A new treatment effect model: the analysis of the rapid railroad network

Koji Miyawaki *

July 31, 2012

Abstract

This article proposes a new treatment effect model and applies it to analyze the effect of a rapid railroad network on the population density. When we evaluate a large infrastructure, such as a rapid railroad network, by using the typical treatment effect model, the construction is considered as the treatment, and it usually takes time until the construction has finished. To address such an issue, this article extends the Roy model, which is usually used for the treatment effect model, and, to the model, incorporates individuals who are treated but their treatment has not finished. Proposed model is applied to evaluate the impact of the Shinkansen network on the population density.

Key words: Treatment effect; Rapid railroad network; Bayesian analysis.

JEL classification: C21, R53.

*School of Economics, Kwansei Gakuin University, Hyogo 662-8501, Japan. Phone:+81-798-54-6255. E-mail:miyawaki.koji@kwansei.ac.jp.

1 Introduction

A treatment, such as a medical experiment and an economic policy, is generally evaluated by the Rubin causal model (see, e.g., Rubin (1978) for its Bayesian approach). In this model, each individual is assigned two potential outcomes and one of them is observed according to its treatment status, that is, in the treatment arm or in the control arm. When a treatment takes time to be complete, we observe individuals who are treated but their treatment is not complete. Such individuals are neither in the treatment arm nor in the control arm. In this situation, the typical treatment effect model is not simply applicable because of these individuals. Such a situation occurs when we evaluate a large infrastructure, such as a rapid railroad network.

To address the situation, this article extends the so-called Roy model to allow for such individuals who are neither in the treatment arm nor in the control arm. The Roy model is a flexible and typical treatment effect model and its Bayesian estimation method is proposed by Chib and Hamilton (2000). (See also Poirier and Tobias (2003) for the identification of the variance covariance matrix of this model.) The original Roy model includes individuals who are treated and those who are not treated. In addition to these kinds of individuals, we introduce individuals who are treated but their treatment is not complete under a reasonable assumption. Because the estimation of model parameters requires high dimensional integration, we take the Bayesian approach and apply the Markov chain Monte Carlo method to conduct inferences on model parameters.

Proposed model is applied to evaluate the impact of the rapid railroad network (the Shinkansen network) on the population density. The rapid railroad network is one of large infrastructures and its construction usually take long time, which implies that our framework is applicable. The economic impact of the infrastructure is recently discussed by, for example, Baum-Snow (2007), Donaldson (2010), and Faber (2012). In this article, we focus on the population density as an economic impact. It is often observed that the population density decays as the distance to the business district becomes longer. Such a relationship is modeled as the population density function (see, e.g., McDonald (1989) for a survey of the estimation of the population density function). This article analyzes how the population density function is influenced by the introduction of the rapid railroad network.

This article is organized as follows. Section 2 describes the Roy model extended to individuals who are treated but their treatment has not finished. The estimation method of this model is presented in Section 3. Then, we apply the proposed model and its estimation method to evaluate the rapid railroad network in Section 4. Section 5 concludes this article.

2 Extended Roy model

There are n individuals. Let y_i and x_i be the observed response and treatment status of the individual i, respectively ($i = 1, \ldots, n$). If the individual i is in the treatment arm, $x_i = 1$. If it is in the control arm, $x_i = 0$. Potential outcomes for y_i are y^*_{0i} and y^*_{1i}. The treatment is assigned by the sign of the latent variable x^*_i, that is, $x_i = 1$ if $x^*_i > 0$ and $x_i = 0$ otherwise. Further, we introduce another variable c_i to indicate whether the treatment of the individual i is complete or not. If it is complete, $c_i = 1$. If it is not, $c_i = 0$. This c_i is observable. If we assume that the effect of the treatment appears monotonically, we have the following model, which is given by

$$y_i \in \begin{cases} \{y^*_{0i}\}, & \text{if } x_i = 0, \\ (y^*_{Li}, y^*_{Ui}), & \text{if } x_i = 1, c_i = 0, \\ \{y^*_{1i}\}, & \text{if } x_i = 1, c_i = 1, \end{cases} \tag{1}$$

where $y^*_{Li} = \min(y^*_{0i}, y^*_{1i})$ and $y^*_{Ui} = \max(y^*_{0i}, y^*_{1i})$.

The unobserved variables $(y^*_{0i}, y^*_{1i}, x^*_i)$ are modeled by the linear regression. More precisely,

$$y^*_{ji} = \boldsymbol{w}'_{ji}\boldsymbol{\beta}_j + \epsilon_{ji}, \quad (j = 0, 1), \tag{2}$$

$$x_i^* = \boldsymbol{v}_i' \boldsymbol{\gamma} + \eta_i, \tag{3}$$

where $\boldsymbol{w}_{ji}$ and $\boldsymbol{v}_i$ are the k_b and k_g dimensional vectors of explanatory variables, respectively with $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ their respective coefficients vectors. The $(\epsilon_{0i}, \epsilon_{1i}, \eta_i)$ are error terms that jointly follow the multivariate normal distribution with mean $\mathbf{0}$ and variance covariance matrix $\kappa_i^{-1}\boldsymbol{\Omega}$, that is, $N(\mathbf{0}, \kappa_i^{-1}\boldsymbol{\Omega})$. The mixing term κ_i is introduced to relax the normality assumption and is distributed as the gamma distribution with both shape and scale parameters $v/2$ $(v > 0)$, that is, $G(v/2, v/2)$.

Then, the i-th individuals likelihood function augmented by $\boldsymbol{y}_i^* = (y_{0i}^*, y_{1i}^*, x_i^*)'$ is given by

$$p\left(y_i, \boldsymbol{y}_i^*, x_i = 0 \mid \boldsymbol{X}_i, \boldsymbol{\beta}, \kappa_i, \boldsymbol{\Omega}\right) = p^*\left(\boldsymbol{y}_i^*, y_{0i}^* = y_i \mid \boldsymbol{X}_i, \boldsymbol{\beta}, \kappa_i, \boldsymbol{\Omega}\right) I\left(x_i^* \le 0\right), \tag{4}$$

$$p\left(y_i, \boldsymbol{y}_i^*, x_i = 1 \mid \boldsymbol{X}_i, \boldsymbol{\beta}, \kappa_i, \boldsymbol{\Omega}\right) = \begin{cases} p^*\left(\boldsymbol{y}_i^* \mid \boldsymbol{X}_i, \boldsymbol{\beta}, \kappa_i, \boldsymbol{\Omega}\right) I\left(x_i^* > 0\right) I\left(y_{Li}^* < y_i < y_{Ui}^*\right), & \text{if } c_i = 0, \\ p^*\left(\boldsymbol{y}_i^*, y_{1i}^* = y_i \mid \boldsymbol{X}_i, \boldsymbol{\beta}, \kappa_i, \boldsymbol{\Omega}\right) I\left(x_i^* > 0\right), & \text{if } c_i = 1, \end{cases} \tag{5}$$

where

$$\boldsymbol{X}_i = \begin{pmatrix} \boldsymbol{w}_{0i}' & & \boldsymbol{O} \\ & \boldsymbol{w}_{1i}' & \\ \boldsymbol{O} & & \boldsymbol{v}_i' \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_0 \\ \boldsymbol{\beta}_1 \\ \boldsymbol{\gamma} \end{pmatrix}, \tag{6}$$

and $p^*(\boldsymbol{y}_i^* \mid \boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{\Omega}, \kappa_i)$ is the density function of the multivariate normal distribution with mean $\boldsymbol{X}_i\boldsymbol{\beta}$ and variance covariance matrix $\kappa_i^{-1}\boldsymbol{\Omega}$.

For the identification, we set the $(3,3)$ element of $\boldsymbol{\Omega}$ to one. Thus,

$$\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\omega} \\ \boldsymbol{\omega}' & 1 \end{pmatrix}. \tag{7}$$

Such a restriction is often applied in the multinomial probit models. We use the typical decomposition to incorporate this restriction, proposed by McCulloch, Polson, and Rossi (2000), which is given by

$$\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Phi} + \boldsymbol{\omega}\boldsymbol{\omega}' & \boldsymbol{\omega} \\ \boldsymbol{\omega}' & 1 \end{pmatrix}, \tag{8}$$

where $\boldsymbol{\Phi} = \boldsymbol{\Omega}_{11} - \boldsymbol{\omega}\boldsymbol{\omega}'$.

3 Bayesian approach and Gibbs sampler

To conduct inferences on model parameters, we take the Bayesian approach. First, we assume the following proper prior distributions on $\boldsymbol{\beta}$, $\boldsymbol{\Phi}$, and $\boldsymbol{\omega}$, which are given by

$$\boldsymbol{\beta} \sim N(\boldsymbol{b}_0, \boldsymbol{B}_0), \quad \boldsymbol{\Phi} \sim IW(n_0, \boldsymbol{D}_0), \quad \boldsymbol{\omega} \sim N(\boldsymbol{m}_0, \boldsymbol{M}_0), \tag{9}$$

where $IW(n, \boldsymbol{D})$ denote the inverse Wishart distribution with n degrees of freedom and positive definite matrix $\boldsymbol{D}$. Let $\pi(\boldsymbol{\beta}, \boldsymbol{\Phi}, \boldsymbol{\omega})$ be the prior density function as the product of the densities of Equation (9).

Then, the posterior density function is given by

$$\begin{aligned} \pi(\boldsymbol{\beta}, \boldsymbol{\Phi}, \boldsymbol{\omega}, \boldsymbol{\kappa}, \boldsymbol{y}^* \mid \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{c}) \propto \pi(\boldsymbol{\beta}, \boldsymbol{\Phi}, \boldsymbol{\omega}) \prod_{i=1}^{n} \kappa_i^{\nu/2-1} \exp\left(-\frac{\nu}{2}\kappa_i\right) \\ \times \prod_{i \in I_0} p^*\left(\boldsymbol{y}_i^*, y_{0i}^* = y_i \mid \boldsymbol{X}_i, \boldsymbol{\beta}, \kappa_i, \boldsymbol{\Omega}\right) I\left(x_i^* \le 0\right) \\ \times \prod_{i \in I_{10}} p^*\left(\boldsymbol{y}_i^* \mid \boldsymbol{X}_i, \boldsymbol{\beta}, \kappa_i, \boldsymbol{\Omega}\right) I\left(x_i^* > 0\right) I\left(y_{Li}^* < y_i < y_{Ui}^*\right) \\ \times \prod_{i \in I_{11}} p^*\left(\boldsymbol{y}_i^*, y_{1i}^* = y_i \mid \boldsymbol{X}_i, \boldsymbol{\beta}, \kappa_i, \boldsymbol{\Omega}\right) I\left(x_i^* > 0\right), \end{aligned} \tag{10}$$

where $I_0 = \{i \mid x_i^* = 0\}$, $I_{1l} = \{i \mid x_i^* = 1, c_i = l\}$ $(l = 0, 1)$, $\boldsymbol{\kappa} = \{\kappa_i\}_{i=1}^n$, $\boldsymbol{y}^* = \{\boldsymbol{y}_i^*\}_{i=1}^n$, $\boldsymbol{y} = \{\boldsymbol{y}_i\}_{i=1}^n$, $\boldsymbol{X} = \{\boldsymbol{X}_i\}_{i=1}^n$, and $\boldsymbol{c} = \{c_i\}_{i=1}^n$.

To obtain samples from this posterior density function, we apply the Gibbs sampler, which is implemented in six steps.

Step 1. Initialize $\boldsymbol{\beta}$, $\boldsymbol{\Phi}$, $\boldsymbol{\omega}$, $\boldsymbol{\kappa}$, and $\boldsymbol{y}^*$.

Step 2. Generate $\boldsymbol{\beta}$ conditioned on $\boldsymbol{\Phi}, \boldsymbol{\omega}, \boldsymbol{\kappa}, \boldsymbol{y}^*$ from the multivariate normal distribution.

Step 3. Generate $\boldsymbol{\Phi}$ conditioned on $\boldsymbol{\beta}, \boldsymbol{\omega}, \boldsymbol{\kappa}, \boldsymbol{y}^*$ from the inverse Wishart distribution.

Step 4. Generate $\boldsymbol{\omega}$ conditioned on $\boldsymbol{\beta}, \boldsymbol{\Phi}, \boldsymbol{\kappa}, \boldsymbol{y}^*$ from the multivariate normal distribution.

Step 5. Generate $(\kappa_i, \boldsymbol{y}_i^*)$ for $i = 1, \ldots, n$.

Step 5-a. Generate $\boldsymbol{y}_i^*$ conditioned on $\boldsymbol{\beta}, \boldsymbol{\Phi}, \boldsymbol{\omega}$ from the truncated multivariate t distribution.

Step 5-b. Generate κ_i conditioned on $\boldsymbol{\beta}, \boldsymbol{\Phi}, \boldsymbol{\omega}, \boldsymbol{y}_i^*$ from the gamma distribution.

Step 6. Go to Step 2.

Exact expressions of the parameters of these full conditional distributions are presented in Appendix A

4 Rapid railroad network and population density function

This section applies the above treatment effect model to analyze the impact of the rapid railroad network in Japan on the population density. We focus on Kyushu area (the southwest island of Japan) where the latest Shinkansen, which is called Kyushu Shinkansen, runs through from the south (Kagoshima city) to the north (Fukuoka city). Partly because of the fiscal problem, its construction was divided into two parts. That is, the south part (from Kagoshima-Chuo to Shin-Yatsushiro) was opened since 2004, and, after that, the north part (from Shin-Yatsushiro to Fukuoka) followed since 2011. Thus, in the period between 2004 and 2011, there are three types of cities: (1) cities where a Shinkansen station had been opened, (2) cities where a Shinkansen station is under construction, and (3) cities where no Shinkansen station is planned. When we evaluate the economic influence of Kyushu Shinkansen on the population density by using the typical treatment effect model, it is difficult to decide whether the second type is categorized to the first type or the last type. Our framework, however, is able to include such cities as ones which are treated but their treatment is not complete, that is, $x_i = 1$ and $c_i = 0$.

To analyze the impact, we use the city-level data, which consists of 247 cities in Kyushu area. In 2005, there are five cities where a Shinkansen station had been opened and seven cities where a Shinkansen station is under construction. In remaining 235 cities, no Shinkansen station is planned. The dependent variable y_i is the logarithm of the population density in 2005 calculated as the ratio of the population (in thousand) to the habitable area (in square kilometer) subtracting the farming area (in square kilometer). To explain this y_i, we use the distance (in kilometer) to the nearest business districts as well as the constant. According to Kanemoto and Tokuoka (2001), there are 18 business districts in Kyushu area. Thus, for each city, we calculate 18 great-circle distances to these business districts and pick the minimum distance as the explanatory variable. Because the distance is invariant to the treatment status, we simply use it as the common explanatory variable to $\boldsymbol{w}_{0i}$ and $\boldsymbol{w}_{1i}$. As the explanatory variables for x_i^*, we use four variables: the constant, the distance to the nearest business district, the debt expenditure ratio of the city budget in 2001 (in thousand yen), and the absolute value of the population difference between day and night in 2000 (in person). The last two variables are standardized.

As the prior distributions, we use

$$\boldsymbol{\beta} \sim N(\mathbf{0}, 10\boldsymbol{I}), \quad \boldsymbol{\Phi} \sim IW(10, 10\boldsymbol{I}), \quad \boldsymbol{\omega} \sim N(\mathbf{0}, 10\boldsymbol{I}). \tag{11}$$

After deleting 5×10^4 samples, we draw 5×10^5 Markov chain Monte Carlo samples from the posterior distribution. To conduct inferences, we reduce them to 10^4 samples by picking

Table 1: Population density function

Parameter	*Mean*	*SD*	95% *interval*		*INEF*
β_{01} (constant, untreated)	−4.89	.093	[−5.08	−4.71]	1
β_{02} (distance, untreated)	− .010	.003	[− .016	− .005]	1
β_{11} (constant, treated)	.93	1.74	[−1.91	4.95]	23
β_{12} (distance, treated)	− .043	.062	[− .15	.090]	26
γ_1 (constant)	−1.84	.35	[−2.54	−1.15]	2
γ_2 (distance)	− .022	.021	[− .073	.007]	5
γ_3 (debt-ex. ratio)	.18	.23	[− .26	.65]	1
γ_4 (population diff.)	.97	.44	[.21	1.90]	5

* "INEF" denotes the estimated inefficiency factors.

up every 50-th sample.

Results for regression coefficients are summarized in Table 1. From this table, we found that one kilometer away from the nearest business district is associated with the 4.3% and 1% decrease of the population density evaluated at the posterior mean when the city has a Shinkansen station or not, respectively. Thus, this result suggests that, by the introduction of the Shinkansen network in Kyushu area, business districts become more valuable, so that the population agglomerates around them. Among explanatory variables for the assignment of the treatment (x_i^*), it is credible that the absolute value of the population difference between day and night has a positive effect on this assignment because its 95% credible interval does not include zero.

Next, we estimate the posterior means of the average treatment effect (ATE), that is, $E(y_{1i}^*) - E(y_{0i}^*)$, for the selected two cities. Fukuoka (city id is 2) is one of the business districts in Kyushu area and a Shinkansen station was under construction in 2005. Its ATE is estimated to be 3.85 and its 95% credible interval does not include zero. Thus, the Shinkansen network has a positive impact concerning the population density on this city. On the other hand, Kagoshima (city id is 203) is also one of the business districts and already has a Shinkansen station. Its ATE is estimated to be 1.15 with 95% credible interval including zero. This suggests that the Shinkansen network has a different effect on the population

Table 2: Correlation matrix

Parameter	*Mean*	*SD*	95% *interval*		*INEF*
$Corr(y_{0i}^*, y_{1i}^*)$	.004	.19	[−.36	.38]	6
$Corr(y_{0i}^*, x_i^*)$	−.17	.19	[−.51	.24]	2
$Corr(y_{1i}^*, x_i^*)$	−.28	.44	[−.89	.69]	38

* "INEF" denotes the estimated inefficiency factors.

density depending on the characteristics of business districts.

We also estimate the ratio defined as

$$r_i = \frac{y_i - \min(y_{0i}^*, y_{1i}^*)}{\max(y_{0i}^*, y_{1i}^*) - \min(y_{0i}^*, y_{1i}^*)}, \tag{12}$$

for cities where a Shinkansen station is under construction. The posterior means of this ratio are estimated to be $(r_2, r_3, r_4, r_{10}, r_{69}, r_{110}, r_{115}) = (.54, .47, .49, .48, .41, .47, .39)$. These ratios suggest that half of the treatment effect is already observed for such cities.

Finally, we present the estimation results of the correlation matrix, which are given in Table 2. From this table, we found weak and negative correlations between the log of the population density and the assignment of the treatment.

5 Concluding remarks

In this article, we proposed a new treatment effect model that includes individuals who are treated but their treatment is not complete. Such a modeling allows us to address the evaluation of the large infrastructure, such as the rapid railroad network, because the construction of such a large infrastructure usually takes long time. To evaluate the treatment effect of this kind, it is necessary to include in the treatment effect model the individual who are treated but their treatment has not finished. We addressed this issue by extending the Roy model. By using the proposed model, we conducted the empirical analysis of the Shinkansen network and found that the Shinkansen network causes the agglomeration of the population around the business districts.

A Full conditional distributions

Step 2. Generate $\boldsymbol{\beta}$ conditioned on $\boldsymbol{\Phi}, \boldsymbol{\omega}, \boldsymbol{\kappa}, \boldsymbol{y}^*$ from the multivariate normal distribution.

The full conditional distribution for $\boldsymbol{\beta}$ is given by $N(\boldsymbol{b}_1, \boldsymbol{B}_1)$, where

$$\boldsymbol{B}_1^{-1} = \boldsymbol{B}_0^{-1} + \sum_{i=1}^{n} \kappa_i \boldsymbol{X}_i' \boldsymbol{\Omega}^{-1} \boldsymbol{X}_i, \quad \boldsymbol{b}_1 = \boldsymbol{B}_1 \left(\boldsymbol{B}_0^{-1} \boldsymbol{b}_0 + \sum_{i=1}^{n} \kappa_i \boldsymbol{X}_i' \boldsymbol{\Omega}^{-1} \boldsymbol{y}_i^* \right). \tag{13}$$

Step 3. Generate $\boldsymbol{\Phi}$ conditioned on $\boldsymbol{\beta}, \boldsymbol{\omega}, \boldsymbol{\kappa}, \boldsymbol{y}^*$ from the inverse Wishart distribution.

Let $\boldsymbol{e}_i = \boldsymbol{y}_i^* - \boldsymbol{X}_i \boldsymbol{\beta}$ and $\boldsymbol{e}_i = (\boldsymbol{e}_{1i}', e_{2i})'$. The full conditional distribution for $\boldsymbol{\Phi}$ is given by $IW(n_1, \boldsymbol{D}_1)$, where $n_1 = n_0 + n$ and

$$\boldsymbol{D}_1 = \boldsymbol{D}_0 + \sum_{i=1}^{n} \kappa_i (\boldsymbol{e}_{1i} - e_{2i} \boldsymbol{\omega})(\boldsymbol{e}_{1i} - e_{2i} \boldsymbol{\omega})'. \tag{14}$$

Step 4. Generate $\boldsymbol{\omega}$ conditioned on $\boldsymbol{\beta}, \boldsymbol{\Phi}, \boldsymbol{\kappa}, \boldsymbol{y}^*$ from the multivariate normal distribution.

The full conditional distribution for $\boldsymbol{\omega}$ is given by $N(\boldsymbol{m}_1, \boldsymbol{M}_1)$, where

$$\boldsymbol{M}_1^{-1} = \boldsymbol{M}_0^{-1} + \sum_{i=1}^{n} e_{2i}^2 \kappa_i \boldsymbol{\Phi}^{-1}, \quad \boldsymbol{m}_1 = \boldsymbol{M}_1 \left(\boldsymbol{M}_0^{-1} \boldsymbol{m}_0 + \sum_{i=1}^{n} e_{2i} \kappa_i \boldsymbol{\Phi}^{-1} \boldsymbol{e}_{1i} \right). \tag{15}$$

Step 5-a. Generate $\boldsymbol{y}_i^*$ conditioned on $\boldsymbol{\beta}, \boldsymbol{\Phi}, \boldsymbol{\omega}$ from the truncated multivariate t distribution.

We first specify elements in $\boldsymbol{\Omega}$, which are given by

$$\boldsymbol{\Omega} = \begin{pmatrix} \upsilon_0^2 & \xi_{01} & \omega_0 \\ \xi_{01} & \upsilon_1^2 & \omega_1 \\ \omega_0 & \omega_1 & 1 \end{pmatrix}. \tag{16}$$

Let $\phi_j^2 = \upsilon_j^2 - \omega_j^2$, $\zeta_{01} = \xi_{01} - \omega_0\omega_1$, $\boldsymbol{\sigma}_{1-j} = (\zeta_{01} + \omega_0\omega_1, \omega_{1-j})'$, and

$$\boldsymbol{\Phi}_j = \begin{pmatrix} \phi_j^2 + \omega_j^2 & \omega_j \\ \omega_j & 1 \end{pmatrix}, \quad \boldsymbol{X}_{ji} = \begin{pmatrix} \boldsymbol{w}'_{0i} \times (1-j) & \boldsymbol{w}'_{1i} \times j & 0 \\ \mathbf{0}' & \mathbf{0}' & \boldsymbol{v}'_i \end{pmatrix}. \tag{17}$$

Further, let

$$\boldsymbol{S}_j = \boldsymbol{\Phi}_j - \boldsymbol{\sigma}_{1-j}\boldsymbol{\sigma}'_{1-j}\left(\phi_j^2 + \omega_j^2\right)^{-1}, \tag{18}$$

$$\boldsymbol{\mu}_{ji} = \boldsymbol{X}_{ji}\boldsymbol{\beta} + \boldsymbol{\sigma}_{1-j}\left(y_i - \boldsymbol{w}'_{1-j,i}\boldsymbol{\beta}_{1-j}\right)\left(\phi_j^2 + \omega_j^2\right)^{-1}, \tag{19}$$

$$h_{ji} = \nu(\nu+1)^{-1}\left\{1 + \left(y_i - \boldsymbol{w}'_{ji}\boldsymbol{\beta}_j\right)^2\left(\phi_j^2 + \omega_j^2\right)^{-1}/\nu\right\}. \tag{20}$$

Then, we have the full conditional distribution for $\boldsymbol{y}_i^*$ depending on (x_i, c_i).

Case 1: $x_i = 0 \Leftrightarrow i \in I_0$.

Set $y_{0i}^* = y_i$ and draw

$$\left(y_{1i}^*, x_i^*\right)' \mid y_{0i}^* = y_i \sim t_2\left(\boldsymbol{\mu}_{1i}, h_{0i}\boldsymbol{S}_1, \nu+1\right) I\left(x_i^* \le 0\right), \tag{21}$$

where $t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ denotes the p-variate t distribution with mean $\boldsymbol{\mu}$, variance covariance matrix $\boldsymbol{\Sigma}$, and ν degreees of freedom.

Case 2: $x_i = 1$ *and* $c_i = 0 \Leftrightarrow i \in I_{10}$.

Draw

$$\boldsymbol{y}_i^* \sim t_3\left(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{\Omega}, \nu\right) I\left(x_i^* > 0\right) I(y_{Li}^* < y_i < y_{Ui}^*). \tag{22}$$

Case 3: $x_i = 1$ *and* $c_i = 1 \Leftrightarrow i \in I_{11}$.

Set $y_{1i}^* = y_i$ and draw

$$\left(y_{0i}^*, x_i^*\right)' \mid y_{1i}^* = y_i \sim t_2\left(\boldsymbol{\mu}_{0i}, h_{1i}\boldsymbol{S}_0, \nu+1\right) I\left(x_i^* > 0\right). \tag{23}$$

Step 5-b. Generate κ_i conditioned on $\boldsymbol{\beta}, \boldsymbol{\Phi}, \boldsymbol{\omega}, \boldsymbol{y}_i^*$ from the gamma distribution.

The full conditional distribution for κ_i is given by

$$G\left(\frac{\nu+3}{2}, \frac{\nu+\boldsymbol{e}_i'\boldsymbol{\Omega}^{-1}\boldsymbol{e}_i}{2}\right). \quad (24)$$

Acknowledgements

The author thanks Shintaro Minami for his helpful comments and discussions. This work is supported by the Grants-in-Aid for Scientific Research 21830115 from the Japanese Ministry of Education, Science, Sports, Culture and Technology and by the Academic Frontier Project of the Nihon University Population Research Institute. All computational results were obtained using Ox for Linux (see Doornik, 2002).

References

Baum-Snow, N. (2007). Did highways cause suburbanization? *Quarterly Journal of Economics 122*(2), 775–805.

Chib, S. and B. H. Hamilton (2000). Bayesian analysis of cross-section and clustered data treatment models. *Journal of Econometrics 97*(1), 25–50.

Donaldson, D. (2010). Railroads of the raj: Estimating the impact of transportation infrastructure. mimeo.

Doornik, J. A. (2002). *Object-Oriented Matrix Programming Using Ox* (3rd ed.). London: Timberlake Consultants Press and Oxford.

Faber, B. (2012). Trade integration, market size, and industrialization: Evidence from china's national trunk highway system. mimeo.

Kanemoto, Y. and K. Tokuoka (2001). Metropolitan area definitions in Japan. University of Tokyo CIRJE Discussion Paper Series CIRJE-J-55.

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

McDonald, J. F. (1989). Econometric studies of urban population density: a survey. *Journal of Urban Economics 26*, 361–385.

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

Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *Annals of Statistics 6*(1), 34–58.