Summary:

Measuring the inaccurate: Causes and consequences of train delays

TØI Report 1459/2015
Author(s): Askill Harkjerr Halse, Vegard Østli and Marit Killi
Oslo 2015, 71 pages Norwegian language

In this report, we argue that the rich available data on train performance and railway infrastructure should be used to get precise measurements of economic relationships in railway management. As one such exercise, we first show how temporary speed reductions on railway links caused by low infrastructure quality affects running time and delays for Norwegian freight trains. Even though each speed reduction only adds about 44-50 seconds to running time, speed reductions still contribute to delay at the destination. Secondly, we show that delays has a negative effect on demand for passenger and freight trains services. The corresponding demand elasticity is lower than the one implied by willingness-to-pay studies, consistent with evidence from Great Britain.

In is widely acknowledged in the transportation economics literature that more reliable transport time constitutes an economic benefit. In the presence of unreliability, individuals and firms adjust by taking costly measures like departing early or keeping a safety stock of goods. The 'cost' of train delays is therefore the foregone benefits that could have been achieved if all trains were running on time.

Much of the existing literature on railway punctuality is based on optimization and/or simulation, calling for more empirical studies. In the innovation project PRESIS, funded by the Research Council of Norway and the Norwegian National Rail Administration, we have developed methods to survey reliability in the Norwegian rail sector. The tools consists both of automated descriptive analysis carried out in close to real time, and more detailed analysis of relations which are of particular interest.

This report documents two examples of the latter type of analysis. One concerns the effect of speed reductions on running time and reliability, while the other concerns the effect of reliability on demand for railway services. Both are based on econometric methods for analysis of panel data, where we follow the same observational units over many time periods. Using linear regression models with fixed effects, we are able to credibly control for unobserved heterogeneity which is constant over time or across observational units.

Just like users of the transportation system adapt to unreliability when planning their trips and shipments, the infrastructure manager and train operators take the underlying uncertainty in the transport system into account when deciding on timetables. All timetables have some degree of slack which ensures against large consequences of small deviations. This means that effects of measures or incidences on a railway link might not translate into changes in arrival reliability at the destination. Our study of the effect of speed reductions is one way to address the relationship between effects on the link and at destination.

The impact of speed reductions on freight train reliability

Large parts of the Norwegian railway network outside the bigger cities is only operated by freight trains and interregional passenger trains with low service frequency. On these railway lines, there are often temporary speed reductions on certain links due to problems with the infrastructure. Some speed reductions only last for a few days, while others can be in effect for many months or even years. Train timetables are normally not adjusted as a result of speed reductions, since they are assumed to have little impact on arrival time at the destination.

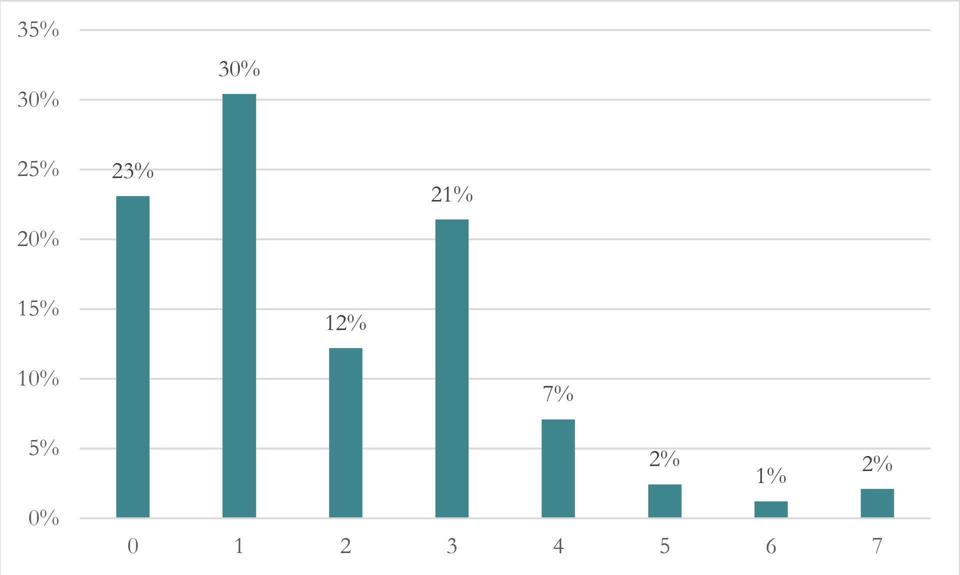

Figure 1. Freight trains on the Oslo-Trondheim line in 2012 by the number of speed reductions the train had to pass between Eidsvoll and Trondheim

We analyze the impact of speed reductions both on the link level and at destination for freight trains. To identify the effect on the link level, we utilize a dataset on running times and speed reductions on the Oslo-Trondheim line ('Dovrebanen') in 2012. Speed reductions are recorded each day and running times for each train is recorded at the link level, allowing to estimate the effect with high precision. Almost all speed reductions are of length 400 meters or less, implying that most of the effect is due to trains having to break and then accelerate again.

To investigate whether speed reductions result in later arrival at the destination, we use data covering the years 2008-2013 and study four of the main railway lines: Oslo-Bergen, Oslo-Stavanger, Oslo-Trondheim and Oslo-Åndalsnes. This data only includes speed reductions recorded each week, and there are likely to be many other factors which explain changes in arrival time. However, by using varying combinations of controls for time, season, time of day and train service, we can evaluate to what extent the estimated effect of speed reductions is sensitive to controlling for other factors.

On the link level, we find a highly statistically significant effect of speed reductions. One speed reduction leads to about 44-50 seconds longer running time. We also find a significant positive effect at destination for all four railway lines. This implies that on average, trains are not able to compensate for the time lost at speed reductions by driving faster or having shorter stops when waiting for crossing or overtaking trains.

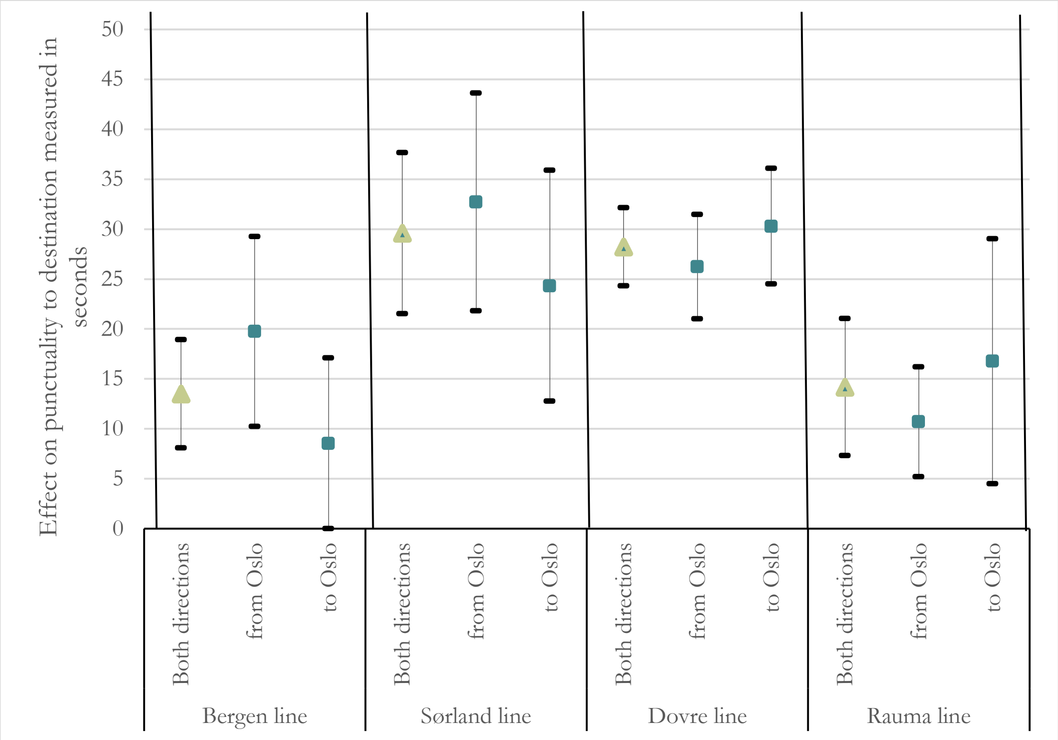

Figure 2. Effect of one more speed reduction on running time from origin to destination terminal. Point estimates and confidence intervals for each freight train line.

Since a large part of the network is single-track, waiting for meeting or overtaking trains is likely to be of high importance for arrival reliability. Some trains have to wait longer than planned for trains meeting or overtaking, while others wait less. Often, the management of crossings depends on to what extent the train is delayed or ahead of schedule already, and many trains do not depart exactly at their scheduled departure time.

As a result of this, we find that there is large heterogeneity across freight trains in running times and arrival delay. We are not able to explain the variation in running times by speed reductions to a large extent. However, there is a small significant impact of speed reductions on the probability of being more than 15 minutes delayed. This indicates that some trains incur long delays as a result of the interaction between speed reductions and meeting and overtakings.

How reliability affects demand for railway services

As shown in the literature, the effect of train reliability on demand for railway services can be estimated both directly and indirectly. The indirect approach involves combining willingness-to-pay for reliability with a known demand elasticity, for instance the price elasticity. The direct approach is to estimate the relationship between reliability and demand using market data.

We estimate the effect directly using market data for both passenger and freight trains. The passenger train data covers interregional trains in the greater Oslo area (not including the inner area) for 2010-2013. To obtain a data set suitable for analysis, we have combined the monthly number of period tickets sold for each of the 444 origin-destination station pairs with reliability data for the trains serving the station pair. In the case of freight trains, we have train-level data on freight volume and reliability for 28 terminal pairs on five railway lines. Both data sets are aggregated to monthly observations.

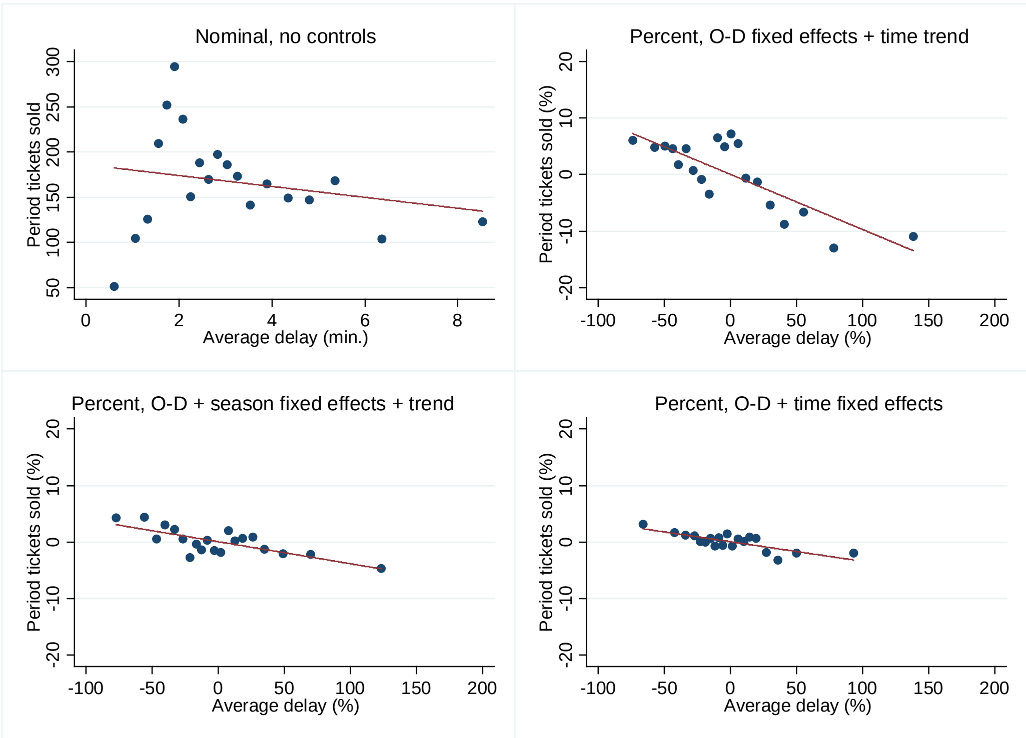

Figure 3. The relationship between reliability on the return trip and the number of period tickets sold on passenger trains for station pairs in the greater Oslo area. (Each point corresponds to about 1000 observations.)

For passenger trains, we find a quite robust negative effect of average delay on demand when considering reliability on return trips only, or when studying morning and return trips jointly. The demand elasticity with respect to average delay is -0,04, which is about half of the elasticity one gets when using the indirect approach based on Norwegian willingness-to-pay studies. This is consistent with the findings by Wardman and Batley (2014) for Great Britain.

The estimated elasticity for freight trains is somewhat less precise, and the results indicate that the share of long delays is a better measure of reliability than average delay. In any case, the results show that the effect on demand seems to be somewhat weaker for freight trains, with an elasticity with respect to average delay of about -0,01 to -0,02. If we want to compare this with the elasticity based on the indirect approach, we have to assume a freight fare. In practice, this varies between contracts. With reasonable assumptions about the fare, it seems that the implied elasticity is somewhat higher than the estimated one also for freight.

An interesting question is whether customers react both to `regular' variation in reliability as well as major disruptions and large delays. Our results for passenger trains do not seem to be driven only by extreme observations, and we find no descriptive evidence than demand has dropped substantially during or after winters with many long delays. In the case of freight trains, it seems more likely that large events like blocked lines due to landslides and flooding play an important role. However, this is difficult to spot in the data.