

家計調査自動車購入額調査法の改善のためのエージェント・ベース・シミュレーション

大貫裕二¹

1. 概要

総務省統計局が実施する家計調査は月次の家計消費動向を示す統計として注目されている。一方、「自動車購入額」をはじめとする単価が大きく購入頻度の低い消費項目は結果が安定しないということが知られている[宇南山 2011:pp. 9-10]。月次公表される二人以上の世帯でも全国約3000万世帯のわずか0.03%のサンプル数に過ぎない。サンプル数の増加は経費（地方公共団体委託費で約10億円[総務省 2010:p. 366]）の面から困難である。そこで、調査サンプル数は現状のまま、過去の購入実績について調査を追加することで精度を高めることが考えられる。本研究ではその可能性を探るため、エージェント・ベース・シミュレーションを用いた基礎的な検討を行った。

2. エージェント・ベース・シミュレーションを用いる意義

統計理論では、標本誤差が統計の精度を示す主要な指標である。仮に調査サンプルの選定がランダムに行われているとすれば、調査手法の変更による調査精度の向上の程度を推定できる。一方、実際の統計調査では、調査を低コストで進めるため厳密な意味ではランダムな調査サンプルの選定を行っていない。家計調査の二人以上世帯については、層化3段抽出により、全国168の市町村（東京特別区を1区域とする）から約8000の調査世帯を選定し、6か月連続して調査を行う。調査世帯は毎月6分の1ずつ入れ替わる。このような複雑な仕組みの下で調査精度の変化を論じるには、実際の統計調査の手法をモデル化してシミュレーションすることが容易であり、かつモデルの持つ限界を明確に意識できる。本研究ではエージェントは購入の意思決定については、購入履歴や環境の影響を受けないとする（エージェント・ベースとする必要は必ずしもない）一方、統計調査への回答について、購入時から回答時までの時間経過による記憶誤りを、月ごとのランダム・ウォーク・モデルにより生成した（エージェント・ベースとすることが必要）。

3. モデル設定と方法

(1) わが国の二人以上世帯の世帯数を次の通り推計した。

総務省統計局「住民基本台帳人口移動報告」の年次データの総世帯数を利用する。

総世帯に占める二人以上世帯の割合は、5年ごとに行われる国勢調査結果を2005年まで線形補間した。

各年、これらの世帯数分の連番を世帯番号とする。年ごとに同一世帯の番号は変化する。

平成17年(2005年)国勢調査の全国都道府県別世帯数と世帯人数が一人の世帯数から「二人以上世帯」の世帯数を求め、世帯番号を各市町村ごとに振り分ける。例えば、北海道の札幌市は世帯番号1から始まり、沖縄県の与那国村は全国の最後の世帯番号で終わる。

(2) 家計調査における調査世帯の選定方法に準じて調査対象世帯を選定した。

家計調査における調査世帯の選定は、層化三段抽出により行われる。第一層は調査市町村の選定であり、47都道府県の都道府県庁所在都市及び政令指定都市はそれぞれ1層とし、それ以外の市町村は規模別に、大都市・中都市・小都市A・小都市B及び町村に分類され、地方ブロックごとに属性ごとに層化された団体の中から抽出される。[総務省 掲載年不詳]

本研究においては、平成20年度標本設計における抽出市町村がシミュレーション実施全期間において対

¹ 筑波大学大学院システム情報工学研究科社会システム・マネジメント専攻 onuki@sk.tsukuba.ac.jp

象であったとした。実際には小規模な市町村などで調査市町村の選定替えが行われる。

平成 20 年度標本設計では、平成 17 年国勢調査結果を平成 20 年(2003 年)に補正したデータに基づき、調査市町村ごとの二人以上世帯数の標本数の設定が行われる。1 人の調査員が 2 つの調査地区を受け持ち、それぞれの調査地区で 6 世帯ずつの二人以上世帯に家計簿の記入を依頼するため、一つの調査市町村の調査世帯数は 12 の倍数として設定される。調査地区が代表する対象地域の世帯数との比率を一定に調整するため、各調査市町村ごとに「調整率」が設定されている。「調整率」は世帯数に比べて調査世帯数が多いほど小さく、調査世帯数が少ないほど大きい。

本研究においては、調査市町村ごとの標本数は平成 20 年度標本設計における標本数で固定とした。世帯数の経年変化については、各調査市町村の二人以上世帯数は全国と同一の伸び率で推移するとした。

1 つの調査地区は 12 ヶ月間連続して調査対象とし、その後別の調査地区に移動する。多くの場合近隣の調査地区が選定される。本研究においては、調査地区の設定を次のように行った。調査市町村の調査対象世帯数を 6 で除した数を、調査市町村のブロック数とする。ブロック数は調査市町村で同時に調査が実施されている調査地区の数に相当する。調査市町村の二人以上世帯数をブロック数で除すことで、ブロックあたりの二人以上世帯数が求められる。この世帯数を 1 調査地区あたり 60 世帯として除しブロックごとの調査地区数を求め連番を付す。各ブロックから 1 つずつの調査地区をランダムに選定する。12 ヶ月ごとの調査地区の選定換えの際には、連番順に選定する。

第三層は、調査地区からの調査世帯の選定である。家計調査では、調査地区内の世帯名簿を作成し、二人以上世帯に連番を振って、一つの調査地区から 6 世帯を選定する。一つ目の世帯のみ乱数(起番号)により決定し、残る 5 世帯は、連番の総数を 6 で除した数を順次加算し四捨五入することにより地域的な偏り無く選定している。同一の世帯を 6 ヶ月間継続調査した後で、同じ世帯名簿から異なる世帯を抽出してさらに 6 世帯を選定する。

本研究においては、調査地区内の世帯数を 60 で一定としているため、調査地区選定ごとに起番号を乱数で選定し、それに 10 ずつ順次加え 60 で除した余りを第一回目の 6 ヶ月間の選定世帯とした。第二回目の起番号は第一回目の起番号に 1 を加えた番号とした。

(3) 乗用車購入世帯の選定

1993 年 4 月～2010 年 3 月の国内における二人以上世帯による乗用車購入世帯数は、自動車販売連合会の普通乗用車・小型乗用車・軽乗用車の国内新車販売登録台数とした。

車種別に(1)で求めた年ごとの世帯数までの乱数を購入台数分発生することにより、購入世帯の番号を特定する。購入は全国で無作為に行われるものとし、地方や地域ごとの差は無いものとした。

(4) 思い出し記入方式における記憶誤り

正しい自動車購入月と記憶している自動車購入月の変数を各エージェントが持つ。記憶している自動車購入月は、自動車購入時点では正しい自動車購入月と合致しており、確率 $p(t)$ で 1 か月前もしくは 1 か月後に移動する。購入月の調査でも記憶違いは生じると設定した。確率 $p(t)$ は自動車購入月に初期値 $p_0(n)$ で開始し、1 か月ごとに定数 $i\%$ ずつ上昇する。ただし $p(t)$ が 50%に達した時点で $p(t)=50\%$ の定数とする。エージェントごとに記憶が異なるモデルとするため $p_0(n)$ は 0 から定数 p_0 までの均一分布とする。このモデルにより各エージェントの記憶する自動車購入月は正しい自動車購入月からランダム・ウォークにより

乖離していく。一度誤った記憶になったあと、正しい購入月に戻るケースもある。初期値はエージェントによって異なるが、購入からの時間経過とともに乖離は進みやすくなる。

(5) シミュレーションに利用するツール

シミュレーションの実施は、SOARS 4.03 で行い、結果の分析はAADL 言語 (Version 1.16 α) で行った。

[SOARS Project, エージェントベース社会システム科学研究センター]

(6) 試行回数及び評価の方法

自動車購入世帯の選定は1回のみ行い、この同一データについて、調査世帯を100通り選定して、調査結果から得られる推定乗用車購入数の分布を設定した購入数と比較する。

4. 結果

(1) 従来の調査月の購入のみによる推定と、過去の購入の正確な報告による推定結果

図1に過去1年の購入情報追加の効果を、図2に情報追加による推定値の平均と標準偏差の変化を示す。

過去の購入に関して正確な報告がなされる場合、情報の追加とともに標準偏差が縮小していく一方、平均値に大きな変化は無い。

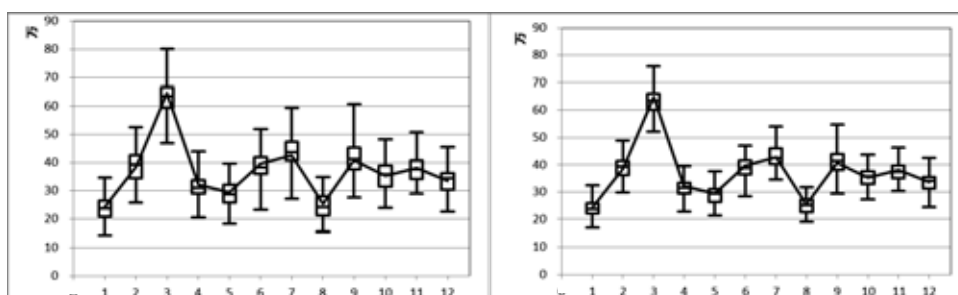


図1 左図は調査月の購入数のみによる推定結果、右図は過去1年間の購入数を追加した推定結果
縦軸は全国の1か月自動車購入台数、横軸は1995年の1-12月。線グラフは設定した真の購入数。
箱髭図は100回のサンプル選定による最大値、第3四分位値、中央値、第1四分位値、最小値を示す。

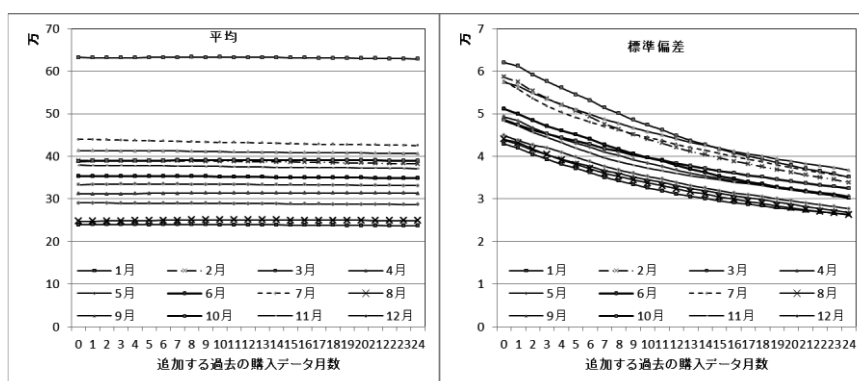


図2 過去の購入情報の追加月数と推定結果の平均 (左) と標準偏差 (右) 1995年1-12月結果

(2) 過去の購入の報告に記憶違いを導入した推定結果

$p_0=10\%$, $i=1\%$ の場合について図3に過去1年の購入情報追加の効果を、図4に情報追加による推定値の平均と標準偏差の変化を示す。正確な報告がなされる場合との最大の相違は、過去の情報の追加に従って、推定結果の平均値が購入数の多い月では減少し、購入数の少ない月では増加する点である。購入月の記憶

のずれにより季節パターンの平準化が生ずるためである。

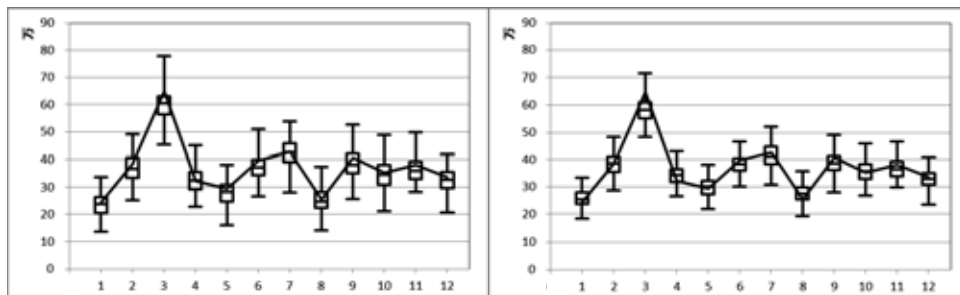


図3 $p_0=10\%$, $i=1\%$ の結果 (説明は図1に同じ)

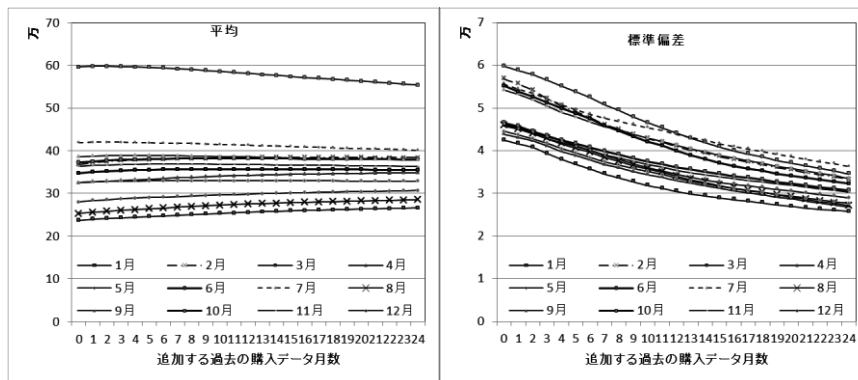


図4 $p_0=10\%$, $i=1\%$ の結果 (説明は図2に同じ)

5. 結論と今後の課題

家計調査の自動車購入額調査法の改善のため、実際の調査方法をモデル化したエージェント・ベース・シミュレーションを行った。基礎的検討のため購入数を対象とした。過去の購入情報の追加による推定結果の改善が確認された一方、購入世帯の記憶誤りにより季節パターンの平準化が生じることを示した。

本モデルにおいては、総世帯数が増加する中、世帯の同定の都合上、増加世帯は市町村ごとに最後尾に追加する設定としている。このため記憶誤りが無い場合でも新規調査世帯に増加世帯分の占める比率が高くなり、平均値の低下が観察された。世帯の新規発生と同定に関する検討とモデルの改善は今後の課題である。一方、実際の調査においても調査地区をランダムに選定していないことが同様の偏りを生じさせる可能性があることを示唆している。

引用文献

- ・宇南山卓「家計調査の課題と改善に向けて」統計と日本経済, 第1巻, 第1号, 2011年4月, 3頁-28頁
<http://www.cirje.e.u-tokyo.ac.jp/journal/20110102.pdf> (2011-9-6 参照)
- ・総務省「平成23年度予算概算要求書」総務省, 2010
http://www.soumu.go.jp/main_content/000087501.pdf (2011-9-6 参照)
- ・総務省「家計調査の概要」総務省, 掲載年不詳
<http://www.stat.go.jp/data/kakei/1.htm> (2011-9-6 参照)
- ・SOARS Project「SOARS Project オフィシャルサイト」
<http://www.soars.jp/ja/> (2011-9-6 参照)
- ・エージェントベース社会システム科学研究センター「AADL」
<http://www.cabsss.titech.ac.jp/aad1.html> (2011-9-6 参照)