

令和元年6月3日現在

機関番号：14501

研究種目：基盤研究(C) (一般)

研究期間：2016～2018

課題番号：16K03931

研究課題名(和文)複数時点の断面データを活用した交通調査費用の削減

研究課題名(英文) Reducing travel survey cost by utilising cross-sectional data from multiple time points

研究代表者

三古 展弘 (SANKO, nobuhiro)

神戸大学・経営学研究科・教授

研究者番号：00403220

交付決定額(研究期間全体)：(直接経費) 3,400,000円

研究成果の概要(和文)：非集計需要予測モデルを、断面データを用いて構築する場合、複数時点からデータが得られていても直近の1時点のデータのみを用いることが多く、貴重な過去のデータを無駄にしていた。本研究では、筆者が提案した複数時点のデータを用いて需要予測精度を向上させる方法を適用し、たとえ各時点から得られているデータは少なくても、複数時点のデータを用いたほうが、直近の1時点のデータを用いるよりも統計的に有意に優れた予測を行えることが示された。交通調査費用を追加することなく、予測精度も向上するという実務的な貢献も示された。

研究成果の学術的意義や社会的意義

交通調査には多大な費用を要する。本研究では、複数時点のデータを用いることで予測精度の向上が可能であり、それは調査済みの過去のデータを用いることで追加費用なく、また、各回の調査規模(調査費用)を削減することによっても実現可能であることを示した。また、大規模低頻度調査よりも小規模高頻度調査のほうが望ましい可能性を示した。

研究成果の概要(英文)：When estimating disaggregate demand models for forecasting, only the most recent data is utilised even when cross-sectional data from multiple time points are available. This is not a good use of the data. This study applied a method, proposed by the author, which improves demand forecasts by utilising data from multiple time points. The study demonstrated that using data from multiple time points, where the number of observations from each time point is smaller, produced statistically significantly better forecasts than using data from the most recent time point. Practical contribution is to improve forecasting performance without any additional cost for the survey.

研究分野：交通 需要予測 意思決定 選択モデル 商学 消費者行動 土木計画学

キーワード：消費者行動 交通行動 需要予測 繰り返し断面データ 非集計モデル サンプル数 調査頻度 モデルの更新

様式 C - 19、F - 19 - 1、Z - 19、CK - 19 (共通)

1. 研究開始当初の背景

非集計需要予測モデルを、断面データを用いて構築する場合、複数時点からデータが得られていても直近の1時点のデータのみを用いることが多く、貴重な過去のデータを無駄にしていた。筆者はこれまで、直近のデータだけではなく過去のデータも同時に用いることで、需要予測の精度が向上することを示してきた。

本研究では、これを以下の観点から発展させる。複数時点のデータを同時に用いる方法により、既に調査済みのデータを利用することや各回の調査規模を削減することによって、調査費用を追加することなく、予測精度の改善が可能であることを示す。また、その予測精度の改善が統計的に有意なものであるかも検討する。

2. 研究の目的

本研究の目的は、直近の1時点のデータを用いるよりも、複数時点のデータを用いることで統計的に有意に良い予測精度が得られることを示すことである。特に、以下の2点に着目し、調査費用を増やすことなく予測精度を改善可能であることを示し、今後の調査方法に関する知見を得ることを目的とする。

具体的には次の研究の問いに答えることを目的とする。

- (1) 直近の1時点のデータを用いた場合、直近の1時点を含む2時点のデータを用いた場合、前者のサンプルがいくらのとき、後方で追加する過去の時点のサンプルをいく以上にすれば、前者のモデルよりも統計的に有意に優れた予測精度をもたらすか。ここでは、現在のデータに加えて調査済みの過去のデータを用いることで、追加の調査費用なく統計的に有意に良い予測が行える場合について分析することを目的としている。
- (2) 直近の1時点のデータを用いた場合、直近の1時点を含む2時点のデータを用いた場合、直近の1時点を含む3時点のデータを用いた場合、直近の1時点のサンプルがいくらのとき、各時点のサンプルをいく以上にすれば、複数時点のデータを用いたモデルのほうが直近の1時点のデータを用いたモデルよりも統計的に有意に優れた予測精度をもたらすか。ここでは、直近の1時点のデータしか用いないのであれば、毎回同規模の調査を行う必要があるのに対し、複数時点のデータを用いることで各回の調査規模を削減つまり各回の調査費用を削減しても統計的に有意に良い予測が行える場合について分析することを目的としている。

3. 研究の方法

下の(1)と(2)では中京都市圏のパーソントリップ調査データ(1971, 1981, 1991, 2001年)を使用し、出勤時の交通手段選択行動を分析対象とする。構築するモデルは多項ロジットモデルである。モデル構築に用いるデータは1971, 1981, 1991年のものであり、予測対象年は2001年とする。複数時点のデータを用いるモデルでは、筆者の提案した、パラメータを1人当たりGDPの関数で表現している(Sanko, 2018)。なお、予測精度が統計的に有意に優れているかどうかを示すため、ブートストラップ法を用いる。

(1) 直近の1時点のデータの総てを含んだ上で過去のデータを用いる場合(図1を参照)

古い時点と新しい時点をそれぞれ y_1, y_2 とし、 y_1, y_2 からそれぞれ n_1, n_2 サンプル得られているものとする。 y_1, y_2 は1971, 1981, 1991年から選ぶ3通りの組み合わせ、 n_1, n_2 は100, 200, ..., 10000の12通りから選ぶ12×12通りを検討する。直近の1時点のモデルでは n_2 サンプル用いるのに対し、2時点のモデルでは $n_1 + n_2$ サンプル用いる。

なお、 $n_1 \geq n_2$ だけではなく、 $n_1 < n_2$ も検討することとする。

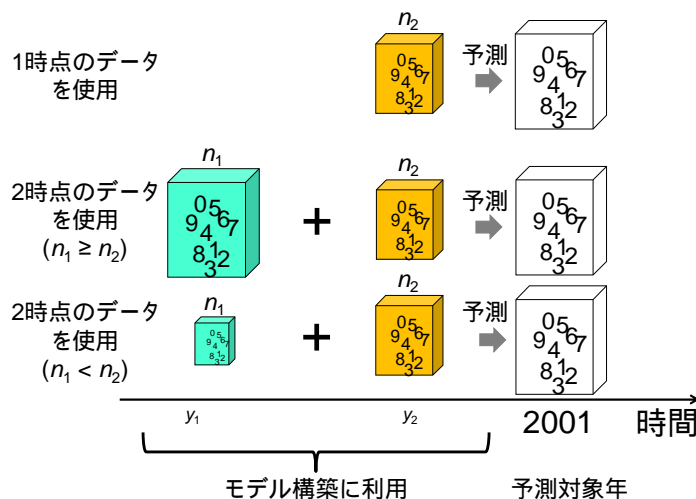


図1 分析(1)のイメージ

(2) 直近の1時点のデータの一部を含んだ上で過去のデータを用いる場合(図2を参照)

y_1, y_2, y_3 を 1971, 1981, 1991 年とする。1 時点のデータを用いる場合では、 y_3 から n_L サンプル、2 時点のデータを用いる場合では、 y_1 と y_3 または y_2 と y_3 の 2 時点から n_M サンプルずつ合計 $2 \times n_M$ サンプル(注: 図2では y_1 と y_3 の場合を例示)、3 時点のデータを用いる場合では、 y_1, y_2, y_3 の 3 時点から n_S サンプルずつ合計 $3 \times n_S$ サンプル得られている場合を考える。 n_S, n_M, n_L に関しては、100, 200, ..., 10000 の 23 通りから、 $n_S \leq n_M \leq n_L$ を満たすように選ばれるものとする。この不等式が意味するのは、直近の1時点のデータのみを用いるときに直近の時点から得られている n_L サンプルは、複数時点のデータを用いるときに直近の時点から得られているサンプルよりも小さくないということである。

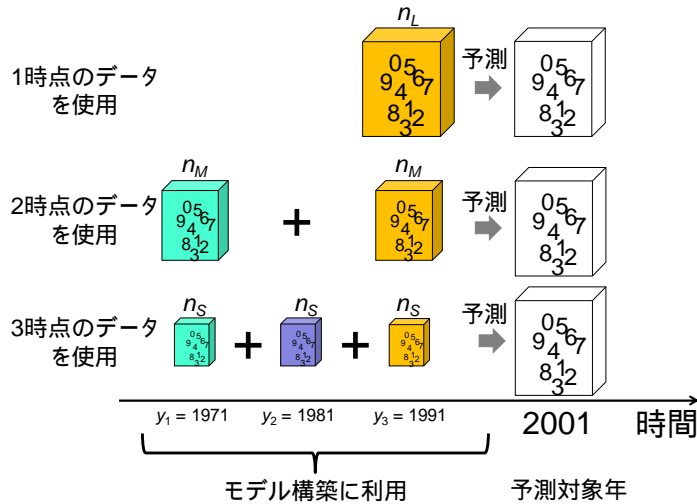


図2 分析(2)のイメージ

(3) 年齢・時代・世代の効果の分離

ここでの使用データは(1)と(2)の場合と同じであるが、1971, 1981, 1991, 2001 年のデータを総てモデルの構築に用いている。

長期間にわたる個人の行動を予測する際に重要な概念に、年齢、時代、世代がある。個人の交通行動の文脈においてこれらの概念の影響を例示すると、年齢は免許の保有など、時代は経済状況や制度など、世代は生まれた時点の経済状況や何歳のときにモータリゼーションを経験したかなど、がある。このうち、年齢や世代の効果は時代を経ても引き継がれる可能性がある。つまり、これらの効果を把握することによって、同一年齢や同一世代の個人の数年後や数十年後の行動の予測精度が向上することが期待できる。これらの効果を特定するには、複数時点のデータを用いることが有効であると考えられ、これによって予測精度が向上するのであれば、各時点での調査規模(調査費用)を削減しても、同程度の予測精度を確保できる可能性がある。しかし、年齢、時代、世代の3変数には線形従属の関係があり、その効果の分離は難しい。そこで、これまでの研究蓄積に基づき、時代の効果を1人当たりGDPで表現することでその分離を試みた。なお、この分析は当初の研究計画には含まれていなかったが、年齢、時代、世代の効果の分離に、本研究の知見を活用できるのではないかと考え、優先して行ったものである。

出勤時の交通手段選択モデル(多項ロジットモデル)を構築し、個人 n の選択肢 m に対する効用の確定項を V_n^m とする。このとき、選択肢固有定数項を年齢、時代、世代の効果に分離することを考える。年齢を i 、時代を j 、世代を k とすると、式(1)のような定式化が可能である。

$$V_n^m = \alpha^m + \alpha_{ai}^m \delta_{ni} + \alpha_{pj}^m \delta_{nj} + \alpha_{ck}^m \delta_{nk} + \beta^m \mathbf{x}_n^m \quad (1)$$

ここに、選択肢 m に対する固有定数項のうち、 α^m は全個人に共通するものであるが、 $\alpha_{ai}^m, \alpha_{pj}^m, \alpha_{ck}^m$ はそれぞれ年齢 i 、時代 j 、世代 k に固有である。 $\delta_{ni}, \delta_{nj}, \delta_{nk}$ はそれぞれ個人 n が年齢 i 、時代 j 、世代 k に該当するとき 1、そうではないとき 0 となる変数である。 \mathbf{x}_n^m は個人 n の選択肢 m に対する説明変数ベクトルであり、 β^m は対応する未知パラメータベクトルである。

一方、式(2)では、時代を表す変数に経済状況を表す1人当たりGDPを用いている。

$$V_n^m = \alpha^m + \alpha_{ai}^m \delta_{ni} + \alpha_p^m gdp_j + \alpha_{ck}^m \delta_{nk} + \beta^m \mathbf{x}_n^m \quad (2)$$

ここに、 gdp_j は時代 j における1人当たりGDP、 α_p^m は対応する未知パラメータである。

なお、識別のため、(1)、(2)のモデルとも、パラメータには適宜制約を加えるものとする。

4. 研究成果

(1) 直近の1時点のデータの総てを含んだ上で過去のデータを用いる場合

図1で示した直近の1時点のデータを用いる場合よりも2時点のデータを用いた場合には、 $n_1 \geq n_2$ だけではなく $n_1 < n_2$ の場合にも、統計的に有意に予測精度が優れている場合があった。これは、モデル更新法の観点から興味深い。モデルの更新は、過去の時点からは多数のデータが得られているが、直近の時点からは少数のデータしか得られていない場合（あるいは独自で追加調査を行うために少数のデータしか得られない場合）過去の多数のデータで構築したモデルを直近の少数のデータで更新する方法である。つまり、 $n_1 \geq n_2$ を対象としていた。しかし、筆者の過去の分析によれば、これまで提案されてきたモデルの更新法では、 $n_1 + n_2$ サンプルを用いてモデルを更新しても、直近の1時点の n_2 サンプルを用いて構築したモデルよりも、統計的に有意に優れた予測を行うことができなかつた。しかし、今回の分析においては、2 時点のデータを用いることで、直近の1時点のデータのみを用いるよりも統計的に有意に良い予測精度が得られる場合があることを示し、新しいモデル更新法として提案することができた。また、 $n_1 < n_2$ はこれまでモデルの更新法が対象としていない文脈であった。しかし、今回の分析では、このような場合においても2 時点のデータを用いることで直近の1 時点のデータのみを用いるよりも統計的に有意に良い予測精度が得られる場合があることを示し、モデルの更新法の適用範囲を拡大した。 $n_1 \geq n_2$, $n_1 < n_2$ のいずれの分析においても、 y_2 の時点では、 y_1 の時点のデータは収集済みであり、追加の調査費用は発生しない。

(2) 直近の1時点のデータの一部を含んだ上で過去のデータを用いる場合

結果の一例として、1 時点の1991年、2 時点の1981と1991年、1971と1991年のデータを用いて2001年の行動を予測した場合を紹介する。1 時点の10000 サンプルを用いるよりも、1981と1991年から各8000(合計16000) サンプルを用いるほうが、また1971と1991年から各4500(合計9000) サンプルを用いるほうが、統計的に有意に良い予測精度が得られた。つまり、1 時点の場合に比べて各時点の調査規模を従来の45~80%にしてもなお、2 時点のデータを用いたほうが統計的に有意に良い予測精度が得られることになり、各回の調査規模を削減できる。また、直近の1 時点のデータに加えて過去のデータを用いる場合には、その過去の時点はより古い時点のほうが望ましい可能性があることが示唆された。

また、1 時点の1991年、2 時点の1971と1991年、3 時点の1971、1981、1991年のデータを用いて2001年の行動を予測した場合を紹介する。1 時点の10000 サンプルを用いるよりも3 時点の各2500(合計7500) サンプルを用いるほうが、統計的に有意に良い予測精度が得られた。また、2 時点の各10000(合計20000) サンプルよりも3 時点の各8000(合計24000) サンプルを用いるほうが、統計的に有意に良い予測精度が得られた。つまり、1 時点の場合に比べて各時点の調査規模を従来の25%にしてもなお、また2 時点の場合に比べて各時点の調査規模を80%にしてもなお、3 時点のデータを用いたほうが統計的に有意に良い予測精度が得られることになり、各回の調査規模を削減できる。

以上をあわせると、大規模調査を低頻度で行うよりも、小規模調査を高頻度で行うことが望ましい可能性が示唆される。

(3) 年齢・時代・世代の効果の分離

式(1)、(2)で示したモデルに加えて、年齢、時代、世代のうち1つの効果のみ、2つの効果のみを導入したモデルも推定した。その結果、式(1)のモデルから得られた年齢、時代、世代の効果は、1つの効果のみ、2つの効果のみを導入したモデルから得られたそれらの効果から大きく乖離する結果となった。一方、式(2)のモデルではそのような乖離は見られなかつた。年齢、時代、世代の効果の分離や、調査費用削減へ向けては今後も分析を継続する必要があるが、そのための予備的な結果を得ることができた。

< 引用文献 >

Sanko, N. Travel demand forecasts improved by using cross-sectional data from multiple time points: enhancing their quality by linkage to gross domestic product, *Transportation*, Volume 45, Issue 3, pp. 905–918, 2018.

5. 主な発表論文等

[雑誌論文](計4件)

三古 展弘, Travel demand forecasts improved by using cross-sectional data from multiple time points: enhancing their quality by linkage to gross domestic product, *Transportation*, Volume 45, Issue 3, pp. 905–918, 2018. (査読有)

DOI: 10.1007/s11116-016-9755-x

三古 展弘, 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, 土木学会論文集 D3 (土木計画学), Vol. 74, No. 1, pp. 21–34, 2018. (査読有)

DOI: <https://doi.org/10.2208/jscejpm.74.21>

三古 展弘, Temporal transferability: trade-off between data newness and the number of observations for forecasting travel demand, *Transportation*, Volume 44, Issue 6, pp. 1403–1420,

2017. (査読有)

DOI: 10.1007/s11116-016-9707-5

三古 展弘, Factors affecting temporal changes in mode choice model parameters, *Transportation Planning and Technology*, Volume 39, Issue 7, pp. 641–652, 2016. (査読有)

DOI: 10.1080/03081060.2016.1204088

〔学会発表〕(計8件)

三古 展弘, Age-period-cohort analysis for alternative-specific constants in commuting mode choice models, presented at the 15th International Conference on Travel Behaviour Research, Santa Barbara, California, USA, July 2018.

三古 展弘, 年齢・時代・世代に着目した出勤交通手段分担率の集計分析, 第56回土木計画学研究発表会, 2017年11月, 岩手大学.

三古 展弘, Trade-off between survey frequency and sample size for forecasting travel demand: a preliminary study using cross-sectional data from three time points, presented at the 11th International Conference on Transport Survey Methods, Estérel, Quebec, Canada, September 2017.

三古 展弘, 交通需要予測における調査頻度とサンプル数のトレードオフ: 3時点の繰り返し断面データを用いた基礎分析, 第55回土木計画学研究発表会, 2017年6月, 愛媛大学.

三古 展弘, Comparing travel demand forecasts between models with larger data from single time point and models with smaller data from two time points, presented at the Fifth International Choice Modelling Conference, Cape Town, South Africa, April 2017.

三古 展弘, A novel model updating method: updating function model with gross domestic product per capita, presented at the 96th Annual Meeting of the Transportation Research Board, Washington D.C., USA, January 2017.

三古 展弘, 1時点大規模データと2時点小規模データによる交通需要予測精度比較, 第54回土木計画学研究発表会, 2016年11月, 長崎大学.

三古 展弘, パーソントリップ調査に関する小論: パラメータの時点変化を考慮したモデル更新法の視点, 第53回土木計画学研究発表会, 2016年5月, 北海道大学.

〔その他〕

研究計画調書でも示したように, 本研究課題の成果を様々な形で社会へ発信した. 特に講義の中では, 学生に最新の研究成果として紹介した. 学部向け講義では, H28年度, 29年度の「市場システム特殊講義(交通計画)」, H30年度の「交通論」において紹介をした. 専門職大学院(MBA)向け講義では, H28年度の「現代経営学応用研究(公益企業論)」, H29年度, 30年度の「需要予測と意思決定」において紹介した. 特に, 「市場システム特殊講義(交通計画)」と「需要予測と意思決定」は筆者が新規開設した科目であり, 需要予測にも重点を置いた講義になっている.

また, 本研究課題の成果を2017年にケープタウン(南アフリカ)で開催された第5回国際選択モデル会議(The Fifth International Choice Modelling Conference)で発表した, そのときに, 次回2019年の第6回会議は神戸で開催されることが決定した. 筆者も主催者の1人として参画している.

6. 研究組織

(1) 研究分担者

研究分担者氏名:

ローマ字氏名:

所属研究機関名:

部局名:

職名:

研究者番号(8桁):

(2) 研究協力者

研究協力者氏名:

ローマ字氏名:

科研費による研究は, 研究者の自覚と責任において実施するものです. そのため, 研究の実施や研究成果の公表等については, 国の要請等に基づくものではなく, その研究成果に関する見解や責任は, 研究者個人に帰属されます.