RAPPORT

Förseningstidsvärdering - FTV

Förstudie av effekterna av olika typ av förseningar på järnväg

Diarienummer: TRV 2015/84817

Sida Jiang, Karin Brundell-Freij & Christer Persson

2017-06-07

Konsulter inom samhällsutveckling

WSP Analys & Strategi är en konsultverksamhet inom samhällsutveckling. Vi arbetar på uppdrag av myndigheter, företag och organisationer för att bidra till ett samhälle anpassat för samtiden såväl som framtiden. Vi förstår de utmaningar som våra uppdragsgivare ställs inför, och bistår med kunskap som hjälper dem hantera det komplexa förhållandet mellan människor, natur och byggd miljö.

Titel: Förseningstidsvärdering - FTV - Förstudie av effekterna av olika typ av förseningar på järnväg

WSP Sverige AB
Arenavägen 7
121 88 Stockholm-Globen
Tel: 010-722 50 00
E-post: sida.jiang@wspgroup.se
Org nr: 556057-4880
Styrelsens säte: Stockholm
www.wspgroup.se/analys

Innehåll

Sammanfattning

Syfte

Syftet med detta forskningsprojekt har varit att studera skillnader i värderingar av förseningar i trafik som beror på om den resenären har fått information i förväg om förseningen eller inte. Syftet har till exempel relevans för avvägningar om hur järnvägsunderhåll ska utföras, till exempel om underhåll bör utföras under ordinarie schemalagd trafik eller om det är bättre att trafiken stängs av samtidigt som resenärerna informeras om de åtföljande förseningarna.

Upplägg av studien

I projektet särskiljs därför mellan två typer av förseningar. *Oinformerade* förseningar som är de förseningar som sker vardagligen i alla former av trafik och som resenären använder för att bygga upp en bild av hur stora restidsvariationerna är inom en viss trafikmiljö. I de analyser som har gjorts inom projektet har oinformerade förseningar representerats i form av restiders standardavvikelse. Den andra typen av förseningar där resenären innan resan startar har fått information som gör det troligare att en försening kommer att ske har vi kallat för *informerade* förseningar. Sådan information kan vara i princip vad som helst som påverkar resenärens bedömning av hur troligt det är att en försening uppstår, till exempel väderlek, rapporterade fel och haverier i trafiksystem och aktuell förseningsinformation. I projektet har fokus varit på den information om försening på det aktuella tåget som finns innan avgång. Vi har dock även testat att analysera hur faktiskt inträffade förseningar (vid slutlig ankomst) ”påverkar” resenärens reseval. (För att resenären skall kunna basera sina val – som görs innan resan startar - på information om den faktiska förseningen vid slutstationen förutsätter det till synes orakelliknande förmågor hos resenären. Men den faktiska ankomstförseningen kan i det här fallet ses som en proxy för de förutsägelser som en kompetent och erfaren resenär kan ha gjort innan resan startar).

De analyser som har gjorts i projektet har varit avgränsade till tågtrafik. Data har hämtats från biljettstatistik som visar enskilda tågresor mellan start- och målstation. Dessa data har påförts trafikeringsdata från det så kallade TFÖR-systemet om enskilda tågs tidföring. Data har begränsats till tre enskilda valsituationer som kan ge information om hur resenärer har agerat avseende oinformerade och informerade förseningar. Varje valsituation består av två stycken alternativ av rutter/avgångar som resenären har ett val mellan. De tre valsituationerna är:

A. Örebro-Stockholm under morgonens högtrafik: alternativ 1 innefattar ett byte i Hallsberg, alternativ 2 är hela sträckan utan byte.
B. Örebro-Stockholm ej-högtrafik under eftermiddagen: alternativ 1 innefattar ett byte i Hallsberg, alternativ 2 är hela sträckan utan byte.
C. Borlänge-Stockholm under morgonens högtrafik: alternativ 1 är regional tåg, alternativ 2 är X2000.

Olika varianter av logitmodeller har estimerats på dessa data, så kallade enkla binomiala logitmodeller och så kallade mixade logitmodeller. De mixade modellerna tar hänsyn till att resenärer har sinsemellan olika värderingar. Båda modellslagen har gett tämligen likartade resultat. Tågens *planerade* restid enligt tidtabellen ingår som förklaringsvariabel i samtliga estimeringar. Om inget annat nämns så är det denna variabel som kallas "restid".

Oinformerade förseningar representeras i estimeringarna genom den så kallade restidsosäkerheten som mäts av standardavvikelsen för tågets restid. Standardavvikelsen har beräknats för alla avgångsdatum under innevarande säsong. Betydelsen av *informerade* förseningar har testats vid estimeringarna genom att inkludera tågets eventuella avgångsförsening från utgångsstationen (i enstaka skattningar användes istället ankomstförsening vid resans målpunkt, se resonemanget ovan).

Reskostnader ingår inte som förklaringsvariabel vid estimeringarna. Det innebär i sin tur att inga enskilda "värderingar" (betalningsvilja) kunnat skattas för variablerna, t.ex. så kallad restidsvärdering. Däremot går kvoten mellan olika värderingar att skatta. Så kallade reliablity ratio (RR) som är kvoten mellan restidsvärdering och värderingen av oinformerade förseningar (standard avvikelse restid) har därför beräknats. RR är ett vanligt mått för att presentera förseningsvärderingar på.

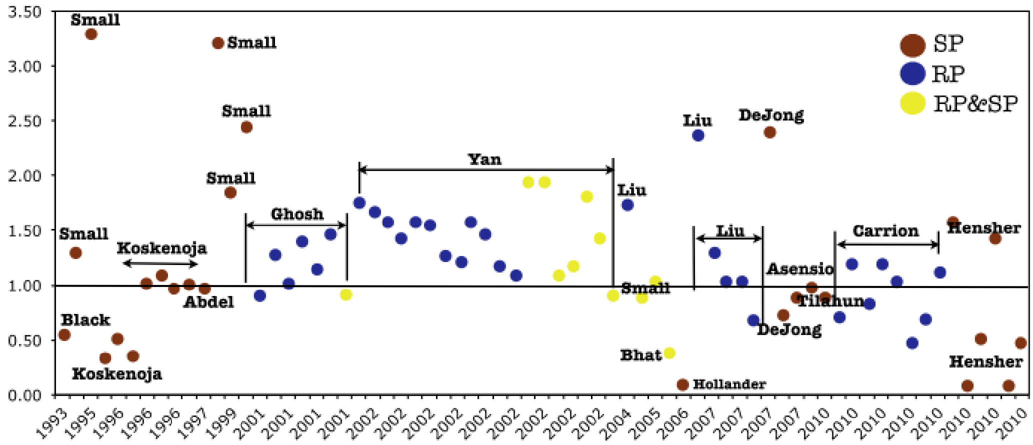
Resultat

I tidigare studier av resenärernas "värdering" (upplevd resuppoffring) kopplad till osäkra restider och förseningar, har man normalt baserat analysen på hypotetiskt beteende – så kallade SP data. På senare år har man kommit att alltmer ifrågasätta validiteten i resultat som baseras på SP-studier. Detta projekts huvudresultat kan därför sägas vara att det visade sig vara möjligt att använda den typ av data om observerat beteende (så kallade RP-data) som har funnits tillgängliga i projektet till att skatta en modell för hur oinformerade förseningar (restidsosäkerhet) påverkar individuellt resebeteende.

Tre variabler har varit i fokus under studien: restid, avgångsförsening (informerad försening), typiska förseningar/restidsosäkerhet (oinformerade förseningar). Data har valts så att vi har tre valsituationer där data är tämligen lika (homogent) för övriga variabler som kan tänkas påverka valet mellan tågavgångar, till exempel reseavstånd. Fördelen med detta upplägg är att det är tämligen enkelt att estimera hur just de tre studerade variablerna påverkar valet mellan tåg. Men det finns också problem med detta upplägg. Ett sådant problem är att vi får låg grad av variation i data som i sin tur kan leda till svårigheter att skilja mellan estimerade effekter för de tre studerade variablerna och variabler som inte ingår explicit i analyserna, till exempel pris eller tid på dagen. Ett annat problem är att homogeniseringen av data leder till att det endast täcker en begränsad del (resor 20-25 mil) av marknaden för tågresor med regional- och fjärrtåg.

För restidsvariabeln har estimeringarna gett korrekt tecken för restidskänsligheten. Restiden varierar bara med enstaka minut <u>inom</u> respektive valsituation, och det estimerade värdet bygger därmed i princip helt på variationen <u>mellan</u> de tre valsituationerna. Därför är det betryggande att det skattade värdet alltså åtminstone har rätt tecken. Men det går inte att komma ifrån att data med betydligt fler valsituationer vore mycket önskvärt. Eftersom ingen kostnadskänslighet har estimerats så går det inte att kontrollera rimligheten i den estimerade restidskänsligheten genom att beräkna restidsvärderingen. Men som beskrivs nedan så ger dess kvot med känsligheten för oinformerad försening ett resultat som ligger i linje med tidigare studier.

Känsligheten för så kallade oinformerade förseningar, mätt som standardavvikelse för restiden per säsong, får också rätt tecken vid estimeringarna. RR (kvoten, känslighet för oinformerad försening delat med känsligheten för restid) har skattats till värdet 1,13 vilket ligger i linje med tidigare studier. Figur 1 reliability ratio (RR) för utvalda tidigare studier. nedan sammanfattar RR erhållna från tidigare studier.



Figur 1 reliability ratio (RR) för utvalda tidigare studier.

Alla specifikationer som har testats där den ”informerade förseningen”, mätt som avgångsförsening från utgångsstationen, ingår ger problem i skattningen. I de flesta specifikationer som variabeln har ingått i så får dess känslighet fel tecken. När vi separerar valsituationerna och skattar separata modeller för dem, får känsligheten visserligen rätt tecken men till priset av att känsligheten för oinformerad försening tappar sin signifikans. Det finns antagligen flera orsaker till problemen. En kan vara att de fall då resenären faktiskt kan utnyttja information om avgångsförsening till att förbättra sitt val av tåg troligen är tämligen sällsynta och dessa behöver troligen identifieras för att en rimlig estimering ska kunna göras. En annan förklaring kan vara att informerade resenärer kan veta att avgångsförseningen inte säger särskilt mycket om vilken ankomstförsening man behöver befara vid slutdestinationen. (I våra data visade sig avgångsförsening och ankomstförsening till slutdestinationen till och med vara negativt korrelerade på övergripande nivå!). Dessutom finns också teoretiska invändningar mot en specifikation där avgångsförsening estimeras med en helt egen parameter (känslighet) samtidigt som den oinformerade förseningen estimeras med hjälp av restidens standardavvikelse. Mycket tyder på att informerad försening istället bör föras in i modellen utan att någon extra parameter behöver estimeras, men genom att standardavvikelsen justeras (Informerad försening innebär att resenären inte är lika osäker på vilken restid som kommer att gälla).

Vad kan göras i framtiden?

Inom studien har en ny typ av data i form av biljettstatistik redovisad på låg nivå i form av enskilda resor, använts. Analyser i form av modellestimeringar på dessa data har utförts för val mellan tågavgångar som en funktion av restid, oinformerad försening och informerad försening. Detta utföll tämligen väl för restid och oinformerad försening. Dock var det alltså inte möjligt att skatta fram trovärdiga värden när det gäller betydelsen av informerad försening. Resultat om informationens värde, det vill säga värderingsskillnader mellan informerade respektive oinformerade förseningar, bör på sikt ses som huvudsyftet för fortsatta studier. Framtida insatser inom området bör därför fokusera på att erhålla resultat om detta.

Innan studien genomfördes fanns osäkerheter om huruvida

1. datakvaliteten var tillräckligt hög i denna situation som kräver att data från flera källor kombineras och att den ena datakällan (biljettstatistik) inte tidigare har använts,
2. data medgav estimering av logit modeller där relevanta variabler ingår
3. tillförlitliga effekter kunde estimeras för de tre relevanta variablerna restid, oinformerad försening och informerad försening.

För dessa tre osäkerheter gäller att efter innevarande studie går det att säga att datakvaliteten troligt är tillräcklig för syftet och att data medger estimering av modeller som är lämpliga för syftet. Däremot så är inte den tredje osäkerheten uträtad eftersom inga tillförlitliga resultat gick att säkerställa för informerad försening.

En särskild svårighet är att vi har studerat hur resenärer väljer mellan tåg. Det innebär att vi missar resenärens möjlighet att anpassa sig till osäkerheter och förseningar genom att avstå från tågresa och antingen ta ett annat färdmedel eller att avstå från resan. Om vi i fortsatta studier kan inkludera analyser av den valsituationen kan det hjälpa till att ge ytterligare variation till data och därmed möjliggöra estimering av fler variabler i modellerna.

Det går därmed att sammanfatta vad som troligen behövs för att öka möjligheterna att estimera effekter/värderingar för informerad försening, i följande

- Ta med alternativet ”inte tåg” i resenärernas val. Det kan man göra genom att ta hänsyn inte bara till hur tågresenärerna fördelar sig mellan tågalternativen, utan också till hur många de är totalt, och hur den totalen varierar. Detta ökar variationen i data.
- Mer data i form av fler valsituationer vilket ger (1) bättre skattningar, och (2) möjlighet att studera hur värderingarna skiljer sig mellan olika typer av resor (finns indikationer i skattningarna på att värdering av osäkerhet är bimodal)
- Använda modellspecifikationer som ligger i linje med rimlig teoretisk förståelse av vad informerade förseningar innebär. Detta handlar om att bättre förstå hur resenärer skapar sig en bild av slutlig försening med hjälp av olika källor som ger in-

formation före resan om den slutliga förseningen (denna bild skapas troligen i ett samspel mellan teoretiska överväganden och empiriska erfarenheter).

1 The scope of Study

With the deregulation of railway operations and modern information system practiced in an increasing manner in Sweden, travelers has more accessibility and flexibility yet inevitably more complicated travel choices to make from time to time. In this environment, travelers encounters reliability issues in the form of delays when using railway. There are reasons to assume that travelers value reliability in form of the expected day-to-day variability differently from delays relating to unexpected or surprising events. We will use the term *informed* for describing the former type of delays or reliability and the term *uninformed* to describe the latter type of delays. This study is an attempt to identify these different type of delays in railway traffic and to provide estimates of their valuation, in economic sense. Areas of application for these valuations are in scheduling of maintenance measures and in providing information to travelers about the scheduling.

Typically, in a Swedish context, value of reliability (VoR) has been included in value of travel time stated preference (SP) studies. There are issues with the SP method for obtaining VoR, for examples how to represent and present reliability which is a measure of a probability distribution and whether there is a firm link between stated preferences in a specific hypothetical choice situation and revealed behavior in a traffic context. These issues will be even larger for this proposed study where we need to distinguish between expected reliability and unexpected delays. Therefore, we have chosen to use data on passengers observed revealed preferences. Literature reviews indicated advanced methods to examine different type of delay using hybrid model (significant developments in Walker & Ben-Akiva, 2002[1]), revealed preference (RP, see reviews in Carrion et al. 2012)[2].

The data in the study has are from two sources, ticket sales data from SJ and data on train movements from Trafikverket's TFÖR database. Both data sources are for the year 2009. The main reason for using relatively old data is that it has given us the opportunity to use detailed ticket sales data on individual train departures, which provide travelers' route choice for the analysis. The data includes 60 545 individual trips on travelers route choice for two specific trip relations. These two relations have been chosen in order to make homogenize the data. Homogenization of data is a reasonable strategy for a new area it will make estimated valuations more accurate but in the same time limit the generalizability of these estimates to trip relations with other characteristics. The chosen trip relations may be described as long-distance non-commuting trips with travel distances between 200 and 250 kilometers.

1 Walker J, Ben-Akiva M. Generalized random utility model [J]. Mathematical Social Sciences, 2002, 43(3): 303-343.

2 Carrion C, Levinson D. Value of travel time reliability: A review of current evidence [J]. Transportation research part A: policy and practice, 2012, 46(4): 720-741.

2 Hypothesis and Model Design

2.1 Empirical Setup of the Hypothesis

To test the hypothesis above on different valuation of uninformed and informed reliability and delays we use route choice data for railway traffic. This is a new area of research and it may be difficult to empirically establish the proposed difference in valuation. Therefore, the hypothesis is tested on a homogenized route choice data set consisting of three choice relations with the central station in Stockholm as one end-point. All three choice relations have similar trip distances 200-250 kilometers. The choice relations are depicted in figure 1 below.

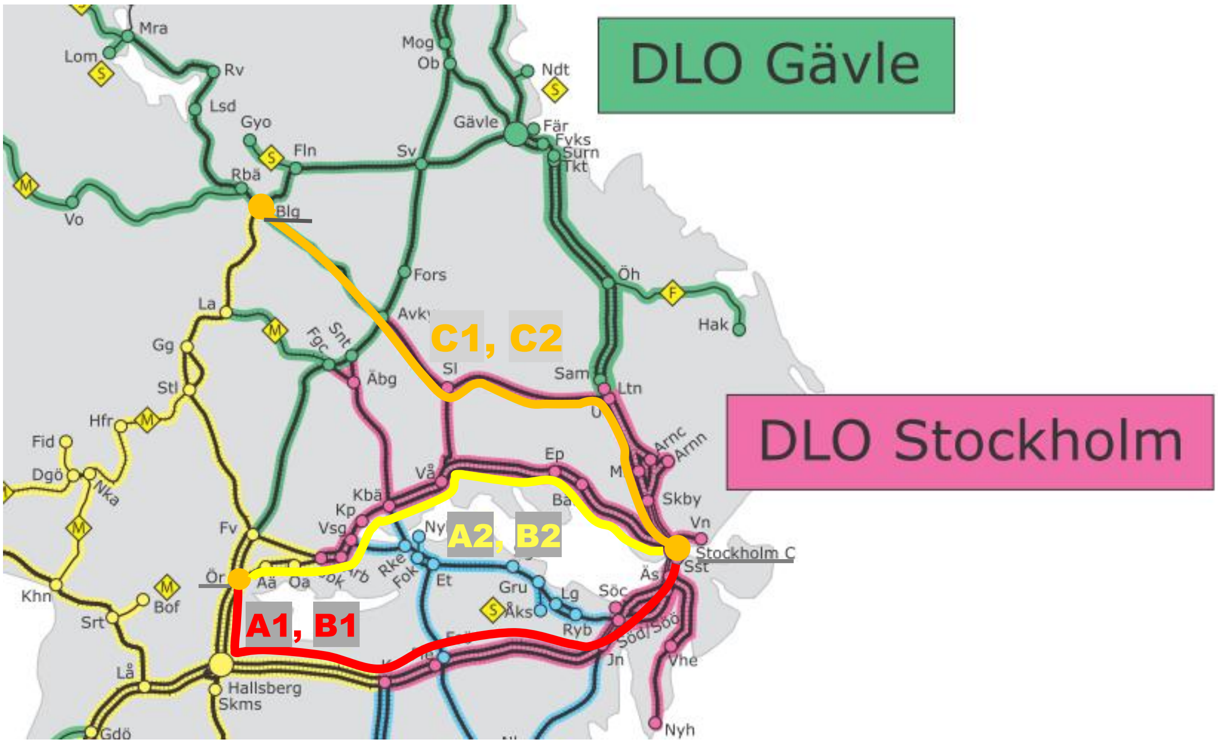


Figure 2 Illustration of studied railway routes

A. From Örebro to Stockholm in the morning peak hours: alternative 1 is a transfer via Hallsberg and alternative 2 is a non-transfer train,
B. From Örebro to Stockholm in the off-peak afternoon: alternative 1 is a transfer via Hallsberg and alternative 2 is a non-transfer train,
C. From Borlänge to Stockholm in the morning peak hours: alternative 1 is regional train and alternative 2 is high-speed x2000.

2.2 Binomial Logit Model

For instance, in choice task A, travelers can only choose either alternative 1 or 2, and both alternatives can be characterized with planed travel time PTT, departure delay (usually informed at the station) D^{inf} and the uninformed but self-estimated travel time reliability – standard deviation of travel time $SD(D^{uninf})$, travel cost C and alternative specific constants ASC. So the utility for both alternatives can be formed linear-additive as follows:

$$U_1 = \boldsymbol{\beta_T} * \mathrm{PTT}_1 + \boldsymbol{\beta_{D^{uninf}}} * \mathrm{SD}(\mathrm{Duninf}_1) + \boldsymbol{\beta_{D^{inf}}} * \mathrm{Dinf}_1 + \boldsymbol{\beta_C} * \mathrm{C}_1$$

$$U_2 = \boldsymbol{\beta_T} * \mathrm{PTT}_2 + \boldsymbol{\beta_{D^{uninf}}} * \mathrm{SD}(\mathrm{Duninf}_2) + \boldsymbol{\beta_{D^{inf}}} * \mathrm{Dinf}_2 + \boldsymbol{\beta_C} * \mathrm{C}_2 + \boldsymbol{ASC}$$

Travel time PTT is the planed travel time or interchangeably time-table travel time, PTT is rather fixed and can be considered only varies over alternatives and choice tasks, thus the difference between PTT over two alternatives can function as choice-task specific constants (CTSC), and due to its importance in explaining the utility and to prevent high correlation with ASC, PTT (or CTSC) is employed to replace ASC.

Travel time reliability or the risk of delay against time table is difficult to model and results in a massive variety of indicators for theoretical and practical appliance. Börjesson & Eliasson (2011)[3] concluded evaluation of travel time reliability varies over the frequency of travel delay; Karlström and Sundberg (2010)[4] are developing an integral of travel time distribution as indicator for travel time reliability. The difficulty to implement the advanced indicator of travel time reliability suggested above lies partly in both requires a-priori expertise to either categorize frequency or the travelers need to take integral of distribution into considerations. Partly also to make the results comparable with other SP/RP studies in Sweden and worldwide, this pre-study begins with standard deviation of travel time as indicator of risk being delayed. Yet it is not informed or deterministic, which invites random effects from person to person, together with the needs of posterior analysis, the project therefore employs also mixed logit model for better design and modal fit.

The concept of informed delay has its very dependence upon when it is informed:

As seen above that the decision usually made ahead of the trip thus informed departure delay at the first station has if existent very marginal effect upon the decision making. But departure delay enters into the utility equations, since departure delay is informed and deterministic and its difference from un-informed delay make it interesting to test, while we need to remember only its relative difference from un-informed delay is interesting, and refinements needs to put into the model before we can draw solid conclusions about how informed delay affects trip choice.

---

3 Börjesson M, Eliasson J. On the use of “average delay” as a measure of train reliability [J]. Transportation Research Part A: Policy and Practice, 2011, 45(3): 171-184.

4 Fosgerau, M., Karlström, A. The value of reliability, Transportation Research Part B: Methodological 44, no. 1, 2010, 38-49.

Travel cost C is not available at the individual purchase level therefore been assumed the same and quantified with an average over all individual tickets in the studied alternative, and in this case it is missing thus enters into the constant term CTSC. Noted that if there is no constant term in utility function and important explanatory variable such as travel cost is missing, will lead to the violation of logit model's basic claims on the error term, different model such as Probit logit is recommended to test and see if can improve the modal fit.

All the parameters are to be estimated using maximum log-likelihood[5] in R[6].

$$LL(\Omega) = \sum_{n=1}^{N} \ln(P_{n,\ i}(\Omega))$$

Where

$$P_{n,\ i}(\Omega) = \int_{\beta_x} \left[\frac{e^{u_i}}{\sum_{j=1}^{2} e^{u_j}}\right] d\beta_x$$

2.3 Mixed Logit Model

The utility equation for mixed logit with randomized effects of travel time and travel time reliability (or interchangeably un-informed travel delay in this case) can be assumed following normal distribution over the travelers, so that the unconditional choice probability:

$$P_{n,\ i}(\Omega) = \int_{\beta} \left[\frac{e^{\beta x_i}}{\sum_{j=1}^{J} e^{\beta x_j}} \boldsymbol{f}(\boldsymbol{\beta_T} |\ \boldsymbol{\Omega})\right] d\beta$$

The difference from ground model is $\boldsymbol{f}(\boldsymbol{\beta_T} |\ \boldsymbol{\Omega})$ - probability density function of randomized parameter been multiplied with logit model. To establish mixed model, firstly to generate uniform random draws using Halton sequences[7] then using inverse function of cumulative density function to derive **normal draws following *N(0, 1)* distribution** – with mean 0 and standard deviation 1. And with standardized normal distribution, the utility function with randomized effects of travel time and travel time reliability can be rewritten as follows:

$U_1 = (\boldsymbol{\beta_T} + \boldsymbol{\sigma_T} * \boldsymbol{draws_T}) * \mathrm{PTT}_1 + (\boldsymbol{\beta_{D^{uninf}}} + \boldsymbol{\sigma_{D^{uninf}}} * \boldsymbol{draws_{D^{uninf}}}) * \mathrm{SD}(\mathrm{Duninf}_1) + \boldsymbol{\beta_{D^{inf}}} * \mathrm{Dinf}_1$

$U_2 = (\boldsymbol{\beta_T} + \boldsymbol{\sigma_T} * \boldsymbol{draws_T}) * \mathrm{PTT}_2 + (\boldsymbol{\beta_{D^{uninf}}} + \boldsymbol{\sigma_{D^{uninf}}} * \boldsymbol{draws_{D^{uninf}}}) * \mathrm{SD}(\mathrm{Duninf}_2) + \boldsymbol{\beta_{D^{inf}}} * \mathrm{Dinf}_2$

---

5 Kleinbaum D G, Klein M. Maximum likelihood techniques: An overview[M]//Logistic regression. Springer New York, 2010: 103-127.

6 R code is origined from Hess et.al., University of Leeds and modified by the authors.

7 Halton method divides 0 − 1 space into p_k segments (with p_k giving prime used as base for parameter k), and by systematically filling the empty spaces, using cycles of length p_k.

So that the choice probability can be reformed with *N(0,1)* random draws as follows:

$$P_{n,\ i}(\beta_x, \sigma_x) = \int_{\varepsilon_x} \left[\frac{e^{\beta x_i + \sigma_x \varepsilon_x x_i}}{\sum_{j=1}^{J} e^{\beta x_j + \sigma_x \varepsilon_x x_j}} \boldsymbol{\phi}(\mathbf{0}, \mathbf{1}) \right] d\varepsilon_x$$

Instead of estimating fixed parameter for travel time and travel time reliability, both the mean $\boldsymbol{\beta_T}$, $\boldsymbol{\beta_{D^{uninf}}}$ and standard deviation $\boldsymbol{\sigma_T}, \boldsymbol{\sigma_{D^{uninf}}}$ is to be estimated based upon simulated log-likelihood maximization (Bhat, 2001)[8].

3 Data

3.1 RP Data processing

Ticket sales data obtained from SJ for three studied choice tasks has been firstly converted to individual choice datasets shown as follows:

Table 1 Raw data from ticket sales from SJ, 2009

date	*tot_ind_train1*	*tot_ind_train2*	*path*	*rate*
2009-09-01	4	148	ÖR-CST	0,026

Table 2 Converted individual choice dataset

date	***ID***	***Choice***	*Path*	*Choice task (1,2,3)*
2009-09-01	**1**	**1**	ÖR-CST	1
2009-09-01	**2**	**1**	ÖR-CST	1
2009-09-01	**3**	**1**	ÖR-CST	1
2009-09-01	**4**	**1**	ÖR-CST	1
2009-09-01	**5**	**2**	ÖR-CST	1
		…		
2009-09-01	**152**	**2**	ÖR-CST	1

Choice task 1 = A, 2 = B, 3 = C

148 rows for 148 individuals that choose train 2

Since we disaggregate the data into individual data without private information on each individual, we may have included the same "real" individual for several trips/choices. Yet in data, we treat all individuals as making separate, uncorrelated decisions. This may lead to small portion of the observation is actually repeating choice of identical individuals, therefore we will estimate the variance 'between individuals' to be smaller than it really is.

8 Bhat C R. Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model [J]. Transportation Research Part B: Methodological, 2001, 35(7): 677-693.

Then the outliers in the choice dataset is removed to minimize its distortion to further model estimation. The top 20 outliers is removed, based on their observed choices being highly improbable according to an initial estimated model, shown as follows:

```
       ID Av prob per choice
 [1,] 34793      1.132710e-06
 [2,] 34794      1.132710e-06
 [3,] 34795      1.132710e-06
 [4,] 34796      1.132710e-06
 [5,] 34797      1.132710e-06
 [6,] 34798      1.132710e-06
 [7,] 34799      1.132710e-06
 [8,] 34800      1.132710e-06
 [9,] 34801      1.132710e-06
[10,] 26375      9.162638e-05
[11,] 26376      9.162638e-05
[12,] 26377      9.162638e-05
[13,] 26378      9.162638e-05
[14,] 26379      9.162638e-05
[15,] 26380      9.162638e-05
[16,] 26381      9.162638e-05
[17,] 26382      9.162638e-05
[18,] 26383      9.162638e-05
[19,] 26384      9.162638e-05
[20,] 26385      9.162638e-05
```

Figure 3 Removal of outliers from RP data

3.2 Integration with TFÖR data

TFÖR data composed of observed train movements at all the stops during the train journey and owned by Trafikverket, train movements of all three choice tasks during 2009 has been processed, including consistency check of departure time, arrival time for observation and corresponding time table. TFÖR data was merged with RP data so that all time variables of all alternative trains over three choice tasks together with individual choices integrated into one dataset, shown as follows:

Table 3 Integration of RP with TFÖR data

date	*ID*	*choice*	*path*	*choice task*	*ptt1*	*ptt2*	*sd1*	*sd2*
2009-09-01	1	1	ÖR-CST	1	120	122	11,99	7,31
2009-09-01	2	1	ÖR-CST	1	120	122	11,99	7,31
2009-09-01	3	1	ÖR-CST	1	120	122	11,99	7,31
2009-09-01	4	1	ÖR-CST	1	120	122	11,99	7,31
2009-09-01	5	2	ÖR-CST	1	120	122	11,99	7,31
		...						
2009-09-01	152	2	ÖR-CST	1	120	122	11,99	7,31

Notes: all time variable in units of minutes

sd is standard deviation of travel time over corresponding season.

TFÖR data

The integration is based upon train number and date.

1, 2 correspond to train 1 and train 2 respectively.

3.3 Descriptive statistics

RP data for three choice tasks obtains from ticket sales from SJ, 2009 and descriptive statistics is summarized in the table:

Table 4 Input variables and distribution of choices. Descriptive statistics

Choice tasks	Train no. (x2000)	pl.tt (min)	Average departure delay (min)	Average choice probability y				Standard deviation of travel time (season-wise)			
				spring	summer	autumn	winter	spring	summer	autumn	winter
1	617 + 420	120	1,05	5,9%	7,2%	5,1%	5,8%	7,73	16,39	11,99	14,38
	714	122	6,16	94,1%	92,8%	94,9%	94,2%	3,32	11,79	7,31	14,74
2	175+638	147	9,64	9,7%	8,2%	5,9%	7,8%	6,02	34,33	14,94	20,04
	176	127	14,53	90,3%	91,8%	94,1%	92,2%	9,43	27,06	22,43	11,5
3	11	151	6,68	10,3%	10,7%	13,8%	12,7%	6,58	66,55	9,25	8,49
	591	125	3,63	89,7%	89,3%	86,2%	87,3%	11,66	28,45	6,09	**12,59**

Notes: in choice task 3 from Borlänge to Stockholm was under maintenance in the month of July so there was no train traffic during this period. Green train number is high speed train X2000 and blue train number is conventional (regional) trains.

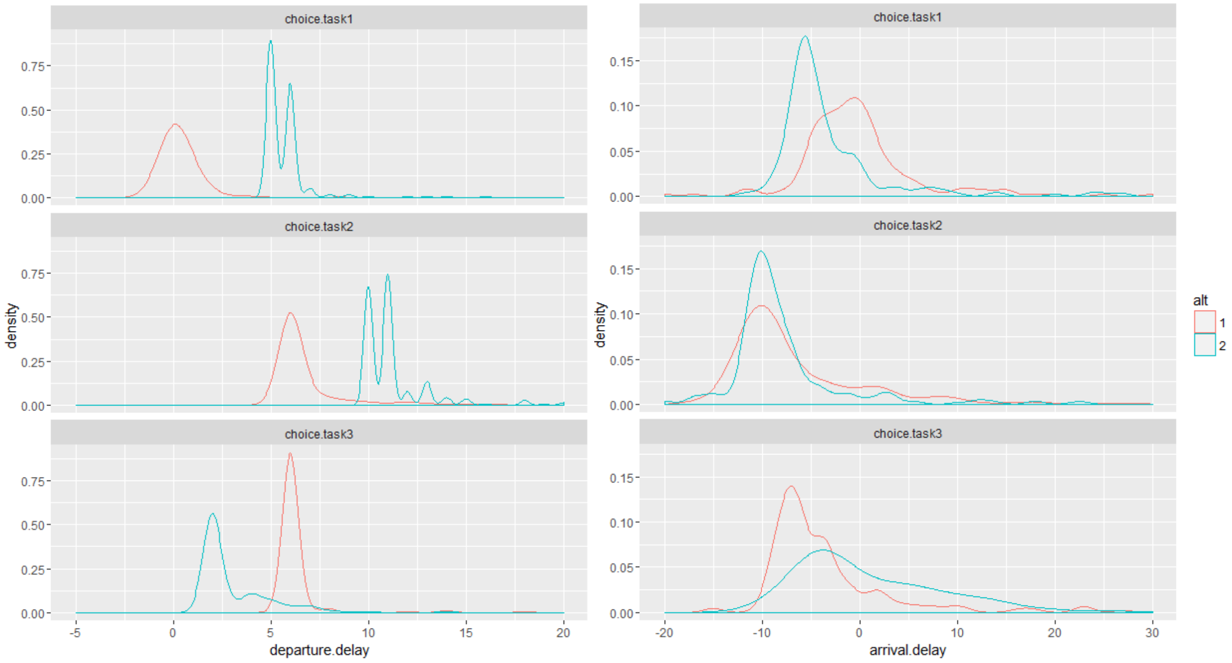
For choice task 1 there are 2 alternative railway routes from Örebro to Stockholm, alternative 1 with combination of train 617 and train 420 transfer at Hallsberg is indirect line. Despite the necessary train transfer, its planned journey time is 2 minutes shorter than the direct alternative with train 714 via Mälarbanan (according to TFÖR). This is because train 420 runs faster (X2000 train) and has fewer stops compared with regional train 714, while the transfer time between train 617 and 420 is only 12 minutes. Obviously, the variables transfer time and number of stops will in themselves contribute to disutility for travelers, but will also affect journey time and travel time reliability. Therefore, to avoid collinearity among variables in the utility function; we have chosen not to single out the effect of transfer and number of stops. Instead, these variables have been included in the comprehensive description of travel time and reliability for the whole journey.

Standard deviation of travel time in summer is usually highest in the year because of maintenance work. The model has been tested with and without summer data, but the estimation results seems not significant from each other so that we still include summer in the analysis.

3.4 Departure delay vs. arrival delay

Departure delay is informed delay upon departure that is different from un-informed delay in its deterministic character for a specific individual. Traveler can try to predict arrival

Figure 4 comparison of departure delay and arrival delay for all the choice tasks delay of interest (final destination) based on departure delay from the start station, but to certain extent, departure delay often can be caught up through the journey, so it is important to investigate the correlation between departure delay and arrival delay of studied routes before drawing conclusions from models.



Note: different x- and y-scale for panel figures due to different variance of departure delay (left panel) and arrival delay (right panel).

Somewhat surprisingly, the correlation between typical departure delays and arrival delays over choice task is seemly negative. For instance in choice task 1, the red-colored alternative 1 is rarely delayed more than 3 minutes on departure, while alternative 2 (blue colored) almost always departs more than 5 minutes late from Örebro. Never the less, it is on average the red colored alternative 1 that ends up with slightly larger arrival delay in Stockholm.. Similar pattern can be found across all the choice tasks. This negative correlation is probably an important explanation behind the counter-intuitive result that increasing departure delay was estimated to contribute positively to utility, in many of our specifications. (see chapter 4).

4 Estimations

4.1 Ground model

Each "observation" in RP data is assigned with unique ID, although we have no information if same individual traveled in different dates in one choice task, or traveled even in more than one choice task. The insufficiency of personal information in this case can lead to the overestimation of statistics t-value; we can only assume that such individuals are rather marginal proportion in the long-distance trip we studied. However we should not forget to bring this insight when examine if coefficient is significant from 0 or 1, and we should therefore apply higher (pseudo) T-values than 1,96 to identify "significance" for robust analysis in this dataset. The estimation data composed of 60 545 observations that is 60 545 individual choice makers and the initial log-likelihood is LL(0): -41966,6 and final log-likelihood for binomial logit model is **LL(final): -26597,8**.

Table 5 Estimates of Binomial Logit model

variable	*est*	*se*	t-value(0)	*robust se*	robust t-value(0)
travel time reliability (sd)	-0,11	0,0025	*-44,19*	0,0047	*-23,56*
planned travel time (ptt)	-0,094	0,0008	*-122,11*	0,0011	*-85,71*
departure delay (ddelay)	0,0693	0,0014	*50,76*	0,0025	*28,25*

Notes: Rho-sq: 0,37; adj. rho-sq: 0,37; AIC: 53201,64 och robust estimation for standard error computed by sandwich matrix.

All the parameters is very significantly different from 0 and 1 even considering same individuals can enters into more than 1 observations. The estimates also suggests that one evaluates travel time reliability (sd) slightly more than travel time (ptt). Further analysis for the ratio between these two valuations ("ratio of reliability") can be found in 4.2.

4.2 Reliability ratio (RR)

In linear-additive specification such as the ground model above, marginal rate of substitution may be calculated to obtain important features as value of travel time (VoT), value of travel time reliability (VoR), and the reliability ratio (RR). This is useful in our case, since travel cost is not available from data, so only RR can be obtained from ground model in the basis of current data.

$$RR = \frac{\partial U / \partial \sigma_T}{\partial U / \partial \mu_T} = \frac{VoR}{VoT}$$

Where $\partial U / \partial \sigma_T$ is coefficient estimated for travel time reliability while $\partial U / \partial \mu_T$ is coefficient for planned travel time. For a comprehensive review of current research over the reliability ratio (Carrion C. & Levinson D., 2012)[2] see the figure:

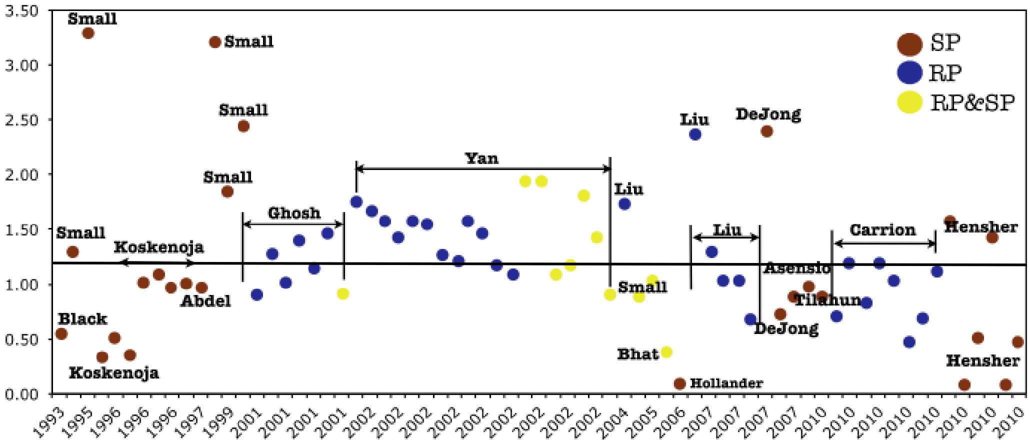


Figure 5 reliability ratio of selected studies

One may argue that estimated reliability ratios have declined over time in both SP and RP data, with RP estimates more constrained in middle of the span. In other words, travel time reliability seems to play relatively less and less roll compared with travel time in utility function. This may be explained by the increase in available travel information in general: traveler can forecast the coming journey and make changes accordingly, so that unreliable travel time become more and more predictable; also substitution with digital activity can help travelers make use of travel delay to the same extent that she can use planned travel time. This declining trend of reliability ratio should probably be considered when conducting Cost-benefit analysis (CBA) with travel time reliability involved.

Results from our ground model is quite in line with other RP studies with RR varied from 0,5 to 2,5. To test also for significance - whether reliability ratio in our case is significantly different from 1 (unreliability valued as travel time), and 0 (unreliability not valued at all), we apply the delta method[9] is implemented and the results can be seen as follows:

Table 6 significance test of RR using delta method

ITEMS	FUNCTION_VALUE	FUNCTION_SE	T-VALUE
sd-ptt	-0,016	0,0039	-4,09 (H_0: = 0)
sd/ptt	1,1701	0,0403	4,22 (H_0: = 1)

RR is exchangeable to sd/ptt in the case the estimated value is 1,17 and with a high t-value strongly suggesting that the value is significantly different from 1. The difference (sd-ptt) has an estimated value of -0,016, with the T-test strongly indicating that travel time reliability (sd) is significantly different from travel time ptt, which means RR is significant from 1 with travel time reliability being further negative compared with travel time. All the analysis using delta method is limited to the ground model that is with linear-additive form. For the more advanced models such as mixed logit model (for which results are

[9] Efron, Bradley. (1982). The jackknife, the bootstrap and other resampling plans. Society for industrial and applied mathematics, Vol. 38, Philadelphia: Society for industrial and applied mathematics., 1982.

presented in the following chapter) we need to simulate reliability ratio at the level of individual travelers, and calculate the mean or median of that distribution to analyses which of the two variables: travel time, or travel time uncertainty, that has the larger impact on utility per minute. That is included in the so called posterior analysis.

4.3 Mixed logit model

The travel time and travel time reliability has been randomized with standard normal draws so that mixed logit model need to estimate both mean and sigma for how the sensitivity of travel time and travel time reliability varies over the observed population. Meanwhile, mixed logit model is non-linear model which can results in numerous local optima instead of single global maximum. To handle local optima issue, different initial values has been firstly run for the first 100 iteration and the initial values with best log-likelihood is chosen for further model estimation. The initial log-likelihood is LL(0): -41966,6 and the model converged at **LL(final): -24 151,5**.

Table 7 Estimation of Mixed Logit Model (with 300 random draws)

	est	*se*	*t-value(0)*	*robust se*	*robust t-value(0)*
sd_mu	-0,46	0,01	*-48,98*	0,01	*-46,72*
ptt_mu	-0,37	0,01	*-49,59*	0,01	*-46,97*
sd_sig	-0,25	0,01	*-29,10*	0,01	*-23,64*
ptt_sig	-0,31	0,01	*-40,95*	0,01	*-33,44*
ddelay	0,17	0,00	*48,22*	0,01	*21,68*

Notes: Rho-sq: 0,42; adj. rho-sq: 0,42; AIC: 48 312,9

Sigma for both travel time (ptt) and travel time reliability (sd) are strongly significant from 0 and 1, which means variation among individuals are significantly existing and random effect can hence improve the modal fit. The same conclusion can be found by comparing LL(final), adjusted rho square and AIC that mixed logit model fits better than basic logit model.

Coefficient for departure delay (ddelay) in both ground model and mixed logit model is positive, and counter with the expectations. Yet the results is quite in line with the preliminary analysis shown in 3.1 of the relation between departure delay and arrival delay, which showed that arrival delay cannot be easily predicted from departure delay. Therefore the explanatory power of departure delay (ddelay) is likely to be confounded by other variables that varies over choice tasks. Because of this problem, inherent in data, the project focus more on the importance of planned travel time (ptt) and travel time reliability (sd). Again, this counter-intuitive finding is most probably limited to the specific study scope and in the many cases with only small departure delay. In the rarer cases where there is a long departure delay, it is most probable that departure delay will be used (righteously) as a predictor by the traveler, indicating that she can expect a longer than usual delay at the final destination also.

4.4 Posterior analysis

The mixing distributions given by the parameters in table 7 can be used directly to construct the distribution of the reliability ratio RR. However, the mixing distributions are assumed to be independent. This assumption is likely to not be fully valid. Since the distribution of especially very high RR as well as very low RR among individuals is highly dependent on the validity of this assumption, the computed RR from the mixing distributions can be expected to deviate considerably from its true value in the population. Further the the mixing distributions are assumed to be Gaussian. Under the estimated parameters given in table 7 there will, for example, be a sizeable proportion of the population for which RR will have a wrong sign. Therefore, it is a need for a more robust estimation of RR. The method used is a so called posterior analysis where, in a Bayesian spirit, the posterior distribution of RR is obtained by applying the mixing distributions to the likelihood of the data. This method can be seen as a way to correct the mixing distributions such that they comply to observed dependencies in the data (i.e the likelihood). In this sense, the obtained distribution for RR can be seen as more robust then the distribution obtained directly from the mixing distributions.

Knowing that different individual has significantly different evaluation of both travel time and travel time reliability, we can further divide the individuals into several groups conditional on their observed choice. In the meantime, the reliability ratio is no longer limited to the average level as earlier illustrated ratio of coefficients, by using posterior analysis upon the mixed logit model. Each individual is assumed to follow a random distribution (with simulated random draws) and each individual is assigned with the expected values of this random distribution termed as conditional mean. Notice that conditional mean is not the actual sensitivities for that individual but the expected mean, in other words, it is associated with different simulation of corresponding distribution and how many random draws one allows for each individual. The study uses 300 as number of random draws, furthermore sensitivity to both travel time and travel time reliability is assumed to be normally distributed.

For more details in posterior analysis of mixed logit model please refer to Greene & Hensher, 2003[10], Hess 2007[11].

The probability of observing the specific value of β given the choice of individual *n:*

---

[10] Greene W H, Hensher D A. A latent class model for discrete choice analysis: contrasts with mixed logit [J]. Transportation Research Part B: Methodological, 2003, 37(8): 681-698.

[11] Hess S. Posterior analysis of random taste coefficients in air travel behaviour modelling [J]. Journal of Air Transport Management, 2007, 13(4): 203-212.

$$\widehat{\beta_n} = \frac{\sum_{r=1}^{R}[L(Y_n|\beta_r)\beta_r]}{\sum_{r=1}^{R} L(Y_n|\beta_r)}$$

Where β_r with $r = 1, \dots, R$ are i.i.d draws, this method will relax the independent assumption of composing variables imposed by unconditional estimation, in other words, the resulting ratio of estimated coefficients supposed to be more robust and fit into the reality revealed by the data. However, this would again leads to problems with data outliers.

The descriptive statistics of posterior analysis results for travel time, travel time reliability and reliability ratio is summarized as follows:

Table 8 descriptive statistics of posterior analysis results for travel time, travel time reliability and reliability ratio

Statistics	*Travel time*	*Travel time reliability*	***Reliability ratio***
1st quartile	-0,207	-0,252	**0,806**
median	-0,106	-0,213	**1,280**
mean	-0,101	-0,214	**1,132**
3rd quartile	-0,040	-0,197	**3,808**
Std. dev. of Sample	0,137	0,050	**94,517**
Std. dev. of Sample mean	0,0006	0,0002	**0,3844**

The mean and median of RR from mixed logit model is slightly above 1 and quite close to RR value from the ground model. But also, a significant spreading of RR between 1 and 4 has been observed, which could potentially partly be explained by different trip purposes and needs to be further examined with complementary information about trip purposes.

The detailed distribution of the estimated reliability ratio is indicated in the following graph:

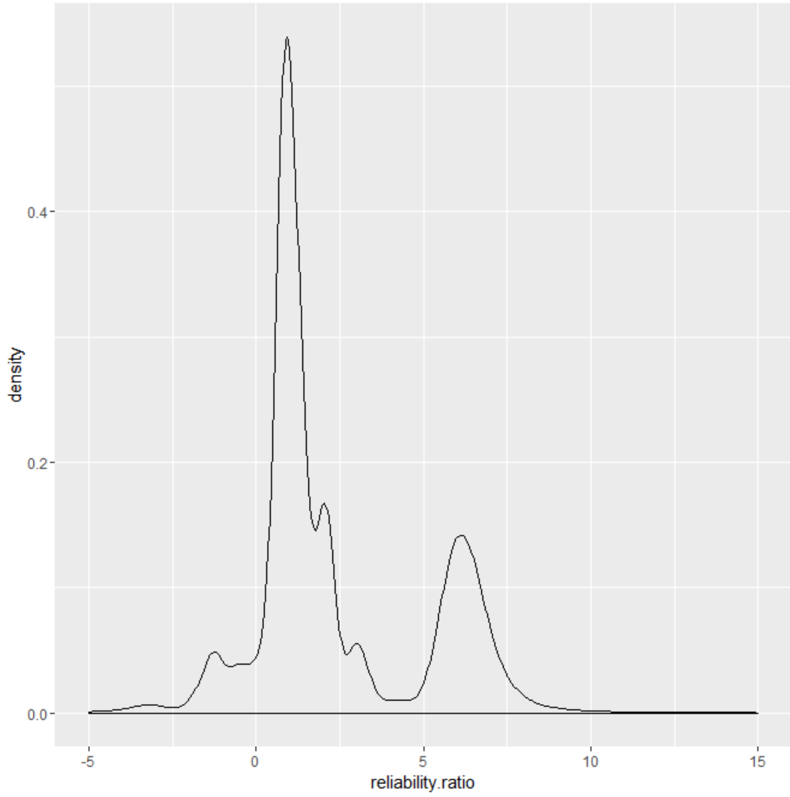


Figure 6 Probability density function of RR from posterior analysis

As expected, the major part of distribution is positive with several modes, the 1st group with more than 50% of the individuals has a RR slightly less than 1. This group of travelers are the ones for which their observed choices indicate that they are relatively less sensitive towards travel time reliability. This group is also the majority in our data, but different data and scope of study can change its dominance and thus yields very different statistics. One important point of our results is therefore that multiple clusters or groups can be seen clearly in the posterior analysis, and we may need further data such as SP to understand better how socio-economic variables or trip purpose divide the sample, and then to specify reliability ratio with respect to e.g. private/business/work trip.

The 2nd largest group with around 15% of individuals has a high RR, with an average very close to 6. This group is thus about 6 times more sensitive of travel time reliability than the overall average. The division into two groups is however not absolute: the analysis also suggests that there are also individuals with RR between 1 and 6.

As with all other results, our conclusions are limited to the range of travel distances for which we have data. Obviously the magnitude of RR can vary with among others travel distance, and the results illustrated above can only draw insights about the trips with distance between 200 and 300 km. Nonlinearity of RR with respect to travel distance can be complemented with data of other routes at different length.

5 Conclusions

In this project, ticket sales data from SJ has been transformed and treated as revealed preference data combined with travel time, travel time reliability and departure delay from TFÖR database. Both data sources can be obtained from historical database and studies over different routes and years can be conducted for other analysis purposes, but one can also – as was done in this study – use the data as “RP” (revealed preferences) to conceptually examine how travelers react to different forms of delay.

Previous studies of how travelers assess the perceived disutility of uncertain travel times and delay, have been based on analyses of hypothetical behavior - so-called SP data. In recent years, the validity of results based on SP-data has been question. This project's main results can therefore be argued to be that it proved possible to use the type of data on observed behavior (RP-data) that was available in the project, to estimate a model of how uninformed delays (travel time uncertainty) affect individual travel behavior.

As a basis for our analyses, we have used the choice task when a traveler chooses between different scheduled travel options. (One of the drawbacks of this approach is that the behavioral response not to go by train is not included at all). Our analysis circles around the travelers’ trade-off between three central qualities of the travel options: Travel time (as planned in the time table), Delay at departure, travel time uncertainty (based on the distribution of real travel times in the recent past)

Our data only comprises only three distinct choice situations (Choice situation = A pair of adjacent scheduled train departures for the same destination). Although the choice tasks have been selected so that they are similar in nature, the alternatives will inevitably differ in many more aspects than the three measured qualities that is introduced in our analyses. Therefore, there is a risk that our results are confounded by other variables with which our explanatory variables co-vary over alternatives and choice tasks. Also, only travel distances in the range 200-300 km is covered in data.

For travel time (as scheduled in time table) our estimations give the intuitively correct sign for the estimated parameter. Travel time varies only with single minutes for the same alternative in a given choice situation (due to minor modifications of the timetable during the observed year). Thus, the estimated value is based almost entirely on the variation between the three choice tasks. Therefore, it is reassuring that the estimated value at least seem to have the correct sign. Never the less, it is clear that a larger data set (that is many more choice tasks) would have been highly desirable. Since it was not possible to estimate parameters for the sensitivity to costs, it is not possible to check whether the estimated sensitivity to travel time is reasonable in terms of “value-of-time”. However, we can conclude that the ratio between the parameters estimated for travel time and travel time reliability, respectively, is very much in line with what has been estimated in previous studies.

The parameter for what was here called “uninformed delay” (Often: travel time uncertainty), measured as the standard deviation of travel time for the relevant season, also get the intuitively correct sign in the estimations. The ratio indicator RR (between the parameters estimated for travel time and travel time uncertainty, respectively) was estimated to 1.13 which is very much in line with previous studies.

Unfortunately, all specifications where the "informed delay" is included (measured as the departure delay from the first boarding station) run into estimation problems. In most cases, the parameter is estimated with a counter-intuitive positive sign. When separate models are estimated for each choice task, correct signs are estimated, but at the cost of the parameter for “uninformed delay” becoming noon-significant.

There are probably several reasons for the problems. One may be that there are rather few situations in reality in which the traveler can actually use information about departure delay to improve their choice of train. Another explanation could be that informed travelers may know that the departure delay is not a strong predictor for the arrival delay they should expect at the final destination. (In our data departure delay and arrival delays at the final destination were even negatively correlated on the overall level!). In addition, there are also theoretical objections to a specification in which sensitivity to departure delays are estimated independently, with a completely separate parameter (sensitivity) at the same time as the sensitivity to “uninformed delay” is reflected by the standard deviation of travel times. Theoretical considerations suggest that informed delay should instead be modeled as a modification of the standard deviation. (Information about delay means that the traveler is no longer as uncertain about travel times).

6 Future work

Thus, the model estimations worked well for the value of travel time and travel time uncertainty. However, it was not possible to estimate credible values regarding the importance of “informed delay”. Results of the “value of information”, as reflected by different values for informed and uninformed delays respectively, should be seen as the main ultimate objective for further studies. Future efforts in this area should focus on obtaining results thereof.

Before our study was conducted, the novel approach we were proposing raised concerns as to whether

1. Data quality was good enough, given that data from multiple sources were combined and one data source (ticket statistics) had not previously been used.
2. The data would allow estimation of logit models in which the relevant variables are included
3. It would be possible to estimate reliable parameter values for the three relevant variables journey time, uninformed delay (travel time uncertainty) and informed delay.

We can now conclude that data quality seems to be sufficient, and that the data allow the estimation of models that are suitable for the purpose. However, relating to the third concern raised, we have not yet been able to show that our approach will allow the estimation of relevant values for all three investigated characteristics (since the effect of “informed delay” was not properly reflected). However, to that specific aim we have proposed modifications of our approach (larger data sets, more variation and different model formulations) that we expect to be successful.

A particular difficulty is that we have studied how travelers choose between trains. This means that we miss in our analysis the traveler’s option to adapt to uncertainties and delays by abandoning the train altogether, either by switching to another mode, or to forgo the trip. If future studies are extended to include also such alternatives, it may help to allow the estimation of more relevant parameters.

To summarize what is probably needed to increase the possibility to estimate the effects / parameters for informed delay also:

- Allow the option "not-use-train" into the described choice situation. (In practice, this can be done by using not only distribution of rail passengers between train alternatives, but also the total number of train travelers)
- Include more data (more choice situations) to provide (1) better estimates, and (2) possibility to study how values differ between different types of travel (there are indications of the estimates on the valuation of uncertainty is bimodal)

- Apply a model formulation based on reasonable theoretical understanding of what "informed delay" means to the traveler. That is to provide a better understanding for how travelers make their predictions of the final delay based on various sources that provide information before the trip. This prediction is most likely created in an interplay between theoretical considerations and empirical experiences).

Mixed logit model has been tested with better fit for the data. In future work it would also be useful to develop that approach further, for example test different random distribution, number of draws as well as modified specifications of utility function for improvement of model fit. In a word, current results have shown differences between how travel time, uninformed delay and informed delay is evaluated, and also significant variance cross observed individuals. Future analysis is to extend the model so that it can utilize RP data to calculate VoT, VoR and RR over different trip purposes, travel distance and for other analysis practices.