データ欠測に着目した居住地・交通行動選択のモデリングとWAICによる評価

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

観測が不完全であることを前提としてモデリングを行うことで居住地選択と交通行動の非観測異質性を直接的に記述できる．この考えに基づき，様々なサンプルセレクションモデルが提案されてきた．しかし，それらのモデルではサンプルの欠測メカニズムを表現するパラメータの識別が困難になるという課題が指摘されている．この点を無視してモデリングを行った場合，モデルの過剰・過小適合を引き起こしてしまい，誤った結論を導く危険性がある．この問題に対処するため，本研究ではベイズモデリングの技術を応用し，その検証を行う．具体的には，情報事前分布を用いてパラメータの識別を行い，Watanabe-Akaike information criterion (WAIC) によりその予測精度を評価する．ベイジアンサンプルセレクションモデルを熊本都市圏PT付帯調査データに適用した結果，情報事前分布の設定によりパラメータの推定値が大きく変化すること，またサンプルへの過剰適合が起こりうることが確認された．同時に，WAICによる情報事前分布の検証はそのような過剰適合を避けるために有効であることが示唆された．

Key Words : *land-use and transport, sample selection, Bayesian informative prior, WAIC*

1. はじめに

居住地選択と交通行動選択を統合したモデリングの重要性は長年議論されており，これまで数多くのモデルが提案されてきた．なかでも，データより得られる居住地選択と交通行動選択の結果を，欠測データ分析の枠組みで自由度の高いモデリングを行う，いわゆる誘導型と呼ばれるモデルの開発が近年盛んに行われている[1]．具体的には，誘導型のモデルは「対象者が現在の居住地とは異なる土地に移住した場合に，移住先で行うであろう交通行動の選択結果が欠測している」という前提で居住地選択と交通行動選択を統合したモデリングを行う．そのため，その欠測メカニズムへの対処が誘導型のモデリングを行う上で重要となる．

誘導型のモデルでは欠測メカニズムを明示的に記述する．欠測メカニズムには主に「完全にランダムな欠測」「ランダムな欠測」「ランダムでない欠測」の3種類に分類され，それぞれ対処方法が異なる[2]．誘導型のモデルが一般的に想定するのは，その中でも最も対処が困難とされる「ランダムではない欠測」である．代表的な誘導型のモデルであるサンプルセレクションモデルは，「ランダムでない欠測」が発生する際の規則性 (非ランダム性) を，多変量正規分布により明示的にモデリングを行うのが一般的である．

サンプルセレクションモデルは，多変量正規分布の分散共分散行列を構成するパラメータの推定が不安定になりやすいという大きな課題を抱えている． Copas and Li[3] は，Heckman[4]のサンプルセレクションモデルでは，欠測メカニズムに関する情報がデータからほとんど得られない状況が起こりうることを指摘した．欠測メカニズムに関する情報がデータからほとんど得られない場合，データのみを用いて分散共分散行列を構成するパラメータの識別を行うことが難しくなる．さらに，仮にパラメータを推定することができたとしても，過剰・過小適合したパラメータは誤った結論につながる危険性があるため，推定されたパラメータの妥当性の確認は必須となる．そのため，サンプルセレクションモデルにより居住地・交通行動選択を統合してモデリングを行う際には，パラメータの識別とその妥当性の確認を併せて行うことが望ましい．

ベイズ推定の技術を応用することで，パラメータの識別とその妥当性の検証を同時に行うことができる．まず，

パラメータに関する事前情報を与えることができる情報事前分布を仮定することで，非ベイズモデルでは識別できないパラメータの識別化が可能となる[5]．次に，仮定した情報事前分布の妥当性を評価するために，事後分布からのサンプリングによりWatanabe-Akaike information criterion (WAIC) [6]が計算できる．WAICは汎化誤差の漸近不偏推定量であり，サンプル外の予測精度に関する評価指標である．WAICにより，仮定した情報事前分布がサンプル外の予測精度の観点から適切であるかどうかを評価することができる．しかし筆者らの知る限り，上述の問題意識に基づき，データの欠測メカニズム表現するパラメータの識別とその妥当性の検証を併せて行った研究は見られない．

本稿では，過去に著者らが提案したベイジアンサンプルセレクションモデルを2012年熊本都市圏PT付帯調査データに適用し，(1) 情報事前分布によるパラメータの識別化，(2) WAICによる情報事前分布の評価，をそれぞれ行う．これにより，識別された欠測メカニズムがサンプル外の予測精度の観点から妥当であるかを検証する．

2. 手法

(1) 離散型サンプルセレクションモデル

筆者ら[7]が提案した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．

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

$$y_{i2}^* = \beta_2 x_{i2} + \varepsilon_2 \ (\text{if } y_{i1} = 1),$$
$$y_{i3}^* = \beta_3 x_{i3} + \varepsilon_3 \ (\text{if } y_{i1} = 0), \tag{2}$$
$$y_{i2} = \begin{cases} 1 \ (\text{if } y_{i2}^* > 0) \\ 0 \ (\text{if } y_{i2}^* \le 0) \end{cases},$$

$$y_{i3} = \begin{cases} 1 \ (\text{if } y_{i3}^* > 0) \\ 0 \ (\text{if } y_{i3}^* \le 0) \end{cases}. \tag{3}$$

ここで，iは対象とする個人，xは説明変数，βはパラメータベクトル，$u_1, \varepsilon_2, \varepsilon_3$は誤差項である．式(1)は離散型サンプルセレクションモデルにおける欠測メカニズムを示す選択方程式である． **図-1**が示すように$y_1^* > 0$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(3)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．

		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，■ : 欠測

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欠測の関係

$u_1, \varepsilon_2, \varepsilon_3$が互いに独立な場合，上述した欠測メカニズムは「ランダムな欠測」に分類される．しかし離散型サンプルセレクションモデルは「ランダムでない欠測」を想定するモデルであるため，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．

$$\begin{pmatrix} u_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix} = N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1+\sigma_2^2+\sigma_3^2 & \sigma_2 & \sigma_3 \\ \sigma_2 & 1 & 0 \\ \sigma_3 & 0 & 1 \end{pmatrix} \right]. \tag{4}$$

σ_2, σ_3は共分散である．よって，離散型サンプルセレクションモデルが想定する欠測メカニズムの誤差成分は，式(4)に示す多変量正規分布に従うと仮定する．そのため，欠測メカニズムのモデリングを行う上でパラメータσ_2, σ_3の推定は極めて重要となる．ここでは計算アルゴリズムの簡便化のため， u_1の分散は$1 + \sigma_2^2 + \sigma_3^2$とする．

提案モデルの尤度関数Lは以下のように定義される．

$$\begin{aligned} L(y_1, y_2, y_3 | \beta, \boldsymbol{\sigma}) = & \prod_{i \ni y_{i1}=1, y_{i2}=1} \mathrm{P}(y_{i1}=1, y_{i2}=1) \\ & \times \prod_{i \ni y_{i1}=1, y_{i2}=0} \mathrm{P}(y_{i1}=1, y_{i2}=0) \\ & \times \prod_{i \ni y_{i1}=0, y_{i3}=1} \mathrm{P}(y_{i1}=0, y_{i3}=1) \\ & \times \prod_{i \ni y_{i1}=0, y_{i3}=0} \mathrm{P}(y_{i1}=0, y_{i3}=0). \end{aligned} \tag{5}$$

このとき

$$\mathrm{P}(y_{i1}, y_{ij}) = \phi_2(y_{i1}, y_{ij} | \beta_1 x_{i1}, \beta_j x_{ij}, \Sigma_j), \tag{6}$$

$$\Sigma_j = \begin{pmatrix} 1+\sigma_2^2+\sigma_3^2 & \sigma_j \\ \sigma_j & 1 \end{pmatrix}. \tag{7}$$

$j \in 2, 3$，$\phi_2(\cdot)$は分散共分散行列Σ_jを持つ2次元の正規分布の確率密度関数である．

(2) 情報事前分布とベイズ推定

前節で説明したように，離散型サンプルセレクションモデルの分散共分散行列はσ_2, σ_3の2つのパラメータで構成される．ベイズ推定ではパラメータの事前分布を仮定する必要があるため，σ_2, σ_3の事前分布をそれぞれ正規分布$\pi(\sigma_2) \sim N(0, G_{02})$，$\pi(\sigma_3) \sim N(0, G_{03})$に従うとする．このとき，分散$G_{02}, G_{03}$を比較的小さい値に設定することで，$\pi(\sigma_2), \pi(\sigma_3)$はパラメータ$\sigma_2, \sigma_3$のスケールに関する情報を含んだ情報事前分布とみなせる．多変量正規分布の分散共分散行列の情報事前分布の分散は，おおよそ0.10～0.50の範囲で設定されることが一般的である[8]．

情報事前分布を用いる際は，その妥当性を検証する必要がある．情報事前分布を用いることでパラメータの識別が可能になることは，長年ベイズモデルの利点として認識されてきた．しかしながらベイズモデルは，識別可能か否かの二元論的な問題を，識別の程度の問題に変換させるだけであり[5]，その識別されたパラメータが妥当であるかを検証する必要がある．特に，欠測データのモデリングのようにパラメータに関する情報が不足する問題では，識別の妥当性に注意する必要がある．Daniels and Hogan[9]は，モデルの推定に本来必要なデータの大部分が欠測しているサンプルでは，パラメータの推定値が情報事前分布に大きく依存する可能性を指摘している．そのため，情報事前分布を用いてパラメータの識別化を行う際には，その妥当性に注意を払う必要がある．

分析者は情報事前分布を任意に設定するのではなく，モデルの評価指標を基になるべく客観的に設定するのが望ましい[10]．情報事前分布を通して分析者のなんらかの主観が入ることは一般的に強く懸念されているため，情報事前分布を用いる際は，情報事前分布に十分に大きな分散を持たせることが必要とされてきた．しかし，その「十分に大きな分散」の程度が議論されることは稀であり，情報事前分布の分散は分析者が任意で設定することがほとんどであった．このように分析者の主観性を排することが難しかった点も，情報事前分布の使用が敬遠されてきた理由の一つとして考えられる．しかし，次節で説明するWAICの開発により，モデルの予測精度を厳密かつ効率的に評価することが可能となった．そのため，候補となる情報事前分布をいくつか想定し，その中から予測精度を最大化させる情報事前分布を選択することで，可能な限り主観性を排して，データに基づいて客観的に情報事前分布を設定することができる．

本稿では，Daniels and Hogan[9]の「ランダムではない欠測」に対応したMarkov Chain Monte Carlo法 (MCMC) による解法アルゴリズムに基づき，パラメータの推定を行う．具体的には，欠測データをデータ拡大と呼ばれる手法により補完し，データが完全に観測されたという前提で多変量プロビットモデル[8,9]の解法アルゴリズムによりパラメータを求める．詳細は渡邉・円山[7]を参照されたい．

(3) Watanabe-Akaike information criterion (WAIC)

WAICはモデルのサンプル外の予測精度を効率的かつ精度よく推定する．Watanabe[6]により提案されたWAICは赤池情報量基準 (AIC) の一般化版と捉えることができ，漸近的に汎化誤差と同じ平均値と分散を持つ．そのため，WAICの減少は汎化誤差の減少と対応する．WAICは，一般的な交差検証手法の中で最も高精度に汎化誤差を推定できる一個抜き交差検証 (LOO-CV) と漸近的に等価である．一個抜き交差検証は計算量が非常に多いのに対し，WAICはMCMCにより得られた事後分布からのサンプルを用いて容易に計算可能である．したがって，WAICはLOO-CVとほぼ同じ精度の推定値をより短時間で計算できる．注意点として，本稿ではWatanabe[6]による本来の定義ではなく，Gelman[11]らによる定義によりWAICを計算する．

4. 実証分析

(1) 2012年熊本都市圏PT調査

前章にて説明した離散型サンプルセレクションモデルを2012年熊本都市圏PT付帯調査データに適用し，(1) 情報事前分布による欠測メカニズムを説明するパラメータの識別化と (2) WAICによる情報事前分布のサンプル外の予測精度の評価を行う．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．

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する．

住まいに関する意識調査から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．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．この因子スコアは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ため，モデルの説明変数として用いる． 例えば，「買い物・通院時のアクセス」の因子スコアが高い対象者は，買い物や通院の際のアクセスを重視する傾向があることを示す．**表-1**が示すように，都市部における「買い物・通院時のアクセス」の因子スコアは正であり，郊外部では負の値となっている．そのため，そのような対象者は郊外部ではなく都市部に居住する傾向があることを示している．

(2) 候補となる情報事前分布

本稿では，情報事前分布の分散G_{02}, G_{03}はそれぞれ0.001，0.01，0.10，0.20，0.50の5つの値を取り得ると想定し，G_{02}, G_{03}の組み合わせが取り得る全てのパターンを候補となる情報事前分布と仮定する．このとき，G_{02}, G_{03}の組み合わせで構成される情報事前分布のパターン数は5×5=25である．この25の候補の中からWAICが最小となるものを，サンプル外の予測精度の観点から最も適切な情報事前分布とする．

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 情報事前分布とWAIC

		G_{03}				
		0.001	0.01	0.10	0.20	0.50
G_{02}	0.001	7434.44	7434.66	—	—	—
	0.01	7423.54	7423.69	—	—	—
	0.10	7403.13	7403.46	—	—	—
	0.20	7401.76	7402.04	—	—	—
	0.50	7403.78	7403.86	—	—	—

表-3 情報事前分布と最終尤度

		G_{03}				
		0.001	0.01	0.10	0.20	0.50
G_{02}	0.001	-3696.96	-3696.91	—	—	—
	0.01	-3691.11	-3691.08	—	—	—
	0.10	-3678.60	-3678.56	—	—	—
	0.20	-3677.14	-3676.94	—	—	—
	0.50	-3676.58	-3676.51	—	—	—

表-4 情報事前分布とσ_2, σ_3の推定値

		G_{03}			
		σ_2	σ_3	σ_2	σ_3
		0.001		0.01	
G_{02}	0.001	0.03	0.003	0.03	0.04
	0.01	0.26	0.003	0.26	0.03
	0.10	1.09	0.002	1.09	0.02
	0.20	1.40	0.002	1.45	0.03
	0.50	1.99	0.002	2.03	0.01

(3) 推定結果

表-2から**表-4**にかけて示すように，10通りの情報事前分布の組み合わせでパラメータの推定を行った．G_{03}が0.10，0.20，0.50の場合の計15通りの候補に関しては，G_{03}の値に伴いWAICが増加することが明らかであるため，省略し計算の効率化を図った．得られたパラメータの事後分布を用いて，候補毎のWAICを計算した．

表-2に候補の情報事前分布とWAICを示す．G_{03}が大きくなるほどWAICは増加するのに対し，G_{02}が0.001から0.20にかけて大きくなるほどWAICは減少している．しかし，G_{02}が0.50の時点でWAICは増加に転じている．これより，情報事前分布の分散$(G_{02}, G_{03}) = (0.20, 0.001)$の場合のサンプル外の予測精度が，候補の中で最も高いことが示されている．$(G_{02}, G_{03}) = (0.20, 0.001)$の時のパラメータの推定結果は**表-5**に示す．

表-3に候補の情報事前分布と最終尤度，**表-4**に候補の情報事前分布とσ_2, σ_3の推定値をそれぞれ示す．**表-3**より，G_{02}, G_{03}がそれぞれ大きくなるほど最終尤度が増加している．これより，表中の計10個の候補の中では，$(G_{02}, G_{03}) = (0.50, 0.01)$の情報事前分布の場合が最もデータ内の適合度が高いことが示されている．**表-4**より，$(G_{02}, G_{03}) = (0.50, 0.01)$の時の$\sigma_2$の推定値は2.03であり，WAICが最小となった$(G_{02}, G_{03}) = (0.20, 0.001)$の時の推定値$\sigma_2$=1.40とは，値が大きく異なっている．

表-5 WAICを最小とする推定結果（*: p<0.05, **: p<0.01）

		パラメータ	t値	
β_1	定数項	1.64	7.93	**
	世帯人数	-0.41	-1.32	
	第一次産業ダミー	-0.91	-6.99	**
	第二次産業ダミー	-0.06	-2.30	*
	共通因子1	0.07	1.64	
	共通因子2	0.21	4.42	**
	共通因子3	0.35	5.99	**
	共通因子4	-0.32	-5.72	**
β_2	定数項	0.28	2.72	**
	年齢/100	-0.23	-1.35	
	男性ダミー	0.13	2.17	*
	世帯人数	-0.05	-2.85	**
	共通因子2	0.01	0.23	
	共通因子3	-0.05	-1.68	
	σ_2	1.40	5.70	**
β_3	定数項	1.63	6.95	**
	年齢/100	-1.26	-3.55	**
	男性ダミー	0.16	1.15	
	世帯人数	-0.06	-1.56	
	共通因子2	-0.06	-1.16	
	共通因子3	-0.13	-2.07	*
	σ_3	0.002	0.07	
サンプルサイズ			3,376	
G_{02}			0.20	
G_{03}			0.001	
最終尤度			-3677.14	
自由度調整済み尤度比			0.210	
WAIC			7401.76	

(4) 考察

WAICを最小化する情報事前分布を採用することで，パラメータの過小・過剰適合を回避することができる．WAICが最小となる情報事前分布と，最終尤度が最大となる情報事前分布が異なるという結果は，パラメータの過剰適合 (過学習) が生じていることを示している．WAICの低下は汎化誤差の減少，最終尤度の増加は訓練誤差の減少にそれぞれ対応している．そのため，情報事前分布の分散が$(G_{02}, G_{03}) = (0.50, 0.01)$の場合では訓練誤差を減少させることには成功しているが，汎化誤差に関しては$(G_{02}, G_{03}) = (0.20, 0.001)$の時よりも増加しており，サンプル外の予測精度は悪化している．これは典型的なデータへの過剰適合 (過学習) である．一般的な交通調査が対象としている母集団は大規模であるため，得られたサンプルに過小・過剰適合したパラメータを用いて，なんらかの結論を導いてしまうことは避けなければならない．そのため，訓練誤差の最小化と併せて汎化誤差をモニタリングすることは，パラメータの過小・過剰適合を回避するために重要である．

同様のモデルにおいてこれまで任意に設定されてきた情報事前分布は，その妥当性を客観的に検証されるべきである．サンプルセレクションモデルを用いた既存研究では，情報事前分布の妥当性の検証は行われておらず，分析者が任意に情報事前分布を設定している[12]．サンプルセレクションモデルでは欠測メカニズムを記述するパラメータの推定が不安定になる危険性が指摘されているため，それらのパラメータの事前分布の妥当性を客観的に検証することは必要不可欠である．少なくとも，**表-4**に示したように，いくつかの異なる情報事前分布を用いて感度分析を実施し，推定値が情報事前分布の設定に対し頑健であるかどうかを確認する必要がある．

情報事前分布の積極的な活用により誘導型のモデルのさらなる拡張が期待できるが，それに伴い情報事前分布への依存度も高くなると考えられる．**図-1**に示すように本稿にて使用した離散型サンプルセレクションモデルは，多くのデータが欠測していることを前提としている．さらに，一般的なHeckman型[4]のサンプルセレクションモデルとは異なり，データから得られる結果変数は連続量ではなく2値データである．この場合，データから得られる欠測メカニズムに関する情報は極めて少ない．そのため推定値が情報事前分布の設定に大きく依存していると考えられる．したがって，モデルを拡張し前提とするデータの欠測の度合いを大きくすると，情報事前分布への依存度がさらに高まることが予想される． しかしそのようなモデリングを行う際でも，WAICにより汎化誤差をモニタリングすることで，柔軟な現象記述と予測精度を両立させることが可能であると考えられる．

5. 結論

欠測データ分析の枠組みで自由度の高いモデリングを行うサンプルセレクションモデルにおいて，サンプルの欠測メカニズムを表現するパラメータの識別が不安定になる点が課題であった．本稿では，(1) 情報事前分布を用いてパラメータの識別を行い，(2) WAICによりサンプル外の予測精度の観点から識別されたパラメータの妥当性を検証した．離散型サンプルセレクションモデルを熊本都市圏PT付帯調査データに適用した結果，パラメータの推定値が情報事前分布の設定に大きく依存していることが示された．そのため，情報事前分布の設定次第では，パラメータの過小・過剰適合を引き起こしてしまうことが示唆される．このとき，WAICによりモデリングを検証することは，過小・過剰適合を避けるために重要であることが示された．

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) 福田大輔，力石真：離散-連続モデルの研究動向に関するレビュー, 土木学会論文集 D3（土木計画学），Vol.69, No.5, pp.I_497-I_510, 2013.
2) 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.
3) Copus, J. B. and Li, H. G.：Inference for non-random samples, *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, Vol.59, No.1, pp.55-95, 1997.
4) Heckman, J. J. : Sample selection as a specification error, *Econometrica*, Vol.47, pp.153-161, 1979.
5) Rossi, P. E. and Allenby, G, M. : Bayesian statistics and Marketing, *Marketing Science*, Vol.22, No.3, pp.304-328, 2003.
6) Watanabe, S. : Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory, *Journal of Machine Learning Research*, Vol.11, pp.3571-3594, 2010.
7) 渡邉萌，円山琢也：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, 土木計画学研究・講演集, Vol.62 (CD-ROM), 2020.
8) Chib, S. and Greenberg, E.: Analysis of multivariate probit models, *Biometrika*, Vol.85, No.2, pp.347-361, 1998
9) Daniels, M. J. and Hogan, J. W.: *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*, New York: Chapman & Hall, 2008.
10) Gelman, A. and Shalizi, C. R. : Philosophy and the practice of Bayesian statistics, *British Journal of Mathematical and Statistical Psychology*, Vol.66, pp.8-38, 2013.
11) Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. and Rubin, D. B. : *Bayesian Data Analysis, Third Edition*, Florida: CRC Press, 2013.

12) Wu, N., Song, X. (Ben), Yao, R., Yu, Q., Tang, C. and Zhao, S. : A Bayesian sample selection model based on normal mixture to investigate household car ownership and usage behavior, *Travel Behaviour and Society*, Vol.20, pp.36-50, 2020.

MODELING AND WAIC EVALUATION OF RESIDENTIAL CHOICE AND TRAVEL BEHAVIOR WITH MISSING DATA

Hajime WATANABE and Takuya MARUYAMA