

PERGAMON

Transportation Research Part B 36 (2002) 37–57

TRANSPORTATION RESEARCH PART B

www.elsevier.com/locate/trb

Dynamic identification of origin–destination matrices in the presence of incomplete observations

Baibing Li *, Bart De Moor

ESAT-SISTA, Department of Electrical Engineering, Katholieke, Universiteit Leuven, Kardinaal Mercierlaan 94, 3001 Leuven, Belgium

Received 10 September 1999; received in revised form 23 May 2000; accepted 6 June 2000

Abstract

It is quite often in practice that observation information of traffic flows is incomplete due to failures of facilities or insufficient installment of sensors. In this paper, we investigate problems of dynamic identification of origin–destination matrices when traffic counts are unavailable in some entrances and/or exits of a traffic system. To deal with non-linearity and inestimability problems in conventional research frameworks, identification problems with incomplete information are reformulated as optimization problems with linear equality constraints and non-negative inequality constraints. An algorithm is then suggested for identification of origin–destination matrices with incomplete observation information. Finally, numerical examples are provided to illustrate the proposed algorithm for some incomplete information patterns. © 2001 Elsevier Science Ltd. All rights reserved.

Keywords: Dynamic identification; Incomplete information; Origin–destination matrix

1. Introduction

Origin–destination (O–D) matrices have a number of uses from long-term planning and design of transportation systems to on-line traffic control. The methods used to estimate O–D matrices then vary depending on their uses: static estimation methods for long-term planning and design, and dynamic estimation methods for on-line traffic control. Dynamic estimations of O–D matrices have also been shown to be fundamental for development of advanced traffic management systems (ATMS) in recent years.

* Corresponding author. Present address: Centre for Process Analytics and Control Technology, Merz Court, University of Newcastle, Newcastle upon Tyne, NE1 7RU, UK.

E-mail address: baibing.li@ncl.ac.uk (B. Li).

0191-2615/01/$ - see front matter © 2001 Elsevier Science Ltd. All rights reserved.
PII: S0191-2615(00)00037-0

Static estimation methods of O–D matrices use accumulated traffic counts to set up a highly underdetermined set of equations for the unknown O–D flows. One important issue in static estimation methods is how to deal with the problem of insufficient information in estimation procedures. Traditionally, static estimation methods were based on the gravity model (Holm et al., 1976; Wilson, 1967) which reduced estimation problems of O–D matrices to calibration problems with small number of unknown parameters. Van Zuzlen and Willumsen (1980) extended the gravity model to entropy-maximizing technique. The problem of insufficient information for determination of O–D matrices was circumvented by maximizing an objective function, entropy. Another type of static estimation methods developed in recent decades is to include additional observation information (traffic counts, parking survey, etc.) and then combine it with priori knowledge of O–D matrices. Cascetta (1984) introduced generalized least squares approach to estimate O–D matrices from both survey data and traffic counts. Bierlaire and Toint (1995) used parking survey to improve estimates of O–D matrices. Mather (1983) proposed the use of a Bayesian approach to the O–D matrix estimation problem which results in functional forms equivalent to the generalized least squares method. Bell (1991a) proposed an algorithm to improve generalized least squares estimates by using inequality constraints. It is clear that the above techniques for dealing with the problem of insufficient information can be used not only in static estimation methods but also, if needed, in dynamic estimation methods in appropriate ways.

In recent decades, a lot of attention has been paid to identification of dynamic O–D matrices. Cremer and Keller (1987) and Nihan and Davis (1987) proposed recursive algorithms to estimate O–D matrices. In each time interval, the first stage of these algorithms is based on recursive least squares estimation, and at the second stage both equality and inequality constraints are enforced. Sherali et al. (1997) developed a constrained optimization algorithm to estimate an O–D matrix dynamically. This algorithm is more accurate than the ones of Cremer and Keller (1987) and Nihan and Davis (1987), while computational costs are very high in comparison with the latter. Recently, Li and De Moor (1999) proposed a fast recursion-based algorithm to estimate intersection O–D matrices. The algorithm of Li and De Moor (1999) is a reasonable compromise between the algorithms of Cremer and Keller (1987) and Nihan and Davis (1987), and the algorithm of Sherali et al. (1997). The accuracy of Li and De Moor's algorithm is almost the same as the algorithm of Sherali et al., while the computational costs of the former is much lower than the latter.

The above algorithms can be used to identify both constant and time-varying O–D matrices, provided that the time-variation of split parameters in O–D matrices is slow. One approach to identify time-varying O–D matrices is to use "forgetting factor" which weights more recent observations more heavily, such as the algorithms of Cremer and Keller (1987), and Nihan and Davis (1987). Another approach is to use rolling horizon framework (Sherali et al., 1997) in which, as time progresses, the horizon used for estimating the split parameters is rolled forward by at least one additional period.

Another important issue for dynamic estimations is how to deal with journey time, τ, that a vehicle needs to traverse through a traffic system. Some algorithms, such as Cremer and Keller (1987), Nihan and Davis (1987), and Li and De Moor (1999), assume τ to be so small (for instance, for an intersection or a short freeway section) that it could be omitted in algorithms. Another approach, adopted by Sherali et al. (1997), is to assume availability of journey time τ from one entrance to one exit pre-estimated based on historical data. Finally, Bell (1991b)

presented two models to deal with journey time τ, one for geometrically distributed travel times, and another for freely distributed travel times.

All of these dynamic estimation methods were based on an assumption that observation information at all entrances and exits of a traffic system is available. However, this is not always the case in practice. In fact, there *do* exist some important circumstances where observation information is incomplete. One typical situation is insufficient installment of sensors in traffic systems. For instance, currently most sections of the rings of Amsterdam and Anterwerp fall into this category. Another possible situation is one or several sensors of a traffic system fail. Clearly, this can occur for any traffic sensor. When observations are incomplete, algorithms proposed in the literature cannot be used.

In practice, there exist many other similar situations where information is incomplete. In response, statistical inferences based on incomplete information have been an important topic in theoretical research for decades (Demster et al., 1977; Little, 1987; McLachlan, 1996). Recently, incomplete-information based signal processing has also been investigated (Frenkel and Feder, 1999). Although dynamic identification of O–D matrices can be regarded as a particular case of signal processing, there exist special structures for identification of O–D matrices that can be exploited to design more efficient algorithms.

In this paper, for situations with incomplete information, where traffic counts at some entrances and/or exits of a traffic system are unavailable, we present a new research formulation to convert an identification problem into an optimization problem with linear equality constraints and non-negative inequality constraints. Then we show that the algorithm proposed by Li and De Moor (1999) can be adopted under certain conditions.

To highlight dynamic identification of O–D matrices with *incomplete* information, and to reduce complexity in notations, we restrict our interests in those situations where an O–D matrix to be estimated is constant and journey time τ through a traffic system is so small that could be omitted (for example for an intersection or a short freeway section) in this paper. Extensions that relax these restrictions will be discussed in Section 6.

Section 2 is devoted to notations. In Section 3, a new research formulation for identification of O–D matrices is introduced. Issues of algorithms are investigated in section 4. The developed algorithm is illustrated through simulation examples in Section 5.

2. Notations

Consider a traffic system (an intersection, a short freeway section, etc.) with m entrances and n exits. The time period of interest is divided into K intervals, each with length T. The journey time that a vehicle needs to traverse through a traffic system is assumed to be so small that could be omitted.

2.1. Notations for general situations

Following the notations by Sherali et al. (1997) and Li and De Moor (1999), let:

- $J_i = \{j : \text{exit } j \text{ is permissible for users entering at entrance } i\}$ $(i = 1, \ldots, m)$;
- $I_j = \{i : \text{entrance } i \text{ admits users that can take exit } j\}$ $(j = 1, \ldots, n)$;

- $y_j(k)$ $(j = 1, \ldots, n)$ denote the traffic volume leaving via the jth exit during the kth time interval $[kT, (k+1)T)$;
- $q_i(k)$ $(i = 1, \ldots, m)$ denote the traffic volume entering the ith entrance during the kth time interval $[kT, (k+1)T)$;
- b_{ij} $(i \in I_j, j = 1, \ldots, n)$ denote the probability (split parameter) a vehicle entering at $i \in I_j$ exiting at j $(j = 1, \ldots, n)$;
- $\boldsymbol{\beta}_j$ be defined by $\boldsymbol{\beta}_j = [b_{ij}]$ $(\forall i \in I_j)$ as $|I_j| \times 1$ vectors such that i is in an increasing order for $j = 1, \ldots, n$, i.e., $\boldsymbol{\beta}_j$ is the jth column of the O–D matrix $[b_{ij}]$;
- $\boldsymbol{\beta} = [\boldsymbol{\beta}_1^{\mathrm{T}}, \ldots, \boldsymbol{\beta}_n^{\mathrm{T}}]_{p\times 1}^{\mathrm{T}}$, where $p = \sum_{j=1}^{n} |I_j|$.

Without loss of generality, we assume that those split parameters corresponding to constantly zero traffic flows (such as b_{ii}, $\forall i$, in a traffic system prohibiting U-turn) are excluded from the vector $\boldsymbol{\beta}$.

2.2. Notations for incomplete information

In this paper, we consider two typical situations of incomplete observation information of traffic flows. The first one is insufficient installment of sensors of traffic flows, and the second one is failures of some sensors in a traffic system. For both situations, there exists at least one entrance or exit where traffic counts are unavailable. Let:

$MI = \{i :$ observations of traffic flows are unavailable at entrance $i\}$;

$MJ = \{j :$ observations of traffic flows are unavailable at exit $j\}$;

$OI = \{1, \ldots, m\} \setminus MI$ and $OJ = \{1, \ldots, n\} \setminus MJ$.

Then OI and OJ are those index sets of entrances and exits where traffic counts are available.

When $y_j(k)$ and $q_i(k)$ are available, we denote them as $y_j^*(k)$ and $q_i^*(k)$, respectively, i.e.,

$$y_j^*(k) \hat{=} y_j(k) \text{ for } j \in OJ \quad \text{and} \quad q_i^*(k) \hat{=} q_i(k) \text{ for } i \in OI.$$

Let $\mathbf{q}^* = [q_i^*]$ $(\forall i \in OI)$ and $\mathbf{y}^* = [y_j^*]$ $(\forall j \in OJ)$ be $\mathbf{q}$ and $\mathbf{y}$, but those elements with $i \in MI$ and $j \in MJ$, respectively, are deleted.

2.3. Some relations between rows and columns of an O–D matrix

For situations with incomplete information, it is more convenient to use those vectors of b_{ij} constituted by transposed rows of the O–D matrix $[b_{ij}]$. Specifically, define

$\mathbf{b}_i = [b_{ij}]$ $(\forall j \in J_i)$ as $|J_i| \times 1$ vectors such that j is in an increasing order for $i = 1, \ldots, m$;

$\mathbf{b} = [\mathbf{b}_1^{\mathrm{T}}, \ldots, \mathbf{b}_m^{\mathrm{T}}]_{p\times 1}^{\mathrm{T}}$.

Define $f_{ij}(k) = q_i(k) b_{ij}, \forall i \in I_j, j = 1, \ldots, n, k = 1, \ldots, K$.

Let $\mathbf{f}_i(k) = [f_{ij}(k)]$ $(\forall j \in J_i)$ as $|J_i| \times 1$ vectors such that j is in an increasing order for $i = 1, \ldots, m$, i.e., $\mathbf{f}_i(k)$ be the jth transposed row of the matrix $[f_{ij}]$. $\mathbf{f}(k) = [\mathbf{f}_1^{\mathrm{T}}(k), \ldots, \mathbf{f}_m^{\mathrm{T}}(k)]_{p\times 1}^{\mathrm{T}}$.

Clearly, $\mathbf{b}$ satisfies normalization conditions and non-negative constraints of probability:

$$\mathbf{R}_0 \mathbf{b} = 1_{m\times 1}, \tag{1a}$$

$$\mathbf{b} \geqslant 0, \tag{1b}$$

where $\mathbf{R}_0 = \mathrm{diag}\{\mathbf{1}^{\mathrm{T}}, \ldots, \mathbf{1}^{\mathrm{T}}\}_{m\times p}$. The ith $\mathbf{1}$ of $\mathbf{R}_0$ is a $|J_i| \times 1$ vector of ones. Throughout this paper, we denote $\mathbf{I}$ as an identity matrix, and $\mathbf{1}$ as a vector of ones, both with appropriate dimensions.

Let $\mathbf{\Lambda}(k) = \mathrm{diag}\{q_1(k)\mathbf{I}, \ldots, q_m(k)\mathbf{I}\}$ and $\mathbf{P}(k) = \mathrm{diag}\{q_1(k), \ldots, q_m(k)\}$. Then we have

$$\mathbf{P}(k)\mathbf{R}_0 = \mathbf{R}_0\mathbf{\Lambda}(k). \tag{2}$$

The equations $f_{ij}(k) = q_i(k)b_{ij}$ can be written as

$$\mathbf{f}(k) = \mathbf{\Lambda}(k)\mathbf{b}. \tag{3}$$

Summation over j for $f_{ij}(k) = q_i(k)b_{ij}$ and from (1a) yield a basic identity

$$q_i = \sum_{j\in J_i} f_{ij} \quad \text{or} \quad \mathbf{q}(k) = \mathbf{R}_0\mathbf{f}(k). \tag{4}$$

Finally, since $\mathbf{b}$ is a rearrangement of $\mathbf{\beta}$, there exists a permutation matrix $\mathbf{\Gamma}$ such that

$$\mathbf{\beta} = \mathbf{\Gamma}\mathbf{b} \quad \text{or} \quad \mathbf{b} = \mathbf{\Gamma}^{\mathrm{T}}\mathbf{\beta}. \tag{5}$$

For example, for a typical 4×4 intersection prohibiting U-turn, we have

$$\mathbf{\beta} = [b_{21}, b_{31}, b_{41}, b_{12}, b_{32}, b_{42}, b_{13}, b_{23}, b_{43}, b_{14}, b_{24}, b_{34}]^{\mathrm{T}},$$

$$\mathbf{b} = [b_{12}, b_{13}, b_{14}, b_{21}, b_{23}, b_{24}, b_{31}, b_{32}, b_{34}, b_{41}, b_{42}, b_{43}]^{\mathrm{T}},$$

$$\Gamma = [\mathbf{t}_4, \mathbf{t}_7, \mathbf{t}_{10}, \mathbf{t}_1, \mathbf{t}_8, \mathbf{t}_{11}, \mathbf{t}_2, \mathbf{t}_5, \mathbf{t}_{12}, \mathbf{t}_3, \mathbf{t}_6, \mathbf{t}_9],$$

where $\mathbf{t}_j$ is the jth column of a 12×12 identity matrix.

Combining (1) and (5), we obtain constraints for the parameter vector $\mathbf{\beta}$

$$\tilde{\mathbf{R}}_0\mathbf{\beta} = \mathbf{1}_{m\times 1} \quad \text{with } \tilde{\mathbf{R}}_0 = \mathbf{R}_0\mathbf{\Gamma}^{\mathrm{T}}, \tag{1a'}$$

$$\mathbf{\beta} \geqslant 0. \tag{1b'}$$

3. Problem formulation

3.1. Statement of problem for complete observation information

When observation information is complete, observation equations can be written as

$$\mathbf{y}(k) = \mathbf{Q}(k)\mathbf{\beta} + \mathbf{e}(k) \quad k = 1, \ldots, K, \tag{6}$$

where $\mathbf{y}(k) = [y_1(k), \ldots, y_n(k)]^{\mathrm{T}} \cdot \mathbf{Q}(k) = \mathrm{diag}\{\mathbf{Q}_1^{\mathrm{T}}(k), \ldots, \mathbf{Q}_n^{\mathrm{T}}(k)\}$. $\mathbf{Q}_j(k) = [q_i(k)]$ $(\forall i \in I_j)$ are $|I_j| \times 1$ vectors such that i is in an increasing order for $j = 1, \ldots, n$. $\mathbf{e}(k) = [e_1(k), \ldots, e_n(k)]^{\mathrm{T}}$ is a noise vector with zero-mean.

Denote the design matrix $[\mathbf{Q}^{\mathrm{T}}(1), \ldots, \mathbf{Q}^{\mathrm{T}}(k)]^{\mathrm{T}}$ at the kth time interval as $\mathbf{X}(k)$. Then

$$\mathbf{X}^{\mathrm{T}}(k)\mathbf{X}(k) = \mathbf{X}^{\mathrm{T}}(k-1)\mathbf{X}(k-1) + \mathbf{Q}^{\mathrm{T}}(k)\mathbf{Q}(k). \tag{7}$$

The identification problem of an O–D matrix $[b_{ij}]$ is usually formulated as solving the following constrained optimization problem (see Sherali et al., 1997; Li and De Moor, 1999)

$$\min \quad F(\boldsymbol{\beta},k)=\sum_{l=1}^{k}\sum_{j=1}^{n}\{y_j(l)-\mathbf{Q}_j^{\mathrm{T}}(l)\boldsymbol{\beta}_j\}^2 \quad k=1,\ldots,K \tag{8}$$

subject to (1a′) and (1b′).

3.2. Problem formulation for incomplete information

For situations with incomplete information, where there exists at least one entrance/exit with unavailable traffic counts, observation equations corresponding to Eqs. (6) become

$$y_j^*(k)=\sum_{i\in OI\cap I_j} q_i^*(k)b_{ij}+\sum_{i\in MI\cap I_j} q_i(k)b_{ij}+e_j(k) \quad \text{for } j\in OJ, \tag{9}$$

while there are no observation equations for those exits with $j\in MJ$.

Difficulties for identification with incomplete information arise from (9):

(i) when traffic counts at some exits are unavailable, the corresponding split parameters b_{ij} $(j\in MJ)$ are inestimable if we adopt cost function (8);

(ii) when observations at some of entrances are unavailable, (9) becomes non-linear observation equations of the unknown b_{ij} $(j\in J_i, i=1,\ldots,m)$ and $q_i(k)$ $(i\in MI)$.

If there exist unavailable observations at some entrances as well as some exits, i.e., a combination of (i) and (ii), problems become more complicated. In this paper, we do not follow the conventional research framework adopted by Sherali et al. (1997) and Li and De Moor (1999) but reformulate the problem.

The basic idea originates from observations on the Eqs. (9). Noting that Eqs. (9) are *non-linear* equations of both the unknown b_{ij} $(j\in J_i, i=1,\ldots,m)$ and $q_i(k)$ $(i\in MI)$, we can eliminate such non-linearity by introducing a set of new variables, $f_{ij}(k)=b_{ij}q_i(k)$ $(\forall i\in I_j, j=1,\ldots,n, k=1,\ldots,K)$, instead of two sets of the unknown variables $\{b_{ij}\}$ and $\{q_i(k)\}$. Eqs. (9) then become a set of linear equations of the unknown $f_{ij}(k)$ after transformation:

$$y_j^*(k)=\sum_{i\in I_j} f_{ij}(k)+e_j(k) \quad \text{for } j\in OJ. \tag{9′}$$

Next, to deal with the above inestimable problems of O–D matrices, we adopt Cascetta (1984) generalized least squares (GLS) approach to estimate O–D matrices by combining observation equations of traffic flows with prior information. Cascetta (1984) introduced GLS to improve survey-data based estimates of O–D matrices by using additional traffic counts. In this paper, we adopt this approach to improve traffic-count based least squares estimate by using additional priori information. Noting that the traffic-count based least squares will result in inestimable problems as mentioned above, the approach by Cascetta (1984) actually provides a solution of eliminating such inestimable problems.

Finally, to obtain recursive estimates, we choose a quadratic cost function at the kth time interval such that it is composed of two terms, one depending on the estimates of split parameters $\hat{b}_{ij}(k-1)$ obtained in the $(k-1)$th time interval, $f_{ij}(k)-\hat{b}_{ij}(k1)q_i(k)$, and another is a correction term based on new observation information $y_j^*(k)-\sum_{i\in I_j}f_{ij}(k)$ in the kth time interval. The split parameters in the kth time interval are then given by $\hat{b}_{ij}(k)=\hat{f}_{ij}(k)/\sum_{j\in J_i}\hat{f}_{ij}(k)$ after the estimates of traffic flows $\hat{f}_{ij}(k)$ are obtained.

Without loss of generality, we assume that observation information at the first r $(0 \leqslant r \leqslant m)$ entrances is unavailable, but at the latter $m - r$ entrances is available. For notational simplicity, we suppress the dependence on the time interval k in the later part of this section. For each time interval k $(k = 1, \ldots, K)$, we formulate identification problem of an O–D matrix as:

Given:
(i) available observations $y_j^*(j \in OJ)$ and q_i^* $(i = r+1, \ldots, m)$;
(ii) estimate of the O–D matrix in the $(k-1)$th time interval, $[\hat{b}_{ij}(k-1)]$, denoted as $[\mathrm{a}_{ij}]$;
(iii) weighting matrix $\mathbf{W} = [w_{ij}] > 0$.

To solve:

$$\min \quad F(\mathbf{f}, \mathbf{q}) = \sum_{j=1}^{n} \sum_{i \in I_j} w_{ij}[f_{ij} - \mathrm{a}_{ij} q_i]^2 + \sum_{j \in OJ} \left[y_j^* - \sum_{i \in I_j} f_{ij} \right]^2 \tag{10}$$

$$\text{subject to} \quad \sum_{j \in J_i} f_{ij} = q_i^* \quad i \in OI \tag{11a}$$

$$f_{ij} \geqslant 0 \quad j \in J_i, i = 1, \ldots, m \tag{11b}$$

and after (sub-)optimal solutions $\hat{f}_{ij}$ are obtained, the split parameters are estimated as

$$\hat{b}_{ij} = \hat{f}_{ij} \Big/ \sum_{j \in J_i} \hat{f}_{ij} \qquad j \in J_i, \quad i = 1, \ldots, m. \tag{12}$$

It is clear that the constraints (11a) correspond to (1) or equivalently, (4), after deleting those equations with $i \in MI$. The cost function (10) means that the estimates of traffic flows are determined by a_{ij} which were obtained in the previous time interval, and then are corrected by new observation information in the current time interval. The weighting matrix $\mathbf{W}$ represents the preference of a decision-maker for priori knowledge and current observation information. From statistical point of view, $\mathbf{W}$ should be chosen according to the covariance matrices of both terms in (10) (see Cascetta, 1984, for details). This will lead to a natural choice of the weighting matrix given in Theorem 2 below.

Let $\mathbf{a} = [\mathbf{a}_1^{\mathrm{T}}, \ldots, \mathbf{a}_m^{\mathrm{T}}]_{p \times 1}^{\mathrm{T}}$, where $\mathbf{a}_i$ are the jth transposed row of the matrix $[\mathrm{a}_{ij}] = [\hat{b}_{ij}(k-1)]$. Following theorems further show justification of the new formulation (10)–(12). The proofs of them are given in Appendix C.

Theorem 1. *For the identification problem of an O–D matrix with completely non-informative situations, i.e., traffic counts at all of entrances and exits are unavailable, the problem* (10)–(12) *gives estimates* $\hat{\mathbf{b}}_i = \mathbf{a}_i$ $(i = 1, \ldots, m)$.

Theorem 1 means that estimates will simply be intact if there is no new information available to correct them in the current time interval.

Theorem 2. *For the identification problem of an O–D matrix with complete observation information, if the design matrix* $\mathbf{X}$ *has full column rank, then the problem* (10)–(12) *gives the same solution as the problem Eqs.* (6) *and* (8) *when the weighting matrix is taken as*

$$\mathbf{W} = \mathbf{\Lambda}^{-1}\mathbf{\Gamma}^{\mathrm{T}}[\mathbf{X}^{\mathrm{T}}(-1)\mathbf{X}(-1)]\mathbf{\Gamma}\mathbf{\Lambda}^{-1},$$

where $\mathbf{X}(-1)$ *and* $\mathbf{X}$ *denote* $[\mathbf{Q}^{\mathrm{T}}(1),\ldots,\mathbf{Q}^{\mathrm{T}}(k-1)]^{\mathrm{T}}$ *and* $[\mathbf{Q}^{\mathrm{T}}(1),\ldots,\mathbf{Q}^{\mathrm{T}}(k)]^{\mathrm{T}}$, *respectively*.

Theorem 2 means that for situations with complete information, solution of the problem (10)–(12) coincides with the solution of the conventional research framework in Sherali et al. (1997) and Li and De Moor (1999) for an appropriate weighting matrix. This is desirable since the latter is a widely adopted research formulation for complete information situations and its solution has some nice properties, for instance, unbiasedness.

It should be noted that, however, unlike the situations where information is complete, the estimates of b_{ij} with incomplete information are non-linear from (12) and then are biased in general. There are exceptions when all of $q_i(k)$ are available and if the initial estimate $\hat{\mathbf{b}}(0)$ is unbiased. For those cases, both the estimates $\hat{f}_{ij}$ derived from Eqs. (10) and (11a) and $\hat{b}_{ij} = \hat{f}_{ij}/\sum_{j\in J_i}\hat{f}_{ij} = \hat{f}_{ij}/q_i^*$ are linear. This means that $\hat{b}_{ij}$ are unbiased if the error $e_j(k)$ in (9) have zero means. We then summarize the above result below.

Lemma 1. *Suppose* $MI = \phi$(*empty*). *If the initial estimate* $\hat{\mathbf{b}}(0)$ *is unbiased, and the errors* $e_j(k)$ *in* (9) *have zero means, then the estimates* $\hat{b}_{ij} = \hat{f}_{ij}/q_i^*$, *are linear and unbiased.*

Bias of estimates of the split parameters for more general situations where some of the entering counts are unavailable will be investigated in Section 5 through simulation study.

3.3. Vector version of problem formulation for incomplete information

A vector version of the formulation (10)–(12) is more convenient in the latter discussion. First of all, consider the equality constraints (11a). Let $\mathbf{R}$ be a matrix of $\mathbf{R}_0$ deleting the first r rows, i.e., $\mathbf{R} = [\mathbf{0}, \mathrm{diag}\{\mathbf{1}^{\mathrm{T}},\ldots,\mathbf{1}^{\mathrm{T}}\}]_{(m-r)\times p}$. Then (11a) can be rewritten as $\mathbf{Rf} = \mathbf{q}^*$.

Next, denote $\mathbf{G} = [\mathbf{G}^{(1)},\ldots,\mathbf{G}^{(m)}]_{|OJ|\times p}$ as a design matrix of (9′) in vector version

$$\mathbf{y}^* = \mathbf{Gf} + \tilde{\mathbf{e}}, \tag{13}$$

where $\tilde{\mathbf{e}}$ is a noise vector. See Appendix A for further discussions about $\mathbf{G}$.

Finally, let

$$\mathbf{f}_{0i} = \begin{cases} \mathbf{0} & \text{if } i \leqslant r \\ q_i^*\mathbf{a}_i & \text{if } i > r \end{cases} \quad \text{and} \quad \mathbf{f}_0 = [\mathbf{f}_{01}^{\mathrm{T}},\ldots,\mathbf{f}_{0m}^{\mathrm{T}}]_{p\times 1}^{\mathrm{T}}, \tag{14}$$

$$\mathbf{Z}_i = \begin{cases} \mathbf{I} - \mathbf{a}_i\mathbf{1}^{\mathrm{T}} & \text{if } i \leqslant r \\ \mathbf{I} & \text{if } i > r \end{cases} \quad \text{and} \quad \mathbf{Z} = \mathrm{diag}\{\mathbf{Z}_1,\ldots,\mathbf{Z}_m\} \tag{15}$$

and insert those equations of (4) with $i \in MI$ into (10) to eliminate the unknown q_i. The optimization problem (10)–(12) can then be rewritten as

$$\min \quad F(\mathbf{f}) = (\mathbf{f}-\mathbf{f}_0)^{\mathrm{T}}\mathbf{Z}^{\mathrm{T}}\mathbf{WZ}(\mathbf{f}-\mathbf{f}_0) + (\mathbf{y}^* - \mathbf{Gf})^{\mathrm{T}}(y^* - \mathbf{Gf}) \tag{10′}$$

$$\text{subject to} \quad \mathbf{Rf} = \mathbf{q}^*, \tag{11a′}$$

$$\mathbf{f} \geqslant \mathbf{0} \tag{11b'}$$

and the estimate of the split parameter vector is given by

$$\hat{\mathbf{b}} = \hat{\mathbf{\Lambda}}^{-1}\hat{\mathbf{f}}, \tag{12'}$$

where $\hat{\mathbf{\Lambda}} = \mathrm{diag}\{\hat{q}_1\mathbf{I}, \ldots, \hat{q}_m\mathbf{I}\}$ and $\hat{\mathbf{q}} = \mathbf{R}_0\hat{\mathbf{f}}$.

4. Algorithm

4.1. Algorithm for incomplete information

Noting that the problem (10′)–(12′) is an optimization problem with both equality-constraints and non-negative inequality constraints, we can solve it by using the algorithm proposed by Li and De Moor (1999) (see Appendix B for details) under certain conditions. Specifically, following the notations of the last section and suppressing the dependence on the time interval k, the Lagrangian equation is given by

$$L(\mathbf{f}; \boldsymbol{\lambda}, \boldsymbol{\mu}) = F(\mathbf{f}) + 2\boldsymbol{\lambda}^{\mathrm{T}}(\mathbf{R}\mathbf{f} - \mathbf{q}^*) - 2\boldsymbol{\mu}^{\mathrm{T}}\mathbf{f}.$$

The Kuhn–Tucker conditions give

$$[\mathbf{Z}^{\mathrm{T}}\mathbf{W}\mathbf{Z} + \mathbf{G}^{\mathrm{T}}\mathbf{G}]\mathbf{f} + \mathbf{R}^{\mathrm{T}}\boldsymbol{\lambda} = \mathbf{G}^{\mathrm{T}}\mathbf{y}^* + \mathbf{Z}^{\mathrm{T}}\mathbf{W}\mathbf{Z}\mathbf{f}_0 + \boldsymbol{\mu}$$

and $\mathbf{R}\mathbf{f} = \mathbf{q}^*$.

If

$$\begin{bmatrix} \mathbf{Z}^{\mathrm{T}}\mathbf{W}\mathbf{Z} + \mathbf{G}^{\mathrm{T}}\mathbf{G} & \mathbf{R}^{\mathrm{T}} \\ \mathbf{R} & \mathbf{0} \end{bmatrix}$$

is non-singular, then $\mathbf{f}$ and $\boldsymbol{\lambda}$ can be determined by the above equations. In particular, if both $\mathbf{U} = \mathbf{Z}^{\mathrm{T}}\mathbf{W}\mathbf{Z} + \mathbf{G}^{\mathrm{T}}\mathbf{G}$ and $\mathbf{R}\mathbf{U}^{-1}\mathbf{R}^{\mathrm{T}}$ are non-singular, we obtain

$$\hat{\mathbf{f}} = \hat{\mathbf{f}}_{LS} - \mathbf{U}^{-1}\mathbf{R}^{\mathrm{T}}[\mathbf{R}\mathbf{U}^{-1}\mathbf{R}^{\mathrm{T}}]^{-1}[\mathbf{R}\hat{\mathbf{f}}_{LS} - \mathbf{q}^*] + H\mathbf{i} \tag{16a}$$

and

$$\hat{\mathbf{f}}_{LS} = [\mathbf{Z}^{\mathrm{T}}\mathbf{W}\mathbf{Z} + \mathbf{G}^{\mathrm{T}}\mathbf{G}]^{-1}[\mathbf{G}^{\mathrm{T}}\mathbf{y}^* + \mathbf{Z}^{\mathrm{T}}\mathbf{W}\mathbf{Z}\mathbf{f}_0], \tag{16b}$$

where $\mathbf{H} = \mathbf{U}^{-1} - \mathbf{U}^{-1}\mathbf{R}^{\mathrm{T}}[\mathbf{R}\mathbf{U}^{-1}\mathbf{R}^{\mathrm{T}}]^{-1}\mathbf{R}\mathbf{U}^{-1}$.

If all of the main diagonal elements of $\mathbf{H} = [h_{ij}]_{p\times p}$ are greater than zero, the Bell correction can be used to choose an appropriate $\boldsymbol{\mu}$ (Li and De Moor, 1999).

The estimate $\hat{\mathbf{f}}$ in (16a) consists of three parts: the first one, $\hat{\mathbf{f}}_{LS}$, is the least squares solution of the cost function (10′); the second one, $\mathbf{U}^{-1}\mathbf{R}^{\mathrm{T}}[\mathbf{R}\mathbf{U}^{-1}\mathbf{R}^{\mathrm{T}}]^{-1}[\mathbf{R}\hat{\mathbf{f}}_{LS} - \mathbf{q}^*]$, is the equality-constraint correction for the least squares solution $\hat{\mathbf{f}}_{LS}$ corresponding to (11a′); the third term, $\mathbf{H}\boldsymbol{\mu}$, for the inequality-constraints (11b′) is used to further improve the solution through the Bell correction. See Li and De Moor (1999).

Denote $\mathbf{T}(\mathbf{v})$ as a vector operator truncating negative elements of $\mathbf{v}$ to zeros. If

$$\begin{bmatrix} \mathbf{Z}^{\mathrm{T}}\mathbf{W}\mathbf{Z} + \mathbf{G}^{\mathrm{T}}\mathbf{G} & \mathbf{R}^{\mathrm{T}} \\ \mathbf{R} & \mathbf{0} \end{bmatrix}$$

is non-singular, and all of the main diagonal elements of $\mathbf{H} = [h_{ij}]_{p\times P}$ are greater than zero, then the algorithm of estimation of an O–D matrix with incomplete information is given below:

Given:

(i) initial estimate of the O–D matrix, $\hat{\mathbf{b}}(0) = [\hat{\mathbf{b}}_1^{\mathrm{T}}(0), \ldots, \hat{\mathbf{b}}_m^{\mathrm{T}}(0)]_{p\times 1}^{\mathrm{T}}$;

(ii) structure matrix for equality-constraints $\mathbf{R}_0$;

(iii) 0/1 matrix $\mathbf{G}$ constituted by coefficients of the observation equations (13);

(iv) weighting matrix $\mathbf{W}(l)$ for $l = 1, \ldots, K$.

Step 0. *Initialization*.

(i) Rearranging the indices of the entrances such that

$$MI = \{1, \ldots, r\} \quad \text{and} \quad OI = \{r+1, \ldots, m\} \ \ (0 \leqslant r \leqslant m).$$

(ii) Deleting the first r rows of $\mathbf{R}_0$ and denoting it as $\mathbf{R} = [\mathbf{0},\ \text{diag}\ \{\mathbf{1}^{\mathrm{T}}, \ldots, \mathbf{1}^{\mathrm{T}}\}]_{(m-r)\times p}$.

Loop for $k = 1, \ldots, K$.

Step 1. For available observations $y_j^*(l)$ $(j \in OJ)$ and $q_i^*(l)$ $(i \in OI)$ $(l = 1, \ldots, k)$, computing

$$\mathbf{Z}_i(k) = \begin{cases} \mathbf{I} - \hat{\mathbf{b}}_i(k-1)\mathbf{1}^{\mathrm{T}} & \text{if } i \leqslant r \\ \mathbf{I} & \text{if } i > r \end{cases} \quad \text{and} \quad \mathbf{Z}(k) = \text{diag}\ \{\mathbf{Z}_1(k), \ldots, \mathbf{Z}_m(k)\},$$

$$\mathbf{f}_{0i}(k) = \begin{cases} \mathbf{0} & \text{if } i \leqslant r \\ q_i^*(k)\hat{\mathbf{b}}_i(k-1) & \text{if } i > r \end{cases} \quad \text{and} \quad \mathbf{f}_0(k) = [\mathbf{f}_{01}^{\mathrm{T}}(k), \ldots, \mathbf{f}_{0m}^{\mathrm{T}}(k)]_{p\times 1}^{\mathrm{T}},$$

$$\mathbf{U}(k) = \mathbf{Z}^{\mathrm{T}}(k)\mathbf{W}(k)\mathbf{Z}(k) + \mathbf{G}^{\mathrm{T}}\mathbf{G}$$

and computing least squares estimate

$$\hat{\mathbf{f}}_{LS}(k) = \mathbf{U}(k)^{-1}[\mathbf{G}^{\mathrm{T}}\mathbf{y}^*(k) + \mathbf{Z}^{\mathrm{T}}(k)\mathbf{W}(k)\mathbf{Z}(k)\mathbf{f}_0(k)].$$

Step 2. Correction of equality-constraints

$$\bar{\mathbf{f}}(k) = \hat{\mathbf{f}}_{LS}(k) - \mathbf{U}^{-1}(k)\mathbf{R}^{\mathrm{T}}[\mathbf{R}\mathbf{U}^{-1}(k)\mathbf{R}^{\mathrm{T}}]^{-1}[\mathbf{R}\hat{\mathbf{f}}_{LS}(k) - \mathbf{q}^*(k)].$$

Step 3. Computation of the Bell correction

$$\mathbf{H}(k) = \mathbf{U}^{-1}(k) - \mathbf{U}^{-1}(k)\mathbf{R}^{\mathrm{T}}[\mathbf{R}\mathbf{U}^{-1}(k)\mathbf{R}^{\mathrm{T}}]^{-1}\mathbf{R}\mathbf{U}^{-1}(k) = [h_{ij}(k)] \text{ and } \mathbf{D}(k) = \text{diag}\{1/h_{ii}(k)\},$$

$$\boldsymbol{\mu}^*(k) = \mathbf{T}(-\mathbf{D}(k)\bar{\mathbf{f}}(k)),$$

$$\tilde{\mathbf{f}}(k) = \bar{\mathbf{f}}(k) + \mathbf{H}(k)\mathbf{i}^*(k).$$

Step 4. Enforcing non-negative constraints (11b′) if necessary: $\hat{\mathbf{f}}(k) = \mathbf{T}(\tilde{\mathbf{f}}(k))$.

Step 5. Computing $\hat{b}_{ij}(k) = \hat{f}_{ij}(k) / \sum_{j\in J_i} \hat{f}_{ij}(k)$ $j \in J_i, i = 1, \ldots, m$

$k = k + 1$.

Until $k = K$.

Remark 1. As mentioned in the last section, a natural choice of weighting matrix is

$$\mathbf{W}(k) = \boldsymbol{\Lambda}^{-1}(k)\boldsymbol{\Gamma}^{\mathrm{T}}[\mathbf{X}^{\mathrm{T}}(k-1)\mathbf{X}(k-1)]\boldsymbol{\Gamma}\boldsymbol{\Lambda}^{-1}(k),$$

where $\mathbf{X}^{\mathrm{T}}(k-1)\mathbf{X}(k-1) = \mathbf{X}^{\mathrm{T}}(k-2)\mathbf{X}(k-2) + \mathbf{Q}^{\mathrm{T}}(k-1)\mathbf{Q}(k-1)$ from (7), and $\hat{q}_j(k-1)$ $(j = 1, \ldots, n)$ in $\mathbf{Q}(k-1)$ are available in the time interval k.

When there exist unobserved entering counts $q_i(k)(i = 1, \ldots, r)$, those unobserved values $q_i(k)$ $(i = 1, \ldots, r)$ can be replaced by their appropriate estimated values, for instance, a weighted average of the estimates $\hat{q}_i(k)$ in the previous time intervals.

4.2. Non-singularity conditions

The positivity of the main diagonal elements of $\mathbf{H}$, h_{ii} $(i = 1, \ldots, n)$, for (16) are ensured by the following result. Its proof is similar to the one of lemma in Li and De Moor (1999).

Lemma 2. *Suppose that* $\mathbf{U}^{-1}$ *is a* $p \times p$ *positive definite matrix, and* $\mathbf{R} = [\mathbf{0},\ \mathrm{diag}\ \{\mathbf{1}^{\mathrm{T}}, \ldots, \mathbf{1}^{\mathrm{T}}\}]_{(m-r)\times p}$, *where the length* m_l *of the lth* $\mathbf{1}$ *of* $\mathbf{R}$ *satisfies* $m_l > 1$. *Then*

$$\mathbf{H} = \mathbf{U}^{-1} - \mathbf{U}^{-1}\mathbf{R}^{\mathrm{T}}[\mathbf{R}\mathbf{U}^{-1}\mathbf{R}^{\mathrm{T}}]^{-1}\mathbf{R}\mathbf{U}^{-1} \geqslant 0$$

and its main diagonal elements are positive, i.e., $h_{ii} > 0$ $(i = 1, \ldots, m)$.

The non-singularity conditions of the problem (10′)–(12′) are given by:

Theorem 3. *Let* $MI = \{1, \ldots, r\}$ $(0 \leqslant r \leqslant m), \mathbf{W} > 0, \mathbf{a} = [\mathbf{a}_1^{\mathrm{T}}, \ldots, \mathbf{a}_m^{\mathrm{T}}]^{\mathrm{T}}_{p\times 1}$ *be the estimate of an O–D matrix obtained in the* $(k-1)$*th time interval. Then for* $\mathbf{\Pi} = [\mathbf{G}^{(1)}\mathbf{a}_1, \ldots, \mathbf{G}^{(r)}\mathbf{a}_r]$ *and* $\mathbf{\Phi} = [\mathbf{Z}^{\mathrm{T}}, \mathbf{G}^{\mathrm{T}}]^{\mathrm{T}}$, *where* $\mathbf{G}$ *and* $\mathbf{Z}$ *are given by Eqs.* (13) *and* (15), *respectively, we have*

(i)

$$\begin{bmatrix} \mathbf{Z}^{\mathrm{T}}\mathbf{W}\mathbf{Z} + \mathbf{G}^{\mathrm{T}}\mathbf{G} & \mathbf{R}^{\mathrm{T}} \\ \mathbf{R} & \mathbf{0} \end{bmatrix}$$

is non-singular if and only if $\mathbf{\Phi} = \left[\mathbf{Z}^{\mathrm{T}}, \mathbf{G}^{\mathrm{T}}\right]^{\mathrm{T}}$ *has full column rank*;

(ii) $\mathbf{\Phi}$ *has full column rank if and only if rank* $(\mathbf{D}) = r$;

(iii) $\mathbf{U}$ *and* $\mathbf{R}\mathbf{U}^{-1}\mathbf{R}^{\mathrm{T}}$ *are non-singular of rank* $(\mathbf{D}) = r$.

The proof of Theorem 3 is given in Appendix C. For example, for a 4×4 intersection prohibiting U- turn, (a) if only $q_1(k)$ and $y_1(k)$ are unavailable, then $\mathbf{\Pi} = \mathbf{G}^{(1)}\mathbf{a}_1 = [a_{12}, a_{13}, a_{14}]^{\mathrm{T}}$ and rank $(\mathbf{\Pi}) \equiv 1$ from Eqs. (1); (b) if only $q_1(k)$ and $y_j(k)$ $(j > 1$, say $j = 4)$ are unavailable, then $\mathbf{\Pi} = \mathbf{G}^{(1)}\mathbf{a}_1 = [0, a_{12}, a_{13}]^{\mathrm{T}}$. Hence, rank $(\mathbf{\Pi}) = 1$ if and only if $a_{14} \neq 1$; (c) if only $q_1(k)$ and $y_j(k)$ $(j = 2, 3, 4)$ are unavailable, then $\mathbf{G}^{(1)} \equiv [0, 0, 0]$, and rank $(\mathbf{\Pi}) = \mathrm{rank}(\mathbf{G}^{(1)}\mathbf{a}_1) \equiv 0$.

We then distinguish three different situations:

(i) $\mathrm{rank}(\mathbf{\Pi}) \equiv r$ for any $\mathbf{a}_i$ satisfying (1).

This corresponds to the situations where there exists sufficient information for updating the O–D matrix (for instance, case (a) in the above example).

(ii) rank $(\mathbf{\Pi}) < r$ for some of $\mathbf{a}_i$ satisfying (1).

It corresponds to the situations where there may exist some time intervals we cannot uniquely determine estimates of traffic flows (for instance, case (b) in the above example). In practice, if the

condition number of the matrix $\mathbf{U} = \mathbf{Z}^{\mathrm{T}}\mathbf{W}\mathbf{Z} + \mathbf{G}^{\mathrm{T}}\mathbf{G}$ is large, for instance, over 1000 (Myers, 1986, pp. 220), we suggest adopting a ridge- regression approach in these time intervals. Note that from the proof of Theorem 2 (Eq. (C.2)), singularity of $\mathbf{U} = \mathbf{Z}^{\mathrm{T}}\mathbf{W}\mathbf{Z} + \mathbf{G}^{\mathrm{T}}\mathbf{G}$ is equivalent to singularity of $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ for situations with complete information. This means that case (ii) is possible even for complete information since $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ may occasionally be singular.

(iii) $\mathrm{rank}(\mathbf{\Pi}) < r$ for any $\mathbf{a}_i$ satisfying (1).

It corresponds to the situations of lacking enough information to improve prior estimate of an O–D matrix. We term an O–D matrix as structurally non-updateable in this case.

A sufficient condition of structural non-updateability for an O–D matrix with incomplete information is that the number of rows of the matrix $\mathbf{\Pi}$ is less than the number of columns. This means that the number of available observation equations (9′) is less than the number of unobserved entrances, or precisely, $|OJ| < |MI|$. In these situations, new information obtained in a time interval, say the kth, is even not enough to determine the unknown $q_i(k)$ $(i \in MI)$, and therefore impossible to update the split parameters b_{ij}.

Note that structural non-updateability is defined for an O–D matrix rather than its individual elements. Occasionally, some of the elements of an O–D matrix may still be estimable even if an O–D matrix itself is non- updateable. For instance, for case (c), the obtained observation equation is

$$y_1^*(k) = b_{21}q_2^*(k) + b_{31}q_3^*(k) + b_{41}q_4^*(k) + e_1(k),$$

which does not provide any information for updating, for instance, b_{12}. Hence, the O–D matrix is non-updateable. However, $(b_{21},\ b_{31},\ b_{41})$ is updateable whenever new observations $\{y_1^*(k), q_2^*(k), q_3^*(k), q_4^*(k)\}$ are obtained at the kth time interval $(k = 1, \ldots, K)$.

Matrix $\mathbf{\Pi}$ represents a characteristic of identification problems of O–D matrices with incomplete information. For the situations with incomplete information, we can firstly determine updateability of an O–D matrix by $\mathbf{\Pi}$ before any further analysis and computation.

4.3. Alternative for computation of $(\mathbf{Z}^{\mathrm{T}}\mathbf{W}\mathbf{Z} + \mathbf{G}^{\mathrm{T}}\mathbf{G})^{-1}$

For simplicity, we suppress the dependence on the time interval k as before in the rest of this section. We will investigate an alternative of computing $[\mathbf{Z}^{\mathrm{T}}\mathbf{W}\mathbf{Z} + \mathbf{G}^{\mathrm{T}}\mathbf{G}]^{-1}$ instead of direct computation of the inverse of $\mathbf{U} = \mathbf{Z}^{\mathrm{T}}\mathbf{W}\mathbf{Z} + \mathbf{G}^{\mathrm{T}}\mathbf{G}$. Note that we cannot adopt the usual approach, i.e., using Eq. (18) below to compute inverse of $\mathbf{U}$ since both $\mathbf{Z}^{\mathrm{T}}\mathbf{W}\mathbf{Z}$ and $\mathbf{G}^{\mathrm{T}}\mathbf{G}$ are normally singular. First of all, we state a lemma for inverse of a partitioned matrix.

Lemma 3. *Let*

$$\mathbf{U} = \begin{bmatrix} \mathbf{U}_{11} & \mathbf{U}_{12} \\ \mathbf{U}_{21} & \mathbf{U}_{22} \end{bmatrix}$$

and $\mathbf{U}_{11.2} = \mathbf{U}_{11} - \mathbf{U}_{12}\mathbf{U}_{22}^{-1}\mathbf{U}_{21}$, *where* $\mathbf{U}_{22}$ *is non-singular, then* $\mathbf{U}_{11.2}$ *is non-singular, and*

$$\text{(i)}\quad \mathbf{U}^{-1} = \begin{bmatrix} \mathbf{U}_{11.2}^{-1} & -\mathbf{U}_{11.2}^{-1}\mathbf{U}_{12}\mathbf{U}_{22}^{-1} \\ -\mathbf{U}_{22}^{-1}\mathbf{U}_{21}\mathbf{U}_{11.2}^{-1} & \mathbf{U}_{22}^{-1} + \mathbf{U}_{22}^{-1}\mathbf{U}_{21}\mathbf{U}_{11.2}^{-1}\mathbf{U}_{12}\mathbf{U}_{22}^{-1} \end{bmatrix}, \tag{17}$$

$$\text{(ii)}\quad [\mathbf{U}_{22}+\mathbf{U}_{21}\mathbf{U}_{11}\mathbf{U}_{12}]^{-1}=\mathbf{U}_{22}^{-1}-\mathbf{U}_{22}^{-1}\mathbf{U}_{21}[\mathbf{U}_{12}\mathbf{U}_{22}^{-1}\mathbf{U}_{21}+\mathbf{U}_{11}]^{-1}\mathbf{U}_{12}\mathbf{U}_{22}^{-1}. \tag{18}$$

The proof is immediate by direct computations.

To apply Lemma 3, let $\mathbf{Z}=\text{diag}\,\{\mathbf{D}_1,\mathbf{D}_2\}$, where $\mathbf{D}_1=\text{diag}\{\mathbf{Z}_1,\ldots,\mathbf{Z}_r\}$ and $\mathbf{D}_2=\text{diag}\,\{\mathbf{I},\ldots,\mathbf{I}\}$. Denote

$$\mathbf{W}=\begin{bmatrix}\mathbf{W}_{11} & \mathbf{W}_{12}\\ \mathbf{W}_{21} & \mathbf{W}_{22}\end{bmatrix}$$

and $\mathbf{G}=[\mathbf{G}_1,\mathbf{G}_2]$ as partitions corresponding to $\mathbf{Z}$. Then we have

$$\mathbf{U}=\mathbf{Z}^{\mathrm{T}}\mathbf{W}\mathbf{Z}+\mathbf{G}^{\mathrm{T}}\mathbf{G}=\begin{bmatrix}\mathbf{D}_1\mathbf{W}_{11}\mathbf{D}_1+\mathbf{G}_1^{\mathrm{T}}\mathbf{G}_1 & \mathbf{D}_1\mathbf{W}_{12}+\mathbf{G}_1^{\mathrm{T}}\mathbf{G}_2\\ \mathbf{W}_{21}\mathbf{D}_1+\mathbf{G}_2^{\mathrm{T}}\mathbf{G}_1 & \mathbf{W}_{22}+\mathbf{G}_2^{\mathrm{T}}\mathbf{G}_2\end{bmatrix}\hat{=}\begin{bmatrix}\mathbf{U}_{11} & \mathbf{U}_{12}\\ \mathbf{U}_{21} & \mathbf{U}_{22}\end{bmatrix}.$$

Let $\mathbf{V}$ denote $\mathbf{W}_{22}^{-1}$. Then if $\mathbf{V}$ is known, from Eq. (18) we have

$$\mathbf{U}_{22}^{-1}=(\mathbf{W}_{22}+\mathbf{G}_2^{\mathrm{T}}\mathbf{G}_2)^{-1}=\mathbf{V}-\mathbf{V}\mathbf{G}_2^{\mathrm{T}}[\mathbf{G}_2\mathbf{V}\mathbf{G}_2^{\mathrm{T}}+\mathbf{I}]^{-1}\mathbf{G}_2\mathbf{V}. \tag{19}$$

Thus, if we have simple means to compute $\mathbf{V}=\mathbf{W}_{22}^{-1}$, computation of $\mathbf{U}^{-1}$ can be converted into two inverse problems with small dimensions, $\mathbf{U}_{11.2}^{-1}$ and $[\mathbf{G}_2\mathbf{V}\mathbf{G}_2^{\mathrm{T}}+\mathbf{I}]^{-1}$ by Eqs. (18) and (19). This approach has advantages when the dimension of the matrix $\mathbf{U}$ is relatively large while the dimension of the matrix $\mathbf{U}_{11}$ is small. For instance, for a 4×4 intersection prohibiting U-turn, $\mathbf{U}$ is a 12×12 matrix. If there exist one entrance and one exit without observation information, we need to compute inverses of two 3×3 matrices by using the above approach.

Finally, we discuss availability of inverse of $\mathbf{W}_{22}$ if we adopt the natural choice of weighting matrix $\mathbf{W}=\mathbf{\Lambda}^{-1}\mathbf{\Gamma}^{\mathrm{T}}[\mathbf{X}(-1)^{\mathrm{T}}\mathbf{X}(-1)]\mathbf{\Gamma}\mathbf{\Lambda}^{-1}$.

Let $\mathbf{\Gamma}=[\mathbf{\Gamma}_1,\mathbf{\Gamma}_2]$ and $\mathbf{\Lambda}=\text{diag}\{\mathbf{\Lambda}_1,\mathbf{\Lambda}_2\}$ be partitioned corresponding to the matrix $\mathbf{Z}$. Then for $\tilde{\mathbf{X}}=\mathbf{X}\mathbf{\Gamma}_2$ which consists of those columns of $\mathbf{X}$ corresponding to the entrances with available traffic counts, we have

$$\mathbf{W}_{22}=\mathbf{\Lambda}_2^{-1}\mathbf{\Gamma}_2^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}(-1)\mathbf{X}(-1)\mathbf{\Gamma}_2\mathbf{\Lambda}_2^{-1}=\mathbf{\Lambda}_2^{-1}\tilde{\mathbf{X}}^{\mathrm{T}}(-1)\tilde{\mathbf{X}}(-1)\mathbf{\Lambda}_2^{-1}$$

and

$$\mathbf{W}_{22}^{-1}=\mathbf{\Lambda}_2[\mathbf{\Gamma}_2^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}(-1)\mathbf{X}(-1)\mathbf{\Gamma}_2]^{-1}\mathbf{\Lambda}_2=\mathbf{\Lambda}_2[\tilde{\mathbf{X}}^{\mathrm{T}}(-1)\tilde{\mathbf{X}}(-1)]^{-1}\mathbf{\Lambda}_2$$

Since $[\tilde{\mathbf{X}}^{\mathrm{T}}(-1)\tilde{\mathbf{X}}(-1)]^{-1}$ can be computed recursively, $\mathbf{W}_{22}^{-1}$ can be easily obtained.

5. Simulation examples

In this section, simulation examples are provided to illustrate dynamic identification of O–D matrices with incomplete information.

5.1. Simulation examples for different information patterns

Consider a typical four-way intersection prohibiting U-turn, $n=m=4$, as shown in Fig 1.

When information is complete, there exist eight information sources: four for entering counts $q_i(k)$ $(i=1,\ldots,4)$ and four for exiting counts $y_j(k)$ $(j=1,\ldots,4)$. Let iEjL denote a missing data

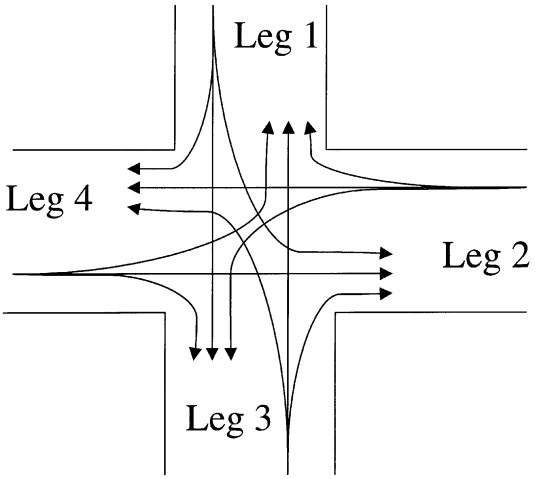

Fig. 1. A typical four-way intersection.

pattern of which traffic counts are unavailable only at the ith entrance and at the jth exit. Let i (or j) $=0$ if no entrance (or exit) with traffic counts unavailable. For instance, 1&2E3L denotes traffic counts are unavailable at the first and second entering legs, and at the third leaving leg; 0E1L denotes traffic counts unavailable only at the first leaving leg.

In this section, we investigate relations of estimations and information patterns. Specifically, following information structures are considered: the patterns 0E0L, 1E0L, 0E1L, 0E1&2L, 1&2E0L, 1E3L. The first one is the complete information pattern used for comparison purpose, while the second and third are patterns with only one unavailable information source, and the last three are patterns with two unavailable information sources.

The "true" parameter matrix $\mathbf{B}$ is given by

$$\mathbf{B} = \begin{bmatrix} 0 & 0.230 & 0.414 & 0.356 \\ 0.051 & 0 & 0.149 & 0.800 \\ 0.352 & 0.358 & 0 & 0.290 \\ 0.083 & 0.843 & 0.074 & 0 \end{bmatrix},$$

which was obtained by Nihan and Davis (1989) for an intersection located at 14th Avenue NW and NW Market Street in Seattle from 6:00 a.m. to 10:00 p.m. of a typical weekday. The simulation data are generated as described by Nihan and Davis (1989). Entering counts $q_i(k)$ are generated as the outcome of a Poisson random variable with mean 100, and the turning movement flows are generated as the outcome of a multinomial random vector with $M(q_i(k), \mathbf{b}_i)$ $(i = 1, \ldots, 4)$. Exiting counts $y_j(k)$ are then given as the sums of all traffic flows exiting at the jth leg during the kth time interval.

Simulation length is set as $M + K = 320$ in which the first $M = 20$ time intervals are used for initial estimation and the latter $K = 300$ are used for performance comparison. The criterion on estimation accuracy is taken as mean absolute value distance over the latter K periods

$$MAVD = \sum_{k=M+1}^{M+K} \left[\sum_{j=1}^{p} \left| \hat{b}^{(j)}(k) - b^{(j)} \right| \Big/ p \right] \Big/ K,$$

where $\mathbf{b} = [b^{(j)}]_{p\times 1} (p = 12)$ is the true split parameter vector constituted by the transposed rows of $\mathbf{B}$ after deleting the constant zeros of the diagonal elements, and $\hat{\mathbf{b}}(k) = [\hat{b}^{(j)}(k)]_{p\times 1}$ is an estimate of the true parameter vector during the kth time interval.

In practice, initial estimates should be taken as prior estimates. In this simulation study, we assume that complete information in the first $M = 20$ time intervals is available, and then adopt least squares estimate as the initial estimate $\hat{b}_{ij}(20)$ and $\mathbf{X}^{\mathrm{T}}(20)\mathbf{X}(20)$ (note that simulations start from $k = 21$ to 320). The weighting matrices are taken as the natural choice, as given by Remark 1.

Fig. 2 gives typical behaviors of estimation accuracy of split parameters across period of simulation time for the 6 information patterns. Theoretically, when some information is lost, the corresponding estimates will have relatively larger variances than the estimates based on complete information. This is verified by Fig. 2. The estimates with one information source unavailable (0E1L and 1E0L) are worse than the estimates with complete information (0E0L), while are better than the estimates with two information sources unavailable (1&2E0L, 0E1&2L, and 1E3L). Hence, the more information we have, the better estimates are in general.

Further simulation experiments show that, if traffic counts in more entrances or exits are unavailable, estimates deteriorate drastically, sometimes even worse than the initial estimates which are obtained by least squares approach based on complete information in the first $M = 20$ time intervals.

The simulation results given in this section show that, for structurally updateable problems, we can obtain estimates of an O–D matrix even if some observation information is unavailable. Estimate accuracy depends on the amount of lost information.

5.2. Simulation study for mean bias

In this section, we investigate mean bias of estimates through simulation.

$L = 20$ and 50 simulation experiments are conducted respectively. The entering and leaving counts are generated as described in the last sub-section. Mean bias is defined as

$$MB(b^{(j)}, k) = \sum_{l=1}^{L} \left\{ \hat{b}^{(j)}(k; l) - b^{(j)} \right\} \Big/ L,$$

where $\hat{b}^{(j)}(k; l)$ is the jth element of the estimate of split parameters at the kth time interval in the lth simulation experiment.

Table 1 gives simulation results for part of split parameters.

First of all, it should be noted that from Lemma 1, estimates for information patterns 0E0L, 0E1L, 0E1&2L are unbiased since all $q_i(k)$ are available for these information patterns. The biases of these estimates then form a basis for comparison with other information patterns where not all $q_i(k)$ are available.

From Table 1 it seems that mean biases of the estimates with the information patterns 1E0L, 1&2E0L, and 1E3L are a little bit larger than the ones with the information patterns 0E0L, 0E1L,

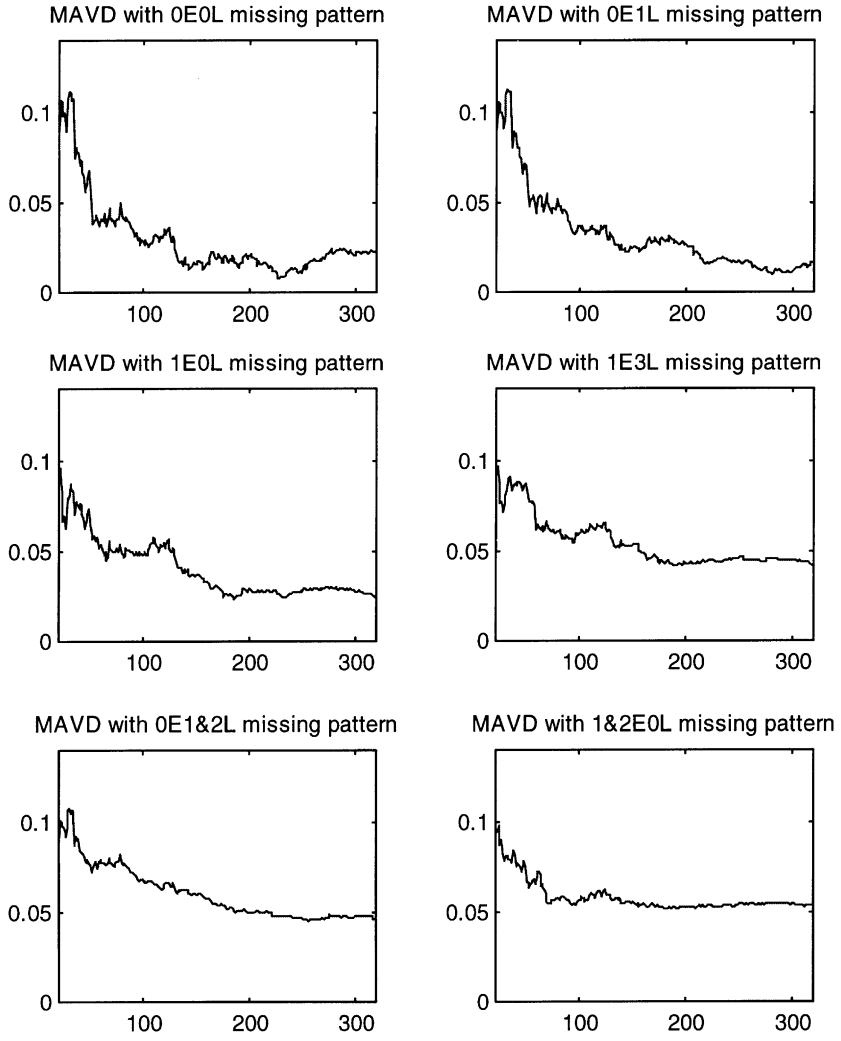

Fig. 2. Estimation accuracy for different missing data patterns.

0E1&2L. This suggests that there possibly exist some biases for those cases although they are quite small.

6. Conclusions and discussions

For the situations with incomplete observation information, where traffic counts are unavailable in some entrances and/or exits of a traffic system, a research formulation is proposed for identification of O–D matrices. An identification problem of O–D matrices with incomplete information is then reformulated as an optimization problem with linear equality and non-negative inequality constraints. An algorithm is suggested to solve this problem.

Table 1
Bias of estimates of split parameters for different information patterns

		MB(b_{12}, k)			MB(b_{43}, k)		
	k	100	200	300	100	200	300
L = 20	0E0L	−0.0012	0.0028	0.0033	−0.0146	−0.0061	−0.0050
	0E1L	−0.0029	0.0026	0.0037	−0.0104	−0.0041	−0.0034
	1E0L	0.0075	0.0074	0.0078	−0.0109	−0.0064	−0.0075
	0E1&2L	−0.0133	−0.0077	−0.0071	−0.0176	−0.0085	−0.0098
	1&2E0L	−0.0211	−0.0124	−0.0122	−0.0100	−0.0079	−0.0067
	1E3L	−0.0143	−0.0146	−0.0138	−0.0314	−0.0310	−0.0279
L = 50	0E0L	0.0051	−0.0031	−0.0061	0.0032	−0.0026	−0.0023
	0E1L	0.0055	−0.0024	−0.0057	0.0019	−0.0049	−0.0032
	1E0L	−0.0020	0.0003	0.0001	−0.0018	−0.0037	−0.0001
	0E1&2L	−0.0106	−0.0102	−0.0104	0.0010	−0.0024	−0.0009
	1&2E0L	−0.0246	−0.0231	−0.0219	0.0084	0.0081	0.0087
	1E3L	−0.0068	−0.0051	−0.0058	−0.0069	−0.0051	−0.0060

So far our discussions have been restricted to those situations where an O–D matrix to be estimated is constant. This can be relaxed to time-varying O–D matrices if the O–D matrix is updateable for given information pattern. The simplest approach is to introduce a forgetting factor $0 < \rho < 1$ in the objective function $(10')$ to give the new observation larger weights. Another often used approach is to use rolling horizon framework (Sherali et al., 1997) in which, as time progresses, the horizon used for estimating the split parameters is rolled forward by at least one additional period. Finally, a more delicate approach may be based on appropriately updating the weighting matrix $\mathbf{W}(k)$ in which $\mathbf{X}^{\mathrm{T}}(k)\mathbf{X}(k)$ are computed for given forgetting factor like those algorithms developed by Cremer and Keller (1987) and Nihan and Davis (1987). Comparison of performances for the above approaches needs more space for investigations and discussions, and then will be the future work. However, two points are quite clear for time- varying O–D matrices. First, if an O–D matrix is not structurally updateable, such tracking is impossible since there exists at least one split parameter whose estimate can not be improved. Second, even if an O–D matrix is structurally updateable, tracking a time- varying O–D matrix with incomplete information will be much more difficult than with complete information. The available information for situations with incomplete observations is less than the one with complete observations at each time interval, which results in larger variances of estimates and worse tracking performances.

Another important issue for dynamically estimating an O–D matrix is the journey time τ that a vehicle needs to traverse through a traffic system. In this paper, it is assumed that τ is small and then omitted. This is reasonable for an intersection or a short freeway section. As the referee pointed out, when entering volumes $q_{ij}(k)$ are not so large such that linear relations (6) still maintain, extension may be made to more general situations, where the journey time through a traffic system is considered. Those extensions may be based on the two approaches of Bell (1991b), one for geometrically distributed travel times, and another for freely distributed travel times. The advantage of the both approaches is that we do not need to assume availability of the journey time from one entrance to one exit.

Acknowledgements

The authors are grateful for the referee's valuable suggestions for the earlier version of this paper. This work was supported by Sustainable Mobility Programme – Project MD/01/24 (Traffic Congestion Problems in Belgium: Mathematical Models, Analysis, Simulation, Control and Actions), Concerted Research Action GOA-MIPS, the FWO project G.0256.97: Numerical Algorithms for Subspace System Identification, the FWO Research Communities: ICCOS and Advanced Numerical Methods for Mathematical Modeling, and the Belgian Government: IUAP P4-02 and P4- 24.

Appendix A. Definition of G

For situations with complete information, $\mathbf{G}$ is defined as

$$\mathbf{G} = \mathbf{Q}\boldsymbol{\Gamma}\boldsymbol{\Lambda}^{-1}, \tag{A.1}$$

which is obtained by inserting (3) and (5) into (13), and comparing (13) with (6).

For situations with incomplete information, where traffic counts are unavailable at some exits, $\mathbf{G}$ is given by (A.1), deleting those rows with $j \in MJ$.

To fix the idea, consider a 4×4 intersection prohibiting U-turn. For situations with complete information, we have $\mathbf{G} = \left[\mathbf{G}^{(1)}, \mathbf{G}^{(2)}, \mathbf{G}^{(3)}, \mathbf{G}^{(4)}\right]$, where

$$\mathbf{G}^{(1)} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \mathbf{G}^{(2)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \mathbf{G}^{(3)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \mathbf{G}^{(4)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

If traffic counts are unavailable at, for instance, the third exit, then the corresponding $\mathbf{G}$ is given by deleting the third row of the above matrices.

Appendix B. A brief summary of Li and De Moor's algorithm

Li and De Moor (1999) proposed a recursion-based algorithm for dynamic estimation of an O–D matrix. For the problem (8), the Lagrangian equation is given by

$$L(\boldsymbol{\beta}, k; \boldsymbol{\lambda}, \boldsymbol{\mu}) = F(\boldsymbol{\beta}, k) + 2\boldsymbol{\lambda}^{\mathrm{T}}(\tilde{\mathbf{R}}_{\mathbf{0}}\boldsymbol{\beta} - 1) - 2\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\beta}.$$

The Kuhn–Tucker conditions gives

$$\hat{\boldsymbol{\beta}}(k) = \hat{\boldsymbol{\beta}}_{LS}(k) - \boldsymbol{\Phi}(k)\tilde{\mathbf{R}}_0^{\mathrm{T}}[\tilde{\mathbf{R}}_0\boldsymbol{\Phi}(k)\tilde{\mathbf{R}}_0^{\mathrm{T}}]^{-1}[\tilde{\mathbf{R}}_0\hat{\boldsymbol{\beta}}_{LS}(k) - \mathbf{1}] + \mathbf{H}(k)\boldsymbol{\mu}, \tag{B.1}$$

where $\boldsymbol{\Phi}(k) = \{\mathbf{X}(k)^{\mathrm{T}}\mathbf{X}(k)\}^{-1}$, $\mathbf{H}(k) = \boldsymbol{\Phi}(k) - \boldsymbol{\Phi}(k)\tilde{\mathbf{R}}_0^{\mathrm{T}}[\tilde{\mathbf{R}}_0\boldsymbol{\Phi}(k)\tilde{\mathbf{R}}_0^{\mathrm{T}}]^{-1}\tilde{\mathbf{R}}_0\boldsymbol{\Phi}(k)$, $\hat{\boldsymbol{\beta}}_{LS}(k)$ is the least squares estimate of $\boldsymbol{\beta}$ based on the observations (6) which can be estimated recursively, and $\boldsymbol{\mu}$ is determined by the Bell correction. The algorithm is given below.

Given: Structure matrix for equality-constraints $\tilde{\mathbf{R}}_{\mathbf{0}}$; $y_j(k)$ $(j = 1, \ldots, n)$ and $q_i(k)$ $(i = 1, \ldots, m)$.

Step 1. Recursively computing least squares estimate $\hat{\boldsymbol{\beta}}_{LS}(k)$ and $\boldsymbol{\Phi}(k)$.

Step 2. Computing equality-constrained least squares estimate

$$\bar{\boldsymbol{\beta}}(k) = \hat{\boldsymbol{\beta}}_{LS}(k) - \boldsymbol{\Phi}(k)\tilde{\mathbf{R}}_0^{\mathrm{T}}[\tilde{\mathbf{R}}_0\boldsymbol{\Phi}(k)\tilde{\mathbf{R}}_0^{\mathrm{T}}]^{-1}[\tilde{\mathbf{R}}_0\hat{\boldsymbol{\beta}}_{LS}(k) - \mathbf{1}].$$

Step 3. For $\mathbf{H}(k) = \boldsymbol{\Phi}(k) - \boldsymbol{\Phi}(k)\tilde{\mathbf{R}}_0^{\mathrm{T}}[\tilde{\mathbf{R}}_0\boldsymbol{\Phi}(k)\tilde{\mathbf{R}}_0^{\mathrm{T}}]^{-1}\tilde{\mathbf{R}}_0\boldsymbol{\Phi}(k)$, and $\mathbf{D}(k) = \mathrm{diag}\{1/h_{ii}(k)\}$, computing the Bell correction for inequality-constraints

$$\tilde{\boldsymbol{\beta}}(k) = \bar{\boldsymbol{\beta}}(k) + \mathbf{H}(k)\mathbf{T}(-\mathbf{D}(k)\bar{\boldsymbol{\beta}}(k)).$$

Step 4. Enforcing non-negative-constraints (1b′) if necessary: $\tilde{\boldsymbol{\beta}}(k) = \mathbf{T}(\tilde{\boldsymbol{\beta}}(k))$.
The estimate $\hat{\beta}_{ij}(k)$ is then given by $\tilde{\beta}_{ij}(k)/\sum_{j=1}^{|J_i|}\tilde{\beta}_{ij}(k)$.

Appendix C. Proofs of theorems

Lemma 4. *For* $\mathbf{S} = \mathbf{I} - \mathbf{s}\mathbf{1}^{\mathrm{T}}$, *where* $\mathbf{s}$ *is a* $t \times 1$ *vector and* $\mathbf{1}^{\mathrm{T}}\boldsymbol{s} = 1$, *solutions of the linear equations* $\mathbf{S}\mathbf{z} = \mathbf{0}$ *are given by* $\mathbf{z}^* = c\mathbf{s}$ *for* $\forall c \in R^1$.

Proof. Any of solutions of $\mathbf{S}\mathbf{z} = \mathbf{0}$ can be written as $\mathbf{z} = (\mathbf{I} - \mathbf{S}^{-}\mathbf{S})\mathbf{r},\ \forall \mathbf{r} \in R^{t'\times 1}$, where $\mathbf{S}^{-}$ is one of the generalized inverses of $\mathbf{S}$. Since $\mathbf{S}^2 = \mathbf{S}$, it follows by noting that $\mathbf{S}$ itself is a generalized inverse of $\mathbf{S}$ and $\mathbf{z} = (\mathbf{I} - \mathbf{S}^{-}\mathbf{S})\mathbf{r} = (\mathbf{1}^{\mathrm{T}}\mathbf{r})\mathbf{s} = c\mathbf{s}\ \forall c \in R^1$.

Proof of Theorem 1. Note that for completely non-informative situations, constraints (11′) are reduced to (11b′), and the cost function (10′) degenerates to min $F(\mathbf{f}) = (\mathbf{f} - \mathbf{f}_0)^{\mathrm{T}}\mathbf{Z}^{\mathrm{T}}\mathbf{W}\mathbf{Z}(\mathbf{f} - \mathbf{f}_0)$. This leads to optimal solutions $\hat{\mathbf{f}}_i = c_i\mathbf{a}_i$ for any $c_i > 0$ from Lemma 4 and (11b′). Then we obtain $\hat{\mathbf{b}}_i = \mathbf{a}_i, \forall i$.

Proof of Theorem 2. The proof of Theorem 2 will be completed by showing that the solution of (12) and (16) is identical to the solution (B.1) at the kth time interval. We suppress the dependence on the time interval k as before. For situations with complete information, we have

$$q_i = q_i^*\ (i = 1, \ldots, m) \quad y_j = y_j^*\ (j = 1, \ldots, n) \quad \mathbf{R} = \mathbf{R}_0 \quad \mathbf{Z} = \mathbf{I} \quad \mathbf{U} = \mathbf{W} + \mathbf{G}^{\mathrm{T}}\mathbf{G}.$$

From (1) and (5)

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\Gamma}\hat{\mathbf{b}} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}^{-1}\hat{\mathbf{f}}. \tag{C.1}$$

For $\boldsymbol{\alpha}$ and $\mathbf{a}$ which are counterparts of $\boldsymbol{\beta}$ and $\mathbf{b}$ at the $(k-1)$th time interval, we have

$$\hat{\boldsymbol{\alpha}} \hat{=} \boldsymbol{\Gamma}\mathbf{a} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}^{-1}\hat{\mathbf{f}}_0. \tag{C.1′}$$

Next, from (7), (A.1), and $\mathbf{Z} = \mathbf{I}$, it is easy to check that

$$\mathbf{U}^{-1} = \boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\boldsymbol{\Gamma}\boldsymbol{\Lambda} \tag{C.2}$$

and from (1) and (2)

$$[\mathbf{R}_0\mathbf{U}^{-1}\mathbf{R}_0^{\mathrm{T}}]^{-1} = \mathbf{P}^{-1}[\tilde{\mathbf{R}}_0(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\tilde{\mathbf{R}}_0^{\mathrm{T}}]^{-1}\mathbf{P}^{-1} \quad \text{and} \quad \mathbf{R}_0\hat{\mathbf{f}}_{LS} - \mathbf{q} = \mathbf{P}(\tilde{\mathbf{R}}_0\hat{\boldsymbol{\beta}}_{LS} - \mathbf{1}). \tag{C.3}$$

Therefore, inserting (16a) into (C.1), we obtain

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}^{-1}\hat{\mathbf{f}} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}^{-1}\hat{\mathbf{f}}_{LS} - \boldsymbol{\Gamma}\boldsymbol{\Lambda}^{-1}\mathbf{U}^{-1}\mathbf{R}_0^{\mathrm{T}}[\mathbf{R}_0\mathbf{U}^{-1}\mathbf{R}_0^{\mathrm{T}}]^{-1}[\mathbf{R}_0\hat{\mathbf{f}}_{LS} - \mathbf{q}] + \boldsymbol{\Gamma}\boldsymbol{\Lambda}^{-1}\mathbf{H}\boldsymbol{\mu}. \tag{C.4}$$

From A.1, B.1, C.1 and C.2, the first term of the right-hand side of (C.4) equals to

$$\hat{\boldsymbol{\alpha}} + (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{Q}^{\mathrm{T}}[\mathbf{y} - \mathbf{Q}\hat{\boldsymbol{\alpha}}_{LS}] = \hat{\boldsymbol{\beta}}_{LS}.$$

From (C.3) and (2), the second term of the right-hand side of (C.4) equals to

$$(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\tilde{\mathbf{R}}_0^{\mathrm{T}}[\tilde{\mathbf{R}}_0(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\tilde{\mathbf{R}}_0^{\mathrm{T}}]^{-1}[\tilde{\mathbf{R}}_0\hat{\boldsymbol{\beta}}_{LS} - \mathbf{1}].$$

From (C.3), and letting $\tilde{\boldsymbol{\mu}} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\mu}$, the third term of the right-hand side of (C.4) equals to

$$\{(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} - (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\tilde{\mathbf{R}}_0^{\mathrm{T}}[\tilde{\mathbf{R}}_0(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\tilde{\mathbf{R}}_0^{\mathrm{T}}]^{-1}\tilde{\mathbf{R}}_0(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\}\tilde{\boldsymbol{\mu}}.$$

Thus, it follows by inserting the above equations into (C.4) and comparing it with (B.1).

Lemma 5. *If* $\mathbf{A} > 0$ *and* $\mathbf{R}$ *has full row rank, then*

$$\tilde{\mathbf{A}} \hat{=} \begin{bmatrix} \mathbf{A} & \mathbf{R}^{\mathrm{T}} \\ \mathbf{R} & \mathbf{0} \end{bmatrix}$$

is non-singular.

Proof. $\tilde{\mathbf{A}}[\mathbf{z}_1^{\mathrm{T}}, \mathbf{z}_2^{\mathrm{T}}]^{\mathrm{T}} = \mathbf{0} \Longleftrightarrow \mathbf{z}_1 = -\mathbf{A}^{-1}\mathbf{R}^{\mathrm{T}}\mathbf{z}_2$ and $\mathbf{R}\mathbf{z}_1 = \mathbf{0} \Longleftrightarrow \mathbf{z}_1 = -\mathbf{A}^{-1}\mathbf{R}^{\mathrm{T}}\mathbf{z}_2$ and $(\mathbf{A}^{-1/2}\mathbf{R}^{\mathrm{T}})^{\mathrm{T}}(\mathbf{A}^{-1/2}\mathbf{R}^{\mathrm{T}})\mathbf{z}_2 = \mathbf{0} \Longleftrightarrow \mathbf{z}_1 = -\mathbf{A}^{-1}\mathbf{R}^{\mathrm{T}}\mathbf{z}_2$ and $\mathbf{z}_2 = \mathbf{0}$ (note $\mathbf{R}$ has full row rank) $\Longleftrightarrow \mathbf{z}_1 = \mathbf{0}$ and $\mathbf{z}_2 = \mathbf{0} \Longleftrightarrow \tilde{\mathbf{A}}$ is non-singular.

Proof of Theorem 3. (i) Let $\tilde{\mathbf{W}} = \mathrm{diag}\{\mathbf{W}, \mathbf{I}\}$. The sufficiency is immediate from $\boldsymbol{\Psi}^{\mathrm{T}}\tilde{\mathbf{W}}\boldsymbol{\Psi} > 0$ and Lemma 5.

Next, consider the necessity. Let $\mathbf{S} = [\mathbf{R}, \mathbf{0}]^{\mathrm{T}}$. Noting $\mathbf{R}^{\mathrm{T}} = [\mathbf{Z}^{\mathrm{T}}, \mathbf{G}^{\mathrm{T}}]\mathbf{S} = \boldsymbol{\Psi}^{\mathrm{T}}\mathbf{S}$, we have

$$\begin{bmatrix} \mathbf{Z}^{\mathrm{T}}\mathbf{W}\mathbf{Z} + \mathbf{G}^{\mathrm{T}}\mathbf{G} & \mathbf{R}^{\mathrm{T}} \\ \mathbf{R} & \mathbf{0} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Psi}^{\mathrm{T}} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{W}} & \mathbf{S} \\ \mathbf{S}^{\mathrm{T}} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Psi} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}. \tag{C.5}$$

If $\boldsymbol{\Psi}$ is column rank deficient, we conclude that

$$\begin{bmatrix} \mathbf{Z}^{\mathrm{T}}\mathbf{W}\mathbf{Z} + \mathbf{G}^{\mathrm{T}}\mathbf{G} & \mathbf{R}^{\mathrm{T}} \\ \mathbf{R} & \mathbf{0} \end{bmatrix}$$

is singular from Eq. (C.5), which then leads to contradiction. This completes the proof of (i).

(ii) Let $\mathbf{z} = [\mathbf{z}_1^{\mathrm{T}}, \ldots, \mathbf{z}_m^{\mathrm{T}}]^{\mathrm{T}}$. Then $\boldsymbol{\Psi}\mathbf{z} = \mathbf{0}$ if and only if

$$\mathbf{Z}_i\mathbf{z}_i = 0 \quad (\text{for } i = 1, \ldots, r) \quad \mathbf{z}_i = 0 \quad (\text{for } i = r+1, \ldots, m), \text{ and } \mathbf{G}\mathbf{z} = \mathbf{0}.$$

From Lemma 4 and $\mathbf{Z}_i\mathbf{z}_i = \mathbf{0}$ (for $i = 1, \ldots, r$), we obtain $\mathbf{z}_i = c_i\mathbf{a}_i, \forall c_i \in R^1 (i = 1, \ldots, r)$. Inserting them into $\mathbf{G}\mathbf{z} = \mathbf{0}$ and noting $\mathbf{z}_i = \mathbf{0}$ (for $i = r+1, \ldots, m$), we have $\sum_{i=1}^{r} c_i\mathbf{G}^{(i)}\mathbf{a}_i = \mathbf{0}$. Hence, $\boldsymbol{\Psi}$ has full column rank $\Longleftrightarrow \boldsymbol{\Psi}\mathbf{z} = \mathbf{0}$ has only zero solution $\mathbf{z}^* = \mathbf{0} \Longleftrightarrow \sum_{i=1}^{r} c_i\mathbf{G}^{(i)}\mathbf{a}_i = \mathbf{0}$ have only zero solution $c_i^* = 0 (i = 1, \ldots, r) \Longleftrightarrow$ rank $(\boldsymbol{\Pi}) = r$. This completes the proof of (ii).

The proof of (iii) is immediate from (ii) and noting that $\mathbf{R}$ has full row rank.

References

Bell, M.G.H., 1991a. The estimation of origin–destination matrices by constrained generalized least squares. Transp. Res. B 25, 13–22.

Bell, M.G.H., 1991b. The real-time estimation of origin–destination flows in the presence of platoon dispersion. Transp. Res. B 25, 115–125.

Bierlaire, M., Toint, Ph.L., 1995. MEUSE: an origin–destination matrix estimator that exploits structure. Transp. Res. B 29, 47–60.

Cascetta, E., 1984. Estimation of trip matrices from traffic counts and survey data: a generalized least squares approach estimator. Transp. Res. B 18, 289–299.

Cremer, M., Keller, H., 1987. A new class of dynamic methods for the identification of origin–destination flows. Transp. Res. B 21, 117–132.

Demster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. 39, 1–38.

Frenkel, L., Feder, M., 1999. Recursive expectation-maximization (EM) algorithm for time-varying parameters with applications to multiple target tracking. IEEE Trans. Signal Process. 47, 306–320.

Holm, J., Jensen, T., Nielsen, S.K., 1976. Calibrating traffic models on traffic census results only. Traffic Eng. Control 17, 137–140.

Li, B., De Moor, B., 1999. Recursive estimation based on the equality-constrained optimization for intersection origin–destination matrices. Transp. Res. B 33, 203–214.

Little, R.J.A., 1987. Statistical Analysis with Missing Data. Wiley, New York.

Mather, M.J., 1983. Inferences on trip matrices from observations on link volumes: a Bayesian statistical approach. Transp. Res. B 17, 435–447.

McLachlan, G.J., 1996. The EM Algorithm and Extensions. Wiley, New York.

Myers, R.H., 1986. Classical and Modern Regression with Applications. Duxbury Press, Boston.

Nihan, N., Davis, G., 1987. Recursive estimation of origin–destination matrices from input/output counts. Transp. Res. B 21, 149–163.

Nihan, N., Davis, G., 1989. Application of prediction-error minimization and maximum likelihood to estimate intersection O–D matrices from traffic counts. Transp. Sci. 23, 77–90.

Sherali, H.D., Arora, N., Hobeika, A.G., 1997. Parameter optimization methods for estimating dynamic origin–destination trip-tables. Transp. Res. B 31, 141–157.

Van Zuzlen, H.J., Willumsen, L.G., 1980. The most likely trip matrix estimated from traffic counts. Transp. Res. B 14, 281–293.

Wilson, A.G., 1967. A statistical theory of spatial distribution models. Transp. Res. 1, 253–269.