Unbiased estimation methods of nonlinear transport models based on linearly projected data

Wai Wong[a] and S.C. Wong[b]

[a]*Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, MI, United States*
[b]*Department of Civil Engineering, The University of Hong Kong, Pokfulam, Hong Kong*

Abstract

Linear data projection is widely used for unbiased traffic data estimation. Nevertheless, recent studies have proven that direct model estimation based on linearly projected data that ignores the scaling factor variability may lead to systematically biased parameters. Adjustment factors were derived for a generalised multivariate polynomial (GMP) function with fixed exponents to remove such biases. However, the methods have not been extended to generic nonlinear transport models necessitating nonlinear regressions. This paper scrutinises the mechanism of systematic data point distortion resulting from linear data projection and identifies the practical difficulties of the adjustment factor approach to other nonlinear models. To reduce such biases in nonlinear transport models, a generic mean value restoration (MVR) method, requiring only the first two moments of the scaling factor, and an extended MVR (EMVR) method, further incorporating higher order moments by assuming a scaling factor distribution, are proposed. Simulation studies are conducted for both GMP functions with relaxed exponents and multivariate exponential decay functions, which are the most commonly adopted nonlinear functions for modeling traffic flow, to examine the effectiveness and robustness of the proposed methods for recovering the assumed true model parameters. Results reveal that the EMVR method generally can achieve higher level of accuracy.

Keywords: Big data era; Linear data projection; Systematic bias; Nonlinear transport models; traffic flow models

1. Introduction

Accurate and reliable transport models are of crucial importance in transportation studies because they help researchers and practitioners develop a better understanding of the interaction between transportation infrastructure, vehicles and road users. Moreover, accurate models are also necessary to ensure the effective and appropriate implementation of urban transport planning, traffic management and control measures.

Accurate transport model estimations require accurate traffic data. With the advent of various high-tech devices, the accuracy and efficiency of traffic data collection has been substantially improved over the past several decades. Nevertheless, various limitations of these detectors and sensors make it impractical to collect data from the entire network. On-road fixed detectors, such as inductive loop detectors, can collect traffic data at an acceptable

level of accuracy, but their high installation and maintenance costs impede their ubiquitous deployment (Herrera and Bayen 2010; Herrera et al. 2010). Thus, their coverage is usually limited to a subset of links (Caceres et al. 2012). A vehicle re-identification system can measure the travel time of a vehicle across a link by matching the vehicle's signature as it passes sensors at the two ends of that link (Kwong et al. 2009). Any available technological utilities that can determine vehicle identities, such as radio frequency identification transponders (Wright and Dahlgren 2001; Ban et al. 2010), licence plate recognition systems (Herrera et al. 2010) and wireless magnetic sensors (Kwong et al. 2009) can be readily adopted for such schemes. However, in addition to the high installation and implementation costs, the risk of compromising privacy issues is a major obstacle to deploying these schemes over an entire network. The cellular systems introduced a decade ago (Bolla and Davoli 2000; Ygnace and Drane 2001; Zhao 2000) provide a potential solution to the cost and coverage problems (Herrera et al. 2010). Nonetheless, their application is discouraged or even prohibited in many countries, because using mobile phones while driving disrupts drivers' attention (Liang, Reyes, and Lee 2007). The advancement of global positioning systems (GPSs) offers another promising means of traffic data collection using probe vehicles covering almost the entire network at a relatively low cost (Miwa et al. 2013). However, GPS data retrieved from vehicle fleets (e.g., FedEx, UPS, or taxis) (Moore et al. 2001; Bertini and Tantiyanugulchai 2004; Schwarzenegger et al. 2009; Wong et al. 2014) could possibly pose bias problems due to their unique operational or travel patterns. Furthermore, the extra capital and installation costs of GPS trackers coupled with the potential privacy issues hinder the application of such systems on a global scale.

With the aim of overcoming the practical issues obstructing direct and accurate traffic data measurement, mathematical techniques such as data scaling, filtering and sampling methods have been developed and leveraged for accurate traffic data estimations. Linear data projection is a ubiquitous and highly transferrable data scaling method that infers unobservable traffic data by projecting observable traffic data using the mean of a set of sampled scaling factors, which are the ratios bridging the unobservable and observable data. We foresee that in the big data era, more and more connected vehicle (or probe vehicle) data will be available and we will often need such a data scaling method to scale up observable traffic data. Moreover, because many transport models heavily rely on a variety of sources of data for estimations, calibrations and validations, it is a timely data scaling method that can compatibly fuse data from different sources for unbiased traffic data estimation. The scaling factors can be defined differently to fit the specified physical context. The unobservable traffic data can usually be expressed as a linear combination of the scaling factors and the observable traffic data. Due to the complexity and stochastic nature of transportation systems, scaling factors are usually random variables rather than constants, and are thus assumed to follow distributions. Depending on the sampling approach, different types of variability, such as spatial or temporal variability, can be measured by the scaling factor variance. In practice, because the value of each scaling factor is unknown, they are replaced by the estimated scaling factor mean in a linear data projection to provide an unbiased estimator for the unobservable traffic data.

Linear data projection has been used in transportation studies to estimate unobservable traffic data. For instance, the unbiased estimator of an hourly total traffic flow across a link that does not have a detector can be obtained using a data scaling method. Given that the total traffic flow in a network is only observable at the subset of links with detectors, and that the probe vehicle flow is observable at every link in the network, the total-traffic-to-probe-vehicle ratio at a link can be defined as the scaling factor in this context. Scaling factors can be sampled from the subset of links with detectors due to the data availability of both total traffic and probe vehicle flows. Because of the heterogeneities of land-use patterns and stochasticity, the scaling factors sampled at different locations are usually different, but they can be assumed to follow a distribution over the network due to geographical proximity. In such cases, the scaling factor variance measures spatial variability across the network. As the scaling factor mean is the most probable observed traffic composition ratio across the network in the long run, if its value is 100 and the hourly probe vehicle flow on the link of interest is 10 veh/h, then the unbiased estimator for total hourly traffic on this link can be evaluated by their product; that is, 1000 veh/h. This method has been employed in many studies to estimate traffic flow (Wong and Wong 2015, 2016a, 2016b, 2016c). Another example that estimates hourly equivalent traffic flow across a road based on linear data projection uses a passenger car unit (PCU) as the scaling factor. Due to the varying traffic composition across time, a PCU is not necessarily static (Chandra, Kumar, and Sikdar 1995). Given that a road is outfitted with a detector counting vehicles 24 hours a day, the hourly equivalent traffic flow across that road for a specific hour is the product of the traffic flow and the PCU value in that hour. However, this PCU value may not be known, because due to budget constraints, surveyors are usually only sent on site to determine vehicle types for several hours a day on certain days in a year, according to a strategic sampling plan. In such cases, the product of an hourly traffic flow and the sampled PCU mean offers an unbiased hourly equivalent traffic flow and the PCU variance measures its temporal variability. Similarly, linear data projection can be leveraged to extrapolating demand data such as projecting travel diary surveys from a sample of households to the zonal level and expanding transit travel distribution from sampled vehicles outfitted with automatic passenger counting system.

Apart from estimating unobservable traffic flows and travel demand, linear data projection can be applied in accident analysis to estimate exposure measures such as vehicle mileage and travel time, which are the typical explanatory variables accounting for the variations in crash counts at specific sites. The corresponding sensitivity parameter in relation to these variables is known as the crash risk. As an example, the annual vehicle mileage on a road is expressed as the product of its annual traffic volume and road length. Because of limited budgets and resources, detailed traffic data collection throughout a year is usually only possible for a subset of roads, and short-term (e.g., one weekday) traffic data are recorded for the remaining roads. If only short-term traffic volume is available for the road of interest, its unbiased vehicle mileage can be estimated via a linear data projection with the annual-to-short-term-traffic-volume ratios sampled from the nearby roads with full-year traffic data as the selected scaling factors. The scaling factor variance quantifies the spatial

variability. Similarly, Meng et al. (2017b) estimated the time exposure based on linear data projection in their recent study modelling multiple-vehicle crash frequency.

With the recent emergence and growing popularity of high-tech devices such as cellular systems and GPSs, linear data projection could be the key to fusing data from different sources and estimating the unobservable but necessary traffic data for different macroscopic transport model estimations. Macroscopic transport models have rapidly gained momentum in recent decades due to their tremendous potential benefits in various applications such as initial land-use planning (Ho and Wong 2007; Yin et al. 2013), urban network traffic control (Daganzo 2007; Aboudolas and Geroliminis 2013; Geroliminis, Haddad, and Ramezani 2013; Zhong et al. 2017) , road pricing schemes (Geroliminis and Levinson 2009; Zheng et al. 2012; Zheng, Rerat, Geroliminis 2016) and dynamic modelling and control of taxi services (Ramezani and Nourinejad 2017). These models can be generally categorised as macroscopic cost flow (MCF) functions and macroscopic fundamental diagrams (MFDs). MCFs are commonly used for static analyses whereas MFDs are usually used for dynamic analyses. Known and accurate macroscopic models are the prerequisite for beneficial applications, and accurate traffic data collected from the entire network are the essential ingredients for their estimation. However, direct measurements of these traffic data are normally not feasible using currently available devices due to the various aforementioned limitations. Linear data projection offers a practical and useful framework to provide unbiased estimators for those unobservable traffic data based on data collected from different sources or methods. In their study of the existence of MFD, Geroliminis and Daganzo (2008) adopted linear data projection to infer accumulation (a proxy measure for traffic density) in a network using a total-traffic-to-occupied-taxi ratio constituted by data retrieved from both detectors and occupied GPS taxis as the scaling factor. For MCFs, Wong and Wong (2015, 2016a, 2016b) estimated macroscopic Bureau of Public Road functions for $1\ \text{km} \times 1\ \text{km}$ sampled networks in Hong Kong based on real-world data acquired from counting stations and GPS-equipped taxis. The total hourly traffic flows entering the sampled networks were estimated by linear data projection with total-traffic-to-occupied-taxi ratio as the chosen scaling factor. The scaling factor variability could be attributed to different factors such as land-use heterogeneity of different lots (Meng et al., 2017a).

As linear data projection offers unbiased estimators for unobservable traffic data, it is intuitive to estimate models based on the linearly projected data. However, the implications – and most importantly, the corresponding effects on the estimated parameters – have been largely ignored in the field. Estimating models directly based on linearly projected data is equivalent to ignoring the effects of scaling factor variability. Ignoring spatial variability implies a uniform scaling factor value across space. Similarly, a constant scaling factor is assumed if temporal variability is ignored. Wong and Wong (2015) recently proved that systematic bias is generically introduced in the parameters of models estimated from linearly projected data, regardless of the scaling factor distribution and the form of the model to be estimated, as long as the scaling factor is subject to variability and the model is a nonlinear function of the scaling factor. The authors, however, only examined generalised multivariate polynomial (GMP) functions with fixed exponents that are essentially linear regression

models. Analytical expressions for quantifying the extents of biases and adjustment factors for reducing such biases were derived for this family of functions. Wong and Wong (2016b) later proved that heteroscedasticity is inherently introduced when linear data projection is adopted, regardless of the form of model to be estimated. This will certainly lead to biased standard error estimation. The authors similarly focused only on the family of GMP functions with fixed exponents. An analytical distribution free method and equivalent scaling factor methods were proposed to accurately estimate the model parameters and their standard errors. In the case studies of estimating MCFs, the usually ignored spatial variability of the scaling factor was incorporated to ensure accurate estimated parameters and standard errors. However, it must be stressed that these methods are applicable only to GMP functions with fixed exponents, which are linear regression models although they can be nonlinear in shape. Methods for reducing such biases have not been extended to the territory of generic nonlinear transport models requiring nonlinear regressions. Despite focusing on network flow and density estimations, Ambühl and Menendez (2016) and Du, Rakha and Gayah (2016), discussed and addressed the unrealistic assumption of a uniform scaling factor (or probe vehicle penetration rate) across the entire network that is usually assumed in MFD estimations. Ambühl and Menendez (2016) proposed a data fusion algorithm to estimate network flows and densities based on probe vehicle data and high-resolution loop detector data. Simulation results revealed that the data fusion algorithm could always reduce the error in data points compared to estimations based on only one data source. Moreover, if the probe vehicle penetration rate was not too low and only median error was considered, the algorithm was found to be efficient within the investigated range of spatial heterogeneity of the probe vehicles. With the distribution of probe vehicle OD pairs, a novel algorithm based on k-means clustering analysis was proposed by Du, Rakha and Gayah (2016) to estimate the unobservable probe vehicle penetration rates of individual OD pairs. Using these estimated rates, equivalent average probe vehicle penetration rates (i.e., harmonic means of the estimated OD pair probe vehicle penetration rates) could be evaluated for network flow and density estimations. Simulation results showed that equivalent average probe vehicle penetration rates can capture the spatial heterogeneity and provided more accurate flow and density estimates compared to cases based on constant probe vehicle penetration rates. Nevertheless, these algorithms focusing on data point estimations typically require relatively detailed data, such as high-resolution loop detector data and distributions of probe vehicle OD pairs, to incorporate the lost scaling factor variability into each of the data points. These algorithms were designed for data point estimations for MFDs, so they are usually not directly transferrable and cannot be easily extrapolated to other situations. Modifications of the algorithms are usually necessary to solve other physical problems. Rather than focusing on data point estimations, this paper aims to develop generic unbiased model estimation methods for nonlinear transport models, based on data estimated from highly transferrable linear data projection.

Many transport models do not belong to the category of GMP functions with fixed exponents, so this study serves as a natural extension of the work of Wong and Wong (2015). Methods are proposed that can reduce the systematic bias arising from the ignorance of scaling factor variability in estimating generic nonlinear transport models that necessitate

nonlinear regressions. Essentially, the size of the model set that can be handled by the methods proposed in this paper is much greater than that in Wong and Wong (2015). In this paper, the mechanism of systematic data point distortion arising from linear data projection is first examined. This provides insights into the direction of biases in parameter estimation from linearly projected data. The inexistence or derivation complexity of simple closed-form adjustment factors and the model specification error induced by linear data projection are identified as the main practical difficulties of the adjustment factor approach adopted in Wong and Wong (2015) for nonlinear transport models. The implication is that the adjustment factor approach is not a feasible solution to cases considering nonlinear regression models. Instead, a generic mean value restoration (MVR) method, requiring only the first two moments of the scaling factor, and an extended MVR (EMVR) method, further incorporating higher order moments by assuming a scaling factor distribution, are proposed. Comprehensive simulation studies with different settings are conducted for both GMP functions with relaxed exponents and multivariate exponential decay functions, to demonstrate and compare the effectiveness and robustness of the proposed methods for recovering the assumed true model parameters. GMP functions with relaxed exponents and multivariate exponential decay functions are the most commonly adopted functional forms for modeling traffic flow relationships such as cost-flow functions and speed-density relationships in their link- and area-based forms. The results show that the EMVR method generally can provide more accurate estimated parameters.

The remainder of the paper is structured as follows. Section 2 explains the mechanism of the systematic distortion of data points led by linear data projection. The practical difficulties identified in Section 3 show that the adjustment factor approach may not be a feasible direction for generic nonlinear transport models that require nonlinear regressions. The MVR and EMVR methods are proposed in Section 4 to remove the embedded systematic biases for all nonlinear models. Section 5 presents the comprehensive simulation studies assessing the effectiveness and robustness of the proposed methods. Simulation results reveal that the EMVR method generally outperforms the MVR method. The final section concludes with the major findings of the paper and discusses the potential directions for future research.

2. The silent distortion of the truth by linear data projection

This section approximates the expectation function of linearly projected data using a Taylor series expansion. To simplify the explanation, quadratic approximation is used to identify the two major factors governing the mechanism of the systematic distortion of data points. A thorough understanding of the mechanism is important because it provides clues to the solution of the problem.

2.1 Approximated expectation functions of linearly projected data

Let $G(\boldsymbol{\beta}; z)$: $\mathbb{R}^m \rightarrow \mathbb{R}$ be a highly differentiable function of any form, where $\boldsymbol{\beta}$ is the vector of the model parameters; the independent variable z is constituted by linear combinations of a set of observable independent variables and a set of scaling factors, i.e.,

$z = \sum_{i=1}^{m} f_i x_i, \forall i \in \{1, 2, \dots, m\}$; x_i is the observable independent variables; f_i is the scaling factor of x_i that is assumed to follow any distribution with mean $\bar{f}$ and variance ${\sigma_f}^2$; and m is the number of terms constructing the quantity z.

In most cases, collecting data for z is impossible or impractically expensive. In contrast, the observable independent variable x_i can usually be collected using much cheaper methods. The scaling factor f_i for each x_i is assumed to follow a distribution. In theory, although it can be assumed to follow any distribution, the properties of the chosen distribution should be in accordance with the conditions of the given situation. For example, if the scaling factor is a non-negative random variable with a relatively lower frequency at high values, then lognormal distribution can usually be assumed to be the candidate distribution. The first and second moments of the assumed distribution can be estimated based on a set of scaling factors sampled from independent sources under similar conditions. In practice, the value of each individual f_i, however, is unknown, thus z can be estimated by using linear data projection in which each individual f_i is replaced by the estimated scaling factor mean $\bar{f}$, i.e., $\bar{z} = \sum_{i=1}^{m} \bar{f} x_i$. Because taking the expectation of z arrives at the same expression, i.e., $E(z) = \bar{z}$, the linearly projected data are unbiased estimates of z in the long run. Nevertheless, the expectation of $G(\boldsymbol{\beta}; z)$ is not only dependent on the scaling factor mean $\bar{f}$, but also its variance ${\sigma_f}^2$, skewness S_f, kurtosis K_f and higher order moments of the scaling factor. As shown in Eq. (1), the expectation function of the linear projected data can be approximated up to different orders. The detailed derivation of Eq. (1) is provided in Appendix A.

$$E[G(\boldsymbol{\beta}; z)] = G(\boldsymbol{\beta}; \bar{\boldsymbol{f}}) + \frac{1}{2!}{\sigma_f}^2 \sum_{i=1}^{m} \frac{\partial^2 G(\boldsymbol{\beta}; \bar{\boldsymbol{f}})}{\partial {f_i}^2} + \frac{1}{3!} S_f \sum_{i=1}^{m} \frac{\partial^3 G(\boldsymbol{\beta}; \bar{\boldsymbol{f}})}{\partial {f_i}^3} + \frac{1}{4!} K_f \sum_{i=1}^{m} \frac{\partial^4 G(\boldsymbol{\beta}; \bar{\boldsymbol{f}})}{\partial {f_i}^4} + \cdots \tag{1}$$

Define $E_r[G(\boldsymbol{\beta}; z)]$ to be the rth order approximation of the expectation function obtained by truncating all the terms behind the rth term in Eq. (1), and $RE_r[G(\boldsymbol{\beta}; z)]$ to be the sum of the corresponding truncated terms, $\forall r \in \mathbb{N}^+$. Thus, $E[G(\boldsymbol{\beta}; z)] = E_r[G(\boldsymbol{\beta}; z)] + RE_r[G(\boldsymbol{\beta}; z)]$. As shown in Eq. (2)–(5), the linear, quadratic, cubic and quartic approximations of the expectation function are represented by $E_1[G(\boldsymbol{\beta}; z)]$, $E_2[G(\boldsymbol{\beta}; z)]$, $E_3[G(\boldsymbol{\beta}; z)]$ and $E_4[G(\boldsymbol{\beta}; z)]$, respectively.

$$E_1[G(\boldsymbol{\beta}; z)] = G(\boldsymbol{\beta}; \bar{\boldsymbol{f}}) \tag{2}$$

$$E_2[G(\boldsymbol{\beta}; z)] = G(\boldsymbol{\beta}; \bar{\boldsymbol{f}}) + \frac{1}{2!}{\sigma_f}^2 \sum_{i=1}^{m} \frac{\partial^2 G(\boldsymbol{\beta}; \bar{\boldsymbol{f}})}{\partial {f_i}^2} \tag{3}$$

$$E_3[G(\boldsymbol{\beta}; z)] = G(\boldsymbol{\beta}; \bar{\boldsymbol{f}}) + \frac{1}{2!}{\sigma_f}^2 \sum_{i=1}^{m} \frac{\partial^2 G(\boldsymbol{\beta}; \bar{\boldsymbol{f}})}{\partial {f_i}^2} + \frac{1}{3!} S_f \sum_{i=1}^{m} \frac{\partial^3 G(\boldsymbol{\beta}; \bar{\boldsymbol{f}})}{\partial {f_i}^3} \tag{4}$$

$$E_4[G(\boldsymbol{\beta};z)] = G(\boldsymbol{\beta};\bar{\boldsymbol{f}}) + \frac{1}{2!}{\sigma_f}^2 \sum_{i=1}^{m} \frac{\partial^2 G(\boldsymbol{\beta};\bar{\boldsymbol{f}})}{\partial {f_i}^2} + \frac{1}{3!} S_f \sum_{i=1}^{m} \frac{\partial^3 G(\boldsymbol{\beta};\bar{\boldsymbol{f}})}{\partial {f_i}^3} + \frac{1}{4!} K_f \sum_{i=1}^{m} \frac{\partial^4 G(\boldsymbol{\beta};\bar{\boldsymbol{f}})}{\partial {f_i}^4} \tag{5}$$

Moreover, $E_r[G(\boldsymbol{\beta};z)]$ is an approximation but not exact expectation function if and only if $RE_r[G(\boldsymbol{\beta};z)]$ is nonzero. Most importantly, it should be noted that although $\boldsymbol{f}$ is replaced by $\bar{\boldsymbol{f}}$ in the linear data projection, the linear approximation, $G(\boldsymbol{\beta};\bar{\boldsymbol{f}})$, is identical to the true model, $G(\boldsymbol{\beta};z)$, because their shapes are determined by the same model form along with the same set of model parameters, $\boldsymbol{\beta}$. In other words, the expectation function of the linearly projected data, $E[G(\boldsymbol{\beta};z)]$, differs from the true model, $G(\boldsymbol{\beta};z)$, if and only if $RE_1[G(\boldsymbol{\beta};z)]$ is not zero.

2.2 Systematic data points distortion mechanism

In the previous subsection, it is shown that $RE_1[G(\boldsymbol{\beta};z)] = E[G(\boldsymbol{\beta};z)] - G(\boldsymbol{\beta};\bar{\boldsymbol{f}})$ is the vertical difference between the expectation function of the linearly projected data and the true model. Because $\boldsymbol{f}$ is replaced by $\bar{\boldsymbol{f}}$ in linear data projection, data points must be shifted horizontally and systematically such that the dislocations result in the vertical difference. This subsection aims to explain the systematic distortion by introducing two major factors governing such a mechanism.

In general, the contribution of each term on the right-hand side of Eq. (1) to $E[G(\boldsymbol{\beta};z)]$ usually decreases with its term order. Therefore, for the sake of simplicity, the expectation function can be approximated by the quadratic approximation (i.e., $E[G(\boldsymbol{\beta};z)] \cong E_2[G(\boldsymbol{\beta};z)]$). The approximate vertical difference shown in Eq. (6) reveals that whether the expectation function is above or below the true model is mainly dependent on the convexity of G over the space of the scaling factor. In particular, if G is linear, the vertical difference should be equal to zero.

$$E[G(\boldsymbol{\beta};z)] - G(\boldsymbol{\beta};\bar{\boldsymbol{f}}) \cong \frac{1}{2!}{\sigma_f}^2 \sum_{i=1}^{m} \frac{\partial^2 G(\boldsymbol{\beta};\bar{\boldsymbol{f}})}{\partial {f_i}^2} \begin{cases} > 0, if\ G\ is\ convex. \\ = 0, if G\ is\ linear. \\ < 0, if\ G\ is\ concave. \end{cases} \tag{6}$$

We now consider that $G(\boldsymbol{\beta};\bar{z})$ is strictly increasing and convex, $\forall \bar{z} \in (\bar{z}_a, \bar{z}_b)$, as illustrated in Fig. 1. $\exists U = \{(z_1, y_1), (z_2, y_2), \dots, (z_l, y_l), \dots, (z_L, y_L)\}$, where U is one of the many sets of unobservable ordered pairs that collects data points lying on $G(\boldsymbol{\beta};\bar{z})$ and L is the number of ordered pairs in U. On most occasions, these data points are not observable because the value of each individual f_i is unknown.

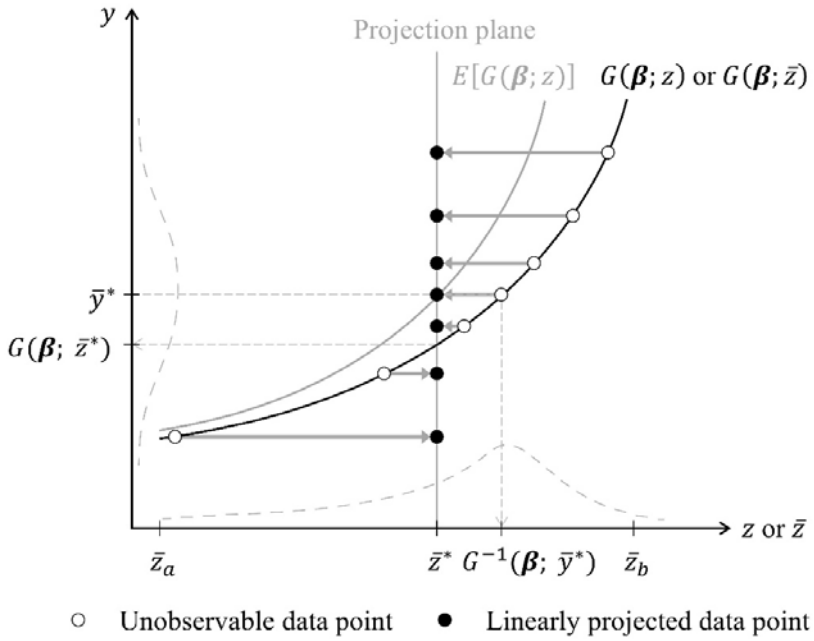

Fig. 1. The mechanism of systematic distortion of data points associated with the virtual projection plane located at $\bar{z} = \bar{z}^*$

Nevertheless, the unbiased estimators for z can be obtained by adopting linear data projection, in which all of the data points in U move horizontally while retaining their y values, and are projected on the virtual projection plane located at $\bar{z} = \bar{z}^*$, that is $E(z_l) = \bar{z}^*, \forall l \in \{1, 2, \dots, L\}$ and $\bar{z}^* \in (\bar{z}_a, \bar{z}_b)$. The ordered pairs set for the observable linearly projected data, $O = \{(\bar{z}^*, y_1), (\bar{z}^*, y_2), \dots, (\bar{z}^*, y_l), \dots, (\bar{z}^*, y_L)\}$, is subsequently obtained. The mean value of y for these linearly projected points is $\bar{y}^* = \sum_{l=1}^{L} y_l / L$. The expectation function, $E[G(\boldsymbol{\beta}; z)]$, must pass through the point $(\bar{z}^*, \bar{y}^*)$, which aggregates all the information of these data points. Because $G(\boldsymbol{\beta}; \bar{z})$ is convex, $\forall \bar{z} \in (\bar{z}_a, \bar{z}_b)$, $(\bar{z}^*, \bar{y}^*)$ must be above the true model and the vertical difference between them is exactly equal to $RE_1[G(\boldsymbol{\beta}; \bar{z}^*)]$ or can be approximated by Eq. (6). Moreover, $G(\boldsymbol{\beta}; \bar{z})$ is strictly increasing, $\forall \bar{z} \in (\bar{z}_a, \bar{z}_b)$, thus the inverse of G at $y = \bar{y}^*$ is greater than $\bar{z}^*$ (i.e., $G^{-1}(\boldsymbol{\beta}; \bar{y}^*) > \bar{z}^*$). That means although some data points shift to the left as others shift to the right upon linear data projection, the net movement of shifting is towards the left such that $(\bar{z}^*, \bar{y}^*)$ is above the true model. Therefore, the systematic distortion is mainly governed by the first and the second derivative of the true model and the resultant dislocation of $(\bar{z}^*, \bar{y}^*)$ is the source of biases introduced in the estimated parameters. Similarly, the net shifting movements of data points associated with any projection plane located at any segment of $G(\boldsymbol{\beta}; \bar{z})$ with different first and second derivatives combinations are summarised in Fig. 2. In particular, if $G''(\boldsymbol{\beta}; \bar{\boldsymbol{f}})$ is zero, $\forall \bar{z} \in (\bar{z}_a, \bar{z}_b)$, there should be no net horizontal movement of data points no matter if $G(\boldsymbol{\beta}; \bar{z})$ is increasing or decreasing for the given interval. These findings are important because they can be used to predict the directions of biases of the estimated parameters (as illustrated in Section 5.3). Linear data projection offers us the unbiased estimator for z, but it silently contorts the truth and what we can observe from the

scatter plot of the linearly projected data does not fully reflect reality. Thus, our research aim is to establish methods that can restore the true relationship of any nonlinear transport models.

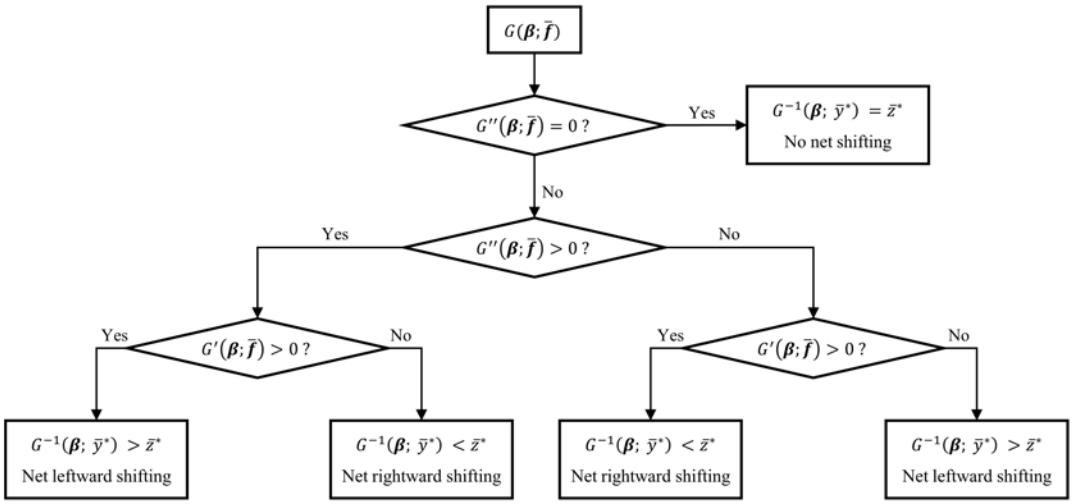

Fig. 2. Summary of the net horizontal movements of data points associated with any virtual projection plane located at any segment of G with different first and second derivatives combinations

3. Practical difficulties of the adjustment factor approach to generic nonlinear transport models

The objective of the least squares method in model estimation is to search for a set of model parameters that can minimise the sum of the squared residuals of the observed data for a given model. For model estimations with linear projected data, Wong and Wong (2015) generically proved that the estimated parameters that can minimise this difference can be systematically biased if the model is a nonlinear function of the scaling factor and the scaling factor is subject to variability. The authors then focused the discussion on GMP functions with fixed exponents, which are linear regression models. Adjustment factors for reducing such biases and analytical expressions for quantifying the biases were derived for this family of functions. However, whether such an adjustment factor approach is also suitable for generic nonlinear transport models remains unknown. This section presents some practical difficulties of using the adjustment factor approach in these cases, suggesting that such an approach may not be a feasible solution for generic nonlinear transport models.

3.1 Inexistence or derivation complexity of simple closed-form adjustment factors

A GMP function with fixed exponents was examined in studies by Wong and Wong (2015, 2016b). Although the function is nonlinear in the independent variable, it is, in fact, a linear regression model because the exponents are fixed and the normal equations are linear in their parameters. The adjustment factors can be easily obtained by taking the expectation of the solution of the normal equations. The derived adjustment factors capture the

information of the ignored higher order scaling factor moments that contribute to the mean values of dependent variables.

However, for other nonlinear transport models that require nonlinear regressions, the normal equations are nonlinear in parameters because at least one of the partial derivatives of their expectation functions with respect to the parameters is a function of the unknown parameters. Thus, these models are categorised as nonlinear regression models. Unlike linear regression models, their normal equations generally cannot be solved by a finite sequence of standard operations, but by iterative methods such as the Gauss-Newton method and the method of steepest descent. As there may be no simple closed-form solution for the nonlinear normal equations, simple closed-form adjustment factors may not exist. Even if they do exist, their derivations could be highly complicated and cumbersome.

3.2 Model specification error induced by linear data projection

In addition to the ignoring of higher order scaling factor moments, model specification error induced by linear data projection is another source of embedded systematic biases. This error arises when the expectation function of linearly projected data and the original function possess different properties, and thus they belong to different families of functions. Under such circumstance, the estimated model parameters can never fully capture the resultant systematic distortion caused by linear data projection, and hence they can never perfectly fit the expectation function.

The possibility of introducing such model specification error can be demonstrated by the existence of a single example in which the model structure of the expectation function of the linearly projected data is changed due to linear data projection. To this end, we consider an exponential decay function as shown in Eq. (7):

$$y = a\exp\left(\frac{-z}{b}\right) = a\exp(\frac{-fx}{b}), \tag{7}$$

where a and b are the model parameters. The exponential decay function often adopted in the natural sciences is a member of the family of exponential functions that is uniquely characterised by the direct proportionality between an exponential function and its first derivative. In other words, the growth rate of an exponential function and the function itself are linked by a proportionality constant.

Assuming that each individual f is unknown and z only can be estimated by linear data projection (i.e., $\bar{z} = \bar{f}x$), the expectation function of the linearly projected data can be obtained by using Eq. (1):

$$E(y) = \left[1 + \frac{1}{2!}\frac{{\sigma_f}^2}{\bar{f}^2}\left(\frac{-\bar{z}}{b}\right)^2 + \frac{1}{3!}\frac{S_f}{\bar{f}^3}\left(\frac{-\bar{z}}{b}\right)^3 + \frac{1}{4!}\frac{K_f}{\bar{f}^4}\left(\frac{-\bar{z}}{b}\right)^4 + \cdots\right] a\exp(\frac{-\bar{z}}{b}) \tag{8}$$

The square bracket in Eq. (10) is a polynomial function of $\bar{z}$, which can be written as $p(\bar{z})$. Thus, Eq. (8) can be expressed as:

$$E(y) = p(\bar{z})\ a \exp(\frac{-\bar{z}}{b}) \tag{9}$$

Given that f is subject to variability (i.e., at least the second moment ${\sigma_f}^2 > 0$), $p'(\bar{z})$ is also a function of $\bar{z}$. The first derivative of $E(y)\ w.r.t.\ \bar{z}$ is given by Eq. (10):

$$\frac{d}{d\bar{z}}E(y) = \left[p'(\bar{z}) - \frac{1}{b}p(\bar{z})\right]\ a \exp(\frac{-\bar{z}}{b}) \tag{10}$$

Because the expression in the square brackets in Eq. (10), $[p'(\bar{z}) - p(\bar{z})/b]$, is a function of $\bar{z}$, the growth rate of the expectation function and the function are not linked by a proportionality constant. In other words, the model structure of the expectation function differs from that of the exponential decay function. The fact that the expectation function in this case does not belong to the family of exponential functions is demonstrated by contradiction. Thus, model specification error could possibly be introduced by linear data projection.

In such cases, biases are derived both from the ignored information of higher order scaling factor moments and the model specification error induced by linear data projection. If an adjustment factor approach is adopted, the derived adjustment factors should be able to account for these two effects simultaneously to remove the systematic biases. It seriously complicates the adjustment factor derivation and their existence could be doubtful. These practical difficulties suggest that the adjustment factor approach may not be feasible for our research question. A generic approach with a solid theoretical foundation that can cater for all nonlinear transport models is preferable.

4. MVR and EMVR methods

This section proposes the MVR and EMVR methods that can be used for unbiased estimations of all nonlinear transport models based on linearly projected data. The proposed methods restore the mean values of the dependent variable during model estimation by accounting for the average vertical difference between the true model and the expectation function of the linearly projected data, and thus reduce the systematic biases embedded in the estimated model parameters. The MVR method, which requires only the first and second moments of the scaling factor, is flexible. It can be widely applied in many situations because, on most occasions, only the first two moments of the scaling factor, but not its entire distribution, are known. In contrast, the EMVR method, which requires an assumed scaling factor distribution, can further capture higher order moments of the scaling factor and achieve higher level of accuracy of the estimated model parameters.

4.1. Formulation of the MVR method

The underlying principle of the MVR method is introduced in this subsection. In Section 2, the resultant dislocation of $(\bar{z}^*, \bar{y}^*)$ is identified as the major cause of systematic biases in the estimated parameters. $RE_1[G(\boldsymbol{\beta}; \bar{z}^*)]$ is the vertical difference between $\bar{y}^*$ and

the true model at $\bar{z} = \bar{z}^*$. Assuming that $E[G(\boldsymbol{\beta}; \bar{z})] \cong E_2[G(\boldsymbol{\beta}; \bar{z})]$, that is, $RE_2[G(\boldsymbol{\beta}; \bar{z})] \cong 0$, the average vertical difference at any $\bar{z} = \bar{z}^*$ (denoted by Δy in Fig. 3) can be approximated by Eq. (6), which is only dependent on the first two moments of the scaling factor. In practical terms, such an assumption is useful and important because, in most cases, only the first two moments are available. The central idea of the MVR method is to restore the dislocated point $(\bar{z}^*, \bar{y}^*)$ to a point lying on the true model, $G(\boldsymbol{\beta}; \bar{z})$, by moving it along the y-axis. To achieve this, for any $\bar{z} = \bar{z}^*$, the average vertical difference, Δy, is subtracted from each element in the range $\{y_1,\ y_2, \dots, y_l, \dots, y_L\}$, such that the new set of the range is given by $\{y_1 - \Delta y,\ y_2 - \Delta y, \dots, y_l - \Delta y, \dots, y_L - \Delta y\}$ as illustrated in Fig. 3. Their mean value is given by $\bar{y}^{**} = \sum_{l=1}^{L}(y_l - \Delta y)\ /L$. The restored point $(\bar{z}^*, \bar{y}^{**})$ should lie on the true model, $G(\boldsymbol{\beta}; \bar{z})$, or very close to it.

The MVR method is a regression analysis based on all N pairs of restored points across the entire range of the independent variable. The corresponding least squares function, S, can be expressed as Eq. (11):

$$S = \sum_{j=1}^{N}\left[(y_j - \Delta y_j) - G\left(\widehat{\boldsymbol{\beta}}; \bar{z}_j\right)\right]^2 \tag{11}$$

where $\widehat{\boldsymbol{\beta}}$ is the vector of parameters to be estimated and $\Delta y_j \cong \frac{1}{2!}{\sigma_f}^2 \sum_{i=1}^{m} \frac{\partial^2 G(\widehat{\boldsymbol{\beta}}; \bar{z}_j)}{\partial {f_i}^2}$. The best-fitted curve should pass through the mean value of the dependent variable at each $\bar{z}^*$. As Δy_j is also a function of $\widehat{\boldsymbol{\beta}}$, both the model and the vertical difference are estimated simultaneously during the regression. Using Eq. (3), the least square function, S, can be alternatively expressed as:

$$S = \sum_{j=1}^{N}\left\{y_j - E_2\left[G\left(\widehat{\boldsymbol{\beta}}; \bar{z}_j\right)\right]\right\}^2 \tag{12}$$

Thus, the MVR method is equivalent to direct model estimation of the linearly projected data based on the quadratic approximation of their expectation function using conventional nonlinear regression packages.

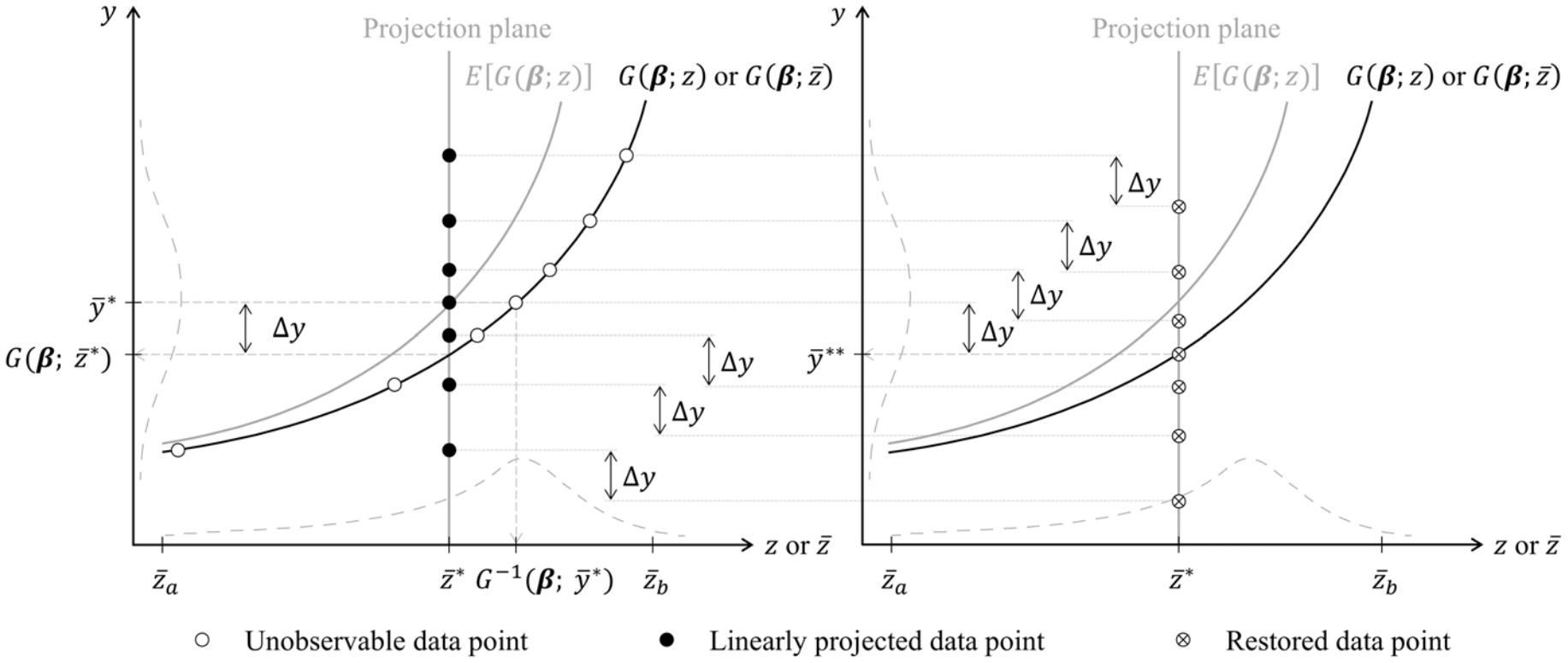

Fig. 3. Restoration of the dislocated point, $(\bar{z}^*, \bar{y}^*)$

4.2. Formulation of the EMVR method

The MVR method proposed in the previous subsection is based on an assumption that $RE_2[G(\boldsymbol{\beta};\bar{z})] \cong 0$ and $E[G(\boldsymbol{\beta};\bar{z})] \cong E_2[G(\boldsymbol{\beta};\bar{z})]$. In general, the assumption is roughly true and acceptable because $RE_2[G(\boldsymbol{\beta};\bar{z})]$ is much smaller than $E_2[G(\boldsymbol{\beta};\bar{z})]$, even if it is not equal to zero. However, if higher order moments can be incorporated in the model estimation procedures, the accuracy of the estimated parameters can be further improved. The EMVR method restores the mean values of the dislocated data points by capturing higher order moments of the scaling factor based on an assumed scaling factor distribution. The assumed scaling factor distribution should be in line with the physical properties of the scaling factor. For example, if the scaling factor is non-negative with lower probabilities at relatively higher values, a lognormal distribution can be an appropriate candidate distribution. In applications, a Kolmogorov-Smirnov or a chi-square goodness-of-fit test can be conducted to check the suitability of the specified distribution with statistical evidence. Depending on the required level of accuracy, a model can be estimated based on an rth order approximated expectation function using the required higher order moments of the scaling factor, where $r > 2$. Similar to Section 4.1, the least square function in this case can be formulated as shown in Eq. (13).

$$S = \sum_{j=1}^{N} \{y_j - E_r[G(\widehat{\boldsymbol{\beta}};\bar{z}_j)]\}^2 \tag{13}$$

In other words, more accurate estimated parameters can be obtained by direct model estimation based on the linearly projected data using the rth order approximated expectation function. The EMVR method can also be applied using conventional nonlinear regression packages.

5. Case studies

To assess the effectiveness and robustness of the proposed MVR and EMVR methods, comprehensive simulation studies estimating both GMP functions with relaxed exponents and multivariate exponential decay functions based on linearly projected data are presented in this section. These two models are the most commonly adopted functional forms for modeling traffic flow relationships. The capability of the proposed methods to recover the assumed true model parameters is examined in various dimensions, including the model forms, scaling factor distributions and numbers of linear combinations constituting the unobservable traffic quantity. This section first introduces the two chosen model forms and describes the physical context of situations in which these two classes of transport models must be estimated based on linearly projected data. With the assumed true model parameters, data are generated to mimic situations that necessitate the use of linear data projection and the proposed methods. Regression analyses are conducted based on the proposed MVR and EMVR methods. The strengths and weaknesses of the two proposed methods are subsequently compared.

5.1.The two chosen classical traffic flow models

The GMP function with relaxed exponent (Eq. 14) and the multivariate exponential decay function (Eq. 15), which are the most commonly adopted functional forms in modeling traffic flow relationships, are the two models used in these simulation studies to examine the effectiveness and robustness of the two proposed methods. The GMP function is defined as follows:

$$y = \beta_0 + \beta_n z^n = \beta_0 + \beta_n \left(\sum_{i=1}^{m} f_i x_i \right)^n \tag{14}$$

where β_0, β_n and n are the model parameters. The GMP function can be used to model monotonically increasing relationships between traffic variables. One typical class of traffic flow models possessing such a relationship is the cost flow function or volume delay function, which can either be link-based (i.e., $m = 1$) or area-based (i.e., $m > 1$). The link-based Bureau of Public Roads (BPR) function adopted in the *Highway Capacity Manual* (Transportation Research Board, 2000) and the area-based macroscopic BPR (MBPR) investigated by Wong and Wong (2015, 2016a, 2016b) share exactly the same model form as the GMP function. In such a context, β_0 is the free-flow travel time per unit distance; β_n is the product of the free-flow travel time and the congestion sensitivity parameter; n is the parameter controlling the model nonlinearity; y is the travel time per unit distance associated with the road or network entity under consideration; and z is the traffic flow associated with the road or network entity under consideration.

Assuming that the total traffic flow in a network is observable at a subset of links outfitted with detectors, and that the probe vehicle flow is observable at every link in the network, which is a common real-world occurrence, the cases of estimating both a link-based BPR function and an area-based MBPR function necessitate the use of linear data projection and the proposed unbiased estimation methods. If the BPR function of a link without a detector is of interest (Scenario (a1) in Fig. 4), the number of links considered, m, is one. In such a case, y is the travel time per unit distance across the link that can be estimated by the reciprocal of the hourly space-mean speed of probe vehicles, under the assumption that those vehicles have traveled at similar speeds to all of the nearby traffic; x is the hourly probe vehicle flow across that link; and f is the scaling factor defined as the total-traffic-to-probe-vehicle ratio at the same link. However, f is unknown due to the absence of a detector. Because of geographical proximity, the scaling factors of nearby links can be assumed to follow a distribution, which can be inferred from the total-traffic-to-probe-vehicle ratios sampled from nearby links with detectors. Both the mean and variance of the scaling factor can thus be estimated. The scaling factor mean can be used to provide unbiased estimation of the hourly total traffic flow across the link in a linear data projection (i.e., $\bar{f}x$). If the exponent of the BPR function is allowed to vary, making it a nonlinear regression model, the proposed MVR or EMVR method is necessary for unbiased estimation.

Conversely, if the MBPR function of a network is of interest (Scenario (a2) shown in Fig. 4), m is the number of links intercepting the boundaries of the selected network (defined as boundary stations in Wong and Wong, 2015, 2016a, 2016b). In such a case, y is the travel time per unit distance across the network that can be approximated by the reciprocal of the hourly space-mean speed of probe vehicles sampled within the network in an hour; x_i is the hourly probe vehicle flow entering the network via boundary station i; and f_i is the corresponding scaling factor, defined as the total-traffic-to-probe-vehicle ratio at boundary station i. Again, the exact value of each of the scaling factors could be unknown due to the absence of a detector. The scaling factor distribution can be inferred from the scaling factors sampled from links with detectors in the network. The scaling factor mean can be used to project the expected hourly total traffic flow entering the network via the boundary stations based on linear data projection (i.e., $\sum_{i=1}^{m} \bar{f} x_i$). The proposed MVR or EMVR method can then be applied for unbiased model estimation.

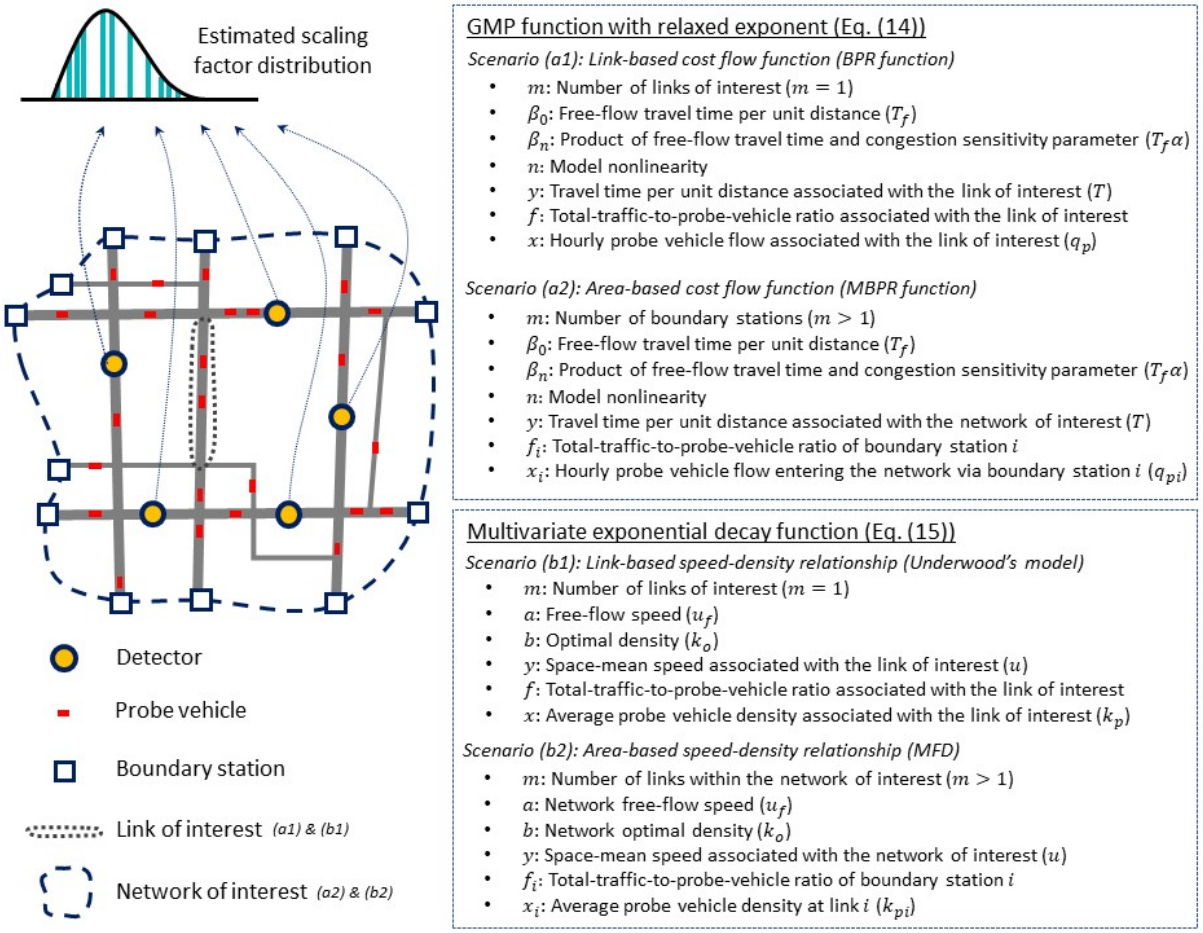

Fig. 4. Real-world scenarios of estimating link- and area-based cost flow functions and speed-density relationships that necessitate the use of linear data projection and the proposed methods for unbiased model estimations

The multivariate exponential decay function is defined as follows:

$$y = a \exp\left(\frac{-z}{b}\right) = a \exp\left(\frac{-\sum_{i=1}^{m} f_i x_i}{b}\right) \quad (15)$$

where a and b are the model parameters. The model can be used to model strictly decreasing relationship between traffic variables. Speed-density relationships, which can either be link-based (i.e., $m = 1$) or area-based (i.e., $m > 1$), are a typical class of traffic flow models possessing a decreasing relationship. Therefore, they can potentially be modeled using the multivariate exponential decay function. When a single link is considered, the multivariate exponential decay function reduces to the classical Underwood's model depicting the link-based speed-density relationship. When an MFD is considered, the multivariate exponential decay function could also be a candidate model describing the decreasing relationship between speed and density over a given area. In particular, urban networks are usually heterogeneously loaded and MFDs have low scatter for homogeneous networks. Different clustering algorithms have been proposed to partition heterogeneous networks into relatively homogenous networks (Saeedmanesh and Geroliminis 2016, 2017; Lopez et al. 2017; An et al. 2018). In the context of speed-density relationships, the y-intercept, a, is the free-flow speed; b is the optimal/critical traffic density at which traffic flow/throughput is at its maximum; y is the space-mean speed associated with the road or network entity under consideration; and z is the traffic density associated with the road or network entity under consideration.

Similar to the cases of cost flow function estimations, the estimations of both link- and area-based speed-density relationships require the use of linear data projection and the proposed MVR or EMVR method when the total traffic flow in a network is only observable at a subset of links installed with detectors and the probe vehicle flow is observable at every link in the network. If the Underwood's model of a link without a detector is of interest (Scenario (b1) in Fig. 4), m is equal to one. In such a case, y is the space-mean speed of traffic across the link within a short period of time (e.g., 5 min) that can be estimated by the space-mean speed of probe vehicles under the assumption that those vehicles have traveled at similar speeds to all of the nearby traffic; x is the average probe vehicle density at the link within the short time period; and f is the scaling factor, defined as the total-traffic-to-probe-vehicle ratio at the same link. Again, f is unknown due to the absence of a detector. The scaling factor distribution could be estimated using the total-traffic-to-probe-vehicle ratios sampled from nearby links with detectors in a similar manner to that described earlier. The scaling factor mean is applied in a linear data projection to provide an unbiased estimation of the total traffic density (i.e., $\bar{f}x$). Because the Underwood's model is a nonlinear regression model, either the MVR or EMVR method must be used for unbiased model estimation.

For the area-wide speed-density relationship (Scenario (b2) shown in Fig. 4), m is the number of links within the sampled network; y is the space-mean speed of all traffic within the network in a short period of time (e.g., 5 min) that can be approximated by the space-mean speed of the probe vehicles within the sampled network in that time period; x_i is the average probe vehicle density at link i in the sampled network within that time period; and f_i

is the corresponding scaling factor, defined as the total-traffic-to-probe-vehicle ratio at link i. Because of the absence of detectors, the individual values of scaling factors are usually unknown. The scaling factor distribution can be inferred from the total-traffic-to-probe-vehicle ratios of links with detectors sampled within the network. Linear data projection can be applied for unbiased estimation of network traffic density (i.e., $\sum_{i=1}^{m} \bar{f} x_i$). As the multivariate exponential decay function is a nonlinear regression model, the proposed MVR or EMVR method must then be used for unbiased model estimation.

5.2.Data generation

To assess the effectiveness and robustness of the proposed MVR and EMVR methods under different conditions, 12 simulation cases with different combinations of the chosen models (GMP functions with relaxed exponents and multivariate exponential decay functions), scaling factor distributions (normal and lognormal) and numbers of linear combinations of the scaling factors and observable independent variables ($m = 1, 2$ and 3) were considered.

The simulation studies estimating the GMP functions and the multivariate exponential decay functions based on linearly projected data could be used to mimic the situations estimating the BPR functions and speed-density relationships in their link- and area-based forms, as described. We set $\beta_0 = 0.025$, $\beta_n = 0.01$, $n = 3$, $a = 30$ and $b = 2000$ and examined the performance of the proposed MVR and EMVR methods in recovering these true model parameters. The observable independent variable, x_i, could be used to mimic the hourly probe vehicle flow on link i or the average density of probe vehicle present on link i in real-world scenarios. In principle, the observable independent variable can be any traffic variable that follows any distribution. For $m = 1$, each observation $\boldsymbol{x}$ comprised one x. Similarly, when $m = 2$ or 3, two or three x were independently sampled from the chosen distribution for each observation $\boldsymbol{x}$. Uniform distributions were chosen for the purpose of data generation. For the GMP function, $10{,}000$ observations of the observable independent variable x were sampled from a $Unif(0, 1)$. For the multivariate exponential decay function, $10{,}000$ observations of the observable independent variable x were sampled from a $Unif(0, 100)$. To avoid any sampling error derived from data generation of the observable independent variables, these six sets of sampled observable independent variables were used throughout all of the simulations. The scaling factor f of each x, which could be defined as the total-traffic-to-probe-vehicle ratio, was sampled from either a normal or lognormal distribution, with $\bar{f} = 2$ and $\sigma_f = 0.4$ for the cases of GMP functions and with $\bar{f} = 100$ and $\sigma_f = 20$ for the multivariate exponential decay functions. For each simulation case, the corresponding $10{,}000$ y, which could serve as the observable travel time per unit distance or the observable space-mean speed, were evaluated based on the values of the true model parameters, the sampled x and f. Assuming that the value of each f was no longer available, which is a common real-world occurrence, z could only be estimated using linear data projection based on the mean of the scaling factor (i.e., $\bar{z} = \sum_{i=1}^{m} \bar{f} x_i$). The proposed MVR or EMVR method had to be employed for unbiased model estimations.

5.3.Regression analyses based on the MVR method

In this subsection, regression analyses were conducted based on the linearly projected data using the original models (Eq. (14) and Eq. (15)) and the MVR method. The quadratic approximation of the expectation functions of the GMP and multivariate exponential decay functions, as shown in Eq. (16) and Eq. (17), respectively, can be obtained by adopting Eq.(3).

$$E_2[y] = \beta_0 + \beta_n \left[1 + \frac{n(n-1)}{2}\left(\frac{\sigma_f}{\bar{f}}\right)^2 \frac{\sum_{i=1}^{m} {x_i}^2}{(\sum_{i=1}^{m} x_i)^2}\right]\left(\sum_{i=1}^{m} \bar{f} x_i\right)^n \tag{16}$$

$$E_2[y] = a \left[1 + \frac{1}{2}\left(\frac{\sigma_f}{\bar{f}}\right)^2 \sum_{i=1}^{m}\left(\frac{-\bar{f} x_i}{b}\right)^2\right] \exp\left(\frac{-\sum_{i=1}^{m} \bar{f} x_i}{b}\right) \tag{17}$$

Each of the simulation cases was repeated 10,000 times with resampled scaling factors to obtain the means of the estimated model parameters. Tables 1 and 2 summarise the results of the six simulation cases for the GMP function and the multivariate exponential decay function, respectively.

Table 1
Estimated model parameters of the GMP functions based on the original model and the MVR method

Assumed scaling factor distribution	m	Original model or linear approximation			Quadratic approximation		
		$\hat{\beta}_0$ (% error)	$\hat{\beta}_n$ (% error)	$\hat{n}$ (% error)	$\hat{\beta}_0$ (% error)	$\hat{\beta}_n$ (% error)	$\hat{n}$ (% error)
Normal distribution	1	0.0250(−0.01%)	0.0112(+12.06%)	2.9997(−0.01%)	0.0250(−0.01%)	0.0100(+0.07%)	2.9997(−0.01%)
	2	0.0251(+0.39%)	0.0109(+8.88%)	2.9806(−0.65%)	0.0250(−0.03%)	0.0100(+0.21%)	3.0007(+0.02%)
	3	0.0252(+0.62%)	0.0107(+7.45%)	2.9832(−0.56%)	0.0250(+0.08%)	0.0100(+0.32%)	3.0013(+0.04%)
Lognormal distribution	1	0.0250(+0.01%)	0.0112(+12.44%)	3.0022(+0.07%)	0.0250(+0.01%)	0.0100(+0.40%)	3.0022(+0.07%)
	2	0.0251(+0.49%)	0.0109(+9.08%)	2.9804(−0.65%)	0.0250(+0.08%)	0.0100(+0.39%)	3.0005(+0.02%)
	3	0.0252(+0.72%)	0.0108(+7.64%)	2.9827(−0.58%)	0.0250(+0.19%)	0.0100(+0.50%)	3.0007(+0.02%)

Table 2
Estimated model parameters of the multivariate exponential decay functions based on the original model and the MVR method

Assumed scaling factor distribution	m	Original model or linear approximation		Quadratic approximation	
		$\hat{a}$ (% error)	$\hat{b}$ (% error)	$\hat{a}$ (% error)	$\hat{b}$ (% error)
Normal distribution	1	29.6806(−1.06%)	2086.6541(+4.33%)	29.9713(−0.10%)	2005.1767(+0.26%)
	2	29.3731(−2.09%)	2086.0795(+4.30%)	29.9376(−0.21%)	2005.6811(+0.28%)
	3	29.1208(−2.93%)	2085.3468(+4.27%)	29.9161(−0.28%)	2006.2485(+0.31%)
Lognormal distribution	1	29.7159(−0.95%)	2079.2287(+3.96%)	30.0087(+0.03%)	1997.6377(−0.12%)
	2	29.4557(−1.81%)	2077.5473(+3.88%)	30.0073(+0.02%)	1998.3882(−0.08%)
	3	29.0931(−3.02%)	2081.2316(+4.06%)	30.0101(+0.03%)	1999.1007(−0.04%)

The model estimation results based on the linearly projected data using the original models of the GMP functions revealed that the estimated parameters, $\hat{\beta}_0$ and $\hat{n}$, were unbiased because their values were extremely close to their true values (i.e., $\beta_0 = 0.025$ and $n = 3$). In terms of magnitude, the percentage errors of $\hat{\beta}_0$ and $\hat{n}$ were less than or equal to 0.72% and 0.65%, respectively, for all the simulation cases. $\hat{\beta}_n$ was the only biased parameter. The magnitudes of percentage errors of $\hat{\beta}_n$ decreased with the number of linear combination of the scaling factor and the observable independent variables (i.e., m). Taking the cases with normally distributed scaling factor as examples, the magnitudes of the percentage errors of $\hat{\beta}_n$ gradually decreased from 12.06% when $m = 1$ to 7.45% when $m = 3$. A similar pattern persisted for the cases with lognormally distributed scaling factors. The random effect cancellation among scaling factors as m escalates could be the explanation for the decreasing percentage errors. Moreover, it should also be noted that $\hat{\beta}_n$ was overestimated for all cases. According to the finding in Section 2.2, as the GMP functions under consideration were strictly increasing convex functions $\forall z > 0$, the net movement of data point shifting was to the left such that the expectation functions of the linearly projected data were above their true functions $\forall z > 0$. Thus, $\hat{\beta}_n$ had to be overestimated to account for the vertical difference and minimise the least squares functions.

The last three columns of Table 1 present the results obtained using the MVR method. Although $\hat{\beta}_0$ and $\hat{n}$ were originally unbiased, their extremely small percentage errors with magnitudes less than or equal to 0.19% for all the cases show that the MVR method further improved their accuracy. Most importantly, the MVR method significantly reduced the systematic biases embedded in $\hat{\beta}_n$. The magnitudes of the percentage errors of $\hat{\beta}_n$ for all of the cases with normally distributed scaling factors were well confined to between 0.07% and 0.32%. Theoretically, the expected value of $\hat{\beta}_n$ should exactly equal its true value (i.e., $\beta_n = 0.01$) because the skewness of the normally distributed scaling factor was zero and $n = 3$, and hence the quadratic approximations used in the MVR method were the exact expectation functions. The small discrepancies were derived from the sampling errors of the scaling factor among the $10{,}000$ repeated simulations for each case, because all of the repeated simulations were conducted using the same set of observable independent variable x. Conversely, the magnitudes of the percentage errors of $\hat{\beta}_n$ were slightly greater for the cases with lognormally distributed scaling factors because their quadratic approximations were not exact due to the nonzero skewness of their scaling factors. The slightly larger percentage errors originated from both the sampling errors of scaling factors and unrecovered minimal effects due to the third moments of the scaling factors on $\hat{\beta}_n$. Nevertheless, in practical terms, such small percentage errors with magnitudes less than or equal to 0.5% were acceptable.

According to the results from estimations based on the original models of the multivariate exponential decay functions, it is apparent that both the estimated parameters, $\hat{a}$ and $\hat{b}$, were biased, which indicates that the least squares functions can be minimised when they are biased. Taking the cases with normally distributed scaling factors as examples, the magnitudes of percentage errors of $\hat{a}$ and $\hat{b}$, respectively, increased from around 1% to 3% and oscillated at about 4.3%. The parameter a controls the y-intercept of a multivariate

exponential decay function. As m increased, the sampled data points close to the y-axis became sparser. The sparser sampled data points at high m and the relatively larger gradient at the y-intercept of a multivariate exponential decay function could be the reason for the increasing percentage error in magnitudes of $\hat{a}$. Similar patterns appeared for $\hat{a}$ and $\hat{b}$ in the cases with lognormally distributed scaling factors. The directions of biases of $\hat{a}$ and $\hat{b}$ should also be noted: $\hat{a}$ was underestimated and $\hat{b}$ overestimated in all cases. As the multivariate exponential decay functions are strictly decreasing convex functions, the net movement of data point shifting was to the right, such that the expectation functions were above their true functions. Moreover, according to Section 3.2, the expectation functions did not belong to the family of exponential functions; thus, model specification errors were induced in these cases. To better fit the flattened expectation functions (relatively flatter than their corresponding true functions), $\hat{a}$ had to be underestimated and $\hat{b}$ overestimated to account for the vertical difference and minimise the least squares function. Therefore, in addition to the sampling errors of the scaling factor and the unrecovered effects of higher order moments, the change in the model structure of the expectation functions, which is categorised as a model specification error induced by linear data projection, was also a source of bias.

The last two columns in Table 2 show the results of $\hat{a}$ and $\hat{b}$ based on the MVR method. The percentage errors of $\hat{a}$ and $\hat{b}$ were significantly reduced in magnitude for all of the cases. Taking the case with $m = 3$ and normally distributed scaling factors as an example, the magnitudes of the percentage errors of $\hat{a}$ and $\hat{b}$ reduced, respectively, from 2.93% to 0.28% and from 4.27% to 0.31%. Similarly, for the case with $m = 3$ and lognormally distributed scaling factors, the percentage errors of $\hat{a}$ and $\hat{b}$ dropped significantly from 3.02% to 0.03% and from 4.06% to 0.04%, respectively. One may observe that the magnitudes of percentage errors for cases with normally distributed scaling factors were generally slightly greater than those for cases with lognormally distributed scaling factors. This is because, for multivariate exponential decay functions, the effects of higher order moments of normally distributed scaling factors on the systematic data point distortion are relatively greater than those of lognormally distributed scaling factors. Thus, fitting the distorted data points using quadratic approximations led to slightly greater percentage errors in these cases. This was indirectly evidenced by the dissipation of such a pattern in the simulations with exactly the same settings presented in Section 5.4 where quartic approximations of the expectation functions. However, in practical terms, the minimal percentage error of less than 0.3% for cases with normally distributed scaling factors were acceptable.

5.4.Regression analyses based on the EMVR method

This subsection assesses and demonstrates the estimation accuracy of the proposed EMVR method. Simulations of model estimations for the GMP function and multivariate exponential decay functions were conducted with the same settings as those presented in the Section 5.3. To enable comparison, the same six sets of observable independent variables were used for all simulations in this subsection.

The application of the proposed EMVR method requires the higher order approximation of the expectation function, and hence, the higher order moments of the scaling factor. For the GMP function, as the chosen value of n is 3, its expectation function can only be approximated up to its cubic form, as shown in Eq. (18), using Eq. (4). In particular, $S_f = 0$ if the scaling factor is assumed to follow a normal distribution. The cubic approximation automatically reduces to a quadratic approximation in these cases. Therefore, only those simulation cases associated with lognormally distributed scaling factors are relevant and presented in this subsection because of their nonzero skewness.

$$E_3[y] = \beta_0 + \beta_n \left[1 + \frac{n(n-1)}{2}\left(\frac{\sigma_f}{\bar{f}}\right)^2 \frac{\sum_{i=1}^{m} {x_i}^2}{(\sum_{i=1}^{m} x_i)^2} + \frac{n(n-1)(n-2)}{6}\frac{S_f}{\bar{f}^3}\frac{\sum_{i=1}^{m} {x_i}^3}{(\sum_{i=1}^{m} x_i)^3}\right]\left(\sum_{i=1}^{m} \bar{f} x_i\right)^n \tag{18}$$

In contrast, the multivariate exponential decay function is infinitely differentiable. As the fourth order approximation is generally considered to have a satisfactory level of accuracy, the quartic approximation of the expectation function, which can be obtained by using Eq. (5), was used to demonstrate the effectiveness of the proposed EMVR method. Moreover, for the cases with the normally distributed scaling factor, the term associated with S_f in the square brackets of Eq. (19) reduced to zero due to the zero skewness.

$$E_4[y] = a\left[1 + \frac{1}{2}\left(\frac{\sigma_f}{\bar{f}}\right)^2 \sum_{i=1}^{m}\left(\frac{-\bar{f}x_i}{b}\right)^2 + \frac{1}{6}\frac{S_f}{\bar{f}^3}\sum_{i=1}^{m}\left(\frac{-\bar{f}x_i}{b}\right)^3 + \frac{1}{24}\frac{K_f}{\bar{f}^4}\sum_{i=1}^{m}\left(\frac{-\bar{f}x_i}{b}\right)^4\right]\exp\left(\frac{-\sum_{i=1}^{m}\bar{f}x_i}{b}\right) \tag{19}$$

Similar to Section 5.3, each of the simulation cases was repeated 10,000 times with the same sets of sampled observable independent variables and resampled scaling factors to obtain the means of the estimated model parameters. Table 3 presents the results of the three simulation cases with lognormally distributed scaling factor for the GMP function, and Table 4 reveals the results of the six simulation cases based on the quartic approximation for the multivariate exponential decay function.

Table 3
Estimated model parameters of the GMP functions based on the EMVR method

Assumed scaling factor distribution	m	Cubic approximation		
		$\hat{\beta}_0$ (% error)	$\hat{\beta}_n$ (% error)	$\hat{n}$ (% error)
Lognormal distribution	1	$0.0250(+0.01\%)$	$0.0100(-0.04\%)$	$3.0022(+0.07\%)$
	2	$0.0250(+0.05\%)$	$0.0100(+0.12\%)$	$3.0017(+0.06\%)$
	3	$0.0250(+0.12\%)$	$0.0100(+0.34\%)$	$3.0014(+0.05\%)$

The model estimation results of the GMP functions show the significant correction power of the proposed EMVR method. All of the estimated parameters were extremely close to their true values, with most of the percentage errors close to 0%. Because $n = 3$, the cubic approximations were the exact expectation functions of the linearly projected data. Any remaining percentage errors in the estimated parameters should arise from the sampling errors of the scaling factors.

Table 4
Estimated model parameters of the multivariate exponential decay functions based on the EMVR method

Assumed scaling factor distribution	m	Quartic approximation	
		$\hat{a}$ (% error)	$\hat{b}$ (% error)
Normal distribution	1	$29.9971(-0.01\%)$	$2000.3991(+0.02\%)$
	2	$29.9868(-0.04\%)$	$2001.0856(+0.05\%)$
	3	$29.9643(-0.05\%)$	$2001.7658(+0.09\%)$
Lognormal distribution	1	$30.0046(+0.02\%)$	$1999.1886(-0.04\%)$
	2	$29.9995(0.00\%)$	$1999.9005(0.00\%)$
	3	$29.9970(-0.01\%)$	$2000.6292(+0.03\%)$

The estimated model parameters of the multivariate exponential decay functions based on the EMVR method were remarkably close to their true values ($a = 30$ and $b = 2000$) with percentage errors well below 0.1% for all cases. The remaining tiny percentage errors should mainly stem from the sampling errors of the scaling factors and any unrecovered higher order moments. The results show the significant correction power of the proposed EMVR method. It should be noted that the pattern of slightly greater percentage errors for cases with normally distributed scaling factors observed in Section 5.3 dissipated here, providing indirect empirical evidence that the relatively greater effects of higher order moments of normally distributed scaling factors on the systematic data point distortion for these functions was indeed the reason for that pattern.

5.5. Comparison of the MVR and EMVR methods

Both the proposed MVR and EMVR methods can significantly reduce systematic biases in estimated parameters for all nonlinear transport models. They can be easily adopted with the use of standard nonlinear regression packages. However, they have different properties, and hence different strengths and weaknesses. The most suitable method should be selected for a given situation.

The MVR method is generic and flexible because only the first two moments of the scaling factor are required and no assumption about the distribution of the scaling factor needs to be made in applications. However, such flexibility is achieved at the cost of the accuracy of the estimated model parameters, if the higher order moments of the scaling factor do effectively play roles in the systematic data point distortion.

In contrast, an assumption of the scaling factor distribution is indispensable when the EMVR method is adopted. The assumed distribution should be in line with the properties of the scaling factor in the given situation. Statistical tests such as a Kolmogorov-Smirnov or a chi-square goodness-of-fit test can be performed to validate the assumed distribution in applications. In general, the EMVR method can further improve the accuracy of the estimated parameters compared with the MVR method.

In other words, the selection of the MVR or EMVR method is a trade-off between flexibility and accuracy. As the EMVR method is generally superior to the MVR method, it should be used to achieve a higher level of accuracy in parameter estimations as long as sufficient information concerning the properties of the scaling factor is available.

6. Conclusions

The various limitations of different high-tech devices have long been the obstacles hindering direct traffic data collection in many real-world situations. This makes traffic data estimation inevitable. Linear data projection is a data scaling method that can fuse data collected from different sources and offer unbiased traffic data estimation. Due to its simplicity, it has been widely adopted in many transportation studies. Although it is intuitive to estimate models based on unbiased data, Wong and Wong (2015) recently proved that this approach can lead to systematically biased estimated parameters because information about the variability of the scaling factor used in linear data projection is lost. GMP functions with fixed exponents were chosen for examination and adjustment factors reducing biases of the estimated parameters were derived. However, it must be stressed that the method is only applicable to GMP functions with fixed exponents that are typical linear regression models characterised by linear normal equations in model estimation.

Because many transport models do not belong to the family of GMP functions with fixed exponents, it is necessary to develop methods that address generic nonlinear transport models requiring nonlinear regression, so the benefits of this powerful tool can be maximised and fully realised. This paper is dedicated to filling this research gap and makes several contributions to the existing body of knowledge in our field. We first investigated the mechanism of the systematic data point data distortion resulting from linear data projection, which is the root cause of systematic bias. The first and second derivatives were found to be the governing factors determining the data point shifting. This finding is important because it can be used to predict the directions of biases of parameters estimated from linearly projected data. We then identified two major practical difficulties in applying the adjustment factor approach (previously adopted in Wong and Wong (2015) for GMP functions with fixed

exponents) to generic nonlinear transport models: the inexistence or derivation complexity of simple closed-form adjustment factors and the model specification error induced by linear data projection. This not only illustrates the complex nature of the problem for nonlinear models, but also reveals a new type of error constituting the systematic bias.

Based on the data point distortion mechanism, a generic MVR method, only requiring the first two scaling factor moments, and an EMVR method, further incorporating higher order moments, were proposed to reduce or avoid systematic bias in nonlinear transport model estimation. To demonstrate the effectiveness and robustness of the proposed methods, simulation studies were conducted, estimating both the GMP functions with relaxed exponents and the multivariate exponential decay functions based on linearly projected data. The GMP function and the multivariate exponential decay function are the most commonly adopted functional forms in modeling traffic flow relationships such as cost-flow functions and speed-density relationships in their link- and area-based forms. The results revealed that the EMVR method usually provides more accurate estimations and should be adopted if sufficient information concerning the scaling factor properties is available. Both MVR and EMVR methods can be easily applied using any standard statistical package with an optimization library. With these methods, analysts who leverage linear data projection are better supported to make practical decisions. These proposed methods are generic and easy to apply, and hence powerful.

According to Wong and Wong (2016b), heteroscedasticity is inherently introduced by linear data projection if the scaling factor is subject to variability. That means, even if the proposed MVR or EMVR method is employed for unbiased parameter estimations, the estimated parameter standard errors are biased. To ensure valid statistical tests based on these statistics that quantify the dispersion of and confidence in the estimated model parameters, the effects of non-homoscedasticity must be taken into consideration. We are currently working on standard error estimations for the estimated parameters of nonlinear transport models based on the proposed MVR and EMVR methods.

Acknowledgements

The work presented in this study was supported by a Research Postgraduate Studentship and by grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. HKU 17208614). The second author was also supported by the Francis S Y Bong Professorship in Engineering. We would like to express our sincere gratitude to Concord Pacific Satellite Technologies Limited and Motion Power Media Limited for their provision of the taxi GPS data, and to the Transport Department of the HKSAR government for providing the traffic flow data from the ATC that were used in this research.

Appendix A. The expectation function of the linearly projected data

Here, we provide the expectation function of the linearly projected data. Approximate $G(\boldsymbol{\beta}; z)$ by a Taylor series expansion with a centre at $f_i = \bar{f}, \forall i \in \{1, 2, \dots, m\}$.

$$\begin{aligned} G(\boldsymbol{\beta}; z) = G(\boldsymbol{\beta}; \bar{\boldsymbol{f}}) &+ \sum_{i=1}^{m} \frac{\partial G(\boldsymbol{\beta}; \bar{\boldsymbol{f}})}{\partial f_i}(f_i - \bar{f}) \\ &+ \frac{1}{2!}\sum_{i=1}^{m}\sum_{j=1}^{m} \frac{\partial^2 G(\boldsymbol{\beta}; \bar{\boldsymbol{f}})}{\partial f_i \partial f_j}(f_i - \bar{f})(f_j - \bar{f}) \\ &+ \frac{1}{3!}\sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{s=1}^{m} \frac{\partial^3 G(\boldsymbol{\beta}; \bar{\boldsymbol{f}})}{\partial f_i \partial f_j \partial f_s}(f_i - \bar{f})(f_j - \bar{f})(f_s - \bar{f}) \\ &+ \frac{1}{4!}\sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{s=1}^{m}\sum_{t=1}^{m} \frac{\partial^4 G(\boldsymbol{\beta}; \bar{\boldsymbol{f}})}{\partial f_i \partial f_j \partial f_s \partial f_t}(f_i - \bar{f})(f_j - \bar{f})(f_s - \bar{f})(f_t - \bar{f}) \\ &+ \cdots \end{aligned} \tag{A1}$$

Taking expectation on both sides of Eq. (A1),

$$\begin{aligned} E[G(\boldsymbol{\beta}; z)] = G(\boldsymbol{\beta}; \bar{\boldsymbol{f}}) &+ \sum_{i=1}^{m} \frac{\partial G(\boldsymbol{\beta}; \bar{\boldsymbol{f}})}{\partial f_i} E(f_i - \bar{f}) \\ &+ \frac{1}{2!}\sum_{i=1}^{m}\sum_{j=1}^{m} \frac{\partial^2 G(\boldsymbol{\beta}; \bar{\boldsymbol{f}})}{\partial f_i \partial f_j} E[(f_i - \bar{f})(f_j - \bar{f})] \\ &+ \frac{1}{3!}\sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{s=1}^{m} \frac{\partial^3 G(\boldsymbol{\beta}; \bar{\boldsymbol{f}})}{\partial f_i \partial f_j \partial f_s} E[(f_i - \bar{f})(f_j - \bar{f})(f_s - \bar{f})] \\ &+ \frac{1}{4!}\sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{s=1}^{m}\sum_{t=1}^{m} \frac{\partial^4 G(\boldsymbol{\beta}; \bar{\boldsymbol{f}})}{\partial f_i \partial f_j \partial f_s \partial f_t} E[(f_i - \bar{f})(f_j - \bar{f})(f_s - \bar{f})(f_t - \bar{f})] \\ &+ \cdots. \end{aligned}$$

Because $E(f_i) = \bar{f}, \forall i \in \{1, 2, \dots, m\}, E(f_i - \bar{f}) = 0$. Assuming $f_i, \forall i \in \{1, 2, \dots, m\}$, are independent of each other,

$E[(f_i - \bar{f})(f_j - \bar{f})] = 0, if\ i \neq j, \forall i, j \in \{1, 2, \dots, m\}$;

$E[(f_i - \bar{f})(f_j - \bar{f})(f_s - \bar{f})] = 0, if\ i \neq j \neq s, \forall i, j, s \in \{1, 2, \dots, m\}$; and

$E[(f_i - \bar{f})(f_j - \bar{f})(f_s - \bar{f})(f_t - \bar{f})] = 0, if\ i \neq j \neq s \neq t, \forall i, j, s, t \in \{1, 2, \dots, m\}$.

Moreover,

$E\left[(f_i - \bar{f})^2\right] = {\sigma_f}^2, if\ i = j, \forall i, j \in \{1, 2, \dots, m\}$;

$E\left[(f_i - \bar{f})^3\right] = S_f, if\ i = j = s, \forall i, j, s \in \{1, 2, \dots, m\}$; and

$$E\left[\left(f_i - \bar{f}\right)^4\right] = K_f, if\ i = j = s = t, \forall i, j, s, t \in \{1, 2, \dots, m\}.$$

The higher order terms, involving the expectation of the random variable f_i, can be simplified in a similar manner. Therefore, the expectation function of the linearly projected data is given by Eq. (A2). It shows that not only the first moment of the scaling factor $\bar{f}$, but also its higher order moments such as variance ${\sigma_f}^2$, skewness S_f, kurtosis K_f, contribute to the mean values of the linearly projected data.

$$E[G(\boldsymbol{\beta}; z)] = G\left(\boldsymbol{\beta}; \bar{\boldsymbol{f}}\right) + \frac{1}{2!}{\sigma_f}^2 \sum_{i=1}^{m} \frac{\partial^2 G\left(\boldsymbol{\beta}; \bar{\boldsymbol{f}}\right)}{\partial {f_i}^2} + \frac{1}{3!} S_f \sum_{i=1}^{m} \frac{\partial^3 G\left(\boldsymbol{\beta}; \bar{\boldsymbol{f}}\right)}{\partial {f_i}^3} + \frac{1}{4!} K_f \sum_{i=1}^{m} \frac{\partial^4 G\left(\boldsymbol{\beta}; \bar{\boldsymbol{f}}\right)}{\partial {f_i}^4} + \cdots \quad \text{(A2)}$$

References

Aboudolas K, Geroliminis N (2013) Perimeter and boundary flow control in multi-reservoir heterogeneous networks. *Transportation Research Part B: Methodological* 55: 265-281.

Ambühl L, Menendez M (2016) Data fusion algorithm for macroscopic fundamental diagram estimation. *Transportation Research Part C: Emerging Technologies* 71: 184-197.

An K, Chiu Y, Hu X, Chen X (2017) A network partitioning algorithmic approach for macroscopic fundamental diagram-based hierarchical traffic network management. *IEEE Transactions on Intelligent Transportation Systems* 19(4): 1130-1139.

Ban XJ, Li Y, Skabardonis A, Margulici JD (2010) Performance evaluation of travel-time estimation methods for real-time traffic applications. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations* 14(2): 54-67.

Bertini RL, Tantiyanugulchai S (2004) Transit buses as traffic probes: Use of geolocation data for empirical evaluation. *Transportation Research Record: Journal of the Transportation Research Board* 1870: 35-45.

Bolla R, Davoli F (2000) Road traffic estimation from location tracking data in the mobile cellular network. *Proceedings of the Wireless Communications and Networking Conference, IEEE*, Chicago, USA, 3: 1107-1112.

Caceres N, Romero LM, Benitez FG, del Castillo JM (2012) Traffic flow estimation models using cellular phone data. *IEEE Transactions on Intelligent Transportation Systems* 13(3): 1430-1441.

Chandra S, Kumar V, Sikdar PK (1995) Dynamic PCU and estimation of capacity of urban roads. *Indian Highways* 23(4): 17-28.

Daganzo CF (2007) Urban gridlock: Macroscopic modeling and mitigation approaches. *Transportation Research Part B: Methodological* 41(1): 49-62.

Du J, Rakha H, Gayah VV (2016) Deriving macroscopic fundamental diagrams from probe data: Issues and proposed solutions. *Transportation Research Part C: Emerging Technologies* 66: 136-149.

Geroliminis N, Daganzo CF (2008) Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B: Methodological* 42(9): 759-770.

Geroliminis N, Haddad J, Ramezani M (2013) Optimal perimeter control for two urban regions with macroscopic fundamental diagrams: A model predictive approach. *IEEE Transactions on Intelligent Transportation System* 14(1): 348-359.

Geroliminis N, Levinson DM (2009) Cordon pricing consistent with the physics of overcrowding. In Lam WHK, Wong SC, Lo HK, (eds). *Transportation and Traffic Theory 2009: Golden Jubilee* (Springer US, New York), 219-240.

Herrera JC, Bayen AM (2010) Incorporation of Lagrangian measurements in freeway traffic state estimation. *Transportation Research Part B: Methodological* 44(4): 460-481.

Herrera JC, Work DB, Herring R, Ban XJ, Jacobson Q, Bayen AM (2010) Evaluation of traffic data obtained via GPS-enabled mobile phones: The Mobile Century field experiment. *Transportation Research Part C: Emerging Technologies* 18(4): 568-583.

Ho HW, Wong SC (2007) Housing allocation problem in a continuum transportation system. *Transportmetrica* 3(1): 21-39.

Kwong K, Kavaler R, Rajagopal R, Varaiya P (2009) Arterial travel time estimation based on vehicle re-identification using wireless magnetic sensors. *Transportation Research Part C: Emerging Technologies* 17(6): 586-606.

Liang Y, Reyes ML, Lee JD (2007) Real-time detection of driver cognitive distraction using support vector machines. *IEEE Transactions on Intelligent Transportation Systems* 8(2): 340-350.

Lopez C, Krishnakurmari P, Leclercq L, Chiabaut N, Lint Hv (2017) Spatiotemporal partitioning of transportation network using travel time data. *Transportation Research Record: Journal of the Transportation Research Board* 2623: 98-107.

Meng F, Wong SC, Wong W, Li YC (2017a) Estimation of scaling factors for traffic counts based on stationary and mobile sources of data. *International Journal of Intelligent Transportation Systems Research* 15(3): 180-191.

Meng F, Wong W, Wong SC, Pei X, Li YC, Huang H (2017b) Gas dynamic analogous exposure approach to interaction intensity in multiple-vehicle crash: Case study of crashes involving taxis. *Analytic Methods in Accident Research* 16: 90-103.

Miwa T, Ishiguro Y, Yamamoto T, Morikawa T (2013) Allocation planning for probe taxi devices based on information reliability. *Transportation Research Part C: Emerging Technologies* 34: 55-69.

Moore II JE, Cho S, Basu A, Mezger DB (2001) *Use of Los Angeles Freeway Service Patrol Vehicles as Probe Vehicles*. California PATH Research Report UCB-ITS-PRR-2001-5, California PATH Program, Institute of Transportation Studies, University of California, Berkeley, CA.

Ramezani M, Nourinejad M (2017) Dynamic modelling and control of taxi services in large-scale urban networks: A macroscopic approach. Paper presented at the *22nd International Symposium on Transportation and Traffic Theory*, Chicago, IL.

Saeedmanesh M, Geroliminis N (2016) Clustering of heterogeneous networks with directional flows based on "Snake" similarities. *Transportation Research Part B: Methodological* 91: 250-269.

Saeedmanesh M, Geroliminis N (2017) Dynamic clustering and propagation of congestion in heterogeneously congested urban traffic networks. *Transportation Research Part B: Methodological* 105: 193-211.

Schwarzenegger A, Bonner DE, Iwasaki RH, Copp R (2009) *2008 State Highway Congestion Monitoring Program (HICOMP)*, Annual Data Compilation, Caltrans, Sacramento, CA.

Transportation Research Board (2000) *Highway Capacity Manual*. National Research Council, Washington, D.C.

Wong RCP, Szeto WY, Wong SC, Yang H (2014) Modelling multi-period customer-searching behaviour of taxi drivers. *Transportmetrica B: Transport Dynamics* 2(1): 40-59.

Wong W, Wong SC (2015) Systematic bias in transport model calibration arising from the variability of linear data projection. *Transportation Research Part B: Methodological* 75: 1-18.

Wong W, Wong SC (2016a) Network topological effects on the macroscopic Bureau of Public Roads function. *Transportmetrica A: Transport Science* 12(3): 272-296.

Wong W, Wong SC (2016b) Biased standard error estimations in transport model calibration due to heteroscedasticity arising from the variability of linear data projection. *Transportation Research Part B: Methodological* 88: 72-92.

Wong W, Wong SC (2016c) Evaluation of the impact of traffic incidents using GPS data. *Proceedings of the Institution of Civil Engineers:Transport* 169(3), 148-162.

Wright J, Dahlgren J (2001) *Using Vehicles Equipped with Toll Tags as Probes for Providing Travel Times*. California PATH Working Paper UCB-ITS-PWP-2001-13, California PATH Program, Institute of Transportation Studies, University of California, Berkeley, CA.

Ygnace JL, Drane C (2001) Cellular telecommunication and transportation convergence: A case study of a research conducted in California and in France on cellular positioning techniques and transportation issues. *2001 IEEE Intelligent Transportation Systems Conference Proceedings*, Oakland, CA: 16-22.

Yin J, Wong SC, Sze NN, Ho HW (2013) A continuum model for housing allocation and transportation emission problems in a polycentric city. *International Journal of Sustainable Transportation* 7(4): 275-298.

Zhao Y (2000) Mobile phone location determination and its impact on intelligent transportation systems. *IEEE Transactions on Intelligent Transportation Systems* 1(1): 55-64.

Zheng N, Rerat G, Geroliminis N (2016) Time-dependent area-based pricing for multimodal systems with heterogeneous users in an agent-based environment. *Transportation Research Part C: Emerging Technologies* 62: 133-148.

Zheng N, Waraich RA, Axhausen KW, Geroliminis N (2012) A dynamic cordon pricing scheme combining the macroscopic fundamental diagram and an agent-based traffic model. *Transportation Research Part A: Policy and Practice* 46(8): 1291-1303.

Zhong RX, Chen C, Huang YP, Sumalee A, Lam WHK, Xu DB (2017) Robust perimeter control for two urban regions with macroscopic fundamental diagrams: A control-Lyapunov function approach. Paper presented at the *22nd International Symposium on Transportation and Traffic Theory*, Chicago, IL.