

時間移転性向上のための モデル更新法の選択基準

三古 展弘¹

¹正会員 神戸大学大学院准教授 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

需要予測において、古い時点と新しい時点のデータが利用可能な場合、2時点のデータを用いるモデル更新法（その場合、どのモデル更新法）を利用するか、新しい時点のデータのみを利用するか検討が必要である。しかし、2時点のサンプル数がどのような場合に、どの方法を採用すべきかは明らかではない。本研究では、中京圏の通勤交通手段選択行動を対象にブートストラップ法を用いて分析する。その結果、(1) 新しい時点のサンプル数が多い場合には新しい時点のサンプルのみを用いるのが良く、(2) 新しい時点のサンプル数が少ない場合、2時点の文脈の差と古い時点のサンプル数に応じてTransfer scaling, Joint context estimation, Combined transfer estimationを用いるのが良い。

Key Words : temporal transferability, model updating, travel demand forecast, sample size, bootstrap, mode choice model

1. はじめに

交通需要予測においては新しい時点の多数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、前者のほうにより良い予測をもたらすことが報告されている (Dissanayake et al, 2012¹⁾, Duffus et al, 1987²⁾, Elmi et al, 1999³⁾, Sanko, 2014⁴⁾, Sanko et al, 2009⁵⁾)。また、多数のデータを用いてモデルを構築するほうがより良い予測を行えることが知られている (Hensher et al, 2005⁶⁾)。もちろん、サンプル数が同じであれば、新しい時点のデータを用いるほうが良く、また、データ収集時点が同じであれば、サンプル数が多いほうが良い。しかし、分析者が常に新しい時点の多数のサンプルを使って予測モデルを構築できるわけではない。

分析者が古い時点のサンプルには多数アクセスできるが、新しい時点のサンプルには少数しかアクセスできない場合は多いと思われる。これは、次のようなときに発生すると考えられる。1つ目に、既に手元に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ある場合である。2つ目に、既に手元に古い時点の多数のサンプルがあるが、新たにデータを収集しようとする場合である。2つ目のケースは、多くの都市圏では大規模な交通調査が低い頻度で行われていることを考えると、より

現実的な設定であると言える。例えば、パーソントリップ調査は日本の3大都市圏では10年に1回実施されており、最新のデータでも10年前に収集されたものであることがある。このとき、新しいデータの収集を検討することもあると考えるが、大規模な調査を独自に行うことは困難であり、独自調査のサンプル数は小さいものにならざるを得ない。

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った場合、データの利用方法は次の3つが考えられる。

- [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を両方使用
- [2] 新しい時点の少数のデータのみを使用
- [3] 古い時点の多数のデータのみを使用

[1]の方法はモデルの更新法として知られており、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ものである。モデルの更新法としてはいくつかの方法が提案されているが、どのような場合に（具体的には、古い時点と新しい時点のサンプル数がいくらの場合に）どのモデル更新法が適切かは分かっていない。さらに、もし、1時点のデータのみを用いた[2]や[3]の方法が優れているのであれば、2つの

時点のデータが利用可能であるからといって、[1]の方法を用いる必要は全くない。したがって、[1]~[3]のうち、最も良い方法を適切に選択する必要がある。

筆者は1時点のデータのみを用いる[2]と[3]の場合を比較し、[3]の古い時点の多数のサンプルのみを用いたモデルが[2]の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統計的に有意に高い予測精度をもたらすことは全くないことを示した（三古⁷⁾、Sanko⁸⁾）。したがって、本研究では、[3]を除外し、[1]のいくつかのモデル更新法と[2]の新しい時点の少数のサンプルのみを用いた場合を検討する。なお、三古⁹⁾、Sanko¹⁰⁾では、[1]と[2]の比較を行っていたが、[1]のモデル更新法として Transfer scalingしか検討していなかった。本研究ではこれを発展させ、モデルの更新法[1]について Transfer scaling, Joint context estimation, Bayesian updating, Combined transfer estimationを検討する。

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を更新するのが良いのか（もしそうならどのモデル更新法が良いのか）。
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を更新するのが良いのか（もしそうならどのモデル更新法が良いのか）。また、小規模な調査を行うときにどの程度のサンプル数を得ることを目標に調査を行えばよいのか。

いま、この問題を分析するに当たって、中京都市圏で得られた1971, 1981, 1991, 2001年のパーソントリップ調査データが利用可能である。ここでは、1971, 1981, 1991年の3時点のデータをモデルの構築と更新に用い、2001年のデータを予測の検証のみに用いる。1971, 1981, 1991年から2つの時点（1つは古い時点、1つは新しい時点）を抽出し、それぞれの時点から様々なサンプル数（100から10000の範囲の12通り）のデータをランダムに抽出する。したがって、この分析では、古い時点と新しい時点のサンプル数がどの程度の場合に、どのモデル更新法または新しい時点のデータのみを用いたモデルが2001年の予測を最も良く行うことができるか、に対する答えを導くことが可能である。さらに、今回はモデルの

構築と更新に用いるデータが得られた2時点として、1971と1981年、1971と1991年、1981と1991年の3通りを検討することができる。また、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。パーソントリップ調査はこれまで同一の政府機関によって実施されており、調査方法が時点間で安定している。そのため、今回の分析でサンプル数以外の要因がコントロールされていると仮定することは妥当と考える。

本研究は、この問題について統計的検定を用いて知見を得ようとする初めての試みであることから、分析を可能な限り簡略化する。モデルの複雑性は主たる興味の対象ではなく、簡単な多項ロジットモデルを用いる。使用するデータは1971~2001年のパーソントリップ調査データであり、最新の2001年のデータも14年前のものであるが、これも主たる興味の対象ではない。

本論文は以下のように構成される。まず、2章において関連する既存研究を紹介する。3章ではデータを説明する。4章では方法論について説明する。多項ロジットモデル、モデル更新法、ブートストラップ、仮説の検定法について説明する。5章ではパラメータ推定値、予測精度の特徴と統計的検定を行い、モデル更新法選択の基準について検討する。最後に、6章で結論を述べる。

2. 既存研究のレビュー

これまで、時間移転性を対象にどのモデル更新法が適切かを分析した研究は多くないが、以下の2つを挙げることができる。

Badoe and Miller¹¹⁾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年の予測精度を比較している。分析の対象はピーク時の出勤交通手段選択行動である。なお、この研究ではデータが2時点からしか得られていないため、2時点目の1986年のデータをモデルの更新にも予測の検証にも用いているという問題がある。予測に用いるモデルは、4種類のモデル更新手法を用いたモデルと1986年の少数のサンプルのみを利用したモデルである。その結果、1986年のサンプルが400から500程度あれば、1986年のデータのみを用いたモデルと、このデータに加えて1986年の多数のサンプルを用いたモデルの間に、予測精度にほとんど差がないという結論が得られている。

同様の分析をKarasmaa and Pursula¹²⁾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。

本研究の基本的な考え方は、2つの既存研究と類似し

ているが、以下のような工夫をし、より意味のある知見が得られるようにしている。

1. 既存研究では古い時点のサンプル数が固定されているため、古い時点のサンプル数が変わった場合については検討できない。本研究では古い時点と新しい時点の両方のサンプル数を変化させる。
2. 既存研究ではデータが2時点からしか得られていないため、2時点目のデータをモデルの更新にも予測の検証にも用いている。本研究では4時点のデータを使って3時点モデルの構築と更新に使用し、残った1時点モデルの検証に使用する。
3. 既存研究では新しい時点での様々なサンプル数をランダムに抽出しているが、その抽出を1回しか行っておらず偶然性に左右される。本研究では、ブートストラップ法を用いて統計的に意味のある結論を導き出す。

3. データ

中京都市圏において1971, 1981, 1991, 2001年の4時点を得られた繰り返し断面データである。パーソントリップ(PT)調査データを用いる。モデルの構築と更新には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。データの詳細については三古⁷⁾またはSanko⁸⁾を参照されたい。しかし、ここでは2点について再度強調しておく。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。2つ目にモータリゼーションが急激に進んだことである。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

本研究の目的は、統計的な検定を用いて複数のモデル更新法を比較することである。このような観点からの研究はこれまでにないので、本研究では分析を簡略化するため、多項ロジットモデルを採用する。しかし、ここで説明する方法論は、他のモデル構造の場合にも適用可能である。本章では、(1)節で多項ロジットモデルを説明し、(2)節でモデルの更新法を説明する。(3)節では今回の分析におけるブートストラップ法の適用について説明し、(4)節ではブートストラップ法を利用したモデル更新法の優劣の検定について説明する。なお、以下の説明

で、古い時点と新しい時点をそれぞれ $t1$ と $t2$ と表記する。

(1) モデル

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、 μ^t は時点 t のスケールパラメータ、 α_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、 β_{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。パラメータは(3)式を最大化することによって推定される。

(2) モデルの更新

ここでは、以下の5つのモデル更新法を検討する。なお、新しい時点のデータのみを用いたモデルもモデル更新法の1つとしてここで説明する。それぞれのモデル更新法については簡単のため、以下では適宜略称を使うことにする。Transfer scalingは定数項とスケールパラメータを推定することから、constantの cst 、Joint context estimationは jnt 、Bayesian updatingは bay 、Combined transfer estimationは com である。また、新しい時点のデータのみを用いたモデルは新しい時点のサンプルは小さいことから $small$ の sma とする。

a) Transfer scaling (略称：cst)

上の(1)節で $t=t1$ を代入してモデルを推定する。ただし、識別のため、スケールパラメータは1に固定する。このとき、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。

$$P_{ip}^{t2} = \frac{\exp\left(\mu^{t2}\left(\alpha_i^{t2} + \sum_k (\hat{\beta}_{ik}^{t1} x_{ikp}^{t2})\right)\right)}{\sum_j \exp\left(\mu^{t2}\left(\alpha_j^{t2} + \sum_k (\hat{\beta}_{jk}^{t1} x_{jkp}^{t2})\right)\right)} \quad (4)$$

ここに、 $\wedge$ は推定値を意味する。

ここで、 $\mathbf{x}$ と $\mathbf{y}$ については $t=2$ のデータを用い、また、式(3)と(4)を用いることで、 μ^{t2} と α^{t2} のみを推定する。なお、説明変数に関するパラメータ $\hat{\beta}^{t1}$ はそのまま用いる。

モデルの予測精度は式(3)の2001年のデータへの対数尤度で表現される。これは、推定されたパラメータ $\hat{\beta}^{t1}$ 、 $\hat{\alpha}^{t2}$ 、 $\hat{\mu}^{t2}$ 、2001年の説明変数 $\mathbf{x}$ と選択結果 $\mathbf{y}$ を式(1)～(3)に代入することによって計算される。

b) Joint context estimation (略称: jnt)

これは、2つの時点のデータを同時に使ってモデルを推定する方法である。時点 $t1$ のデータを用いて式(1)を書き、時点 $t2$ のデータを用いて式(1)を書く。ただし、時点 $t1$ と $t2$ で説明変数に関するパラメータは共通とする($\beta^{t1} = \beta^{t2}$)が、定数項(α^{t1} 、 α^{t2})は時点 $t1$ と $t2$ で異なると考える。スケールパラメータは識別の問題から、時点 $t1$ においては1に固定する($\mu^{t1} = 1$)が、時点 $t2$ においては μ^{t2} をデータから推定する。

モデルの推定は $t1$ のデータを用いた式(3)、 $t2$ のデータを用いた式(3)を書き、その和を最大にするようにして行う。

モデルの予測精度は式(3)の2001年のデータへの対数尤度で表現される。これは、推定された説明変数のパラメータ($\hat{\beta}^{t1} = \hat{\beta}^{t2}$)、時点 $t2$ の定数項($\hat{\alpha}^{t2}$)、時点 $t2$ のスケールパラメータ($\hat{\mu}^{t2}$)、2001年の説明変数 $\mathbf{x}$ と選択結果 $\mathbf{y}$ を式(1)～(3)に代入することによって計算される。

c) Bayesian updating (略称: bay)

上の(1)節で $t=t1$ を代入してモデルを推定する。ただし、識別のため、スケールパラメータは1に固定する。このときのモデルのパラメータを($\hat{\alpha}^{t1} | \hat{\beta}^{t1}$) (これは列ベクトル $\hat{\alpha}^{t1}$ の下に列ベクトル $\hat{\beta}^{t1}$ をつけた列ベクトルを表す)、分散共分散行列を $\hat{\Sigma}^{t1}$ とする。同様に、上の(1)節で $t=t2$ を代入してモデルを推定する。識別のため、スケールパラメータは1に固定する。このときのモデルのパラメータを($\hat{\alpha}^{t2} | \hat{\beta}^{t2}$)、分散共分散行列を $\hat{\Sigma}^{t2}$ とする。

このとき、更新されたパラメータは式(5)で示される。

$$\begin{pmatrix} \alpha^{up} \\ \beta^{up} \end{pmatrix} = \left((\hat{\Sigma}^{t1})^{-1} + (\hat{\Sigma}^{t2})^{-1} \right)^{-1} \left((\hat{\Sigma}^{t1})^{-1} \begin{pmatrix} \hat{\alpha}^{t1} \\ \hat{\beta}^{t1} \end{pmatrix} + (\hat{\Sigma}^{t2})^{-1} \begin{pmatrix} \hat{\alpha}^{t2} \\ \hat{\beta}^{t2} \end{pmatrix} \right) \quad (5)$$

モデルの予測精度は式(3)の2001年のデータへの対数尤度で表現される。これは、更新されたパラメータ α^{up} 、 β^{up} 、スケールパラメータ1、2001年の説明変数 $\mathbf{x}$ と選択結果 $\mathbf{y}$ を式(1)～(3)に代入することによって計算される。

d) Combined transfer estimation (略称: com)

c)のBayesian updatingと同じように $t=t1$ と $t=t2$ においてモデルを構築する。ここで、Bayesian updatingでは、($\alpha^{t1} | \beta^{t1}$) = ($\alpha^{t2} | \beta^{t2}$) = ($\alpha^{up} | \beta^{up}$)となることを仮定していたが、Combined transfer estimationでは、バイアスとして $\Delta = (\alpha^{t2} | \beta^{t2}) - (\alpha^{t1} | \beta^{t1})$ が存在すると考える。

このとき、更新されたパラメータは式(6)で示される。

$$\begin{pmatrix} \alpha^{up} \\ \beta^{up} \end{pmatrix} = \left((\hat{\Sigma}^{t1} + \Delta \Delta')^{-1} + (\hat{\Sigma}^{t2})^{-1} \right)^{-1} \left((\hat{\Sigma}^{t1} + \Delta \Delta')^{-1} \begin{pmatrix} \hat{\alpha}^{t1} \\ \hat{\beta}^{t1} \end{pmatrix} + (\hat{\Sigma}^{t2})^{-1} \begin{pmatrix} \hat{\alpha}^{t2} \\ \hat{\beta}^{t2} \end{pmatrix} \right) \quad (6)$$

このとき $\Delta = \mathbf{0}$ とおけば、c)と一致するので、これはBayesian updatingの一般化である。

モデルの予測精度に関する計算はc)のBayesian updatingで説明したのと全く同じである。

e) Small sample, 新しい時点の小さいサンプルのモデル (略称: sma)

上の(1)節で $t=t2$ を代入してモデルを推定する。ただし、識別のため、スケールパラメータは1に固定する。モデルの予測精度は式(3)の2001年のデータへの対数尤度で表現される。これは、推定されたパラメータ($\hat{\beta}^{t2}$ 、 $\hat{\alpha}^{t2}$)、スケールパラメータ1、2001年の説明変数 $\mathbf{x}$ と選択結果 $\mathbf{y}$ を式(1)～(3)に代入することによって計算される。

(3) ブートストラップ¹³⁾

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。各年から同じサンプル数を抽出するのは、そこからブートストラップを行うことになる、サンプル数の違いが結果に与える影響を避けるためである。また、10000サンプルとしたのは、ブートストラップにおける計算時間を節約するた

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、3つの変数 y , n , b を導入する。

- y はデータ収集年であり、以下の3時点を考える：
1971, 1981, 1991
- n はサンプル数であり、以下の12通りを考える：100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 10000
- b はブートストラップの繰り返し回数であり200回繰り返す： $b=1, 2, \dots, 200$

まず、それぞれのデータ年 y (1971, 1981, 1991の3通り)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200($b=1, 2, \dots, 200$)回ランダムに復元抽出する。このとき、同じデータ年 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 $3 \times 12 \times 200 = 7200$ 通りのデータが生成された。

本研究では、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同時に検討するので、さらに、次の変数を導入する。まず、 y に関して、古い時点と新しい時点をそれぞれ y_1 , y_2 とする。また、古い時点と新しい時点から得られたサンプル数をそれぞれ m_1 , m_2 とする。数式で表現すると、 $y_1 < y_2$ および $m_1 \geq m_2$ となる。1つ目の不等式は y_1 よりも y_2 のほうが新しいことを示しており、2つ目の不等式は古い時点のサンプル数 m_1 が新しい時点のサンプル数 m_2 と等しいかまたは大きいことを示している。

y_1 , y_2 , m_1 , m_2 について考えられる組み合わせは、 y に関する3通りの組み合わせ($(y_1, y_2) = (1971, 1981), (1971, 1991), (1981, 1991)$)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。ここで、それぞれの y_1 , y_2 , m_1 , m_2 について必要となる作業を説明する。

まず、 y_1 の時点のデータのみを用いて200回、 y_2 の時点のデータのみを用いて200回の推定を行う。そして、(2)節のa)~e)で必要になる作業を説明する。

- a)のTransfer scalingでは、 y_1 の時点のデータを用いた200回の推定結果のそれぞれについて y_2 の時点のデータを用いて更新するために推定が必要である。
- b)のJoint context estimationでは y_1 と y_2 の時点のデータである $m_1 + m_2$ サンプルを同時に使ってモデルを200回推定する必要がある。
- c)のBayesian updatingおよびd)のCombined transfer estimationにおいては y_1 の時点のデータ、 y_2 の時点のデータを用いたモデルの推定値を使って計算するだけなので新たに推定を行う必要はない。
- e)については既に行った y_2 の時点のデータを用いた推定結果を用いればよいので新たに推定をする必要はない。

以上をまとめると、次のそれぞれの場合について200回ずつ推定を行う必要がある。つまり、 y_1 のみを使った場合、 y_2 のみを使った場合、Transfer scaling, Joint context estimationである。したがって、 $200 \times 4 = 800$ 回の推定が必要になる。先ほど述べたように、本研究では、 y_1 , y_2 , m_1 , m_2 に関しては234通りの場合があることから、合計で $800 \times 234 = 187200$ 回の推定を行う必要がある。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ので、実際に作業をするときには187200回推定を行う必要はない。

予測精度を2001年のデータへの対数尤度で表現する。 y_1 時点のサンプル数 m_1 , y_2 時点のサンプル数 m_2 における b 回目の抽出データにおいて、モデル更新法 m による予測精度を $Lm(y_1, y_2, m_1, m_2, b)$ と表現する。ただし、 m はモデル更新法を示し、*cst* (Transfer scaling), *jnt* (Joint context estimation), *bay* (Bayesian updating), *com* (Combined transfer estimation), *sma* (Small sample, 新しい時点のデータのみを用いたモデル) である。なお、*sma*では新しい時点のデータしか用いていないということを明示的に示すため、 $Lsma(\bullet, y_2, \bullet, m_2, b)$ と書くこともできる。

(4) 仮説検定

ここでは、 y_1 , y_2 , m_1 , m_2 が同じときに、どのモデル更新法が統計的に優れているかを検定する方法を説明する。いま、 $y_1 < y_2$ かつ $m_1 \geq m_2$ を満足する y_1 , y_2 , m_1 , m_2 の組み合わせのそれぞれについて、異なるモデルの更新法 m_1 と m_2 によって将来予測が行われたとき、以下の変数を定義する。

$$x_b = Lm_1(y_1, y_2, m_1, m_2, b) - Lm_2(y_1, y_2, m_1, m_2, b) \quad (7)$$

なお、ここで総ての b ($= 1, 2, \dots, 200$)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(7)の x_b は、 b 回目のランダム抽出のときに Lm_1 と Lm_2 の両方が計算されたときにのみ、定義されることとする。

x_b の意味しているところは、この値が正であれば、更新法 m_1 のほうが m_2 よりも優れているということである。ここで、帰無仮説 H_0 と対立仮説 H_1 を下に示す。

$$H_0: x_b = 0$$

$$H_1: x_b \neq 0$$

ここで、次の z を定義する。

表一1 各時点のデータを個別に用いたモデルの推定結果

Variables	1971		1981		1991		2001 ^a	
	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(β)	-7776.86		-5985.02		-5300.58		-4716.28	
L(θ)	-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	

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

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

$$z = \frac{\bar{x}_b}{s(x_b)} \quad (8)$$

ここに、 $\bar{x}_b$ と $s(x_b)$ はそれぞれ、 x_b の平均と標準偏差。

ここで、 x_b が標準正規分布に従っていると仮定すると、 $z > 1.96$ のときに更新法 $m1$ のほうが更新法 $m2$ よりも 5% の有意水準で予測精度が高いことを示している。

5. 結果

ここでは、まず 10000 サンプルを使ったモデルの推定結果について述べ、予測精度の特徴と統計的検定の結果を説明する。

(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とその解釈は三古⁷⁾とSanko⁸⁾に示されているが、重要な点についてのみ再度説明する。1971、1981、1991、2001年の順に、20歳以上のシェアは94.1%、96.2%、97.1%、98.8%、65歳以上のシェアは1.5%、1.6%、2.1%、3.1%となっている。極めて高い20歳以上のシェアときわめて低い65歳以上のシェア、3章で示された小さいバスのシェアの影響により、以下の分析においてサンプル数が小さいときに推定に問題が発生することがある。

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一1に示す。1971、1981、

1991年のデータを用いたモデル推定結果のほかに予測対象時点の2001年のデータを用いたモデルの推定結果も示す。本論文では以降もここでのモデル特定化を用いて分析を進める。なお、3章で述べたように費用の変数は含まれていないことに注意が必要である。モデルの解釈は三古⁷⁾とSanko⁸⁾を参照されたいが、重要な点について再度述べておく。2001年に近い時点のモデルほど推定値が2001年の値に近い。したがって、2001年のデータに適用した予測精度（Log-likelihood on 2001 dataの行）は、高い順に1991、1981、1971年の順になっている。推定値の増加あるいは減少傾向はSanko⁴⁾によって内生的にモデル化され、それが予測精度の向上に寄与することが示されているが、これは本研究での興味の対象外である。

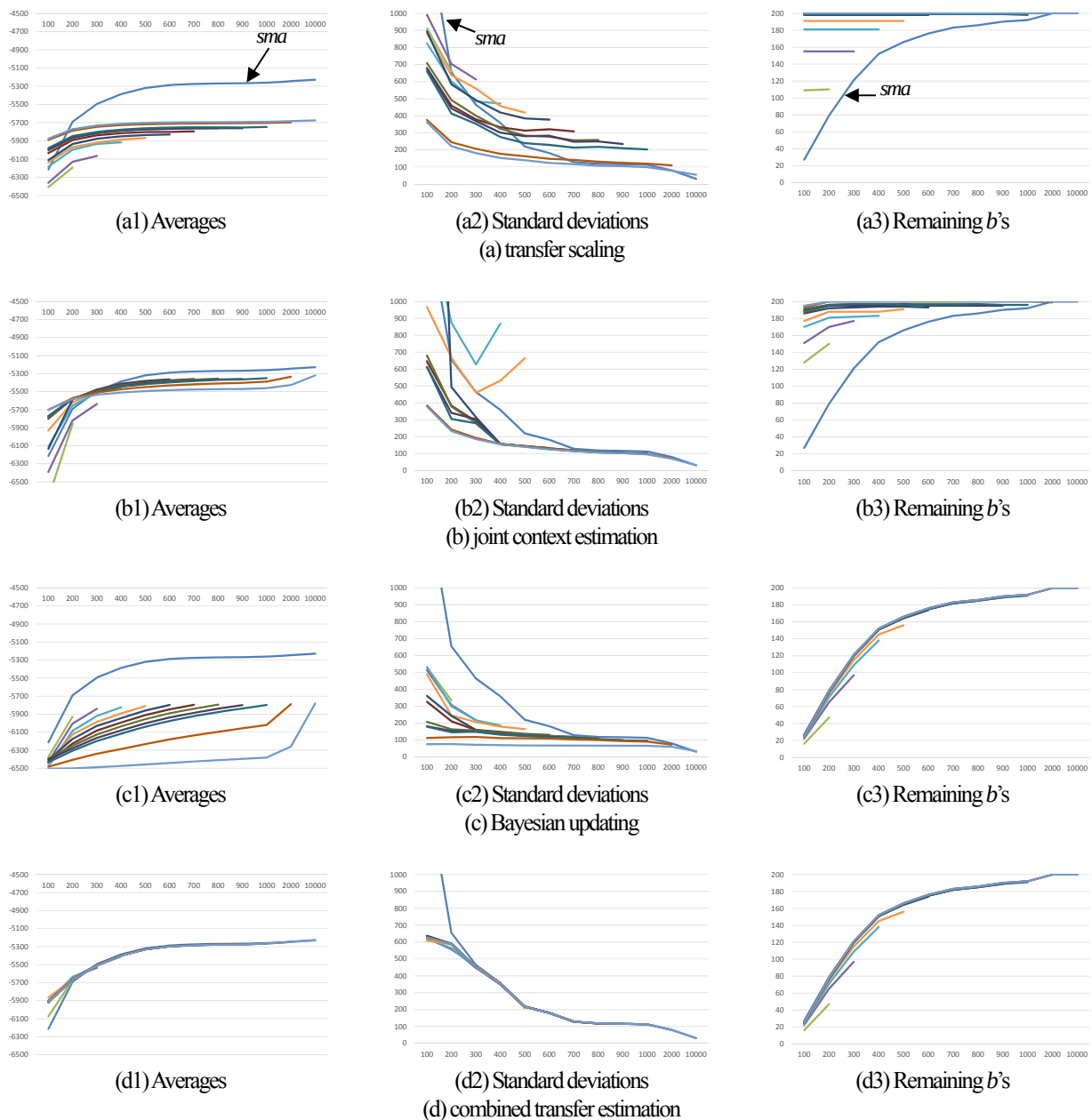
表一1の1971年のモデルを1981、1991年の10000サンプルのデータを用いて定数項とスケールパラメータを更新した場合、1981年のモデルを1991年の10000サンプルのデータを用いて更新した場合の、定数項とスケールパラメータの推定値、またそのモデルによる予測精度は三古⁹⁾を参照されたい。

それ以外のモデル更新法の場合に全10000サンプルを用いた推定結果と予測精度については紙幅の都合で省略する。

(2) 予測精度の特徴

本節ではブートストラップによる予測精度の特徴を示す。ここでは、 $(y_1, y_2) = (1971, 1981)$ の結果のみを紹介する。 $(y_1, y_2) = (1971, 1991), (1981, 1991)$ については紙幅の都合により省略し、必要な場合には本文で紹介する。

図一1に、2001年データへの対数尤度で表される予測精度 $L_m(y_1 = 1971, y_2 = 1981, n_1, n_2, b)$ の平均値と標準偏差を



Note: Horizontal axis, representing the number of observations from the recent 1981 (n_2), does not have an equal interval. Lines of both $n_1 = 10000$ and $m = sma$ appear in the entire 100–10000; both lines are in blue, but the darker and widely fluctuating one is the $m = sma$.

図—1 予測精度の特徴 ($y_1 = 1971$, $y_2 = 1981$)

示す。先にも述べたように、 m_1 や n_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 $b = 1, 2, \dots, 200$ のうち、推定結果および予測結果が得られたもののみについて算出した。また、問題の発生した場合を除いたあとに残った b の数についても示した。

図—1は12枚のパネルから構成され、それぞれのパネルは(a)~(d)および(1)~(3)の2つの記号の組み合わせで表現される。(a), (b), (c), (d)はそれぞれ、*cst*, *jnt*, *bay*, *com*のモデル更新法を示している。(smaについては別のパネルを設けなくて、全部のパネルに比較の基準として示している。)(1), (2), (3)はそれぞれ、 Lm の平均、

Lm の標準偏差、残った b の回数を示している。それぞれのパネルにおいて、横軸には新しい時点1981年のサンプル数 n_2 がとられていて、縦軸にはパネル(1), (2), (3)のそれぞれについて Lm の平均、 Lm の標準偏差、残った b の回数を示している。

それぞれのパネルにおいて、12本の線が12通りの m_1 (=100, 200, ..., 10000)について描かれている。これに加えて、 $m = sma$ の線も描かれている。この図には凡例がないので分かりづらいようにも思うが、 $n_1 \geq n_2$ が満たされる n_2 の範囲においてのみ描かれているという特徴を理解すれば、その読み取りは容易である。 $n_2 = 100$ まで描かれている線は1本しかなく、これは $n_1 = 100$ を表している。

$n_2 = 200$ まで描かれている線は1本しかなく、これは $m_1 = 200$ を表している。以下 m_1 が大きくなったときも同様である。ただし、注意が必要なのは $n_2 = 100 \sim 10000$ の範囲で描かれている線は2本存在することであり、1本は $m_1 = 10000$ であり、もう1本は $m = sma$ である。どちらも青系統の色で描かれているが、濃いほうの図の中での触れ幅の大きいほうが $m = sma$ である。

(a1), (b1), (c1), (d1)の総てのパネルにおいて、*cst*, *jnt*, *bay*, *com*, *sma*は概ね右肩上がりである。つまり、 m_1 が同じであれば、新しい時点からのサンプル数 n_2 が大きいほど、予測精度が平均してよくなることを示している。一方、古い時点のサンプル数 m_1 が大きい場合の線のほうが m_1 が小さい場合の線の上方に描かれているという特徴が*cst*において見られる。これは、古い時点のサンプル数が大きくなると定数項以外のパラメータの移転性が高まるためと考えられる。同様の傾向が m_1 が小さいときの*jnt*においても見られるが、反対の傾向が m_1 が大きいときの*jnt*においては見られた。Joint context estimationにおいて古い時点のサンプル数 m_1 が大きくなることは次の2通りの効果があると考えられる。1つ目に、サンプル数が増えてパラメータが精度良く推定されるということである。2つ目に、2つの時点でパラメータを同一にするという制約の下で古い時点のサンプルが増えると、古い時点のデータを説明するようにパラメータが推定されるが、古い時点のパラメータの移転性のほうが低いので、全体として移転性が低くなるということである。 m_1 が小さいときには前者の効果が卓越し、 m_1 が大きいときには後者の効果が卓越すると解釈できる。古い時点のサンプル数 m_1 が大きくなると*bay*の予測精度は悪くなった。古い時点のサンプル数が増えると古い時点のパラメータが精度良く推定されるが、移転性の低い古い時点のパラメータに重みをおくように更新するためであると考えられる。*com*については、 m_1 の違いはあまり影響を与えなかった。

ところで、*sma*の線が他のモデル更新法の線よりも上方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、古いデータで構築した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。なお、1971と1981年の結果では、*bay*は必ず*sma*の下に現れるという結果になっていたが、1981と1991年のデータを使った場合にはそうではない場合もあった。

(a2), (b2), (c2), (d2)の総てのパネルにおいて、*cst*, *jnt*, *bay*, *com*, *sma*は概ね右肩下がりであり、 m_1 が同じであれば、新しい時点のサンプル数 n_2 が大きいほど、予測精度のばらつきが小さくなる傾向にある。(ただし、パネル(b2), また、1971と1991年の(b2)と(c2)でも例外的な傾向が見られた。)このような傾向は200回のブートストラップの限界とも考えられる。一方、古い時点のサン

プル数 m_1 が大きい場合の線のほうが m_1 が小さい場合の線の下方に描かれているという特徴が*cst*において見られる。しかし、それ以外のモデル更新法の場合、はっきりした傾向は見られなかった。

(a3), (b3), (c3), (d3)の総てのパネルにおいて、*cst*, *jnt*, *bay*, *com*, *sma*は概ね右肩上がりであり、新しい時点のサンプル数が大きくなると、残ったブートストラップの回数も多いことを示している。一方、古い時点のサンプル数 m_1 が大きい場合の線のほうが m_1 が小さい場合の線の上方に描かれている。

このことから、*sma*と比較した他のモデル更新法の特徴について次のように整理できる。

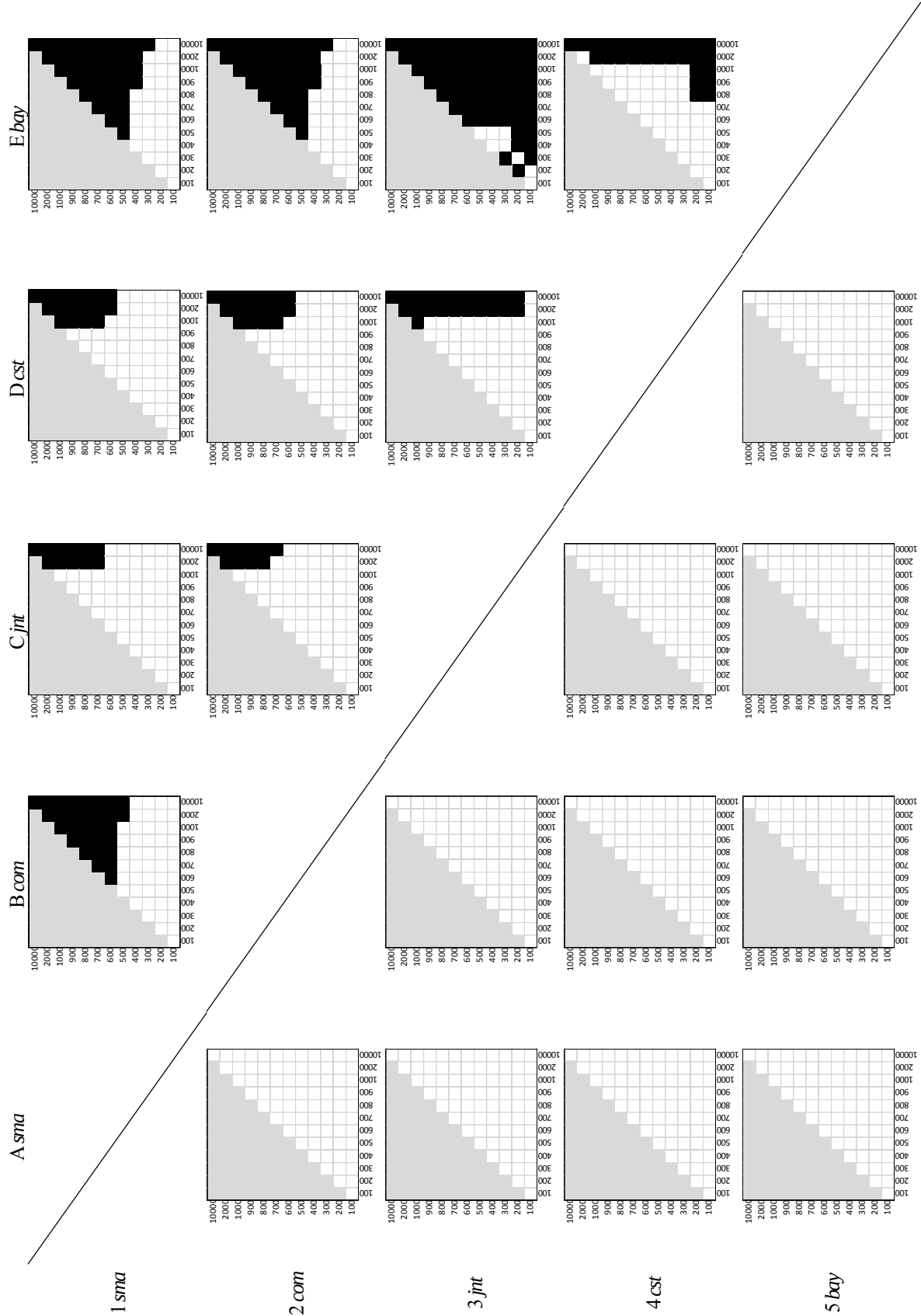
- *cst*と*jnt*は m_1 が大きく n_2 が小さいときに*sma*よりも予測精度が平均して良い(パネル(a1)と(b1))、予測精度のばらつきが小さい(パネル(a2)と(b2))、残った*b*の回数が大きい(パネル(a3)と(b3))。
- *bay*は n_2 が小さいときに*sma*よりも予測精度が平均して良い(今回は示していない $(y_1, y_2) = (1981, 1991)$ の場合のパネル(c1))、予測精度のばらつきが小さい(パネル(c2))、残った*b*の回数が小さい(パネル(c3))。なお、予測精度のばらつきと残った*b*の回数については3通りの (y_1, y_2) の全部において確認できた。

(3) 予測精度の差の検定

モデル更新法の優劣を検定した結果を図—2に示す。この検定は、先に定義した x_0 に基づいているが、それは比較の対象となる2つのモデル更新法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定義される。先ほどの図—1のパネル(a3), (b3), (c3), (d3)でそれぞれのモデル更新法を単独で行ったときの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。

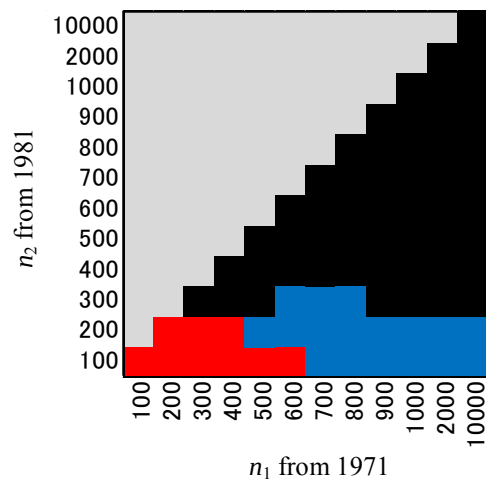
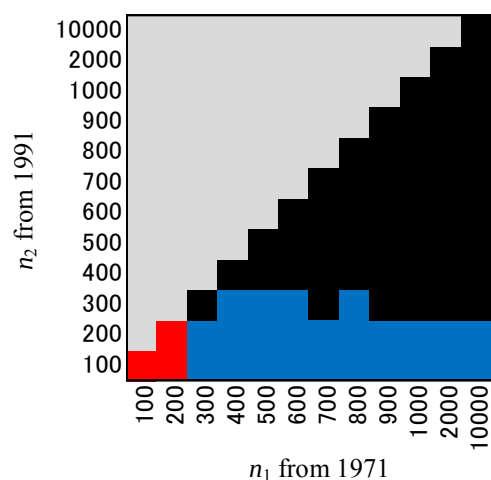
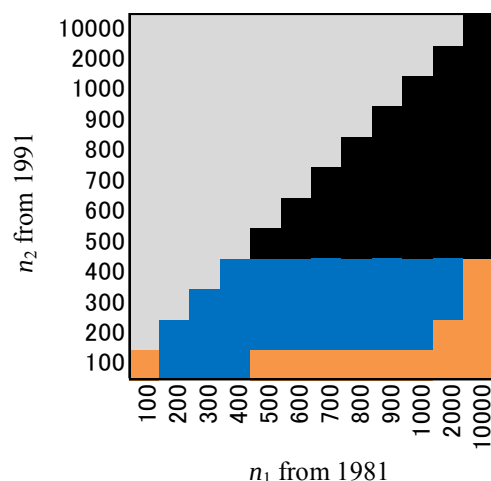
ここでは、 $(y_1, y_2) = (1971, 1981)$ の結果のみを紹介する。 $(y_1, y_2) = (1971, 1991), (1981, 1991)$ については紙幅の都合により省略し、必要な場合には本文で紹介する。

それぞれのパネルの中で、横軸は古い時点つまり1971年のサンプル数 m_1 、縦軸は新しい時点つまり1981年のサンプル数 n_2 を示している。各パネルの左上方にある灰色に塗りつぶされている領域は $m_1 < n_2$ (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 $z > 1.96$ となり、図の左側の1~5と書かれているモデル更新法が、図の上のA~Eと書かれているモデル更新法よりも有意に良い予測精度をもつことを示している。なお、見やすくするために、*sma*, *com*, *jnt*, *cst*, *bay*の順に並べ替えた。



Note: In each panel, horizontal and vertical axes represent n_1 from 1971 and n_2 from 1981, respectively. Grey cells indicate $n_1 < n_2$, which is out of interest of the present study; black cells indicate that models noted as 1 thought 5 are statistically significantly better than those noted as A through E.

図一2 モデル更新法の優劣に関する統計的検定 ($y_1 = 1971$, $y_2 = 1981$)

(a) $y_1=1971$ と $y_2=1981$ (b) $y_1=1971$ と $y_2=1991$ (c) $y_1=1981$ と $y_2=1991$

注：灰色は $n_1 < n_2$ であり、今回の興味の対象外。それ以外の領域は Lm の平均値が最大となるモデル更新法で着色している。黒は *sma*、青は *jnt*、橙は *cst*、赤は *com* であり、*bay* は該当するものがなかった。

図—3 最も予測精度の平均が良いモデル

今回の結果で興味深いのは、図—2の右上のパネルにしか黒色で塗りつぶされたセルが見当たらないことである。つまり、2つのモデル更新法を比較したとき、一方のモデル更新法がある n_1 と n_2 のときには有意に良くて、別の n_1 と n_2 のときには有意に悪いということは1回もなかったということである。このことから、統計的検定によれば、モデル更新法は最も優れているものから、*sma*, *com*, *jnt*, *cst*, *bay* という順になった。このことは、1971と1991年の場合と1981と1991年の場合についても当てはまった。

sma について、1行B列、1行C列、1行D列、1行E列のパネルで黒色に塗りつぶされているセルの合計が $(y_1, y_2) = (1971, 1991)$ の場合に最も多くて、その後は $(1971, 1981)$, $(1981, 1991)$ の順であった。中京都市圏における出勤交通行動は1971年から1981年に大きく変わり、その後、ややおだやかに1991年、2001年と推移したと考えられる。そのため、出勤交通行動における文脈は、違いの大きい順に $(1971, 1991)$, $(1971, 1981)$, $(1981, 1991)$ となり、文脈の違いの大きいときには古いサンプルを利用する価値は少ないと考えられる。

このことから、統計的検定から判断すると、古い時点と新しい時点のデータを両方用いるモデル更新法の利用価値はないということになる。もし、2つの時点のデータを用いるモデル更新法の利用価値があるとするならば、その価値は統計的検定以外に求めなければならない。ここで、図—3に古い時点と新しい時点のサンプル数の様々な組み合わせにおいて、どのモデル更新法が最も高い Lm の平均値をもたらしたかを示した。なお、ここでは (y_1, y_2) について $(1971, 1981)$ だけではなく、 $(1971, 1991)$, $(1981, 1991)$ の場合についても、それぞれパネル(a), (b), (c)に示す。これによると、*sma* 以外の方法が最も優れているのは次のような場合である。

- $(1971, 1981)$ と $(1971, 1991)$ の場合には、 n_1 と n_2 がともに小さい場合には *com* が最も良く、 n_1 は大きい n_2 が小さい場合には *jnt* が良かった。
- $(1981, 1991)$ の場合には、 n_1 は大きい n_2 が小さい場合には *cst* が良く、 n_1 がそれよりも若干小さく n_2 が若干大きい場合には *jnt* が良かった。

1981と1991年では文脈が似ているため、定数項以外のパラメータの移転性が比較的高いと考えられるので、*cst* の入る余地があったと考えられる。

以上を見ると、新しい時点である一定のサンプルがあれば *sma* が最も良いと考えられる。1971と1981年では新しい時点のサンプルが2000あれば古い時点のサンプルがいくらであっても *sma* が他のどの更新法よりも有意に良い(図—2の1行目)、700あれば古い時点のサンプル数によっては *sma* が他のどの更新法よりも有意に良い(図—2の1行目)、400サンプルあれば有意ではないが古い

時点のサンプル数に関係なく *sma* の予測精度が最も高い (図—3パネル(a))。なお、ここでの2000, 700, 400というサンプル数は、1971と1991年の場合には900, 800, 400であり、1981と1991年の場合には、10000, 2000, 500であった。

モデル更新法の採択の基準としてはまず、新しい時点でのサンプル数がどの程度あるかによる。十分に多ければ、新しい時点のサンプルのみを用いるのが良い。新しい時点のサンプル数が小さい場合には2時点のデータを同時に用いる意義がある。古い時点と新しい時点の文脈に差がないときには、古い時点のサンプルが多ければ *Transfer scaling* が、古い時点のサンプルが少なければ *Joint context estimation* が優れている。文脈に差があるときには、古い時点のサンプルが多ければ *Joint context estimation* が、古い時点のサンプルが少なければ *Combined transfer estimation* が優れている。

6. おわりに

本研究はさまざまなモデル更新法による予測精度の差異を検証した。検討したモデル更新法は *Transfer scaling*, *Joint context estimation*, *Bayesian updating*, *Combined transfer estimation*, *Small sample* である。このうち最初の4つ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の両方を用いるが、最後の1つは新しい時点の少数のデータしか用いない。本研究の目的はどのような場合に (具体的にはそれぞれの時点でのサンプル数がどのくらいの場合に)、どのモデル更新法が高い予測精度を持つかを分析し、モデル更新法選択の基準を示すことである。中京都市圏でのパーソントリップ調査データを用いることで、古い時点と新しい時点の組み合わせに関する3通り、それぞれの時点からのサンプル数の組み合わせに関する78通り、からなる合計234通りを検討した。

ブートストラップ法による予測精度の特徴は統計的な検定の結果ではないが、次のように整理される。

- 古い時点のサンプル数を固定した場合、新しい時点のサンプル数が大きいほど予測精度が平均して優れていた。これは5つのモデル更新法に共通して見られた。
- 新しい時点のサンプル数を固定した場合、古い時点のサンプル数が大きいほど予測精度が高いことが *Transfer scaling* の場合において見られた。同様の傾向が古い時点のサンプル数が少ない場合の *Joint context estimation* において見られた。古い時点のサンプル数が多い場合の *Joint context estimation* および *Bayesian updating* では逆の傾向が見られた。

ブートストラップ法による予測精度の統計的な検定の結果、次の知見が得られた。

- Small sample* は他の4つの更新法よりも統計的に有意に優れていることがあるが、その逆は決していない。
- 残された4つの更新法のうち *Combined transfer estimation* が他の3つの更新法よりも統計的に有意に優れていることがあるが、その逆は決していない。
- 残された3つの更新法のうち *Joint context estimation* が他の2つの更新法よりも統計的に有意に優れていることがあるが、その逆は決していない。
- 残された2つの更新法のうち *Transfer scaling* が統計的に有意に優れていることがあるがその逆は決していない。
- つまり、統計的な検定では、優れている順に、*Small sample*, *Combined transfer estimation*, *Joint context estimation*, *Transfer scaling*, *Bayesian updating* である。

古い時点のデータと新しい時点のデータを両方用いるモデル更新法は新しい時点のデータが少ない場合に意味がある。このことから検討したモデル更新法の選択基準は次の通りである。

- 新しい時点のサンプル数が一定以上あるときには *Small sample* を用いる。
- 新しい時点のサンプル数が一定以上ないときは次の2つに場合分けされる。古い時点と新しい時点の文脈に差がないときには、古い時点のサンプルが多ければ *Transfer scaling* を、古い時点のサンプルが少なければ *Joint context estimation* を用いる。文脈に差があるときには、古い時点のサンプルが多ければ *Joint context estimation* を、古い時点のサンプルが少なければ *Combined transfer estimation* を用いる。

なお、今回の事例は3つの選択肢からなる多項ロジットモデルというシンプルなモデル構造を持っていた。モデルの複雑性が与える影響についても検討する必要がある。今回は手元にあるデータを用いたため、最も新しいデータでも14年前に収集されたものである。新しいデータを用い、近年の *Peak car* のような現象を対象とした分析も今後の課題である。このような研究により、効率的な調査設計を行うことができれば、費用や時間の制約が多い近年の交通調査において貢献が大きいと考える。

謝辞： 本研究はJSPS科研費25380564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。

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*, 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) 三古展弘：交通需要予測におけるデータの新鮮さとサンプル数のトレードオフ，土木計画学研究・講演集，No. 50 (CD-ROM)，2014.
 - 8) Sanko, N.: Trade-off between data newness and number of observations for travel demand forecasting, *Compendium of Papers of the 94th Annual Meeting of the Transportation Research Board*, Washington D.C., U.S.A., Jan. 2015.
 - 9) 三古展弘：新しい小さいサンプルは古い大きいサンプルと同時に使うべきか：定数項の修正によるモデル更新の適用可能性，土木計画学研究・講演集，No. 51 (CD-ROM)，2015.
 - 10) Sanko, N.: Should small samples from recent time point be used with older data? Applicability of updating models by transfer scaling, to be presented at the hEART 2015.
 - 11) Badoe, D.A. and Miller, E.J.: Comparison of alternative methods for updating disaggregate logit mode choice models, *Transportation Research Record*, 1493, pp. 90–100, 1995.
 - 12) Karasmaa, N. and Pursula, M.: Empirical studies of transferability of Helsinki metropolitan area travel forecasting models, *Transportation Research Record*, 1607, pp. 38–44, 1997.
 - 13) Efron, B. and Tibshirani, R.J.: *An Introduction to the Bootstrap*. Chapman & Hall, London, 1993.

(2015. 7. 31 受付)

CRITERIA OF SELECTING MODEL UPDATING METHODS FOR BETTER TEMPORAL TRANSFERABILITY

Nobuhiro SANKO