離散型サンプルセレクションモデルとベイズ推定: 都市環境が自動車保有に及ぼす因果効果の検証

渡邉　萌[1]・円山　琢也[2]

[1]学生会員　熊本大学　大学院自然科学教育部（〒860-8555 熊本市中央区黒髪2-39-1）
E-mail: 197d9225@st.kumamoto-u.ac.jp

[2]正会員　熊本大学准教授　くまもと水循環・減災研究教育センター（〒860-8555 熊本市中央区黒髪2-39-1）
E-mail: takumaru@kumamoto-u.ac.jp

土地利用パターンや道路幅員，アクセシビリティなどの都市環境の違いが，住民の自動車保有に与える影響は大きい．それらの影響を正しく把握することは，自動車依存社会からの脱却や，コンパクトシティ施策の効果予測のために必要となる．そのためには，(a) 共変量の影響の除去，(b) 選択バイアスの補正の二つを考慮した解析が求められる．既存の手法を用いて都市環境が自動車保有などの離散的な交通行動に与える因果効果を明らかにする際，選択バイアスと関係する未観測の変数を考慮できない点が大きな課題であった．本研究では，未観測変数によるバイアスを明示的に補正する因果推論手法であるサンプルセレクションモデルの枠組みで，離散型の結果変数を扱うことができるモデルとベイズ推定による解法アルゴリズムを提案した．また，熊本都市圏PT調査データへの適用を行い，熊本市周辺地域から熊本市内への転居により，世帯主の自動車保有確率が19.2%減少することが明らかとなった．選択バイアスの補正を行わない場合の因果効果は14.5%の減少と算出され，提案モデルの有効性が示された．

Key Words : *causal inference, sample selection model, self-selection, built environment, car ownership*

1. はじめに

自動車依存型社会からの脱却やコンパクトシティ施策の効果計測には，居住地の都市環境が自動車保有に与える影響を明らかにすることが重要である．一般的に，都市部と郊外部では住民の自動車保有割合は異なり，都市部の住民の自動車保有割合は低く，郊外部の住民の自動車保有割合は高い傾向にある[1]．これは，土地利用パターンや道路幅員，商業施設や公共交通機関までのアクセス等，様々な都市環境の差異が要因として考えられる．ここで郊外部居住者が都市部に住み替えを促す政策によって，自動車保有率が変化するかどうかを検討するとする．このとき，都市部と郊外部の自動車保有割合の差を，予想される変化として直接的に解釈することはできない．なぜなら，都市部と郊外部における都市環境の差異と自動車保有の間の相関関係は，交絡や逆の因果性が介在する可能性があり，必ずしも因果関係を意味しないためである．より信頼性の高い結果に基づく政策的示唆を得るためには，統計的因果推論手法に基づいた因果効果の算出が必要である[2]．統計的因果推論手法に基づき，都市環境がトリップ数や自動車走行距離等の連続的な交通行動に及ぼす因果効果に関する実証研究が進められている[3]．その一方で，自動車保有などの離散的な交通行動を扱う既存の因果推論手法は限られており，また既存の分析法では人々の交通行動と居住地選択に関する多くの情報が必要となる．そのため，これまで都市環境が離散的な交通行動に与える因果効果に関する知見が蓄積されていない．よって，離散的な交通行動をより少ない情報量で分析することのできる手法の開発が望まれる．

本研究では，サンプルセレクションモデルの枠組みで，離散的な結果変数を扱うための統計的因果推論手法を提案する．具体的には，居住地の都市環境の差異 (都市部・郊外部) が与える，住民の自動車保有確率への因果効果を推定する手法を提案する．さらに，熊本都市圏居住者の自動車保有行動へ適用し，手法の妥当性を検討する．本稿では，離散的な結果変数を扱えるサンプルセレクションモデルを「離散型サンプルセレクションモデル」と呼称する．

2. 既存の統計的因果推論手法の整理と課題

本研究で提案する手法は，Rubin[4]によって提案された反実仮想アプローチ[5]に分類される． 以下，一般的な反実仮想アプローチの具体的な流れと既存手法の限界を踏まえた上で，本研究で提案するモデルの意義を述べる．

(1) 反実仮想アプローチによる因果効果の推定法

a) 共変量による調整

本研究のように都市環境の差異が自動車保有に与える効果を推定する際，独立変数である都市環境の差異 (都市部・郊外部) と，結果変数である自動車保有の両方に関係する共変量の影響を除去する必要がある．例えば，都市部における18歳未満の人口比率が高ければ，都市部の自動車保有割合は当然低くなる．このような影響が除去された効果を算出しなければ，都市環境の差異が自動車保有に与える因果効果を明らかにすることはできない．本研究で提案するモデルでは，都市部と郊外部の結果変数それぞれに共変量を説明変数とした潜在的な回帰関数を仮定し，都市部と郊外部の結果変数の期待値の差を因果効果とすることで共変量による調整を行う．

b) 選択バイアスの補正

都市環境の差異と自動車保有との間の因果効果を算出する際には，都市環境の差異が自動車保有に与える効果に介在する選択バイアスを補正する必要がある．一般的に反実仮想アプローチにおいて対処が必要となる選択バイアスとは，対象者が観測されるかどうかが観測値以外の要因にも依存している状況で，単純な解析を行うことにより生じるバイアスである．具体的な例として，車を保有せず日常的に公共交通機関を利用する人は，より公共交通機関が整備された都市部を居住地として選択する可能性がある (residential self-selection: 居住地自己選択)[3]．この場合，人々の居住地選択 (都市部・郊外部) は彼らの自動車保有に対する選好や習慣と強く相関してしまい，単純な解析ではバイアスのかかった結果となる．

(2) 選択バイアスの補正方法

分析の際，選択バイアスが生じない場合は上述した共変量による調整のみで適切な解析が可能である．しかし，本研究のように選択バイアスが存在する可能性を無視できない場合は，共変量による調整と同時に選択バイアスを補正する必要がある．このとき，反実仮想アプローチでは主に以下の二つの補正の方法が挙げられる[6]．

a) 観測値による補正

一つ目は，選択バイアスを誘発すると考えられる選好や習慣，その他様々な要因に関する情報を調査により取得する方法である．すなわち，対象者が観測されるかどうかに影響を及ぼす要因を可能な限り全て観測する方法である．これにより，統合離散選択モデル[7]や傾向スコアをはじめとした様々な手法を用いた解析による推論が可能になる．しかし，これらの解析手法は選択バイアスの要因が全て観測できていることを前提としているため，観測値以外の要因が存在する場合，結果の信頼性が損なわれる．そのため，結果の信頼性を高めるには質問事項の数は多くなり，高い調査コストを必要とする．

b) サンプルセレクションモデル

二つ目は，未観測な要因による選択バイアスを明示的に補正する手法を用いて解析を行う方法である．具体的な手法として，サンプルセレクションモデルが挙げられる．この手法はヘックマンのプロビット選択モデル[8]を統計的因果推論の考え方を導入して拡張したものであり，交通分野での適用研究も見られる[9]．サンプルセレクションモデルでは，一般的には結果変数に重回帰モデルを仮定する．それゆえ，サンプルセレクションモデルの結果変数は連続量である必要があり，自動車保有 (保有する，保有しない) のような離散的な結果変数を持つ問題にそのまま適用することができない．

(3) 既存手法の課題

現状，本研究で対象とする，都市環境の差異が自動車保有のような離散的な交通行動に与える因果効果を明らかにするための手法は限られている．また，それらの既存の手法に関して，以下に示す課題が挙げられる．

a) 解析手法の課題

上述したように，未観測な要因による選択バイアスを明示的に補正するサンプルセレクションモデルは，連続量の結果変数を持つ交通行動 (トリップ数，自動車保有台数など) しか扱うことができない．したがって，都市環境の差異が自動車保有のような離散的な交通行動に与える因果効果を推定する際に，未観測な選択バイアス要因が存在する場合，因果効果を正しく推定する手法は現状存在しない．

b) データ上の課題

選択バイアスと関係する選好や習慣等に関する情報は，パーソントリップ調査(以下PT調査) などの都市圏を対象とした伝統的な大規模調査では十分に取得しないケースがほとんどである．また，選択バイアスと関係する選好や習慣は，対象とする交通行動によって異なるため，一般的に分析の度に新しく調査を実施し，取得する必要がある．

(4) 離散型サンプルセレクションモデルの意義

上述したように，既存手法の限界と大きな調査コストにより，離散的な交通行動における都市環境の差異による因果効果を明らかにした既存研究は限られている．そのため，これまで知見があまり蓄積されていない．しか

し，離散型サンプルセレクションモデルの開発により，様々な分析が既存のデータを用いて可能になる．日本においても，PT調査をはじめとした様々な交通行動調査が長年実施されている．一般的に これらの調査では選択バイアスの要因は十分に観測されていない．しかし，上述の通りサンプルセレクションモデルは観測値以外の要因によるバイアスを明示的に補正することができるため，過去に取得されたそれらのデータを用いて推論を行うことができる．しかしながら，筆者の知る限り，離散的な結果変数を扱えるサンプルセレクションモデルへの拡張を試みた研究は存在しない．次章では，一般的なサンプルセレクションモデルの拡張である離散型サンプルセレクションモデルの提案を行う．

3. 離散型サンプルセレクションモデル

(1) 一般的なサンプルセレクションモデル

まず，一般的なサンプルセレクションモデルを説明する．ここでは，本提案モデルにおける，(a) 共変量による影響の除去，(b) 選択バイアスの補正方法の2点に焦点を当てる．基本的にこの2点は，次節で説明する離散型サンプルセレクションモデルでも共通である．

一般的なサンプルセレクションモデルは以下のように定式化される．

$$y_{i1}^* = \beta_1 x_{i1} + u_1,$$
$$y_{i1} = \begin{cases} 1 \ (\text{if } y_{i1}^* > 0) \\ 0 \ (\text{if } y_{i1}^* \le 0) \end{cases}, \tag{1}$$

$$y_{i2} = \beta_2 x_{i2} + \varepsilon_2 \ (\text{if } y_{i1} = 1),$$
$$y_{i3} = \beta_3 x_{i3} + \varepsilon_3 \ (\text{if } y_{i1} = 0). \tag{2}$$

ここで，iは対象とする個人，xは共変量，βは係数ベクトル，$u_1, \varepsilon_2, \varepsilon_3$は誤差項，$y_1^*$は潜在変数である．式(1)は一般的に選択方程式と呼ばれる．連続量の観測値であるy_2とy_3が同時に観測されることはなく，$y_1^* > 0$の時にy_2が観測され，$y_1^* \le 0$の時にy_3が観測されるとする．

サンプルセレクションモデルの枠組みでは，共変量を条件付きとしたy_2とy_3の期待値の差の標本平均を，共変量の影響が除去された因果効果とするのが一般的である．ただし，このとき誤差項間で相関が生じている場合，y_2とy_3は不偏推定量でも一致推定量でもなくなる．これにより生じるバイアスが，サンプルセレクションモデルの枠組みにおける選択バイアスとなる．よって誤差項間の相関を明示的に考慮することにより，選択バイアスの補正を行う．誤差項の同時分布は，次のような多変量正規分布による分散共分散行列を仮定するのが一般的である．

$$\begin{pmatrix} u_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix} = N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \sigma_2 & \sigma_3 \\ \sigma_2 & \sigma_{02}^2 & 0 \\ \sigma_3 & 0 & \sigma_{03}^2 \end{pmatrix} \right]. \tag{3}$$

識別性の問題からu_1の分散は1とする．$\sigma_{02}^2, \sigma_{03}^2$は結果変数である$y_2$と$y_3$の分散，$\sigma_2, \sigma_3$は共分散である．一般的なサンプルセレクションモデルのさらなる詳細と推定法については，既存研究[8),9)]を参照されたい．

(2) 離散型サンプルセレクションモデル

本節では離散型サンプルセレクションモデルを提案する．上で示した一般的なサンプルセレクションモデルとの違いは，(a) 結果変数，(b) 誤差項の同時分布，の2点である．順を追って説明する．

まず，離散型サンプルセレクションモデルでは三種類の潜在変数y_1^*, y_2^*, y_3^* を以下のように仮定する．

$$y_{i1}^* = \beta_1 x_{i1} + u_1,$$
$$y_{i1} = \begin{cases} 1 \ (\text{if } y_{i1}^* > 0) \\ 0 \ (\text{if } y_{i1}^* \le 0) \end{cases}, \tag{4}$$

$$y_{i2}^* = \beta_2 x_{i2} + \varepsilon_2 \ (\text{if } y_{i1} = 1),$$
$$y_{i3}^* = \beta_3 x_{i3} + \varepsilon_3 \ (\text{if } y_{i1} = 0), \tag{5}$$

$$y_{i2} = \begin{cases} 1 \ (\text{if } y_{i2}^* > 0) \\ 0 \ (\text{if } y_{i2}^* \le 0) \end{cases},$$
$$y_{i3} = \begin{cases} 1 \ (\text{if } y_{i3}^* > 0) \\ 0 \ (\text{if } y_{i3}^* \le 0) \end{cases}. \tag{6}$$

ここで，iは対象とする個人，xは共変量，βは係数ベクトル，$u_1, \varepsilon_2, \varepsilon_3$は誤差項である．ただし，両モデルの共通事項として，上記の因果効果の算出には$x_{i2} = x_{i3}$である必要がある．本稿では説明の都合上x_{i2}とx_{i3}を区別しているが，実際の推定では全く同じ共変量のセットを使用する．式(4)は選択方程式であり，$y_1^* > 0$の時にy_2が観測され，$y_1^* \le 0$の時にy_3が観測されるとする．また，y_2とy_3が同時に観測されることはない．離散型サンプルセレクションモデルでは，結果変数の観測値は離散変数y_2, y_3である．式(6)に示すように，それぞれ$y^* > 0$の時に1が観測され，$y^* \le 0$の時に0が観測されると仮定する．この点において，結果変数が連続量の観測値である一般的なサンプルセレクションモデルと異なる．そのため離散型サンプルセレクションモデルでは，共変量を条件付きとした潜在変数y_2^*, y_3^*の期待値の差の標本平均を因果効果と定義する． 因果効果の定義と算出については，5章にて詳述する．

誤差項の同時分布は以下のような多変量正規分布を仮定する．

$$\begin{pmatrix} u_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix} = N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 + \sigma_2^2 + \sigma_3^2 & \sigma_2 & \sigma_3 \\ \sigma_2 & 1 & 0 \\ \sigma_3 & 0 & 1 \end{pmatrix} \right]. \tag{7}$$

ここで，σ_2, σ_3は共分散である．結果変数の分散$\varepsilon_2, \varepsilon_3$は，パラメータの識別性の問題のため1に固定する[10)]．また，計算アルゴリズムの簡便化のため，潜在的な結果変数y_2^*, y_3^*が与えられた時の潜在変数y_1^*の条件付き分散$\mathrm{var}(y_1^* | y_2^*, y_3^*)$が1となるよう，$u_1$の分散は$1 + \sigma_2^2 + \sigma_3^2$

とする．したがって，離散型サンプルセレクションモデルと一般的なサンプルセレクションモデルでは，誤差項の同時分布が大きく異なる．この点が二つ目の違いである．

次に，提案モデルの尤度関数Lを，本稿では以下のように定義する．

$$\begin{aligned} L(y_1,y_2,y_3|\beta,\boldsymbol{\sigma}) = &\prod_{i\ni y_{i1}=1,y_{i2}=1} \mathrm{P}(y_{i1}=1|y_{i2},y_{i3})\mathrm{P}(y_{i2}=1) \\ &\times \prod_{i\ni y_{i1}=1,y_{i2}=0} \mathrm{P}(y_{i1}=1|y_{i2},y_{i3})\mathrm{P}(y_{i2}=0) \\ &\times \prod_{i\ni y_{i1}=0,y_{i3}=1} \mathrm{P}(y_{i1}=0|y_{i2},y_{i3})\mathrm{P}(y_{i3}=1) \\ &\times \prod_{i\ni y_{i1}=0,y_{i3}=0} \mathrm{P}(y_{i1}=0|y_{i2},y_{i3})\mathrm{P}(y_{i3}=0). \end{aligned} \tag{8}$$

$\boldsymbol{\sigma}=(\sigma_2,\sigma_3)$，$\Sigma_{23}$を式(7)の分散共分散行列$\Sigma$の右下$2\times 2$の単位行列とするとき，上記で述べたように

$$\mathrm{var}(y_1^*|y_2^*,y_3^*) = 1+\sigma_2^2+\sigma_3^2-\boldsymbol{\sigma}\Sigma_{23}\boldsymbol{\sigma}^{\mathrm{T}} = 1, \tag{9}$$

であるから，y_{i2}とy_{i3}が与えられた場合の$y_{i1}=1$となる確率は以下の二項プロビットモデルで表される．

$$\begin{aligned} \mathrm{P}(y_{i1}=1|y_{i2},y_{i3}) &= \mathrm{P}(y_{i1}^*>0|y_{i2}^*,y_{i3}^*) \\ &= \Phi\{\beta_1 x_{i1}+\sigma_2(y_{i2}^*-\beta_2 x_{i2})+\sigma_3(y_{i3}^*-\beta_3 x_{i3})\}. \end{aligned} \tag{10}$$

ここで$\Phi(\cdot)$は標準正規分布の累積分布関数(CDF)である．

本稿では，居住地（都市部，郊外部）と自動車保有（保有する，保有しない）を観測値とする．具体的には，都市部に居住している住民は$y_{i1}=1$，郊外部に居住している住民は$y_{i1}=0$が観測されたとし，自動車を保有している住民はそれぞれ$y_{i2}=1$，$y_{i3}=1$，自動車を保有していない住民はそれぞれ$y_{i2}=0$，$y_{i3}=0$が観測されたと定義する．

離散型サンプルセレクションモデルは結果変数の生成過程に潜在変数を仮定しているため，一般的なサンプルセレクションモデルと異なり，最尤推定法によりパラメータを推定することが困難である．このようなモデルの最尤推定についてシミュレーションを行った既存研究では，初期値に大きく依存した局所解が出力されたことが報告されている[11]．また，Markov Chain Monte Carlo法(MCMC)を用いる場合であっても，式(8)に示す尤度関数を直接評価することは非常に高い計算コストを要する．このようなパラメータ推定の困難さが，サンプルセレクションモデルを離散型へ拡張する上で大きな課題であった．そこで本研究では，Daniels and Hogan (2008)[12]に基づき，複数の事後分布から求めたいパラメータ$\beta,\boldsymbol{\sigma}$のサンプリングを別々に行う解法アルゴリズムを構築し，パラメータ推定の効率化を試みる．

4. ベイズ推定

(1) 事後分布と解法アルゴリズム

本提案モデルでは，居住地選択y_1と自動車保有y_2,y_3の

		y_1^*	
		$y_1^*>0$	$y_1^*\le 0$
y_2^*	$y_2^*>0$	$y_1=1,y_2=1$	$y_1=0,y_2=1$
	$y_2^*\le 0$	$y_1=1,y_2=0$	$y_1=0,y_2=0$
y_3^*	$y_3^*>0$	$y_1=1,y_3=1$	$y_1=0,y_3=1$
	$y_3^*\le 0$	$y_1=1,y_3=0$	$y_1=0,y_3=0$

□:観測された，■:観測されていない

図-1 離散型サンプルセレクションモデルにおける潜在変数と観測変数の関係

生成過程には，式(7)に示すような非独立な関係を仮定している．そのため，y_2,y_3が観測されるかどうかは，それらの潜在変数であるy_2^*,y_3^*の値と関係している．このような場合，図-1に示す本提案モデルにおける欠測は，ランダムでない欠測 (MNAR: missing not at random) に分類される[5]．以下では，ランダムでない欠測を補完し推定の効率化を行うDaniels and Hogan (2008)[12]に基づき，本提案モデルにおける事後分布を導出する．

まず，観測されたy_2,y_3をy_{o2},y_{o3}，観測されていないy_2,y_3をy_{m2},y_{m3}として区別する．また，ここでは欠測データy_{m2},y_{m3}を潜在的な確率変数として扱う．$\pi(\cdot)$を確率分布とするとき，求めたいパラメータ$\beta,\boldsymbol{\sigma}$の事後分布は

$$\pi(\beta,\boldsymbol{\sigma}|y_1,y_{o2},y_{o3}) \propto \pi(\beta,\boldsymbol{\sigma})\mathrm{P}(y_1|y_2,y_3)\mathrm{P}(y_{o2},y_{o3}), \tag{11}$$

であり，このとき

$$\mathrm{P}(y_1|y_2,y_3) = \prod \mathrm{P}(y_{i1}|\beta_1,\boldsymbol{\sigma},y_{i2},y_{i3}), \tag{12}$$

$$\mathrm{P}(y_{o2},y_{o3}) = \prod_{i\ni y_{i1}=1}\mathrm{P}(y_{io2})\prod_{i\ni y_{i1}=0}\mathrm{P}(y_{io3}). \tag{13}$$

とする．ここで，パラメータ$\beta,\boldsymbol{\sigma}$と欠測データ$y_{m2},y_{m3}$の結合事後分布を考える．ベイズの定理より，以下が成立する．

$$\begin{aligned} &\pi(y_{m2},y_{m3},\beta,\boldsymbol{\sigma}|y_1,y_{o2},y_{o3}) \\ &\propto \pi(\beta,\boldsymbol{\sigma})\mathrm{P}(y_1|y_2,y_3)\mathrm{P}(y_{m2},y_{m3}|y_1,y_{o2},y_{o3})\mathrm{P}(y_{o2},y_{o3}). \end{aligned} \tag{14}$$

データ拡大法 (data augmentation) により，欠測データy_{m2},y_{m3}を補完すると，パラメータ$\beta,\boldsymbol{\sigma}$のサンプリングは以下の通りとなる．

$$\begin{aligned} &\pi(y_{m2},y_{m3},\beta,\boldsymbol{\sigma}|y_1,y_{o2},y_{o3}) \\ &\propto \pi(\beta_1,\boldsymbol{\sigma})\mathrm{P}(y_1|\beta_1,\boldsymbol{\sigma},y_2,y_3)\pi(\beta_2,\beta_3)\mathrm{P}(y_2,y_3|\beta_2,\beta_3). \end{aligned} \tag{15}$$

データ拡大法とは，パラメータを条件付けた上で潜在変数を事後分布からサンプリングし，サンプリングされた潜在変数を条件付けた上でパラメータを事後分布からサンプリングするというアルゴリズムである．プロビットモデルに代表される非線形モデルでは，潜在変数を導入することで，求めたいパラメータのサンプリングが容易になることがあり，多くの活用事例がある[10),13),14)]．なお，データ拡大法によるy_{m2},y_{m3}のサンプリングとギブスサンプリングによるパラメータ$\beta,\boldsymbol{\sigma}$のサンプリングは，確率変数を事後分布からサンプリングするという点で同じ

ものである．

したがって，提案モデルの解法アルゴリズムは以下の通りとなる．

Step 0: $k = 0$

Step 1: $y_{im2}^{(k)}, y_{im3}^{(k)} \sim \pi\left(y_{im2}, y_{im3} \middle| y_{i1}, y_{io2}, y_{io3}, \beta^{(k-1)}, \boldsymbol{\sigma}^{(k-1)}\right)$

Step 2: $\beta_1^{(k)}, \boldsymbol{\sigma}^{(k)} \sim \pi\left(\beta_1, \boldsymbol{\sigma} \middle| y_1, y_{o2}, y_{o3}, y_{m2}^{(k)}, y_{m3}^{(k)}, \beta_2^{(k-1)}, \beta_3^{(k-1)}\right)$

Step 3: $\beta_2^{(k)}, \beta_3^{(k)} \sim \pi\left(\beta_2, \beta_3 \middle| y_{o2}, y_{o3}, y_{m2}^{(k)}, y_{m3}^{(k)}\right)$

Step 4: $k = k + 1$とし，Step 1に戻る

kはMCMCの繰り返し数である．上記のステップを十分な回数繰り返すことにより，求めたいパラメータの事後分布を生成する．次節では，上記のStep 1~3を詳細に説明する．

⑵ サンプリング

まず，上記のStep 1に関する説明を行う．ベイズの定理より，欠測データy_{m2}, y_{m3}はそれぞれ以下の確率を持つベルヌーイ分布からサンプリングされる．

$$\mathrm{P}(y_{im2} = 1 | y_{i1}, y_{io3}, \beta, \Sigma) = \frac{\mathrm{P}(y_{i1}, y_{im2} = 1, y_{io3} | \beta, \Sigma)}{\sum_{z=0}^{1} \mathrm{P}(y_{i1}, y_{im2} = z, y_{io3} | \beta, \Sigma)}, \quad i \ni y_{i1} = 0, \tag{16}$$

$$\mathrm{P}(y_{im3} = 1 | y_{i1}, y_{io2}, \beta, \Sigma) = \frac{\mathrm{P}(y_{i1}, y_{io2}, y_{im3} = 1 | \beta, \Sigma)}{\sum_{z=0}^{1} \mathrm{P}(y_{i1}, y_{io2}, y_{im3} = z | \beta, \Sigma)}, \quad i \ni y_{i1} = 1. \tag{17}$$

ここで$\mathrm{P}(y_{i1}, y_{im2}, y_{io3} | \beta, \Sigma), \mathrm{P}(y_{i1}, y_{io2}, y_{im3} | \beta, \Sigma)$は3次元の多変量正規分布の確率である[15]．欠測データをサンプリングすることにより，全ての個人においてy_2, y_3が完全に得られる．

次に，Step 2~3を説明する．まず，以下に示すように，潜在変数y_{i2}^*, y_{i3}^*をデータ拡大法により生成する[10]．

$$f(y_{i2}^*, y_{i3}^* | y_{i1}, y_{i2}, y_{i3}, \beta, \Sigma) \propto \phi_3(y_i^* | \beta X_i, \Sigma) \prod_{j=1}^{3} \{\mathrm{I}(y_{ij}^* > 0)\mathrm{I}(y_{ij} = 1) + \mathrm{I}(y_{ij}^* \leq 0)\mathrm{I}(y_{ij} = 0)\}. \tag{18}$$

ここで$X_i = (x_{i1}, x_{i2}, x_{i3})$であり，$\phi_3(y_i^* | \beta X_i, \Sigma)$は平均$\beta X_i$と，式(7)に示す分散共分散行列$\Sigma$を持つ3次元の正規分布の確率密度関数である．$\mathrm{I}(\cdot)$は指示関数であり，この3次元正規分布から$\prod_{j=1}^{3}\{\mathrm{I}(y_{ij}^* > 0)\mathrm{I}(y_{ij} = 1) + \mathrm{I}(y_{ij}^* \leq 0)\mathrm{I}(y_{ij} = 0)\} = 1$を満たすまで$y_{i1}^*, y_{i2}^*, y_{i3}^*$のサンプリングを行い，求めたい$y_{i2}^*, y_{i3}^*$を生成する．

生成されたy_2^*, y_3^*を用いて，$\beta_1, \boldsymbol{\sigma}$と$\beta_2, \beta_3$を別々にサンプリングする．具体的には，2項プロビットモデル[13]のアルゴリズムを用いて式(10)より$\beta_1, \boldsymbol{\sigma}$を，多変量プロビットモデル[10]のアルゴリズムに基づきβ_2, β_3を二変量プロビットモデル$\mathrm{P}(y_2, y_3 | \beta_2, \beta_3, \Sigma_{23})$よりサンプリングする．ここでは詳細は省略するが，どちらもギブスサンプリングにより容易にサンプリングを行うことができる．

5. 因果効果の算出

3章で説明した一般的なサンプルセレクションモデルによる因果効果の算出は，Heckman et al. (2001, 2003)[16,17]にて詳しく説明されている．因果効果の指標としては，母集団における因果効果 (ATE) や処置群における因果効果 (TT) が挙げられる[2]．一般的に，求めたい因果効果とは母集団における因果効果であるATEのことを指す．本章では，提案した離散型サンプルセレクションモデルにおける，これらの因果効果の指標の計算方法について説明する．

離散型サンプルセレクションモデルにおけるATEとTTの算出の基本的な考え方は，一般的なサンプルセレクションモデルと同じである．本稿ではATEを，「対象とする母集団からランダムに抽出された個人が，郊外部から都市部に移住した場合の自動車保有確率の減少幅」と定義する．郊外部から都市部に移住した場合の自動車保有確率の減少幅を$\Delta \equiv \mathrm{P}(y_2 = 1) - \mathrm{P}(y_3 = 1)$とし，結果変数$y_2^*, y_3^*$における共変量を$\mathbf{x} = x_2 = x_3$とするとき，提案モデルにおける$\mathbf{x} = x$の時のATEは以下のように定義される．

$$\mathrm{ATE}(x) = \mathrm{E}(\Delta | \mathbf{x} = x) = \Phi(\beta_2 x) - \Phi(\beta_3 x). \tag{19}$$

一般的なサンプルセレクションモデルにおけるATEと異なる点として，標準正規分布の累積分布関数$\Phi(\cdot)$によりプロビット変換を行っている．これにより，提案モデルにおけるATEは0から1の間の数字をとる．式(19)より，提案モデルにおけるATEは以下の確率密度関数で表現される．

$$\mathrm{ATE} = \mathrm{E}[\Delta] = \int \mathrm{ATE}(\mathbf{x}) dF(\mathbf{x}). \tag{20}$$

ATEは，以下の式に示すように，全サンプル$\boldsymbol{N}$における$\mathrm{ATE}(x)$の平均をATEの期待値とみなして点推定により算出するのが一般的である．

$$\int \mathrm{ATE}(\mathbf{x}) dF(\mathbf{x}) \approx \frac{1}{N}\sum_{i=1}^{N} \mathrm{ATE}(x_i) = \frac{1}{N}\sum_{i=1}^{N}\left[\Phi\left(\beta_2 x_i\right) - \Phi\left(\beta_3 x_i\right)\right]. \tag{21}$$

本稿におけるTTは，「都市部に住む個人が，郊外部から都市部に移住した場合の自動車保有確率の減少幅」と定義する．ATEと同様に，TTも式(22)のように確率密度関数となる．

$$\mathrm{TT} = \mathrm{E}(\Delta | y_1 = 1) = \int \mathrm{TT}(x, x_1, y_1 = 1)\, dF(x, x_1 | y_1 = 1). \tag{22}$$

ATEと同様に，式(23)，式(24)より得られた点推定値をTTと算出する．

$$\mathrm{TT} \approx \frac{1}{\sum_{i=1}^{N} y_{1i}} \sum_{i=1}^{N} y_{1i} \mathrm{TT}(x_i, x_{i1}, y_{i1} = 1), \tag{23}$$

$$\mathrm{TT}(x, x_1, y_1 = 1) = \mathrm{P}(y_2 = 1 | y_1 = 1) - \mathrm{P}(y_3 = 1 | y_1 = 1) = \frac{\mathrm{P}(y_1=1, y_2=1)}{\mathrm{P}(y_1=1)} - \frac{\mathrm{P}(y_1=1, y_3=1)}{\mathrm{P}(y_1=1)}. \quad (24)$$

式(25)より，算出されたATEとTTから居住地自己選択による影響度を示す指標RSSを算出することができる[9].

$$\mathrm{RSS} = 1 - \frac{\mathrm{ATE}}{\mathrm{TT}}. \quad (25)$$

ATE/TTはBEP (Built Environment Proportion) と呼ばれ，本稿においては，居住地自己選択がコントロールされた場合の居住地の違い (都市部・郊外部) が自動車保有に与える影響度を示す.

6. モデルの適用

(1) 2012年熊本都市圏PT調査

ここでは，提案モデルを2012年熊本都市圏PT調査の付帯調査データに適用し，その有効性を検証する．対象となるサンプルは，付帯調査である「住まいに関する意識調査」に回答した世帯主であり，有効サンプルサイズは3,376である．そのうち，都市部に居住する回答者数は2,560，郊外部に居住する回答者数は816である．使用するデータは，対象者の自動車保有状況に関する情報や，個人属性，世帯属性，住まいに関する意識調査から得られた回答を用いる．これらのデータのうち，モデルの共変量として用いる変数の候補として使用したデータを**表-1**に示す．本分析では，「対象者がほぼ自分専用の自動車を保有しているかどうか」を自動車保有の有無と定義する．また，熊本市内を都市部，熊本市外を郊外部と定義し，**図-2**では自動車保有割合を居住地別 (都市部・郊外部) に示す．これより単純集計では，都市部と郊外部との間の自動車保有割合の差は16.1%である.

(2) 因子分析による因子スコアの算出

住まいに関する意識調査から得られたデータを基に，モデルに導入する変数を作成した．住まいに関する意識調査では，調査時に居住していた居住地を選択した際に重視した点とその程度を，「1. あまり重視してない，2. やや重視した，3. 重視した」の3段階で答える質問を設けている．対象者は**表-2**に示す14の項目から，最大3つまで選び，それぞれ上記の1~3の番号を回答している．得られた回答に対し，対象者の居住地選択の背後にある潜在的な居住地に対する選好を把握することを目的として，因子分析を行った．因子分析は，潜在的な居住地に対する選好 (共通因子) の個数を分析前に仮定する必要がある．共通因子の個数を決定する際によく用いられる指標として，「相関行列において値が1以上である固有値の個数」があり，この指標に基づいて本分析では4つと仮定した．因子分析により，共通因子と**表-2**に示す項目との関係を表す因子負荷量を計算できる．因子負荷量の絶対値が大きいほど，共通因子はその項目に対し強く影響を及ぼしていると解釈する．本分析では，因子負荷量の絶対値が0.2以上の場合は，その共通因子と項目は関係があると解釈した．得られた共通因子と因子負荷量を**表-3**に示す．本分析では，4つの共通因子をそれぞれ，「核家族としての生活」，「買い物・通院時のアクセス」，「公共交通へのアクセス」，「郊外部での生活」と解釈しラベル付けを行った．因子負荷量の絶対値が0.2未満の項目については，**表-3**上では省略している．得られた因子負荷量から，4つの共通因子の因子スコア (因子得点) をそれぞれ対象者毎に算出できる．この因子スコアにより，一人一人の対象者がどの共通因子を重視しているのかを数値化することができる.

(3) モデルの推定

調査より得られた個人属性と前節で算出した因子スコアを，対象者の居住地選択 (都市部・郊外部) と自動車保有の有無を説明する変数として推定に使用する．MCMCの繰り返し数は35,000回とし，burn-in 区間を5,000と設定する.

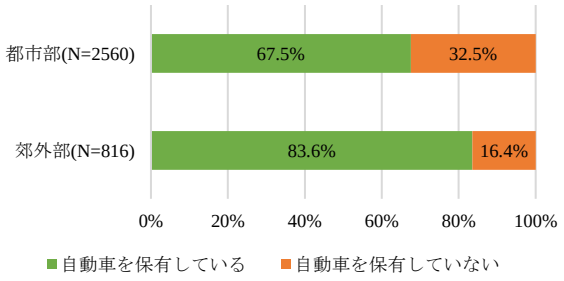

図-2 居住地別の自動車保有割合

表-1 変数の説明

変数名	説明	% or Mean (SD)	
		都市部	郊外部
男性ダミー	対象者(世帯主)が男性の場合1，それ以外は0	76.5%	82.6%
年齢	対象者の年齢	48.2 (14.7)	47.1 (14.6)
世帯人数	対象者が属する世帯の人数	3.11 (1.46)	3.31 (1.46)
第一次産業ダミー	対象者の職業が第一次産業に分類される場合1，それ以外0	1.2%	1.8%
第二次産業ダミー	対象者の職業が第二次産業に分類される場合1，それ以外0	15.1%	28.1%
共通因子1	共通因子「核家族としての生活」の因子スコア	-0.02 (0.98)	0.06 (1.00)
共通因子2	共通因子「買い物・通院時のアクセス」の因子スコア	0.04 (1.11)	-0.13 (1.08)
共通因子3	共通因子「公共交通へのアクセス」の因子スコア	0.07 (0.99)	-0.20 (0.89)
共通因子4	共通因子「郊外部での生活」の因子スコア	-0.06 (0.84)	0.20 (0.90)

表-2 現在の居住地を選んだ際に重視した点

項目
駅・電停までの距離
バス停までの距離
スーパー等までの距離
かかりつけの病院までの距離
通勤・通学先までの距離
熊本市中心部までの距離
小中学校までの距離
自動車の利用のしやすさ
価格・家賃
住宅の広さ・間取り
親からの相続や同居
親・親戚までの距離
緑豊かな自然
災害に強い場所

表-3 共通因子と因子負荷量

共通因子	項目	因子負荷量
核家族としての生活	かかりつけの病院までの距離	-0.23
	通勤・通学先までの距離	0.43
	熊本市中心部までの距離	0.20
	自動車の利用のしやすさ	0.27
	価格・家賃	0.77
	住宅の広さ・間取り	0.76
	親からの相続や同居	-0.31
買い物・通院時のアクセス	バス停までの距離	0.29
	スーパーまでの距離	0.75
	かかりつけの病院までの距離	0.78
	自動車の利用のしやすさ	0.22
公共交通へのアクセス	駅・電停までの距離	0.85
	バス停までの距離	0.58
	熊本市中心部までの距離	0.40
郊外部での生活	自動車の利用のしやすさ	0.24
	親からの相続や同居	0.33
	親・親戚までの距離	0.39
	緑豊かな自然	0.67
	災害に強い場所	0.56

表-4 事前分布

パラメータ	平均	分散
β	0	100
σ_2	0	0.01
σ_3	0	0.01

表-5 推定結果($\sigma = 0$に固定)

共変量		パラメータ	t値	
β_2	定数項	0.61	5.71	**
	年齢/100	-0.44	-2.49	*
	男性ダミー	0.25	4.05	**
	世帯人数	-0.03	-1.90	
	共通因子2	-0.06	-2.36	*
	共通因子3	-0.16	-5.75	**
β_3	定数項	1.65	7.09	**
	年齢/100	-1.29	-3.64	**
	男性ダミー	0.16	1.18	
	世帯人数	-0.06	-1.55	
	共通因子2	-0.06	-1.14	
	共通因子3	-0.13	-2.01	*
サンプルサイズ			3,376	
自由度調整済み尤度比			0.280	

*: p<0.05, **: p<0.01

表-6 推定結果

共変量		パラメータ	t値	
β_1	定数項	1.00	14.69	**
	世帯人数	-0.04	-2.25	*
	第一次産業ダミー	-0.33	-1.58	
	第二次産業ダミー	-0.53	-8.12	**
	共通因子1	0.06	1.97	*
	共通因子2	0.14	4.75	**
	共通因子3	0.22	7.21	**
	共通因子4	-0.19	-5.76	**
β_2	定数項	0.52	4.68	**
	年齢/100	-0.39	-2.19	*
	男性ダミー	0.22	3.58	**
	世帯人数	-0.04	-2.18	*
	共通因子2	-0.04	-1.70	
	共通因子3	-0.14	-4.69	**
	σ_2	0.25	2.65	**
β_3	定数項	1.67	6.54	**
	年齢/100	-1.20	-3.16	**
	男性ダミー	0.14	0.98	
	世帯人数	-0.06	-1.60	
	共通因子2	-0.06	-1.07	
	共通因子3	-0.12	-1.98	*
	σ_3	0.04	0.47	
サンプルサイズ			3,376	
自由度調整済み尤度比			0.282	

*: p<0.05, **: p<0.01

表-7 因果効果

指標	値
ATE	-0.192
TT	-0.161
RSS	-0.192

7. 推定結果

提案モデルの推定に用いた事前分布を**表-4**に，提案モデルによる推定結果と算出した因果効果に関する指標をそれぞれ**表-5**，**表-6**，**表-7**に示す．**表-5**は$\boldsymbol{\sigma}$を0に固定した場合のβ_2, β_3の推定結果を示している．

前章で作成した4種類全ての因子スコアは，居住地選択行動を説明することに成功している (p<0.05)．具体的な傾向としては，「核家族としての生活」，「買い物・通院時のアクセス」，「公共交通へのアクセス」を重視する対象者は都市部に居住する傾向がみられた．また，「郊外部での生活」を重視する対象者は，実際に郊外部に居住している傾向が見られた．

自動車保有行動において，都市部と郊外部に共通する傾向としては，対象者の年齢が高いほど車を保有しない傾向にあること，また「公共交通へのアクセス」を重視する対象者は車を保有しない傾向にあることが示された．

共分散σ_2, σ_3はそれぞれ正の符号を示した．σ_2は正に有意であり (p<0.01)，これは都市部への居住と自動車保有の両方に正の影響を与える未観測な選択バイアス要因が存在することを示している．σ_3に関しては，値の絶対値が小さく，また統計的に有意でないことから

(p>0.05)，郊外部への居住と自動車保有の間には選択バイアス要因の影響は確認されなかった．

因果効果の指標に関しては，**表-7**に示すようにATEとTTはそれぞれ-0.19と-0.16であり，BEP=1.19，RSS=–0.19となった．5章で定義したように，ATEとTTは自動車保有確率の減少幅であるため， ATEによると都市部と郊外部の自動車保有確率の差は19.2%であり，TTでは16.1%である．また，**表-5**で示すパラメータを用いて，$\boldsymbol{\sigma} = 0$の場合，すなわち共変量の調整のみを行いバイアスの補正は行わない場合のATEの算出を行った結果，ATE=–0.145，すなわち自動車保有確率の減少幅は14.5%であった．

8. 考察

Mokhtarian and Cao (2008)[6]は，居住地自己選択を引き起こす要因を，(1) 欠測した社会経済的属性，(2) 観測されない態度変数の2種類に分類している．本稿で提案した離散型サンプルセレクションモデルは，これらの未観測な要因が起因して生じるバイアスを，$\boldsymbol{\sigma}$により明示的に補正した上で因果効果を算出することができる．しかし，提案モデルは上記の2種類によるバイアスを区別しない．そのため後者の方に関心がある場合には，社会経済的属性は全て観測しモデル上で考慮する必要がある．

σ_2が正に有意であることから，都市部への居住と自動車保有の両方に正の影響を与える未観測要因が存在することを示している．その未観測要因として，年収等の対象者の経済的な要因が挙げられる．2012年熊本都市圏PT調査とその付帯調査では年収に関するデータを取得できておらず，それらの要因はモデル上で考慮されていない．日本における交通調査では，年収等，回答者に関する一部の社会経済的属性は取得しないことが多い．提案モデルはそのような社会経済的属性の欠測が起因して生じるバイアスを明示的に補正するため，過去に実施された調査データへの適用を含め，様々な都市・時点における実証研究の蓄積が期待される．

一方，提案モデルが採用している，未観測要因により生じる誤差相関が多変量正規分布に従うという仮定には批判もある[18]．そのため，提案モデルの信頼性は慎重に検証されるべきである．今後の課題として，社会経済的属性が未観測であることによる影響を本分析のように$\boldsymbol{\sigma}$で捉えた場合と，観測しモデル上で共変量として用いた場合とで算出される因果効果がどの程度異なるかを検証する必要がある．

本分析により得られたATE (19.2%) は，**図-2**に示す単純集計における自動車保有割合の差 (16.1%) とは3.1%異なり，さらにはバイアスの補正を行わなかった場合のATE (14.5%) とは4.7%異なる．これより，バイアスの補正を行わず共変量の調整のみを行った分析では，都市部と郊外部の都市環境の違いが自動車保有に与える影響を過小に評価してしまうことが示唆される．これは，$\boldsymbol{\sigma} = 0$に固定した場合は，上述した都市部における自動車保有確率に正の影響を与えている未観測な経済的要因を無視することになり，その影響を補う形でパラメータβ_2を過大推計してしまうためである．

RSSは一般的に0から1の間の値を取るとされている[9]．これは，基本的にはTT>ATEが成り立つと考えられているためである．しかしながら，本分析で算出したRSSはその区間[0, 1]の外にある．これは，TT<ATE，すなわち都市部に居住する世帯主の「郊外部から都市部に移住した場合の自動車保有確率の減少幅」が対象全体と比較して小さいということを示す．したがって，本来，都市部に移住することによる自動車保有の減少が比較的期待できない世帯主が，現在の都市部に居住していることになる．これは，前述したように経済的に余裕のある世帯主が都市部に居住する傾向があるためだと考えられる．理想的には，自動車保有志向のある住民は自動車の利用がしやすい郊外部へ，公共交通利用志向のある住民は公共交通の発達した都市部に居住するのが都市計画上望ましい．しかし，一般的に都市部が比較的地価が高く，自動車を保有できるような経済的に余裕のある世帯の方が都市部に居住する傾向にあり，これは熊本都市圏でも同様であると考えられる．したがって，都市部への移転を促進するだけでは，そのような世帯が流入する傾向にあるため，その場合，都市部への移住が世帯主の自動車保有確率に与える因果効果 (19.2%の減少) が完全には発揮されない可能性が示唆される．

コンパクトシティ施策に期待される効果の一つに，自動車依存型社会からの脱却が挙げられる．熊本市でも中心市街地や公共交通軸沿線への居住促進と公共交通網の維持・開発を同時に進めることが計画されている[19]．このような住民の移住と公共交通への転換が同時に起こりうる状況下で，公共交通の整備が住民の交通行動に与える因果効果を算出する際には，居住地自己選択によるバイアスを補正する必要がある．一般的に大都市の居住者を対象として回答項目の多い交通調査を実施するのはコスト面で効率的でなく，過去に実施された調査データを用いて因果効果を算出できる本提案モデルの意義は大きいと考えられる．また，本稿では一時点における分析を行ったが，過去複数時点での調査から算出されたATEを比較することで，都市構造と住民の交通行動の関係性の長期的な変化を明らかにすることができる． ただし，提案モデルのようなサンプルセレクションモデルによる選択バイアスの補正の精度は，選択方程式である居住地選択モデルの精度に依存する[9]．そのため，居住地選択を説明することのできる情報を十分に取得できているかどうかに注意を払う必要がある．

9. 結論

本研究では，離散的な交通行動における都市環境の差異による因果効果を算出する離散型サンプルセレクションモデルの提案を行った．提案モデルにより，未観測な選択バイアス要因が存在する場合であっても，それにより生じるバイアスを明示的に補正した上で因果効果を算出することができる．提案モデルは既存の手法と比較して必要となる情報量が少ないため，既存手法の適用においては情報量が不足しているとされてきた調査データを用いた実証研究の蓄積が期待される．

本稿では提案モデルを熊本都市圏PT調査データに適用した．居住地の違い(熊本市内またはその周辺地域)が自動車保有に与える因果効果を示すATEを算出し，住民の自動車保有確率に19.2%の差があることが明らかになった．これは，選択バイアスの補正を行わず共変量の調整だけを行った場合の因果効果(14.5%)とは4.7%異なり，提案モデルの有効性が示された．

現状では，自動車保有志向のある世帯主(例えば年収の高い世帯主)が都市部に居住する傾向にあることが明らかとなった．これより，都市部への移転を促進するだけでは，都市部が自動車保有の減少に与える因果効果が完全には発揮されないことが示唆される．

謝辞：本研究は日本学術振興会特別研究員奨励費(DC2: 20J15157)とJSPS科研費(18H01561)の助成を受けている．ここに感謝の意を表します．

参考文献

1) 自動車検査登録情報協会：都市別の自家用乗用車の普及状況, 2019.
2) 織田澤利守, 大平悠季：交通インフラ整備効果の因果推論：論点整理と展望, 土木学会論文集 D3（土木計画学）, Vol.75, No.5, pp.I_1-I_15, 2019.
3) Cao, X., Mokhtarian, P. L. and Susan L. Handy.: Examining the impacts of residential self-selection on travel behavior: a focus on empirical findings, *Transport Reviews*, Vol.29, No.3, pp.359-395, 2009.
4) Rubin, D. B.: Estimating causal effects of treatments in randomized and nonrandomized studies, *Journal of Educational Psychology*, Vol.66, pp.688-701, 1974.
5) 星野崇宏：調査観察データの統計科学：因果推論・選択バイアス・データ融合, 岩波書店, 2009.
6) Mokhtarian, P. L. and Cao, X.: Examining the impacts of residential self-selection on travel behavior: A focus on methodologies, *Transportation Research Part B: Methodological*, Vol.42, No.3, pp.204-228, 2008.
7) Bhat, C. R. and Guo, J. Y.: A comprehensive analysis of built environment characteristics on household residential choice and auto ownership levels, *Transportation Research Part B: Methodological*, Vol.41, No.5, pp.506-526, 2007.
8) Heckman, J. J.: Sample selection as a specification error, *Econometrica*, Vol.47, pp.153-161, 1979.
9) Mokhtarian, P. L. and van Herick, D.: Quantifying residential self-selection effects: A review of methods and findings from applications of propensity score and sample selection approaches, *Journal of Transport and Land Use*, Vol.9, No.1, pp.9-28, 2016.
10) Chib, S. and Greenberg, E.: Analysis of multivariate probit models, *Biometrika*, Vol.85, No.2, pp.347-361, 1998.
11) Rajbhandari, A.: Identification and MCMC estimation of bivariate probit models with partial observability, In Jeliazkov, I. and Yang, X.-S. (Eds.) *Bayesian Inference in the Social Sciences* (pp.299-316), Wiley.
12) Daniels, M. J., Hogan, J. W.: *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*, New York: Chapman & Hall, 2008.
13) Albert, J. H., Chib, S.: Bayesian analysis of binary and polychotomous response data, *Journal of the American Statistical Association*, Vol.88, No.422, pp.669-679, 1993.
14) Fang, H. A.: A discrete-continuous model of households' vehicle choice and usage, with an application to the effects of residential density, *Transportation Research Part B*, Vol.42, No.9, pp.736-758, 2008.
15) Genz, A.: Numerical computation of multivariate normal probabilities, *Journal of Computational and Graphical Statistics*, Vol.1, No.2, pp.141-149, 1992.
16) Heckman, J., Tobias, J. L., Vytlacil, E.: Four parameters of interest in the evaluation of social programs, *Southern Economic Journal*, Vol.68, No.2, pp.210-223, 2001.
17) Heckman, J., Tobias, J. L., Vytlacil, E.: Simple estimators for treatment parameters in a latent-variable framework, *Review of Economics and Statistics*, Vol.85, No.3, pp.748-755, 2003.
18) Bhat, C. R., Eluru, N.: A copula-based approach to accommodate residential self-selection effects in travel behavior modeling, *Transportation Research Part B*, Vol.43, No.7, pp.749-765, 2009.
19) 熊本市：熊本市立地適正化計画，2016. https://www.city.kumamoto.jp/common/UploadFileDsp.aspx?c_id=5&id=9398&sub_id=4&flid=80022

DEVELOPING A DISCRETE CHOICE MODEL WITH SAMPLE SELECTION AND ITS BAYESIAN ESTIMATION: EXAMINING THE BUILT ENVIRONMENT IMPACTS ON CAR OWNERSHIP

Hajime WATANABE and Takuya MARUYAMA