A Model for Built Environment Effects on Mode Usage*

Kai Yoshioka[†1] and Tomomi Miyazaki[‡2]

[1]Department of Economics, University of California, Irvine
[2]Department of Economics, Kobe University

May 30, 2020

Abstract

This paper develops an econometric framework for estimating the effect of the built environment on transportation mode choice and usage when a large fraction of the population under study is nonlicensed. We use a multivariate ordinal outcomes model with a binary selection component to allow for (1) heterogeneity in built environment effects and (2) indirect effects urban form has on mode usage via license choice. Our exposition focuses on the joint modeling of correlated and discrete outcomes (binary and ordinal), strategizing with identification restrictions and nonidentification, and the efficient estimation of model parameters. The current analysis is Bayesian in large part because estimation via Markov chain Monte Carlo (MCMC) simulation is more practical than maximizing a likelihood function consisting of tens of thousands of high-dimensional integrals. We present an efficient estimation algorithm for our model and discuss a way to sample latent data from the truncated multivariate normal distribution in a computationally efficient manner. A separate contribution this paper makes to the urban/transportation economics literature is the cross entropy index for land use imbalance, which we propose as a replacement to the entropy index for land use mix/balance. Using individual/household-level data from the 5th Nationwide Person Trip Survey (NPTS), we investigate whether the built environment is a policy-relevant determinant of travel behavior in the Japanese elderly. Effects are found to be nonzero but modest at best. Our results and conclusions are broadly consistent with those based on the United States.

1 Introduction

We consider in this paper the problem of estimating the effect of the built environment on transportation mode choice and usage when a large fraction of the population under study does not have the option to drive. We address in particular two important ways in which a mode usage model could be misspecified in this context: (1) assuming that built environment effects are the same for those who can and cannot legally drive, and (2) treating the choice to be licensed as

*We would like to thank Professors David Brownstone and Ivan Jeliazkov for their many insightful comments, and Koji Adachi and Tatsuki Enomoto for data cleaning and handling. We would also like to acknowledge seminar participants at the University of Hawaiʻi at Mānoa. The main author is responsible for any errors.

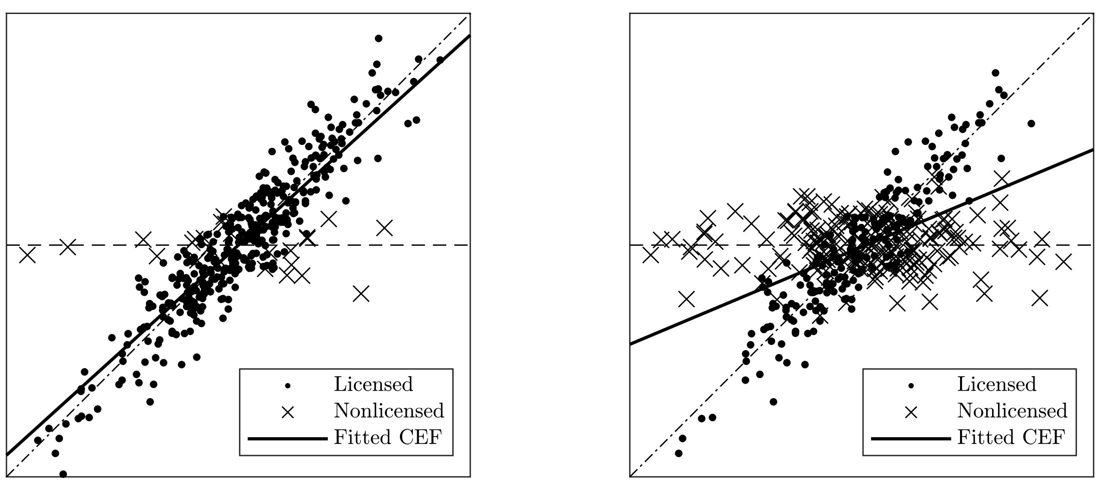
†Email: klyoshio@uci.edu

‡Email: miyazaki@econ.kobe-u.ac.jp

exogenous. We present an econometric framework to overcome these misspecification concerns and apply it to travel diary data to study whether the transportation habits of the Japanese elderly can be shaped using the tools of urban planning. To motivate our model, we begin by elaborating on how (1) and (2) could lead to incorrect inference.

Previous studies have found that an increase in density leads to more nonmotorized travel and less driving (Bento et al., 2005; Leck, 2006; Parady et al., 2015). These two effects seem highly plausible because jointly, they appeal to the notion of a substitution effect: all else held constant, improving accessibility to goods and services results in people opting to walk or bike instead of driving. Of course, if this were true, then densification should have a different effect on a person who cannot drive. If those who cannot drive make up a large portion of the population of interest, then failing to distinguish the density effect between the nonlicensed and licensed groups may lead to a serious case of confounding. This is shown in Figure 1. In the left pane, the licensed far outnumber the unlicensed, resulting in a line of best fit that resembles the former's conditional expectation function (CEF). Confounding is inconsequential here insofar as the average effect is concerned. If the two groups are similar in size, however, then confounding can be severe as shown in the right pane.

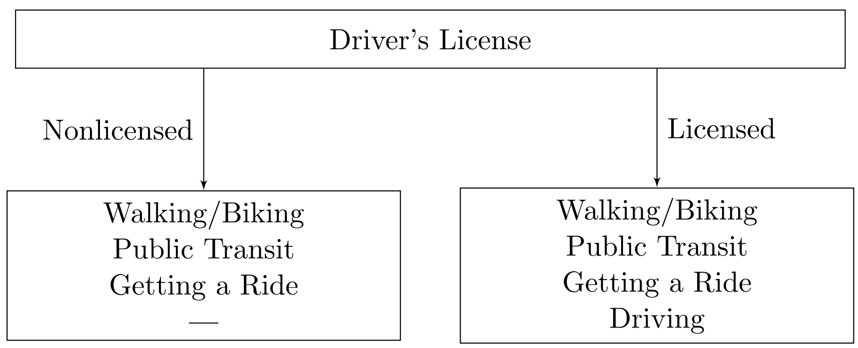
Figure 1: Confounding Effects



One possible solution to the problem of confounding is to fit two models, one for each group, so that built environment effects are estimated separately. Fitting two models also allows for different unobserved substitution patterns in the error term, and furthermore, renders trivial the task of excluding driving as a feasible transportation mode option for those not licensed. The issue with this strategy is that it treat license choice as exogenous. To see why this is problematic, suppose that improving accessibility to rail leads to less driving by drivers and also fewer drivers overall. Fitting a model for each group may reveal the former effect but not the latter.

To address these misspecification concerns, we build a multivariate ordinal outcomes model with a binary selection mechanism to estimate the effect of urban form on transportation mode choice and usage. A graphical representation of our model is given in Figure 2. This figure shows that, in the first stage, individuals select into either the nonlicensed group or the licensed group. This decision is observed for all units in the sample. In the second stage, individuals decide on how much of each available mode option to use. Available mode options are walking/biking, public transit, getting a ride, and for those who can do so legally, driving. Mode usage is measured on a three-bin ordinal scale. This is done for practical and theoretical reasons, as we will explain later. A key feature of our model is that mode usage is modeled jointly. Correlation patterns not captured by the included covariates appear in the covariance matrix of the error vector.

Figure 2: Multivariate Ordinal Outcomes Model with a Binary Selection Mechanism



Despite its intuitive appeal and innocuous appearance, the econometric modeling of the proposed model is not straightforward: license outcome is binary whereas mode usage are ordinal, all outcomes are discrete and correlated, the likelihood function is comprised of many high-dimensional integrals, and the modeling of counterfactuals yields unidentified parameters. Fortunately, the Bayesian paradigm offers practical solutions to all of these problems. Our estimation strategy relies on those of Albert and Chib (1993), Chib (2007), and Chib et al. (2009).

Empirically, our econometric model is motivated by growing concerns for the future of road safety in Japan, where babyboomers will reach the age of 75+ by 2025 (the so-called 2025 problem). A report by the National Police Agency Transportation Bureau (NPATB) in Japan found that in 2017, the average number of fatal accidents involving drivers age 75+ was double that of drivers under the age of 75. It also found that the number of individuals 75+ with driver's licenses doubled in the last decade (NPATB, 2018). The ultimate goal of this paper is to determine whether urban planning can help ensure a high traffic safety standard in Japan in the face of the 2025 problem.

A separate contribution we make to the urban/transportation economics literature is the cross entropy index for measuring land use imbalance. The literature has pointed out the many problems with the entropy index, which numerous studies have used to measure land use mix/balance. In this paper, we argue that if a reference distribution is available, a natural alternative to the entropy index is the cross entropy index. We do this by showing that the entropy and cross entropy indices are elegantly linked: If the reference distribution is the uniform distribution, then the entropy index is a valid measure of land use balance, and furthermore, balance and imbalance sum to unity.

The remainder of this paper is organized as follows: Section 2 elaborates on the proposed econometric model with emphasis on strategizing with identification restrictions and nonidentification. MCMC estimation is discussed in Section 3. The cross entropy index and data are covered in Sections 4 and 5, respectively. Results are given in Section 6, and Section 7 concludes.

2 Model

The econometric model represented graphically by Figure 2 contains eight equations: a selection equation, three mode usage equations for the nonlicensed group, and four mode usage equations for the licensed group. Following the latent variables framework of Albert and Chib (1993), we have for sample unit $i \in \{1, \ldots, n\}$:

$$\text{Selection Mechanism - License}: z_{i1} = \boldsymbol{x}_{i1}'\boldsymbol{\beta}_1 + \epsilon_{i1} \tag{1}$$
$$\text{Nonlicensed - Walking/Biking}: z_{i2} = \boldsymbol{x}_{i2}'\boldsymbol{\beta}_2 + \epsilon_{i2} \tag{2}$$
$$\text{Nonlicensed - Public Transit}: z_{i3} = \boldsymbol{x}_{i3}'\boldsymbol{\beta}_3 + \epsilon_{i3} \tag{3}$$
$$\text{Nonlicensed - Riding}: z_{i4} = \boldsymbol{x}_{i4}'\boldsymbol{\beta}_4 + \epsilon_{i4} \tag{4}$$
$$\text{Licensed - Walking/Biking}: z_{i5} = \boldsymbol{x}_{i5}'\boldsymbol{\beta}_5 + \epsilon_{i5} \tag{5}$$
$$\text{Licensed - Public Transit}: z_{i6} = \boldsymbol{x}_{i6}'\boldsymbol{\beta}_6 + \epsilon_{i6} \tag{6}$$
$$\text{Licensed - Riding}: z_{i7} = \boldsymbol{x}_{i7}'\boldsymbol{\beta}_7 + \epsilon_{i7} \tag{7}$$
$$\text{Licensed - Driving}: z_{i8} = \boldsymbol{x}_{i8}'\boldsymbol{\beta}_8 + \epsilon_{i8}, \tag{8}$$

where, for equation $j \in \{1, \ldots, 8\}$, $\boldsymbol{x}_{ij}$ is a vector of covariates, $\boldsymbol{\beta}_j$ is a vector of coefficients, and ϵ_{ij} is an error term. Let $\mathcal{N}$ and $\mathcal{L}$ denote the nonlicensed and licensed groups, respectively. Due to the nature of the data generating process, the latent data $(z_{i5}, z_{i6}, z_{i7}, z_{i8})$ is missing for $i \in \mathcal{N}$. Likewise, (z_{i2}, z_{i3}, z_{i4}) is missing for $i \in \mathcal{L}$.

The non-missing latent data vectors $\boldsymbol{z}_{i\mathcal{N}} = (z_{i1}, z_{i2}, z_{i3}, z_{i4})$ and $\boldsymbol{z}_{i\mathcal{L}} = (z_{i1}, z_{i5}, z_{i6}, z_{i7}, z_{i8})$ relate to their discrete observed counterparts $\boldsymbol{y}_{i\mathcal{N}} = (y_{i1}, y_{i2}, y_{i3}, y_{i4})$ and $\boldsymbol{y}_{i\mathcal{L}} = (y_{i1}, y_{i5}, y_{i6}, y_{i7}, y_{i8})$ through the link functions

$$y_{i1} = 1\{z_{i1} > 0\} \quad \text{and} \quad y_{ij} = 1\{z_{ij} > 0\} + 1\{z_{ij} > 1\} \text{ for } j > 1. \tag{9}$$

Here, $1\{\cdot\}$ is the indicator function which equals one when its argument is true and zero otherwise. The link functions in (9) imply that $y_{i1} \in \{0, 1\}$ is binary, $y_{ij} \in \{0, 1, 2\}$ for $j > 1$ is ordinal, and every cutpoint is set to either zero or one.

Fixing every cutpoint identifies Equations (2) through (8), freeing up all but the top-left corner of the error vector covariance matrix (Nandram and Chen, 1996). Although restricted covariance matrices are harder to deal with than unrestricted covariance matrices, the restriction here actually simplifies both the specification and estimation of the model. To see why, suppose that the error vector $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \epsilon_{i2}, \epsilon_{i3}, \epsilon_{i4}, \epsilon_{i5}, \epsilon_{i6}, \epsilon_{i7}, \epsilon_{i8})'$ has the normal distribution $N(0, \boldsymbol{\Omega})$, where the restricted covariance matrix $\boldsymbol{\Omega}$ is given by

$$\boldsymbol{\Omega} = \begin{bmatrix} 1 & \Omega_{12} & \Omega_{13} & \Omega_{14} & \Omega_{15} & \Omega_{16} & \Omega_{17} & \Omega_{18} \\ \Omega_{21} & \Omega_{22} & \Omega_{23} & \Omega_{24} & \cdot & \cdot & \cdot & \cdot \\ \Omega_{31} & \Omega_{32} & \Omega_{33} & \Omega_{34} & \cdot & \cdot & \cdot & \cdot \\ \Omega_{41} & \Omega_{42} & \Omega_{43} & \Omega_{44} & \cdot & \cdot & \cdot & \cdot \\ \Omega_{51} & \cdot & \cdot & \cdot & \Omega_{55} & \Omega_{56} & \Omega_{57} & \Omega_{58} \\ \Omega_{61} & \cdot & \cdot & \cdot & \Omega_{65} & \Omega_{66} & \Omega_{67} & \Omega_{68} \\ \Omega_{71} & \cdot & \cdot & \cdot & \Omega_{75} & \Omega_{76} & \Omega_{77} & \Omega_{78} \\ \Omega_{81} & \cdot & \cdot & \cdot & \Omega_{85} & \Omega_{86} & \Omega_{87} & \Omega_{88} \end{bmatrix}.$$

The "1" in the top-left corner is the usual scaling restriction for the binary probit model, and the dots " $\cdot$ " represent unidentified parameters. Following Chib (2007), we do not model unidentified covariance terms. This is different from assuming that the unidentified covariance terms are zero, nor does it mean that the data is uninformative about the unidentified parameters (see Koop and Poirier (1997) and Poirier and Tobias (2003) for a discussion on the topic). "Ignoring" the unidentified terms is convenient because the identified portion of $\boldsymbol{\Omega}$ can be partitioned into $\boldsymbol{\Omega}_{\mathcal{N}}$ and $\boldsymbol{\Omega}_{\mathcal{L}}$, where

$$\boldsymbol{\Omega}_{\mathcal{N}} = \begin{bmatrix} 1 & \Omega_{12} & \Omega_{13} & \Omega_{14} \\ \Omega_{21} & \Omega_{22} & \Omega_{23} & \Omega_{24} \\ \Omega_{31} & \Omega_{32} & \Omega_{33} & \Omega_{34} \\ \Omega_{41} & \Omega_{42} & \Omega_{43} & \Omega_{44} \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Omega}_{\mathcal{L}} = \begin{bmatrix} 1 & \Omega_{15} & \Omega_{16} & \Omega_{17} & \Omega_{18} \\ \Omega_{51} & \Omega_{55} & \Omega_{56} & \Omega_{57} & \Omega_{58} \\ \Omega_{61} & \Omega_{65} & \Omega_{66} & \Omega_{67} & \Omega_{68} \\ \Omega_{71} & \Omega_{75} & \Omega_{76} & \Omega_{77} & \Omega_{78} \\ \Omega_{81} & \Omega_{85} & \Omega_{86} & \Omega_{87} & \Omega_{88} \end{bmatrix}.$$

The sole overlap between $\boldsymbol{\Omega}_{\mathcal{N}}$ and $\boldsymbol{\Omega}_{\mathcal{L}}$ is Ω_{11}, which is already set to one. The scaling restriction simplifies the estimation of the covariance matrix by making it so that the identified terms in $\boldsymbol{\Omega}$ update using data from either $\mathcal{N}$ or $\mathcal{L}$ and not both. For ease of specification, we work with $\{\boldsymbol{\Omega}_{\mathcal{N}}, \boldsymbol{\Omega}_{\mathcal{L}}\}$ instead of $\boldsymbol{\Omega}$. Group-specific error vectors can now be defined as $\boldsymbol{\epsilon}_{i\mathcal{N}} = (\epsilon_{i1}, \epsilon_{i2}, \epsilon_{i3}, \epsilon_{i4})' \sim N(0, \boldsymbol{\Omega}_{\mathcal{N}})$ and $\boldsymbol{\epsilon}_{i\mathcal{L}} = (\epsilon_{i1}, \epsilon_{i5}, \epsilon_{i6}, \epsilon_{i7}, \epsilon_{i8})' \sim N(0, \boldsymbol{\Omega}_{\mathcal{L}})$. There are several ways to deal with covariance matrices that have a single on-diagonal restriction; see McCulloch et al. (2000), Chan and Jeliazkov (2009), and Chib et al. (2009). This paper uses Chib et al. (2009) as it is more efficient than McCulloch et al. (2000) and easier to implement than Chan and Jeliazkov (2009).

As a final note, specification and estimation can be simplified in the way described above as long as mode usage is measured on an ordinal scale with at least three bins. The number of bins may vary across modes and groups. The benefit to using three bins, however, is that the cutpoints need not be sampled (Nandram and Chen, 1996; Jeliazkov et al., 2008).

2.1 Likelihood Function

Let $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \boldsymbol{\beta}_3', \boldsymbol{\beta}_4', \boldsymbol{\beta}_5', \boldsymbol{\beta}_6', \boldsymbol{\beta}_7', \boldsymbol{\beta}_8')'$ denote the coefficient vector. The selection matrices

$$\boldsymbol{S}_{\mathcal{N}} = \begin{bmatrix} I & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I & 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \boldsymbol{S}_{\mathcal{L}} = \begin{bmatrix} I & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & I & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & I \end{bmatrix}$$

select those $\boldsymbol{\beta}_j$'s in the coefficient vector pertaining to each group so that $\boldsymbol{S}_{\mathcal{N}}\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \boldsymbol{\beta}_3', \boldsymbol{\beta}_4')'$ and $\boldsymbol{S}_{\mathcal{L}}\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_5', \boldsymbol{\beta}_6', \boldsymbol{\beta}_7', \boldsymbol{\beta}_8')'$. Data matrices are in seemingly unrelated regression (SUR) form:

$$\boldsymbol{X}_{i\mathcal{N}} = \begin{bmatrix} \boldsymbol{x}_{i1}' & 0 & 0 & 0 \\ 0 & \boldsymbol{x}_{i2}' & 0 & 0 \\ 0 & 0 & \boldsymbol{x}_{i3}' & 0 \\ 0 & 0 & 0 & \boldsymbol{x}_{i4}' \end{bmatrix} \quad \text{and} \quad \boldsymbol{X}_{i\mathcal{L}} = \begin{bmatrix} \boldsymbol{x}_{i1}' & 0 & 0 & 0 & 0 \\ 0 & \boldsymbol{x}_{i5}' & 0 & 0 & 0 \\ 0 & 0 & \boldsymbol{x}_{i6}' & 0 & 0 \\ 0 & 0 & 0 & \boldsymbol{x}_{i7}' & 0 \\ 0 & 0 & 0 & 0 & \boldsymbol{x}_{i8}' \end{bmatrix}.$$

Let $g \in \{\mathcal{N}, \mathcal{L}\}$, $N_{\mathcal{N}} = |\mathcal{N}|$, and $N_{\mathcal{L}} = |\mathcal{L}|$. Define the vectors and matrices

$$\boldsymbol{z}_g = \begin{bmatrix} \boldsymbol{z}_{1g} \\ \vdots \\ \boldsymbol{z}_{N_g g} \end{bmatrix}, \quad \boldsymbol{X}_g = \begin{bmatrix} \boldsymbol{X}_{1g} \\ \vdots \\ \boldsymbol{X}_{N_g g} \end{bmatrix}, \quad \text{and} \quad \boldsymbol{\epsilon}_g = \begin{bmatrix} \boldsymbol{\epsilon}_{1g} \\ \vdots \\ \boldsymbol{\epsilon}_{N_g g} \end{bmatrix}.$$

The latent data generating process is given by

$$\boldsymbol{z}_g = \boldsymbol{X}_g \boldsymbol{S}_g \boldsymbol{\beta} + \boldsymbol{\epsilon}_g \sim N(\boldsymbol{X}_g \boldsymbol{S}_g \boldsymbol{\beta}, I_{N_g} \otimes \boldsymbol{\Omega}_g),$$

where "$\otimes$" is the Kronecker product.

For notational ease and faster runtimes, we also make use of the matrix normal representation of the latent data generating process: Let $\text{vec}_{p,q}^{-1}(\cdot)$ denote the inverse of the vectorization function $\text{vec}(\cdot)$ that takes a column vector as its input and outputs a $p \times q$ matrix. Let J_g denote the number of equations that pertain to group g so that $J_{\mathcal{N}} = 4$ and $J_{\mathcal{L}} = 5$. Define the matrices

$$\boldsymbol{Z}_g' = \text{vec}_{J_g,N_g}^{-1}(\boldsymbol{z}_g), \quad \boldsymbol{M}_g' = \text{vec}_{J_g,N_g}^{-1}(\boldsymbol{X}_g \boldsymbol{S}_g \boldsymbol{\beta}), \quad \text{and} \quad \boldsymbol{E}_g' = \text{vec}_{J_g,N_g}^{-1}(\boldsymbol{\epsilon}_g).$$

Then

$$\boldsymbol{Z}_g = \boldsymbol{M}_g + \boldsymbol{E}_g \sim MN_{N_g,J_g}(\boldsymbol{M}_g, I_{N_g} \otimes \boldsymbol{\Omega}_g),$$

where MN denotes the matrix normal distribution.

Regions of truncation are given by $\mathcal{B}_{i\mathcal{N}} = \mathcal{B}_{i1} \times \mathcal{B}_{i2} \times \mathcal{B}_{i3} \times \mathcal{B}_{i4}$ and $\mathcal{B}_{i\mathcal{L}} = \mathcal{B}_{i1} \times \mathcal{B}_{i5} \times \mathcal{B}_{i6} \times \mathcal{B}_{i7} \times \mathcal{B}_{i8}$, where

$$\mathcal{B}_{i1} = \begin{cases} (-\infty, 0] & \text{if } y_{i1} = 0 \\ (0, \infty) & \text{if } y_{i1} = 1 \end{cases} \quad \text{and} \quad \mathcal{B}_{ij} = \begin{cases} (-\infty, 0] & \text{if } y_{ij} = 0 \\ (0, 1] & \text{if } y_{ij} = 1 \\ (1, \infty) & \text{if } y_{ij} = 2 \end{cases} \quad \text{for } j > 1.$$

Let $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \boldsymbol{\Omega}_{\mathcal{N}}, \boldsymbol{\Omega}_{\mathcal{L}}\}$ denote the set of identified model parameters, and let $\mathcal{B}_g$ be the collection of truncation regions for all individuals in group g. The data-augmented likelihood is given by

$$f(\boldsymbol{y}, \boldsymbol{z}|\boldsymbol{\theta}) = \prod_g 1\{\boldsymbol{z}_g \in \mathcal{B}_g\} f_N\big(\boldsymbol{z}_g \big| \boldsymbol{X}_g \boldsymbol{S}_g \boldsymbol{\beta}, I_{N_g} \otimes \boldsymbol{\Omega}_g\big),$$

where $\boldsymbol{y}$ and $\boldsymbol{z}$ are the observed data and the non-missing latent data, respectively. Note that, because the likelihood is not augmented with missing latent data, it is not a complete-data likelihood.

2.2 Prior Distribution

The model is completed by specifying a prior distribution over $\boldsymbol{\theta}$. Let $\boldsymbol{\beta}$, $\boldsymbol{\Omega}_{\mathcal{N}}$ and $\boldsymbol{\Omega}_{\mathcal{L}}$ be independent *a priori*. As per convention, the coefficient vector has a normal prior: $\boldsymbol{\beta} \sim N(\boldsymbol{b}_0, \boldsymbol{B}_0)$. We follow Chib et al. (2009) in specifying a prior for $\boldsymbol{\Omega}_g$: Partition $\boldsymbol{\Omega}_g$ as

$$\boldsymbol{\Omega}_g = \begin{bmatrix} 1 & \boldsymbol{\Omega}_{g12} \\ \boldsymbol{\Omega}_{g21} & \boldsymbol{\Omega}_{g22} \end{bmatrix}$$

and let $\boldsymbol{\Omega}_{g22\cdot 1} = \boldsymbol{\Omega}_{g22} - \boldsymbol{\Omega}_{g21}\boldsymbol{\Omega}_{g12}$. Note that the Jacobian of the one-to-one transformation $\{\boldsymbol{\Omega}_{g21}, \boldsymbol{\Omega}_{g22}\} \to \{\boldsymbol{\Omega}_{g21}, \boldsymbol{\Omega}_{g22\cdot 1}\}$ is 1. Let $\boldsymbol{Q}_g$ be a positive definite matrix of the same size as $\boldsymbol{\Omega}_g$. We induce a prior for $\boldsymbol{\Omega}_g$ by specifying the following prior over the set of transformed parameters:

$$\boldsymbol{\Omega}_{g22\cdot 1} \sim IW(\boldsymbol{Q}_{g22\cdot 1}, \nu_g) \quad \text{and} \quad (\boldsymbol{\Omega}_{g21}|\boldsymbol{\Omega}_{g22\cdot 1}) \sim N(\boldsymbol{Q}_{g21}Q_{g11}^{-1}, \boldsymbol{\Omega}_{g22\cdot 1}Q_{g11}^{-1}).$$

Our prior is given by

$$\pi(\boldsymbol{\theta}) = f_N(\boldsymbol{\beta}|\boldsymbol{b}_0, \boldsymbol{B}_0) \prod_g f_N\Big(\boldsymbol{\Omega}_{g21}\Big|\boldsymbol{Q}_{g21}Q_{g11}^{-1}, \boldsymbol{\Omega}_{g22\cdot 1}Q_{g11}^{-1}\Big) f_{IW}(\boldsymbol{\Omega}_{g22\cdot 1}|\boldsymbol{Q}_{g22\cdot 1}, \nu_g).$$

3 Estimation

Putting together the data-augmented likelihood and prior, we get the data-augmented posterior

$$\pi(\boldsymbol{\theta}, \boldsymbol{z}|\boldsymbol{y}) \propto f(\boldsymbol{y}, \boldsymbol{z}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}).$$

Estimation and inference is facilitated via posterior Gibbs sampling (Gelfand and Smith, 1990). Our Gibbs estimation algorithm can be summarized in the following way:

Gibbs Sampling Algorithm

In each MCMC iteration:

1. Sample $\boldsymbol{\beta} \sim (\boldsymbol{\beta}|\boldsymbol{Z}, \boldsymbol{\theta} \setminus \boldsymbol{\beta})$
2. For $g \in \{\mathcal{N}, \mathcal{L}\}$:
   (a) Sample $\boldsymbol{\Omega}_{g22\cdot1} \sim (\boldsymbol{\Omega}_{g22\cdot1}|\boldsymbol{Z}_g, \boldsymbol{\beta})$
   (b) Sample $\boldsymbol{\Omega}_{g21} \sim (\boldsymbol{\Omega}_{g21}|\boldsymbol{Z}_g, \boldsymbol{\beta}, \boldsymbol{\Omega}_{g22\cdot1})$
   (c) Construct $\boldsymbol{\Omega}_g$ from $\boldsymbol{\Omega}_{g22\cdot1}$ and $\boldsymbol{\Omega}_{g21}$
3. For each $i \in \mathcal{N} \cup \mathcal{L}$:
   (a) For each $j \in \{1, 2, 3, 4\}$ for $i \in \mathcal{N}$ and $j \in \{1, 5, 6, 7, 8\}$ for $i \in \mathcal{L}$:
      i. Sample $z_{ij} \sim (z_{ij}|y_{ij}, \boldsymbol{z}_{ig} \setminus z_{ij}, \boldsymbol{\beta}, \boldsymbol{\Omega}_g)$

Note the following: First, following Chib et al. (2009), counterfactuals are marginalized out of the sampler. Collapsing the Gibbs sampler in this way simplifies prior inputs, reduces storage costs as well as computational demands, and improves mixing (Liu, 1994; Chib, 2007; Li, 2011). Second, covariance matrices are sampled in a single block, which improves the mixing of the MCMC chain. Third, latent data is sampled within three nested for-loops. The outer, middle, and inner loops correspond to the MCMC iteration, sample unit, and equation, respectively. This type of triple for-loop setup is common in the Bayesian analysis of multivariate discrete data, and is slow to implement. We describe a way to vectorize out the middle for-loop in subsection 3.3 to cut code run time. Implementation details for our Gibbs sampler are given below.

3.1 Sampling β

The posterior full conditional distribution for $\boldsymbol{\beta}$ is $N(\boldsymbol{b}, \boldsymbol{B})$, where

$$\boldsymbol{b} = \boldsymbol{B}\left(\boldsymbol{B}_0^{-1}\boldsymbol{b}_0 + \sum_g \boldsymbol{S}_g'\boldsymbol{X}_g'\text{vec}(\boldsymbol{\Omega}_g^{-1}\boldsymbol{Z}_g')\right) \quad \text{and} \quad \boldsymbol{B} = \left(\boldsymbol{B}_0^{-1} + \sum_g \boldsymbol{S}_g'\boldsymbol{X}_g'(I_{N_g} \otimes \boldsymbol{\Omega}_g^{-1})\boldsymbol{X}_g\boldsymbol{S}_g\right)^{-1}.$$

3.2 Sampling Ω_g

Let $\boldsymbol{R}_g = \boldsymbol{Q}_g + \boldsymbol{E}_g'\boldsymbol{E}_g$. Further let $\boldsymbol{R}_{g22\cdot1} = \boldsymbol{R}_{g22} - \boldsymbol{R}_{g21}\boldsymbol{R}_{g12}/R_{g11}$, where

$$\boldsymbol{R}_g = \begin{bmatrix} R_{g11} & \boldsymbol{R}_{g12} \\ \boldsymbol{R}_{g21} & \boldsymbol{R}_{g22} \end{bmatrix}.$$

To simulate $\boldsymbol{\Omega}_g$, draw $\boldsymbol{\Omega}_{g22\cdot1}$ from $IW(\boldsymbol{R}_{g22\cdot1}, \nu_g + n_g)$ and then $\boldsymbol{\Omega}_{g21}$ from $N(\boldsymbol{R}_{g21}R_{g11}^{-1}, \boldsymbol{\Omega}_{g22\cdot1}R_{g11}^{-1})$. Recover $\boldsymbol{\Omega}_g$ using the inverse transform

$$\boldsymbol{\Omega}_g = \begin{bmatrix} 1 & \boldsymbol{\Omega}_{g12} \\ \boldsymbol{\Omega}_{g21} & \boldsymbol{\Omega}_{g22\cdot1} + \boldsymbol{\Omega}_{g21}\boldsymbol{\Omega}_{g12} \end{bmatrix}.$$

3.3 Sampling z

Let $\mu_{igj\cdot\backslash j}$ and $\Omega_{gj\cdot\backslash j}$ denote the conditional mean and conditional variance of z_{ij}. The non-missing latent data $\boldsymbol{z}$ is generated from truncated univariate conditional normal distributions:

1. For $i \in \mathcal{N}$,

   i. Sample $(z_{i1}|y_{i1}, \boldsymbol{z}_{i\mathcal{N}} \setminus z_{i1}, \boldsymbol{\beta}, \boldsymbol{\Omega}_{\mathcal{N}}) \sim TN_{\mathcal{B}_{i1}}(\mu_{i\mathcal{N}1\cdot 234}, \Omega_{\mathcal{N}1\cdot 234})$

   ii. Sample $(z_{i2}|y_{i2}, \boldsymbol{z}_{i\mathcal{N}} \setminus z_{i2}, \boldsymbol{\beta}, \boldsymbol{\Omega}_{\mathcal{N}}) \sim TN_{\mathcal{B}_{i2}}(\mu_{i\mathcal{N}2\cdot 134}, \Omega_{\mathcal{N}2\cdot 134})$

   iii. Sample $(z_{i3}|y_{i3}, \boldsymbol{z}_{i\mathcal{N}} \setminus z_{i3}, \boldsymbol{\beta}, \boldsymbol{\Omega}_{\mathcal{N}}) \sim TN_{\mathcal{B}_{i3}}(\mu_{i\mathcal{N}3\cdot 124}, \Omega_{\mathcal{N}3\cdot 124})$

   iv. Sample $(z_{i4}|y_{i4}, \boldsymbol{z}_{i\mathcal{N}} \setminus z_{i4}, \boldsymbol{\beta}, \boldsymbol{\Omega}_{\mathcal{N}}) \sim TN_{\mathcal{B}_{i4}}(\mu_{i\mathcal{N}4\cdot 123}, \Omega_{\mathcal{N}4\cdot 123})$

2. For $i \in \mathcal{L}$,

   i. Sample $(z_{i1}|y_{i1}, \boldsymbol{z}_{i\mathcal{L}} \setminus z_{i1}, \boldsymbol{\beta}, \boldsymbol{\Omega}_{\mathcal{L}}) \sim TN_{\mathcal{B}_{i1}}(\mu_{i\mathcal{L}1\cdot 5678}, \Omega_{\mathcal{L}1\cdot 5678})$

   ii. Sample $(z_{i5}|y_{i5}, \boldsymbol{z}_{i\mathcal{L}} \setminus z_{i5}, \boldsymbol{\beta}, \boldsymbol{\Omega}_{\mathcal{L}}) \sim TN_{\mathcal{B}_{i5}}(\mu_{i\mathcal{L}5\cdot 1678}, \Omega_{\mathcal{L}5\cdot 1678})$

   iii. Sample $(z_{i6}|y_{i6}, \boldsymbol{z}_{i\mathcal{L}} \setminus z_{i6}, \boldsymbol{\beta}, \boldsymbol{\Omega}_{\mathcal{L}}) \sim TN_{\mathcal{B}_{i6}}(\mu_{i\mathcal{L}6\cdot 1578}, \Omega_{\mathcal{L}6\cdot 1578})$

   iv. Sample $(z_{i7}|y_{i7}, \boldsymbol{z}_{i\mathcal{L}} \setminus z_{i7}, \boldsymbol{\beta}, \boldsymbol{\Omega}_{\mathcal{L}}) \sim TN_{\mathcal{B}_{i7}}(\mu_{i\mathcal{L}7\cdot 1568}, \Omega_{\mathcal{L}7\cdot 1568})$

   v. Sample $(z_{i8}|y_{i8}, \boldsymbol{z}_{i\mathcal{L}} \setminus z_{i8}, \boldsymbol{\beta}, \boldsymbol{\Omega}_{\mathcal{L}}) \sim TN_{\mathcal{B}_{i8}}(\mu_{i\mathcal{L}8\cdot 1567}, \Omega_{\mathcal{L}8\cdot 1567})$

Sampling the latent data using a triple for-loop is slow, especially when the sample size and the number of equations are both large. To cut runtime, we vectorize out the middle for-loop using the following multivariate truncated normal sampling technique. This technique requires code that efficiently generates a column vector of inid (independent, not identically distributed) univariate truncated normals. Our code uses a modified version of Robert (1995) to sample from truncation regions that are deep in the tails of the normal distribution. Let $\boldsymbol{x} \sim TN_{[\boldsymbol{a},\boldsymbol{b}]}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$ denote a column vector of inid univariate truncated normal draws, where $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{\mu}$, and $\boldsymbol{\sigma}^2$ are lower bound, upper bound, mean, and variance vectors, respectively, such that $x_i \overset{inid}{\sim} TN_{[a_i,b_i]}(\mu_i, \sigma_i^2)$.

Let $\boldsymbol{Z}_{gj}$ denote the jth column of $\boldsymbol{Z}_g$. Let $\boldsymbol{Z}_{g\backslash j}$ denote $\boldsymbol{Z}_g$ without its jth column. Define $\boldsymbol{M}_{gj}$ and $\boldsymbol{M}_{g\backslash j}$ accordingly. Let Ω_{gjj} be the jjth element of $\boldsymbol{\Omega}_g$, $\boldsymbol{\Omega}_{g\backslash jj}$ be the jth column of $\boldsymbol{\Omega}_g$ omitting the jth row, $\boldsymbol{\Omega}_{gj\backslash j}$ be the jth row of $\boldsymbol{\Omega}_g$ omitting the jth column, and $\boldsymbol{\Omega}_{g\backslash j\backslash j}$ be $\boldsymbol{\Omega}_g$ with the jth row and jth column omitted. Let $\Omega_{gjj\cdot\backslash j} = \Omega_{gjj} - \boldsymbol{\Omega}_{gj\backslash j}\boldsymbol{\Omega}^{-1}_{g\backslash j\backslash j}\boldsymbol{\Omega}_{g\backslash jj}$. By the property of the matrix normal distribution we have that

$$\left(\boldsymbol{Z}_{gj}\middle|\boldsymbol{Z}_{g\backslash j}, \boldsymbol{\beta}, \boldsymbol{\Omega}_g\right) \sim MN_{N_g,1}\Big(\boldsymbol{M}_{gj} + \left(\boldsymbol{Z}_{g\backslash j} - \boldsymbol{M}_{g\backslash j}\right)\boldsymbol{\Omega}^{-1}_{g\backslash j\backslash j}\boldsymbol{\Omega}_{g\backslash jj}, I_{N_g}\Omega_{gjj\cdot\backslash j}\Big),$$

which is a column vector of inid univariate normal distributions. Let $\mathcal{B}_{gj}$ be the set of equation j truncation regions for all individuals in group g. The latent data z_{igj} for all sample units i can be drawn at once using

$$\left(\boldsymbol{Z}_{gj}\middle|\boldsymbol{Y}_{gj}, \boldsymbol{Z}_{g\backslash j}, \boldsymbol{\beta}, \boldsymbol{\Omega}_g\right) \sim TN_{\mathcal{B}_{gj}}\Big(\boldsymbol{M}_{gj} + \left(\boldsymbol{Z}_{g\backslash j} - \boldsymbol{M}_{g\backslash j}\right)\boldsymbol{\Omega}^{-1}_{g\backslash j\backslash j}\boldsymbol{\Omega}_{g\backslash jj}, \iota_{N_g}\Omega_{gjj\cdot\backslash j}\Big),$$

where ι_{N_g} is the $N_g \times 1$ vector of ones.

4 Cross Entropy Index for Land Use Imbalance

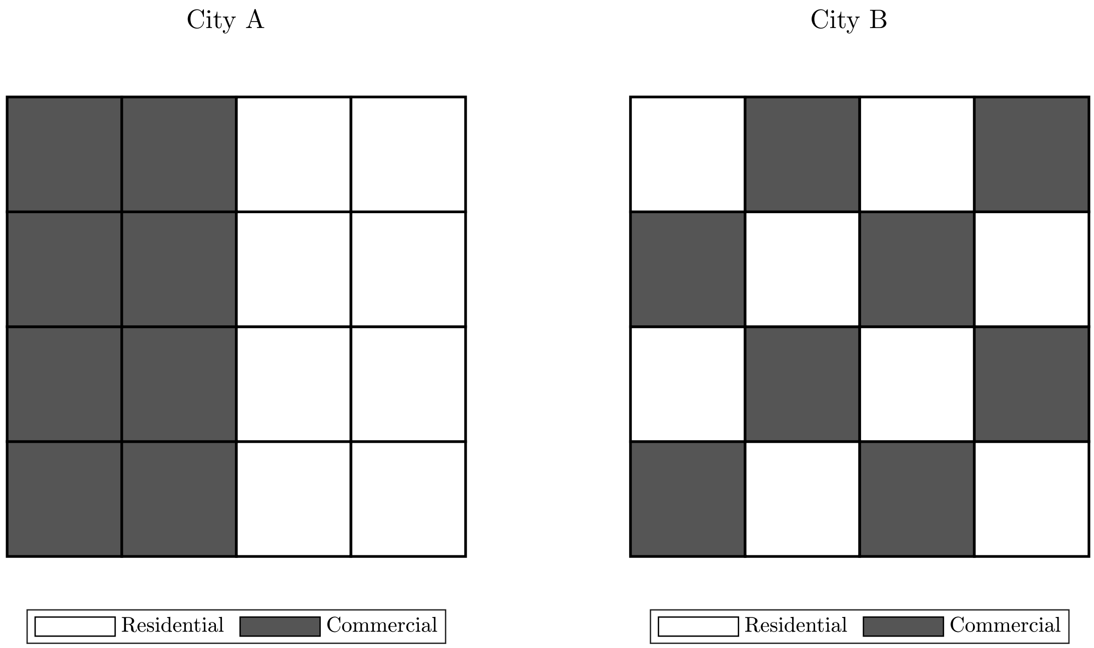
In this section, we introduce the cross entropy index for land use imbalance and recommend it as an alternative to the entropy index for land use mix/balance. Along with density, land use mix/balance is a popular urban landscape measure of accessibility to goods and services. It is typically associated with the notion of entropy from information theory, which, for a random variable X with finite support $\mathcal{X}$ and probability mass function p, is given by

$$H(p) = -\sum_{x \in \mathcal{X}} p(x) \ln p(x). \tag{10}$$

This formula can be traced back to Shannon and Weaver (1963) in the field of communication, who relate their work to Boltzmann's H-theorem from statistical mechanics. Cervero (1989) was the first in the area of transportation studies to use Equation (10) to measure land use integration. This was done by setting $\mathcal{X}$ to the set of all possible land uses in a given area and $p(x)$ to the proportion of land in said area devoted to use $x \in \mathcal{X}$. The entropy index $H^*(p) = H(p)/\ln(|\mathcal{X}|)$ in popular use today is due to Kockelman (1997), who proposed standardizing H to H^* so that H^* lies between 0 and 1, where $H^* = 1$ represents optimal mix/balance.

The entropy index, while popular, has three known issues: (i) it is not a valid measure of land use integration; (ii) it prescribes the discrete uniform distribution, $u(x) = \frac{1}{|\mathcal{X}|}$, as the optimal distribution of land uses over $\mathcal{X}$; and (iii) it is symmetric with respect to land use types (Kockelman, 1997; Mitchell Hess et al., 2001; Song et al., 2013). To see why the entropy index is not a valid measure of land use integration, consider two square cities, city A and city B, whose land areas are allocated to residential and commercial uses as shown in Figure 3. Clearly, city B has better land use mix than city A. The entropy index value is 1 for both cities, however, indicating that the two are equally (and also maximally) mixed. Evidently, the entropy index is not a measure of land use integration. Some authors attempt to rectify this problem by referring to the entropy index as a measure of land use balance rather than land use mix. Unless there is good reason to believe that the optimal distribution of land uses is the discrete uniform distribution, however, it is not a valid measure of balance either.

Figure 3



To understand the land use symmetry problem, suppose that there are two other cities C and D whose land areas are split 60%/30%/10% and 10%/30%/60%, respectively, between commercial, residential, and recreational use, in that order. One would expect city C to have better balance than city D; yet, they generate the same entropy index value. This is because land uses are treated symmetrically, i.e., land use labels can be switched around without affecting the value of the index.

Existing solutions to the aforementioned validity and land use symmetry problems involve modifying the entropy index so that balance is measured relative to some reference distribution; see, for example, Kockelman (1997) and Song et al. (2013). In the remainder of this section, we demonstrate that if a reference distribution is available, the natural solution should be to abandon balance and entropy altogether and to instead measure land use imbalance using cross entropy.

The cross entropy of probability mass function p relative to the reference distribution q is

$$D(p||q) = \sum_{x \in \mathcal{X}} p(x) \ln \frac{p(x)}{q(x)}. \tag{11}$$

Also known as relative entropy and Kullback-Leibler (KL) divergence, cross entropy has many uses in the fields of information theory and statistics. Its purpose here is to measure land use imbalance. As before, let $p(x)$ denote the proportion of land area devoted to use $x \in \mathcal{X}$. Let $q(x)$ denote the reference or optimal counterpart to $p(x)$. Then, Equation (11) quantifies the discrepancy between the actual and ideal land use distributions. Perfect balance is achieved when there is no imbalance, i.e., $D(p||q) = 0$, which occurs at $p = q$. When $p \neq q$, there is imbalance, and $D(p||q)$ takes a positive value. To match the entropy index H^*, we refer to $D^*(p||q) = D(p||q)/\ln(|\mathcal{X}|)$ as the cross entropy index.

We believe that the cross entropy index is the natural alternative to the entropy index because the two can be linked elegantly in the following fashion: For any pmf q over $\mathcal{X}$, we have

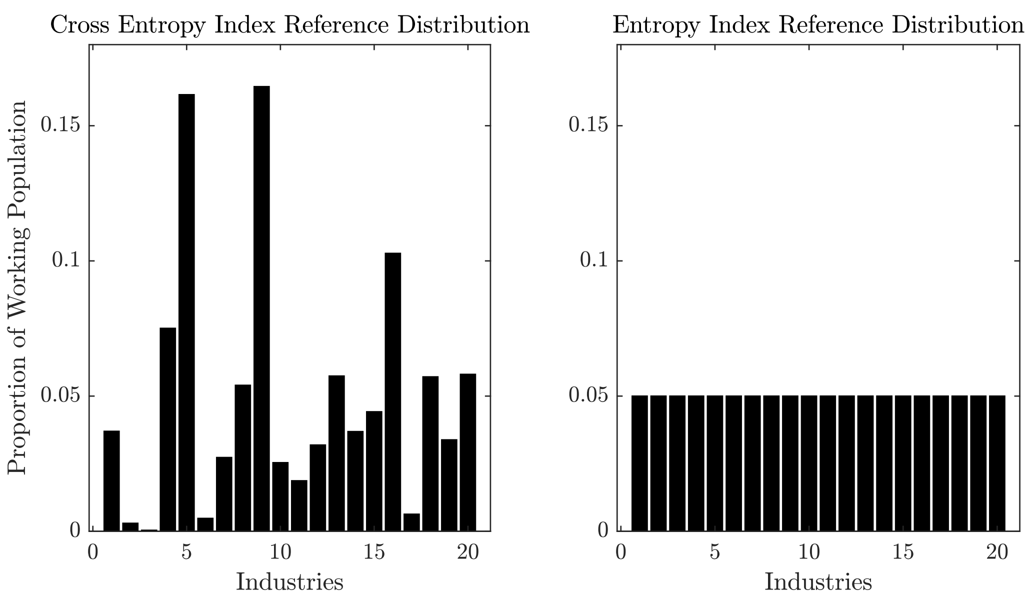
$$D^*(p||q) = -\sum_{x\in\mathcal{X}} \frac{p(x)\ln q(x)}{\ln(|\mathcal{X}|)} - H^*(p).$$

Now suppose that q is the discrete uniform distribution, u, as prescribed by the entropy index. Then, the entropy index H^* is a valid measure of balance, and the above expression simplifies to

$$D^*(p||u) = 1 - H^*(p). \tag{12}$$

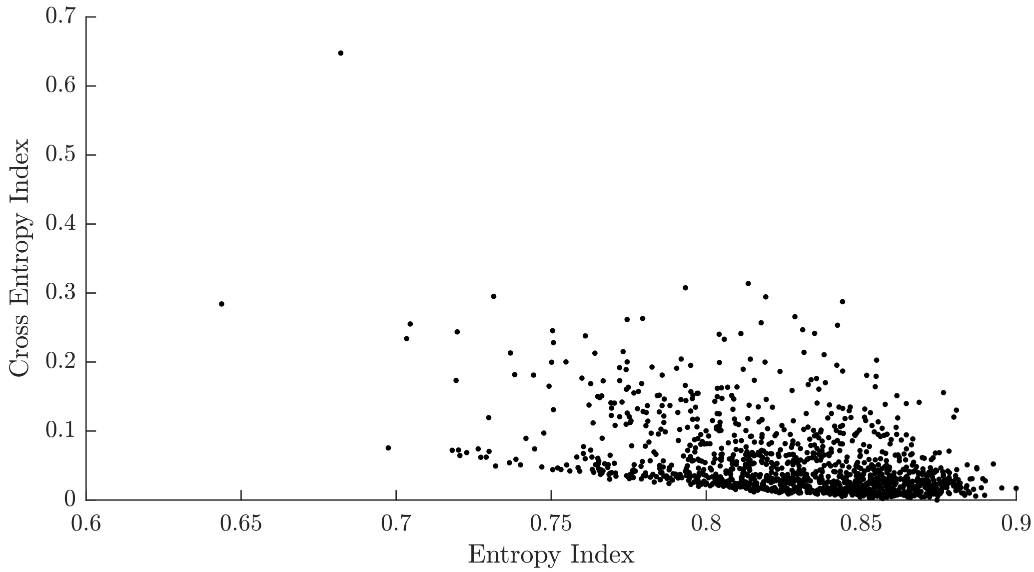
In other words, balance and imbalance sum to unity. Equation (12) shows that the uniform ideal $q \sim u$ implies $p \sim u$ achieves perfect balance, $H^*(p) = 1$, and zero imbalance, $D^*(p||u) = 0$.

Figure 4: Comparison of Reference Distributions



The cross entropy index requires a choice for q. The go-to reference distribution in the literature is the land use distribution for the greater region, such as that of the state/prefecture or country under study. Note that the reference distribution may vary from one place to another; for example, reference distributions can be region-specific to incorporate regional variation in the optimal land use distribution. For our application, we set the reference distribution to the country-level distribution. Figure 4 shows the reference distributions for the cross entropy index (left) and the entropy index (right). We proxy land use proportions with the proportion of the working population in twenty industries (see the next section for a list of these industries). A scatterplot comparing the cross entropy index D^* to the entropy index H^* for every city in Japan is given in Figure 5. Although we can visually verify that the two are inversely related, there is substantial variability around this relationship.

Figure 5: Entropy Index vs. Cross Entropy Index



5 Data

To study the transportation habits of the Japanese elderly, we collect individual and household-level data from the 5th Nationwide Person Trip Survey (NPTS), which was carried out by the Ministry of Land, Infrastructure, Transport and Tourism (MLIT) in 2010. The study surveyed 35,000 households from 70 cities (see Table 1 for a list of surveyed cities). Members from participating households kept a travel diary for one "typical" day, either a midweek day (Tuesday, Wednesday, Thursday) or a Sunday in a two-day weekend, in either October or November (MLIT, 2012). The travel diary and household survey can be found in MLIT (2012).

We restrict our sample to participants between the ages of 65 and 100, which yields a sample of size $n = 25{,}743$. Data from the 4th NPTS was used to study the relationship between the built environment and non-work trip frequency in Parady et al. (2015). City-level built environment measures are constructed using national census and geographic information system (GIS) data.

Table 1: Cities Surveyed in the 5th Nationwide Person Trip Survey

Group		Cities
Three Major Metropolitan Areas	Central	Saitama, Chiba, Tokyo, Yokohama, Kawasaki, Nagoya, Kyoto, Osaka, Kobe
	Peripheral 1	Toride, Tokorozawa, Matsudo, Inagi, Sakai, Toyonaka, Nara
	Peripheral 2	Ome, Odawara, Gifu, Toyohashi, Kasugai, Tsushima, Tokai, Yokkaichi, Kameyama, Omihachiman, Uji, Izumisano, Akashi
Regional Urban Area I (Central City Pop. $\geq$ 1M)	Central	Sapporo, Sendai, Hiroshima, Kitakyushu, Fukuoka
	Peripheral	Otaru, Chitose, Shiogama, Kure, Otake, Dazaifu
Regional Urban Area II (Central City Pop. $\geq$ 400K)	Central	Utsunomiya, Kanazawa, Shizuoka, Matsuyama, Kumamoto, Kagoshima
	Peripheral	Oyabe, Komatsu, Iwata, Soja, Isahaya, Usuki
Regional Urban Area III (Central City Pop. $<$ 400K)	Central	Hirosaki, Morioka, Koriyama, Matsue, Tokushima, Kochi
	Peripheral	Takasaki, Yamanashi, Kainan, Yasugi, Nangoku, Urasoe
Regional Area		Yuzawa, Ina, Joetsu, Nagato, Imabari, Hitoyoshi

Adapted from MLIT (2007).

5.1 Dependent Variables

Dependent variables include a binary indicator for whether or not a person is licensed and three/four ordinal mode usage measures. Mode usage measures are generated from the travel diary portion of the NPTS. Available transportation modes are walking/biking, public transit, getting a ride, and for those who can do so legally, driving. Since a trip may involve multiple modes of travel, we follow MLIT's reporting guidelines in prioritizing public transit over the use of a private vehicle, which in turn has priority over nonmotorized travel (MLIT, 2012).

Whereas Parady et al. (2015) uses count models to measure usage, this paper uses a three-bin ordinal scale, with zero representing no trips, one representing one or two trips, and two representing three or more trips. We do this for several reasons. First, multivariate ordinal data models are more flexible than traditional multivariate count data models. The former can accommodate positive correlations, negative correlations, over-dispersion, and under-dispersion. In comparison, the multivariate Poisson model requires correlations to be positive, and the multivariate Poisson-lognormal model, while capable of handling negative correlations, assumes that the data is overdispersed (Jeliazkov et al., 2008). Second, because a typical day's worth of travelling is split between three to four transportation mode categories, the vast majority of trip counts are zero, and the non-zero counts tend to be small. Count data models are not suited to deal with this kind of data. Lastly, as mentioned in Section 2, the three-bin ordinal scale is convenient for the purposes of model specification and estimation.

Table 2: Usage by Travel Mode and License Status

Walking/Biking

	No Trips	1 or 2 Trips	3+ Trips	Total
Nonlicensed	7,994 (31.05%)	2,089 (8.11%)	788 (3.06%)	10,871 (42.23%)
Licensed	12,220 (47.47%)	1,997 (7.76%)	655 (2.54%)	14,872 (57.77%)
Total	20,214 (78.52%)	4,086 (15.87%)	1,443 (5.61%)	25,743 (100.00%)

Transit

	No Trips	1 or 2 Trips	3+ Trips	Total
Nonlicensed	10,189 (39.58%)	599 (2.33%)	83 (0.32%)	10,871 (42.23%)
Licensed	14,312 (55.60%)	491 (1.91%)	69 (0.27%)	14,872 (57.77%)
Total	24,501 (95.18%)	1,090 (4.23%)	152 (0.59%)	25,743 (100.00%)

Getting a Ride

	No Trips	1 or 2 Trips	3+ Trips	Total
Nonlicensed	8,558 (33.24%)	1,624 (6.31%)	689 (2.68%)	10,871 (42.23%)
Licensed	13,187 (51.23%)	1,214 (4.72%)	471 (1.83%)	14,872 (57.77%)
Total	21,745 (84.47%)	2,838 (11.02%)	1,160 (4.51%)	25,743 (100.00%)

Drive

	No Trips	1 or 2 Trips	3+ Trips	Total
Nonlicensed	10,871 (42.23%)	0 (0.00%)	0 (0.00%)	10,871 (42.23%)
Licensed	6,662 (25.88%)	4,614 (17.92%)	3,596 (13.97%)	14,872 (57.77%)
Total	17,533 (68.11%)	4,614 (17.92%)	3,596 (13.97%)	25,743 (100.00%)

Table 2 tabulates mode usage by travel mode and license status, and offers two important insights: First, nearly half (42.23%) of the sample cannot drive, yet driving is the most popular mode option (it has the smallest no trip count). Second, those who can drive report less usage on all non-driving modes than those who cannot drive. This suggests that driving is a substitute to all other modes. Since mode usage is modeled jointly, any substitution effects unaccounted for by the included covariates is captured by the covariance matrix.

5.2 Independent Variables

Individual and household-level controls are generated from the household survey portion of the NPTS. These are age, sex, employment status, household size, homeownership status, vehicle count, and an indicator for the household having a bicycle. An indicator for weekday is also included in the mode use equations. Unlike the National Household Travel Survey (NHTS), the NPTS does not collect information on household income and household members' education levels.

Built environment features considered here are population density, land use imbalance, and accessibility to public transit. Measures for the first two are constructed using data from the 2010 national census. The proportion of land devoted to different uses is proxied by the proportion of the working population in twenty industries. These industries are: (1) agriculture and forestry; (2) fisheries; (3) mining and quarrying of stone and gravel; (4) construction; (5) manufacturing; (6) electricity, gas, heat suppy and water; (7) information and communications; (8) transport and postal activities; (9) wholesale and retail trade; (10) finance and insurance; (11) real estate and goods rental and leasing; (12) scientific research, professional and technical services; (13) accommodations, eating and drinking services; (14) living-related and personal services and amusement services; (15) education, learning support; (16) medical, health care and welfare; (17) compound services; (18) services, N.E.C.; (19) public services; and (20) miscellaneous.

Table 3: Summary Statistics

Variable	Mean	S.D.	Min	Max
Individual-level variables				
Age	73.93	7.04	65	100
Male	0.51	—	—	—
Unemployed	0.58	—	—	—
Weekday	0.48	—	—	—
Household-level variables				
Household Size	2.75	1.35	1	7
Homeowner	0.94	—	—	—
Vehicle Count	1.58	1.12	0	5
Bicycle	0.64	—	—	—
City-level variables				
Population Density (per m^2)	1.29	1.55	0.06	7.90
Land Use Imbalance	0.08	0.06	0.02	0.26
Bus Stops (per km^2)	1.83	1.38	0.26	7.27
Train Stations (per $km^2 \times 10$)	0.85	0.93	0	4.24

Accessibility to transit is measured as the number of bus stops and train stations in the city normalized by city area. Data on transit facilities can be found on the National Land Numerical Information Download Service website. Summary statistics for our independent variables are reported in Table 3.

6 Results

Regression results are based on an uninformative prior that uses the following hyperparameters: $\boldsymbol{b}_0 = 0 \times \iota$, $\boldsymbol{B}_0 = 100 \times I$, $\boldsymbol{Q}_g = I$ for $g \in \{\mathcal{N}, \mathcal{L}\}$, $\nu_{\mathcal{N}} = 6$, and $\nu_{\mathcal{L}} = 7$. The MCMC chain ran for 10,000 iterations after a burn-in of 2,000 draws. Priors and posteriors are given in the Appendix.

6.1 Results for β

Posterior means and standard deviations for the coefficient vector $\boldsymbol{\beta}$ are given in Table 4. The coefficient estimates themselves are neither interpretable nor comparable with other estimates due to the discrete/non-linear nature of the data. Effects that are decisively positive or negative (posterior sign probabilities of at least 0.975) are indicated by a star ($\star$). We comment on the direction of effects here and on policy-relevance in the next subsection.

For the elderly who cannot drive, density is positively associated with transit use and negatively associated with getting rides. For the elderly who can drive, densification leads to more walking and biking, more transit use, and less driving. Private vehicle use in general goes down (riding for those who cannot drive and driving for those who can), and the use of other transport modes go up. This suggests that the effect of density on mode choice and usage is facilitated by a substitution mechanism that involves the use of a private vehicle.

As for land use balance, we find that better balance (less imbalance) leads to more nonmotorized travel for both groups and also more transit use by the licensed group. In constrast with density, we find no evidence suggesting that substitution effects are at play here.

Bus accessibility is found to have either no effect or an indeterministic effect everywhere. In comparison, increased train accessibility leads to less driving by license holders and fewer license holders overall.

Table 4: Posterior Means and Standard Deviations for $\boldsymbol{\beta}$

		Nonlicensed			Licensed			
	License	Walking/Biking	Transit	Getting a Ride	Walking/Biking	Transit	Getting a Ride	Driving
Age	-0.105⋆	-0.056⋆	-0.045⋆	-0.045⋆	-0.011⋆	-0.027⋆	-0.009	-0.014⋆
	(0.002)	(0.004)	(0.008)	(0.005)	(0.005)	(0.006)	(0.006)	(0.003)
Male	1.885⋆	0.563⋆	0.441⋆	-0.047	-0.058	0.090	-0.902⋆	0.354⋆
	(0.023)	(0.075)	(0.149)	(0.100)	(0.076)	(0.089)	(0.097)	(0.043)
Unemployed	-0.178⋆	-0.205⋆	-0.245⋆	-0.184⋆	0.099⋆	-0.178⋆	-0.062	-0.144⋆
	(0.022)	(0.032)	(0.048)	(0.039)	(0.031)	(0.048)	(0.042)	(0.023)
Weekday		0.271⋆	0.155⋆	-0.020	0.026	0.108⋆	-0.221⋆	0.329⋆
		(0.029)	(0.041)	(0.035)	(0.029)	(0.044)	(0.040)	(0.022)
Household Size	-0.469⋆	-0.063⋆	-0.012	-0.171⋆	0.045⋆	0.048	-0.004	-0.085⋆
	(0.011)	(0.021)	(0.038)	(0.027)	(0.022)	(0.029)	(0.029)	(0.014)
Homeowner	0.348⋆	0.111⋆	0.187⋆	0.445⋆	0.020	0.067	0.059	0.027
	(0.046)	(0.056)	(0.078)	(0.077)	(0.072)	(0.101)	(0.100)	(0.054)
Vehicle Count	0.809⋆	-0.014	-0.204⋆	0.306⋆	-0.208⋆	-0.274⋆	0.001	0.106⋆
	(0.015)	(0.033)	(0.063)	(0.041)	(0.031)	(0.041)	(0.041)	(0.018)
Bicycle	-0.262⋆	0.402⋆	-0.096	-0.172⋆	0.375⋆	-0.072	0.025	0.013
	(0.023)	(0.035)	(0.050)	(0.040)	(0.033)	(0.048)	(0.042)	(0.023)
Population Density (per m^2)	0.025	0.019	0.074⋆	-0.086⋆	0.052⋆	0.072⋆	-0.001	-0.042⋆
	(0.014)	(0.020)	(0.025)	(0.026)	(0.020)	(0.026)	(0.029)	(0.016)
Land Use Imbalance	-0.143	-1.602⋆	0.333	0.524	-1.267⋆	-2.473⋆	-0.528	0.154
	(0.193)	(0.276)	(0.406)	(0.316)	(0.278)	(0.540)	(0.380)	(0.200)
Bus Stops (per km^2)	0.012	-0.003	0.021	0.026	-0.021	0.046	0.000	-0.021
	(0.013)	(0.018)	(0.023)	(0.025)	(0.019)	(0.025)	(0.028)	(0.015)
Train Stations (per $km^2 \times 10$)	-0.066⋆	-0.008	-0.006	0.008	0.041	0.097⋆	0.006	-0.048⋆
	(0.021)	(0.030)	(0.041)	(0.037)	(0.029)	(0.041)	(0.040)	(0.022)

Star (⋆) indicates posterior sign probabilities that are > 0.975.

6.2 Policy Simulation Results

Regression tables for generalized linear models (GLMs) are useful in that they convey information about statistical significance and the direction of effects. They are not in themselves useful for policymaking purposes, however, as they provide no information on effect sizes (Brownstone, 2008). In this section, we discuss two popular covariate effect estimation methods, the partial effect at the average (PEA) and the average partial effect (APE). We then describe a third estimation method that overcomes the shortcomings of the first two. The third method is used to determine whether urban form effect sizes are sufficiently large to be considered policy relevant.

To motivate the discussion, suppose that there is an exogenous urban form shock S such that $S : \boldsymbol{x}_i^{\text{pre}} \to \boldsymbol{x}_i^{\text{post}}$ (i.e., S transforms the pre-shock covariate vector $\boldsymbol{x}_i^{\text{pre}}$ into its post-shock form $\boldsymbol{x}_i^{\text{post}}$) for all i. Note that S can represent the built environment changing in numerous ways at once (e.g., population density doubling and land use imbalance halving at the same time). Further suppose we want to estimate the overall effect that S has on the probability of being licensed ($y_{i1} = 1$). Given $\boldsymbol{\beta}_1$, the effect of S for subject i is

$$(\text{Effect}_i|\boldsymbol{\beta}_1) = P(y_{i1} = 1|\boldsymbol{\beta}_1, \boldsymbol{x}_{i1}^{\text{post}}) - P(y_{i1} = 1|\boldsymbol{\beta}_1, \boldsymbol{x}_{i1}^{\text{pre}}) = \Phi(\boldsymbol{x}_{i1}^{\text{post}\prime}\boldsymbol{\beta}_1) - \Phi(\boldsymbol{x}_{i1}^{\text{pre}\prime}\boldsymbol{\beta}_1). \quad (13)$$

Note that, because the effect of S depends on the covariates, it varies from person to person.

Let $\overline{\boldsymbol{x}_1^{\text{pre}}}$ and $\overline{\boldsymbol{x}_1^{\text{post}}}$ denote the sample means of $\boldsymbol{x}_{i1}^{\text{pre}}$ and $\boldsymbol{x}_{i1}^{\text{post}}$, respectively, and let $\hat{\boldsymbol{\beta}}_1$ be a point estimate for $\boldsymbol{\beta}_1$. One way to find an overall effect based on Equation (13) is to compute the partial effect at the average:

$$\widehat{\text{Effect}} = \Phi\left(\overline{\boldsymbol{x}_1^{\text{post}}}'\hat{\boldsymbol{\beta}}_1\right) - \Phi\left(\overline{\boldsymbol{x}_1^{\text{pre}}}'\hat{\boldsymbol{\beta}}_1\right).$$

The problem with finding the effect at the average is that the average covariate vector typically corresponds to an uninteresting, unrepresentative, and/or nonexistent individual. A better estimator for the overall effect is the average partial effect, which takes the average of individual effects:

$$\widehat{\text{Effect}} = \overline{\Phi(\boldsymbol{x}_1^{\text{post}\prime}\hat{\boldsymbol{\beta}}_1)} - \overline{\Phi(\boldsymbol{x}_1^{\text{pre}\prime}\hat{\boldsymbol{\beta}}_1)}.$$

Neither PEA nor APE account for estimation uncertainty, however, and this can yield misleading effect estimates (Jeliazkov and Vossmeyer, 2016). Moreover, statistical software packages typically do not automatically give confidence intervals for the APE and PEA estimators, and policy relevance cannot be judged on point estimates alone.

To account for both data variability and estimation uncertainty, we follow Fang (2008) and Brownstone and Fang (2014) in examining the posterior distribution of the average treatment effect (ATE), where the average is taken with respect to the sample:

$$(\text{Effect}|\boldsymbol{\beta}_1) = \frac{1}{N}\sum_{i=1}^{N}(\text{Effect}_i|\boldsymbol{\beta}_1) = \overline{\Phi(\boldsymbol{x}_1^{\text{post}\prime}\boldsymbol{\beta}_1)} - \overline{\Phi(\boldsymbol{x}_1^{\text{pre}\prime}\boldsymbol{\beta}_1)}.$$

The ATE formula is similar to that of APE, the difference being that the former is evaluated at every $\boldsymbol{\beta}_1$ drawn from the posterior distribution $\pi(\boldsymbol{\beta}_1|\boldsymbol{y})$. The non-Bayesian analog to this is to take draws of $\hat{\boldsymbol{\beta}}_1$ from its sampling distribution and plug them into the APE formula.

Extending the ATE approach to the multivariate case requires only straightforward modifications. For example, suppose we are interested in studying the effect of S on the joint probability of not having a driver's license ($y_{i1} = 0$) and taking zero nonmotorized trips ($y_{i2} = 0$). Then, given $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \Omega_{12}, \Omega_{22})$, the effect of S for subject i is

$$\begin{aligned}
(\text{Effect}_i|\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \Omega_{12}, \Omega_{22}) &= P(y_{i1} = y_{i2} = 0|\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \Omega_{12}, \Omega_{22}, \boldsymbol{x}_{i1}^{\text{post}}, \boldsymbol{x}_{i2}^{\text{post}}) \\
&\quad - P(y_{i1} = y_{i2} = 0|\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \Omega_{12}, \Omega_{22}, \boldsymbol{x}_{i1}^{\text{pre}}, \boldsymbol{x}_{i2}^{\text{pre}}) \\
&= \Phi\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix} \middle| \begin{bmatrix} \boldsymbol{x}_{i1}^{\text{post}\prime}\boldsymbol{\beta}_1 \\ \boldsymbol{x}_{i2}^{\text{post}\prime}\boldsymbol{\beta}_2 \end{bmatrix}, \begin{bmatrix} 1 & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix}\right) - \Phi\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix} \middle| \begin{bmatrix} \boldsymbol{x}_{i1}^{\text{pre}\prime}\boldsymbol{\beta}_1 \\ \boldsymbol{x}_{i2}^{\text{pre}\prime}\boldsymbol{\beta}_2 \end{bmatrix}, \begin{bmatrix} 1 & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix}\right).
\end{aligned}$$

We can obtain the posterior distribution for the ATE by plugging draws of $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \Omega_{12}, \Omega_{22})$ from its posterior $\pi(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \Omega_{12}, \Omega_{22}|\boldsymbol{y})$ into the expression above.

The primary disadvantage with the method we employ is that it is computationally demanding relative to the PEA and APE methods. To cut code runtime, we use a thinned posterior sample of the model parameters.

Table 5 reports probability changes following the simultaneous doubling of population density and the halving of land use imbalance. Effect sizes are estimated precisely and found to be very small. This suggests that urban planning measures have virtually no effect at changing the transportation habits of the Japanese elderly. Our results and conclusions are broadly consistent with those based on the United States (Bento et al., 2005; Fang, 2008; Ewing and Cervero, 2010; Brownstone and Fang, 2014).

Table 5: Effect of Simultaneously Doubling Density and Halving Land Use Imbalance

Walking/Biking	No Trips	1 or 2 Trips	3+ Trips
Nonlicensed	-0.015 (0.004)	0.003 (0.002)	0.003 (0.002)
Licensed	-0.009 (0.005)	0.011 (0.003)	0.007 (0.002)
Transit			
Nonlicensed	-0.013 (0.005)	0.004 (0.002)	0.001 (0.001)
Licensed	-0.003 (0.005)	0.008 (0.002)	0.003 (0.001)
Getting a Ride			
Nonlicensed	0.004 (0.004)	-0.007 (0.002)	-0.005 (0.001)
Licensed	0.006 (0.005)	0.002 (0.002)	0.001 (0.001)
Driving			
Nonlicensed	—	—	—
Licensed	0.015 (0.005)	-0.001 (0.002)	-0.006 (0.003)

Posterior standard deviations given in parentheses.

6.3 Results for Ω

A benefit of modeling mode usage jointly rather than separately is that substitution effects not accounted for by the covariates in the model appear in $\mathbf{\Omega}$. To study how mode usage relate to each other in the error term, we find the percentage of posterior draws that are negative (indicating substitution) for select covariance terms. These are reported in Table 6. Our findings are as follows: Controlling for covariates, nonmotorized travel and getting rides are viewed as substitutes. The elderly who drive view transit as complementary to both nonmotorized travel and getting rides. Driving is also viewed as a substitute to all other mode options even after accounting for the included covariates.

Table 6: Posterior Probability of $\Omega_{jk} < 0$ in $\mathbf{\Omega}$

Nonlicensed

	Transit	Ride
Nonmotorized	38%	100%
Transit		26%

Licensed

	Transit	Ride	Drive
Nonmotorized	0%	98%	100%
Transit		0%	100%
Ride			100%

7 Conclusion

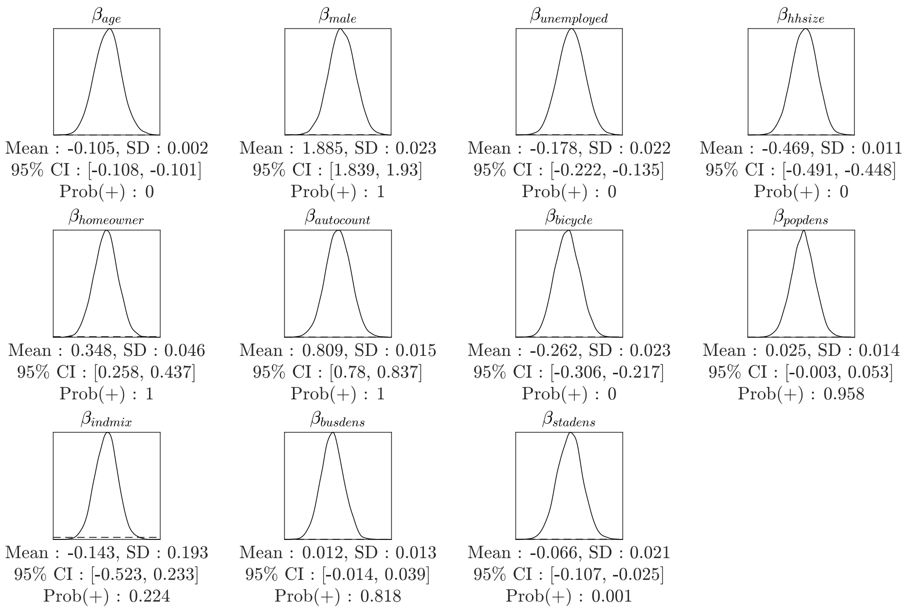
This paper presents an econometric framework to study the effects of the built environment on transportation mode choice and usage when a large fraction of the population under study is nonlicensed. We use a multivariate ordinal outcomes model with a binary selection mechanism to allow for both heterogeneous and indirect urban form effects. Emphasis was placed on the joint modeling of correlated discrete outcomes, strategizing with identification restrictions and the lack of identification, and the efficient estimation of model parameters. We also discussed the computationally efficient sampling of truncated multivariate normal latent data.

This paper introduces the cross entropy index for land use imbalance and recommends it as an alternative to the entropy index for land use mix/balance. Whereas entropy imposes the assumption that the optimal land use distribution is given by the uniform distribution, cross entropy gives the researcher the ability to choose a reference distribution. We show that the two indices are connected in that if the uniform distribution is optimal, then the entropy index is a valid measure of balance, and furthermore balance and imbalance sum to unity. The topic of reference distribution selection is not pursued here and is open to future research.

We apply our model on a sample of Japanese elderly to investigate whether urban planning tools can be used to improve traffic safety conditions in Japan. Although we are successful at identifying heterogeneous and indirect urban form effects, ultimately we find that built environment effects are too small to warrant attention from policymakers. Our conclusion is similar to those based on the United States, where built environment effects are reportedly nonzero but economically irrelevant.

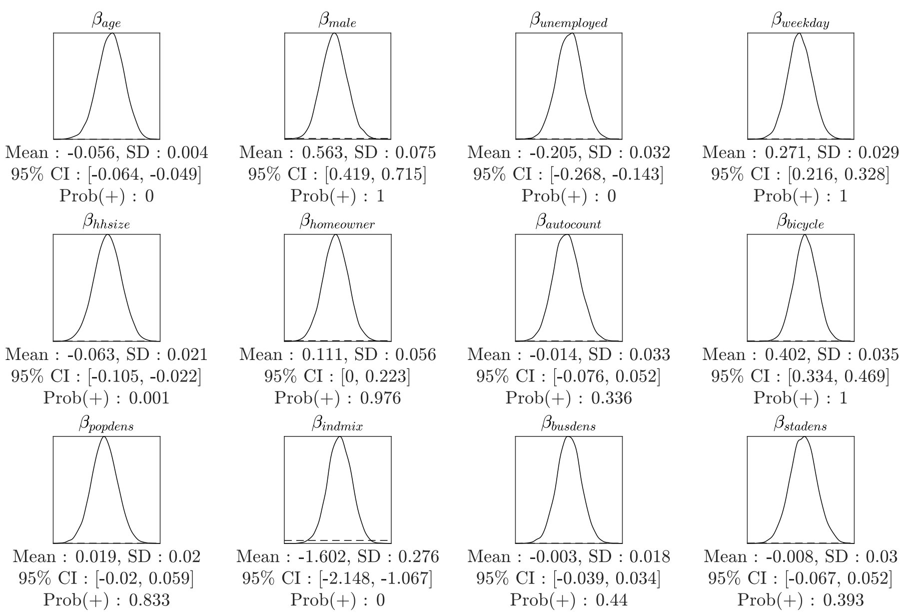
8 Appendix

Figure 6: Posteriors for Equation 1 (License)



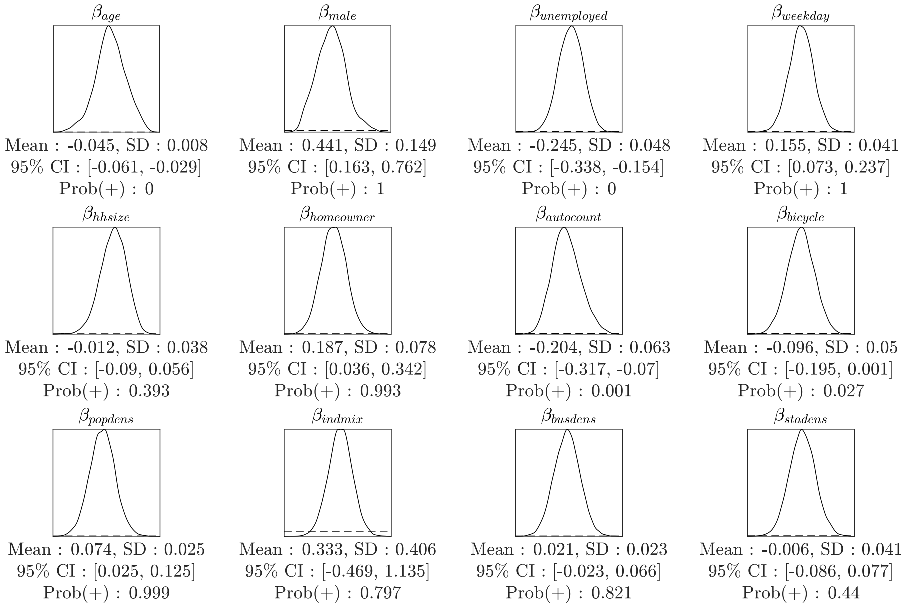
Posterior given by solid line, prior by dashed line.

Figure 7: Posteriors for Equation 2 (Nonlicensed - Walking/Biking)



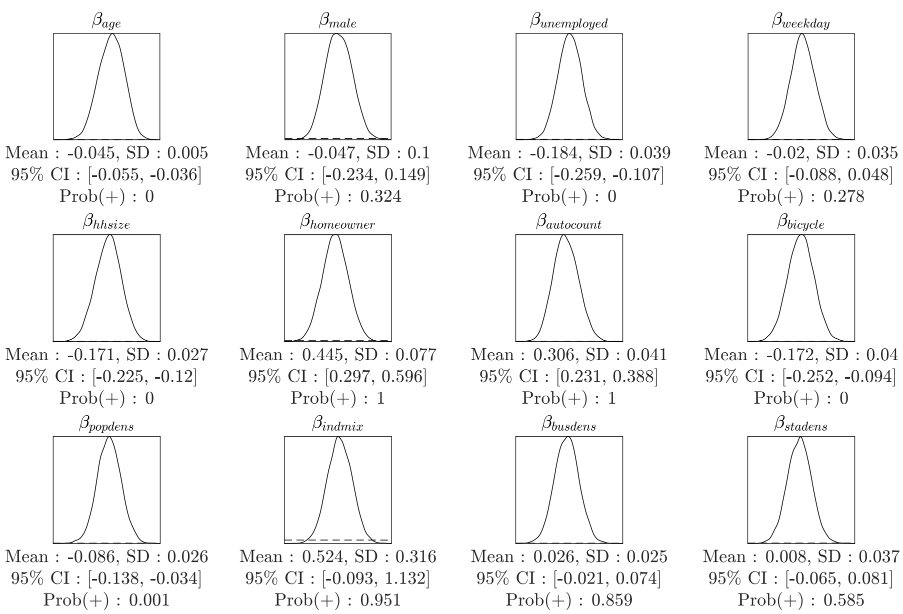
Posterior given by solid line, prior by dashed line.

Figure 8: Posteriors for Equation 3 (Nonlicensed - Transit)



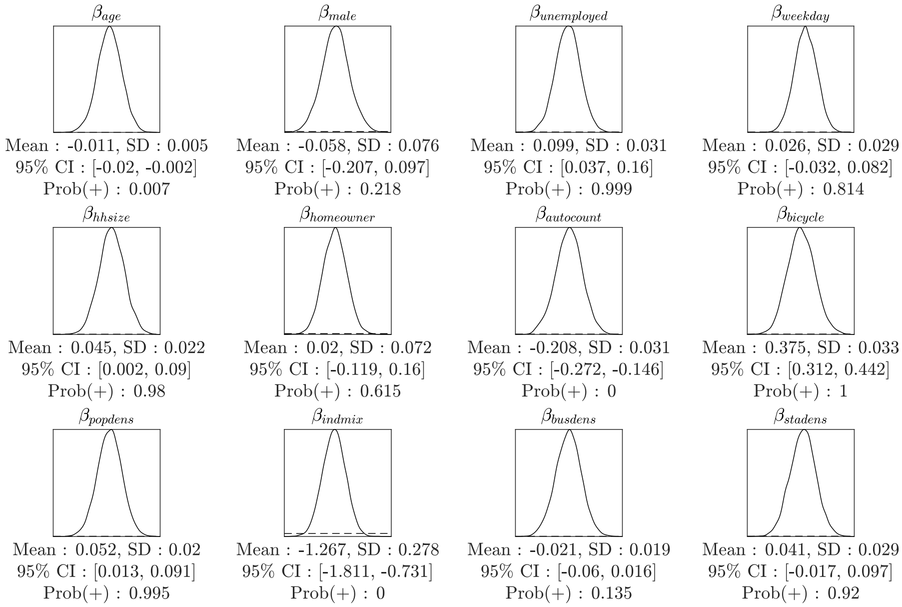
Posterior given by solid line, prior by dashed line.

Figure 9: Posteriors for Equation 4 (Nonlicensed - Ride)



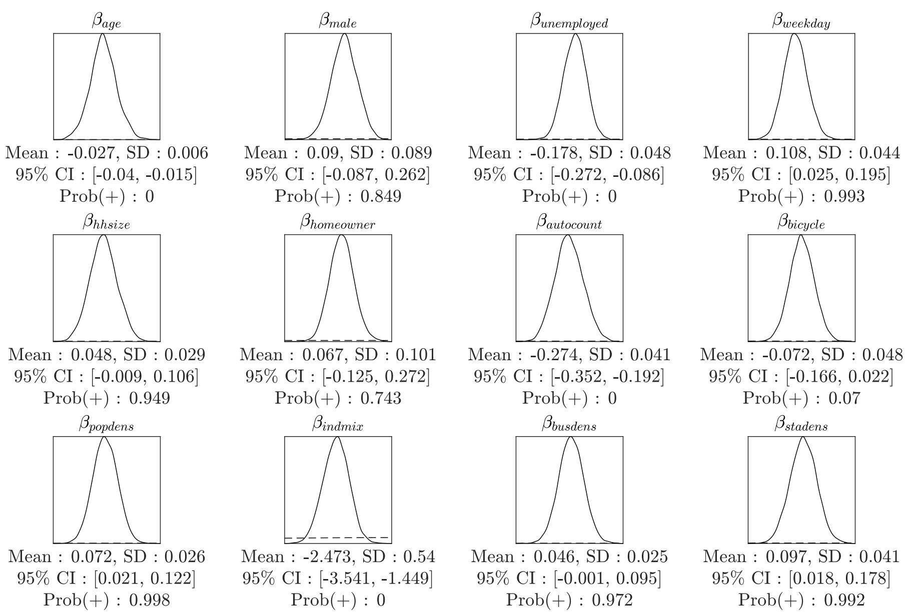
Posterior given by solid line, prior by dashed line.

Figure 10: Posteriors for Equation 5 (Licensed - Walking/Biking)



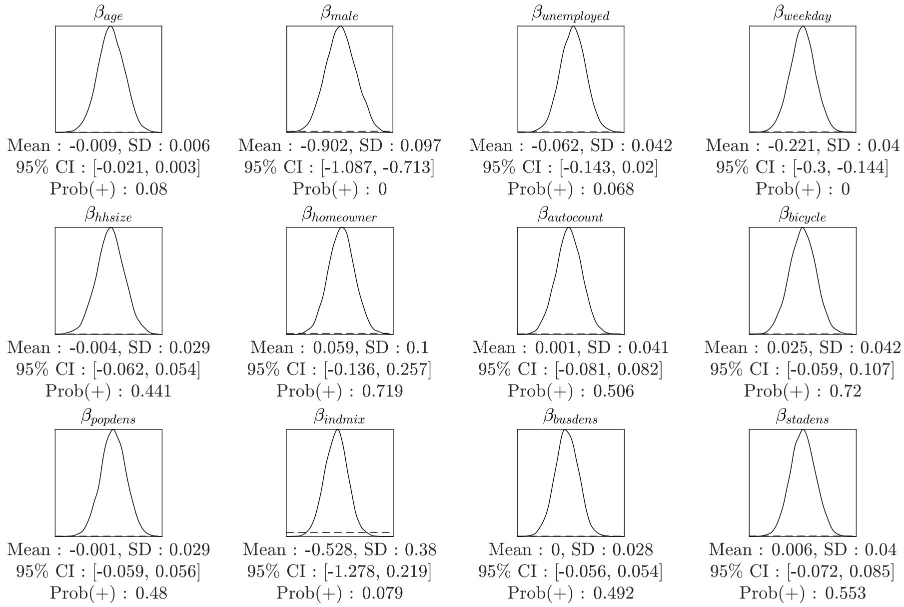
Posterior given by solid line, prior by dashed line.

Figure 11: Posteriors for Equation 6 (Licensed - Transit)



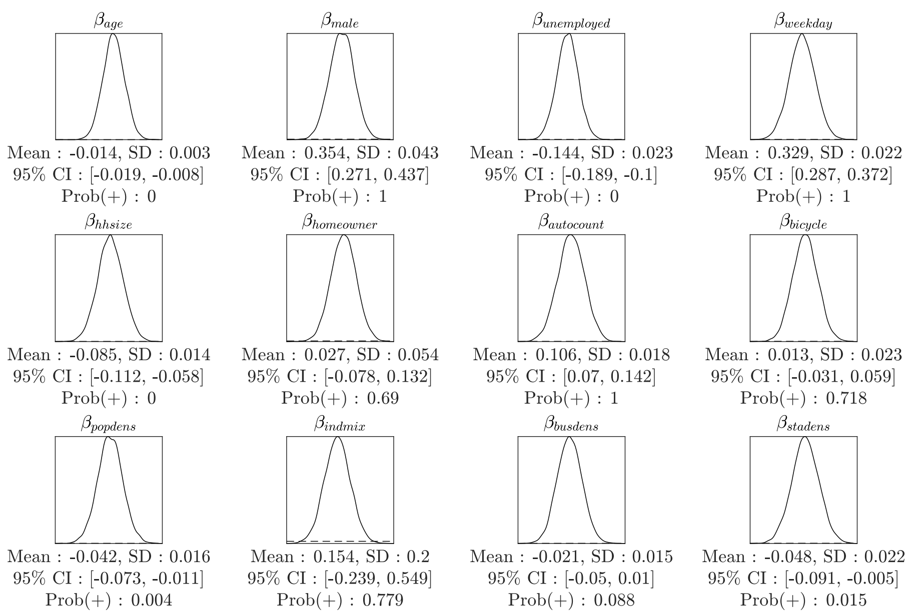
Posterior given by solid line, prior by dashed line.

Figure 12: Posteriors for Equation 7 (Licensed - Ride)



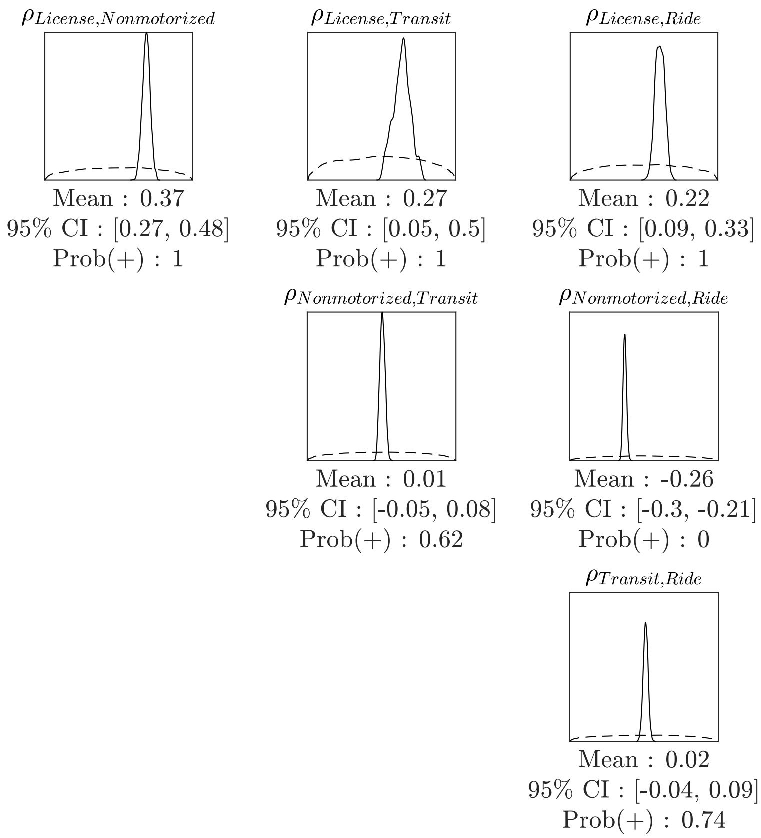
Posterior given by solid line, prior by dashed line.

Figure 13: Posteriors for Equation 8 (Licensed - Drive)



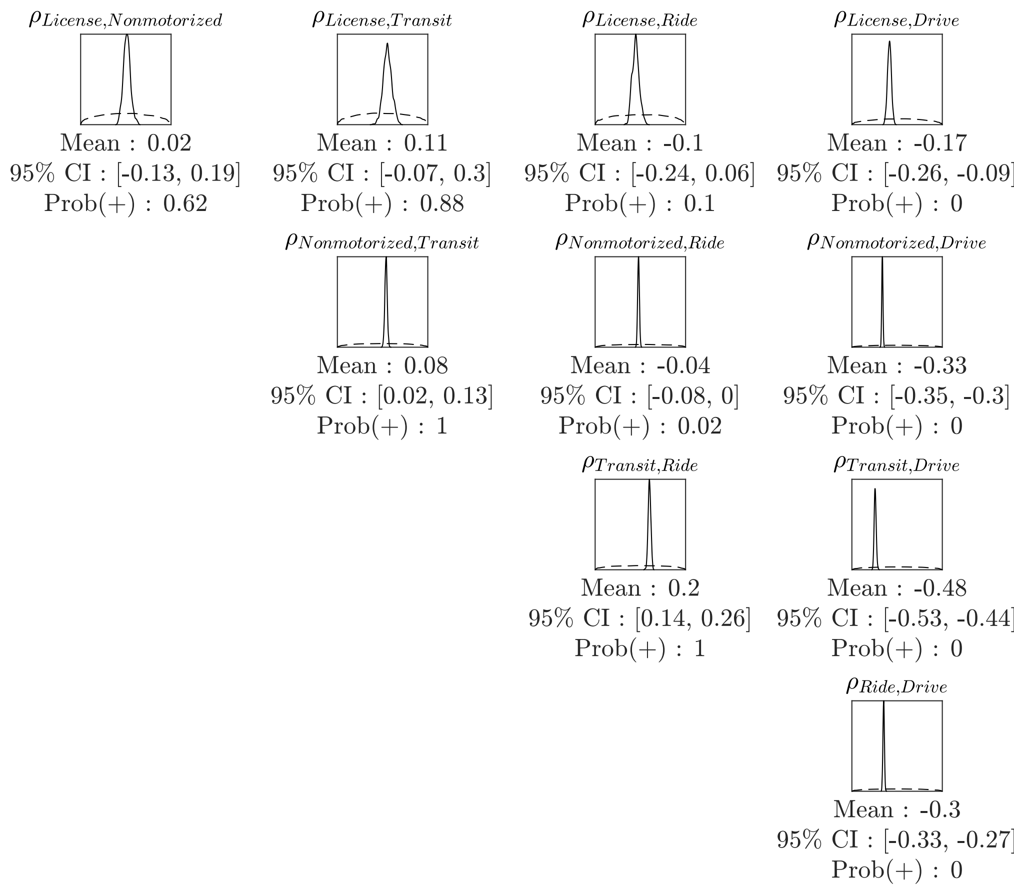
Posterior given by solid line, prior by dashed line.

Figure 14: Posteriors for Implied Correlation Matrix (Nonlicensed)



Posterior given by solid line, prior by dashed line.

Figure 15: Posteriors for Implied Correlation Matrix (Licensed)



Posterior given by solid line, prior by dashed line.

References

Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88(422):669–679.

Bento, A. M., Cropper, M. L., Mobarak, A. M., and Vinha, K. (2005). The effects of urban spatial structure on travel demand in the United States. *Review of Economics and Statistics*, 87(3):466–478.

Brownstone, D. (2008). Key relationships between the built environment and VMT. special report 298: Driving and the built enviroment: The effects of compact development on motorized travel, energy use, and CO2 emissions. Irvine, CA, committee on the relationships among development patterns, vehicle miles traveled, and energy consumption. *Transportation Research Board and the Division on Engineering and Physical Sciences*.

Brownstone, D. and Fang, H. (2014). A vehicle ownership and utilization choice model with endogenous residential density. *Journal of Transport and Land Use*, 7(2):135–151.

Cervero, R. (1989). *America's suburban centers*. Routledge.

Chan, J. C.-C. and Jeliazkov, I. (2009). MCMC estimation of restricted covariance matrices. *Journal of Computational and Graphical Statistics*, 18(2):457–480.

Chib, S. (2007). Analysis of treatment response data without the joint distribution of potential outcomes. *Journal of Econometrics*, 140(2):401–412.

Chib, S., Greenberg, E., and Jeliazkov, I. (2009). Estimation of semiparametric models in the presence of endogeneity and sample selection. *Journal of Computational and Graphical Statistics*, 18(2):321–348.

Ewing, R. and Cervero, R. (2010). Travel and the built environment: a meta-analysis. *Journal of the American Planning Association*, 76(3):265–294.

Fang, H. A. (2008). A discrete–continuous model of households' vehicle choice and usage, with an application to the effects of residential density. *Transportation Research Part B: Methodological*, 42(9):736–758.

Gelfand, A. E. and Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association*, 85(410):398–409.

Jeliazkov, I., Graves, J., and Kutzbach, M. (2008). Fitting and comparison of models for multivariate ordinal outcomes. In *Bayesian Econometrics*, pages 115–156. Emerald Group Publishing Limited.

Jeliazkov, I. and Vossmeyer, A. (2016). The impact of estimation uncertainty on covariate effects in nonlinear models. *Statistical Papers*, pages 1–12.

Kockelman, K. (1997). Travel behavior as function of accessibility, land use mixing, and land use balance: evidence from San Francisco Bay Area. *Transportation Research Record: Journal of the Transportation Research Board*, (1607):116–125.

Koop, G. and Poirier, D. J. (1997). Learning about the across-regime correlation in switching regression models. *Journal of Econometrics*, 78(2):217–227.

Leck, E. (2006). The impact of urban form on travel behavior: A meta-analysis. *Berkeley Planning Journal*, 19(1).

Li, P. (2011). Estimation of sample selection models with two selection mechanisms. *Computational Statistics & Data Analysis*, 55(2):1099–1108.

Liu, J. S. (1994). The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association*, 89(427):958–966.

McCulloch, R. E., Polson, N. G., and Rossi, P. E. (2000). A Bayesian analysis of the multinomial probit model with fully identified parameters. *Journal of Econometrics*, 99(1):173–193.

Mitchell Hess, P., Vernez Moudon, A., and Logsdon, M. (2001). Measuring land use patterns for transportation research. *Transportation Research Record: Journal of the Transportation Research Board*, (1780):17–24.

MLIT (2007). Results from the 4th Nationwide Person Trip Survey (press release).

MLIT (2012). Toshi ni okeru hito no ugoki -heisei 22nen zenkoku toshi koutsuu tokusei chousa shuukei kekka kara-.

Nandram, B. and Chen, M.-H. (1996). Reparameterizing the generalized linear model to accelerate Gibbs sampler convergence. *Journal of Statistical Computation and Simulation*, 54(1-3):129–144.

NPATB (2018). Heisei 29 nen ni okeru koutsuu shibou jiko no tokuchou nado ni tsuite.

Parady, G. T., Chikaraishi, M., Takami, K., Ohmori, N., and Harata, N. (2015). On the effect of the built environment and preferences on non-work travel: Evidence from Japan. *European Journal of Transport & Infrastructure Research*, 15(1).

Poirier, D. J. and Tobias, J. L. (2003). On the predictive distributions of outcome gains in the presence of an unidentified parameter. *Journal of Business & Economic Statistics*, 21(2):258–268.

Robert, C. P. (1995). Simulation of truncated normal variables. *Statistics and Computing*, 5(2):121–125.

Shannon, C. E. and Weaver, W. (1963). The mathematical theory of communication. 1949. *Urbana, IL: University of Illinois Press.*

Song, Y., Merlin, L., and Rodriguez, D. (2013). Comparing measures of urban land use mix. *Computers, Environment and Urban Systems*, 42:1–13.