

# Ansats för Järnvägsresor med transaktionsdata

*För information*

---

I denna PM summeras analysarbetet med transaktionsdata för järnvägsresor. Med utgångspunkt från tidigare nämnddiskussioner presenteras den nu använda metoden, samt olika tänkbara alternativ. Nämnden inbjuds att komma med synpunkter.

---

## Bakgrund

Järnvägsresor har 4,28 promille i vägningstal KPI. Från 2019 används transaktionsdata från en reseaktör i beräkning av index för järnvägsresor. Den nya designen föredrogs för nämnden under höstmötet 2018 (se Bubuioc m.fl., 2018) och även på vårmötet 2019 (se Bubuioc och Tongur, 2019). Efter diskussioner i nämnden angående de utformade produkterbjudandenas homogenitet, återkommer enheten med en PM för att belysa utformningen av undersökningen.

## Syfte

I denna PM redovisas nuvarande beräkningsmetod, samt några alternativ för utformningen av produkterbjudandena för järnvägsresor och hur index kan påverkas av de olika alternativen.

# Nuvarande metod

De transaktionsdata som SCB använder från och med 2019 innehåller daglig omsättning av alla sålda resor till privatpersoner, per destination och timme, fördelat mellan resenärstyperna vuxen och ungdom. Underlaget har tagits fram/designats i samråd med uppgiftslämnaren som beaktat SCB:s behov.

Produkterbjudandena för vilka priser beräknas från transaktionsdata baseras på följande kombinationer:

*Sträcka* - resa från en avresedestination till en slutdestination

*Avresetid* - i kategorierna morgon, dag, kväll, natt

*Dag i veckan* - mån. – sön.

*Resenärstyp* - vuxen el. ungdom

Indelningen i kategorier av avresetid har gjorts, efter analys och diskussioner med sakkunniga i branschen, enligt resonemanget att resenärerna har ett ärende/en aktivitet på destinationsorten och anpassar sitt val av avresetid till detta med viss flexibilitet. En förutsättning för att fånga prisutvecklingen över tid är att det också finns faktiska avgångar att följa, vilket betyder att tidsintervallen inte kan göras alltför specifika/snäva.

För *avresetid* har följande tidsintervall definierats för de fyra kategorierna:

Morgon - från och med kl. 05:00 till klockan 09:00

Dag - från och med klockan 09:00 till klockan 15:00

Kväll - från och med klockan 15:00 till klockan 20:00

Natt - från och med klockan 20:00 fram till 05:00

Materialet är mycket omfattande och därför tillämpas ett cut-off urval som täcker c:a 60% av total omsättning. Variabeln pris, som används i indexberäkningen, beräknas från all omsättning/antal resor under månaden för varje kombination av specifikationerna ovan för urvalet av resor.

Eftersom varje veckodag förekommer 4-5 gånger under månaden sammanvägs dessa aritmetiskt med respektive dags omsättning, till exempel summan av (första tisdagen + andra tisdagen + ... + sista tisdagen i månaden), delat med summan av antalet resor för samma dagar, och snittpriset baseras på hela månadens omsättning för kombinationen (=produkterbjudandet).

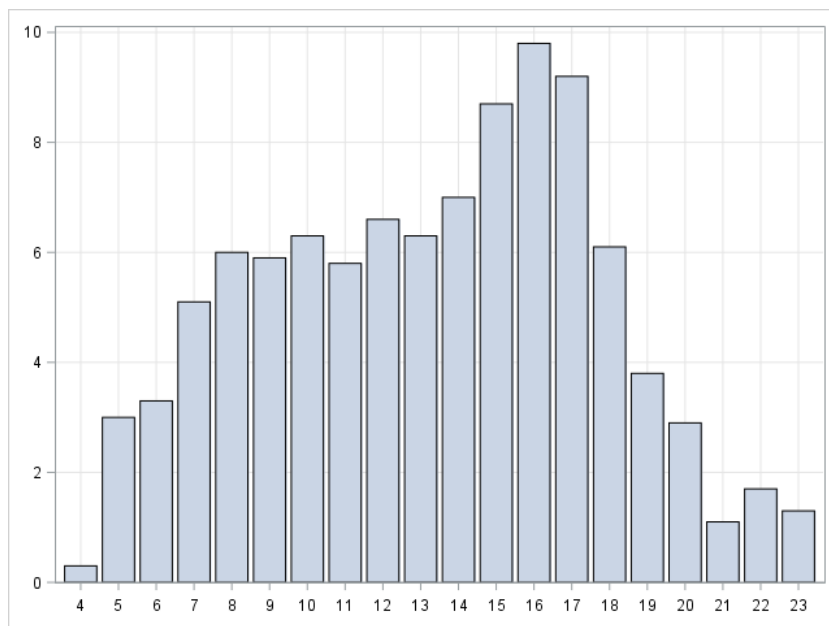
I ord kan ett produkterbjudande beskrivas som ”genomsnittspriset under månaden för att åka från Stockholm till Uppsala en tisdag förmiddag för en vuxen resenär”, dvs. ett enhetsvärde för specifikationen.

Index för järnvägsresor (elementäraggregatet) beräknas enligt standardförfarande som priser aktuell månad jämfört med basmånaden (Jevons) över samtliga specificerade produkterbjudanden för ingående tågoperatörer. Både transaktionsdata och centralt insamlade priser ingår i indexberäkningen.

# Analys

I graf 1 nedan visas på procentuell fördelning (omsättning) av resor efter avgångstid under dygnet, avgränsat till de 50% mest omsatta resorna i urvalet.

**Graf 1** Omsättningens fördelning i procent över avgångstider under dygnet

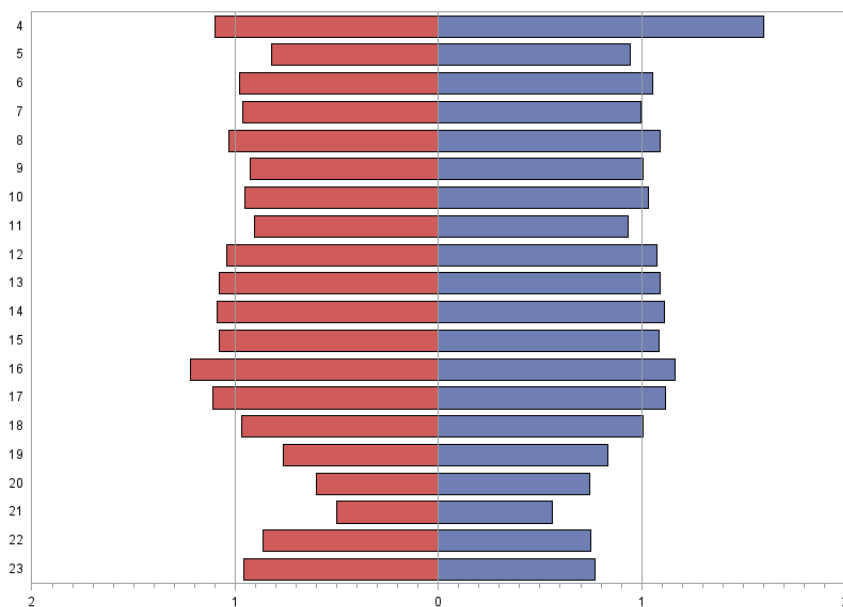


**Not:** vertikal axel är procent, horisontell axel är timmar.

Graf 1 kan jämföras med de definierade kategorierna för avresetid. Till synes finns toppar på eftermiddagen och mindre intensitet på sena/tidiga tider under dygnet.

I graf 2 visas relativa medelvärden och relativa standardavvikelse per timme i relation till dygnsmedelvärde respektive dygnsstandardavvikelse, avgränsat till de 50% mest omsatta resorna i urvalet. Vänster del (röd) är relativt medelvärde, höger del (blå) är relativ standardavvikelse.

**Graf 2** Relativ avvikelse per avgångstimme mot dygnsmedelvärde.  
Medelpris och standardavvikelse för medelpris, per timme relativt dygnsmedel.  
Avser de 50% mest omsatta resorna i urvalet.



**Notering:** Vertikal axel är avgångstimme (4-23), där resor mellan klockan 23 och 4 ingår i resorna för klockan 23. Horisontell axel anger relativt värde. Vänster sida (röd) är relativt medelvärde, höger sida (blå) är relativ standardavvikelse.

Som kan ses i graf 2, är relationerna stabila under merparten av dagen med tydliga avvikelser dels på eftermiddagen kl. 16, dels tidig morgon/sena kvällar. En möjlig förklaring till jämnheterna övriga tider kan vara att det finns en anpassning i termer av antal avgångar/tillgängliga platser, liksom flexibilitet hos privatresenärer.

# Möjligheter i vald beräkningsväg

Nämnden ställde frågan på mötet i maj 2019 om huruvida de valda grupperingarna för *avresetid* (morgon, dag, kväll, natt) med respektive tidsintervall resulterar i tillräckligt homogena grupperingar.

I nämnden diskuterades tre alternativ till befintlig specifikation, fallen 2-4 nedan. Ytterligare möjligheter finns, av vilka några beaktas i analysen:

Fall 1: befintlig specifikation

Fall 2: distinkta avgångstimmar för befintligt urval

Fall 3: som fall 2 men avgångstimmar enbart för resor med stor omsättning

Fall 4: definitionen av *natt* utökas till kl. 6 på morgonen istället för kl. 5

Fall 5: avgångstider slås ihop till tvåtimmarsintervall

Fall 6: avgångstider används inte alls. Enbart dag beaktas.

Det sista fallet (fall 6) är det minst "homogena" alternativet eftersom avresetid inte alls används i specifikationen, vilket betyder att privatresenärer kan tänkas välja resa helt efter lägsta pris/parera sitt köp efter pris. Däremot är det intuitivt något motsägelsefullt vad gäller konsumentnyttan att enbart agera pris-sökare i sammanhanget. Fall 6 var den initiala utformningen som diskuterades i nämnden på höstmötet 2018 (Bubuioc m.fl., 2018). Nämnden påpekade att det kunde innebära så kallat mix-problem vilket föranledde nytt dataunderlag och design med tidsindelning.

## Resultat

I tabell 1 visas relativa prisindex beräknat enligt ovan fem nämnda fall (fall 2-6) relativt den aktuella periodens prisindex med befintlig specifikation (fall 1, alltså fyra tidsintervall för avresetid). Dataunderlaget täcker perioden december 2017- november 2018 och användes för utformningen av nuvarande undersökning. För indextal har relativ framställning valts med anledning av att underlaget kommer från en enskild uppgiftslämnare och därmed innebär risk för röjande. Relationsuttrycket i tabell 1 för fallen 2-6 är enligt "Rel # = (Index Fall # / Index Fall 1)".

**Tabell 1** Relativa index för de fem fallen mot fall 1 (befintlig specifikation)

Period	Fall 2	Fall 3	Fall 4	Fall 5	Fall 6
1	1,001	1,003	1,002	1,000	0,995
2	1,012	1,009	1,009	1,008	1,006
3	1,007	1,005	1,005	1,003	1,004
4	1,006	1,010	1,005	1,004	1,000
5	1,000	1,007	1,000	1,001	0,996
6	1,000	1,005	0,998	1,001	0,996
7	0,992	0,999	0,991	0,996	1,005
8	0,992	1,004	0,994	0,992	1,000
9	1,009	1,013	1,007	1,005	1,007
10	1,014	1,011	1,010	1,008	1,007
11	1,016	1,012	1,012	1,007	1,013
Medel	1,004	1,007	1,003	1,002	1,003
Std. avv.	1,131	1,059	1,078	1,080	1,094

**Not:** De alternativa beräkningsvägarna, fem fall enligt beskrivning i text, visas relativt indexvärdet för fall 1 samma månad.

Det kan ses från tabell 1 att samtliga alternativa fall (2-6) resulterar i snarlika index – de relativa värdena är nära befintlig specifikation (fall 1) men ligger i samtliga fall högre i genomsnitt, om än marginellt. Fall 2 uppvisar störst variation relativt befintlig specifikation vilket kan utläsas från standardavvikelsen som är 13,1% större.

I tabell 2 visas på bortfallsandelarna (i procent) av varukorgen relativt månad januari (som är den första prisjämförelsemånaden). I jämförelse mot befintlig specifikation, fall 1, får övriga specifikationer av avgångstid (fall 2- fall 5) genomgående högre bortfall eftersom de är mer detaljerade. Fall 6 har lägst bortfall med anledning av dess grovhet i definitionen.

**Tabell 2** Bortfall i procent för de fem fallen jämte befintlig specifikation

Period	Fall 1	Fall 2	Fall 3	Fall 4	Fall 5	Fall 6
1	0	0	0	0	0	0
2	0,58	1,14	0,61	0,75	0,77	0
3	0,04	0,54	0,24	0,36	0,35	0
4	0,08	1,01	0,32	0,34	0,74	0
5	0,42	1,64	0,77	0,58	1,57	0
6	0,25	0	0,65	0,1	0,26	0
7	14,89	29,81	16,16	15,72	25,73	9,82
8	0	2,4	0,37	0,48	1,64	0
9	0,29	2,48	0,85	0,8	1,53	0
10	0,17	3,17	0,65	1,04	2,21	0
11	5,09	13,72	5,4	8,11	10,27	2,53
Medel	2,18	5,59	2,60	2,82	4,51	1,24

**Not:** Bortfall har räknats som andel icke-funna prisnoteringar i aktuell månad relativt antal prisnoteringar i januari enligt specifikationerna (fall 1- fall 6).

I tabell 3 återges utfallet från en variansanalys där prisvariabeln (beroende variabel) utformats enligt de fem fallen (2-6) och kan jämföras med befintlig specifikation (fall 1). Beroende variabel blir logaritmerad priskvot (aktuell

månads pris/basmånadens pris) enligt specifikation i respektive fall. Varianserna skiljer sig åt i storlek/nivå mellan fallen, liksom respektive motsvarande mängd observationer med anledning av skillnaderna i utformning.

**Tabell 3** Förklaringsgrader för samtliga fall. Beroendevariabel log(Priskvot) enligt respektive fall.

Fall	R2	Noteringar	Kvadratsumma
1	0,2054	25 847	3 276
2	0,1235	82 691	5 688
3	0,1828	26 452	3 424
4	0,1552	44 251	4 485
5	0,1529	57 241	4 825
6	0,4089	7 309	2 041

**Not:** Kolumn R2 anger förklaringsgrad, alltså enligt modellen förklarad variation i total kvadratsumma (sista kolumnen). I samtliga fall är resa (destinationer), veckodag, resenärstyp och period beaktade som effekter.

I tolkningen av tabell 3 bör inses att den ökande förklaringsgraden i fall 6, som också är den mest aggregerade formen (och därmed något intet-sägande), kommer från att dagsmönstren kan vara mer lika varandra än exempelvis timme-mönstren liksom att timprisernas avvikelser är fler (fler timmar på dygnet än dagar i veckan).

## Diskussion

Transaktionsdata är en ny datakälla för järnvägsresor och har implementerats enligt resonemanget i avsnittet om *Produkterbjudande*, alltså understödd av branschkunskap och analys. Vi ser i denna PM och från tidigare diskussioner i nämnden att det finns flera möjligheter att identifiera prisvariabeln mot bakgrund av vad som bedöms vara homogent/tillräckligt homogent. Att göra en helt uttömmande analys över alla dimensioner är svårt och tidskrävande. Utformningen bör vara förnuftig och relevant för syftet att följa prisutvecklingen för järnvägsresor.

## Referenser

Bubuioc, R. och Tongur, C. (2019). *Järnvägsresor i KPI – ändringar från 2019*. PM till nämnden för KPI, möte nr. 6, 2019-05-15.

Bubuioc, R., Sandén, B., Tareke, M., Johansson, J. och Tongur, C. (2018). *Ny datakälla för Järnvägs-, flyg- och charterresor*. PM till nämnden för KPI, möte nr. 5, 2018-10-16.