

# 複数時点の断面データを用いた 交通需要予測精度の向上

三古 展弘<sup>1</sup>

<sup>1</sup>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

非集計モデルを構築するためのデータは、将来時点の文脈と似た文脈で入手されることが望ましいため、できるだけ新しいものを用いるべきだと考えられてきた。従って、複数時点でデータが得られていても、直近のデータのみを用いてモデルを構築することが多かった。本論文では、複数時点の断面データを効率的に用いて予測精度を向上させることを考えた。その結果、パラメータを時間の関数で表現することにより、複数時点のデータを用いたモデルのほうが直近のデータのみを用いたモデルよりも、予測精度が高いことが示された。

**Key Words :** *multiple cross-sectional data, predictive performance, transferability, mode choice model, parameter change*

## 1. はじめに

非集計モデルを需要予測に用いるためには、将来時点の文脈と似た文脈のデータを用いることが望ましいため、できるだけ新しいデータを用いるべきだと考えられてきた。従って、複数時点でデータが得られていても、直近のデータのみを用いてモデルを構築することが多かった。ここで、次のような疑問が生じる。

**需要予測のモデル構築には直近のデータだけを用い、古いデータを用いないことは妥当なのか。**

本研究では、データの種類として多時点断面データを考える。断面データはクロスセクション・データとも呼ばれ、パーソントリップ(PT)調査のデータなどがこれに該当する。PT調査は日本の3大都市圏では既に4回以上(30年以上にわたって)実施されており、貴重なデータであるが、直近のデータしか用いないのであれば、貴重な過去のデータを無駄にしていることになる。

この疑問に答えようとした既存研究にBadoe and Wadhawan<sup>1)</sup>がある。彼らの論文では、カナダの大トロント地域(The Greater Toronto Area)で1964年、1986年、1991年の3時点で得られた多時点断面データを用いて分析している。論文の中では、1991年の朝ピークの出勤交通手段選択行動を予測する際に、直近の1986年のデータのみを用いたモ

デルと、1964年と1986年の2つの時点のデータを用いたモデルによる予測精度を比較している。1964年と1986年のデータを同時に用いたモデルは、Ben-Akiva and Morikawa<sup>2)</sup>においてRP (Revealed Preference)データとSP (Stated Preference)データを同時に用いているのと同様の考えで、パラメータを2つの時点で共通としている。また、スケールパラメータや定数項は2つの時点で共通とした場合や、異なるとした場合などの複数の場合を検討している。予測精度の評価は2つの指標で行われている。1つ目は非集計ベースの評価指標で、1991年のデータへの対数尤度を計算している。その結果は非常に明瞭で、直近の1986年のデータのみを用いたモデルのほうが、2時点のデータを同時に用いたどのモデルよりも優れていた。2つ目は集計ベースの評価指標であり、MAE (Mean Absolute Error)を計算している。その結果は、2時点のデータを同時に用いたモデルのうちのいくつかは、直近の1986年のデータのみを用いたモデルよりも優れているというものであり、はっきりしたことは言えなかった。しかし、集計ベースの評価指標は、予測対象地域を彼らの基準で7分割し、分割された地域での手段ごとの予測シェアと実績シェアの差を計算し、それを合計するという形になっている。つまり、MAEは彼らの設定した地域分割に依存することになり、非集計ベースの指標よりも説得力において劣ると考えられる。

ここで、筆者が着目したのは、Badoe and Wadhawanの

仮定である「パラメータを2つの時点で共通」にしている点である。これは、言い換えると、2つの時点で文脈が全く同じであると仮定していることになる。Ben-Akiva and Morikawaでは、複数のデータソース利用の考え方として、一方のデータの短所を他方のデータの長所で補うということを挙げている。しかし、Badoe and Wadhawanの場合、1986年の文脈が1964年の文脈に比べて1991年の文脈に近く、1986年のデータの短所を1964年のデータの長所で補うという形にはなっていないと考える。

本研究の目的は、需要予測において直近のデータだけではなく古いデータも活用することで、予測精度が向上することを示すことである。筆者の定式化では、パラメータが時間の関数になっていると仮定する。この考えは筆者らの過去の研究において、交通手段選択モデルや自動車保有モデルのパラメータが時代とともに一貫して増加傾向や減少傾向にあるものが見られることに基づいている<sup>3)</sup>。たとえば、モータリゼーションの進行につれて、交通手段選択モデルにおける自動車の選択肢固有定数項が増加する、ということが挙げられる。

## 2. データ

データは、中京都市圏において1971年、1981年、1991年、2001年の4時点で得られた多時点断面データである、PTデータを用いる。構築されるモデルは出勤目的の交通手段選択モデルであり、考慮される選択肢は鉄道、バス、自動車の3種類である。モデルとしては3項選択のロジットモデルを考える。

## 3. 方法論

時点 $t$ における選択肢 $i$ に対する効用関数の確定項 $V_i^t$ を式(1)のように定式化する。

$$V_i^t = \mu^t \sum_k \beta_{ik}^t x_{ik}^t \quad (1)$$

ここに、 $\mu^t$ 、 $\beta_{ik}^t$ 、 $x_{ik}^t$ はそれぞれ、時点 $t$ におけるスケールパラメータ、時点 $t$ における選択肢 $i$ に対する $k$ 番目説明変数のパラメータ、時点 $t$ における選択肢 $i$ に対する $k$ 番目説明変数、である。

推定するモデルを以下に示す。1時点のデータのみを用いた「個別時点モデル」3種類、3時点のデータを用いた「3時点モデル」5種類である。なお、以下の説明で括弧内のI、Jで始まる記号は略称を表す。

- 個別時点モデル：1971年のデータのみ(I-71)、1981年のデータのみ(I-81)、1991年のデータのみ(I-91)を用いたモデル
- 3時点モデル

$$J-1: \beta_{ik}^t = \beta_{ik} \forall t, \quad \mu^t = 1 \forall t$$

$$J-2: \beta_{ik}^t = \beta_{ik} \forall t, \quad \mu^t \text{を推定} (\mu^{71} = 1)$$

$$J-3: \beta_{ik}^t = \beta_{ik} \forall t \quad (\text{ただし、選択肢固有定数項は時点ごとに推定}), \quad \mu^t \text{を推定} (\mu^{71} = 1 \text{とし、サービスレベル(LOS)と社会経済(SE)変数で別のスケールを仮定する})$$

$$J-4: \beta_{ik}^t = \beta_{ik} + (t - 1971)/10 \times \tilde{\beta}_{ik}, \quad \mu^t = 1 \forall t$$

$$J-5: \beta_{ik}^t = \beta_{ik} + (t - 1971)/10 \times \tilde{\beta}_{ik}, \quad \mu^t = 1 + (t - 1971)/10 \times \tilde{\mu}$$

ここに、 $\beta_{ik}$ は時点に関係なく一定な、選択肢 $i$ に対する $k$ 番目説明変数のパラメータ、 $\tilde{\beta}_{ik}$ は時点によって変化する、選択肢 $i$ に対する $k$ 番目説明変数のパラメータ、 $\tilde{\mu}$ は時点によって変化するスケールパラメータ。

J-1、J-2、J-3のモデルはBadoe and Wadhawanの考えに基づいたモデルであり、J-3は彼らの研究で非集計ベースの予測精度が最も良かったモデルである。J-2とJ-3による将来予測では、定数項やスケールは1991年の値を用いる。J-4とJ-5が筆者の考えに基づくモデルである。J-4とJ-5による将来予測では、まず2001年のパラメータを計算し、そのパラメータを予測に適用する。これらのモデルの予測精度は非集計ベース（2001年のデータに対する対数尤度）と式(2)で示される集計ベース(MAE)によって評価する。なお、MAEの計算にはBadoe and Wadhawanとは異なりセグメントの設定はしていない。

$$MAE = \sum_m \left| S_m^p - S_m^a \right| \quad (2)$$

ここに、 $m$ は交通手段を表し、 $S_m^p$ 、 $S_m^a$ はそれぞれ2001年での手段 $m$ の予測、実績のシェアを表す。

## 4. 結果

推定時間を短縮するため、個別時点モデルの推定にはランダムに抽出した10,000サンプルを用い、3時点モデルの推定には各時点から抽出した10,000サンプルの合計である30,000サンプルを用いた。予測対象時点である2001年のデータもランダムに10,000サンプル抽出し、そのデータに対する非集計レベル、集計レベルの予測精度を評価した。各モデルの予測精度を表-1に示す。参考のため、2001年のデータでモデルを構築した場合(I-01)も表中に示した。

まず、3つの個別時点モデルを見る。非集計、集計ベースのどちらでもI-91が最も予測精度がよく、続いてI-81、I-71となった。1時点のデータのみが利用可能な場合には、その時点に近いほうが良いことを示している。

表-1 予測精度

	対数尤度 <sup>a</sup>	MAE
I-71	-6,521.95	0.372
I-81	-5,225.15	0.195
I-91	-4,801.79	0.045
I-01	-4,716.28	0
J-1: $\beta$ 一定, $\mu$ 一定	-5,291.29	0.191
J-2: $\beta$ 一定, $\mu$ 推定	-5,000.68	0.108
J-3: $\beta$ 一定, $\mu$ LOS/SE別・定数別推定	-4,876.13	0.048
J-4: $\beta$ 関数, $\mu$ 一定	-4,789.61	0.042
J-5: $\beta$ 関数, $\mu$ 関数	-4,737.49	0.013

<sup>a</sup>: 2001年のデータに対する対数尤度.

次に, Badoe and Wadhawanの考えによる, 3時点のデータを使って, パラメータを時点間で共有するモデルを見る. J-1~J-3は1時点のデータを用いたモデルの中で最も良いI-91よりも非集計レベル, 集計レベルともに悪い結果になった. これは, 複数時点のデータがあっても最新の1時点のデータのみを用いた場合のほうが予測精度が高いということであり, Badoe and Wadhawanの知見と一致している.

さらに, 筆者の考えによる, 3時点のデータを使って, パラメータを時間の関数で表現するモデルを見る. J-4, J-5は1時点のデータを用いたモデルの中で最も良いI-91よりも非集計レベル, 集計レベルにおいて良い結果になった. これは, 複数時点のデータがある場合には, 最新の1時点のデータのみではなく, 古いデータを用いるほうが予測精度が高くなることを示す新しい知見である. さらに, モデルJ-5はモデルJ-4よりもよく, パラメータだけではなく, スケールも時間の関数として表現することが望ましいと言える.

最も予測精度の良かったJ-5の推定結果を表-2に示す. 表-2において, 時点に関係なく一定なパラメータは総て有意であった. 一方, 時点によって変化するパラメータは定数項, 所要時間, 男性ダミー, 名古屋発着ダミー, スケールについて有意であり, 20歳以上ダミー, 65歳以

表-2 J-5 ( $\beta$ ,  $\mu$ ともに時間の関数) モデルの推定結果

説明変数	$\beta_{ik}$		$\tilde{\beta}_{ik}$	
	推定値	t値	推定値	t値
定数項(B)	0.104	2.07	-0.508	-8.73
定数項(C)	-1.25	-11.06	0.760	6.57
所要時間[hr]	-0.732	-8.62	-0.882	-7.72
男性ダミー(R)	0.574	8.89	0.316	3.90
男性ダミー(C)	2.04	31.63	0.301	2.83
20歳以上ダミー(C)	0.875	8.33	0.0812	0.70
65歳以上ダミー(B)	1.84	9.14	-0.0705	-0.36
名古屋発着ダミー(C)	-1.13	-24.90	-0.942	-9.14
スケール	1	—	-0.141	-5.83
N	30,000			
L( $\beta$ )	-19,084.69			
L(0)	-25,940.98			
Adjusted rho-squared	0.264			

(注) 説明変数の括弧内のR, B, Cは, それぞれ鉄道, バス, 自動車の選択肢固有であり, 表記のないものは選択肢共通である.

上ダミーについては有意ではなかった. 定数項に関してパラメータの符号から判断すると, 他の要因が同じなら, バスは利用されないようになり, 自動車は利用されるようになることを示しており, モータリゼーションによる影響と考えられる. 所要時間に関しては, 旅行者がより時間に対して敏感になっていると解釈できる. 本研究では出勤交通を分析対象としているが, 日本では通勤費用は会社から補填されることが多いため, 費用の変数を説明変数に含んでいない. そのため, 時間価値の変化について考察することはできない. しかし, 旅行者が時間により敏感になっているということは, 時間価値が増加しているということと矛盾しない. Abrantes and Wardman<sup>4)</sup>がメタ分析によって, 時間価値はGDPが増えるにつれて増加するというを示しているが, 日本でもこの時期にGDPが増加していることを考えると矛盾しない知見である. 男性ダミーについては, パラメータの符号から判断すると, 他の要因が同じなら, 男性は鉄道や自動車を利用するようになることを示している. バスに定数項が含まれていないことから, 男性はバスを利用しなくなっているとも解釈できる. 20歳以上ダミーと65歳以上ダミーについては20年間で交通手段選択に与える影響が有意に変わらないことを示している. 名古屋発着ダミーは, 名古屋市を出発地または到着地にもつ場合には, 自動車を利用しなくなることを示している. 名古屋市内の道路混雑や公共交通ネットワーク整備のうち, 所要時間で表されている要因以外の効果と考えられる. スケールパラメータは20年間で減少しており, 20年間で誤差項の分散が大きくなっていることを示している.

この結果について, 移転性の観点から考察する. J-5のモデルにおいていくつかのパラメータが有意に変化することが明らかになった. このため, このようなパラメータを移転する際には注意を要することが明らかになり, どのパラメータが移転性を持つかの検討も必要である(定数項の場合の検討はSanko and Morikawa<sup>5)</sup>を参照).

ここで問題になるのは, パラメータの関数形である. 今回はパラメータの変化を線形関係で表したが, 長い期間に適用する場合には, パラメータの符号が逆転することがある. つまり, 所要時間のパラメータが正になったり, スケールパラメータが負になったりする場合が生ずる. そのため, このモデルを無限に長い期間に適用することはできない. しかしながら, 今回のような線形という単純な仮定を置いた場合でさえも, 予測精度の向上に寄与することができるということは, 本研究の大きな貢献であると言えるだろう.

## 5. おわりに

断面データは多くの都市圏で複数時点において得られ

ているため、これらのデータを無駄なく利用し、高い予測精度を得る方法を検討することを目的とした。複数時点のデータを同時に用いた場合、パラメータを時点間で共通とした場合では、直近の1時点のデータを使ったモデルよりも高い予測精度を得ることができなかった。本研究では、説明変数のパラメータとスケールパラメータを時間の関数として表現することで、将来時点のパラメータを予測し、そのパラメータで手段選択行動を予測することを考えた。その結果、直近の1時点のデータを使ったモデルよりも高い予測精度を得ることができた。これは、今回の分析で得られた新しい知見であり、需要予測の精緻化のために有益であると考えられる。今後の課題としては、今回仮定した関数形である線形関係の改良が挙げられる。

**謝辞**：本研究は科学研究費補助金（課題番号：22730334）からの支援を受けた。パラメータを時間の関数として表現することは三輪富生氏（名古屋大学）との議論の中で思いついたものである。また、この考えに関して、山本俊行氏（名古屋大学）、Stephane Hess氏（英国リーズ大学）およびAndrew Daly氏（同、RAND Europe）よりコメントをいただいた。ここに記して謝意を表する。

## 付録 個別時点モデルの推定結果

個別時点モデルの推定結果を表-3に示す。

### 参考文献

- 1) Badoe, D. A. and Wadhawan, B.: Jointly estimated cross-sectional mode choice models: Specification and forecast performance, *ASCE Journal of Transportation Engineering*, Vol. 128, No. 3, pp. 259–269, May/June 2002.
- 2) Ben-Akiva, M. and Morikawa, T.: Estimation of travel demand models from multiple data sources, in: Koshi, M. (Ed.), *Transportation and Traffic Theory (Proceedings of the 11th International Symposium on Transportation and Traffic Theory)*, pp. 461–476, Elsevier, New York, 1990.
- 3) 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.
- 4) Abrantes, P. A. L. and Wardman, M. R.: Meta-analysis of UK values of travel time: an update, *Transportation Research Part A*, Vol. 45, No. 1, pp. 1–17, 2011.
- 5) 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.

(2012.5.7 受付)

## IMPROVING TRAVEL DEMAND FORECAST PERFORMANCE USING CROSS-SECTIONAL DATA FROM MULTIPLE TIME POINTS

Nobuhiro SANKO

Forecasting is often based on data from the most recent time point, even when cross-sectional data is available from multiple time points. This is because forecasting models estimated in a similar context have higher transferability; and the context of the most recent time point is believed to have most similarity to that of a future time point. In this paper, the author aims at improving predictive performance of a model by utilising an older data set. In the author's formulation, parameters are assumed to be functions of time, and the model based on this formulation resulted in a better predictive performance.

表-3 個別時点モデルの推定結果

説明変数	1971 (I-71)		1981 (I-81)		1991 (I-91)		2001 (I-01)	
	推定値	t-値	推定値	t-値	推定値	t-値	推定値	t-値
定数項(B)	0.127	2.42	-0.392	-6.21	-0.638	-8.98	-1.03	-12.11
定数項(C)	-1.15	-9.84	-0.645	-4.65	0.301	1.96	0.560	2.23
所要時間[hr]	-0.606	-6.94	-1.81	-16.47	-1.59	-15.71	-2.60	-20.48
男性ダミー(R)	0.577	8.59	0.787	8.70	0.812	7.53	0.511	3.89
男性ダミー(C)	1.97	29.44	2.17	25.22	1.78	17.30	1.38	10.91
20歳以上ダミー(C)	0.900	8.28	0.764	5.78	0.776	5.18	0.511	2.06
65歳以上ダミー(B)	1.91	8.89	1.37	5.73	1.33	5.59	0.561	2.05
名古屋発着ダミー(C)	-1.12	-24.08	-1.77	-33.21	-2.18	-37.81	-2.21	-36.70
N	10,000		10,000		10,000		10,000	
L( $\beta$ )	-7,776.86		-5,985.02		-5,300.58		-4,716.28	
L(0)	-8,948.26		-8,593.88		-8,398.85		-8,159.63	
Adjusted rho-squared	0.130		0.303		0.368		0.421	

(注) 表-2の注と同じ。