ESTIMATION OF CAR-FOLLOWING MODEL PARAMETERS DISTRIBUTION USING BOOTSTRAP METHOD

Fan Wu
Master Student of Transportation College of Southeast University,
Nanjing, China, e-mail: figo@seu.edu.cn

Jian Lu
Professor of Transportation College of Southeast University,
Nanjing, China, e-mail: lujian_1972@seu.edu.cn

Jun Jiang
PhD Student of Transportation College of Southeast University,
Nanjing, China, e-mail: xuewuhen530@163.com

Submitted to the 3rd International Conference on Road Safety and Simulation,
September 14-16, 2011, Indianapolis, USA

ABSTRACT

Acknowledging driver’s car-following behavior heterogeneity is crucial to the understanding of drivers’ car-following behavior. In naturalistic driving, driver behavior also shows randomness. Empirical studies have been carried out to calibrate car-following behavior parameters in order to extract information about driver behavior characteristics. However, no explicit paradigm is available for quantitative assessment of driver behavior heterogeneity. Based on the assumption that driver's car-following behavior is stochastic, an approach utilizing the bootstrap resampling technique was proposed to investigate how drivers’ car-following model parameters are distributed. Instead of calibrating parameters for a few long trajectories, trajectories were cut into shorter segmented trajectories as calibration source, and then bootstrap was performed to estimate the sampling distribution of the calibrated driver’s car-following model parameters. A case study was presented to illustrate the process. 26 drivers were recruited in an active mode car-following experiment, and trajectory data collected in the experiment were used as data source to calibrate the intelligent driver model and to estimate the parameter distributions using the proposed approach. The method was able to find differences of parameters’ mean statistic between two driver groups. Finally, the stability of estimation was discussed regarding the trajectory length used in the calibration. The proposed method turned out to be relatively reliable and would aid the study of driver behavior heterogeneity.

Keywords: car-following model, calibration, bootstrap, driver behavior heterogeneity

INTRODUCTION

Driver behavior has long been an important topic for transportation research. The human factors influence roadway designs, traffic flow stability and safety. Recently much attention has been focused on driver behavior heterogeneity. And acknowledging driver's car-following behavior heterogeneity is crucial to the understanding of drivers' car-following behavior. Studies on driver heterogeneity suggest that there are both certain degree of inter-driver and intra-driver differences in car-following. Ossen and Hoogendoorn (2007) reports a high degree of inter-driver heterogeneity in car-following behavior. Wang et al. (2010) studies the intra-driver heterogeneity of driving behavior between the acceleration process and the deceleration process and found that obvious intra-driver heterogeneities exist in driving behaviors between acceleration processes and deceleration processes of car-following. The intra-driver heterogeneity can also be seen as a result of the stochastic nature of driver behavior.

In naturalistic driving, drivers would not perform deterministically and randomness is always observed. The stochastic behavior of driving has been studied by (Jost and Nagel 2003; Wagner 2005). Several approaches also model driver behavior as stochastic process. For example, Yang et al. (2008) proposes a stochastic driver model based on the assumption that the driver normally has intention to achieve a desired vehicle speed and as long as this state was roughly achieved, some deviations would be acceptable. Hamdar et al. (2008) model drivers' car-following as utility maximization with a stochastic choice between different acceleration alternatives.

Based on the assumption that drivers' car-following behavior is stochastic, it is possible to assume that a driver's car-following behaviors at different time are independent. Then the parameters of a car-following model become random variables following certain distributions. To quantitatively assess driver behavior heterogeneity in car-following, it would be helpful to investigate how drivers' car-following model parameters are distributed.

By simply estimating the parameter distributions of car-following models from many independent trajectories, the deterministic assumption of traditional car-following models is partially relaxed. In this work, an approach utilizing the bootstrap resampling technique was proposed. Instead of calibrating parameters for a few long trajectories, trajectories of the same driver or same group of drivers were cut into shorter segmented trajectories as calibration source, and then bootstrap was performed to estimate the sampling distribution of the calibrated driver's car-following model parameters.

The calibration and bootstrap approach is to be presented in the following section.

Then a case study to investigate the influence of driving experience on drivers' car-following behavior was explicated, which employed the proposed method. The conclusions and future research shall be discussed in the last section.

THE CALIBRATION AND BOOTSTRAPPING APPROACH

The method includes two steps. First the parameters were calibrated independently for each short trajectory. Then the calibrated values of same driver or same driver group were selected and bootstrapped to give estimates of model parameters for driver or driver groups.

Car-following model calibration

The general framework for calibrating car-following models proposed by Ossen (2008) is here used to find optimal values of car-following model parameters. The framework is described as follows. Let $z_n(t)$ be real state of driver n at time t, the state variables normally include position $x_n(t)$ and speed $v_n(t)$, then vector $\xi_n(t)$ which consists of the states of j vehicles ahead represents the traffic condition driver n is facing at time t:

$$\xi_n(t) = (z_{n-j}(t), \cdots z_{n-1}(t)) \tag{1}$$

A general car-following model can be abstracted as function f which takes traffic condition at time $t - T_r$ (T_r is driver's reaction time delay) $\xi_n(t - T_r)$ as variables and a set of parameters β_n as model parameters. The evolution of state of driver n then can be described as the differential equation as Equation (2). The discretized predicted state of driver n at time t_k $\hat{z}_n(t_k)$ can be calculated through integration in Equation (3).

$$\frac{d}{dt}z_n(t) = f(\xi_n(t - T_r \mid \beta_n)) \tag{2}$$

$$\hat{z}_n(t_k) = \begin{cases} y_n(t_1)) & k = 1 \\ \hat{z}_n(t_{k-1}) + \int_{t_{k-1}}^{t_k} f(\xi_n(s - T_r) \mid \hat{\beta}_n)ds & \text{otherwise} \end{cases} \tag{3}$$

The optimal parameter set β_n^* will be the one that minimize the objective function g which measures the difference of observed state y_n and predicted state $\hat{z_n}$.

$$\beta_n^* = argmin_{\hat{\beta}_n} g(y_n, \hat{z_n}) \tag{4}$$

The optimization algorithm for the objective function needs to be selected. In the case study the genetic algorithm was used.

Bootstrapping

After the calibration, the calibrated values of same driver or same driver group were selected and bootstrapped to give estimates of model parameters for drivers or driver groups. Bootstrap method was first introduced by Efron (1979). In statistics, bootstrapping is a computer-based method for assigning measures of accuracy to sample estimates. The technique allows estimation of the sample distribution of almost any statistics using only very simple methods. When the theoretical distribution of a statistic of interest is complicated or unknown, it is the case for car-following model parameters. The bootstrapping procedure is distribution-independent. It provides an indirect method to assess the properties of the distribution underlying the sample and the parameters of interest that are derived from this distribution.

Given the assumption that drivers' car-following behavior is stochastic, the independent and identically distributed condition is automatically satisfied. Many statistics of each parameter can be bootstrapped to give estimates of the subject driver. To assess how drivers’ car-following model parameters are distributed, the nonparametric bootstrap for the mean and standard deviation statistics was performed.

CASE STUDY

A case study was performed to investigate the influence of driving experience on drivers’ car-following behavior.

Data collection and processing

Trajectory data were collected from an active mode car-following experiment. This type of experiment utilizes GPS and various ranging technologies (for this study laser ranging was used) to measure driver behavior and relative motion of longitudinal adjacent vehicle pairs. Active mode means that it is the driver driving the experiment vehicle who is being observed. Active car-following experiment is able to cover a relatively long time of driver’s behavior and thus able to characterize driver’s behavioral patterns. The specifics of the experiment are described as follows.

Equipments

The experiment vehicle was an automatic transmission sedan with a discharge capacity of 1.8L. The length of the vehicle is 4.18m, and the width is 1.70m. The experiment system consisted of two essential parts: a GPS receiver and a laser rangefinder. The GPS receiver was used to measure the position of experiment vehicle, and the laser rangefinder was used to measure the space gap. A micro-computer was used to connect the GPS receiver and the laser rangefinder, running the experiment software. A video camera was used to record the experiment scenario and behavior of test drivers. The experiment data were monitored on the laptop. The laser rangefinder was mounted on the front deck where it was 0.8m away from the front bumper of the vehicle. The equipments and experiment scenario are shown in Figure 1.

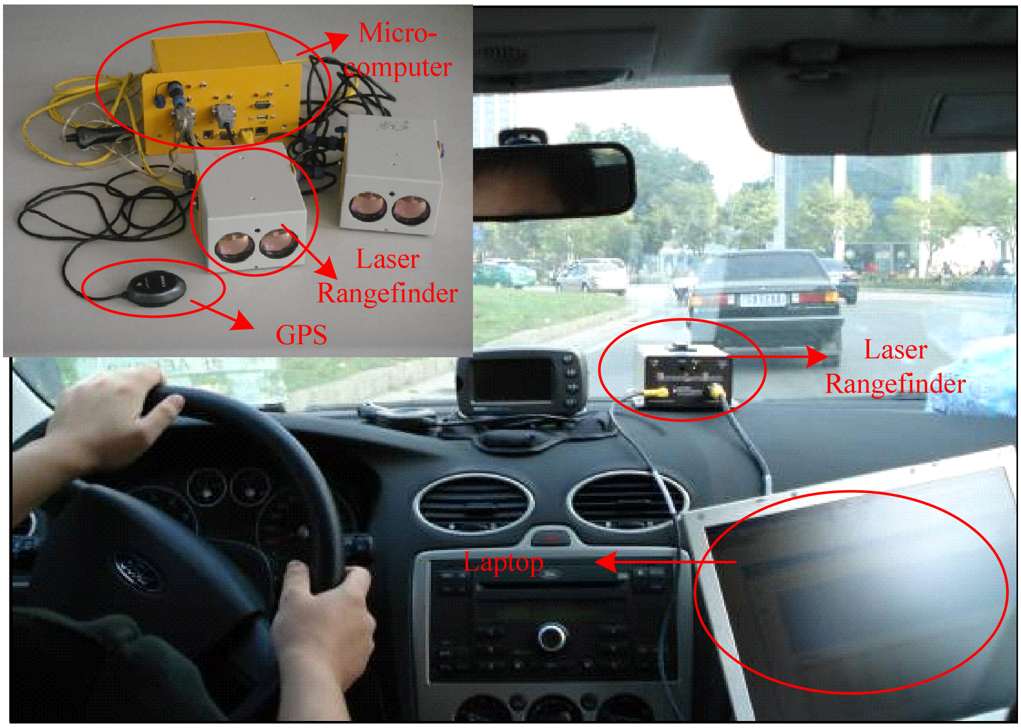

Figure 1 Equipments and the Experiment Scenario

Drivers

10 experienced drivers and 16 beginner drivers were recruited in the experiment. Experienced drivers and beginner drivers were distinguished considering following factors: a) cumulative mileage b) years that drivers have held their driver's license for. Drivers who have held their driver's licenses for more than 5 years with a cumulative mileage of more than 150,000 km were classified as experienced driver, otherwise as beginner driver.

Age of experienced driver ranged from 33 to 53, beginner drivers' age varied among 23 to 29. The experienced group had an average 12.6 years history of holding a driver's license and an average 672,200 km cumulative mileage in contrast to 3.1 years and 20,070 km for beginners' group. In all tested drivers only 3 beginner drivers were female. The occupations of all the drivers covered taxi driver, teacher, officer, engineer and student.

Data Collection

The experiment was carried out in sunny days during daytime (08:00 a.m. to 06:00 p.m.) on urban 4-6 lane roads in Nanjing, China. Before the experiment, drivers were informed with the same designated route and were instructed to relax and drive as usual, and then each driver drove the experiment vehicle for about 45 minutes following the same designated route. Traffic conditions such as flow rate or density were not measured, and the planned route were moderately congested (no gridlock) during peak hours. Distance between experiment vehicle and frontal vehicle were measured by the laser rangefinder every 0.1 second, and positions of experiment vehicle were reported by the GPS every 1 second. The experiment lasted about 15 days in September, 2010.

Data Processing

The raw data collected in the experiment contained missing and false data due to the hardware limitations of the laser rangefinder. Missing data were reported when the rangefinder’s laser beam was projected onto low reflective rate surface or the distance exceeded the hardware limit. When vehicle turned or changed lanes the rangefinder would project its laser beam onto objects other than frontal vehicle, thus false data were collected. Also the GPS and distance measurements contained random errors.

To tackle the problem of discontinuity and lessen the effect of measurement error, a method combining local regression and manual processing was applied. According to Toledo et al. (2007), the local regression method developed by (Cleveland 1979; Cleveland and Devlin 1988) is suitable for processing trajectory data. The procedure is able to recover short missing data and cancel out random errors; also local regression performs better than simple moving averaging in terms of robustness.

First, the trajectories of experiment vehicle and frontal vehicle were calculated from raw data, and then local regression was applied to both trajectories. The polynomial degree was set to 3 to ensure that the derived acceleration values from differentiation do not degenerate to a constant; windows size was set as 11 covering about 1s most of the case and covering longer time when there is missing value in the window. Tricube kernel function (shown in Figure 2) was used as weight function. An example of local regression result is given in Figure 3.

Then manual labor was applied to pick out reasonable data. Through visual inspection, trajectories well mapped by local regression and also monotonically nondecreasing were picked out. The speed, acceleration, time gap and following distance were calculated for each observation from the processed trajectory data. Finally, observations containing anomaly values (speed>100km/h, acceleration falls out the physical possible range [-5, 3] m/s2) were removed.

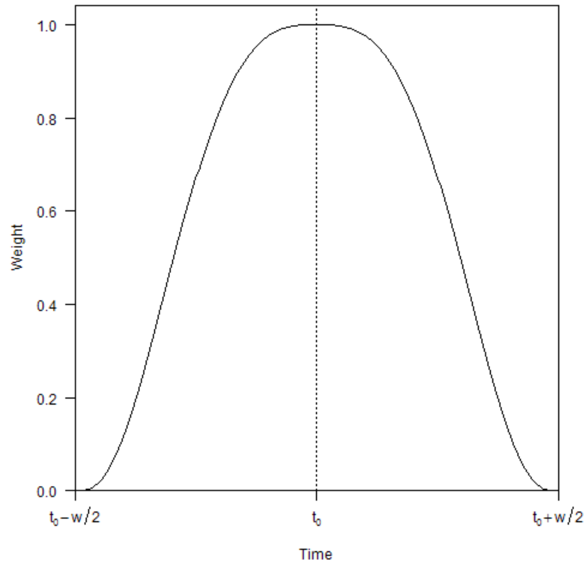

Figure 2 Tricube Weight Function (w is the window size)

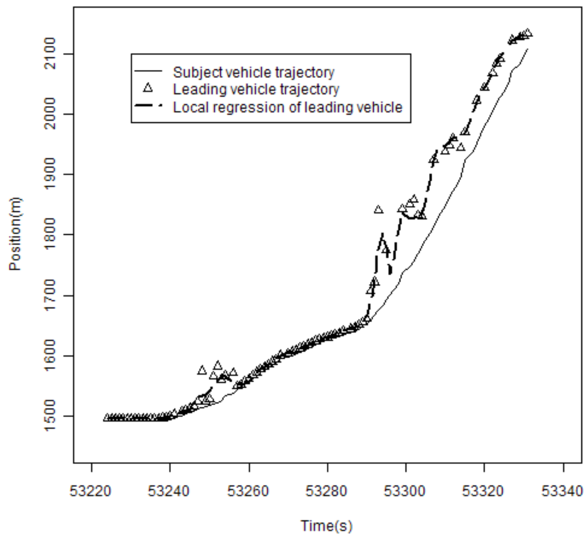

Figure 3 Example of Local Regression on Trajectories

Calibration for the IDM model

In the case study the Intelligent Driver Model (IDM) was used as car-following model. IDM is proposed by Treiber et al. (2000). The original IDM has 5 parameters. The 5 parameters are desired speed DV, maximum acceleration rate a_{max}, maximum deceleration rate b_{max}, minimal space gap ΔX^* and desired time headway T.

The objective function form used is Theil's U (Theil 1961) which is given in equation (5), the state variable y_n and $\hat{z}_n$ used in the objective function is space gap.

$$g(y_n, \hat{z}_n) = \frac{\sqrt{\frac{1}{K}\sum_{k=1}^{K}(y_n(t_k) - \hat{z}_n(t_k))^2}}{\sqrt{\frac{1}{K}\sum_{k=1}^{K}(y_n(t_k))^2} + \sqrt{\frac{1}{K}\sum_{k=1}^{K}(\hat{z}_n(t_k))^2}} \tag{5}$$

The search range of parameters is shown in Table 1. The search range of parameters defines the search space for the genetic algorithm. By using the prior information the calibrated values will be guaranteed in a reasonable range.

Table 1 : Search Range of Parameters

Parameter	$DV(m/s)$	$a_{max}(m/s^2)$	$b_{max}(m/s^2)$	$\Delta X^*(m)$	$T(s)$
Range	[0,30]	[0.3,3]	[0.5,10]	[0,5]	[0,5]

Results

The average length of 815 trajectories used in calibration is 21.4s. The average value of objective function is 6.6%. Estimation of parameter means for IDM is shown in Table 2. Estimation of parameter standard deviation for IDM is shown in Table 3. The result of stratified bootstrap of difference of mean of the two driver groups is shown in Table 4. The 95% confidence intervals are normal bootstrap confidence intervals in Table 2-4. The 95% confidence intervals of differences in mean value of desired speed and maximum deceleration rate are entirely at one side of the zero point. These results suggest that comparing to beginner drivers experienced driver have lower desired speed and higher maximum deceleration rate.

Table 2 : Estimation of Parameter Means for IDM

	Beginner driver group		Experienced driver group	
	Sample means	95% C.I.	Sample means	95% C.I.
$DV(m/s)$	8.09	(6.978,9.236)	6.48	(5.549,7.421)
$a_{max}(m/s^2)$	0.63	(0.531,0.718)	0.66	(0.523,0.802)
$b_{max}(m/s^2)$	2.33	(1.362,3.295)	4.29	(2.621,5.949)
$\Delta X^*(m)$	2.50	(2.124,2.878)	2.41	(1.905,2.931)
$T(s)$	2.46	(2.140,2.783)	2.34	(1.810,2.886)

Table 3 : Estimation of Parameter Standard Deviation for IDM

	Beginner driver group		Experienced driver group	
	Sample S.D.	95% C.I	Sample S.D.	95% C.I
$DV(m/s)$	5.55	(4.338,6.866)	4.83	(4.286,5.440)
$a_{max}(m/s^2)$	0.34	(0.229,0.474)	0.46	(0.243,0.724)
$b_{max}(m/s^2)$	2.31	(1.310,3.531)	3.69	(3.053,4.593)
$\Delta X^*(m)$	1.27	(1.043,1.523)	1.50	(1.305,1.749)
$T(s)$	1.12	(0.936,1.333)	1.32	(1.052,1.643)

Table 4 : Estimation of Difference of Mean of IDM Parameters

	Difference of mean	95% C.I
$DV(m/s)$	1.60	(0.072,3.048)
$a_{max}(m/s^2)$	- 0.03	(- 0.204,0.137)
$b_{max}(m/s^2)$	- 1.95	(- 3.774, - 0.115)
$\Delta X^*(m)$	0.09	(- 0.552,0.751)
$T(s)$	0.12	(- 0.519, - 0.742)

Influence of trajectory length

An important concern about the validity of the proposed method is whether the average trajectory length has an impact on the result. To address this problem an experiment was carried out. One of the 26 drivers was picked as the subject. The trajectories data was further evenly cut into quarter length. The original average trajectory length is 19.6s while the average trajectory length for processed data set is 4.9s. Both the original and processed dataset were calibrated and bootstrapped. The results are presented in Table 5 and Table 6. The average objective function value is 14.8% for the original dataset and 1.5% for the quarter length dataset. The obvious lower objective function value for the quarter length dataset suggests shorter trajectories are better fitted by the model than longer trajectories. To assess the similarity of two confidence interval estimations, a simple variable K quantifying the proportion of overlapping interval length in the total covering interval length was calculated. The maximum value of K is 1 where two intervals are exactly the same.

The experiment results are provided in Table 5 and Table 6. In Table 6 while some of the bootstrap distributions are not normal, bias-corrected, accelerated (BCa) confidence intervals are given instead of normal confidence intervals. The calculated K values are given in Table 7. For the mean statistic confidence interval quarter length dataset seems to get better results as the confidence intervals (all except that of a_{max}) are within the original estimated intervals and are generally narrower. The experiment results suggest the proposed method is relatively stable for most of the parameters in terms of trajectory length for the estimation of car-following model parameter distributions.

Table 5 : Comparison of Parameter Means of Different Trajectory Length

	Original length		Quarter length	
	Sample means	95% C.I.	Sample means	95% C.I.
$DV(m/s)$	6.70	(2.208,11.092)	5.32	(3.657,7.039)
$a_{max}(m/s^2)$	0.46	(0.329,0.588)	0.57	(0.478,0.657)
$b_{max}(m/s^2)$	1.59	(0.927,2.257)	1.49	(0.796,2.204)
$\Delta X^*(m)$	3.01	(1.627,4.425)	2.26	(1.980,2.543)
$T(s)$	1.49	(0.537,2.406)	0.99	(0.656,1.308)

Table 6 : Comparison of Parameter Standard Deviation of Different Trajectory Length

	Original length		Quarter length	
	Sample S.D.	95% Bca C.I	Sample S.D.	95% Bca C.I
$DV(m/s)$	6.42	(4.178,8.550)	5.21	(3.974,7.152)
$a_{max}(m/s^2)$	0.18	(0.039,0.262)	0.16	(0.121,0.220)
$b_{max}(m/s^2)$	0.98	(0.722,1.226)	1.24	(0.682,1.860)
$\Delta X^*(m)$	1.87	(1.290,2.313)	1.10	(0.910,1.313)
$T(s)$	1.29	(0.764,1.828)	1.06	(0.748,1.525)

Table 7: K value for Evaluation of Different Estimations

	K for means	K for S.D.
$DV(m/s)$	0.38	0.65
$a_{max}(m/s^2)$	0.34	0.44
$b_{max}(m/s^2)$	0.87	0.43
$\Delta X^*(m)$	0.20	0.02
$T(s)$	0.35	0.70

CONCLUSIONS

An approach using the bootstrap resampling method was presented in this paper to estimate car-following model parameters' distributions. Based on the assumption that drivers' car-following behavior is stochastic, short trajectories are fed as independent calibration sources to calculate car-following model optimal parameters and bootstrap was performed to estimate the parameter distribution statistics of drivers or driver groups.

A case study employing the proposed method was conducted in China to investigate whether drivers' driving experience has an effect on the car-following behavior. The results of the case study suggest a difference of desired speed and maximum deceleration rate, two parameters of the IDM model, between beginner and experienced driver groups. An experiment to study the influence of trajectory on the result was also carried out. The results of the experiment suggest that the method is relatively reliable and yet shorter trajectories seem to provide better estimations.

The findings of this paper suggest that the proposed method would be useful for the study of driver behavior heterogeneity. The approach also has its advantages for its applicability to low quality trajectory data since long consecutive trajectories are not always available. Finally, the study raises a question that what would be the optimal trajectory length for microscopic car-following model calibrations and further studies on this topic would be beneficial.

ACKNOWLEDGEMENT

Work reported in this paper has been funded by the National Science Foundation of China (No.50708019), Program for New Century Excellent Talents in University （NCET-08-0115）; Jiangsu Province Ordinary University Graduate Student Scientific Research Innovation Plan (No. CX08B_130Z) and "Blue Project" of Jiangsu Province. The authors thank the support of the above foundations.

REFERENCES

Cleveland, W. S. (1979). "Robust Locally Weighted Regression and Smoothing Scatterplots." *Journal of the American Statistical Association*, 74(368), 829-836.

Cleveland, W. S., and Devlin, S. J. (1988). "Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting." *Journal of the American Statistical Association*, 83(403), 596-610.

Efron, B. (1979). "Bootstrap Methods: Another Look at the Jackknife." *The Annals of Statistics*, 7(1), 1-26.

Hamdar, S., Treiber, M., Mahmassani, H., and Kesting, A. (2008). "Modeling Driver Behavior as Sequential Risk-Taking Task." *Transportation Research Record: Journal of the Transportation Research Board*, 2088(-1), 208-217.

Jost, D., and Nagel, K. (2003). "Probabilistic Traffic Flow Breakdown in Stochastic Car-Following Models." *Transportation Research Record: Journal of the Transportation Research Board*, 1852(-1), 152-158.

Ossen, S. (2008). "Longitudinal Driving Behavior: Theory and Empirics," PhD thesis, TU Delft, The Netherlands.

Ossen, S., and Hoogendoorn, S. (2007). "Driver Heterogeneity in Car Following and Its Impact on Modeling Traffic Dynamics." *Transportation Research Record: Journal of the Transportation Research Board*, 1999(-1), 95-103.

Theil, H. (1961). *Economic Forecasts and Policy*. North Holland, Amsterdam.

Toledo, T., Koutsopoulos, H., and Ahmed, K. (2007). "Estimation of Vehicle Trajectories with Locally Weighted Regression." *Transportation Research Record: Journal of the Transportation Research Board*, 1999(-1), 161-169.

Treiber, M., Hennecke, A., and Helbing, D. (2000). "Congested traffic states in empirical observations and microscopic simulations." *Physical Review E*, 62(2), 1805.

Wagner, P. (2005). "Empirical Description of Car-Following." Traffic and Granular Flow ’03, S. P. Hoogendoorn, S. Luding, P. H. L. Bovy, M. Schreckenberg, and D. E. Wolf, eds., Springer Berlin Heidelberg, 15-27.

Wang, H., Wang, W., Chen, J., and Jing, M. (2010). "Using Trajectory Data to Analyze Intradriver Heterogeneity in Car-Following." *Transportation Research Record: Journal of the Transportation Research Board*, 2188(-1), 85-95.

Yang, H. H., Peng, H., Gordon, T. J., and Leblanc, D. (2008). "Development and validation of an errorable car-following driver model." *American Control Conference, 2008*, 3927-3932.