Predicting travel time variability for cost-benefit analysis

Katrine Hjorth, Thomas C. Jensen, Niels Framroze Møller
Department of Management Engineering, Technical University of Denmark

Short paper submitted to hEART 2019 (2935 words, excluding references)

Abstract

This paper describes an analysis of the relation between travel time variability and mean delay, based on GPS data of travel times on a large set of Danish roads. Our overall aim is to develop a model to predict the extent of travel time variability for road traffic, based on information about average travel time and delay, inspired by the approach applied in the Netherlands. Our results suggest that this is indeed possible, but that further analysis is needed before the method can be applied in practice.

1. Introduction

Travel time variability (TTV) measures the extent of unpredictability in travel times that travellers face when they plan their journey. It arises due to unpredictable day-to-day fluctuations in traffic demand, traffic incidents, transitory roadworks and weather conditions, affecting road capacity. TTV is costly for society, since the unpredictability causes people to either risk arriving later than desired or to depart earlier than optimal to avoid being late (potentially wasting their time).

As a cost for society, TTV should be included in cost-benefit analyses of transport and infrastructure projects, in order to evaluate potential gains of reducing TTV along with gains of reducing average travel times. This requires that we can predict TTV in different future scenarios. Usually we apply traffic models to forecast average travel times on the road and rail networks. However, in a Danish context, not even the newly developed and highly advanced National Traffic Model is yet able to predict the extent of TTV. Inspired by the Dutch forecasting approach (cf. Kouwenhoven & Warffemius, 2016), the current study analyses the relation between travel time variability and mean delay, with the overall aim of developing prediction methods that are sufficiently simple to be implemented on a countrywide basis and using existing data sources. Our study differs from Kouwenhoven & Warffemius (2016) in two ways: First, we measure TTV by the difference between the 90% and 50% quantiles of the normalised travel time distribution (travel time per kilometre) rather than the standard deviation. Second, we compute delays and TTV based on individual travel times, whereas they use aggregated travel times over 15 minute intervals.

Our results are promising, but also raise a number of methodological issues to be solved before applying the methods in practice. This short paper presents our main results, and a discussion of these methodological issues.

2. Method

Our goal is a simple prediction model that relies on output from a traffic model: This could be traffic flows (number of cars and passengers) or average travel times in the network.

Theoretically, it is preferable to model travel times and TTV as a function of traffic demand, which is the main determinant of congestion. The classic speed-flow curves simply model the relation between speed, i.e. inverse travel time, and observed flow. However, on congested roads these speed-flow curves fail to account for the following issues:

1. Dynamic effects and spillbacks: A high demand in one time period may affect travel times in several time periods afterwards, as queues at bottlenecks take time to dissolve. Hence, observations of travel time and traffic flow are not independent over time, but are related through a dynamic process. Moreover, capacity shortage and bottlenecks affect travel times not only locally, but also upstream, creating dynamic effects both in time and in space.
2. Endogeneity: Traffic demand affects observed traffic flows and travel times, but travel times may also affect the observed traffic flow, because congestion and queues cause fewer cars to pass a certain road segment within a given time period.

It is crucial to distinguish between demand and observed traffic flows here: During congestion, demand may exceed the road capacity, in which case the observed traffic flow only reflects part of the demand. Thus, demand is unobserved, and we only observe the realized flow, which may be endogenous, as it may depend on the travel time we wish to model (cf. Fosgerau & Small, 2012).

It is possible to build more advanced models based on queueing theory (see e.g. Hyder Consulting, 2010, or Overgaard et al, 2014), but this demands a quite detailed modelling of the road network, including locations of all bottlenecks, merges and diverges, together with observations of traffic flow on all road segments upstream of a bottleneck far enough to guarantee that we observe the entire queue. In Denmark, this would only be feasible for a very limited part of the network.

Fosgerau & Small (2012) and Hyder Consulting (2010) attempt to handle the issues with dynamic effects and endogeneity. Compared to our goal (application on a national level without modelling the individual characteristics of each specific road segment), the Hyder Consulting (2010) study is of a smaller scale and focus on modelling specific roads, where all relevant upstream flows can be observed. Fosgerau & Small (2012) model a section of the Danish motorway network and approximate the entry flow (demand) based on observed exit flows on upstream links (not including entry ramps), since they cannot observe all relevant flows. They instrument flow and traffic density using the flow two links upstream and the density two links downstream, to control for endogeneity. In an earlier research project, we have attempted to develop a simple model to predict TTV on a Danish highway based on traffic flow, controlling for endogeneity bias by using total morning traffic volume as an instrumental variable (DTU Transport, 2015); however, the resulting predictions were not satisfactory.

An alternative approach is to model TTV as a function of average delay, which is the difference between average travel time and free flow travel time (travel time in uncongested conditions). Such a model is not a causal relationship, but merely a relation between two features of the statistical distribution of travel times. In principle, we cannot exclude endogeneity issues here as well, but this has yet to be investigated. This approach has been adopted in the Netherlands (Kouwenhoven & Warffemius, 2016), and has the advantage that it does not rely on traffic counts, which are only available at a limited set of measuring points. The only traffic information used is information about travel times, which can be obtained from a variety of sources: Stationary loop detectors or Bluetooth or GPS trackings. Such a model can be applied more widely and potentially cover most of the road network.

To measure TTV, the Dutch approach (Kouwenhoven & Warffemius, 2016) uses the standard deviation of the travel time distribution, which has also been analysed in other studies (e.g. Eliasson 2006, Peer et al. 2012). They find that the resulting model is not very stable, as the removal of very few outliers can change the results considerably. For this reason, we have chosen to use the difference between the 90% and 50% quantiles of the normalised distribution (travel time per km), which is a more robust measure, as it is not affected by rare extreme events.

3. Data

We used a dataset with travel times measured from GPS-logs from a fleet of cars. The data were collected and processed by Hermes Traffic Intelligence for the Danish Road Directorate. We attempted to minimize potential bias stemming from spillback effects by defining road segments such that the traffic at the end points of a road segment is not affected by bottlenecks.

The dataset contains a set of roughly 1900 one-way road segments that are between 5 and 15 kilometres long and part of the main road network.[1] Segments are defined by sub-dividing roads (defined by administrative road number) into smaller units. This implies that while segments can contain intersections, they do not contain turns from one road to another.[2] The dataset contains different road types, but mostly road segments outside cities.

The raw dataset contains records at the trip level. It has information about segment id, segment entry time and segment travel time. We also have information about segment characteristics: Administrative road number, exact location, segment length, annual traffic volume, number of lanes and the speed limit. Hermes has tagged trucks as far as possible, based on the observed travel time patterns. The raw dataset contains roughly 8.7 million travel time observations from 2014 and 2015, but almost half of these stem from vehicles tagged as trucks, which we have removed from the analysis, since they are subject to other speed limits than other vehicles.

Very often, speed limits and the number of lanes vary within a segment. We have information about the values of these attributes and the share of segment length they represent. We computed the minimum number of lanes, the average number of lanes and the average speed limit for each segment.[3] If information about speed limit or lanes is missing for part of the segment, the measures are computed based on the available information.

We distinguish between three classes of roads, defined by the administrative road number:

- Motorways
- Other state-owned roads
- Municipality-owned roads

We note that the category "Motorways" include not only motorways, but also some so-called motor-traffic-ways, that typically have fewer lanes and lower speed limits. These are easy to identify, and so this does not invalidate the analysis.

We generated an estimation dataset by dividing the data into 19 time bands (18 1-hour intervals from 6AM to midnight, and a single time band for the period from midnight to 6AM), excluding observations

[1] The so-called "route numbered road network".

[2] Meaning that travel times are only measured for through-going traffic, and not for vehicles turning.

[3] We used a harmonic average, where outcomes are weighted not by the length share, but by the travel time share.

from weekends and holidays, computing normalised travel times in minutes per kilometre by dividing with segment length, and computing average (normalised) travel time, average (normalised) delay and (normalised) TTV for each combination of segment and time band. In the rest of the paper, we take travel time to mean the normalised travel time, unless we explicitly mention otherwise (and likewise for averages and quantiles).

The free-flow travel time is defined as the 5% quantile of travel times in a segment (across all time bands), and the mean delay in a given time band is computed as the difference between average travel time and the free-flow travel time. Note that this definition causes some observations of mean delay to be negative, as the average travel time in some time bands with only few observations may actually be less than the 5% quantile over the entire day. TTV is computed as the difference between the 90% and the 50% quantiles.

In our analysis, we omitted data points with negative mean delay (as we use a logarithmic model) and data points based on 20 or fewer trips. This leaves roughly 24,000 data points.

4. Analysis and results

4.1 Nonparametric analysis

We used nonparametric (local constant) regressions to estimate the relation between the logarithm of TTV and the logarithm of mean delay. We made separate regressions depending on the segment length, shown in Figures 1-3 below.

In general, we observe a clear positive relation between TTV and mean delay, with narrow 95% confidence bands except for high values of mean delay.[4] The regression curve tends to be upward bending or linear.

It seems there is some effect of segment length, despite the rather limited variation in the dataset. In uncongested conditions (low values of mean delay), we generally observed a lower TTV for a given mean delay on longer segments. Possibly, this is because some of the speed variation may be averaged out on longer segments compared to shorter segments. The differences are not always significant, and there are exceptions. In congested conditions, segments of 10-15km are different from the shorter segments, but the road types differ: For motorways, we see higher TTV values (for a given value of mean delay) for long segments; while we find the opposite for the other road types.

We also made an initial investigation of the effect of the number of lanes. Here, the results varied considerably between road types.

[4] We note that the standard errors are not adjusted to control for the fact that several data points stem from the same road segment.

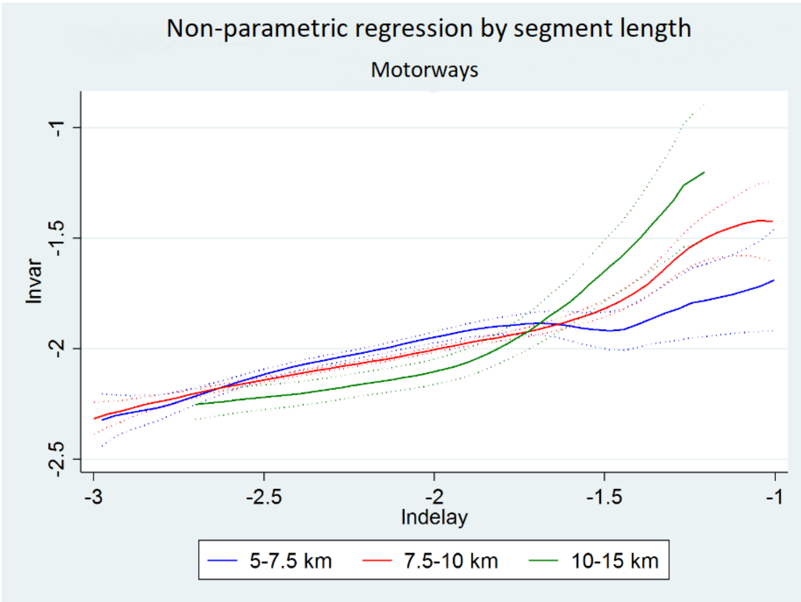

Figure 1: Local constant regression for motorways, by segment length. Regressions curves with 95% confidence intervals (dotted lines).

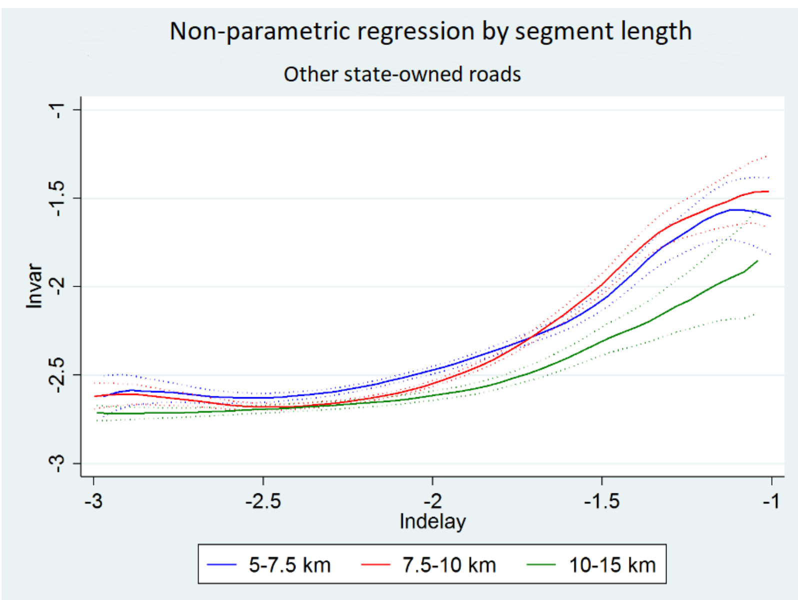

Figure 2: Local constant regression for other state-owned roads, by segment length. Regressions curves with 95% confidence intervals (dotted lines).

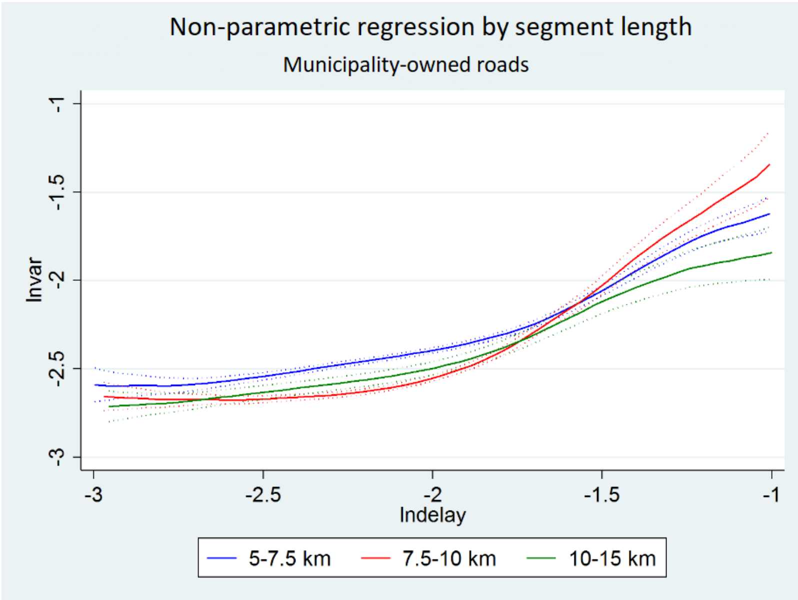

Figure 3: Local constant regression for municipality-owned roads, by segment length. Regressions curves with 95% confidence intervals (dotted lines).

4.2 Parametric analysis – motorway data

We estimated a parametric model of the relation between log TTV and log mean delay for motorways. We chose a second-order polynomial, which allows for the types of curvature observed in the nonparametric plots, and we allowed the shape of the polynomial to depend on segment length and the number of lanes.

We let i=1,…,N index the observations, and introduce the following notation:

- y_i: Travel time variability (milliseconds per kilometre),
- x_i: Mean delay (milliseconds per kilometre),
- z_i: Segment length (kilometres),
- w_i: Average number of lanes.

The regression model is:

$$ln y_i = \beta_0 + \beta_1 ln x_i + \beta_2 (ln x_i)^2 + \varepsilon_i \quad (1)$$

where

$$\beta_0 = \gamma_0 + \gamma_1 ln z_i + \gamma_2 w_i \quad (2a)$$

$$\beta_1 = \lambda_0 + \lambda_1 ln z_i + \lambda_2 w_i \quad (2b)$$

$$\beta_2 = \phi_0 + \phi_1 ln z_i + \phi_2 w_i \quad (2c)$$

and ε_i is an error term (assumed to be independent across observations). The logarithmic form implies that the estimated parameters on x_i, z_i and w_i do not depend on the chosen time unit (milliseconds or minutes): Only the constant term is affected. Inserting (2a)-(2c) into (1), we get

$$lny_i = \gamma_0 + \gamma_1 lnz_i + \gamma_2 w_i + \lambda_0 lnx_i + \phi_0(lnx_i)^2 + \phi_2 w_i(lnx_i)^2 \\ + \phi_1 lnz_i(lnx_i)^2 + \lambda_2 w_i lnx_i + \lambda_1 lnz_i lnx_i + \varepsilon_i \qquad (3)$$

We estimated (3) using OLS and the data for motorways, excluding observations with missing values (N = 7248). The results are presented in Table 1. The OLS residuals indicate heteroscedasticity in the error terms; hence we report the heteroscedasticity-robust standard errors (HCSE).5

Table 1. Estimated parameters with heteroscedasticity-robust standard errors (JHCSE), t and p values.

	Estimate	Rob. Std. Error (JHCSE)	t (JHCSE)	t-prob
γ_0	-195,66	52,81	-3,70	0,00
γ_1	58,30	24,75	2,36	0,02
γ_2	21,21	3,08	6,88	0,00
λ_0	46,39	11,91	3,90	0,00
ϕ_0	-2,62	0,67	-3,91	0,00
ϕ_2	0,26	0,04	6,79	0,00
ϕ_1	0,80	0,32	2,53	0,01
λ_2	-4,68	0,68	-6,84	0,00
λ_1	-13,68	5,59	-2,45	0,01

All parameters are significantly different from zero. The estimated values of ϕ_1 and ϕ_2 are positive, meaning that longer segment length and more lanes increases the upward bending of the polynomial.

We used an F-test to test the hypothesis ($\gamma_1, \phi_1, \lambda_1$)=(0,0,0), i.e. that segment length does not affect the relation between TTV and mean delay. The hypothesis is strongly rejected. In summary, both the nonparametric and the parametric analysis shows that even though we normalised travel times, and analyse travel times per kilometre, the relation between TTV and mean delay still depends on segment length. This may be problematic, which we discuss in the following section.

5. Conclusions

We have analysed the relation between TTV and mean delay, measuring TTV by the difference between the 90% and 50% quantiles of the normalised distribution (travel time per km), which is a more robust measure than the standard deviation. Overall, our results appear reliable and even reasonably robust over different road types.

However, stability is not the only criterion that matters: It would be ideal to have a measure that did not depend on the length of the analysed road segment. First, it would be easy to apply in practice in cost-benefit analysis. Second, the partitioning of the road network into segments or links in our analysis as well as in traffic models is to some degree arbitrary, and it is indeed problematic if the predicted variability

[5] We note that the standard errors are not adjusted to control for the fact that several data points stem from the same road segment.

depends on this. We expect this issue is not limited to our analysis: Kouwenhoven & Warffemius (2016) also report length dependence in their model; however, this relates to a much wider range of length than we consider, so the results are not directly comparable.

We encourage more research into constructing a TTV measure that is both robust and length-invariant, such that we can obtain a model that gives the same predicted TTV, regardless of how we partition the road network. As a first attempt, we replicated our analysis using the variance of travel time as a measure of TTV instead of the quantile difference, but this did not result in a model invariant of length. We cannot rule out that the observed length dependence could be caused by something else and may be partly explained by other variables - this should be investigated before applying the method in practice. Moreover, even if length dependence remains statistically significant, its consequences must be judged based on the numerical size of the effect.

Other issues demand attention as well:

1) Handling of travel time observations exceeding the speed limit. As a test, we truncated these travel times (fixing them at the time corresponding to the speed limit). However, this created an artificial relation between the variables of interest, and so we moved away from this solution again.
2) Our model is a rough aggregate model that distinguishes between the administrative road types, rather than between actual road features that are likely to affect TTV. This modelling should be improved, e.g. by defining more meaningful road types based on speed limits and the number of lanes.
3) In some cases, it is difficult to distinguish between unobserved segment characteristics and the effect of other characteristics, e.g. the number of lanes, because there are quite few road segments with more than six lanes (three in each direction).

Acknowledgements

This research project received financial support from the Danish Road Directorate. The research has previously been published in a report in Danish (Transport DTU, 2017).

References

de Jong, G.C., Bliemer, M.C.J., 2015: *On including travel time reliability of road traffic in appraisal*, Transportation Research Part A, vol 73, pp. 80-95.

DTU Transport, 2015. *Prediction model for travel time variability*, a report for the Danish Ministry of Transport and the Danish Road Directorate.

Eliasson, J., 2006. *Forecasting travel time variability*. European Transport Conference 2006.

Fosgerau, M., Small, K., 2012: *Marginal Congestion Cost on a Dynamic Expressway Network*, Journal of Transport Economics and Policy, vol. 46(3), pp. 431-450.

Hyder Consulting, 2010. *Motorway Travel Time Variability - Final Report*. A report for the The Highways Agency.

Kouwenhoven, M., og Warffemius, P.M.J., 2016. *Forecasting Travel Time Reliability in Road Transport: A new Model for The Netherlands*, International Transport Forum, Discussion Paper 2016-02.

Overgaard, C., Filges, D., Sørensen, H., Brun, B., 2014. *Modellering af kø-opstuvning i en statisk rutevalgsmodel*. Paper presented at the conference Trafikdage at Aalborg University.

Peer, S. Koopmans , C. C., Verhoef E. T., 2012. *Prediction of travel time variability for cost-benefit analysis*. Transportation Research Part A, vol. 46, p. 79-90.

Transport DTU, 2017. *Rejsetidsvariabilitet på strækningsniveau*. A report for the Danish Road Directorate.