Predicting Travel Time Variability for Cost-Benefit-Analysis[1]

Stefanie Peer,
Department for Spatial Economics, VU University, Amsterdam, The Netherlands, speer@feweb.vu.nl
Carl Koopmans,
Department for Spatial Economics, VU University, Amsterdam, The Netherlands, ckoopmans@feweb.vu.nl
Erik Verhoef,
Department for Spatial Economics, VU University, Amsterdam, The Netherlands, everhoef@feweb.vu.nl

Abstract

The goal of this paper is to develop an econometric model that can be used to predict travel time variability for cost-benefit-analysis (CBA). The model explains travel time variability by the size of (mean) delays, as well as by other time-variant and invariant road characteristics. The data set used for the analysis is based on travel time data of 146 Dutch highway links for the year 2008.

The resulting relationship between travel time variability and its explanatory variables can be used in transport-related cost-benefit-analyses. Up till now, most CBAs do not account for travel time variability, or only in very rough ways. One reason is the lack of straightforward ways on how to include travel time variability. While forecasts of mean travel times can be derived from traffic assignment models, this is usually not the case for travel time variability.

The regression analysis shows that indeed a very significant relationship between travel time variability and delay can be found. It is also fairly robust against changes in the underlying dataset, for instance in respect to days of the week, months or weather conditions. Explanatory variables other than delay and flow-capacity-ratio contribute only little to the predictive power of the model.

Draft 10/2009

An earlier version of this paper has been presented at the ERSA conference Lodz, Poland (August 2009).

[1] The authors wish to thank Peter Schout and Henk Taale (both DVS, Delft) for their support in the data collection.

1. Introduction

The goal of this paper is to develop an econometric model that predicts travel time variability. The emphasis is hereby on applicability in cost-benefit-analyses (CBA). The results presented in this paper shall give insights in the empirical relationship between travel time variability and mean travel times as well as other variables, such as road characteristics. The analysis is based on travel time data of 146 highway links in the Netherlands.

So far, only little research has been done in the field of explaining and forecasting travel time variability. Unreliable travel times and costs related to them have attracted the attention of researchers and CBA practitioners only in recent years. Due to the lack of knowledge on how to predict and value travel time variability, the costs related to unreliable travel times are not yet included into most transport-related CBAs. In cases when they are taken into account, mostly rather imprecise rules of thumb are used[2]. Fosgerau and Karlström (2007) as well as Eliasson (2009) point out the importance of including these costs into CBA, showing that they can be quite substantial. In these papers they are citing the amount of 10-15% of the costs associated with travel time gains or losses, respectively.

A paper that is similar in its focus as this one has been written by Eliasson (2006). He also uses an econometric model and finds a non-linear relationship between the relative standard deviation of travel time (standard deviation divided by travel time) and the relative increase in travel time (travel time divided by free-flow time) on urban roads. Interestingly, he shows that these two variables are not positively related for all congestion levels. If congestion is very severe, relative standard deviation is a decreasing function of the relative increase in travel time. Other explanatory variables are not included in the model, as their contribution in increasing the explanatory power of the model was only minor. Also Gilliam et al. (2008) research travel time variability in an urban context by means of regression models. They relate a coefficient of variation of travel times (defined as standard deviation divided by mean travel time) and a congestion index (actual travel time divided by free flow time) to each other, using GPS data. Not only do they find a positive relation between these two factors but also between the coefficient of variation and the length of the road stretch. Li (2005) uses individual car data on urban toll roads to investigate vehicle-to-vehicle variability (travel time variability between drivers who depart within the same time slot) and variability between different time slots. The latter he again decomposes into variability induced by demand and supply related factors. He finds that during morning peaks demand related factors play a major role, whereas in the afternoon peak supply related factors become more important As intended in this paper, Kouwenhoven et al. (2005) also use regression analysis as a mean to predict travel time variability. They find that variability is strongly correlated to mean speed but do not find a correlation between variability and exogenous factors such as incidents or weather. However, the latter shows to influence mean speed resulting in an indirect effect on travel time variability.

Unlike the papers mentioned up to here, Fosgerau (2008) does not use regression models to explain travel time variability. Instead he uses Jensen´s inequality to theoretically prove that whenever mean travel times are increasing between two time periods, the standard deviation of travel times increases as well during the same time. As this is also the case at the point where mean travel time peaks, a loop comes into being as soon as mean travel times decrease. Hence, Fosgerau provides a theoretical explanation for the looping phenomenon that is often observed in travel time distributions (e.g. Eliasson, 2006, Franklin, 2008).

[2] For instance, in the Netherlands a guideline for CBA has been issued that suggests to include additional costs for unreliable travel time equal to 25% of the costs that are due to changes in mean travel times (CPB)

Another line of research that is more technical in its origin is not to forecast travel time variability by regression models but rather to find ways of predicting it directly within traffic assignment models. Although in most cases traffic assignment models with the ability to forecast travel time variability have not become fully operational yet, considerable advancement can be seen in this area. (E.g. Li, 2009).

The results presented in later sections of this paper are empirical. Emphasis is not put on modeling structural relationships between single events (e.g. weather conditions, incidents) and congestion pattern but rather on the analysis of aggregate travel time distributions. This corresponds to the focus of CBAs, which usually are based on aggregate traffic conditions.

The paper is organized as follows. Section 2 gives a short introduction to the valuation of travel time variability. Section 3 provides an overview of the data. Section 4 covers the econometric analysis. Different setups are compared to each other and robustness as well as sensitivity of the results is investigated. In section 5, implications for cost-benefit analysis are derived. Finally, section 6 offers some concluding remarks as well as suggestions for further research.

2. Valuation of Travel Time Variability

Since the goal of this paper is not only to predict travel time variability but also to use these predictions to calculate costs and benefits resulting from changes in travel time variability, the valuation of travel time variability plays a relevant role.

The empirical analysis in this paper uses the standard deviation of travel time distributions to represent travel time variablity. This is consistent with the prevailing theories on the valuation of travel time variability, which are also based on the standard deviation of travel times. In the classical mean-variance approach of valuing travel time, the disutility U(t) of the driver departing at time t depends on the standard deviation (σ) and the mean (μ) in a linear way:

$$U(t) = VOT * \mu + \beta * \sigma$$

where α is the value of time and β the value of (un)reliability. Corresponding to this theory, many studies have derived empirical values for α and β (e.g. by means of stated-choice experiments or expert opinions[3]).

Fosgerau and Karlström (2007) show that the mean-variance approach can be combined with the scheduling model. The latter model is based on the rationale that the utility of drivers depends on how much they arrive early (schedule delay early: SDE) or late (schedule delay late: SDL) in respect to their preferred arrival time:

$$U(t) = VOT * \mu + VSDE * SDE + VSDL * SDL$$

Here, μ is again the mean travel time and VOT the value of time, whereas VSDE is the value of SDE and VSDL the value of SDL. It is commonly assumed and empirically proven that VSDE<VOT<VSDL. Fosgerau and Karlström find that under the assumption of a constant (standardized) travel time distribution, the utility function is still linear in the standard deviation of the travel time distribution. It has been criticized by Van Lint et al. (2008) that measuring travel time variability in terms of the standard deviation disregards the fact that most travel distributions are left-skewed, indicating a large probability of long

[3] E.g. in the Netherlands: De Jong et al. Preliminary Monetary Values for the Reliability of Travel Times

delays (compared to the probability of travel times much shorter than average travel times). However the valuation approach developed by Fosgerau and Karlström (2007) indirectly accounts for the skewness of the distribution by including mean delay (H). This is the average delay a driver faces once he is late, which is determined by the cumulative (standardized) travel time distribution Φ (which depends on the skewness) and the optimal probability of being late (p).

In practice, these valuation methods are hardly used. Reasons are the lack of appropriate data as well as constraints in time and resources to adequately use these approaches. As alternative, costs of travel time variability are sometimes expressed as a percentage of the costs of travel time gains or losses. Hence: $C_{TTV} = C_{TT}(\mu)$, where C_{TTV} indicates the costs of travel time variability and C_{TT} the costs of travel time. The latter are a function of mean travel time μ. Such a setup gives the possibility to express C_{TTV} in terms of mean travel times: $C_{TTV} = f(\mu)$. This approach is reasonable as long as there are no factors (e.g. traffic management measures) whose influence on travel time variability and on mean travel time differ considerably in size and (or) direction. For instance, for factors, which have no influence on travel time variability but only on mean travel times, changes in the costs of travel time variability are assumed although they do not come into being in reality.

The focus of this paper is not to derive the exact effects of such factors on mean travel time and travel time variability, but rather to investigate whether these factors add any explanatory power to the model. If they do, this can be taken as an indication that it is not always sufficient to base the predictions of travel time variability on mean travel times only.

3. Data and Data Adjustments

Since the focus of this paper lies on predicting travel time variability in a way that is useful for CBAs, only such variables are included for which cost-benefit analysts usually have data available and which are also useful to include in CBA. In order to compute benefits and costs of transport-related projects, CBA requires forecasts of changes that occur in case of project implementation. For travel time data these forecasts are usually derived from traffic assignment models. Although advancements have been made in recent years that make model simulations more flexible and realistic, estimates of travel times are often still obtained at rather aggregate levels as output from these models. Predictions are therefore not made at an hourly or even more disaggregated level. Instead, they represent several hours or even entire days of traffic conditions. This restrains us from using time-lagged variables at a very small temporal scale. Instead we approximate time patterns by using dummies for the morning peak as well as for the rising and declining segments of the peak periods.

The econometric model does not include variables that are too disaggregate to use for CBAs either. Examples are for instance different weekdays, months or weather conditions. However, since drivers expect travel time distributions not only to vary over the day but also according to these factors, we will test whether the relationship between travel time variability and mean travel time varies across these potential influence factors.

3.1. *Highway Data*

For the econometric analysis we use travel time distributions of 146 highway links in the Netherlands. The data are collected from loop detectors, which measure the speed of the vehicles passing. From these speed measurements travel times can be derived using the piecewise-linear-speed-based (PLSB) trajectory algorithm, which has been developed by Van Lint and Van der Zijpp (2003). It generates almost unbiased estimates of travel times with small residual variance. Using loop measurement data implies that the resulting data are not individual vehicle travel times but instead they indicate an average travel time at a specific departure time. Consequently, differences in travel time times between vehicles that depart at the same moment (vehicle-to-vehicle variability) are not taken into account in this analysis.

Travel time estimates are aggregated into 15 min intervals. Hence, for each link 96 travel time observations are available per day. As Eliasson (2006) argues, the variability is smaller, the shorter the time intervals chosen for aggregation are. The decision to use 15 minute intervals seems reasonable as most likely drivers are not aware of differences in travel time distributions at a more disaggregate level (e.g. at minute-to-minute level).

We collect the travel time data for the entire year 2008, however, only taking into account working days (255 days). Since underlying demand patterns differ considerably between working days and non-working days, they should be analyzed separately. For instance, it probably does not come at a surprise to most drivers that almost all highways exhibit less congestion at 9.00 a.m. during weekends than they do at the same time during working days. Thus, including both weekends as well as working days into the analysis would most likely increase the standard deviation, however, not for reasons that are unknown to drivers (which is not the type of variability this research focuses on). As mentioned above, we will also test whether other days need to be excluded from the sample due to day-of-the week, month or weather effects.

Figure 1 shows a map of the Netherlands, indicating all road stretches included in the analysis. Most of them are concentrated in the West of the country, which is one of the most densely populated areas in Europe, known as the Randstad. It includes mayor cities such as Amsterdam, Rotterdam, Utrecht and The Hague.

FIGURE 1 ABOUT HERE

These road stretches have been selected due to the high density of loop detectors installed, which allows us to gather travel time data of very good quality. The length of the road stretches has been chosen in a way such that each link starts and ends after a highway intersection[4]. Thus, each link comprises one intersection point at its end. The link length varies between 2.2 km and 37.1 km, with an average of 13.2 km. To define road stretches in this way is motivated by the idea that for drivers these intersection points are natural points of reference since they frequently are bottlenecks. Often they are also distinct landmarks and hence easy to remember for drivers. In addition, most journeys include at least one highway intersection point, since people usually do not enter a highway just to drive very short distances. For these reasons, it is likely that many drivers base their departure time choice and maybe even route and mode choices on the expected delay and travel time variability in front of the intersection point. Hence, in contrast to concepts of defining the link length according to the distance between ramps or such that all links are of equal length, the concept employed here reflects the intuition of drivers in highway networks: Road stretches comprise the dynamics going on in front of intersections. Defining road stretches in this way has the disadvantage that causes of travel time variability are not uniformly distributed in case. Hence, the results of the analysis should not be directly applied to predict variability in a network.

[4]www.autosnelwegen.nl

Kouwenhoven et al. (2005) who perform a similar analysis also using different road links of different length, find that their model fit improves if they scale all road stretches to the same length of 20 km. Since our analysis should be applicable in CBAs, which usually deal with a wide range of different link lengths, we do not follow this procedure. Instead, we include length as explanatory variable in order to control for possible effects.

3.2. Travel Time Distributions

On the basis of the travel time data of the highway links, travel time distributions (consisting of one data points per day in the sample) for each quarter of an hour per link can be derived. Thus, for each link there are 96 travel time distributions in the dataset. We then use the first moment of the distribution (mean) as an explanatory variable of the second moment of the distribution (standard deviation[5], *'Stdev'*). Mean travel time is standardized by subtracting free flow time, which is defined as the lowest of the 96 observations for mean travel times for each link. Negative delays are therefore not possible. Unlike other authors, we choose to subtract free flow time from the mean travel times instead of dividing them by each other (e.g. Eliasson, 2006). Hence, the calculations here do not yield a percentage value of delay in respect to free flow time but a delay in minutes. At first sight one might argue that 1-minute-delay on a short road stretch is not comparable to a 1-minute-delay on a long road stretch, since people will perceive it differently. This is indeed worth to investigate. However, for now we assume that these delays cause equal costs to the driver, namely, arriving one minute later than without delay.

Taking a closer look on the resulting dataset, we find that for many observations, delays are rather short. If these delays are sufficiently short, they do not cause costs to drivers and therefore should not enter the analysis, which should ultimately be used to determine the costs of travel time variability in CBA. We find that many of these short delays are observations from the nighttime. It is likely that travel time variability during night hours is not caused by delays due to congestion but rather by day-to-day differences in the composition of the vehicles that are observed. Clearly this type of variability does not result in any costs to the drivers and is therefore not useful to include into CBA. Therefore, we decide to exclude all observations where the delay per km is lower than 0.1 minutes. Figure 2 shows that this causes a major part of the observations not to be used any longer (left of the black line). This causes the remaining observations to be more consistent in the causes of travel time variability.

FIGURE 2 ABOUT HERE

The remaining observations are aggregated over an entire year, hence, across different week days, months and weather conditions. These attributes are frequently known to drivers and therefore, drivers can adjust their departure time choice accordingly. Research on travel time variability does not focus on these recurrent traffic patterns, but on non-recurrent ones. We will investigate if the results of the empirical analysis in this paper are robust against these factors.

[5] Standard deviation (of travel times) and travel time variability will be used as synonyms in the rest of the paper.

3.3. Flow-capacity-ratio[6]('fcr')

A second important characteristic of traffic conditions is the flow-capacity-ratio, measured in the number of vehicles passing on one lane per hour. Hence, the variable used in the empirical analysis is equal to the ratio of the average flow (per 15 minute interval) divided by the number of lanes (also including lanes that are only open when there are more than 1500 drivers per hour per lane). Clearly, the flow-capacity-ratio does not only differ between road links but also over time. A positive coefficient of this variable can be expected, indicating that *ceteris paribus* a road with higher volume-capacity ratio is likely to be subject to a higher level of travel time variability. The reasoning is that road links have different maximum capacities of flow per lane. This can also be observed from Figure 3, where the red lines indicate mean flow observations for three different road links. One can easily see that for a given delay, the flow-capacity-ratio differs considerably across the sample and also that in general the curves corresponding to different road links begin to rise steeply at different levels of delay (indicating maximum capacity of flow has been reached). Reasons might for instance be differences in speed limits or road geometry. Roads with a high maximum capacity are then expected to exhibit higher travel time variability due to an increased risk of incidents as well as a larger impact of incidents.

FIGURE 3 ABOUT HERE

4. Analysis

This section consists of three main parts. First, different models including delay and flow-capacity-ratio as explanatory variables are estimated. Second, some remarks are made about the robustness of the results using samples from which travel time observations with specific properties are eliminated. Finally, models including additional explanatory variables are estimated and discussed.

4.1. Main Model

The full data set comprises 146 road stretches and has a time dimension of 96 quarters of an hour. As mentioned before, observations with a delay/km of less than 0.1 minutes are not considered in the estimations. Of the remaining observations, a random sample of 80% of observations is taken and the remaining 20% are kept as holdout sample to use for prediction. This enables us to compare models in respect to the difference between the fitted and the actual values of the standard deviation of travel times (the regarding formulas for the root-mean-squared-error (RMSE) and the bias can be found in Table 1).

TABLE 1 ABOUT HERE

After implementing several different setups, we found that a nonlinear multiplicative model of the following structure performs best in terms of predictive power:

$$Stdev = \beta_1 * Delay^{\beta_2} * VCR^{\beta_3}$$

[6] Data from www.autosnelwegen.nl: A weighted average of the number of lanes is used if it is not constant over the link

Besides the model associated with this formula, a model taking into account delay as only explanatory variable is estimated. Also a linear model that only takes into account delay is estimated. These alternative setups should give an indication how much predictive power is lost if simpler model formulations are used. Table 2 shows the results of these 3 regressions:

TABLE 2 ABOUT HERE

As expected, we find in all three regressions a positive relationship between delay and travel time variability. From the non-linear setups it becomes clear that the marginal effects of delay on travel time variability are decreasing in delay ($\beta_2 < 1$). The same is true for the flow-capacity-ratio (fcr). It is positively related to the standard deviation, exhibiting decreasing marginal effects.

Comparing the two non-linear regressions with each other, we indeed find that adding the volume-capacity ratio decreases the mean-squared error considerably. As expected, the linear model is worse in predictive power compared to the non-linear models. The same is true for the explanatory power, as the R-squared falls from more than 90% to about 70% when switching from the non-linear models to linear ones.

By comparing the actual to the fitted values of travel time variability in the holdout sample, we can also derive the bias. It is positive for all three regressions indicating that actual values are higher than fitted values. Hence, predictions based on these model results will tend to be conservative. The bias ranges from 0.06 (linear model) to about 0.1 (non-linear models). Comparing it to the RMSE of 1.7 and 1.4, respectively, the size of the bias is not negligible, however, also not very high. It might result from variables omitted in these regressions, as it becomes smaller when more explanatory variables are included (as demonstrated in later parts of the paper).

4.2. Robustness of Results

In this section, we will test whether the above results change significantly if travel time observations with specific properties that are potentially related to recurrent traffic patterns are eliminated from the estimation sample. Examples are specific weather condition, days of the week or months. If there are significant differences, it is important to assure that the travel time distribution on whose basis predictions of travel time variability are derived corresponds to the distribution on whose basis the valuation parameter of travel time variability have been derived. In most cases, this means to look at the concept of variability that underlies the stated-preference-survey that is used to determine the costs of unreliable travel times. Often in such surveys respondents are asked to imagine travel time variability to be due to different traffic patterns across week days, months or weather conditions. In this case, it is reasonable that also the “volume” of unreliability shall be based on the entire travel time distribution across the year, without taking any selection procedures.

The following “properties potentially related to recurrent traffic patterns” have been selected:

1. *Adverse weather conditions*[7](Days with a precipitation duration of more than 2 hours): We expect higher variability for days with bad weather conditions and, hence, lower overall variability if the observations of days with adverse weather are taken out of the sample.

2. *Day of the week:* Removing Fridays from the sample is expected to lead to lower overall variability, since Fridays often exhibit different congestion patterns than other weekdays.

3. *Months:*

    a. *Winter months: December, January*: In these months, the weather is usually more adverse to driving (low temperatures, high precipitation) compared to the rest of the year. Due to these weather conditions there might also be an increased demand for car use (as a substitute for public transport or cycling).

    b. *Summer months: July, August:* In the summer months, most road links show lower demand levels as a large share of workers is on vacation and therefore not commuting.

    c. *Summer and winter months:* All four months are removed from the sample.

We then test hypotheses of the following form:

Stdev (delay=d|all data) = Stdev (delay=d|data without "properties potentially related to recurrent traffic patterns"),

where *Stdev* is the standard deviation of the travel time distribution, and *d* any constant value of delay. Thus, the hypothesis is that for a given value of delay d predicted travel time variability is equal regardless of the underlying data-set. The model to be estimated is then:

$$StDev = D_1(\beta_{11} * Delay^{\beta_{21}} * VCR^{\beta_{31}}) + D_2(\beta_{12} * Delay^{\beta_{22}} * VCR^{\beta_{32}})$$

Here D_i (for i=1, 2) indicates a dummy equal to 1 for observations of group i and 0 otherwise. One of the groups is associated with the travel time distribution based on all observations, whereas the second group is associated with travel time distributions based on observations that do not have properties that are potentially linked to recurrent traffic conditions. Therefore, we can test whether the corresponding coefficients of these two groups are significantly different from each other. This can be done by means of t-tests: $\beta_{11} = \beta_{12}, \beta_{21} = \beta_{22}$ and $\beta_{31} = \beta_{32}$.

TABLE 3 ABOUT HERE

From table 3 we see that only few of the coefficients are significantly different from each other (grey shading). All of the differences significant at the 5% level (rain duration; summer months; summer and winter months) can be attributed to the exponent of delay (β_2). For the rain duration, the difference in the coefficients is positive indicating a higher variability in the case that the observations from days with more than two hours of rain are removed from the sample. The opposite is true for the months.

[7]Data from the Koninklijk Nederlands Meteorologisch Instituut (KNMI). For each road link the closest weather station among the stations in Schiphol, De Bilt, Rotterdam and Gilze-Rijen has been determined. Weather data are accumulated on a day-to-day basis.

Removing the months with specific supply (winter) and demand (summer) conditions leads to a lower overall variability.

Looking at the differences in β_1, one can see that all of them are negative, indicating that the removal of data tends to decrease predicted travel time variability (although this effect might be counterbalanced by a positive difference in the exponentials on delay and flow-capacity-ratio). However, in general differences seem not to be too large and for most cases it is seems to be acceptable to use the full distribution of travel time observations. This is especially true if also the corresponding valuation parameters of travel time variability do not rule out variability derived from recurrent traffic patterns.

4.3. *Sensitivity of the results in respect to other explanatory factors*

Travel time variability might not only be influenced by delays and the flow-capacity-ratio but also by other factors. In this section the results of regressions that do not only take into account delay and the flow-capacity-ratio, but also additional explanatory variables are presented and analyzed.

This section does not aim at explaining travel time variability by drawing precise conclusions on causal relationships between travel time variability and other explanatory variables. It rather focuses on if the addition of other explanatory variables improves the predictive power of the model and to provide an intuition on whether it is sufficient to use models of the shape $C_{TTV} = f(\mu)$ for calculating the costs of travel time variability.

As the goal of this paper is to derive a model to *predict* travel time variability for CBA, it has been chosen to use a rather aggregate dataset, with travel time distributions formed from observations over an entire year. Hence, causal effects at a temporal and spatial level are not taken into account (e.g. incidents and their impacts on traffic conditions in other parts of the road network). This renders it a very difficult task to draw causal conditions at an aggregate level, as there is no possibility to eliminate endogeneity issues from the data set. Instead, we can only check for correlations between variability and other explanatory variables.

The following variables are taken into account in the analysis:

Time-invariant variables

- *Length:* We expect the standard deviation of travel times to increase with the length of the road stretch. The longer the link, the more potential locations at which congestion can form (for instance after an incident). Nevertheless, we do not expect travel time variability to double if distance doubles since there are counterbalancing effects ("averaging out"). This means that on parts of a link travel time is faster and on other ones slower than expected. For instance, an incident might induce congestion before the incident location but almost free-flow time after the incident location.
- *Number of ramps per km[8] ('rampskm'):* The number of ramps per km, on the one hand, can give an indication of the robustness of the transport network, i.e. which possibilities drivers have to avoid congestion. On the other hand, much of the congestion occurs close to ramps since drivers need to change lanes, which renders traffic flow unstable.

[8] Data from www.autosnelwegen.net

- *Freight[9]:* The impact of freight transport is included as percentage of freight traffic on a specific road stretch aggregated over the year. Thus, no differentiation according to the time of the day takes place. The percentages are also aggregated across the two driving directions of a highway. The hypothesis here is that higher freight percentages lead to higher standard deviations of travel times because differences in the speeds of vehicles render the traffic flow more unstable and therefore lead to higher standard deviations.
- *Speed limits[10]:* Lower speed limits might decrease standard deviations since traffic flow gets more uniform and the accident risk decreases (keeping everything else constant). However, a problem of correlation might exist here, leading to a positive relationship between the speed limit and variability since lower speed limits are often evident on links with a high probability of delays.
- *Variable Speed limits ('varspeed'):* Variable speed limits have recently been introduced on a large number of roads in the Netherlands. Via electronic signs along the highways, speed limits can be adjusted according to traffic and weather conditions. They are expected to decrease variability as they shall both prevent incidents from happening and lower the impacts of incidents on other vehicles. The existence of variable speed limits is taken into account in form of a dummy variable, since no sufficient data could be found on when exactly which speed limit is in place.
- *Shoulder used as driving lane[11]('shoulder'):* Although shoulders used as driving lanes might decrease mean travel times due to higher road capacity, we can expect them to increase travel time variability since they might worsen the incident risk as well as the incidents' impact on traffic. A dummy is included for all time periods for which the shoulder is used as driving lane[12].

Time-variant variables

- *Morning peak:* We investigate whether there are differences between the morning and evening peak periods. Therefore we include a dummy for all observations before 12:00. The morning peak might has higher travel time variability (at the same level of delay) than the evening peak for the reason that many drivers are less flexible and need to adhere to specific arrival times at work in the mornings. In the evenings in contrast, many drivers face less strict requirements on when to arrive home, inducing them to adjust their departure time and route choices to traffic conditions.
- *Maximum Delay ('maxpeakdelay')* : It is expected to make a difference whether 5 minutes delay just occur at the onset of the peak and delay rises further to 20 minutes at the top of the peak or whether this 5 minutes delay is already the maximum delay being observed at a road link. If an observation is part of a higher peak (compared to a lower one) in terms of maximum delay, we expect a higher standard deviation due to the severeness of the congestion pattern on this road. For each observation before 12:00 we include the highest delay between 7:00 and 12:00 as maximum delay and for each observation after 12:00 the highest delay between 16:00and 19:00 in order to check for this effect.
- *Loop-Effect ('loop'):* The variable is included to account for the loop effect analyzed by Fosgerau (2008). It is implemented as a dummy variable that is equal to 1 in case delay has been increasing in the quarter before the actual time period considered. The hypothesis is that travel time variability is

[9] Data from DVS, Delft
[10] Data from www.autosnelwegen.net: In the case of multiple speed limits on one link, a weighted average is used.
[11] Data from www.autosnelwegen.nl
[12] This is the case when the number of cars/hour/lane exceeds 1500 (Source:Rijkswaterstraat).

smaller in time periods when delays have been increasing during the last quarter of an hour rather than decreasing.

Omitted Variables

There are a considerable number of further explanatory variables that could have been included in the analysis. However, first, it is not useful for our purpose to include data that are in most cases not available or useful for cost-benefit analysis. Second, including more variables in the models might lead us to over-fit the model on the current database and make it less generalizable to other roads. Third, some variables have been included in the analysis but showed to be hardly significant. This is for instance the case for location specific variables such as the geographical position of a road in respect to major urban centers (i.e. circumferential, inbound, etc.). As hardly any effects on travel time variability were found, they were left out in the final analysis presented in the following section of the paper.

Results

We estimate a linear (additive) model here for practical reasons. Since the power of the delay is similar to 0.5 in previous calculations, we include the square root of the delay as explanatory variable. For all variables we include the variable itself as well as its interaction term with the delay (indicated by the first letter *D* in the regression table). Almost all variables (except for freight) show to be significant either themselves or as an interaction factor with delay or in both cases. We find that the effect of including these additional variables on the RMSE is rather small compared to the model estimated before. However, it should be kept in mind that these models were non-linear and the square-root included in the models estimated in this section of the paper is only a rough estimate of the coefficient of the exponent of delay.

TABLE 4 ABOUT HERE

Since all explanatory variables enter the model not only linearly but also in form of interaction or squared terms, interpretation of the size and sign of the overall influence of this variable on travel time variability can be done more easily by means of examples. Table 5 gives some intuition on the economic significance of the coefficients.

TABLE 5 ABOUT HERE

For each variable the following analysis is done: All variables except for the one analyzed are kept at their base level (approximately their average value in the sample). For each variable, one scenario is calculated under the assumption that it assumes a rather low value and another scenario when it assumes a rather high value (for dummy variables this is by nature 0 and 1, respectively). These two scenarios are then compared with each other and the percentage change is provided in Table 5. However, since all variables are also included as an interaction terms with delay, the above described analysis is done for three different levels of delay, namely 2, 5 and 10 minutes. Thus, the table allows one to compare if and how the marginal effect of each variable depends on the size of the delay.

We find that travel time variability is slightly increasing in the length of the road stretch for short delays. It comes at a surprise that the estimated relationship between variability and delay is very stable across different lengths of the road stretches. A graphical indication of this is given by Figure 4 showing the relationship between delay and travel time variability for different lengths of road stretches. Hardly any difference can be noticed in the distribution of scatter points among the three categories of road lengths.

FIGURE 4 ABOUT HERE

In respect to the number of ramps per km, we find that given a level of delay, variability is decreasing. On possible explanation is that robustness of the network is improved as a consequence of a higher number of ramps, as drivers have the possibility to exit the road in case of congestion. The coefficients of the percentage of freight traffic are not significant. This might be due to the aggregate nature of the data, which do not distinguish between different levels of freight traffic over the day.

Speed limits show to be negatively related to variability for short delays, and positively for longer ones. Variable speed limits are positively correlated with travel time variability for all levels of delay in the analysis. For the cases of positive relationships, endogeneity is likely to play a role. Lower speed limits are implemented usually on road stretches that are frequently subject to congestion. Evidence is also mixed for the use of the shoulder as driving lane. Variability increases in the use of the shoulder when delays are short, but decreases for longer delays. This might give an indication that at least for smaller delays the use of shoulders makes the road network less robust and more variable.

For the other explanatory variables, results are not very surprising. Morning peaks tend to exhibit stronger variability than evening peak. This is probably due to higher strictness of schedules for most drivers during the morning peak compared to the evening peak, where congestion can be more easily circumvented by departing earlier or later. Also the maximum delay of the corresponding peak is positively related to travel time variability. This makes sense since peaks that are in general more distinct are expected to show higher levels of variability. As discussed before, the maximum flow-capacity-ratio per lane varies across road stretches. Those with higher maximum capacity are then likely to show higher variability. Finally, for the loop effect also the expected sign is found. If delays are rising in time, travel time variability is lower than when they are decreasing.

Fairly similar results are obtained when only taking into account time-invariant variables (right column of table 4).

Concluding, we find that the additional predictive power added to the model by including this range of variables is rather small. However, most of these variables show to be related to travel time variability in a specific way. Research should be done at a more disaggregate level to find out more about these relationships as well as about the corresponding causalities.

5. Implications for CBA

This section provides some suggestions on what the results presented in the previous section imply for cost-benefit analysis. We will compare the current Dutch guideline on the valuation of travel time reliability to the results implied by the regression analysis.

For simplicity we use the nonlinear regression that does not take into account the flow-capacity-ratio (Table 2: 2nd estimation result). To determine costs, the scheduling is model is not employed here as the required coefficients have not yet been estimated for Dutch highways. Instead the mean-variance model is employed. A hypothetical VOT of 10c/minute and a VOR of 8c/minute are assumed (implying a reliability ratio= VOR/VOT=0.8, which is consistent with what has been found in previous research (e.g.

Rand Europe, 2005). For each point in the sample we can calculate the costs (C) a driver faces by the formula*: C= 10*Delay+8*Stdev.* The costs calculated this way (blue line in Figure 5) are compared to the costs yielded when applying the Dutch guideline on how the cost of unreliable travel times shall be included in CBA. This guideline suggests that the costs of travel time variability are equal to 25% of the costs associated with delays, regardless of the size of the delays (red line in Figure 5). It shows that the Dutch approach has a strong tendency towards underestimating the cost of travel time variability (at least for the chosen cost ratio).

FIGURE 5 ABOUT HERE

For CBA, levels of travel time variability do not matter as much as do changes. The non-linear shape of the relationship between delay and standard deviation has as a consequence the following: If delay is small and it is increased (by for instance 1 minute), this leads to higher absolute increase in travel time variability than if the initial delay had been larger. Figure 6 shows the marginal increase of travel time variability if delay increases for a delay range of 0.5 to 30 minutes. It shows that for very small delays, marginal increases are above 1, indicating that travel time variability rises faster than delay. The underlying assumption, of course is that the drivers' valuation of a decrease in variability is independent from the initial size of the delay.

FIGURE 6 ABOUT HERE

6. Conclusions

In the precedent analysis we showed that the econometric models used in this paper have a high explanatory as well as predictive power. They unanimously show that variability is closely related to delays and that the marginal effect of the delay on travel time variability decreases in the delay. The same is true for the flow-capacity-ratio. Comparing the results of the estimations to the current Dutch guideline on how to include travel time variability in CBA, we find that it widely underestimates the costs of travel time variability.

From the analysis on the robustness of the estimation results towards different underlying datasets, it seems that in general it is safe to use travel time distributions over an entire year without risking overestimating travel time variability considerably. From analyzing the sensitivity of the relationship between delay and variability to other explanatory variables we find that predictions of variability do not change a lot if these variables are added. However, as most of them are significant, this is an indication that travel time variability is not only determined by delay but also by other variables. Hence, expressing costs of travel time variability in terms of the costs associated with travel times is a useful, however, not generally valid approach.

It has been mentioned in the introduction that Eliasson(2009) as well as Fosgerau and Karlström (2007) come to the conclusion that the costs related to travel time variability are around 10-15% of the costs related to changes in mean travel times. Here we find that for small delays, the costs associated with unreliable travel times even exceed the costs related to mean travel times. One possible explanation is the analysis in this paper refers to highways, whereas Eliasson and Fosgerau/Karlström analyze travel time variability on urban roads. In urban networks travel time variability is already at a high level for no or very small delays (e.g. due to traffic lights). As a consequence, an increase in delay might hardly increase travel time variability further.

Due to the variation in the sample, the results of the analysis can be used to predict travel times on other Dutch highways and, with some caution, probably also to urban highways in other countries. In order to yield similar results for other road types, the analysis as in this paper can be done using an according data set.

When applying the results of the paper, it is important to remember that all road segments have been defined between to lie within two highway intersection points. This implies that for most road stretches in the sample, (potential) causes of variability are not uniformly distributed across but rather accumulated at the beginning and the end of the road stretches (hence, at the intersection points). This analysis is therefore not directly applicable to sub-segments of such road stretches that do not involve intersection points or only involve intersection points. The analysis in this paper refers to an average of these two cases.

The model developed in this paper is an improvement to existing rules-of-thumb on how to include variability in cost-benefit-analysis. However, more research has to be done to investigate relationships between variability and other factors (such as traffic management measures), which might have an impact on variability. Thus, less aggregate analyses need to be done taking into account spatial and temporal correlations among observations. An improved understanding of such underlying dynamics can then help to develop better prediction models.

However, econometric prediction models based on delays as explanatory variables will hopefully not be the last step in this area of research, as they require using mean travel times predicted by traffic assignment models as an input. However, travel time predictions generated by traffic assignment models have rather stringent underlying assumptions, for instance on the number of drivers using a highway. For obvious reasons, the number of drivers is based on forecasts not taking into account that more or less drivers might decide to use a road stretch depending on changes in travel time variability. It should therefore be an important goal of the research on travel time variability to develop models that allow predicting travel time variability in more elaborate ways rather than only on the basis of delay.

References

Eliasson, J. (2006). Forecasting Travel Time Variability. Working Paper.

Eliasson, J. (2009). A cost-benefit analysis of the Stockholm congestion charging system. Transportation Research Part A 43: 468-480

Fosgerau, M. (2008). On the relation between the mean and variance of delay in dynamic queues with random capacity and demand. MPRA Paper.

Fosgerau, M., Karström, A.(2007). The Value of Reliability. Working Paper.

Franklin, J.P., Karlström, A. (2008). Travel Time Variability for Stockholm Roadways: Modeling the Mean Lateness Factor. Working Paper.

Gilliam, C. (2008). Forecasting and appraising travel time variability in urban areas. Association for European Transport and contributors 2008

Kouwenhoven, M. et al. (2005). Development of a tool to assess the reliability of Dutch road networks. Association for European Transport and contributors 2005

Li, H. (2009). Reliability-based Dynamic Network Design with Stochastic Networks. Ph.d.Thesis. TRAIL Research School, Delft

Li,R. (2004). Examining travel time variability using AVI data. Paper presented at CAITR 2004.

RAND Europe (2005). The Value of Reliability in Transport. Provisional Values for the Netherlands based on expert opinion. Transport Research Center (AVV)

Van Lint, J.W.C, et al. (2008). Travel time unreliability on freeways: Why measures based on variance tell only half the story. Transportation Research Part A 42: 258-277

Van Lint, J.W.C., van der Zijpp, N.J.(2003). An Improved Travel-time Estimation Algorithm using Dual Loop Detectors. Paper presented at TRB 2003.

Figures and Tables

Figure 1: Sample Road Stretches

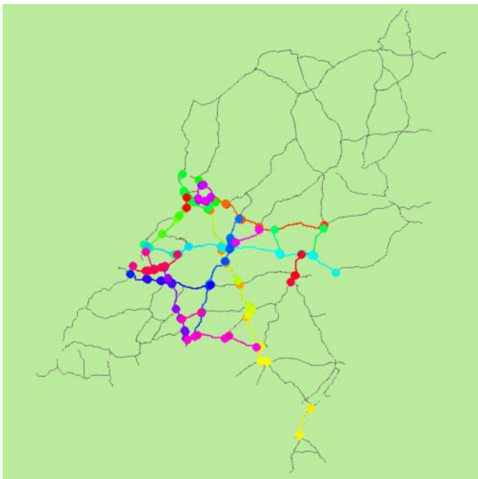

Figure 2: Distribution of Delays/km

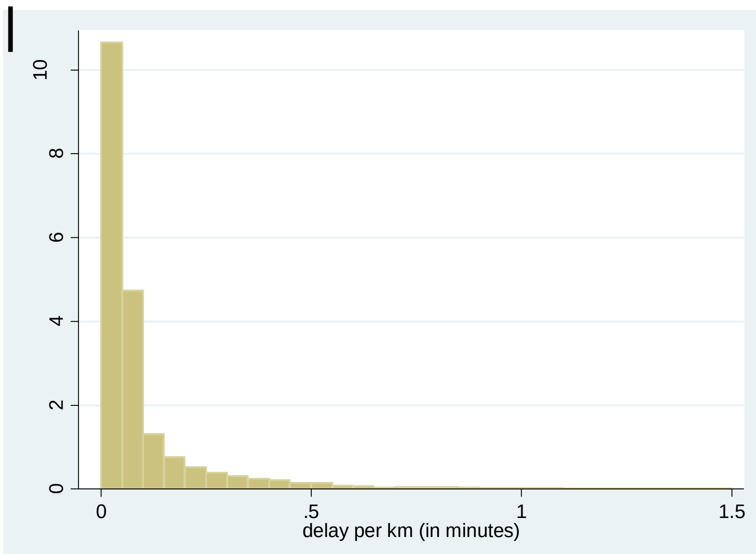

Figure 3: The relation between delay and the flow-capacity-ratio

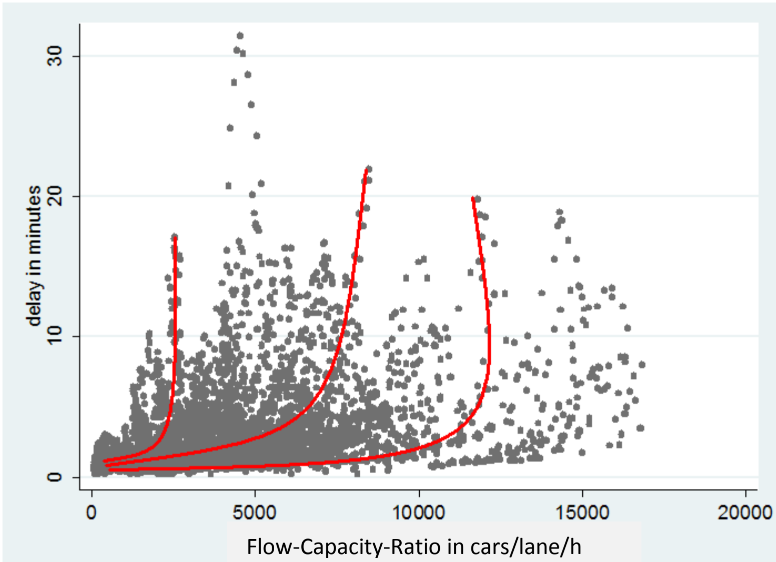

Table 1: Formulas for Prediction

Root Mean Squared Error (RMSE):

$$\sqrt{\frac{1}{N} * \left(\sum_i (StDev_i - StDev_i^f)\right)^2}$$

Bias:

$$\frac{1}{N} * \sum_i StDev_i - \frac{1}{N} * \sum_i StDev_i^f$$

StDev denotes the standard deviation of travel times with *f* indicating the fitted values. *N* is the number of observations.

Table 2: Main Regression Results

	NL (incl. FCR)		NL(without FCR)		Linear	
VARIABLES	*coef*	*se*	*coef*	*se*	*coef*	*se*
$\boldsymbol{\beta_1}$	0.5443***	(0.0375)	2.0460***	(0.0269)		
$\boldsymbol{\beta_2}$	0.5971***	(0.0067)	0.6631***	(0.0066)		
$\boldsymbol{\beta_3}$	0.1696***	(0.0084)				
delay					0.7763***	(0.0085)
constant					1.5173***	(0.0404)
Observations	3451		3451		3451	
R^2	0.9196		0.9076		0.7041	
R^2 adj.	0.9195		0.9075		0.7041	
RMSE (holdout.s.)	1.4393		1.5434		1.6888	
Bias (holdout.s.)	0.1003		0.1016		0.0633	

Table 3: Robustness Analysis: Comparing the model based on all observations to the model based only on observations that do NOT have certain properties.

Properties of the observations taken out of the sample	**Average number of observations left in the sample per traject**	**P-values of the t-test. The signs in the brackets indicates the direction of the coefficient change in respect to the model based on all observations.**		
		β_1	β_2	β_3
Rain Duration>2h	175	0.2348 (-)	0.0043 (+)	0.6786 (+)
Fridays	204	0.8877 (-)	0.2397 (+)	0.6150 (-)
July, August	211	0.5186 (-)	0.0014 (-)	0.4577 (+)
December, January	212	0.6987 (-)	0.0781 (-)	0.5770 (+)
July, Aug, Dec., Jan.	168	0.1837 (-)	0.0000 (-)	0.0914 (+)

Note: Shading indicates p-values <0.05

Table 4: Regressions with additional explanatory variables

	All variables		Time- invariant var.	
VARIABLES	*coef*	*se*	*coef*	*se*
delay	-0.8973***	(0.1202)	-1.0681***	(0.1219)
Root(delay)	3.3949***	(0.2272)	4.4426***	(0.1782)
length	0.0204**	(0.0083)	0.0116	(0.0091)
Dlength	-0.0044***	(0.0014)	-0.0023	(0.0015)
rampskm	-1.5140***	(0.1876)	-1.6044***	(0.2075)
Drampskm	0.0773	(0.0522)	0.1044*	(0.0560)
freight	0.0071	(0.0054)	0.0180***	(0.0060)
Dfreight	-0.0010	(0.0012)	-0.0026**	(0.0012)
speedlimit	-0.0357***	(0.0045)	-0.0388***	(0.0049)
Dspeedlimit	0.0073***	(0.0010)	0.0085***	(0.0010)
varspeed	-0.5150**	(0.2373)	-0.1656	(0.2618)
Dvarspeed	0.2232***	(0.0592)	0.0954	(0.0649)
shoulder	0.2622**	(0.1287)	0.2713*	(0.1416)
Dshoulder	-0.0617**	(0.0311)	-0.1084***	(0.0338)
morningpeak	-0.9629***	(0.0678)		
Dmorningpeak	0.3017***	(0.0156)		
maxpeakdelay	0.0213***	(0.0073)		
Dmaxpeakdelay	0.0057***	(0.0015)		
fcr	0.0002***	(0.0000)		
Dfcr	-0.0000***	(0.0000)		
loop	-0.2448***	(0.0644)		
Dloop	-0.0024	(0.0138)		
Constant	2.1531***	(0.4797)	1.5276***	(0.5088)
Observations	3451		3451	
R^2	0.8199		0.7744	
R^2 adj.	0.8187		0.7735	
RMSE (holdout.s.)	1.3846		`1.5036	
Bias (holdout.s.)	0.0739		0.0488	

Table 5: Sensitivity of travel time variability predictions

Variable	Units	Base Scenario	Values			Changes in Predictions Minutes of Delay		
			Small		*High*	*2*	*5*	*10*
length	km	10	7	→	15	1.55%	-0.23%	-1.91%
rampskm	number	0.2	0.1	→	0.3	-5.35%	-3.24%	-1.20%
freight	percentage	10	5	→	15	0.88%	0.31%	-0.24%
speedlimit	km/h	100	90	→	120	-8.61%	0.35%	9.37%
varspeed	dummy	0	0	→	1	3.31%	8.77%	13.95%
shoulder	dummy	0	0	→	1	1.65%	-0.68%	-2.88%
morningpeak	dummy	1	0	→	1	-1.22%	8.65%	20.03%
maxpeakdelay	minutes	10	7	→	15	6.75%	5.94%	5.19%
fcr	veh/h/lane	4000	2000	→	7000	23.43%	15.50%	8.40%
loop	km/h	1	0	→	1	-5.12%	-3.61%	-2.14%

*the grey shaded areas indicate increases in travel time variability

Figure 4: The influence of the length of road stretches on the relation between travel time variability and delay

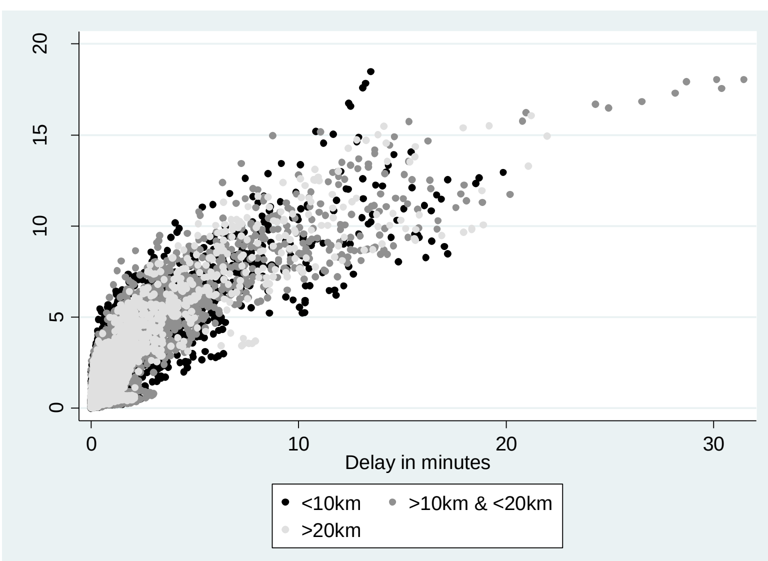

Figure 5: Cost comparison

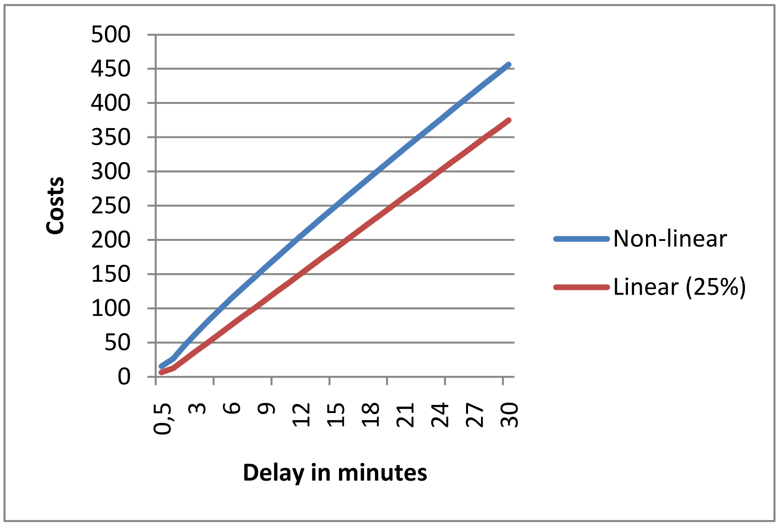

Figure 6: Marginal increases in variability depending on the delay