A LEAST SQUARES METHOD FOR ORIGIN-DESTINATION ESTIMATION INCORPORATING VARIABILITY OF DAY-TO-DAY TRAVEL DEMAND

Tao WEN[a,b] (corresponding author), Chen CAI[b], Lauren GARDNER[a], Vinayak DIXIT[a], S. Travis WALLER[a,b], and Fang CHEN[b]

[a] School of Civil and Environmental Engineering, UNSW Australia, Australia; [b] National ICT Australia (NICTA), Australia,
Corresponding author email: t.wen@unsw.edu.au

ABSTRACT:

The estimation of an origin-destination (O-D) trip matrix from observed link flow is an essential part of the 4-step transport planning process. Conventional methods for estimating O-D trip matrices usually consider the fluctuation of link flows observed from loop counts as an error term that can be ignored, resulting in a significant loss of information. This work, instead, explicitly considers the fluctuation of link flows as a result of the volatility of demand. A strategic user equilibrium (StrUE) model is implemented, which aims to minimize the expected travel time, assuming users have learned the actual demand distribution. Expressions for the analytical link travel times and corresponding variability are derived from the StrUE model. This information is utilized in the least squares method proposed in this paper. The objective of the proposed model is therefore to estimate a demand distribution that minimizes the distances between the two terms: 1) the mean of the observed link flows and the estimated mean of the link flows; 2) standard deviation of observed link flows and estimated standard deviation of the link flows. A solution algorithm is proposed, and results are presented from numerical experiments are implemented on a virtual network.

Keywords: strategic user equilibrium, O-D estimation, least squares method

Subject area: (Please put a “X” as appropriate, you may choose more than one)

	Transportation Surveys
	Travel Behaviour Modelling
X	Transportation, Land Use and Built Environment
	Transportation and Quality of Life
	Transportation and Climate Change
	Transportation Infrastructure
	Technology, Transportation and Telecommunications
	Logistics and Supply Chain Management
	Transport Dynamics
	Regional and International Transportation
X	Others

Preferred mode of presentation: (Please put a “X” as appropriate)

	Oral presentation only		Poster presentation only
X	No preference		

Name of presenter/key speaker:

Tao Wen

Request for the consideration of publication in the Proceedings: (Please put a “X” as appropriate)

X	Yes		No

Submission for HKSTS Outstanding Student Paper Award (You are a postgraduate student and the sole author of the paper): (Please put a “X” as appropriate)

X	No		Yes

1. INTRODUCTION AND BACKGROUND:

Enhanced origin-destination (O-D) matrix estimation methodologies would be extremely useful for transportation planning. Traditionally, the O-D matrix is obtained from plate surveys, household surveys or roadside surveys. Such survey activities may be financially expensive for large size networks at frequent intervals, and usually suffer from limited response and sampling coverage. As an alternative, O-D matrix estimation provides a statistical approach for estimating or calibrating an O-D matrix from observed link flows and some prior knowledge of the O-D demand. In this paper, we propose a framework that combines the least squares method for O-D matrix estimation and the strategic user equilibrium model (StrUE) for traffic assignment (Dixit et al., 2013). This framework is hereby referred as LSStrUE. The StrUE model is defined such that "at strategic user equilibrium all used paths have equal and minimal expected cost". For each user present in a given demand scenario, their chosen route is followed regardless of the realized travel demand on a given day. Therefore the link flows will not result in an equilibrium state in any particular demand realization, but instead equilibrium exists stochastically overall demand realizations. The StrUE model was proposed to capture the impact of day to day demand volatility on reliability, and eventually route choice. Therefore, apart from the O-D splits, a fundamental parameter that needs to be estimated is the variance in the total trip demand distribution.

An important aspect of the StrUE model is that the total trip demand is assumed to follow a certain statistical distribution; traditionally a lognormal distribution has been used (Wen et al., 2014, Duell et al., 2014). Under the assumption of a lognormal distribution for demand, this paper focuses on estimating the demand distribution parameters. Only the distribution of the total trip demand needs to be estimated because for the StrUE model the O-D proportions are assumed to be fixed. Furthermore, for the StrUE model a lognormal-distributed total trip demand allows for a closed form probability density function of the link flow, which can be shown to follow a lognormal distribution as well. The direct relationship between the link flow variables and the total demand in StrUE allows for the use of both expected link flow and standard deviation of link flow to estimate the total demand distribution. In this study this estimation is accomplished by implementing the least squares estimation method, in which we minimize the squared different between estimated and observed expected link flows and standard deviations of link flows. A bi-level programming method is proposed to eliminate the impact of strongly biased initial estimates, where the upper level provides the total mean demand and variance to the StrUE model, from which the StrUE model can provide link flow mean and variance to the upper level. A benefit of the proposed modelling framework includes the incorporation of actual day-to-day observed link flows and the corresponding variations. Additionally, the performance of the LSStrUE approach can be assessed based on the accuracy of its estimations for both expected link flows and link flow variations, which are a direct output of the StrUE model.

Historically O-D matrix estimation has focused on statistical approaches based on loop counts. however a wide range of alternative methods have been explored in previous studies, including the generalized least square method(Cascetta, 1984, Bell, 1991), the maximum likelihood(Spiess, 1987), bi-level programming approach(Yang et al., 1992) and maximum entropy(Fisk, 1988). Generally the problem is to identify an O-D matrix which optimizes an objective function subject to a set of constraints. However, the problem is often challenging due to the number of observable links in a traffic network, which is typically much smaller than the number of O-D pair demands. Thus, it may not be possible to obtain a unique solution from a single set of link counts alone. The problem was further extended to account for the stochastic nature of observed flows (Lo et al., 1996, Lo et al., 1999). Recently, dynamic approaches were introduced to account for the time dependent characteristics in the network (Frederix et al., 2011, Bierlaire and Crittin, 2004). However, the application to large scale network and the computation complexity still remains a problem.

Among the research, relatively little attention has been paid to the higher order of the variables in a network, such as their variance and covariance that can potentially provide more constraints to the optimization problem. Cremer and Keller demonstrated that aggregating or averaging link count data collected over a sequence of time period may lose some important information.(Cremer and Keller, 1987). Hazelton (Hazelton, 2003) proposed a weighted least squares method to account for the covariance of links, and assumed a parameter to explain the circumstances when the variance exceeds the mean if a Poisson distribution is used. Bell (Bell, 1983) proposed a maximum likelihood method and found the analytical solution to the covariance of O-D matrix by using a Taylor approximation. However, these research contributions still have some limitations in the assumptions. For example, the O-D demand was assumed to follow the Poisson or multinomial distribution, which stipulates certain relationships between the mean and variance of the O-D demand. In the LSStrUE, the O-D demand is assumed to follow a lognormal distribution, which allows the mean and variance of total demand to be independent of each other, and assures the non-negativity of the demand. In a well-constructed network, loop detectors can easily provide link counts on a day-to-day basis; therefore, it is important to consider the variation of link flows and the distribution of O-D total demand as effective information to calibrate the O-D trip matrix. The proposed LSStrUE framework estimates the distribution of the total O-D demand and thus significantly reduces computation complexity.

Estimation of the O-D trip matrix also requires a proper assignment model. When applying the assignment model to a large network, realism and computational complexity are both critical and must be equally considered to determine a model's practical applicability. Further, a major complication in transportation modelling is the ability to properly account for the inherent uncertainties regarding demand (Kim et al., 2009, Bellei et al., 2006) and capacity levels (Brilon et al., 2005, Wu et al., 2010). Additionally, as has been noted, uncertainty on these variables directly affects route choice behaviour (Uchida and Iida, 1993). It is, therefore, necessary to incorporate these stochastic elements into models to ensure robust planning capabilities, but to do so in a manner that maintains computational tractability. The strategic user equilibrium (Dixit et al., 2013) effectively accounts for the impact of demand uncertainty; the model relies on users minimizing their expected travel time based on their previous trip experiences in which they have gathered knowledge on demand (daily trips). The user's knowledge of each can be represented by a given distribution, with a known expected value and variance. Based on these known distributions, each user selects a travel route to minimize their expected travel time subject to Wardrop's UE conditions.

This study expands on previous OD estimation research by (i) considering the day-to-day link flow variations and using a least squares method to relate this information to the total demand distribution, (ii) assuming that the total demand follows a lognormal distribution, allowing the variance and the mean of the total demand to be different, and (iii) applying the strategic user equilibrium model to account for the impact of variation in demand, and link flows according to the estimated total demand distribution. The remainder of this paper is structured as follows: Section 2 defines the mathematical model and includes a derivation for the analytical solution to the total demand estimation. Numerical analysis is demonstrated in Section 3; conclusions and future research are presented in Section 4. A Table in the Appendix defines the notation used in this paper.

2 MODEL FORMULATION:

This section defines the mathematical concept of the LSStrUE framework. A least squares method is implemented to minimize the differences in mean and standard deviation:

$$\min : J\left(m_T, s_T\right) = \sum_n \left(p_n m_T - m_n\right)^2 + \sum_n \left(p_n s_T - s_n\right)^2 \quad n \in N \tag{1}$$

The first order derivative can be obtained by taking partial derivatives with respect to $\mathrm{m_T}$ and $\mathrm{s_T}$:

$$\frac{\partial J\left(m_T, s_T\right)}{\partial m_T} = \sum_{n \in N} 2 p_n \left(p_n m_T - m_n\right) \tag{2}$$

$$\frac{\partial J\left(m_T, s_T\right)}{\partial s_T} = \sum_{n \in N} 2 p_n \left(p_n s_T - m_n\right) \tag{3}$$

The Hessian matrix of the objective function is obtained by taking partial derivative with respect to the variable m_T and s_T:

$$H = \begin{bmatrix} \sum_n 2p_n^2 & 0 \\ 0 & \sum_n 2p_n^2 \end{bmatrix} > 0 \quad (4)$$

It is shown that the Hessian matrix is strictly positive, therefore the objective function has unique optimal solution, which can be found when the first derivative of the objective function is equal to zero:

$$\frac{\partial J(m_T, s_T)}{\partial m_T} = 0 \rightarrow m_T = \frac{\sum_n p_n m_n}{\sum_n p_n^2} \quad (5)$$

$$\frac{\partial J(m_T, s_T)}{\partial m_T} = 0 \rightarrow s_T = \frac{\sum_n p_n s_n}{\sum_n p_n^2} \quad (6)$$

The parameters of the lognormal distribution can be obtained by the definition of the lognormal distribution:

$$\mu_T = \ln m_T - \frac{1}{2}\ln\left(1 + \frac{s_T^2}{m_T^2}\right) \quad (7)$$

$$\sigma_T^2 = \ln\left(1 + \frac{s_T^2}{m_T^2}\right) \quad (8)$$

Assuming that the StrUE model represents the route choice behaviour, we can then formulate a bi-level programming problem, where the upper level is the least squares demand estimation problem; the lower level is the StrUE model, which has the objective function:

$$Minimize: z(f) = \sum_{n \in N} \int_0^{f_n} \int_0^{\infty} \int_0^{\infty} t_n(p_n T) g(T) dT df \quad (9)$$

Subject to:

$$\sum_k v_k^{rs} = q_{rs} \qquad \forall k, r, s \quad (10)$$

$$v_k^{rs} \geq 0 \qquad \forall k, r, s \quad (11)$$

$$p_n = \sum_r \sum_s \sum_k v_k^{rs} \delta_{n,k}^{rs} \qquad \forall k, r, s \quad (12)$$

The fraction of the total demand between O-D pair *r-s*, namely q_{rs} , can be obtained from the prior estimates, i.e. household survey data, or field experiments. The link travel time function for the StrUE model is defined by the U.S. Bureau of Public Roads (U.S, 1964) cost function due to its widespread use in transport planning models:

$$t_n(flow) = t_{nf}\left[1 + \alpha t_{nf}\left(\frac{f_n T}{C_n}\right)^{\beta}\right] \quad (13)$$

Where α and β are the parameters for the BPR function. The objective functions of the upper and the lower levels are both strictly convex, therefore the model always has feasible solutions. A solution algorithm has been proposed to the bi-level programming:

Step 1: (Initialization) k=0. Start from the prior O-D matrix; obtain the fraction of total trips q_{rs} and initial values for the mean and the variance of the total demand. Produce a set of link proportions from the StrUE model. Note that q_{rs} will be kept invariant over the bi-level iterations while μ_T^k and σ_T^k will be calibrated.

Step 2: (Optimization) Substituting the link-flow proportion matrix P_k, solve the upper-level to obtain μ_T^k and σ_T^k of the total demand.

Step 3: (Simulation) Using μ_T^k and σ_T^k, apply the StrUE model to produce a new set of link flow proportions P_{k+1}.

Step 4: (Convergence test) Calculate the deviation between simulated and observed link flows, and the deviation between estimated and target O-D matrices. If stopping criterion is met, stop.

3 NUMERICAL ANALYSIS:

The objective of the analysis is to test if the LSStrUE can effectively estimate the total demand distribution from observed link flows. The estimated total demand distribution should closely approximate the actual total demand distribution; the mean and variance of link flow produced by the StrUE model should also closely match the observed link flows. The main idea is to artificially determine the total demand distribution and generate random link flow samples accordingly. The LSStrUE will reproduce the desired total demand distribution from the random samples with perturbed prior estimates. The test should also reflect the scalability of the LSStrUE to networks of substantial complexity. Numerical tests are conducted on the Sioux Falls network (24 nodes and 76 links). The network properties are pre-defined in (Bar-Gera, 2012). The notations used in this section are defined in the Appendix. The O-D demand is specified as proportions of the total network demand, therefore the demand for a given O-D pair is the O-D proportion multiplied by the total demand. The BPR parameters α and β are equal to 0.15 and 4, respectively. The observed link flows are generated by the following way:

Step 1: The true μ_T, σ_T are determined for the total demand.

Step 2: We implement the StrUE based on the total demand distribution and obtain a set of link proportions.

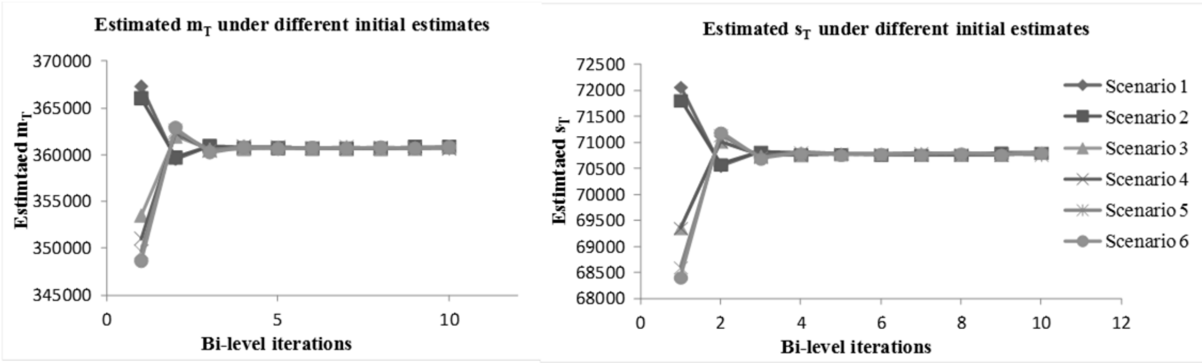
Step 3: We generate 100 samples of the total demand from the lognormal distribution using μ_T, σ_T as parameters and each sample total demand is assigned to the network using the pre-calculated link proportions, thus providing multiple sets of link flows.

The actual expected total demand of the Sioux Falls network is m_A=360600, and the coefficient of variation cov is equal to 0.2, i.e. the standard deviation is 20% of the expected total demand. In Table 1, each scenario represents a different initial estimate of the total demand distribution.

TABLE 1 Different scenarios of initial estimation of O-D matrix

Scenario	Scenario description	m_T	s_T
1	$m_T = 0.8m_A \text{ and } cov = 0.1$	288480	28848
2	$m_T = 0.8m_A \text{ and } cov = 0.3$	288480	86544
3	$m_T = 1.2m_A \text{ and } cov = 0.1$	432720	43272
4	$m_T = 1.2m_A \text{ and } cov = 0.3$	432720	129816
5	$m_T = 1.5m_A \text{ and } cov = 0.1$	540900	54090
6	$m_T = 1.5m_A \text{ and } cov = 0.3$	540900	162270

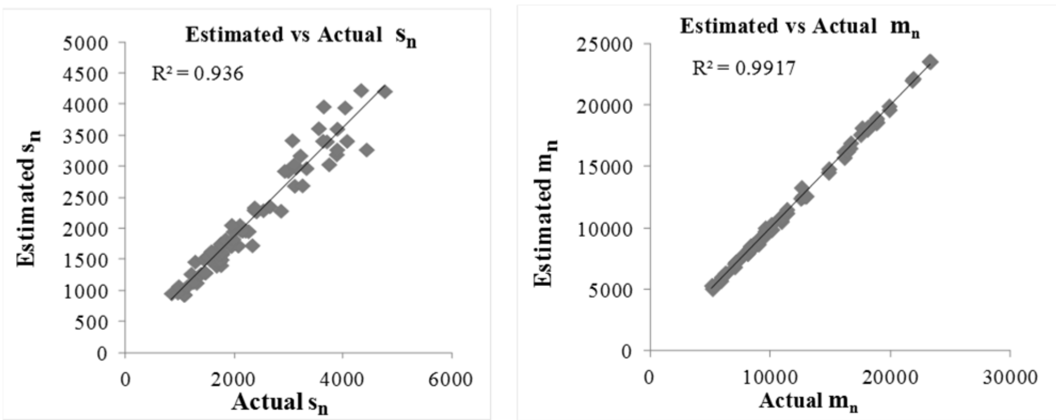
In Fig.2.a and Fig.2.b, the x-axis represents the number of iterations of the bi-level programming. In Fig.2.a, the y-axis represents the estimated expected total demand; In Fig.2.b, the y-axis represents the estimated standard deviation of the total demand. Both figures show that the estimated results converge to the actual ones in less than 3 iterations. This indicates that the LSStrUE's robust performance against biased initial estimates, and demonstrates the efficiency in arriving at convergence. In scenario 1 and 2, the estimated results of the first iteration in both figures are very different from the actual ones. This is mainly because the link proportions of the first iteration are obtained based on the initial estimates, and the initial estimates in scenarios 1 and 2 are very biased, and therefore the results of the first iteration will be inaccurate. If the initial estimates of total demand distribution are quite different from the actual distribution, the estimation of both expected total demand and standard deviation of total demand will be very inaccurate; therefore we have shown that applying the bi-level programming can reduce the impact of biased initial estimates.



(a) (b)

FIGURE 2 Estimated (a) expected total demand and (b) standard deviation of total demand under different scenarios of initial estimation; results of 10 bi-level iterations are presented.

Fig.3 (a) demonstrates the performance of the LSStrUE at the link level; the *x*-axis indicates the actual expected link flow while the *y*-axis represents the estimated expected link flow. In Fig.3 (b) the *x*-axis indicates the actual standard deviation of link flow while the *y*-axis represents that of the estimated link flow. The estimated results are analytically produced by the StrUE model based on the total demand distribution after the convergence criterion has been met. The estimated expected link flows and the corresponding actual expected link flows are sorted from the smallest to the largest, so are the estimated standard deviations of link flows. The R squared value of the results in Fig.3 (a) is equal to 0.9837 while in Fig.3 (b) the R squared value is 0.936. Despite the R squared value in Fig.3 (b) is slightly smaller than that in Fig.3 (a), both R squared values are close to 1. This indicates that the estimated results produced by the StrUE based on estimated demand distribution closely approximate the actual values, the LSStrUE is capable of providing a robust estimation.



(a) (b)

FIGURE 3 (a) The estimated and actual expected link flow comparison, estimated results are produced by the StrUE model based on estimated demand distribution; (b) The estimated and actual standard deviation of link flow comparison, estimated results are produced by the StrUE model based on estimated demand distribution.

4 CONCLUSION:

This paper proposes a methodological framework (LSStrUE) to estimate the mean and standard deviation of the total travel demand from observed link flows. A least squares method is proposed to estimate the total demand distribution. A bi-level programming method is also included to reduce the impact of biased initial estimates. A numerical analysis is conducted on a test network, and results for both the system level and the link level demonstrated robust performance of the LSStrUE framework. In the numerical experiment, the estimated mean and standard deviation of the total demand converged to the desired values regardless of the initial estimates after 2 or 3 iterations. Similarly, the link level analysis produced R squared values of 0.9916 and 0.937, for the expected value and standard deviation of link flow, respectively. Based on the results, the estimated link flow distribution closely approximates the actual link flow distribution, suggesting that the LSStrUE can calibrate the total demand effectively and efficiently. Future research will investigate the use of the covariance of loop counts, that is, each element in the covariance matrix should be taken into consideration in the least squares method. This can potentially provide much more information than only mean and variance of link flows. Furthermore, the O-D demand may be assumed to follow a multivariate lognormal distribution, in this way the O-D demand is no longer aggregated as was the case with uni-variate lognormal distribution, possibly providing the covariance matrix of the link flows. Since the OD estimation problem is a combination of a statistical optimization model and a traffic assignment model, an improvement in either process warrants further research.

5 ACKNOWLEDGEMENTS:

This research was in part made possible by the generous support of NICTA. NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy and the Australian Research Council through the ICT Centre of Excellence program.

6 REFERENCES:

BAR-GERA, H. 2012. Transportation test problems of Sioux falls network. *website* http://www.bgu.ac.il/~bargera/tntp/.

BELL, M. G. H. 1983. The Estimation of an Origin-Destination Matrix from Traffic Counts. *Transportation Science,* 17**,** 198-217.

BELL, M. G. H. 1991. The estimation of origin-destination matrices by constrained generalised least squares. *Transportation Research Part B: Methodological,* 25**,** 13-22.

BELLEI, G., GENTILE, G., MESCHINI, L. & PAPOLA, N. 2006. A demand model with departure time choice for within-day dynamic traffic assignment. *European Journal of Operational Research,* 175**,** 1557-1576.

BIERLAIRE, M. & CRITTIN, F. 2004. An efficient algorithm for real-time estimation and prediction of dynamic OD tables. *Operations Research,* 52**,** 116-127.

BRILON, W., GEISTEFELDT, J. & REGLER, M. Reliability of freeway traffic flow: a stochastic concept of capacity. Proceedings of the 16th International symposium on transportation and traffic theory, 2005.

CASCETTA, E. 1984. Estimation of trip matrices from traffic counts and survey data: A generalized least squares estimator. *Transportation Research Part B: Methodological,* 18**,** 289-299.

CREMER, M. & KELLER, H. 1987. A new class of dynamic methods for the identification of origin-destination flows. *Transportation Research Part B: Methodological,* 21**,** 117-132.

DIXIT, V., GARDNER, L. M. & WALLER, S. T. Strategic User Equilibrium Assignment Under Trip Variability. Transportation Research Board 92nd Annual Meeting, 2013.

DUELL, M., GARDNER, L., DIXIT, V. & WALLER, S. T. Evaluation of a strategic road pricing scheme accounting for day-to-day and long term demand uncertainty. Transportation Research Board 93rd Annual Meeting, 2014.

FISK, C. S. 1988. On combining maximum entropy trip matrix estimation with user optimal assignment. *Transportation Research Part B: Methodological,* 22**,** 69-73.

FREDERIX, R., VITI, F. & TAMPÈRE, C. M. J. 2011. Dynamic origin–destination estimation in congested networks: theoretical findings and implications in practice. *Transportmetrica A: Transport Science,* 9**,** 494-513.

HAZELTON, M. L. 2003. Some comments on origin–destination matrix estimation. *Transportation Research Part A: Policy and Practice,* 37**,** 811-822.

KIM, J., KURAUCHI, F. & UNO, N. 2009. Analysis of variation in demand and performance of urban expressways using dynamic path flow estimation. *Transportmetrica,* 7**,** 63-84.

LO, H., ZHANG, N. & LAM, W. 1999. Decomposition algorithm for statistical estimation of OD matrix with random link choice proportions from traffic counts. *Transportation Research Part B: Methodological,* 33**,** 369-385.

LO, H., ZHANG, N. & LAM, W. H. 1996. Estimation of an origin-destination matrix with random link choice proportions: a statistical approach. *Transportation Research Part B: Methodological,* 30**,** 309-324.
SPIESS, H. 1987. A maximum likelihood model for estimating origin-destination matrices. *Transportation Research Part B: Methodological,* 21**,** 395-412.
U.S 1964. Department of Commerce. Bureau of Public Roads. Traffic Assignment Manual Urban Planning Division, US Department of Commerce, Washington, DC.
UCHIDA, T. & IIDA, Y. Risk assignment. A new traffic assignment model considering the risk of travel time variation. Proceedings of the 12th International Symposium on the Theory of Traffic Flow and Transportation, Jul 21 - 23 1993, 1993 Berkeley, CA, United states. Publ by Elsevier Science Publishers B.V., 89-89.
WEN, T., GARDNER, L., DIXIT, V., DUELL, M. & WALLER, S. T. A Strategic User Equilibrium Model Incorporating Both Demand and Capacity 2 Uncertainty 3. Transportation Research Board 93rd Annual Meeting, 2014.
WU, X., MICHALOPOULOS, P. & LIU, H. X. 2010. Stochasticity of freeway operational capacity and chance-constrained ramp metering. *Transportation Research Part C: Emerging Technologies,* 18**,** 741-756.
YANG, H., SASAKI, T., IIDA, Y. & ASAKURA, Y. 1992. Estimation of origin-destination matrices from link traffic counts on congested networks. *Transportation Research Part B: Methodological,* 26**,** 417-434.

7 APPENDIX:

TABLE: Summary of notations.

N	Link (index) set
K_{RS}	Path set
p_n	Proportion of total demand on link n; $P=(p_1,\dots,p_n)$
t_n	Travel time on link n; $\mathbf{t} = (\dots, t_n, \dots)$
t_{nf}	Free flow travel time on link n
c_k^{rs}	Travel time on path k connecting O-D pair r-s; $\mathbf{c}^{rs} = (\dots, c_k^{rs})$; $\mathbf{c} = (\dots, c^{rs}, \dots)$
q_{rs}	Fraction of total trips that are between O-D pair r-s; $\sum_{\forall rs} q_{rs} = 1$
T	Random variable for total trips with probability distribution $g(T)$
$g(T)$	Lognormal probability density function of the total demand
v_k^{rs}	Proportion of flow on path k, connecting O-D pair r-s
C_n	The capacity on link n
$\delta_{n,k}^{rs}$	Indicator variable $\delta_{a,k}^{rs} = \begin{cases} 1 & if\ n\ is\ included\ in\ path\ k \\ 0 & otherwise \end{cases}$
μ	Mean of the corresponding normal distribution, also called the location parameter for the lognormal distribution.
σ^2	Variance (= the second central moment) of the corresponding normal distribution, also called the scale parameter for the lognormal distribution.
m_T	The expected total demand
v_T	Variance of the lognormal distribution
m_n	The expected link flow on link n.
v_n	The variance of link flow on link n.
s_T	The standard deviation of the total demand
s_n	The standard deviation of link flow on link n.
cov	Coefficient of variation