



# 交通需要予測モデルにおける定数項の修正：データ 収集時点とサンプル数に基づくその意義

三古，展弘

---

(Citation)

土木学会論文集D3(土木計画学), 74(1):21-34

(Issue Date)

2018

(Resource Type)

journal article

(Version)

Version of Record

(Rights)

© 2018 公益社団法人 土木学会

(URL)

<https://hdl.handle.net/20.500.14094/90007008>



# 交通需要予測モデルにおける定数項の修正： データ収集時点とサンプル数に基づくその意義

三古 展弘<sup>1</sup>

<sup>1</sup>正会員 神戸大学大学院教授 経営学研究科 (〒657-8501 神戸市灘区六甲台町2-1)

E-mail: sanko@kobe-u.ac.jp

古い時点の多数のデータと新しい時点の少数のデータが利用可能なとき、2時点のデータを同時に用いる方法として定数項の修正がある。しかし、新しい時点のデータのみを用いたほうが定数項の修正よりも予測精度が優れているという報告もある。本研究は、定数項の修正を用いる意義を、データ収集時点とサンプル数に着目して分析する。その結果、いかなるデータ収集時点とサンプル数の組み合わせでも定数項の修正のほうが有意に良い予測をもたらすことはなかった。しかし、新しい時点のサンプル数が極めて小さいとき、統計的に有意ではないものの定数項の修正のほうが良い予測をもたらす、その予測値のばらつきが小さく、また推定と予測の計算に問題が少ないことが明らかになった。

**Key Words :** travel demand forecast, transfer scaling, temporal transferability, data collection time points, number of observations, bootstrap

## 1. はじめに

交通需要予測においては新しい時点のデータを用いることが重要である。新しい時点のデータを用いたモデルと古い時点のデータを用いたモデルを比較し、前者のほうがより良い予測をもたらすことが報告されている<sup>1)</sup>。しかし、これらは十分なサンプル数のデータが得られている複数の時点でモデルを構築した場合に得られる知見である。一方で、サンプル数が多ければ、モデルがより精度良く推定され<sup>2)</sup>、その結果として予測精度も高くなることが期待される。確かに、データサイズが同じであればデータは新しいほうが予測精度が高く、データ収集時点が同じであればデータサイズが大きいほうが予測精度が高い。ところが、実際には、古い時点のサンプル数は大きい、新しい時点のサンプル数は小さい場合がある。

サンプル数の大きい古いデータとサンプル数の小さい新しいデータがあった場合、分析におけるデータの利用方法は次の3つが考えられる。

- [1] サンプル数の大きい古いデータとサンプル数の小さい新しいデータを両方使用
  - [2] サンプル数の小さい新しいデータのみを使用
  - [3] サンプル数の大きい古いデータのみを使用
- もし、1時点のデータしか用いないのであれば、[2]

と[3]の比較が必要であり、既に筆者<sup>3)</sup>が報告している。主な知見としては、新しい時点のサンプル数が小さくても、古い時点の多数のサンプルのモデルよりも統計的に有意に良い予測精度を持つ場合があることが確認できた。一方、古い時点の多数のサンプルのモデルが新しい時点の少数のサンプルのモデルよりも統計的に有意に良い予測精度を持つ場合はなかった。しかし、新しい時点のサンプル数が小さければ、古い時点の多数のサンプルによるモデルよりも統計的に有意に良い予測をもたらさない場合がある。また、古い時点の多数のサンプルによるモデルのほうが統計的に有意ではないものの新しい時点の少数のサンプルによるモデルよりも良い予測を行える場合がある。つまり、古い時点のデータを利用する意義が全くないわけではない。

ここで、2つの時点のデータを用いた場合（上の[1]）を考える。これはモデルの更新として知られ、古い時点の多数のサンプルで構築したモデルを、新しい時点の少数のサンプルで更新する。つまり、古い時点のサンプルで構築したモデルを最小限の労力で収集した新しいデータを利用することによって少なくとも古い時点のデータのみを用いたモデルよりも改善しようという試みである。これまでに、定数項の修正(transfer scaling)、joint context estimation、Bayesian updating、combined transfer estimationというモデル更新法が提案されてきた。実際、モデルの更新

は上の[3]よりも[1]のほうが良くなることを目的としている。

本研究で対象とするのは上の[1]と[2]の比較である。

(上の[1]と[3]の比較については本論文では取り扱わない。筆者<sup>7)</sup>は[3]のほうが[2]よりも予測精度において統計的に優れている場合はないことを示したので、[1]と[2]の比較のほうが[1]と[3]の比較より意義があると考えられる。) モデルの更新は2つの時点のデータを両方用いているが、新しい時点の1つのデータを用いるだけでもモデルの更新よりもよい予測精度をもたらす場合があると報告されている。Badoe and Miller<sup>8)</sup>とKarasmaa and Pursula<sup>9)</sup>は新しい時点のサンプル数が少ない場合でも(例えば400~500サンプル)、新しい時点に加えて古い時点のデータを利用することで予測精度の向上にはほとんど貢献しないと主張している。

今回対象とするモデルの更新法は定数項の修正である。これは、Atherton and Ben-Akiva<sup>10)</sup>によって提案され、古い時点の多数のサンプルによって構築されたモデルの選択肢固有定数項とスケールパラメータのみを新しい時点の少数のサンプルによって再推定し、それ以外の説明変数のパラメータは完全に移転可能とするものである。定数項の修正という方法は、非常に簡便にモデルを更新できる方法で理解もしやすいので、実務的にも利用価値が高いと考えられる。なお、本研究では選択肢固有定数項とスケールパラメータの両方を修正するが、以降では誤解のない場合には単に定数項の修正と記す。

本研究は実務的には次の問いに答えることができると考える。

1. 古い時点の多数のデータと新しい時点の少数のデータがあるとき、新しい時点のデータのみでモデルを構築するのが良いのか、古い時点のデータを用いたモデルの定数項とスケールパラメータを新しい時点のデータを用いて更新するのが良いのか。
2. 既存のデータがあるが、その後の交通状況の変化を反映するために新しいデータで分析したい。しかし、時間や費用などの制約から小規模な調査しか行えない。このとき、どのくらいのサンプル数のときに新しい時点のデータのみでモデルを構築するのが良いのか、古い時点のデータを用いたモデルの定数項とスケールパラメータを更新するのが良いのか。

上の2の状況は、パーソントリップ調査のような大規模だが低頻度(3大都市圏では10年に1回)で実施される調査において多く発生すると考えられる。このようなデータを用いて需要予測を行う場合、最新のデータでも10年以上前に収集されたものであることがある。このとき、既存の古いデータを使うか、新しいデータを独自に収集するかという判断が必要になる。モデルの更新の目的か

ら考えると、定数項の修正は新しい時点単独のデータではサンプル数が少ないために十分に良い予測を行えないことを前提としている。もし、新しい時点から十分なサンプル数が得られているのであれば、新しい時点のデータのみを用いるべきである。つまり、定数項の修正は、新しい時点単独で定数項の修正よりも良い予測を行えない場合に検討するべきであり、そのような状況が発生するデータ収集時点とサンプル数を把握する必要がある。

本研究のリサーチクエスションは以下に整理される。

リサーチクエスション：古い時点の多数のサンプルと新しい時点の少数のサンプルが得られたときに、2時点のデータを用いて定数項とスケールパラメータを修正したモデルと新しい時点のデータのみを用いたモデルを構築し、将来予測を行う。このとき、いかなるデータ収集年とサンプル数のときに、どちらが良い予測を行うことができるか、また、それは統計的に有意であるか。

筆者は、この問題を中京都市圏で1971, 1981, 1991, 2001年に得られたパーソントリップ調査データを用いて出勤交通手段選択行動を対象に分析する。データは4時点から得られているが、最初の3時点のデータをモデルの推定に用い、最後の1時点のデータは予測精度の検証のみに用いる。古い時点と新しい時点の組み合わせとして1971/1981, 1971/1991, 1981/1991の3通りを考える。また、古い時点と新しい時点から得られるサンプル数については100~10000の範囲の12通りを考える。統計的に意味のある知見を得るために、ブートストラップ法を用いる。本研究で用いられるデータは、同一の公的機関によって統一されたフォーマットで調査されたものである。従って、データの質に調査時点間で大きな違いはないと考えられる。よって、今回のデータ収集時点とサンプル数を分析するのに適切であると考えられる。なお、本研究のデータは1971年から2001年に得られたものであり、最も新しい2001年のデータでも15年以上前に得られたものである。しかし、これは主たる関心の対象ではない。

本研究は、2時点のデータを用いて定数項を修正したモデルと直近の1時点のデータのみを用いるモデルの予測精度の優劣を統計的に検討する初めての試みである<sup>[註1]</sup>。そのため、可能な限り分析を簡略化し、また、筆者の利用可能なデータを用いることにする。データ収集時点とサンプル数が本研究で関心の2つの次元である。移転性に影響を与えると考えられる他の次元には、(i) 交通行動の理論(例えば、意思決定方略は補償型か非補償型か、トリップベースかツアーベースか)、(ii) モデルの数学的構造(例えば、ロジットモデルかネスティ

ッドロジットモデルか), (iii) モデルの特定化 (例えば, 説明変数の選択, 説明変数の影響は線形か非線形か, 異質性の考慮)<sup>[12]</sup>がある. 本研究では, 線形の効用関数を仮定した多項ロジットモデルを用いて効用最大化のフレームワークで分析する. また, 論文全体を通して統一したモデルの特定化を採用する. モデルの特定化は移転性に影響を与えることが報告されており, 多くの研究は説明変数を増やすことで移転性が向上することを示している<sup>[9, 11, 12]</sup>. 一方で, *overfitting*の危険性を孕んでいる<sup>[13]</sup>. 本研究では論文全体で共通の1つのモデル特定化を採用しているので, モデルの特定化が予測精度に与える影響は考慮しない. 採用した特定化は, 筆者が試みた中で最も良いと判断したものであるが, より高い時間移転性を持つ特定化が存在する可能性は否定できない. しかし, 仮にそのような特定化が存在しても, そのモデルが今回の比較の対象とする2つのモデル (2時点のデータを用いた定数項の修正によるモデルと新しい時点のデータのみを用いたモデル) の両方の予測精度を同程度に向上させるのであれば, 特定化の効果は相殺されと考えられる. もし, 2つのモデルの予測精度を異なる程度に向上させるのであれば, 特定化が結果に影響を与える可能性がある. この点は今回の分析の対象外であり, 今後の課題である.

本論文は以下のように構成される. 第2章では既存研究のレビューを行う. 第3章では使用データを説明する. 第4章では方法論として, まず多項ロジットモデルと定数項の修正について説明した後, ブートストラップ法, ブートストラップ法を用いた統計的検定について説明する. 第5章では推定値, 予測精度の特徴, 検定の結果について説明する. 最後に第6章で結論を述べる.

## 2. 既存研究のレビュー

定数項の修正は, Atherton and Ben-Akiva<sup>[10]</sup>から引用すると, 以下の動機によって提案された.

ほとんどのモデルの特定化は, モデルによって明示的に説明されない要因を説明するために定数項を含んでいる. このような定数項の存在は, モデルが選択過程の総ての側面を捉えていないことを意味し, その捉えられていない要因は地域 [時点] 間で異なる可能性があるため, ある地域 [時点] の定数項の推定値が別の地域 [時点] で適当かは不明である. 従って, 時間, 費用, 所得, 自動車利用可能性などの間の推定された関係が移転可能であるとする理論的根拠はあるが, 定数項が移転可能であるという根拠はない. ([ ] 内筆者.)

定数項の修正は, 非常に簡便にモデルを更新できる方

法で理解もしやすいので, 実務的にも利用価値が高いと考えられる. ここで, これに関連した既存研究をいくつか紹介する. 第1章で述べたように, モデルの修正は古い時点の多数のデータのみを用いるよりも, 古い時点の多数のデータに加えて新しい時点の少数のデータを用いるほうが予測精度が高くなることを期待している. 古い時点のデータのみを用いたモデルと, 古い時点のデータに加えて新しい時点のデータを用いて定数項の修正をしたモデルの比較については例えば, Sanko and Morikawa<sup>[14]</sup>に譲り, ここでは, 本研究の主題である, 古い時点のデータと新しい時点のデータを用いて定数項を修正したモデルと, 新しい時点のデータのみを用いたモデルを比較した研究を紹介する.

Badoe and Miller<sup>[8]</sup>はカナダのトロントで1964年から得られた多数のサンプルと1986年から得られた少数のサンプルを用い, 1964年のサンプル数は変化させずに1986年のサンプル数のみを変化させたときの1986年の予測精度を検討している (注: データが2時点からしか得られていないので, 1986年のデータはモデルの推定にも予測精度の検証にも用いられている). 分析の対象は朝ピーク時の出勤交通手段選択行動である. 彼らは以下の2つのモデルを推定している. (i) 1964年の多数のサンプルと1986年の少数のサンプルの両方を用いてモデル更新法を使ったモデル, (ii) 1986年の少数のサンプルのみを用いたモデル. そして, (i)と(ii)のモデルを用いて1986年の行動を予測し比較した. (i)のモデルとしては, 定数項の修正を含む4種類のモデル更新法を検討している. モデルの特定化について, サービスレベル変数のみを用いた特定化とサービスレベル変数に加えて社会経済属性変数を用いた特定化を行っている. 予測精度に優れている後者の特定化において, 1986年のサンプル数が400の場合でも, (ii)のモデルのほうが(i)の定数項の修正よりも予測精度において優れていることを示している.

同様の分析をKarasmaa and Pursula<sup>[9]</sup>はフィンランドのヘルシンキの1981年と1988年のデータを用いて出勤時の交通手段・目的地選択モデルを対象に行い, 類似の結論を得ている.

まとめると, 新しい時点の少数のサンプルに加えて, 古い時点の多数のサンプルを用いて定数項を修正する意義は, 予測精度の向上という観点からは極めて小さいのではないかと考えられる.

しかし, これらの研究ではデータが2時点からしか得られておらず, 2時点目のデータをモデルの推定と評価の両方に用いているという問題がある. また, 新しい時点の様々なサンプル数を検討しているが, 新しい時点からのサンプルの抽出を1回しか行っておらず, 結論に統計的な視点を含んでいない.

### 3. データ

中京都市圏において1971, 1981, 1991, 2001年の4時点で得られた繰り返し断面データである, パーソントリップ調査データを用いる. モデルの構築には1971, 1981, 1991年のデータを用い, 2001年のデータはモデルの予測精度の検証にのみ用いる. 本研究で分析の対象とするのは鉄道, バス, 自動車の3選択肢からの通勤交通手段選択行動である.

分析対象範囲は, 4時点総ての調査圏域に含まれる1971年調査の調査圏域である. 1971年の調査圏域全体から抽出される通勤交通手段選択行動を一括して分析しており, 距離帯などによるセグメンテーションは行っていない. ゾーン区分は1991年調査の基本ゾーン(行政区画を考慮して概ね人口2万人を目処に分割)を採用した. サービスレベル変数(所要時間)は基本ゾーンのOD(出発地・到着地)ペア毎に回答者の報告値を平均した(Sanko and Morikawa<sup>14)</sup>でも同様の方法を採用). また, 交通手段の利用可能性については, ODペアで回答者のうち1人も利用のない交通手段は利用不可能として選択肢集合から除いている. 都市圏の全域を一括した分析, サービスレベルの算出および利用可能性の考慮は必ずしも実務で用いられている方法と合致しないとも考えられるが, 4時点で統一した方法を用いることで調査時点とサンプル数以外をコントロールしており, 本研究の意義を損なわないと考える.

その他のデータの詳細および記述統計についてはSanko<sup>9)</sup>を参照されたいが, 2点について再度触れておく. 1つ目に, 通勤の費用については通勤手当が支給されることが多いため考慮しない<sup>13)</sup>. 2つ目に, 各交通手段のシェアが1971~1981年にかけて大きく変化し, その後は緩やかに変化したことである. 推定のためにデータを整理した後のサンプルを見ると1971, 1981, 1991, 2001年の交通手段のシェアは, 鉄道: 28%, 28%, 26%, 25%, バス: 21%, 9%, 5%, 3%, 自動車: 51%, 63%, 68%, 72%となっている.

### 4. 方法論

本研究は, 古い時点の多数のデータと新しい時点の少数のデータが利用可能なとき, 2時点のデータを用いて定数項の修正というモデル更新法を利用する場合と, 新しい時点の少数のデータのみを用いる場合の予測精度を統計的に検定する初めての試みである. そこで, データ収集時点とサンプル数以外についてはできる限り簡略化した設定にしている. 本章では, まずモデルについて説明し, ブートストラップ, 仮説検定について説明する.

#### (1) モデル

2つの時点として $t1$ (古い時点)と $t2$ (新しい時点)を考える. [1] $t1$ の時点で構築したモデルの定数項とスケールパラメータを $t2$ の時点のデータを用いて更新した場合, [2] $t2$ の時点のデータのみを用いてモデルを構築した場合を比較する. まず, ロジットモデルについて説明し, その後で[1]と[2]の具体的手順について説明する.

ランダム効用理論に基づき, 全効用を確定項と誤差項に分けて表現する. 個人 $p$ の選択肢 $i$ に対する時点 $t$ (ここでは $t1$ と $t2$ を区別しないで定式化する)における効用関数の確定項 $V_{ip}^t$ を式(1)のように定式化する.

$$V_{ip}^t = \mu^t \left( \alpha_i^t + \sum_k \beta_{ik}^t x_{ikp}^t \right) \quad (1)$$

ここに,  $\mu^t$ は時点 $t$ のスケールパラメータ,  $\alpha_i^t$ は時点 $t$ の選択肢 $i$ の選択肢固有定数項,  $x_{ikp}^t$ は時点 $t$ の個人 $p$ の選択肢 $i$ に対する $k$ 番目説明変数,  $\beta_{ik}^t$ はそれに対応するパラメータである.

誤差項に独立で同一なばらつきを持つガンベル分布を仮定すると, 時点 $t$ において個人 $p$ が選択肢 $i$ を選択する確率 $P_{ip}^t$ は式(2)のロジット式で表現される.

$$P_{ip}^t = \frac{\exp(V_{ip}^t)}{\sum_j \exp(V_{jp}^t)} \quad (2)$$

このとき, 対数尤度関数は式(3)で表現され, これを最大化することによってパラメータを推定する.

$$L^t = \sum_p \sum_j y_{jp}^t \ln(P_{jp}^t) \quad (3)$$

ここに,  $y_{jp}^t$ は時点 $t$ で個人 $p$ の選択結果が選択肢 $j$ であったとき1, そうではないとき0となるダミー変数.

#### a) 2時点のデータを用いた定数項の修正

先の[1]の場合である, 2時点のデータを用いた定数項の修正を説明する.

まず, 式(1)~(3)に $t=t1$ を代入して, モデルを推定する. なお, スケールパラメータと変数のパラメータを識別することはできないので, スケールパラメータを $\mu^{t1}=1$ に固定する.

次に,  $t2$ のデータを用いて時点 $t2$ において個人 $p$ が選択肢 $i$ を選択する確率 $P_{ip}^{t2}$ は式(4)で表現される.

$$P_{ip}^{t^2} = \frac{\exp\left(\mu^{t^2}\left(\alpha_i^{t^2} + \sum_k (\hat{\beta}_{ik}^{t^1} x_{ikp}^{t^2})\right)\right)}{\sum_j \exp\left(\mu^{t^2}\left(\alpha_j^{t^2} + \sum_k (\hat{\beta}_{jk}^{t^1} x_{jkp}^{t^2})\right)\right)} \quad (4)$$

ここに、 $\wedge$ は推定値を意味する。ここで、式(3)において  $t=2$  にすることで  $\mu^{t^2}$  と  $\alpha^{t^2}$  を推定する。

モデルの予測精度は推定されたパラメータ  $\hat{\beta}^{t^1}$ ,  $\hat{\alpha}^{t^2}$ ,  $\hat{\mu}^{t^2}$ , 2001年の説明変数  $\mathbf{x}$  と選択結果  $\mathbf{y}$  を式(1)~(3)に代入したときの式(3)の値  $L^{2001}$  で表現される。

#### b) 新しい時点の少数のデータのみを用いたモデル

先の[2]の場合である、新しい時点の少数のデータのみを用いたモデルを説明する。

式(1)~(3)に  $t=2$  を代入して、モデルを推定する。なお、スケールパラメータはa)項で述べたのと同様の理由で  $\mu^{t^2}=1$  に固定する。モデルの予測精度は推定されたパラメータ  $\hat{\beta}^{t^2}$ ,  $\hat{\alpha}^{t^2}$ ,  $\mu^{t^2}=1$ , 2001年の説明変数  $\mathbf{x}$  と選択結果  $\mathbf{y}$  を式(1)~(3)に代入したときの式(3)の値  $L^{2001}$  で表現される。

## (2) ブートストラップ<sup>15)</sup>

まず、1971, 1981, 1991年のデータから通勤トリップをランダムに10000サンプルずつ抽出した。以後の分析では100~10000サンプルの範囲でサンプル数を変化させるが、それらは、この10000サンプルからランダムに抽出される。各時点から同じ10000サンプルをまず抽出するのは、以後で100~10000サンプルを抽出する元のサンプル数が異なることによる結果への影響を避けるためである。また、予測対象年の2001年からもランダムに10000サンプルを抽出して検証に用いる。

ここで、5つの変数 ( $y_1$ ,  $y_2$ ,  $m_1$ ,  $m_2$ ,  $b$ ) を定義する。 $y_1$  と  $y_2$  ( $y_1 < y_2$ ) はデータ収集年、 $m_1$  と  $m_2$  ( $m_1 \geq m_2$ ) はそれぞれ  $y_1$ ,  $y_2$  において得られたデータのサンプル数、 $b$  はブートストラップの繰り返し回数である。ここで、上の2つの不等式は、古い時点  $y_1$  のサンプル数  $m_1$  で構築したモデルを新しい時点  $y_2$  の同じまたは小さいサンプル数  $m_2$  で更新することを意味している。

まず、分析の準備として、それぞれのデータ年  $y$  (1971, 1981, 1991) においてサンプル数  $n$  (100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 10000 の12通り) のデータを先に抽出した10000サンプルから1000 ( $b=1, 2, \dots, 1000$ ) 回ランダムに復元抽出する。このとき、同一のデータ年  $y$  の  $b$  回目の抽出において、 $n$  が小さいサンプルは  $n$  が大きいサンプルの一部分になるように抽出している。これによって  $y$ ,  $n$  および  $b$  の組み合わせからなる  $3 \times 12 \times 1000 = 36000$  通りのデータが生成された。

$y_1$ ,  $y_2$ ,  $m_1$ ,  $m_2$  について考えられる組み合わせは、 $y_1$  に

関する3通りの組み合わせ ( $(y_1, y_2) = (1971, 1981), (1971, 1991), \text{ and } (1981, 1991)$ ) と  $n$  に関する78通りの組み合わせ ( $12 \times 13/2$ ) の  $3 \times 78 = 234$  通り考えられる。この234通りのそれぞれについて1000 ( $b=1, 2, \dots, 1000$ ) 回、[1]  $y_1$  時点の  $m_1$  サンプル ( $b$  回目抽出) で推定した多項ロジットモデルを  $y_2$  時点の  $m_2$  サンプル ( $b$  回目抽出) で更新、[2]  $y_2$  時点の  $m_2$  サンプル ( $b$  回目抽出) で多項ロジットモデルを推定、した。[1]の計算をするのに  $y_1$  の時点のデータを用いて  $234 \times 1000 = 234000$  回、 $y_2$  の時点のデータを用いて更新するのに同じく234000回の合計468000回の推定を行う必要がある。[2]の計算をするのに  $y_2$  の時点のデータを用いて234000回の推定を行う必要がある。つまり、延べ702000回の推定が必要になる。しかし、1971年のデータを用いてモデルを推定すればそれは1981年のデータで更新する場合にも1991年のデータで更新する場合にも使えるなど、重複する場合もあるので、実際に作業をするときには702000回推定を行う必要はない。

予測精度は2001年のデータへの対数尤度で表現する。上の[1]と[2]の方法による2001年のデータへの対数尤度をそれぞれ  $L1(y_1, y_2, m_1, m_2, b)$  と  $L2(y_2, m_2, b)$  と表現する。

## (3) 仮説検定

ここでは、今回対象とする2つのモデル、つまり、2つの時点のデータを用いた定数項の修正によるモデルと、新しい時点のデータのみを用いたモデルの予測精度の差異を統計的に検定する方法を説明する。

ここで、帰無仮説と対立仮説を、今回対象とする2つのモデルの予測精度を母数として、次のように記述する。帰無仮説は今回対象とする2つのモデルの予測精度が等しいということであり、対立仮説は2つのモデルの予測精度が等しくない (一方のモデルの予測精度が優れている)、ということである。今回は、予測精度を2001年のデータへの対数尤度で評価するので、この観点から書き直すと、帰無仮説は今回対象とする2つのモデルの2001年のデータへの対数尤度が等しいということであり、対立仮説は2つのモデルの2001年のデータへの対数尤度が等しくない (一方のモデルの2001年のデータへの対数尤度が大きい)、ということである。

この検定は、検定統計量として、2つのモデルによる2001年のデータへの対数尤度の差を考えることで実行できる。今回は、この統計量の標本分布をブートストラップ法によって近似する。具体的には、次の式(5)で示される  $x_b$  を繰り返し抽出されたブートストラップ標本を用いて計算する。

$y$  と  $n$  に関して2つの場合、 $y_1$  と  $m_1$  (古い時点の多数のサンプル)、 $y_2$  と  $m_2$  (新しい時点の少数のサンプル) を考える。ここで、 $y_1 < y_2$  か  $m_1 \geq m_2$  となる  $y_1$ ,  $y_2$ ,  $m_1$ ,  $m_2$  について、以下の変数  $x_b$  ( $b=1, 2, \dots, 1000$ ) を定義する。なお、

ここで、 $L2(y_2, m_2, b)$ と $L1(y_1, y_2, m_1, m_2, b)$ の計算において、同一の $b$ が使われている。

$$x_b = L2(y_2, m_2, b) - L1(y_1, y_2, m_1, m_2, b) \quad (5)$$

なお、ここで総ての $b(=1, 2, \dots, 1000)$ に対して良好な推定結果が得られるわけではない。特に $m_1$ や $m_2$ が小さい場合においてこのことは問題となる。そこで、式(5)の $x_b$ は、 $b$ 回目のランダム抽出のときに $L1$ と $L2$ の両方が計算されたときにのみ、定義されることとする。もし、新しい時点の少数のデータのみを用いるモデルのほうが予測精度が高いのであれば、 $x_b$ は正の値をとるはずである。

2つのモデルによる2001年のデータへの対数尤度差の95%信頼区間は、 $x_b$ の2.5パーセンタイル値と97.5パーセンタイル値をそれぞれ $x_{b(0.025)}$ と $x_{b(0.975)}$ と定義すると、式(6)のように表現される。

$$(x_{b(0.025)}, x_{b(0.975)}) \quad (6)$$

ここで、 $x_{b(0.025)} > 0.0$ または、 $x_{b(0.975)} < 0.0$ のとき、5%の有意水準で帰無仮説を棄却することができる。

## 5. 結果

本章では、まず各時点からの10000サンプルのデータを総て用いた場合の推定結果について述べる。次に、データ収集時点とサンプル数が異なる場合の予測精度の特徴を比較する。最後に、データ収集時点とサンプル数が異なる場合の予測精度の違いを統計的に検定する。

### (1) 推定結果

モデル化にあたって、男性ダミー（男性=1, 女性=0）、20歳以上ダミー（20歳以上=1, 19歳以下=0）、65歳以上ダミー（65歳以上=1, 64歳以下=0）、名古屋ダミー（名古屋市を出発地または到着地とする=1, そうではない=0）を定義した。モデルの変数の記述統計はSanko<sup>4)</sup>に示されているが、1点だけ再度触れておく。このデータでは、20歳以上が多く、65歳以上が少なくなっている。20歳以上の比率は1971, 1981, 1991, 2001年の順に94.1%, 96.2%, 97.1%, 98.8%であり、同様に65歳以上の比率は1.5%, 1.6%, 2.1%, 3.1%である。

各時点から抽出した10000サンプルのデータを用いてモデルを構築した結果を表-1に示す。1971, 1981, 1991, 2001年のデータを個別に用いたモデル推定結果<sup>4)</sup>を表の上方に示す。なお、2001年のデータは予測の検証にしか用いないが、参考のために示す。また、1971年と1981年のモデルについて、前者については1981年と1991年

の10000サンプルを使って定数項とスケールパラメータを修正、後者については1991年の10000サンプルを使って同様の修正を行った結果を表の下方に示す。

モデルの特定化は、1章で述べたように、本研究では主たる分析の対象としておらず、以降の分析は総てここで示したモデルの特定化に基づいている。しかし、2点について触れておく。1点目に、自動車保有の変数は、交通手段選択（自動車選択）と密接な関係があり、内生性の問題を含むため説明変数に採用していない。2点目に、旅行費用については3章で説明した理由で説明変数に含んでいない。なお、モデルの解釈はSanko<sup>4)</sup>を参照されたい。

それぞれのモデルを2001年のデータに適用した予測精度も表-1に併記する（Log-likelihood on 2001 dataの行）。各時点のデータを個別に用いたモデルを比較すると、2001年のデータへの予測精度は1991年が最も良く、以下、1981年, 1971年となる。また、定数項を修正したモデルを比較に含めると、1971年のモデルよりも1971年のモデルを1981年のデータで更新したほうが良く、また1971年のモデルを1991年のデータで更新したほうがさらに良い。しかし、これは、1981年のデータや1991年のデータを個別に用いたときの予測精度には及ばない。つまり、2時点から同一のサンプル数のデータが得られている場合、新しい時点のデータのみを用いたほうが新しい時点のデータを古い時点のデータと一緒に用いるよりも予測精度が高い。同様の議論が1981年のモデルを1991年のデータで更新した場合にも当てはまる。

### (2) 予測精度の特徴

次にブートストラップによる結果を示す。表-2は行番号(1)~(3)と列番号(a)~(c)の組み合わせからなっている。行番号はデータ年の組み合わせを示しており、列番号は表示している結果を示している。まず、予測精度（ $L1(y_1, y_2, m_1, m_2, b)$ と $L2(y_2, m_2, b)$ で定義される2001年データへの対数尤度）の平均値と標準偏差をそれぞれパネル(a), (b)に示す。それぞれの $y_1, m_1, y_2, m_2$ について1000回ずつモデルを構築し予測をしているが、 $m_1$ や $m_2$ が小さい場合、モデルの推定に問題が発生することもある。ここでの平均値と標準偏差は $b = 1, 2, \dots, 1000$ のうち、推定結果および予測結果が得られたものについてのみ算出した。また、問題のある場合を除いた後の残った繰り返し回数についてパネル(c)に示した。

$L1(y_1, y_2, m_1, m_2, b)$ と $L2(y_2, m_2, b)$ の平均値を表-2のパネル(1-a)~(3-a)に示す。この図の見方を $y_1 = 1971, y_2 = 1981$ の場合のパネル(1-a)を例に説明する。横軸に新しい時点( $y_2 = 1981$ )のサンプル数 $m_2$ をとって $L2(y_2 = 1981, m_2, b)$ の平均値を縦軸に描く。 $m_2$ は100から10000まで変化させているので100~10000の範囲で $L2$ の平均値は得られている。今度

表-1 推定結果

	1971		1981		1991		2001 <sup>a</sup>	
Variables	Est.	t-stat.	Est.	t-stat.	Est.	t-stat.	Est.	t-stat.
Constant (B)	0.127	2.42	-0.392	-6.21	-0.638	-8.98	-1.03	-12.11
Constant (C)	-1.15	-9.84	-0.645	-4.65	0.301	1.96	0.560	2.23
Travel time [hr]	-0.606	-6.94	-1.81	-16.47	-1.59	-15.71	-2.60	-20.48
Male dummy (R)	0.577	8.59	0.787	8.70	0.812	7.53	0.511	3.89
Male dummy (C)	1.97	29.44	2.17	25.22	1.78	17.30	1.38	10.91
20 years old or older dummy (C)	0.900	8.28	0.764	5.78	0.776	5.18	0.511	2.06
65 years old or older dummy (B)	1.91	8.89	1.37	5.73	1.33	5.59	0.561	2.05
Nagoya dummy (C)	-1.12	-24.08	-1.77	-33.21	-2.18	-37.81	-2.21	-36.70
N (randomly drawn)	10000		10000		10000		10000	
L( $\beta$ )	-7776.86		-5985.02		-5300.58		-4716.28	
L( $\theta$ )	-8948.26		-8593.88		-8398.85		-8159.63	
Adj rho-squared	0.130		0.303		0.368		0.421	
Log-likelihood on 2001 data	-6521.95		-5225.15		-4801.79		Not applicable	
<i>Updated by 1981 data</i>								
Constant (B)	-0.441	-11.91		—		—		—
Constant (C)	-0.986	-43.63		—		—		—
Scale parameter	1.27	44.49		—		—		—
N (randomly drawn)	10000			—		—		—
L( $\beta$ )	-6138.88			—		—		—
Log-likelihood on 2001 data	-5672.04			—		—		—
<i>Updated by 1991 data</i>								
Constant (B)	-0.751	-16.18	-0.732	-13.26		—		—
Constant (C)	-0.667	-26.30	-0.271	-9.13		—		—
Scale parameter	1.21	42.41	1.02	45.29		—		—
N (randomly drawn)	10000		10000			—		—
L( $\beta$ )	-5662.31		-5366.31			—		—
Log-likelihood on 2001 data	-5314.67		-4894.30			—		—

Note: (R), (B), and (C) notations refer to alternative-specific variables for rail, bus, and car, respectively. Variables without notations are generic.

<sup>a</sup> 2001 is the target year of forecast, and a model from 2001 is not required but is presented for a comparison purpose.

はL1の結果について $m_1 = 100$ の場合についてL1 ( $y_1 = 1971, y_2 = 1981, m_1 = 100, n_2, b$ )の平均値を縦軸に描く。ここで、 $m_1 \geq m_2$ なので $m_2 = 100$ の場合のみ描かれる。同様に $m_1 = 200 \sim 10000$ の場合についても描く。これらの線については凡例がないので分かりづらいようにも見えるが、線が $m_2 = 100$ まで描かれているときは $m_1 = 100$ のときのL1の平均値、線が $m_2 = 200$ まで描かれているときは $m_1 = 200$ のときのL1の平均値、というように考えると分かりやすい。なお、 $m_2 = 100 \sim 10000$ の範囲で描かれているのはL2 ( $y_2 = 1981, n_2, b$ )とL1 ( $y_1 = 1971, y_2 = 1981, m_1 = 10000, n_2, b$ )の2つである。どちらも青系統の色で描かれているが、点線で描かれているものがL2 ( $y_2 = 1981, n_2, b$ )である。

(1-a)～(3-a)のパネルを見ると、全体的な傾向として右肩上がりであり、L2の場合には $m_2$ が大きいほど、またL1の場合には $m_1$ が同じであれば $m_2$ が大きいほど予測精度が平均して高いことが分かる。また、右肩上がりの傾向はL2においてより顕著である。さらに、これも全体的な傾向としてL1は $m_2$ が同じであれば $m_1$ が大きいほど予測精度が平均して高いことが分かる。ところがL2の線がL1の線

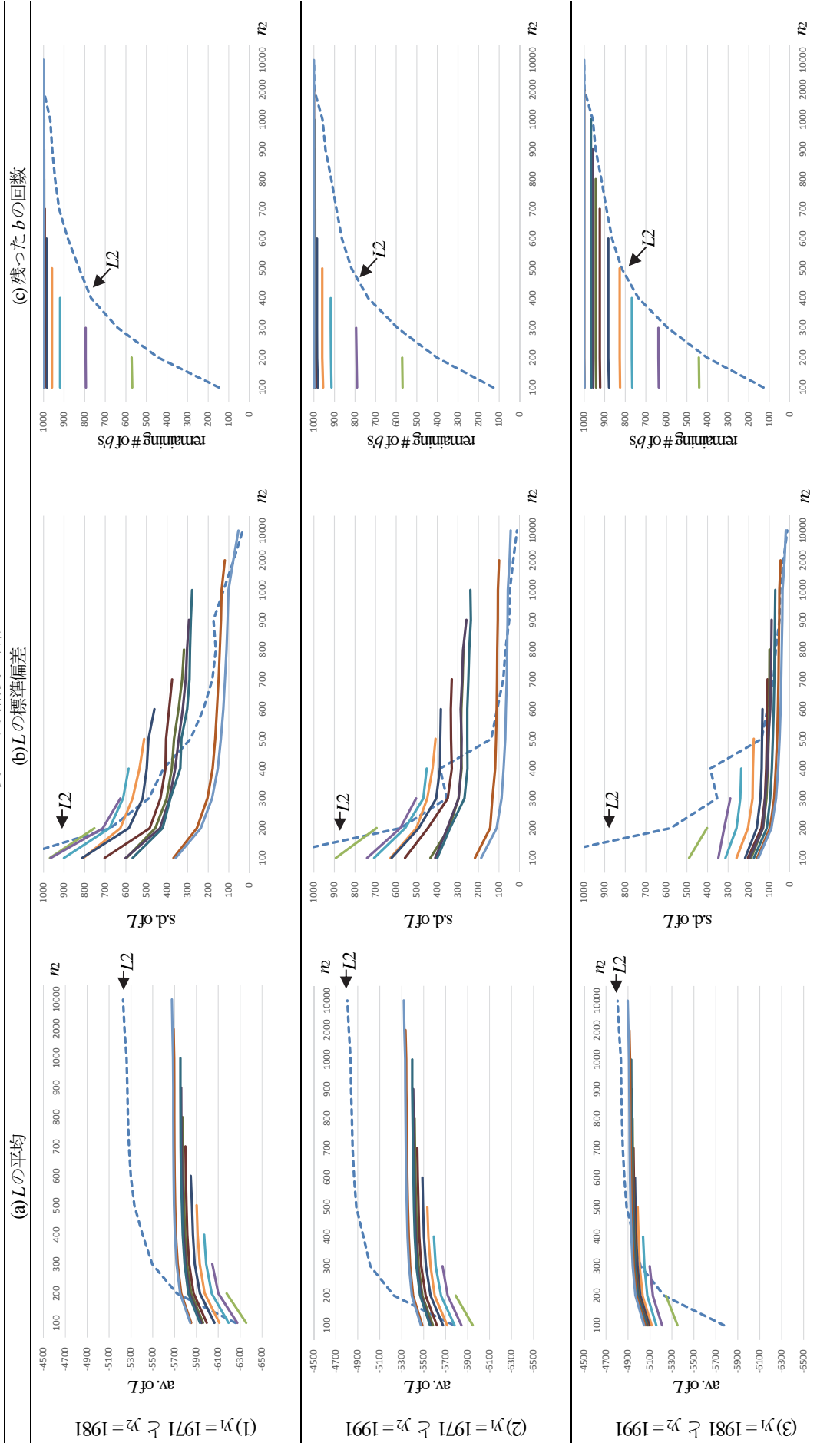
よりも上に現れることがある。これは、新しいデータのみを用いてモデルを構築したほうが古いモデルを新しいデータで更新するよりも良い予測を平均して行えることを示している。

次に、L1とL2の標準偏差を表-2のパネル(1-b)～(3-b)に描く。今度は、全体的な傾向として右肩下がりであり、L2の場合には $m_2$ が大きいほど、またL1の場合には $m_1$ が同じであれば $m_2$ が大きいほど予測精度のばらつきが小さいことを示している。また、右肩下がり傾向はL2においてより顕著である。さらに、これも全体的な傾向としてL1は $m_2$ が同じであれば $m_1$ が大きいほど予測精度のばらつきが小さいことを示している。先ほどとは少し違うのが、L1とL2の線が交差するときの新しい時点からのサンプル数 $m_2$ は平均値の場合よりも大きいことである。

ここで、パネル(1-a)～(3-a)とパネル(1-b)～(3-b)の考察を重ねると、次のことが分かる。 $m_2$ が小さいときには定数項の修正を用いたほうが予測精度は平均して高くそしてそのばらつきも小さい。 $m_2$ が増えると定数項の修正を用いたほうが予測精度は平均して低いとそのばらつきは小



表-2 予測精度の特徴



さい。さらに $m_2$ が増えると定数項の修正を用いたほうが予測精度は平均して低くそのばらつきも大きい。

表-2のパネル(1-c)~(3-c)は、それぞれの場合で試みた1000回のブートストラップのうち、何回で推定と予測が正しく行えたかを示したものである。定数項の修正を行う場合のほうが推定と予測を正しく行えた。表-1が示すように、モデルは3つの選択肢のうち2つの選択肢に定数項を含んでいる。そのため、モデルが正しく推定されるためには、それぞれの選択肢が最低でも1回選ばれていなければならない。また、同様のことがダミー変数についても当てはまる。定数項の修正については、古い時点のデータにおいて定数項とダミー変数の制約、新しい時点のデータにおいて定数項の制約を満たしている必要がある。一方、新しい時点のデータのみを用いる方法では、新しい時点のデータにおいて定数項とダミー変数の制約を満たさなければならない。今回の分析では新しい時点のデータ数のほうが古い時点のデータ数よりも小さいあるいは等しいという設定をしているので、新しい時点のデータで制約を満たすほうが難しい。新しい時点のデータで制約を満たすことが容易である定数項の修正のほうが問題が少なかったと考えられる。もちろん、古い時点のデータのほうがより偏った分布をしているのであれば、今回とは逆の結果が得られる可能性もあるが、今回はそうはならなかった。

次に、データ年の違いについて考察する。パネル(1-a)と(2-a)では $m_2 = 200$ 程度以上において $L2$ のほうが平均して高い予測精度をもたらすのに対して、パネル(3-a)では $m_2 = 500$ 程度以上において $L2$ のほうが平均して高い予測精度をもたらす。定数項の修正において、古い時点のサンプル数が多いことは、古い時点の文脈でパラメータを精度良く推定することに貢献している。定数項の修正においては、定数項以外の移転可能性を仮定することから、定数項以外のパラメータに移転可能性があるのであれば、これを多数のサンプルで精度良く推定することは予測精度の向上につながる。表-1において、修正前の1971年のモデルよりも修正前の1981年のモデルのほうが2001年の予測精度が極めて高い。この要因の一部は定数項以外の移転性の高さによるものと考えられる。また、1971年のモデルと1981年のモデルを同じ1991年のデータで更新したときの予測精度は、それぞれ-5314.67と-4894.30となっている。定数項を修正しても1971年モデルの予測精度は1981年モデルのそれに比べて劣っており、1971年モデルの定数項以外の移転性が低いことがうかがえる。また、定数項以外のパラメータ推定値を見ても、1971年モデルよりも1981年モデルのほうが、予測対象年である2001年モデルに近いものが多い。以上のことから、1971年のデータを増やして定数項以外のパラメータを精度よく推定することは、1981年のデータを増やして定数項以外のパ

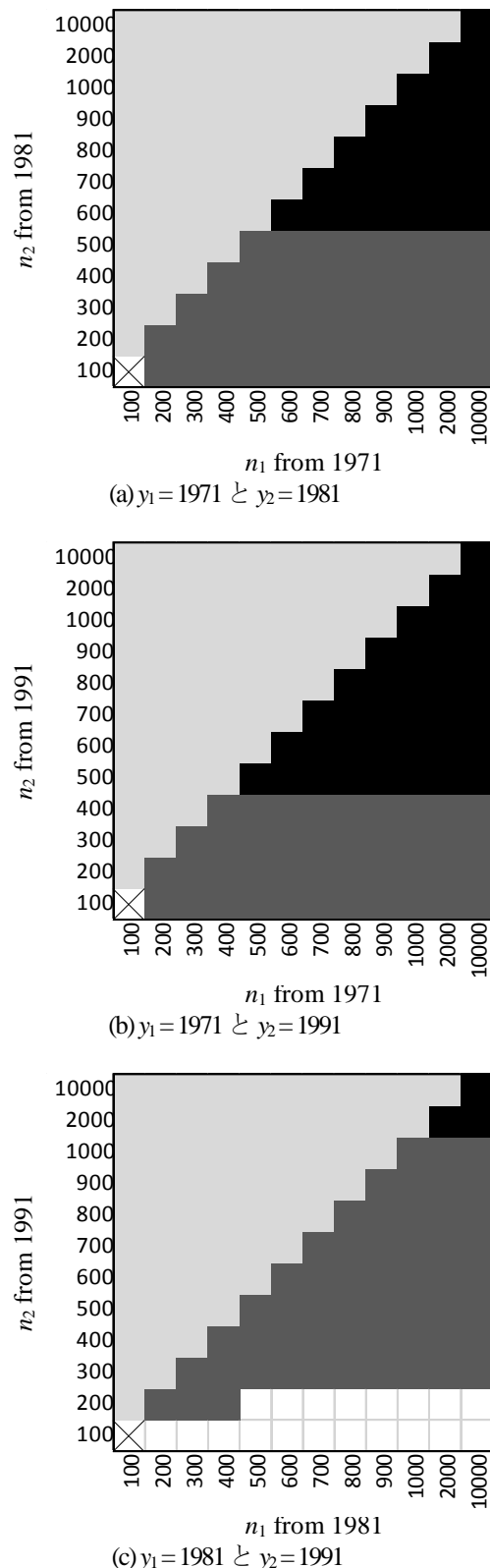
ラメータを精度良く推定することに比べて予測精度の向上への貢献が小さいと解釈できる。

### (3) 予測精度の差の検定

古い時点の多数のサンプルと新しい時点の少数のサンプルを両方用いて定数項を修正したモデルと、新しい時点の少数のサンプルのみを用いたモデル、の2つのどちらが予測精度が良いのかを検定した結果を図-1に示す。1971年と1981年の場合をパネル(a)に、1971年と1991年の場合をパネル(b)に、1981年と1991年の場合をパネル(c)に示す。なお、ここで示す結果は、繰り返しのbのうち、古い時点の多数のサンプルと新しい時点の少数のサンプルを両方用いて定数項を修正したモデルと、新しい時点の少数のサンプルのみを用いたモデル、の両方で正しくモデルが構築されて、予測が得られた場合のみに基づいている。先ほどの表-2のパネル(1-c)~(3-c)で $L1$ または $L2$ それぞれについて残っている繰り返し回数について示したが、ここでの残っている繰り返し回数は2つの方法を同時に考慮しているので注意が必要である。

図-1では、横軸に古い時点のサンプル数、縦軸に新しい時点のサンプル数をとった。各パネルの左上方にある薄い灰色で塗りつぶされている領域は $m_1 < m_2$  (古い時点のサンプル数よりも新しい時点のサンプル数のほうが大きい) であり、今回の興味の対象外である。黒色で塗りつぶされている領域は、式(6)で示した95%信頼区間において、 $x_b(0.025) > 0.0$ を満たしており、新しい時点の小さいサンプルを用いたほうが、2つの時点のデータを用いて定数項の修正をするよりも5%の有意水準で予測精度が良いことを示している。濃い灰色で塗りつぶされている領域は、 $x_b$ の中央値に関して $x_b(0.5) > 0.0$ を満たしており、統計的に5%有意ではないものの、新しい時点の小さいサンプルを用いたほうが、2つの時点のデータを用いて定数項の修正をするよりも予測精度が良いことが多かったことを示している。白色の領域は統計的に5%有意ではないものの、2つの時点のデータを用いて定数項の修正をするほうが、新しい時点の小さいサンプルを用いるよりも、予測精度が良いことを示している ( $x_b(0.5) < 0.0$ )。なお、式(6)で示した95%信頼区間において、 $x_b(0.975) < 0.0$ となることは1回もなかった。つまり、古い時点のサンプルと新しい時点のサンプルを両方用いる定数項の修正のほうが、新しい時点のサンプルのみを用いるよりも有意に優れた予測を行う場合は全くなかった。

まず、図-1の黒く塗りつぶされている新しい時点の小さいサンプルのみを用いたほうが統計的に有意に良い予測を行える場合を検討する。いずれのパネルにおいても、古い時点のサンプル数が一定程度あり、新しい時点のサンプル数も一定程度あるときに、新しい時点のサンプルのみで有意に良い予測を行うことができた。パネル(a)を



注：薄い灰色は  $n_1 < n_2$  であり、今回の興味の対象外。黒と濃い灰色は新しい時点のサンプルのみを用いたほうが定数項の修正よりも予測精度が高い。ただし、黒は5%有意（式(6)で示した2.5パーセンタイル値が正）であり、濃い灰色は統計的に有意ではない（ $x_0$ の中央値が正）。×印で示した  $n_1=n_2=100$  は正しく計算された  $x_0$  が40回に満たないため検定できていない。

図-1 予測精度の差の検定

見ると、古い時点1971年のデータが600サンプル以上、新しい時点1981年のデータが600サンプル以上あれば、新しい時点のデータのみを用いるほうが有意に良い予測を行える。興味深いのは、古い時点のサンプル数が大きく変化しても（1971年のサンプル数が600～10000の範囲で変化しても）、古い時点のサンプルと新しい時点のサンプルを同時に用いるよりも有意に良い予測を行うために必要な最小限の新しい時点のサンプル数は驚くほど安定していることである（1981年の600サンプル）。このことは、新しい時点でのサンプル数をいくらにしたらよいか、という定量的な知見をもたらすのに役立つと考えられる。同様の傾向がパネル(b)と(c)でも見られた。

これを概念として図-2を用いて説明し、これを実際のデータで確認した例を図-3に示す。なお、図-1で示した検定においては、定数項の修正と新しい時点のデータのみを用いるという2つの手法の両方で推定と予測に問題のなかったブートストラップの繰り返し回数を用いている。しかし、図-2の考察においては、2つの手法のそれぞれ独立に推定と予測に問題のなかったブートストラップの繰り返し回数を用いて作成した表-2の結果を利用している。また、図-3の実際のデータを用いた例についても、表-2を作成したのと同様の方法で作成している。図-2, 3は結果そのものというよりは、結果の解釈を考察する概念を示し、それを実際のデータで確認するというものであるため、この点についての厳密性は追求しないものとする。

まず、図-2を説明する。ここでは簡便のため正規分布の形状を利用しているが、本研究では分布形を仮定していないため、必ずしも正規分布である必要はない。2つのパネルがあるが、両者に共通して、横軸はlog-likelihood on 2001、縦軸はその密度を示している。定数項の修正によるモデルの予測精度の分布をL1、新しい時点のデータのみによるモデルの予測精度の分布をL2として表現する（例えば、図-1のパネル(a)の  $m_1=2000, m_2=500$  の場合）。ここで、これらの分布から計算された  $x_0$  によって帰無仮説が棄却されなかったとする。

ここで、新しい時点からのサンプル数が増えた場合のL1とL2の分布の変化の様子を示したのがパネル(a)である。L1とL2はそれぞれL1'とL2'のように分布は右へ移動し、その分布のばらつきは小さくなる（例えば、図-1のパネル(a)の  $m_1=2000, m_2=600$  の場合）。これは、表-2のパネル(a)と(b)から得られた知見であり、新しい時点のサンプル数  $m_2$  が増えるにつれて、予測精度は良くなり、そのばらつきは小さくなる。また、L2のほうが、予測精度の向上およびばらつきの減少が顕著である。つまり、2つの分布の間の差はL2のほうがより大きく右へ移動するために広がり、また、2つの分布ともにばらつきが小さくなっているために、帰無仮説が棄却されたと判断される。

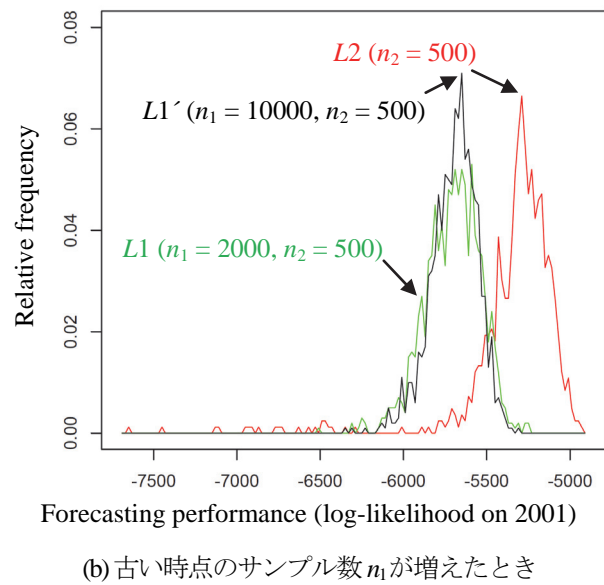
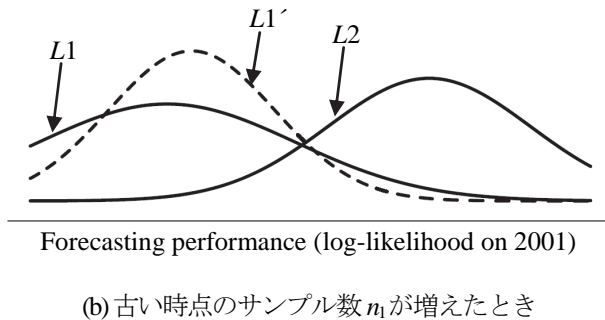
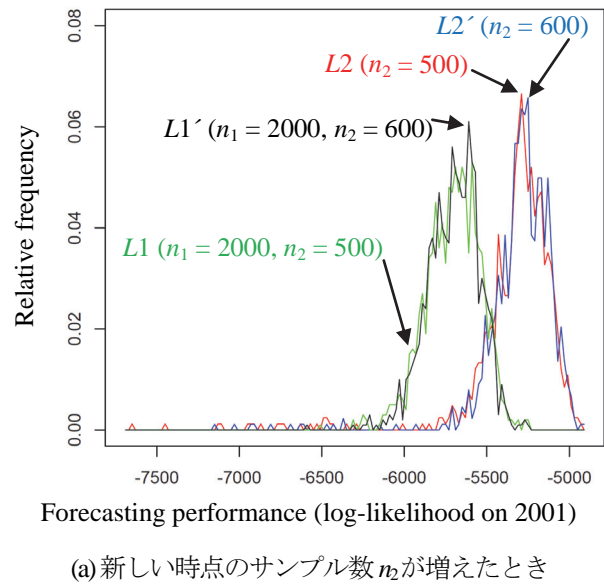
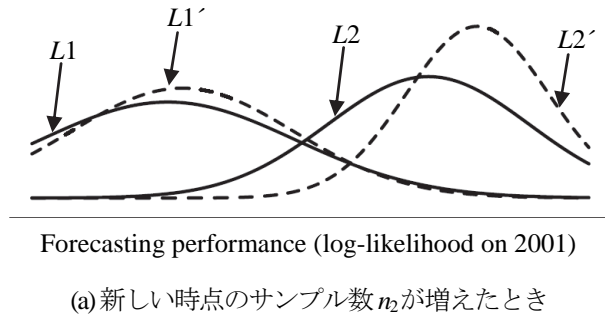


図-2 予測精度の分布 (概念図)

一方、古い時点からのサンプル数が増えた場合の  $L1$  と  $L2$  の分布の変化の様子を示したのがパネル(b)である。 $L2$  は新しい時点のサンプルしか用いないので、古い時点のサンプル数が増えても分布に変化はない。一方、 $L1$  は  $L1'$  のように分布は右へ移動し、その分布のばらつきは小さくなる (例えば、図-1のパネル(a)の  $m_1 = 10000$ ,  $m_2 = 500$  の場合)。これも、表-2のパネル(a)と(b)から得られた知見であり、古い時点のサンプル数が増えるにつれて、予測精度は良くなり、そのばらつきは小さくなる。つまり、2つの分布の間の差は小さくなっており、ここで仮に  $L1$  のばらつきが、分布の間の差が小さくなる影響以上に、小さくなっていけば、帰無仮説が棄却されたと判断される。しかし、今回は古い時点のサンプル数を増やし

注：簡便のため図中で  $y_1 = 1971$ ,  $y_2 = 1981$ ,  $b$  の記述を省略。

図-3 予測精度の分布 (実際)

ても検定の結果が5%の有意水準で変化することはなかった。(同様に、例えば、図-1のパネル(a)の  $m_1 = 2000$ ,  $m_2 = 600$  と  $m_1 = 10000$ ,  $m_2 = 600$  の場合において、図-2のパネル(b)を作図した場合、古い時点のサンプル数が増えるにつれて、予測精度は良くなり、そのばらつきは小さくなる。つまり、2つの分布の間の差は小さくなっており、ここで仮に分布の間の差が、 $L1$  のばらつきが小さくなる影響以上に小さくなっていけば、帰無仮説が棄却されないと判断される。しかし、今回は古い時点のサンプル数を増やしても検定の結果が5%の有意水準で変化することはない。これが、古い時点のサンプルと新しい時点のサンプルを同時に用いるよりも有意により予測を行うために必要な最小限の新しい時点のサンプル数が驚

くほど安定している, という結果につながると解釈できる。

このことを実際のデータで確認するために, 横軸に Log-likelihood on 2001 を階級幅20でとり, 縦軸にその相対度数をとったものを図-3に示す。図-3のパネル(a)および(b)はそれぞれ図-2のパネル(a)および(b)の説明で示した例を実際のデータで確認したものである。ただし, 1000回の繰り返し回数のうち推定と予測に問題がなかった回数が異なる場合があったため, 縦軸には度数ではなく, 相対度数を用いた。パネル(a)では,  $L1(y_1=1971, y_2=1981, m_1=2000, m_2=500, b)$  および  $L2(y_2=1981, m_2=500, b)$  が, 新しい時点のサンプル数  $m_2$  が増えると, それぞれ  $L1'(y_1=1971, y_2=1981, m_1=2000, m_2=600, b)$  および  $L2'(y_2=1981, m_2=600, b)$  に移動する様子を示している。一方, パネル(b)では,  $L1(y_1=1971, y_2=1981, m_1=2000, m_2=500, b)$  および  $L2(y_2=1981, m_2=500, b)$  が古い時点のサンプル数  $m_1$  が増えると, それぞれ  $L1'(y_1=1971, y_2=1981, m_1=10000, m_2=500, b)$  に移動および変化なしとなる様子を示している。概ね, 先に示した概念図の説明が当てはまる結果が得られることが確認できた。

もう1つの帰無仮説を棄却しない理由は, サンプルの代表性によるものである。今回の分析では, (2)節で示したように, モデルが正しく推定されて予測が行われるためには, 定数項とダミー変数に関して満たさなければならない制約がある。しかし, サンプル数が少ないときにはこの制約を満たさない場合が増え(表-2のパネル(c)の知見) ブートストラップ抽出は  $x_b$  の計算から除かれる。したがって, 小さいサンプル数のときに計算される  $x_b$  はサンプルの代表性を満たしていない可能性がある。

2時点の間隔が10年である図-1のパネル(a), (c)と20年であるパネル(b)を比較する。20年間隔であるパネル(b)のほうが新しい時点のサンプル数がより少なくても, 2時点のデータを用いる定数項の修正よりも有意に良い予測を行える。例えば, パネル(a)と(b)は古い時点のデータは共に1971年のものであるが, 新しい時点のデータはパネル(b)はそれより20年若い1991年のもの, パネル(a)は10年若い1981年のものである。データ収集間隔が長くて文脈が大きく異なると考えられるときには, 新しいデータにより価値があるために, 新しい時点のデータが少なくても有意に良い予測をもたらすことができると考えられる。

2時点の間隔がともに10年である図-1のパネル(a)と(c)を比較する。このとき, 有意に良い予測をもたらす新しい時点のサンプル数はパネル(a)の1971/1981のほうがパネル(c)の1981/1991よりも小さい。これは, 交通行動の変化のペースが均一ではないことと関係している。3章でも示したように, 交通手段のシェアは1971年から1981年にかけては大きく変化するが, それ以降の変化のペースは緩やかである。交通行動が大きく変化している時期には,

新しいデータがより大きな意味を持ち, 新しい小さいデータのみを用いる場合でも有意に良い予測ができることを示している。

以上の考察をもとに, 定数項の修正を用いる意義について議論する。まず, 図-1の統計的検定によると, 新しい時点と古い時点のサンプルを同時に用いる定数項の修正が新しい時点のデータのみを用いる場合よりも統計的に有意に良い予測をもたらすことは, いかなるデータ収集年とサンプル数の組み合わせにおいても存在しなかった。つまり, 定数項の修正を用いる意義は統計的検定に求めることはできない。次に, 統計的に有意ではないが, 定数項の修正のほうが良い予測をもたらす場合について検討する。これは, 表-2のパネル(a)において  $L2$  のほうが上方に描かれている場合, 図-1において着色されていない場合に相当する。表-2からは, 1971/1981と1971/1991の場合は新しい時点のサンプル数が100まで, 1981/1991の場合には新しい時点のサンプル数が400までの極めて限られた範囲では, 定数項の修正を用いるほうが平均的に良い予測を行える。一方, 図-1からは, 1981/1991の場合には新しい時点のサンプル数が200までの極めて限られた範囲では定数項の修正を用いるほうが良い予測を行える回数が多かった。また, 表-2のパネル(b)から判断すると, 定数項の修正を用いたほうが, 新しい時点のサンプル数が上で示した100~400という限られた範囲よりも広い範囲において予測のばらつきが小さい。また, 表-2のパネル(c)から判断すると, 定数項の修正のほうが推定および予測に関する問題が少ない。以上の内容をまとめると, 定数項の修正を用いる意義は, 新しい時点のサンプル数が極めて少なく平均的に優れた予測を行いたい場合や多くの回数で優れた予測を行いたい場合, 新しい時点のサンプル数がそれよりも多く予測精度は平均的に劣っているがばらつきが小さいという意味で安定した予測を行いたい場合, 予測精度に関係なく推定および予測に問題が少ないという意味で安定した予測を行いたい場合, にあると考える。

## 6. おわりに

本研究は, 2時点のデータを用いる定数項の修正(選択肢固有定数項とスケールパラメータの修正)が, 直近の1時点のデータのみを用いるモデルよりも統計的に有意に良い予測精度を持つかを検討する初めての試みである。そのため, 可能な限り分析を簡略化し, また, 筆者の利用可能なデータを用いて分析した。1章でも述べたように, データ収集時点とサンプル数が本研究で関心のある2つの次元であり, 交通行動の理論, モデルの数学的構造, モデルの特定化が与える影響については検討の

対象外であり、今後の研究課題である。

中京都市圏の4時点のパーソントリップ調査データを用いて通勤交通手段選択行動を対象に分析した。ブートストラップ法を用いることで、データ収集時点(1971, 1981, 1991年)の組み合わせとサンプル数(100から10000の範囲の12通り)の組み合わせが2001年の予測精度に与える影響を評価した。

統計的な検定による結果ではないが、ブートストラップによって以下の知見が得られた。

- 古い時点のサンプル数が同じであれば、新しい時点のサンプル数が多いほど、定数項の修正による予測精度は高く、そのばらつきは小さい。
- 新しい時点のデータのみを用いるモデルでも、新しい時点のサンプル数が多いほど、予測精度は高く、そのばらつきは小さい (Sanko<sup>7)</sup>の再確認)。
- 新しい時点のデータが増えるときの予測精度の向上とばらつきの減少の程度は、新しい時点のデータのみを用いるモデルのほうが、定数項の修正によるモデルよりも顕著である。
- 新しい時点のサンプル数が同じであれば、古い時点のサンプル数が多いほど、定数項の修正による予測精度は高く、そのばらつきは小さい。

統計的な検定によると、以下の知見が得られた。

- 新しい時点の少数のサンプルのみを用いたモデルのほうが、古い時点の多数のデータで推定したモデルの定数項とスケールパラメータを新しい時点の少数のデータを用いて更新するよりも有意に良い予測を行えることがある。このような知見が得られる新しい時点のサンプル数は、古い時点のサンプル数が大きく変わっても驚くほど安定している。
- 定数項の修正のほう新しい時点の少数のサンプルのみを用いたモデルよりも有意に予測精度が高いことはいかなる時点とサンプル数の組み合わせにおいても存在しなかった。

以上の議論から、定数項の修正によるメリットは、定数項を修正することによって新しい時点のデータのみを用いるよりも統計的に良い予測を行うことができる、ということには求められない。

定数項の修正のメリットは以下の点にあると考える。

- 新しい時点のサンプル数が極めて少ない場合、定数項の修正のほうが良い予測を平均して行える、あるいは良い予測を行える回数が多い。
- 新しい時点のサンプル数が少ない場合(上の「極めて少ない」よりは多い場合でも)、定数項の修正のほうが予測のばらつきが小さい。
- 定数項の修正のほうが推定と予測の計算に関する問題が少ない。

2時点のデータを用いるモデル更新法については、今

回報った定数項の修正(transfer scaling)以外にも、joint context estimation, Bayesian updating, combined transfer estimation が知られている。これらの場合のデータ収集時点とサンプル数が予測精度に与える影響については、筆者の最新の学会報告<sup>16, 17)</sup>を参照されたい。交通調査に関する予算制約の観点からも本研究のような分析が、効率的な調査の設計のためにも必要と考える。

**謝辞：**本研究はJSPS科研費25380564, 16K03931の助成を受けている。データ使用に関して、中京都市圏総合都市交通計画協議会と名古屋大学森川研究室の支援を受けた。また、3名の匿名の査読員からの指摘および各務和彦准教授(神戸大学大学院経営学研究科)との議論が本論文の改善に有益であり、謝意を表す。本論文は第51回土木計画学研究発表会(2015年6月, 九州大学)とhEART 2015 - 4th Symposium of the European Association for Research in Transportation(2015年9月, デンマーク・コペンハーゲン)での報告内容を基にしている。

## 注

[注1] 実務的には、一定の予測精度を得ることを基準に、定数項の修正におけるサンプル数、新しい時点のデータのみを用いるモデルにおけるサンプル数を議論することもある必要であると考えられる。ただし、本研究では、2時点のデータが利用可能で今回取り扱っている2つの手法による予測が可能であることを仮定しているため、直接的にはこの点について扱わない。ただし、サンプル数を変化させたときの2つの手法の予測精度については表-2に示している。

[注2] ここでの議論は、Sikder<sup>18)</sup>を参考に行っている。Sikderは移転性を分析するときに考慮すべき点として4つを挙げている。(Sikderは地域移転性を分析しているが、この議論は時間移転性にも適用可能と考える。)4つのうちの3つは本文中で述べた(i)-(iii)である。本研究の興味の対象は4つ目の(iv)モデルのパラメータ推定値(例えば、説明変数の係数や弾性値や時間価値などの移転性)である。データ収集時点とサンプル数という2つの次元は、まず、パラメータ推定値に影響を与え、それが予測精度に影響を与えるという構造になっている。なお、本研究はモデルの推定よりも調査に焦点を当てている。(i)-(iii)はいずれもモデル推定に関することで、これらは調査後に分析者が操作することが可能である。しかし、データ収集時点とサンプル数は調査に関する問題で、調査後に分析者が操作することはできない。

[注3] 筆者ら<sup>19)</sup>は中京都市圏のデータを用いて自動車と公共交通の2つの選択肢を持つ通勤交通手段選択モデルを構築し、自動車費用のパラメータは有意に推定されないことを示している。さらに、公共交通の費用のパラメー

タは正に推定された。このことは、費用のパラメータを含まないことを実証的に支持している。

#### 参考文献

- 1) Dissanayake, D., Kurauchi, S., Morikawa, T. and Ohashi, S.: Inter-regional and inter-temporal analysis of travel behaviour for Asian metropolitan cities: case studies of Bangkok, Kuala Lumpur, Manila, and Nagoya, *Transport Policy*, Vol. 19, No. 1, pp. 36–46, 2012.
- 2) Duffus, L. N., Alfa, A. S. and Soliman, A. H.: The reliability of using the gravity model for forecasting trip distribution, *Transportation*, Vol. 14, No. 3, pp. 175–192, 1987.
- 3) Elmi, A. M., Badoe, D. A. and Miller, E. J.: Transferability analysis of work-trip-distribution models, *Transportation Research Record*, Vol. 1676, pp. 169–176, 1999.
- 4) Sanko, N.: Travel demand forecasts improved by using cross-sectional data from multiple time points, *Transportation*, Vol. 41, No. 4, pp. 673–695, 2014.
- 5) Sanko, N., Dissanayake D., Kurauchi, S., Maesoba, H., Yamamoto, T. and Morikawa, T.: Inter-temporal analysis of household car and motorcycle ownership behaviors - the case in the Nagoya metropolitan area of Japan, 1981–2001 -, *IATSS Research*, Vol. 33, No. 2, pp. 39–53, 2009.
- 6) Hensher, D. A., Rose, J. M. and Greene, W. H.: *Applied Choice Analysis: A Primer*, Cambridge University Press, Cambridge, 2005.
- 7) Sanko, N.: Temporal transferability: trade-off between data newness and the number of observations for forecasting travel demand, *Transportation*, Vol. 44, No. 6, pp. 1403–1420, 2017.
- 8) Badoe, D. A. and Miller, E. J.: Comparison of alternative methods for updating disaggregate logit mode choice models, *Transportation Research Record*, Vol. 1493, pp. 90–100, 1995.
- 9) Karasmaa, N. and Pursula, M.: Empirical studies of transferability of Helsinki metropolitan area travel forecasting models, *Transportation Research Record*, Vol. 1607, pp. 38–44, 1997.
- 10) Atherton, T. J. and Ben-Akiva, M. E.: Transferability and updating of disaggregate travel demand models, *Transportation Research Record*, Vol. 610, pp. 12–18, 1976.
- 11) Fox, J., Daly, A., Hess, S. and Miller, E.: Temporal transferability of models of mode-destination choice for the Greater Toronto and Hamilton area, *The Journal of Transport and Land Use*, Vol. 7, No. 2, pp. 41–62, 2014.
- 12) Train, K. E.: A comparison of the predictive ability of mode choice models with various levels of complexity, *Transportation Research Part A*, Vol. 13, No. 1, pp. 11–16, 1979.
- 13) Badoe, D. A. and Miller, E. J.: Analysis of the temporal transferability of disaggregate work trip mode choice models, *Transportation Research Record*, Vol. 1493, pp. 1–11, 1995.
- 14) Sanko, N. and Morikawa, T.: Temporal transferability of updated alternative-specific constants in disaggregate mode choice models, *Transportation*, Vol. 37, No. 2, pp. 203–219, 2010.
- 15) Efron, B. and Tibshirani, R. J.: *An Introduction to the Bootstrap*, Chapman & Hall, London, 1993.
- 16) 三古展弘：時間移転性向上のためのモデル更新法の選択基準, 土木計画学研究・講演集, Vol. 52, 2015.
- 17) Sanko, N.: Criteria for selecting model updating methods for better temporal transferability, *Compendium of Papers of the 95th Annual Meeting of the Transportation Research Board*, Washington D.C., U.S.A., 2016.
- 18) Sikder, S.: Spatial Transferability of Activity-Based Travel Forecasting Models, Ph.D. Dissertation, University of South Florida, Tampa, U.S.A., 2013.
- 19) Sanko, N., Morikawa, T. and Kurauchi, S.: Mode choice models' ability to express intention to change travel behaviour considering non-compensatory rules and latent variables, *IATSS Research*, Vol. 36, No. 2, pp. 129–138, 2013.

(2016. 10. 6 受付)

## TRANSFER SCALING IN TRAVEL DEMAND FORECASTING MODELS: ITS USEFULNESS BASED ON DATA COLLECTION TIME POINTS AND THE NUMBERS OF OBSERVATIONS

Nobuhiro SANKO

Transfer scaling method utilises both a larger data from older time point and a smaller data from more recent time point. A use of only the smaller data from more recent time point, however, reportedly produced better forecasts than the transfer scaling. This study analyses usefulness of the transfer scaling focusing on data collection time points and the numbers of observations. In any combinations of time points and the numbers of observations, the transfer scaling never produced statistically significantly better forecasts than the model using data from only the more recent time point. The transfer scaling method has advantages in cases where the number of observations from the more recent time point is substantially small; it produced better forecasts on average with smaller variance, and it is less likely to produce poor estimates and forecasts.