0,15
2022	0,60	0,63
2023	0,60	0,74

Vidare testar vi Transfer learning (TL) som en implementeringsmetod för estimering av antal resor per månad. För en beskrivning av hur vi använt metoden Transfer learning, se avsnitt 4.2. För detta test analyserar vi måttet Relativ effektivitet(TL) = MSE(TL) / Var(RVU). Detta mått indikerar hur effektiv TL-metoden är jämfört med den direkta RVU-estimatorn, där ett lägre värde innebär en högre effektivitet och därmed en vinst i precision.

Tabell 3 visar Relativ effektivitet med en TL-estimator jämfört med RVU-skattningarna, från upprepade urval (100 000) med återläggning, så kallad bootstrapping, under RVUs design där vi använder RVUs varianser.

Tabell 3. Relativ effektivitet med en Transfer learning-estimator jämfört med RVU-skattningarna, brutet på månad.

År	Relativ effektivitet
2020	0,24
2021	0,077
2022	0,00039
2023	0,0099

Ju lägre Relativ effektivitet i denna Transfer learning-metod (TL), desto mer tillfredsställande resultat. Denna metod ger mer tillfredsställande värden för alla fyra år jämfört med att endast använda RVU som datakälla, inklusive tillhörande osäkerhetsintervall. Exempelvis innebär ett värde på 0,24 för år 2020 att variansen för TL-skattningarna endast är ungefär en fjärdedel av variansen för RVU-skattningarna. I våra tidigare hypotesprövningar (avseende Mobilnätsdata = RVU samt Mobilnätsdata $\propto$ RVU) konstaterade vi att resultaten för år 2020 inte var tillfredsställande och att år 2021 inte hade helt tillfredsställande p-värden. Med TL-metoden förbättras resultaten avsevärt vid kombination av Mobilnätsdata och RVU när data bryts ner på månad och år, jämfört med att endast använda RVU som källa.

För år 2022 och 2023 ser vi mycket låga värden för Relativ effektivitet (TL), vilket indikerar hög effektivitet. Det kan dock finnas än mer robusta modeller. Låga p-värden för Relativ effektivitet behöver inte garantera att modellen är robust. För att förbättra robustheten i resultaten vidareutvecklar vi modellen genom att kombinera en linjär blandad modell med Transfer Learning.

Den linjära modellen skattas med linjär prediktion, vi kallar denna metod för LP i denna PM. Utöver detta kan vi även testa Transfer learning (TL) och slutligen en robust estimator, som vi kallar för modell MX, som kombinerar LP och TL. Med robust menar vi i detta fall att vi kombinerar både linjär prediktion och transfer learning till en modell (MX), i stället för att lita helt på endast Transfer learning.

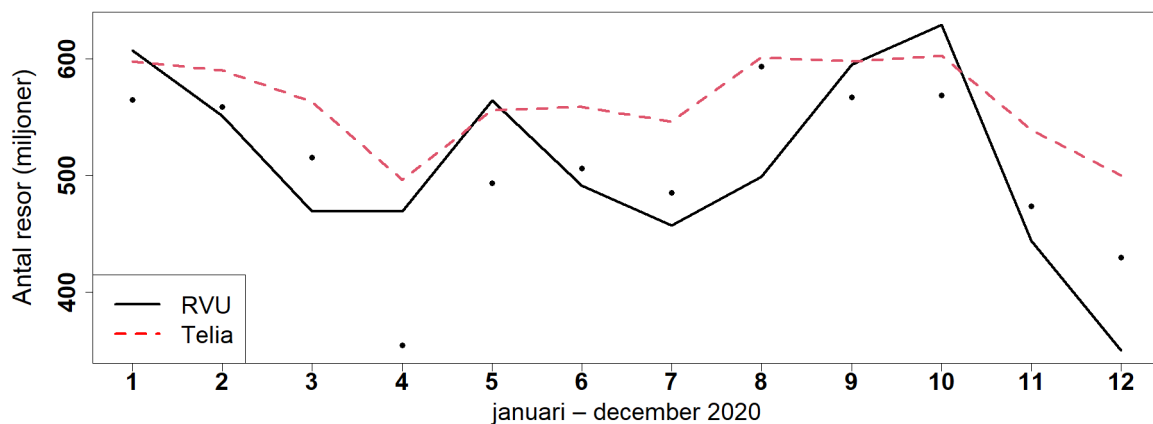
MSE används för att utvärdera en estimators prestanda. Om vi vill jämföra MSE mellan olika datakällor och modeller, som i vårt fall när vi har både RVU, Mobilnätsdata samt kombinationer av dessa, är det mer fördelaktigt att använda Average Relative Root Mean Square Error (ARRMSE). Då normaliserar vi de faktiska värdena som MSE mäter, vilket innebär att skalan av antal resor inte påverkar resultatet utan vi gör en relativ jämförelse mellan källorna. Vi genomför ARRMSE för respektive metod i följande ordning: RVU (som enskild källa), Mobilnätsdata (som enskild källa), LP (skattningar i linjär blandad modell), TL (Transfer learning), MX (Robust estimator som kombinerar LP och TL). Resultaten redovisas i Tabell 4. Skattningar av ARRMSE har tagits fram genom bootstrapping (100 000) under RVU:s design, vid samtliga fem metoder.

Tabell 4. Average Relative Root Mean Square Error (ARRMSE) för resor brutet på månad, per källa/metod och år.

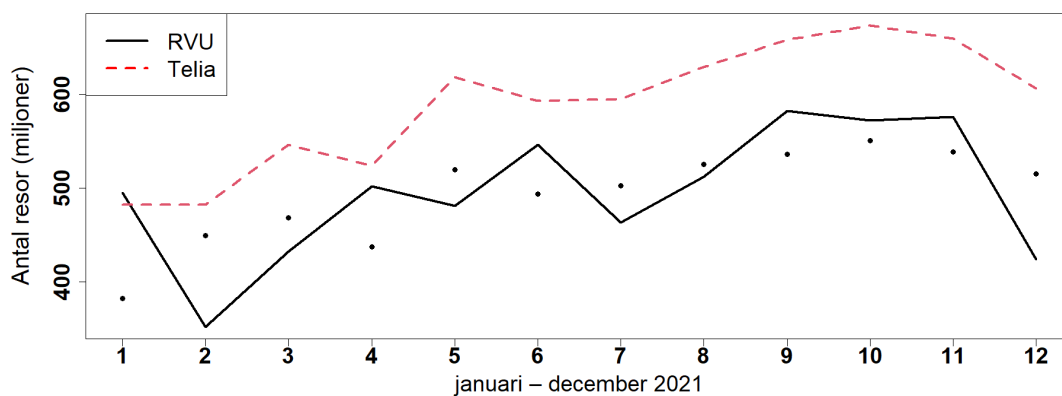
År	RVU	Mobilnätsdata	LP	TL	MX
2020	0,052	0,073	0,026	0,022	0,024
2021	0,083	0,026	0,022	0,021	0,022
2022	0,082	0,0017	0,000059	0,0014	0,00066
2023	0,085	0,0093	0,000079	0,0077	0,0033

ARRMSE är som nämnt ett mått på en estimators prestanda, där ett lägre värde indikerar bättre precision. Vi observerar att RVU har det högsta ARRMSE-värdet åren 2021–2023 jämfört med Mobilnätsdata och övriga metoder. Det betyder att användning av RVU direkt ger minst tillfredsställande resultat avseende högst medelkvadrattfel vid nedbrytning på månad för åren 2021–2023. Vi kan även se att RVU har lägst ARRMSE 2020, och även något lägre än Mobilnätsdata det året, jämfört med övriga år. Mobilnätsdata och övriga metoder har tydligt lägst ARRMSE 2022 och 2023 jämfört med de två första åren.

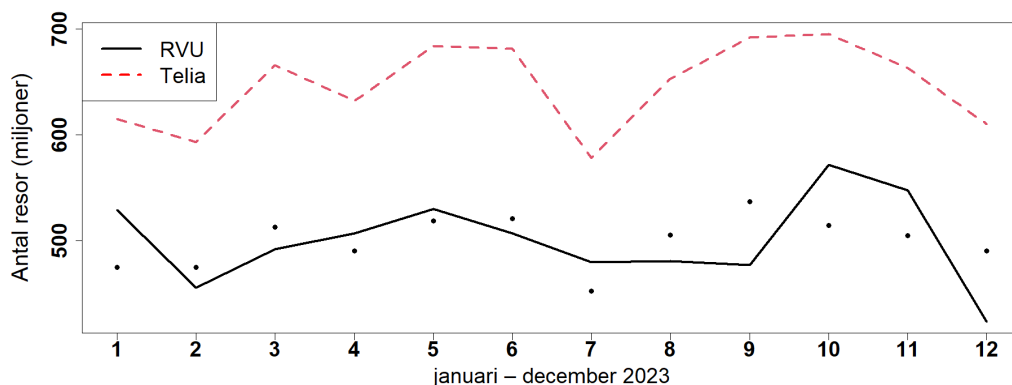
I Figur 4 – Figur 7 visas resultaten per år. De svarta datapunkterna representerar en linjär modell där båda datakällorna är integrerade, den röda streckade linjen visar Mobilnätsdata och den svarta heldragna linjen visar RVU, för januari till december varje år. Vi identifierar vissa outliers, men inga av dessa bedöms som influencers (outliers som har betydande påverkan på en linjär modell).



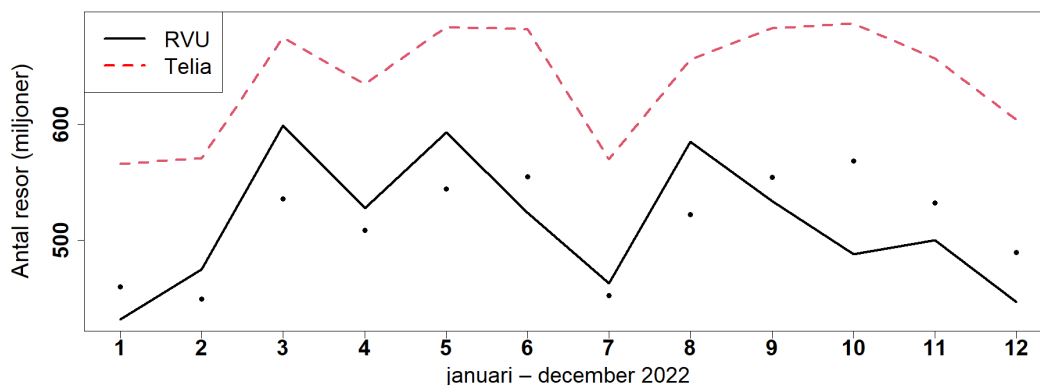
Figur 4. Antal resor per månad år 2020, där 1=januari, 12=december. De svarta datapunkterna i diagrammet markerar skattningarna för en linjär modell.



Figur 5. Antal resor per månad år 2021, där 1=januari, 12=december. De svarta datapunkterna i diagrammet markerar skattningarna för en linjär modell.



Figur 6. Antal resor per månad år 2022, där 1=januari, 12=december. De svarta datapunkterna i diagrammet markerar skattningarna för en linjär modell.



Figur 7. Antal resor per månad år 2023, där 1=januari, 12=december. De svarta datapunkterna i diagrammet markerar skattningarna för en linjär modell.

För att styrka att den linjära modellen bör ingå i metoden ska värden på interceptet eller lutningskoefficienten för regressionslinjen inte uppvisa betydande variation mellan månader. Det innebär att vi inte vill se några stora avvikelser för en enskild månad.

I Tabell 5 redovisas hur interceptet (β_0) och lutningskoefficienten (β_i) förändras när enskilda måtpunkter exkluderas, för åren 2020–2023.

Tabell 5. Värden på raderna β_0 visar hur mycket interceptet förändras och värden på β_i visar hur mycket lutningen förändras om månaden utesluts, i en linjär modell för resor brutet på månad. Åren 2020–2023. Kolumnerna visar månaderna i ordning, där 1 = januari och 12 = december.

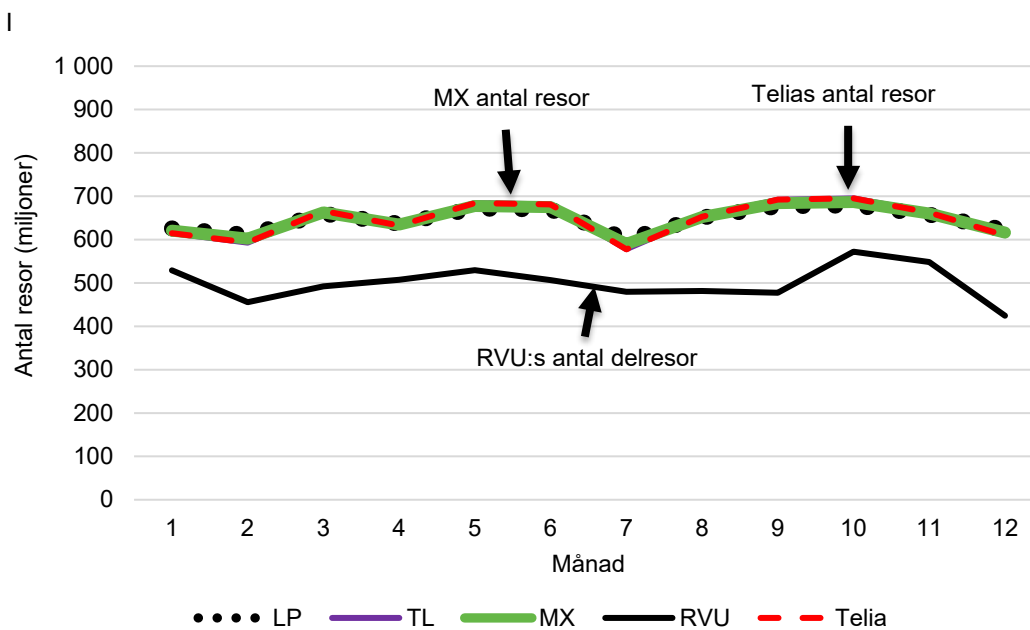
År, parameter	1	2	3	4	5	6	7	8	9	10	11	12
2020, β_0	-0,10	0,02	0,00	0,62	0,05	0,00	-0,03	0,27	-0,06	-0,17	-0,05	-0,40
2020, β_i	-0,05	0,01	0,00	0,28	0,02	0,00	-0,01	0,14	-0,03	-0,08	-0,02	-0,18
2021, β_0	-2,14	1,85	0,30	-0,77	-0,14	-0,01	0,00	-0,06	0,48	0,28	0,40	-0,15
2021, β_i	0,35	-0,29	-0,03	0,12	0,03	0,00	0,01	0,02	-0,08	-0,04	-0,07	0,05
2022, β_0	-1,79	1,55	-1,71	0,12	-1,69	1,04	0,65	-0,70	0,70	3,01	0,39	-1,38
2022, β_i	-0,08	0,08	-0,09	0,00	-0,09	0,06	0,03	-0,04	0,04	0,17	0,03	-0,06
2023, β_0	-0,45	0,27	-0,08	-0,06	0,10	-0,10	-0,48	-0,01	-0,64	0,67	0,15	0,66
2023, β_i	0,17	-0,10	0,04	0,02	-0,03	0,05	0,19	0,02	0,28	-0,27	-0,06	-0,24

Exempelvis visar tabellen att om april 2023 (kolumn 4) exkluderas kommer interceptet att minska med 6 procent och lutningskoefficienten att öka med 2 procent. Vi ser inget tydligt mönster i dessa avvikelser i Tabell 5, vilket gäller för samtliga fyra år.

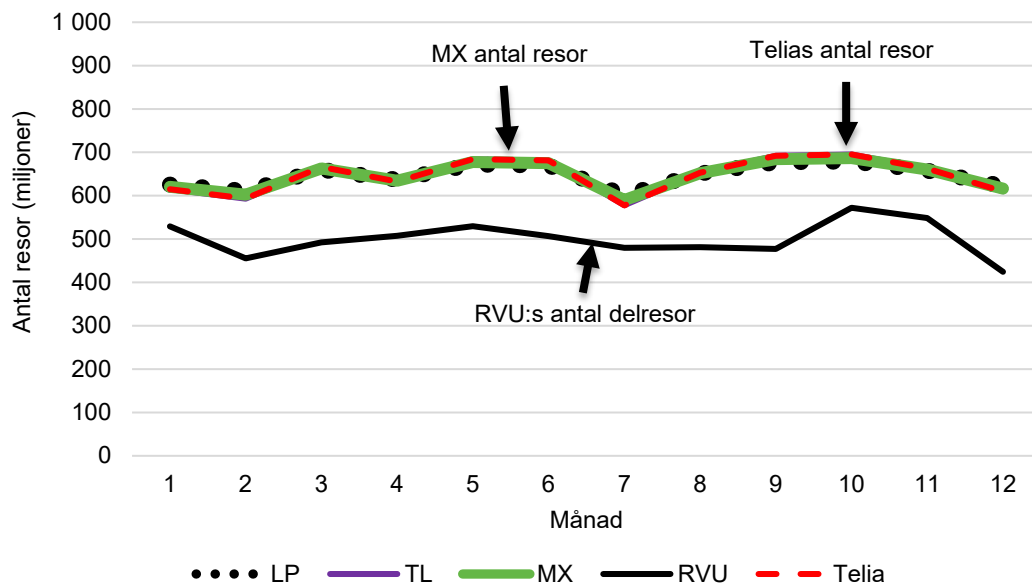
Kort sammanfattning:

- Resultaten av Relativ effektivitet från bootstrapping indikerar att Transfer Learning (TL) är mer effektiv än RVU vad gäller varians.
- Vid jämförelse av de fem metodernas resultat från bootstrapping (endast RVU som källa, endast Mobilnätsdata som källa, LP, TL, MX) ser vi att MX konsekvent ger det lägsta ARRMSE-värdet. MX-modellen indikerar därmed på bättre precision än vad RVU gör på helår när vi jämför med RVU:s osäkerhetsmått.
- Utifrån visuell analys av Figur 4, Figur 5, Figur 6 och Figur 7 samt från Tabell 5 ser vi att den linjära modellen kan inkluderas i vår modellering.
- Om skattningar i en linjär blandad modell kan kombineras med Transfer learning (LP+TL=MX) utan att linjär prediktion blir osäker innebär det en fördel i form av det ger en mer stabil skattning, jämfört med att endast använda Transfer learning.

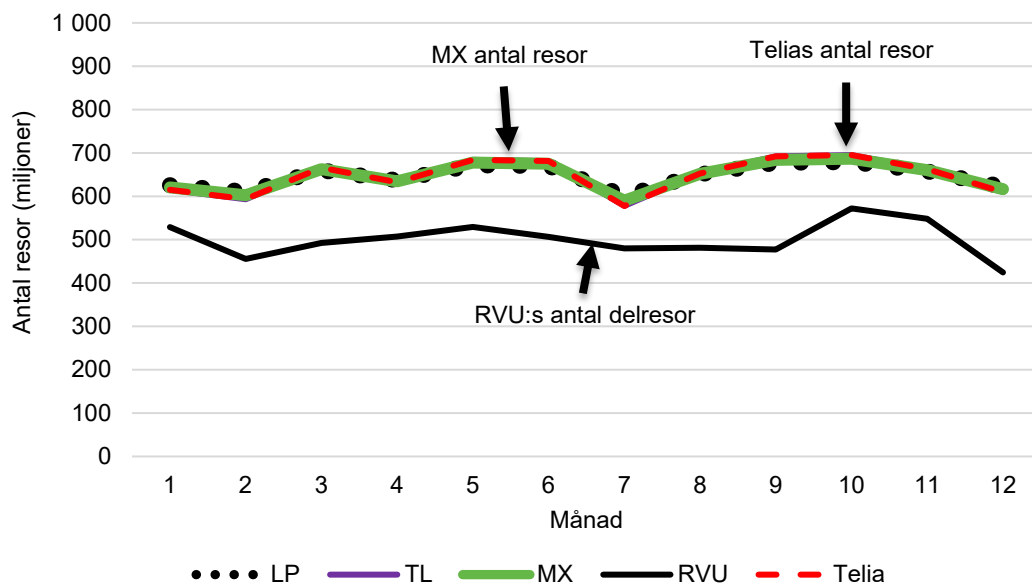
Därmed anser vi att den mest robusta modellen att använda för att skatta antal resor per månad är med estimatören MX, som kombinerar skattningar i den linjära blandade modellen (LP) och Transfer learning (TL).



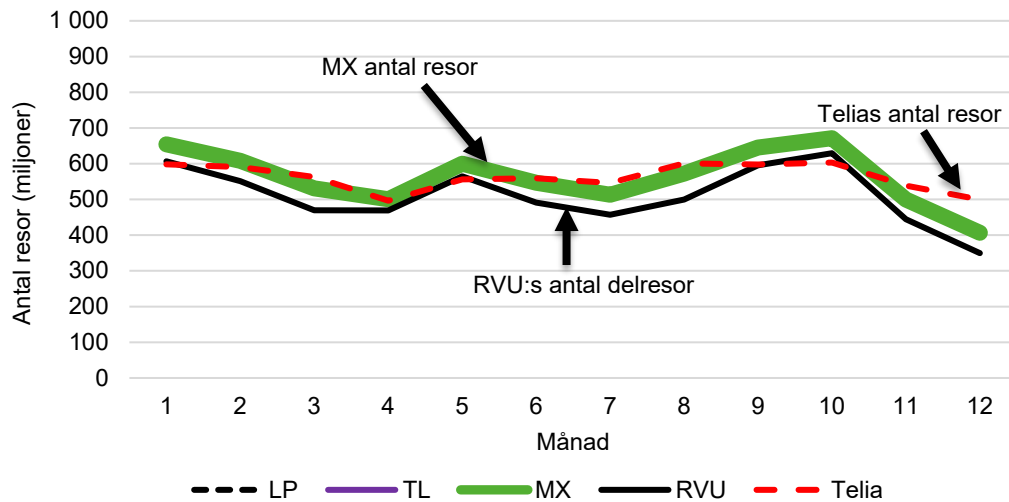
Figur 8 presenteras antal resor per månad för respektive metod och datakälla för år 2023. MX är den metod vi valt för skattning av resor per månad, och observera att resorna har viktats till Mobilnätsdatas totalt antal resor. En diskussion om av valet av källa för totalt antal resor finns i kapitel 6.



Figur 8. Antal resor per månad för de två källorna samt för de tre metoderna, viktat till Telias antal resor totalt. År 2023.



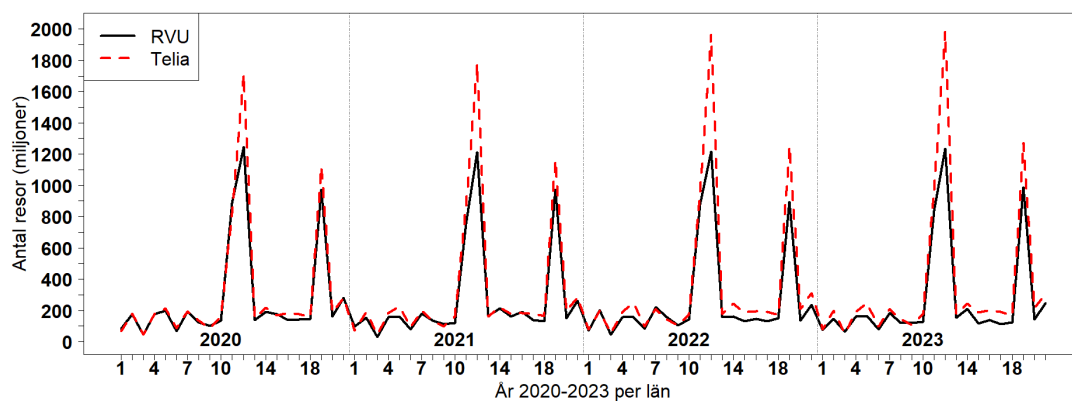
I Figur 8 ser vi att Mobilnätdata följer MX relativt väl under 2023. Däremot visar resultaten för 2020 (se Figur 9) att RVU snarare följer MX relativt väl.



Figur 9. Antal resor per månad för de två källorna samt för de tre metoderna, viktat till Telias antal resor totalt. År 2020.

5.2 Korrelation och testning mellan RVU och Mobilnätsdata, brutet på län

Figur 10 visar ett diagram över de fyra åren 2020–2023, där Mobilnätsdatas antal resor visas av en röd streckad linje och RVU:s antal resor av en svart linje. Diagrammet visar antal resor inom län, där länen är sorterade i bokstavsordning och åren i stigande ordning från vänster.



Figur 10. Antal resor per inom län och år för de två källorna, åren 2020–2023. Förklaring till x-axelns länsnummer finns i Tabell 11 i Tabellbilaga.

Vi testar hur väl de två datakällorna överensstämmer med varandra per år för att kunna bestämma den mest lämpliga implementeringsmetoden¹⁸. Samtliga tester i detta avsnitt har utförts i R.

Vi sätter Y som RVU-data och X som Mobilnätsdata. RVU:s länsestimatorer för antal delresor har approximerade normalfördelningar och estimerade varianser. RVU:s länsestimatorer behandlas som icke-biased. Mobilnätsdatas antal resor behandlas som konstanter, med

¹⁸ Zhang, "Disaggregation of trips using MNO data".

negligerbara varianser men med bias. För hypotesprövningarna har vi använt oss av χ^2 -test (Chi²-test). Då kan nollhypotesen med tillhörande hypotesprövningar uttryckas enligt följande:

- a) $H_0: X_t = Y_t$, som med ord blir H_0 : Mobilnätsdatas antal resor är lika med RVU:s antal delresor över alla 21 län ett givet år, där endast resor inom län mäts och där vi antar att RVU:s antal delresor motsvarar det verkliga antalet.
- b) $H_0: X_t \propto Y_t \Leftrightarrow \frac{X_t}{X} = \frac{Y_t}{Y}$, som med ord blir H_0 : Mobilnätsdatas antal resor är proportionell mot RVU:s antal delresor över alla 21 län ett givet år, där endast resor inom län mäts och där vi antar att RVU:s fördelning över resor speglar den verkliga fördelningen.

Alternativa hypotesen, H_1 , är att det finns en signifikant skillnad mellan datakällorna avseende antal resor respektive fördelningen av resor per län.

Sammanfattningsvis testar vi om Mobilnätsdata skiljer sig från RVU när det gäller resor inom län och år, eller om skillnaderna faller inom ramen för urvalsosäkerheten i RVU. Liksom i avsnitt 0 vill vi att p-värdena ska vara höga eftersom det innebär att vi inte kan förkasta nollhypoteserna. Med andra ord vill vi inte att källorna ska skilja sig från varandra avseende antal resor och fördelningen av resor per inom län. Resultaten är dock inte tillfredsställande eftersom samtliga nollhypoteser förkastas (se p-värden i Tabell 6).

Tabell 6. p-värden för hypotestest a) Mobilnätsdatas antal resor = RVU:s antal delresor över alla 21 län ett givet år, samt för hypotestest b) Mobilnätsdatas antal resor är proportionell mot RVU:s antal delresor över alla 21 län ett givet år.

År	Nollhypotes a)	Nollhypotes b)
2020	0,00	0,00
2021	0,00	$5,61 \times 10^{-5}$
2022	0,00	$2,00 \times 10^{-7}$
2023	0,00	0,00022

Vi testar även Transfer learning (TL) för att estimerar antal resor. Måttet Relativ effektivitet(TL) = $MSE(TL) / Var(RVU)$ används för detta. Ett lägre värde innebär, som bekant, en större effektivitet jämfört med den direkta RVU-estimatoren, från bootstrapping (100 000) under RVU:s design.

Tabell 7 visar Relativ effektivitet med en TL-estimator jämfört med RVU-skattningarna, från bootstrapping (100 000) under RVU:s design.

Tabell 7. Relativ effektivitet med en Transfer learning-estimator jämfört med RVU-skattningarna, brutet på län.

År	Relativ effektivitet
2020	0,029
2021	0,17
2022	0,089
2023	0,20

Denna metod (TL) ger mer tillfredsställande resultat för samtliga fyra år jämfört med en skattning baserad enbart på RVU som källa, inklusive tillhörande osäkerhetsintervall. Exempelvis innebär ett värde på 0,029 för år 2020 att MSE endast utgör cirka 3 procent av variansen i RVU-skattningarna. Det innebär att TL ger bättre precision vid skattning av antal

resor per län jämfört med vad RVU enskilt ger. Därmed ger TL-metoden mer tillförlitliga resultat vid nedbrytning på län och år än en modell som enbart bygger på RVU.

I likhet med testerna utförda i avsnitt 0 vill vi ytterligare stärka modellens robusthet genom att kombinera skattningar i en linjär blandad modell med en Transfer learning-metod. Vi testar därmed Linjär prediktion (LP), Transfer learning (TL) och slutligen en robust estimator, som vi kallar för modell MX, som kombinerar LP och TL.

Att använda mobilnätsdata eller RVU direkt ger sämst resultat vid nedbrytning på län, där mobilnätsdata uppvisar något högre (och därmed sämre) värden än RVU på Average Root Mean Square Error (ARRMSE). Som vi redogjort för i avsnitt 0 används ARRMSE för att bedöma en estimators prestanda. Skattningar av ARRMSE har tagits fram genom bootstrapping (100 000) under RVU:s design, vid samtliga fem metoder.

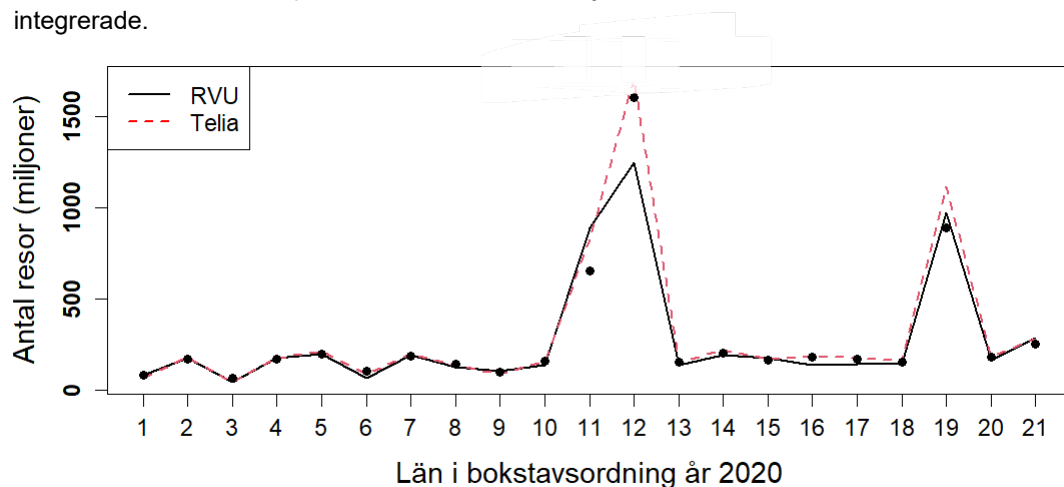
Tabell 8 nedan redovisas ARRMSE för följande metoder: RVU (som enskild källa), Mobilnätsdata (som enskild källa), LP (skattningar i linjär blandad modell), TL (Transfer learning), MX (Robust estimator som kombinerar LP och TL). Skattningar av ARRMSE har tagits fram genom bootstrapping (100 000) under RVU:s design, vid samtliga fem metoder.

Tabell 8. Average Relative Root Mean Square Error (ARRMSE) för resor brutet på län, per källa/metod och år.

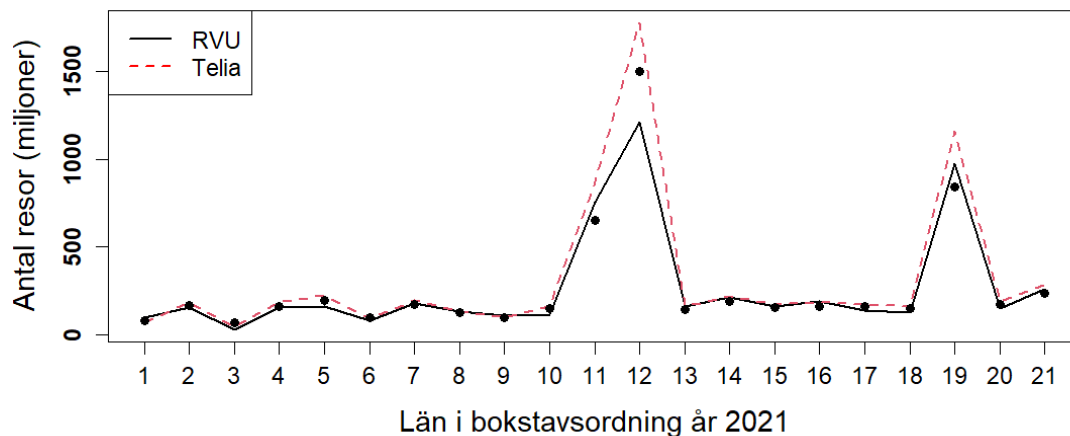
År	RVU	Mobilnätsdata	LP	TL	MX
2020	0,075	0,11	0,016	0,0071	0,0096
2021	0,11	0,11	0,059	0,034	0,051
2022	0,10	0,13	0,032	0,024	0,027
2023	0,11	0,11	0,044	0,040	0,042

Resultaten visar att Transfer learning (TL) och den robusta metoden MX (som kombinerar LP och TL) uppvisar de lägsta (bästa) ARRMSE-värdena vid estimering av antal resor inom län. Detta innebär att TL och MX är mer fördelaktiga än att enbart använda RVU eller Mobilnätsdata.

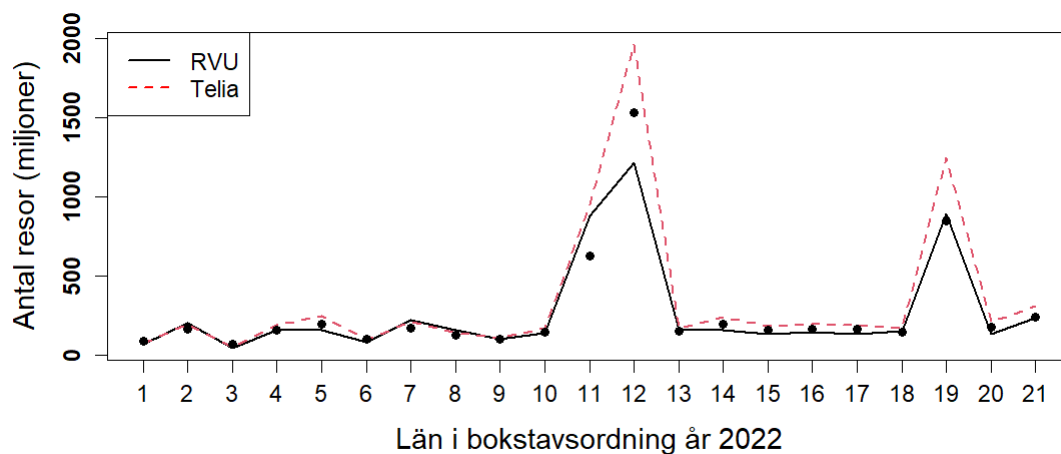
I Figur 11 – Figur 14 nedan visas hur en linjär modell presterar per *inom* län i diagram. Den röda streckade linjen representerar Mobilnätsdatas antal resor, den svarta linjen RVU:s antal resor, och de svarta datapunkterna motsvarar en linjär modell där båda källorna är integrerade.



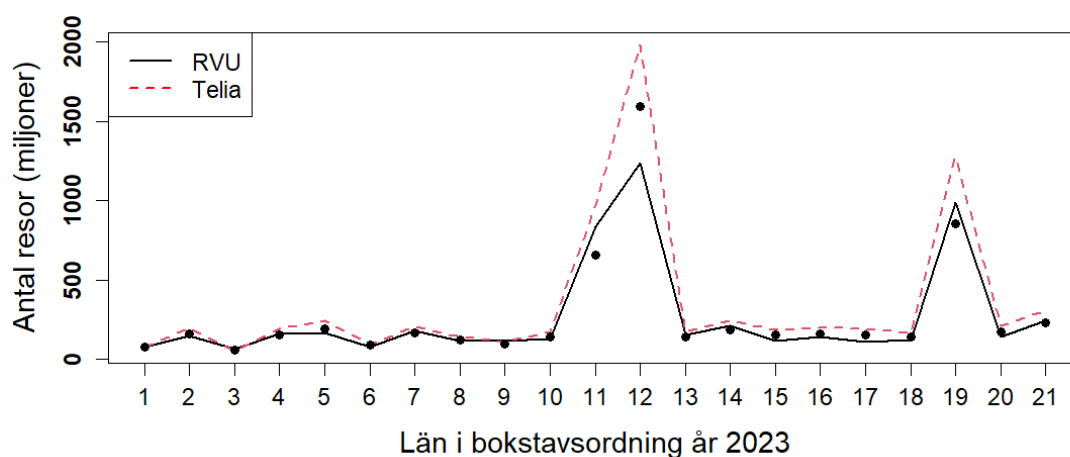
Figur 11. Antal resor per län år 2020. Förklaring till x-axelns länsnummer finns i Tabell 11 i Tabellbilaga. De svarta datapunkterna i diagrammet markerar skattningarna för en linjär modell.



Figur 12. Antal resor per län år 2021. Förklaring till x-axelns länsnummer finns i Tabell 11 i Tabellbilaga. De svarta datapunkterna i diagrammet markerar skattningarna för en linjär modell.



Figur 13. Antal resor per län år 2022. Förklaring till x-axelns länsnummer finns i Tabell 11 i Tabellbilaga. De svarta datapunkterna i diagrammet markerar skattningarna för en linjär modell.



Figur 14. Antal resor per län år 2023. Förklaring till x-axelns länsnummer finns i Tabell 11 i Tabellbilaga. De svarta datapunkterna i diagrammet markerar skattningarna för en linjär modell.

Vid analys av de tre län med flest resor: Skåne län (11), Stockholms län (12) och Västra Götalands län (19), ser vi att den linjära modellen är riskabel. Dessa tre län innehåller så kallade influencers, det vill säga outliers som har betydande påverkan på en linjär modell, och det gäller konsekvent över samtliga fyra år. I Tabell 9 redovisas hur interceptet (raderna β_0) och lutningskoefficienten (raderna β_i) i den linjära modellen förändras när en enskild mätpunkt (län) exkluderas, per år.

Tabell 9. Värden på raderna β_0 visar hur mycket interceptet förändras och värden på β_i visar hur mycket lutningen förändras om länet utesluts, i en linjär modell för resor brutet på inom län. Åren 2020–2023. Kolumnerna visar länen i bokstavsordning, där 1 = Blekinge län och 21 = Östergötlands län.

År, Parameter	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
2020, β_0	0,00	0,00	0,05	0,00	0,00	0,08	-0,01	0,03	0,00	0,05	0,00	-0,94	0,03	0,02	-0,01	0,08	0,05	0,02	0,07	0,04	-0,04
2020, β_i	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	-0,04	0,19	0,00	0,00	0,00	0,00	0,00	0,00	-0,02	0,00	0,00
2021, β_0	-0,02	0,03	0,08	0,01	0,06	0,04	0,00	0,00	-0,01	0,06	0,00	-0,71	-0,02	-0,03	0,00	-0,04	0,05	0,04	0,11	0,04	-0,03
2021, β_i	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	-0,01	0,16	0,00	0,00	0,00	0,00	0,00	0,00	-0,03	0,00	0,00
2022, β_0	0,03	-0,06	0,05	0,00	0,07	0,04	-0,07	-0,04	0,00	0,01	0,00	-0,73	0,00	0,06	0,05	0,03	0,05	0,00	0,03	0,07	0,00
2022, β_i	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	-0,04	0,18	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
2023, β_0	0,01	0,03	0,00	0,00	0,06	0,04	-0,03	0,00	-0,05	0,04	0,01	-1,14	-0,01	-0,04	0,09	0,05	0,11	0,04	0,13	0,07	-0,02
2023, β_i	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	-0,03	0,19	0,00	0,00	0,00	0,00	0,00	0,00	-0,03	0,00	0,00

För att en linjär modell ska fungera tillfredsställande bör skillnaderna mellan länen (kolumnerna) inte vara alltför stora. I detta fall ser vi dock att kolumn 12 (Stockholms län) år 2023 har ett interceptvärde på $-1,14$ (dvs. interceptet minskar med 114 procent om denna mätpunkt exkluderas) samt en lutningskoefficient på $0,19$ (dvs. lutningen ökar med 19 procent vid exkludering av denna mätpunkt). Dessutom är det endast tre mätpunkter där lutningen skiljer sig från 0. Motsvarande mönster som år 2023 ser vi även övriga tre år. Denna tydliga skillnad mellan länen innebär en risk eftersom det kan leda till stora fel i modellen om en linjär modell används i vår modellering. Transfer learning-metoden tillåter visserligen bias, men är samtidigt mindre känslig än en linjär modell vid en dåligt anpassad modell.

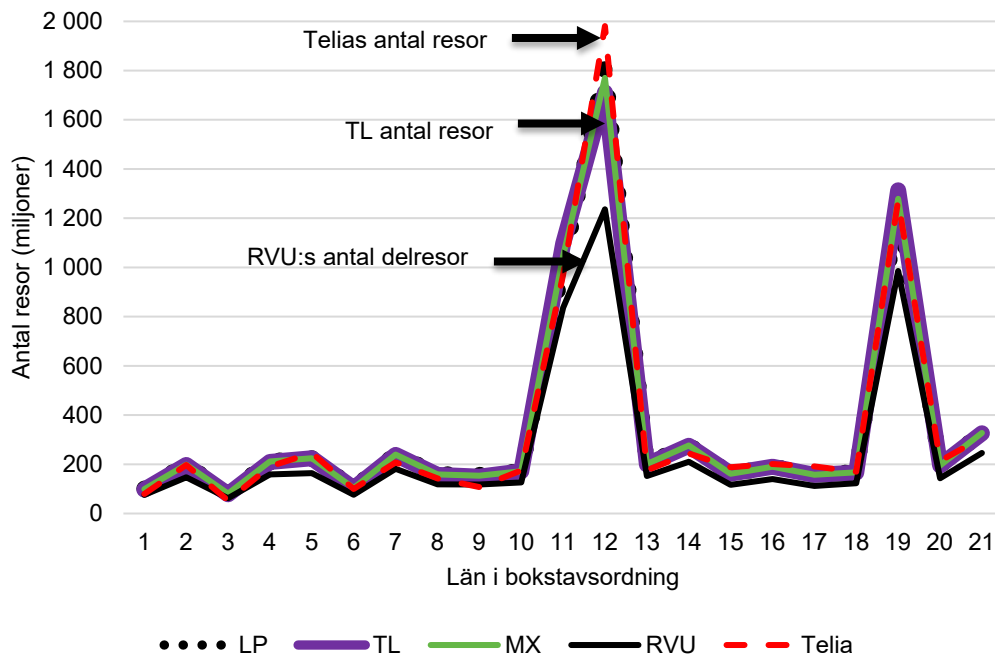
Transfer learning är mer motståndskraftig mot influencers. Detta beror på att Transfer learning-metoden inte kräver antagandet om en linjär modell.

Kort sammanfattning:

- Resultaten av Relativ effektivitet från bootstrapping visar att TL är mer effektiv avseende varians jämfört med RVU.
- Vi har sett att TL och MX ger lägst ARRME över lag vid jämförelse av samtliga fem metoder (endast RVU som källa, endast Mobilnätdata som källa, LP, TL, MX) från bootstrapping.
- MX-metoden är mer känslig, och därav mer riskabel, än TL eftersom den påverkas relativt mycket vid exkludering av mätvärden för de tre län med flest antal resor.
- Utifrån visuell analys av Figur 11, Figur 12, Figur 13 och Figur 14 samt analys av Tabell 9 ser vi att vi inte kan ta med oss den linjära modellen i vår modellering.
- I början av avsnittet såg vi att de två datakällorna skiljer sig åt avseende antal resor samt fördelningen av resor per inom län. Det betyder att linjär modellering kan vara olämplig.

Med hänsyn till dessa resultat bedömer vi att Transfer learning (TL) är den mest lämpliga implementeringsmetoden för skattning av antal resor brutet på län.

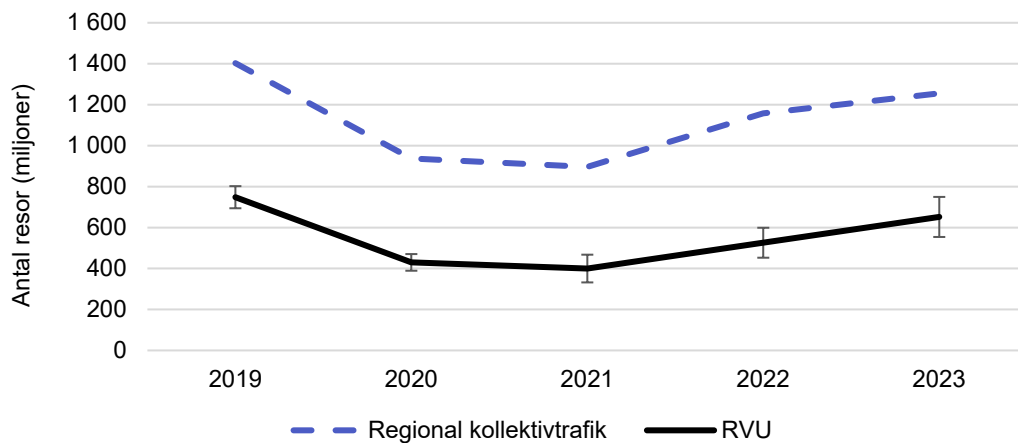
I Figur 15 presenteras antal resor per län för respektive metod och källa för år 2023. TL har valts som metod för resor per inom län. Observera att resorna i metoderna viktats till Mobilnätdataas totalt antal resor, i likhet med vad som gjordes i avsnitt 0. En diskussion om av valet av källa för totalt antal resor finns i kapitel 6.



Figur 15. Antal resor per län för de två källorna samt för de tre metoderna, viktat till Telias antal resor totalt. År 2023.

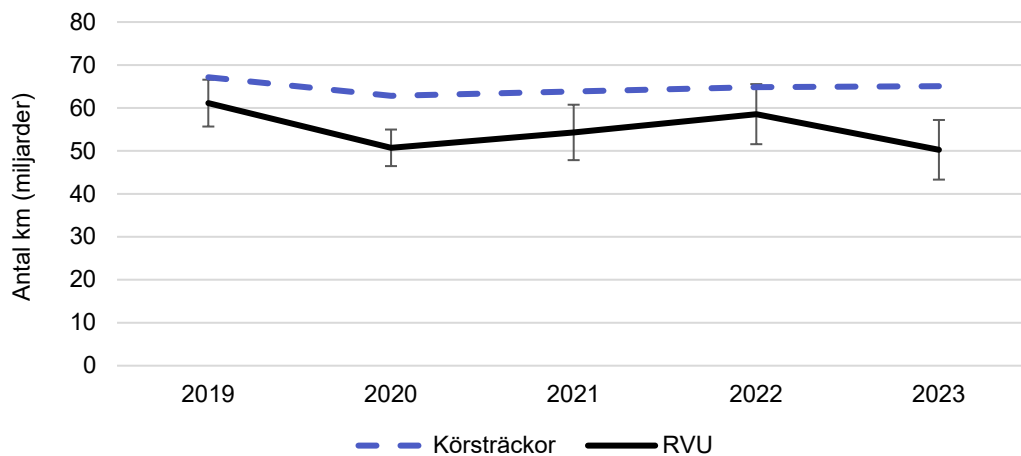
5.3 Korrelation mellan RVU och övrig statistik

I avsnitt 3.1 och Figur 1 såg vi att korrelationen mellan RVU och Mobilnätsdata var väldigt låg sett till den årliga utvecklingen av antal resor. Det är dock svårt att säga vilken av kurvorna som visar den mest rimliga utvecklingen. Med Mobilnätsdata går det inte att dela upp resandet på färdssätt. För RVU är det möjligt att dela upp resorna på färdssätt och det finns andra källor att jämföra med. För färdssätten buss, tunnelbana och spårvagn finns officiell statistik över påstigningar. Eftersom en delresa i RVU kan innehålla flera påstigningar är det förväntat att nivåerna ska skilja sig åt. Jämför vi utvecklingen för påstigningar från och med 2019 med RVU:s uppgifter om delresor ser vi en förväntad skillnad i antal resor, men kurvorna korrelerar bra med varandra (korrelationen är 0,99), se Figur 16.



Figur 16. Antal delresor i RVU och påstigningar enligt Regional kollektivtrafik i miljoner, åren 2019–2023.

Även körsträckorna enligt den officiella statistiken, som bygger på uppgifter från besiktningar av fordonsflottan, följer utvecklingen av körsträckor med personbil enligt RVU (korrelationen är 0,73), se Figur 17.



Figur 17. Körd sträcka med personbil som förare enligt RVU och körsträckor enligt den officiella statistiken, åren 2019–2023. Miljarder km.

Orsaken till att resandet enligt RVU inte har ökat så mycket efter pandemin beror framför allt på minskat resande med cykel och till fots. För dessa färdssätt har vi dock ingen annan källa att jämföra med för att se om utvecklingen verkar rimlig. Däremot ser vi en rimlig utveckling för de andra stora färdssätten, personbil och kollektivtrafik.

Under antagandet att RVU visar korrekta andelar för olika färdssätt går det att skapa index över hur resandet utvecklats sig enligt annan officiell statistik (hädanefter kallat index för officiell statistik). För att skapa index för officiell statistik antas antal personkilometer med bantrafik, påstigningar med regional kollektivtrafik, körsträckor i personbil och inrikes passagerare med flyg under 2019 motsvara andelen delresor med tåg, regional kollektivtrafik, personbil och flyg enligt RVU för samma år (se *Andel exklusive annat* i Tabell 10). Körsträckor för personbil får således en stor inverkan på indexet för officiell statistik (79,4 procent) medan inrikes flyg får en liten inverkan (0,3 procent).

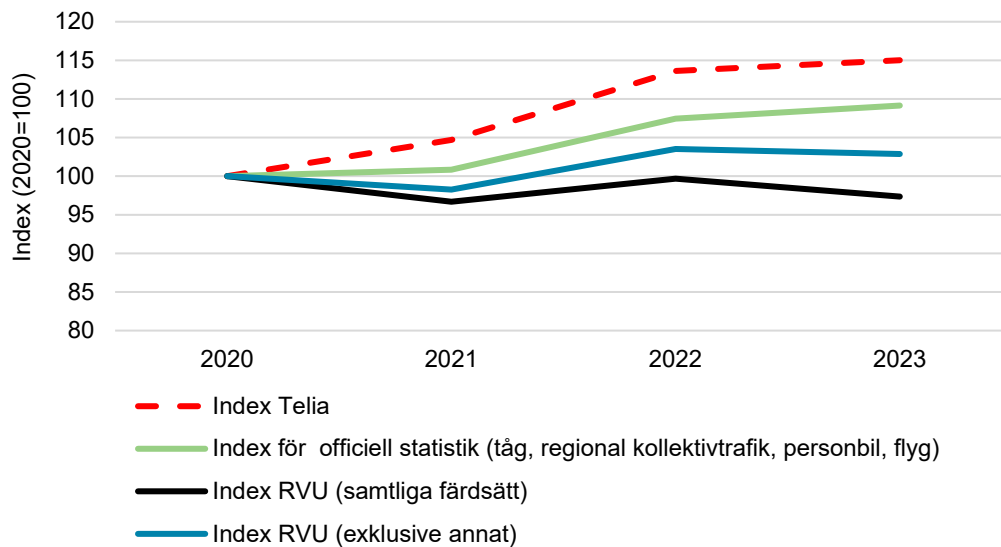
Eftersom vi saknar officiell statistik för gång och cykel kommer indexet för officiell statistik inte ta hänsyn till förändringar i resandet med dessa färdssätt, trots att det står för en relativt stor del av resandet (30,3 procent). Det innebär att det finns en relativt stor osäkerhet i indexet för officiell statistik.

Tabell 10. Färdssättens andelar enligt RVU 2019.

Färdssätt	Andel	Andel exklusive annat
Tåg	3,4%	4,8%
Regional kollektivtrafik	10,8%	15,4%
Personbil	55,3%	79,4%
Flyg	0,2%	0,3%
Annat	30,3%	

När indexet för officiell statistik jämförs med utvecklingen av antal resor enligt Mobilnätsdata är korrelationen relativt hög (0,98). Vi har även hög korrelation mellan indexet för officiell statistik och RVU exklusive annat färdssätt än tåg, regional kollektivtrafik, personbil och tåg (0,90).

Däremot korrelerar RVU dåligt med övriga index om samtliga färdssätt ingår, med en korrelation på $-0,21$. Mellan RVU, där samtliga färdssätt ingår, och Mobilnätsdata är korrelationen $-0,09$, se Figur 18.



Figur 18. Utveckling för resandet enligt index (2000=100) perioden 2020–2023.

Att RVU stämmer relativt bra med övrig jämförbar statistik avseende trafikslag tyder på att RVU är en god källa för att följa utvecklingen för resandet.

I Mobilnätdata har vi, som vi tidigare konstaterat, inte tillgång till vare sig färdssätt eller syfte med resan. Utifrån antagandet att RVU fångar flöden för färdssätt på ett med verklighet överensstämmande sätt har vi valt att använda RVU:s andelar för att beskriva resvolym per färdssätt, årligen.

Även syfte med resan publiceras i RVU, och denna indelning håller standarden för officiell statistik. Därför har vi valt att använda RVU:s andelar för att beskriva resvolym per syfte och år. Färdssätt och syfte har därmed fått RVU:s andelar och dessa andelar har vi, liksom vi gjort för brytningar på månad och län, viktat till Mobilnätdatas antal resor totalt per år. Den metoden kallar vi för en rak uppräkningsmetod.

6 Diskussion och slutsatser

I kapitel 5 har vi analyserat de mest lämpade implementeringsmetoderna för nedbrytning av materialet på månad och län. Resultaten från analyserna visade att metoder som kombinerar de två datakällorna gav mer stabila resultat jämfört med att använda datakällorna separat.

För nedbrytning per månad visade analyserna att en robust blandad modell, som kombinerar Transfer learning med skattningar i en linjär blandad modell, gav de bästa resultaten för att skatta antal resor med båda datakällorna. Se avsnitt 0 för analyserna.

För nedbrytning på län visade det sig däremot att Transfer learning-metoden, då den används enskilt och inte tillsammans med en linjär blandad modell, var mest lämpad för att skatta antal resor med båda datakällorna. Se avsnitt 5.2 för analyserna. Datat nedbrutet på län, till skillnad från data nedbrutet på månad, innehåller influencers som skulle påverka en linjär blandad modell relativt mycket. Vi bedömer därför att det är en säkrare metod att använda oss av endast Transfer learning vid nedbrytning på län, säkrare i den mening att inte riskera lika stora skattningsfel i modellen som det kan bli med en linjär modell i det fallet.

Utöver nedbrytning på tid och geografi har vi i projektet även delat in antal resor på färdsätt och syfte, genom att kombinera de två datakällorna. Vi har konstaterat att det är möjligt att redovisa färdsätt och syfte med hjälp av RVU och som vi sedan kan kombinera med Mobilnätsdata. Genom en rak uppräknings kombinerar vi RVU:s andelar med det totala antalet resor från Mobilnätsdata för att beskriva resor per färdsätt och syfte. Se avsnitt 5.3 för analyserna.

Vi har kommit fram till att antal resor per år beskrivs av Mobilnätsdata i den kombinerade datamodellen. Ett argument för att använda Mobilnätsdata som källa att vikta upp resorna till är den bredare målpopulationen i Mobilnätsdata, som omfattar befolkningen från 6 år och äldre jämfört med 6–84 år i RVU. Vi vet dock om att det finns frågetecken kring mätningen i Mobilnätsdata och vi finner den källan som osäkrare än vår produkt RVU över lag. Det är även en "black box" för oss hur uppräknings till totalt antal resor i Mobilnätsdata skett. Vi betraktar vår redovisning som ett resultat av detta projekt som experimentell statistik och inte officiell statistik. Vi har därför valt att vikta till totalt antal resor baserat på Mobilnätsdata och inte till totalt antal delresor baserat på RVU-data. Vi vill för statistikanvändare också var tydliga med skillnaderna mellan publiceringen från detta projekt och den publicering som utgörs av officiell statistik i den nationella resvaneundersökningen, RVU för användaren. Det är också en orsak till att vi inte använder RVU-data för att vikta upp antal resor totalt.

Om vi i framtiden får tillgång till ytterligare mobilnätsdatakällor från fler operatörer kan det vara än mer fördelaktigt att använda mobilnätsdata som källa för att vikta till totalt antal resor. Ett annat alternativ för att vikta till totalt antal resor kan vara att kombinera RVU och mobilnätsdata.

Det finns nackdelar med traditionella urvalsundersökningar, exempelvis sjunkande svarsfrekvenser som bland annat bidrar till större osäkerhetsmått. Detta blir särskilt utmanande vid nedbrytningar av statistiken på tidsperioder och geografi. En fördel med Mobilnätsdata, förutsatt att den är korrekt, är att det stora dataunderlaget ger möjlighet till nedbrytning av datamaterialet, och där överträffar Mobilnätsdata RVU. Vi är medvetna om att

det finns frågetecken kring Mobilnätsdata, som vi tidigare nämnt, vilket ligger till grund för nuvarande beslut att inte publicera statistik *mellan* geografiska platser.

Utifrån slutsatserna innebär det att vi redovisar följande nya tabeller:

- a) Antal resor per månad
- b) Antal resor per inom län
- c) Antal resor per färdstätt och syfte

Följande metoder har använts för skattningarna i respektive tabell:

- a) en robust blandad modell (en MX-estimator) som kombinerar skattningar i linjär blandad modell (LP) med Transfer learning (TL), vid nedbrytning på månad.
- b) en metod med Transfer learning (TL) vid nedbrytning på län.
- c) en rak uppräknings där vi använder RVU:s andelar vid nedbrytning på färdstätt och syfte.

7 Publicering

Resultatet från arbetet med att kombinera RVU med mobilnätdata redovisas via den visualiseringsplattform som Trafikanalys använder för att publicera statistik i form av tabeller och figurer, och finns på följande webbsida: www.trafa.se/transportmonster/RVU-Sverige/kombinerade-mobilnatsdata-och-enkatdata-beskriver-resmonster-15129/

Vi redovisar resultatet i form av tabeller och figurer för år 2020–2023 med följande uppdelningar:

- Antal resor per månad.
- Antal resor per län.
- Antal resor per färdstätt och syfte.

Figur 19 visar ett skärmdokument från publiceringssidan.

Ungefär 95 procent av alla resor har sin start- och slutpunkt inom ett och samma län. Flest resor sker inom Stockholms län och minst antal resor på Gotland. Antalet resor har ökat för varje år under perioden 2020 till 2023. År 2023 hade således flest antal resor och året då pandemin bröt ut, 2020, hade minst antal resor.

Tabell 1. Antal tusen resor inom län per län. År 2020, 2021, 2022 och 2023.



Län	2020	2021	2022	2023
Stockholms län	1 417 863	1 522 583	1 639 991	1 711 320
Uppsala län	216 140	250 596	215 250	275 240
Södermanlands län	154 059	193 263	202 928	199 119
Östergötlands län	315 423	312 309	311 685	326 308
Jönköpings län	216 037	216 167	281 781	237 960
Kronobergs län	112 231	129 211	133 287	149 612
Kalmar län	139 404	158 820	199 384	156 595
Gotlands län	50 253	39 190	58 400	78 070
Blekinge län	90 915	110 871	90 524	96 935
Skåne län	989 284	910 291	1 132 181	1 096 774
Hallands län	222 323	201 296	209 623	224 244
Västra Götalands län	1 097 589	1 176 475	1 184 126	1 313 666
Värmlands län	195 468	192 964	175 013	162 074
Örebro län	181 608	182 771	180 011	195 446
Västmanlands län	163 918	159 555	195 688	165 752
Dalarnas län	197 066	187 960	258 116	197 060
Gävleborgs län	195 787	190 129	205 870	209 445
Västernorrlands län	161 677	167 974	175 456	156 997
Jämtlands län	74 869	96 156	105 718	102 519
Västerbottens län	155 812	223 681	193 667	189 919
Norrbottens län	151 286	146 302	185 049	171 054
Riket	6 499 013	6 768 564	7 333 747	7 416 109

Figur 19. Skärmdokument år 2025 från den visualiseringsplattform som Trafikanalys använder för att publicera statistiken, där publiceringen uteslutande har skett. Tabellen i bilden visar antal resor per län och år med en Transfer learning-metod (TL), åren 2020–2023.

8 Arbetet framåt

I framtiden vill vi fortsätta vidareutveckla statistiken. Bland annat kan det handla om att inhämta data från fler mobilnätoperatörer än den enda operatören, Telia, som har använts i detta projekt. Genom att inkludera fler operatörer skulle vi kunna korsvalidera mobilnätsdata mellan olika operatörers data och RVU. Detta skulle inte bara ge en mer omfattande datamängd, utan även öka robustheten i analyserna och ge bättre förutsättningar för att skapa mer detaljerad statistik av högre kvalitet.

Ett utvecklingsområde är att mer sofistikerat nyttja mobilnätsdatas möjligheter att bryta ner materialet i mer detaljerade uppdelningar samt att undersöka relationer mellan olika platser. För att säkerställa att denna nedbrytning är genomförbar krävs ytterligare analyser av kvaliteten på mobilnätsdata.

Ett annat område för utveckling är att bättre integrera variabler från RVU som inte finns i mobilnätsdata, såsom färd sätt och syfte. Detta skulle dels omfatta att förbättra och förfinas den nuvarande metoden som vi har använt oss av för att dela in resor i färd sätt och ärende, dels att lägga till fler dimensioner, exempelvis uppdelning av resor efter kön och ålder.

Det kan komma att bli aktuellt att ändra källan för det totala antalet resor per år, antingen genom att använda data från fler mobilnätoperatörer eller genom att kombinera mobilnätsdata med RVU.

För att vidareutveckla och förbättra statistiken krävs ytterligare analyser samt utvärdering av de olika datakällorna.

9 Källförteckning

Trafikanalys. "Resvanor, Sverige". Trafikanalys, u.å. www.trafa.se/transportmonster/RVU-Sverige/

Trafikanalys. "PM 2023:6 Hur väl kan mobilnätsdata beskriva våra resvanor", u.å. www.trafa.se/etiketter/transportovergripande/hur-val-kan-mobilnatsdata-beskriva-vara-resvanor-14193/

Trafikanalys. "Regional linjetrafik och Körsträckor". Trafikanalys, u.å. www.trafa.se/kollektivtrafik/kollektivtrafik/ och www.trafa.se/vagtrafik/korstrackor/

Zhang, L.-C. "(2024) Disaggregation of trips using MNO data". Opublicerat, MNO-MINDS.

SCB. "Sveriges officiella statistik". www.scb.se/om-scb/samordning-av-sveriges-officiella-statistik/

Sinno Jialin Pan and Qiang Yang Fellow, IEEE. "(2009) A Survey on Transfer Learning". www.cse.ust.hk/~qyang/Docs/2009/tkde_transfer_learning.pdf

James and Stein, Berkeley Symp. on Math. Statist. and Prob., 1961: 361-379 (1961). "Estimation with Quadratic Loss". <https://projecteuclid.org/ebooks/berkeley-symposium-on-mathematical-statistics-and-probability/Proceedings-of-the-Fourth-Berkeley-Symposium-on-Mathematical-Statistics-and-probability/Proceedings-of-the-Fourth-Berkeley-Symposium-on-Mathematical-Statistics-and-probability/Estimation-with-Quadratic-Loss/bsmsp/1200512173>

Fay, R.E. and Herriot, R.A. (1979). Estimates of income for small places: An application of James-Stein procedures to Census data. Journal of the American Statistical Association, [Vol. 74, No. 366 \(Jun., 1979\)](#), pp. 269-27.

Tabellbilaga

Tabell 11. Länen ordnade i bokstavsordning.

Nummer	Län
1	Blekinge län
2	Dalarnas län
3	Gotlands län
4	Gävleborgs län
5	Hallands län
6	Jämtlands län
7	Jönköpings län
8	Kalmar län
9	Kronobergs län
10	Norrbottnens län
11	Skåne län
12	Stockholms län
13	Södermanlands län
14	Uppsala län
15	Värmlands län
16	Västerbottens län
17	Västernorrlands län
18	Västmanlands län
19	Västra Götalands län
20	Örebro län
21	Östergötlands län

Trafikanalys är en kunskapsmyndighet för transportpolitiken. Vi analyserar och utvärderar föreslagna och genomförda åtgärder inom transportpolitiken. Vi ansvarar även för officiell statistik inom områdena transporter och kommunikationer. Trafikanalys bildades 2010 och har huvudkontor i Stockholm samt kontor i Östersund.