A Bayesian ordered potential outcome analysis:

Treatment effects due to a subsidized health insurance program

November 18, 2016

Abstract

In this paper we use a Bayesian approach to analyse the treatment effects due to patients' status, covered or uncovered by a subsidized health insurance, on the number of preventive healthcare visits to physicians. We propose a Bayesian endogenous switching model that allows interaction effects with observed or unobserved individual characteristics as well as taking into account endogeneity due to patients' status. Our Bayesian approach allows calculating the posterior distributions of treatment effects, and identifies the covariance between the two potential outcomes despite the fact that we do not observe individuals in both states at the same time. We found in our application that there are self-selection effects as well as moral hazard, which may imply bad consequences for the healthcare system.

JEL Classification: C11, H51, I13

Keywords: Bayesian Econometrics, Endogenous Switching, Health Insurance, Ordered Response, Potential Outcome, Treatment Effects.

1 Introduction

This paper is concerned with the use of a Bayesian approach for calculating the treatment effects of a potentially endogenous binary treatment on a discrete ordered response variable representing the number of preventive healthcare visits to physicians. Here, the endogeneity is related to the patients' status of being either covered or uncovered by a subsidized health insurance. We use three specifications: an endogenous switching model, a self-selection model, and an exogenous model. Endogenous switching models are more general than the typical self-selection models since the former takes into account interaction effects with observed or unobserved individual characteristics as well as selection (Maddala, 1983). The latter takes into account only selection effects, whereas exogenous models take into account neither selection nor interaction effects. As we use non-linear models, we can estimate the treatment effects for specific individuals. As a consequence, we can identify who are the most affected by the subsidized health insurance.
The advantage of using a Bayesian approach is that it allows calculating the complete posterior distribution of the treatment effects, which are complicated non-linear functions of the parameters, and as a consequence, we can make probabilistic statements about them. In addition, our Bayesian approach allows identifying the covariance between the potential outcomes for individuals in both states (covered and uncovered), despite the fact that there is no individual in both states at the same time. This parameter allows building the joint distribution of the outcome variable, which involves the response variable for covered and uncovered individuals. So, we can estimate measures like the probability of a positive effect, which in turn has enormous policy implications.

We evaluate the treatment effects of the subsidized health care program in Medellín (Colombia) using data from its Living Standards Survey in 2007. In particular, we found that there are remarkable selection and interaction effects that must be taken into account for estimating the treatment effects on the number of preventive healthcare visits in a year. As a consequence, we consider that more significant results are obtained using the endogenous switching models. In particular, we found that on average the probability that the representative woman and man, who are in the subsidized regime, visit a physician in a year is 8.93 p.p. and 34.34 p.p. lower than that probability would be for the representative woman and man who are not in this regime

if she/he were in it (potential outcome). These figures are very different using the self-selection and exogenous models. The former estimates that the representative woman and man in the program has on average a 3.24 p.p. and 10.17 p.p. higher probability than uncovered individuals, and the latter model estimates 1.11 p.p. and 2.26 p.p. higher probabilities for the covered representative woman and man, respectively. As a consequence, using these specifications would generate bad policy recommendations due to being based on inconsistent parameter estimators.

In addition, we found that the probability of a positive effect on the treated, which is measured as the probability that the number of preventive visits to the doctor in a year for individuals in the subsidized regime is higher than the number of visits would be for individuals who are not covered if she/he were in it, is 0.41% and 3.37% for the representative woman and man, respectively. On the other hand, the probability of a negative effect on the treated is 89.43% and 92.76% for the representative woman and man. This means that individuals in the subsidized regime do not visit the doctor for preventive healthcare reasons as much as would individuals who are not in the program if they were in it. Taking into account that we have found positive self-selection effects ($E(\sigma_{12} = 0.3684)$, $E(\sigma_{13} = 0.9717)$), which suggests that eligible individuals who expect to visit the doctor more frequently have a higher probability of being in the subsidized regime, it seems that there are moral hazard problems. Once individuals obtain healthcare coverage (preventive and curative), they do not care about preventive utilization as much as would the uncovered individuals if they were covered. This is a disturbing situation due to the strong financial problems that the Colombian healthcare system has.

An early discussion regarding endogenous switching was that of Roy (1951), who analysed the problem of agents' choosing between two professions depending on their own information about productivity in each profession, and its impacts on the distribution of income. In our particular application, individuals have private information regarding their health attributes, and conditioned on eligibility criteria, can be involved in the subsidized regime. The problem of selection arises naturally in situations where randomized studies are not possible, so the Rubin Causal Model (Rubin, 1974; Holland, 1986), which is based on potential outcomes expressed in the form of counterfactual conditional statements, is the suitable statistical method to handle

this situation. In particular, selection implies that the ignorability (conditional independence) assumptions, essentially, is ruled out (Rubin, 1978; Rosenbaum and Rubin, 1983), and as a consequence, selection should be taken into account in the inferential procedure. Because of this, we can decompose the total impact of the subsidized health care program into selection and the pure effect of the program. In addition, interaction effects can be very important. Omitting them implies strong constraints in the parameter space that may generate erroneous inference (Maddala, 1983).

Evaluating the treatment effect of health insurance choice on health utilization outcomes controlling by selection have been based on different methodological approaches. In the one hand, there are methodologies based on a frequentist framework such as semi-parametric discrete factor models (Goldman, 1995), moment conditions (Cardon and Hendel, 2001), simultaneous equations (Melo et al., 2002), latent factor structures and simulated likelihood (Deb and Trivedi, 2006), copulas (Prieger, 2002; Smith, 2003; Zimmer and Trivedi, 2006), bivariate count data models (Riphahn and Million, 2003), univariate and multivariate regression analysis (Karve, 2012), Maximum Likelihood (Ramírez Hassan et al., 2013) and Poisson regression and simulation (Frakt et al., 2015).

Regarding the support of the outcome variables, these have been non-negative continuous in the case of health expenditure (Cardon and Hendel, 2001; Karve, 2012), non-negative counts measuring physician visits or ambulatory services (Goldman, 1995; Melo et al., 2002; Deb and Trivedi, 2006; Zimmer and Trivedi, 2006; Deb and P, 2006; Ramírez Hassan et al., 2013; Frakt et al., 2015) and binary outcomes indicating hospitalization (Melo et al., 2002; Deb and Trivedi, 2006; Ramírez Hassan et al., 2013). Regarding the treatment variable, this has been binary (Goldman, 1995; Melo et al., 2002; Zimmer and Trivedi, 2006; Ramírez Hassan et al., 2013) or multinomial (Cardon and Hendel, 2001; Deb and Trivedi, 2006). Most of the literature has found remarkable selection effects (Goldman, 1995; Melo et al., 2002; Deb and Trivedi, 2006; Zimmer and Trivedi, 2006; Ramírez Hassan et al., 2013), which generate enormous distortions in the evaluation of the utilization if they are not taken into account. In addition, curative health care has been extensively analysed (Goldman, 1995; Melo et al., 2002; Cardon and Hendel, 2001;

Deb and Trivedi, 2006; Frakt et al., 2015), whereas preventive health care utilization not (Deb and Trivedi, 2006; Ramírez Hassan et al., 2013).

On the other hand, there are Munkin and Trivedi (2003), Cagatay (2005), Deb and P (2006) and Munkin and Trivedi (2008) using a Bayesian framework. The first models simultaneously the number of physician visits, physicians visit expenditure and private insurance, and finds that there is no conclusive evidence regarding selection. Cagatay (2005) studies the demand for health care in the context of moral hazard in a Bayesian framework. Deb and P (2006) model the number of visits to doctors, finding that selection accounts for more than 50% of the difference observed between the insured and the non-insured. And Munkin and Trivedi (2008) model simultaneously the number of physician visits, through an ordered variable, and a multinomial endogenous variable showing insurance status. These authors find a strong favorable selection effect.

To the best of our knowledge, we did not find any publications in the literature using endogenous switching models to evaluate the treatment effects of the health insurance choice on health utilization.

After this introduction, we present in Section 2 our Bayesian econometric approach. In particular, we present an endogenous switching model, a self-selection model, and an exogenous ordered probit model, their likelihoods, priors and posteriors. In addition, we present in that section the procedures to estimate different treatment effects, such as the Average Treatment Effect, the selection corrected Average Treatment Effect, the Average Treatment Effect on the Treated, and the probability of a positive effect on the treated. In Section 3 we present a brief description of the Colombian health system. In Section 4 we present the econometric results, and finally, we make some concluding remarks.

2 Econometric Framework

2.1 Endogenous Switching Model

A latent variable y_{1i}^* determines whether the outcome observed is y_{2i} or y_{3i}. Specifically, we observe whether y_{1i}^* is positive or negative,

$$y_{1i} = \begin{cases} 1, & y_{1i}^* > 0 \\ 0, & y_{1i}^* \leqslant 0 \end{cases} \tag{1}$$

and observe exactly one of y_{2i} or y_{3i}, according to

$$y_i = \begin{cases} y_{2i}, & y_{1i}^* > 0 \\ y_{3i}, & y_{1i}^* \leqslant 0 \end{cases} \tag{2}$$

where

$y_{2i} = j \Leftrightarrow \alpha_{2,j-1} < y_{2i}^* \leqslant \alpha_{2,j}, j = 1, 2, \ldots, m$; $\alpha_{2,0} = -\infty$, $\alpha_{2,1} = 0$ and $\alpha_{2,m} = \infty$.

$y_{3i} = k \Leftrightarrow \alpha_{3,k-1} < y_{3i}^* \leqslant \alpha_{3,k}, k = 1, 2, \ldots, m$; $\alpha_{3,0} = -\infty$, $\alpha_{3,1} = 0$ and $\alpha_{3,m} = \infty$, y_{2i}^* and y_{3i}^* are latent variables, m is the number of categories, and $\alpha_{2,1} = \alpha_{3,1} = 0$ for identification purposes.

We assume that the latent variables in Definitions 1 and 2 are linear in the regressors with additive errors, i.e., $y_{1i}^* = x_{1i}'\beta_1 + \varepsilon_{1i}$, $y_{2i}^* = x_{2i}'\beta_2 + \varepsilon_{2i}$ and $y_{3i}^* = x_{3i}'\beta_3 + \varepsilon_{3i}$. In addition, the correlated errors are joint normal, $\varepsilon \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \Sigma)$,

$$\begin{bmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \\ \varepsilon_{3i} \end{bmatrix} \overset{i.i.d.}{\sim} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & 1 & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & 1 \end{bmatrix} \right) \tag{3}$$

The Likelihood

Setting $\alpha_2 = \begin{bmatrix}\alpha_{2,2} & \alpha_{2,3} & \dots & \alpha_{2,m-1}\end{bmatrix}'$ and $\alpha_3 = \begin{bmatrix}\alpha_{3,2} & \alpha_{3,3} & \dots & \alpha_{3,m-1}\end{bmatrix}'$ the cut-point vectors of the states, and $\theta = (\beta_1, \beta_2, \beta_3, \alpha_2, \alpha_3, \Sigma)$, the likelihood of this model is

$$f(y_1, y|\theta : x_{1i}, x_{2i}, x_{3i}) = \prod_{i:y_{1i}=1} P(y_{2i} = y_i, y_{1i} = 1|\theta : x_{1i}, x_{2i}) \prod_{i:y_{1i}=0} P(y_{3i} = y_i, y_{1i} = 0|\theta : x_{1i}, x_{3i}) \tag{4}$$

where the joint probabilities are from bivariate Normal cumulative distribution functions (the assumption in 3).

Given that our econometric framework is Bayesian, it is a good idea to reparametrize the model to improve the mixing properties of the algorithm (Nandram and Chen, 1996; Li and Tobias, 2008). In particular, we first separated out the largest cut-point from both states, $\alpha_{2,m-1}$ and $\alpha_{3,m-1}$, and define the transformation:

$$\lambda_2 = \frac{1}{[\alpha_{2,m-1}]^2} \quad and \quad \lambda_3 = \frac{1}{[\alpha_{3,m-1}]^2} \tag{5}$$

let

$$\begin{aligned}
\tilde{\beta}_2 &= \sqrt{\lambda_2}\beta_2, \quad \tilde{y}_{2i} = \sqrt{\lambda_2}y^*_{2i}, \quad \tilde{\varepsilon}_{2i} = \sqrt{\lambda_2}\varepsilon_{2i} \\
\tilde{\beta}_3 &= \sqrt{\lambda_3}\beta_3, \quad \tilde{y}_{3i} = \sqrt{\lambda_3}y^*_{3i}, \quad \tilde{\varepsilon}_{3i} = \sqrt{\lambda_3}\varepsilon_{3i}
\end{aligned} \tag{6}$$

The error variance for the transformed disturbances now takes the following form:

$$\begin{bmatrix}\varepsilon_{1i} \\ \tilde{\varepsilon}_{2i} \\ \tilde{\varepsilon}_{3i}\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}0 \\ 0 \\ 0\end{bmatrix}, \begin{bmatrix}1 & \tilde{\sigma}_{12} & \tilde{\sigma}_{13} \\ \tilde{\sigma}_{12} & \lambda_2 & \tilde{\sigma}_{23} \\ \tilde{\sigma}_{13} & \tilde{\sigma}_{23} & \lambda_3\end{bmatrix}\right) \equiv \mathcal{N}(0, \tilde{\Sigma}) \tag{7}$$

where $\tilde{\sigma}_{12} = \sqrt{\lambda_2}\sigma_{12}$, $\tilde{\sigma}_{13} = \sqrt{\lambda_3}\sigma_{13}$ and $\tilde{\sigma}_{23} = \sqrt{\lambda_2\lambda_3}\sigma_{23}$ (Definitions 5 and 6). The transformed cut-point vectors are defined by $\tilde{\alpha}_2 = \begin{bmatrix}\tilde{\alpha}_{2,2} & \tilde{\alpha}_{2,3} & \dots & \tilde{\alpha}_{2,m-2}\end{bmatrix}'$ and $\tilde{\alpha}_3 = \begin{bmatrix}\tilde{\alpha}_{3,2} & \tilde{\alpha}_{3,3} & \dots & \tilde{\alpha}_{3,m-2}\end{bmatrix}'$.

Prior Distributions

We employ independent priors for the parameters of $\tilde{\theta} = (\beta_1, \tilde{\beta}_2, \tilde{\beta}_3, \tilde{\alpha}_2, \tilde{\alpha}_3, \tilde{\Sigma})$:

$$\pi(\tilde{\theta}) = \pi(\tilde{\beta})\pi(\tilde{\alpha}_2)\pi(\tilde{\alpha}_3)\pi(\tilde{\Sigma}). \tag{8}$$

where $\tilde{\beta} = (\beta_1', \tilde{\beta}_2', \tilde{\beta}_3')'$.

We assume a Gaussian prior for the location parameters, i.e., $\tilde{\beta} \sim \mathcal{N}_K(\tilde{b}_0, \tilde{B}_0)$ where $\tilde{b}_0 = 0_K$ and $\tilde{B}_0 = 1000 I_K$, that is, a priori there is no effect of the regressors on the outcome, and the prior information is vague. The priors of $\tilde{\alpha}_2$ and $\tilde{\alpha}_3$ are assumed to be improper, and finally we assume an inverse Wishart prior $\tilde{\Sigma} \sim \mathcal{IW}(\rho, R)$ with the restriction that the element $(1,1)$ of $\tilde{\Sigma}$ is equal to 1, and $\rho = 6$ and $R = I_3$ (Li and Tobias, 2008).

Posterior Distributions

Setting,

$$\tilde{s} = \begin{bmatrix} y_1^* \\ \tilde{y}_2 \\ \tilde{y}_3 \end{bmatrix}, X = \begin{bmatrix} X_1 & 0 & 0 \\ 0 & X_2 & 0 \\ 0 & 0 & X_3 \end{bmatrix}, \tilde{s}_i = \begin{bmatrix} y_{1i}^* \\ \tilde{y}_{2i} \\ \tilde{y}_{3i} \end{bmatrix}, X_i = \begin{bmatrix} x_{1i}' & 0 & 0 \\ 0 & x_{2i}' & 0 \\ 0 & 0 & x_{3i}' \end{bmatrix} \tag{9}$$

we assume the same set of controls for the outcome variable, i.e., $x_{2i} = x_{3i}$, with $\tilde{\theta}_{-\tau}$ denoting all parameters other than τ. Using the likelihood (4), the priors (8), and the previous definitions (9), we can follow the algorithm proposed by Li and Tobias (2008) to sample from the posterior distributions.

Algorithm:

1. Draw $\tilde{\beta}$ from the conditional posterior distribution:

$$\tilde{\beta} \mid \tilde{s}, \tilde{\theta}_{-\tilde{\beta}}, X \sim \mathcal{N}(\tilde{b}, \tilde{B}_1)$$

where

$$\tilde{B}_1 = \left[X'(I_n \otimes \tilde{\Sigma}^{-1})X + \tilde{B}_0^{-1}\right]^{-1} \quad and \quad \tilde{b} = \left[X'(I_n \otimes \tilde{\Sigma}^{-1})X + \tilde{B}_0^{-1}\right]^{-1}\left[X'(I_n \otimes \tilde{\Sigma}^{-1})\tilde{s} + \tilde{B}_0^{-1}\tilde{b}_0\right]$$

2. Draw $\tilde{\Sigma}$ from the conditional posterior distribution:

$$\tilde{\Sigma} \mid \tilde{s}, \tilde{\theta}_{-\tilde{\Sigma}}, X \sim \mathcal{IW}\left(N + \rho, \left[\sum_i (\tilde{s}_i - X_i\tilde{\beta})(\tilde{s}_i - X_i\tilde{\beta})' + R\right]\right) I(\tilde{\Sigma}_{11} = 1).$$

Algorithm 3 in Nobile (2000) is used to generate variables from this inverted Wishart distribution, conditioned on the $(1,1)$ element's being equal to 1.

3. We use the suggestion of Nandram and Chen (1996) to sample from the density below,

$$\tilde{\alpha}_2 \mid \tilde{\theta}_{-\tilde{\alpha}_2}, \tilde{y}_2, y_1^*, X \sim \prod_{i:y_1=1} \Phi\left(\frac{\tilde{\alpha}_{2,y_i} - \tilde{\mu}_{2i}}{\sqrt{\tilde{\sigma}_2^2}}\right) - \Phi\left(\frac{\tilde{\alpha}_{2,y_{i-1}} - \tilde{\mu}_{2i}}{\sqrt{\tilde{\sigma}_2^2}}\right)$$

where $\Phi(.)$ is the cumulative normal distribution,

$$\tilde{\mu}_2 = X_2\tilde{\beta}_2 + \left(I_N \otimes \begin{bmatrix}\tilde{\sigma}_{12} & \tilde{\sigma}_{23}\end{bmatrix}\right)\left(I_N \otimes \begin{bmatrix}1 & \tilde{\sigma}_{13} \\ \tilde{\sigma}_{13} & \lambda_3\end{bmatrix}^{-1}\right) H_2$$

$\tilde{\mu}_{i2}$ being the component i of the vector $\tilde{\mu}_2$, $H_2 = [y_{11}^* - x_{11}'\beta_1, \tilde{y}_{31} - x_{31}'\tilde{\beta}_3, y_{12}^* - x_{12}'\beta_1, \tilde{y}_{32} - x_{32}'\tilde{\beta}_3, \cdots, y_{1N}^* - x_{1N}'\beta_1, \tilde{y}_{3N} - x_{3N}'\tilde{\beta}_3]'$ of dimension $2N \times 1$ and

$$\tilde{\sigma}_2^2 = \lambda_2 - \begin{bmatrix}\tilde{\sigma}_{12} & \tilde{\sigma}_{23}\end{bmatrix}\begin{bmatrix}1 & \tilde{\sigma}_{13} \\ \tilde{\sigma}_{13} & \lambda_3\end{bmatrix}^{-1}\begin{bmatrix}\tilde{\sigma}_{12} \\ \tilde{\sigma}_{23}\end{bmatrix}$$

4. Sample $\tilde{y}_{2i}$ independently from:

$$\tilde{y}_{2i} \mid \tilde{\theta}, \tilde{y}_{3i}, y_{1i}^*, X \sim \begin{cases} \mathcal{TN}_{(\tilde{\alpha}_{2,y_{i-1}}, \tilde{\alpha}_{2,y_i})}(\tilde{\mu}_{i2}, \tilde{\sigma}_2^2) & \text{if } y_{1i} = 1 \\ \mathcal{N}(\tilde{\mu}_2, \tilde{\sigma}_2^2) & \text{if } y_{1i} = 0 \end{cases}, i = 1, 2, ..., n$$

where $\mathcal{TN}_{(\tilde{\alpha}_{2,y_{i-1}}, \tilde{\alpha}_{2,y_i})}(\tilde{\mu}_{i2}, \tilde{\sigma}_2^2)$ is a truncated normal distribution in the interval $(\tilde{\alpha}_{2,y_{i-1}}, \tilde{\alpha}_{2,y_i})$

with mean $\tilde{\mu}_{i2}$ and variance $\tilde{\sigma}_2^2$.

5. By similar arguments to those in stage 3,

$$\tilde{\alpha}_3 \mid \tilde{\theta}_{-\tilde{\alpha}_3}, \tilde{y}_3, y_1^*, X \sim \prod_{i:y_1=0} \Phi\left(\frac{\tilde{\alpha}_{3,y_i} - \tilde{\mu}_{i3}}{\sqrt{\tilde{\sigma}_3^2}}\right) - \Phi\left(\frac{\tilde{\alpha}_{3,y_{i-1}} - \tilde{\mu}_{i3}}{\sqrt{\tilde{\sigma}_3^2}}\right)$$

where

$$\tilde{\mu}_3 = X_3\tilde{\beta}_3 + \left(I_N \otimes \begin{bmatrix}\tilde{\sigma}_{13} & \tilde{\sigma}_{23}\end{bmatrix}\right)\left(I_N \otimes \begin{bmatrix}1 & \tilde{\sigma}_{12} \\ \tilde{\sigma}_{12} & \lambda_2\end{bmatrix}^{-1}\right) H_3$$

$\tilde{\mu}_{i3}$ is the component i of the vector $\tilde{\mu}_3$, $H_3 = [y_{11}^* - x_{11}'\beta_1, \tilde{y}_{21} - x_{21}'\tilde{\beta}_2, y_{12}^* - x_{12}'\beta_1, \tilde{y}_{22} - x_{22}'\tilde{\beta}_2, \cdots, y_{1,N}^* - x_{1N}'\beta_1, \tilde{y}_{2N} - x_{2N}'\tilde{\beta}_2]'$ of dimension $2N \times 1$ and

$$\tilde{\sigma}_3^2 = \lambda_2 - \begin{bmatrix}\tilde{\sigma}_{13} & \tilde{\sigma}_{23}\end{bmatrix}\begin{bmatrix}1 & \tilde{\sigma}_{12} \\ \tilde{\sigma}_{12} & \lambda_2\end{bmatrix}^{-1}\begin{bmatrix}\tilde{\sigma}_{13} \\ \tilde{\sigma}_{23}\end{bmatrix}$$

6. Sample $\tilde{y}_{3i}$ independently from:

$$\tilde{y}_{3i} \mid \tilde{\theta}, \tilde{y}_{2i}, y_{1i}^*, X \sim \begin{cases}\mathcal{TN}_{(\tilde{\alpha}_{3,y_{i-1}}, \tilde{\alpha}_{3,y_i})}(\tilde{\mu}_{i3}, \tilde{\sigma}_3^2) & \text{if } y_{1i} = 0 \\ \mathcal{N}(\tilde{\mu}_3, \tilde{\sigma}_3^2) & \text{if } y_{1i} = 1\end{cases}, i = 1, 2, ..., n$$

7. Sample y_{1i}^* independently from:

$$y_{1i}^* \mid \tilde{\theta}, \tilde{y}_{2i}, \tilde{y}_{3i}, X \sim \begin{cases}\mathcal{TN}_{(0,\infty)}(\tilde{\mu}_{i1}, \tilde{\sigma}_1^2) & \text{if } y_{1i} = 1 \\ \mathcal{TN}_{(-\infty,0)}(\tilde{\mu}_{i1}, \tilde{\sigma}_1^2) & \text{if } y_{1i} = 0\end{cases}, i = 1, 2, ..., n$$

where

$$\tilde{\mu}_1 = X_1\beta_1 + \left(I_N \otimes \begin{bmatrix}\tilde{\sigma}_{12} & \tilde{\sigma}_{13}\end{bmatrix}\right)\left(I_N \otimes \begin{bmatrix}\lambda_2 & \tilde{\sigma}_{23} \\ \tilde{\sigma}_{23} & \lambda_3\end{bmatrix}^{-1}\right) H_1$$

$\tilde{\mu}_{i1}$ is the component i of the vector $\tilde{\mu}_1$, $H_1 = [\tilde{y}_{21} - x_{21}'\tilde{\beta}_2, \tilde{y}_{31} - x_{31}'\tilde{\beta}_3, \tilde{y}_{22} - x_{22'}\tilde{\beta}_2, \tilde{y}_{32} -$

$x'_{22}\tilde{\beta}_3, \cdots, \tilde{y}_{2N} - x'_{2N}\tilde{\beta}_2, \tilde{y}_{3N} - x'_{3N}\tilde{\beta}_3]'$ and

$$\tilde{\sigma}_1^2 = 1 - \begin{bmatrix} \tilde{\sigma}_{12} & \tilde{\sigma}_{13} \end{bmatrix} \begin{bmatrix} \lambda_2 & \tilde{\sigma}_{23} \\ \tilde{\sigma}_{32} & \lambda_3 \end{bmatrix}^{-1} \begin{bmatrix} \tilde{\sigma}_{12} \\ \tilde{\sigma}_{13} \end{bmatrix}$$

Finally, we invert the mappings described by the reparametrization to recover the structural coefficients of interest.

2.2 Self-Selection Model

The model specification is as follows:

$$y_{1i} = \begin{cases} 1, & y_{1i}^* > 0 \\ 0, & y_{1i}^* \leqslant 0 \end{cases} \tag{10}$$

where

$y_i = j \Leftrightarrow \alpha_{j-1} < y_i^* \leqslant \alpha_j, j = 1, 2, \ldots, m$; $\alpha_0 = -\infty$, $\alpha_1 = 0$ and $\alpha_m = \infty$, y_{1i}^* and y_i^* are latent variables, $\alpha_1 = 0$ for identification purposes, and m is the number of categories.

We assume that the latent treatment variable in 10 is linear, that is, $y_{1i}^* = x'_{1i}\beta_1 + \varepsilon_1$, as well as the latent variable associated with the outcome, $y_i^* = x'_{2i}\beta_2 + \delta y_{1i} + \varepsilon_{2i}$. In addition,

$$\begin{bmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \sigma_{12} \\ \sigma_{21} & 1 \end{bmatrix} \right) \tag{11}$$

The Likelihood

Setting $\theta = [\beta, \Sigma]'$ and $\beta = [\beta_1', \beta_2', \delta]'$,

$$f(y_1, y | \theta : x_{1i}, x_{2i}) = \prod_{i:=1}^{N} P(y_i, y_{1i} | \theta : x_{1i}, x_{2i}) \tag{12}$$

where the joint probabilities are from a bivariate Normal cumulative distribution function (the assumption in 11).

Again we reparametrize our model to improve the mixing properties of our algorithm (Nandram

and Chen, 1996; Li and Tobias, 2008). In particular, we define the transformation

$$\lambda = \frac{1}{[\alpha_{m-1}]^2} \tag{13}$$

and let

$$\tilde{\beta}_2 = \sqrt{\lambda}[\beta_2', \delta]', \quad \tilde{y}_i = \sqrt{\lambda} y_i^*, \quad \tilde{\varepsilon}_{2i} = \sqrt{\lambda}\varepsilon_{2i} \tag{14}$$

The error variance for the transformed disturbances now takes the following form:

$$\begin{bmatrix} \varepsilon_{1i} \\ \tilde{\varepsilon}_{2i} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \tilde{\sigma}_{12} \\ \tilde{\sigma}_{12} & \lambda \end{bmatrix} \right) \equiv \mathcal{N}(0, \tilde{\Sigma}) \tag{15}$$

where

$\tilde{\sigma}_{12} = \sqrt{\lambda}\sigma_{12}$ (using definitions 13 and 14), and the transformed cut-point vector is defined as $\tilde{\alpha} = \begin{bmatrix} \tilde{\alpha}_2 & \tilde{\alpha}_3 & \dots & \tilde{\alpha}_{m-2} \end{bmatrix}$.

The Prior Distributions

We employ independent priors for the parameters of $\tilde{\theta}$:

$$\pi(\tilde{\theta}) = \pi(\tilde{\beta})\pi(\tilde{\alpha})\pi(\tilde{\Sigma}). \tag{16}$$

where $\tilde{\beta} = (\beta_1', \tilde{\beta}_2')'$.

We assume a Gaussian prior for the location parameters, i.e., $\tilde{\beta} \sim \mathcal{N}_K(\tilde{b}_0, \tilde{B}_0)$ where $\tilde{b}_0 = 0_K$ and $\tilde{B}_0 = 1000 I_K$. The prior of $\tilde{\alpha}$ is assumed to be improper, and finally we assume an inverse Wishart prior $\tilde{\Sigma} \sim \mathcal{IW}(\rho, R)$ with the restriction that the element $(1,1)$ of $\tilde{\Sigma}$ is equal to 1, and $\rho = 6$ and $R = I_2$.

The Posterior Distributions

Setting

$$\tilde{s} = \begin{bmatrix} y_1^* \\ \tilde{y} \end{bmatrix}, X = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix}, \tilde{s}_i = \begin{bmatrix} y_{1i}^* \\ \tilde{y}_i \end{bmatrix}, X_i = \begin{bmatrix} x_{1i}' & 0 \\ 0 & x_{2i}' \end{bmatrix} \tag{17}$$

Given the likelihood (12), the priors (16), and the previous definitions (17), we propose the follow algorithm to sample from the posterior distributions.

Algorithm:

1. Draw $\tilde{\beta}$ from the conditional posterior distribution:

$$\tilde{\beta} \mid \tilde{s}, \tilde{\theta}_{-\tilde{\beta}}, X \sim \mathcal{N}(\tilde{b}, \tilde{B}_1)$$

where

$$\tilde{B}_1 = \left[X'(I_n \otimes \tilde{\Sigma}^{-1})X + \tilde{B}_0^{-1}\right]^{-1} \quad and \quad \tilde{b} = \left[X'(I_n \otimes \tilde{\Sigma}^{-1})X + \tilde{B}_0^{-1}\right]^{-1} \left[X'(I_n \otimes \tilde{\Sigma}^{-1})\tilde{s} + \tilde{B}_0^{-1}\tilde{b}_0\right]$$

2. Draw $\tilde{\Sigma}$ from the conditional posterior distribution:

$$\tilde{\Sigma} \mid \tilde{s}, \tilde{\theta}_{-\tilde{\Sigma}}, X \sim \mathcal{IW}\left(N + \rho, \left[\sum_i (\tilde{s}_i - X_i\tilde{\beta})(\tilde{s}_i - X_i\tilde{\beta})' + R\right]\right) I(\tilde{\Sigma}_{11} = 1).$$

Algorithm 3 in Nobile (2000) is used to generate variables from this inverted Wishart distribution, conditioned on the $(1,1)$ element's being equal to 1.

3. Draw $\tilde{\alpha}$ from the conditional posterior distribution:

$$\tilde{\alpha} \mid \tilde{y}, y_1^*, \tilde{\theta}_{-\tilde{\alpha}} \sim \prod_i \Phi\left(\frac{\tilde{\alpha}_{y_i} - \tilde{\mu}_{i2}}{\sqrt{\tilde{\sigma}_2^2}}\right) - \Phi\left(\frac{\tilde{\alpha}_{y_{i-1}} - \tilde{\mu}_{i2}}{\sqrt{\tilde{\sigma}_2^2}}\right)$$

where $\mu_{i2} = x_{2i}'\tilde{\beta}_2 + \tilde{\sigma}_{12}\tilde{\sigma}_{11}^{-1}(y_1^* - x_{1i}'\tilde{\beta}_1)$ and $\tilde{\sigma}_2^2 = \tilde{\sigma}_{11} - \tilde{\sigma}_{12}\tilde{\sigma}_{11}^{-1}$

We use the suggestion of Nandram and Chen (1996) to sample from the previous distribution.

4. Sample $\tilde{y}_i$ independently from:

$$\tilde{y}_i \mid \tilde{\theta}, y^*_{1i}, X \sim \mathcal{TN}_{(\tilde{\alpha}_{y_{i-1}}, \tilde{\alpha}_{y_i})}(\tilde{\mu}_{i2}, \tilde{\sigma}^2_2), i = 1, 2, ..., N.$$

5. Sample y^*_{1i} independently from:

$$y^*_{1i} \mid \tilde{\theta}, \tilde{y}_{2i}, \tilde{y}_{3i}, X \sim \begin{cases} \mathcal{TN}_{(0,\infty)}(\tilde{\mu}_{i1}, \tilde{\sigma}^2_1) & \text{if } y_{1i} = 1 \\ \mathcal{TN}_{(-\infty,0)}(\tilde{\mu}_{i1}, \tilde{\sigma}^2_1) & \text{if } y_{1i} = 0 \end{cases}, i = 1, 2, ..., n$$

where $\tilde{\mu}_1 = X_1\beta_1 + \tilde{\sigma}_{12}\tilde{\sigma}_{22}(\tilde{y} - X_2\tilde{\beta}_2)$, $\tilde{\sigma}^2_1 = 1 - \tilde{\sigma}^2_{12}\tilde{\sigma}_{22}$ and $\tilde{\mu}_{i1}$ is the component i of the vector $\tilde{\mu}_1$.

Then, we recover the structural parameters reverting the reparametrization.

2.3 Ordered Multinomial Probit

The model specification is as follows:

$$y_i = j \Leftrightarrow \alpha_{j-1} < y^*_i \leqslant \alpha_j, j = 1, 2, \ldots, m \tag{18}$$

where the latent variable for the outcome variable in 18 is linear, that is, $y^*_i = x'_{2i}\beta_2 + \delta y_{1i} + \varepsilon_i$, $\varepsilon_i \sim \mathcal{N}(0, 1)$, $\alpha_0 = -\infty$, $\alpha_1 = 0$ and $\alpha_m = \infty$. We define $\alpha = \begin{bmatrix} \alpha_2 & \alpha_3 & \ldots & \alpha_{m-1} \end{bmatrix}'$.

The Likelihood

Set $\theta = (\beta, \alpha)$ and $\beta = [\beta'_2, \delta]'$ with dimension $k \times 1$. Then

$$f(y|\theta : x_i, y_{1i}) = \prod_{i:=1}^{N} P(y_i|\theta : x_{2i}, y_{1i}) \tag{19}$$

where the probabilities are from a normal cumulative distribution function.

Prior Distributions

We employ independent priors for the parameters of θ:

$$\pi(\theta) = \pi(\beta)\pi(\alpha) \tag{20}$$

We assume a Gaussian prior for the location parameters, i.e., $\beta \sim \mathcal{N}_K(b_0, B_0)$ where $b_0 = 0_K$ and $B_0 = 1000 I_K$. The prior of α is assumed to be improper.

Posterior Distributions

Setting $X = [X_2, y_1]$, and taking into account the likelihood (19) and the priors (20), we define the following algorithm.

Algorithm:

1. Draw β from the conditional posterior distribution

$$\beta \mid y^*, \theta_{-\beta}, X \sim \mathcal{N}(b, B_1)$$

where

$$B_1 = \left[X'X + B_0^{-1}\right]^{-1} \quad and \quad b = \left[X'X + B_0^{-1}\right]^{-1} \left[X'y^* + B_0^{-1} b_0\right]$$

2. Draw α from the conditional posterior distribution

$$\alpha \mid y^*, \beta, \theta_{-\rho} \sim \prod_i \Phi\left(\alpha_{y_i} - \mu_i\right) - \Phi\left(\alpha_{y_{i-1}} - \mu_i\right)$$

where $\mu_i = [x'_{1i}, y_{1i}]\beta$.

We use the suggestion of Rossi et al. (2012) to sample from the previous distribution.

3. Sample y^* from

$$y_i^* \mid \theta, X \sim \mathcal{TN}_{(\alpha_{y_{i-1}}, \alpha_{y_i})}(\mu_i, 1), i = 1, 2, ..., N.$$

2.4 Treatment Effects

The Average Treatment Effect quantifies the expected result gain for a randomly chosen individual, that is, $ATE = y_2(\theta, X) - y_3(\theta, X)$. Given that we have ordered categories, we follow Li and Tobias (2008), who adapt this concept to the differences in probabilities associated with different treatment states. In particular, we analyse the differences in probabilities, between subsidized program members and non-members, of visiting a physician for preventive care in a year. So, we have the following ATE definition for the endogenous switching model taking into account that visiting once a year is the second category (see Section 4),[1]

$$\begin{aligned} ATE(y_1, y, X) &= P(y_2 \geq 2|y_1, y, X) - P(y_3 \geq 2|y_1, y, X) \\ &= \sum_{j=2}^{6} \{P(y_2 = j|y_1, y, X) - P(y_3 = j|y_1, y, X)\} \\ &= \sum_{j=2}^{6} \left\{[\Phi(\alpha_{2,j} - x_{2i}'\beta_2) - \Phi(\alpha_{2,j-1} - x_{2i}'\beta_2)] - [\Phi(\alpha_{3,j} - x_{3i}'\beta_3) - \Phi(\alpha_{3,j-1} - x_{3i}'\beta_3)]\right\} \end{aligned} \tag{21}$$

Here, the $x_{2i} = x_{3i}$ are the characteristics of an individual in the sample.

One great advantage of our Bayesian approach is that we can estimate the complete posterior distribution of the ATE just sampling the parameters from the posterior distribution. As a consequence, we can make different probabilistic statements about this estimand, for instance, what is the probability that the ATE is higher than a specific value? Decision theory under uncertainty indicates that the mean is an optimal point estimate when a squared loss function is used:

$$\widehat{ATE} = E_{\theta|y_1,y,X}[ATE(y_1, y, X)] \approx \frac{1}{L} \sum_{l=1}^{L} ATE(y_1, y, X)$$

[1] We consider that visiting once or more times a year for preventive health care is a good benchmark. However, we can estimate the differences in probabilities associated with any specified category.

where L is the number of samples from the posterior distribution, $\theta|y_1, y, X$.

However, the previous measure of the Average Treatment Effect does not explicitly take into account selection effects. In this case a good measure for the treatment effect of an universally applicable program that is corrected by self-selection would be $ATE^{SC} = y_2(\theta, x_i, y_1 = 1) - y_3(\theta, x_i, y_1 = 0)$. We adapt this measure as in the ATE case, that is, we calculate the differences in probabilities between members and non-members of the subsidized regime conditioned on their regime status.

$$
\begin{aligned}
ATE^{SC}(y_1, y, X) =& P(y_2 \geq 2 \mid y_1 = 1, y, X) - P(y_3 \geq 2 \mid y_1 = 0, y, X) \\
=& \sum_{j=2}^{6} \{P(y_2 = j \mid y_1 = 1, y, X) - P(y_3 = j \mid y_1 = 0, y, X)\} \\
=& \sum_{j=2}^{6} \int_{-x'_{1i}\beta_1}^{\infty} \left[\Phi\left(\frac{\alpha_{2,j} - x'_{2i}\beta_2 - \sigma_{12}\epsilon_1}{(1-\sigma_{12}^2)^{1/2}} \right) - \Phi\left(\frac{\alpha_{2,j-1} - x'_{2i}\beta_2 - \sigma_{12}\epsilon_1}{(1-\sigma_{12}^2)^{1/2}} \right) \right] \frac{f(\epsilon_1)}{P(\epsilon_1 > -x'_{1i}\beta_1)} d\epsilon_1 \\
& - \int_{-\infty}^{-x'_{1i}\beta_1} \left[\Phi\left(\frac{\alpha_{3,j} - x'_{3i}\beta_3 - \sigma_{13}\epsilon_1}{(1-\sigma_{13}^2)^{1/2}} \right) - \Phi\left(\frac{\alpha_{3,j-1} - x'_{3i}\beta_3 - \sigma_{13}\epsilon_1}{(1-\sigma_{13}^2)^{1/2}} \right) \right] \frac{f(\epsilon_1)}{P(\epsilon_1 \leq -x'_{1i}\beta_1)} d\epsilon_1 \\
\approx& \sum_{j=2}^{6} \frac{1}{L} \sum_{l=1}^{L} \left[\Phi\left(\frac{\alpha_{2,j} - x'_{2i}\beta_2 - \sigma_{12}\epsilon_{1l}}{(1-\sigma_{12}^2)^{1/2}} \right) - \Phi\left(\frac{\alpha_{2,j-1} - x'_{2i}\beta_2 - \sigma_{12}\epsilon_{1l}}{(1-\sigma_{12}^2)^{1/2}} \right) \right] \\
& - \frac{1}{K} \sum_{k=1}^{K} \left[\Phi\left(\frac{\alpha_{3,j} - x'_{3i}\beta_3 - \sigma_{13}\epsilon_{1k}}{(1-\sigma_{13}^2)^{1/2}} \right) - \Phi\left(\frac{\alpha_{3,j-1} - x'_{3i}\beta_3 - \sigma_{13}\epsilon_{1k}}{(1-\sigma_{13}^2)^{1/2}} \right) \right]
\end{aligned}
\tag{22}
$$

where ϵ_{1l} and ϵ_{1k} are draws from a truncated standard normal distribution in $(-x'_{1i}\beta_1, \infty)$ and $(-\infty, -x'_{1i}\beta_1)$ respectively.

Observe in the previous expression that if there are no selection effects, $\sigma_{12} = \sigma_{13} = 0$, and the distribution of ϵ_1 would be independent of ϵ_2 and ϵ_3, and we obtain expression 21.

An optimal point estimate for ATE^{SC} is

$$\widehat{ATE}^{SC} = E_{\theta|y_1,y,X}[ATE(y_1,y,X)^{SC}] \approx \frac{1}{L}\sum_{l=1}^{L} ATE(y_1,y,X)^{SC}$$

where L is the number of samples from the posterior distribution, $\theta|y_1,y,X$.

In addition, we can calculate the Average Treatment Effect of Treated, which is the relevant measure when we want to consider the average gain from treatment for the treated (Cameron and Pravin, 2005). This is the expected gross benefit for participating in the subsidized program, and is the relevant effect when the program has eligibility criteria (Maddala, 1983):

$$\begin{aligned}
ATET(y_1,y,X) &= P(y_2 \geq 2|y_1=1,y,X) - P(y_3 \geq 2|y_1=1,y,X) \\
&= \sum_{j=2}^{6}\{P(y_2=j|y_1=1,y,X) - P(y_3=j|y_1=1,y,X)\} \\
&= \sum_{j=2}^{6}\int_{-x'_{1i}\beta_1}^{\infty}\left[\Phi\left(\frac{\alpha_{2,j}-x'_{2i}\beta_2-\sigma_{12}\epsilon_1}{(1-\sigma_{12}^2)^{1/2}}\right) - \Phi\left(\frac{\alpha_{2,j-1}-x'_{2i}\beta_2-\sigma_{12}\epsilon_1}{(1-\sigma_{12}^2)^{1/2}}\right)\right] \\
&- \left[\Phi\left(\frac{\alpha_{3,j}-x'_{3i}\beta_3-\sigma_{13}\epsilon_1}{(1-\sigma_{13}^2)^{1/2}}\right) - \Phi\left(\frac{\alpha_{3,j-1}-x'_{3i}\beta_3-\sigma_{13}\epsilon_1}{(1-\sigma_{13}^2)^{1/2}}\right)\right]\frac{f(\epsilon_1)}{P(\epsilon_1 > -x'_{1i}\beta_1)}d\epsilon_1 \\
&\approx \sum_{j=2}^{6}\frac{1}{L}\sum_{l=1}^{L}\left[\Phi\left(\frac{\alpha_{2,j}-x'_{2i}\beta_2-\sigma_{12}\epsilon_{1l}}{(1-\sigma_{12}^2)^{1/2}}\right) - \Phi\left(\frac{\alpha_{2,j-1}-x'_{2i}\beta_2-\sigma_{12}\epsilon_{1l}}{(1-\sigma_{12}^2)^{1/2}}\right)\right] \\
&- \left[\Phi\left(\frac{\alpha_{3,j}-x'_{3i}\beta_3-\sigma_{13}\epsilon_{1l}}{(1-\sigma_{13}^2)^{1/2}}\right) - \Phi\left(\frac{\alpha_{3,j-1}-x_{3i'\beta_3-\sigma_{13}\epsilon_{1l}}}{(1-\sigma_{13}^2)^{1/2}}\right)\right] \qquad (23)
\end{aligned}$$

where ϵ_{1l} are draws from a truncated standard normal distribution in $(-x'_{i1}\beta_1, \infty)$.

Again, our Bayesian approach allows obtaining the complete posterior distribution. An optimal posterior point estimate of the ATET is

$$\widehat{ATET} = E_{\theta|y_1,y,X}[ATET(y_1,y,X)] \approx \frac{1}{L}\sum_{l=1}^{L} ATET(y_1,y,X)$$

where L is the number of samples from the posterior distribution, $\theta|y_1,y,X$.

Note that we never observe an individual in both of the two states at the same time, and we must use a counterfactual scenario to identify the program effects, that is, we must use the potential outcome framework. As a consequence, we must interpret this measure as what would be the program's effect on an individual who is not in the subsidized regime if she/he were in that program.

In addition, we have Bayesian learning associated with σ_{23} despite the fact that we do not observe individuals in both states at the same time. This allows obtaining relevant measures of policy interest, for instance, the probability of a positive effect on the treated (PPET) due to the program, that is, $P(y_2 > y_3 \mid y_1 = 1)$.

$$\begin{aligned} PPET(y_1, y, X) &= P(y_2 > y_3 | y_1 = 1, y, X) \\ &= P(y_2 - y_3 > 0 | y_1 = 1, y, X) \\ &= \sum_{i=2}^{6} \sum_{i>j} P(y_2 = i, y_3 = j | y_1 = 1, y, X) \end{aligned} \tag{24}$$

where

$$\begin{aligned} P(y_2 = i, y_3 = j | y_1 = 1, y, X) &= P(\alpha_{2,i-1} < x_2'\beta_2 + \epsilon_2 < \alpha_{2,i}, \alpha_{3,j-1} < x_3'\beta_3 + \epsilon_3 < \alpha_{3,j} | \epsilon_1 + x_1'\beta_1 > 0) \\ &= P(\alpha_{2,i-1} - x_2'\beta_2 < \epsilon_2 < \alpha_{2,i} - x_2'\beta_2, \alpha_{3,j-1} - x_3'\beta_3 < \epsilon_3 < \alpha_{3,j} - x_3'\beta_3 | \epsilon_1 > -x_1'\beta_1) \\ &= \int_{-x_1'\beta_1}^{\infty} \int_{\alpha_{2,i-1}-x_2'\beta_2}^{\alpha_{2,i}-x_2'\beta_2} \int_{\alpha_{3,j-1}-x_3'\beta_3}^{\alpha_{3,j}-x_3'\beta_3} \frac{f(\epsilon_2, \epsilon_3, \epsilon_1)}{Pr(\epsilon_1 > -x_1'\beta_1)} d\epsilon_3 d\epsilon_2 d\epsilon_3 \\ &= \int_{-x_1'\beta_1}^{\infty} \int_{\alpha_{2,i-1}-x_2'\beta_2}^{\alpha_{2,i}-x_2'\beta_2} \int_{\alpha_{3,j-1}-x_3'\beta_3}^{\alpha_{3,j}-x_3'\beta_3} \frac{f(\epsilon_2 \mid \epsilon_3, \epsilon_1) f(\epsilon_3|\epsilon_1) f(\epsilon_1)}{Pr(\epsilon_1 > -x_1'\beta_1)} d\epsilon_3 d\epsilon_2 d\epsilon_1 \\ &= \int_{-x_1'\beta_1}^{\infty} \int_{\alpha_{3,j-1}-x_3'\beta_3}^{\alpha_{3,j}-x_3'\beta_3} \left(\Phi\left[\frac{\alpha_{2,i} - x_2'\beta_2 - \mu}{\varsigma^{1/2}} \right] - \Phi\left[\frac{\alpha_{2,i-1} - x_2'\beta_2 - \mu}{\varsigma^{1/2}} \right] \right) \frac{f(\epsilon_3|\epsilon_1) f(\epsilon_1)}{Pr(\epsilon_1 > -x_1'\beta_1)} d\epsilon_3 d\epsilon_1 \end{aligned} \tag{25}$$

with $\mu = \frac{1}{1-\sigma_{13}^2}[\sigma_{23}(\epsilon_3 - \sigma_{13}\epsilon_1) + \sigma_{21}(\epsilon_1 - \sigma_{13}\epsilon_3)]$, $\varsigma = \frac{1-\sigma_{13}^2-\sigma_{23}^2-\sigma_{21}^2+2\sigma_{13}\sigma_{21}\sigma_{23}}{1-\sigma_{13}^2}$ and we multiply

and divide by $P(\alpha_{3,j-1} - x_3'\beta_3 < \epsilon_3 < \alpha_{3,j} - x_3'\beta_3 | \epsilon_1 > -x_1'\beta_1)$ the equation (25),

$$
\begin{aligned}
=& E_{\epsilon_3,\epsilon_1}\left(\Phi\left[\frac{\alpha_{i,2} - x_2'\beta_2 - \mu}{\varsigma^{1/2}}\right] - \Phi\left[\frac{\alpha_{2,i-1} - x_2'\beta_2 - \mu}{\varsigma^{1/2}}\right]\right) \\
& \times P(\alpha_{3,j-1} - x_3'\beta_3 < \epsilon_3 < \alpha_{3,j} - x_3'\beta_3 | \epsilon_1 > -x_1'\beta_1) \\
\approx& \frac{1}{L}\sum_{l=1}^{L}\left(\Phi\left[\frac{\alpha_{i,2} - x_2'\beta_2 - \mu_l}{\varsigma^{1/2}}\right] - \Phi\left[\frac{\alpha_{2,i-1} - x_2'\beta_2 - \mu_l}{\varsigma^{1/2}}\right]\right) \\
& \times P(\alpha_{3,j-1} - x_3'\beta_3 < \epsilon_{3l} < \alpha_{3,j} - x_3'\beta_3 | \epsilon_{1l} > -x_1'\beta_1) \\
=& \frac{1}{L}\sum_{l=1}^{L}\left(\Phi\left[\frac{\alpha_{i,2} - x_2'\beta_2 - \mu_l}{\varsigma^{1/2}}\right] - \Phi\left[\frac{\alpha_{2,i-1} - x_2'\beta_2 - \mu_l}{\varsigma^{1/2}}\right]\right) \\
& \times \left(\Phi\left[\frac{\alpha_{j,3} - x_3'\beta_3 - \sigma_{31}\epsilon_{1l}}{(1-\sigma_{31}^2)^{1/2}}\right] - \Phi\left[\frac{\alpha_{j-1,3} - x_3'\beta_3 - \sigma_{13}\epsilon_{1l}}{(1-\sigma_{13}^2)^{1/2}}\right]\right) \qquad (26)
\end{aligned}
$$

where $\mu_l = \frac{1}{1-\sigma_{13}^2}[\sigma_{23}(\epsilon_{3l} - \sigma_{13}\epsilon_{1l}) + \sigma_{21}(\epsilon_{1l} - \sigma_{13}\epsilon_{3l})]$, $\epsilon_{3l} \sim TN_{(\alpha_{3,j-1} - x_3'\beta_3, \alpha_{3,j} - x_3'\beta_3)}(\sigma_{13}\epsilon_{1l}, 1 - \sigma_{13}^2)$ and $\epsilon_{1l} \sim TN_{(-x_1'\beta_1, \infty)}(0,1)$

An optimal posterior point estimate of the PPET is

$$
\widehat{PPET} = E_{\theta|y_1,y,X}[PPET(y_1,y,X)] \approx \frac{1}{L}\sum_{l=1}^{L} PPET(y_1,y,X)
$$

where L is the number of samples from the posterior distribution, $\theta|y_1,y,X$.

We obtain similar expressions for the treatment effects associated with the self-selection and exogenous ordered probit models as with those obtained in previous paragraphs for the endogenous switching model. This is done just setting $y_2 = x_2'\beta_2 + \delta + \epsilon_2$ and $y_3 = x_2'\beta_2 + \epsilon_2$ for the former models, and taken into account that $\sigma_{12} = \sigma_{13}$ in the case of the self-selection model, and $\sigma_{12} = \sigma_{13} = 0$ for the exogenous probit model. These imply enormous constraints in the models' parameter spaces. The only exception is the probability of a positive effect, something which cannot be calculated for these models. In addition, the treatment effects are the same in the exogenous ordered probit model due to the implicit parameters' constraints.

3 Colombian Health Care Systems

The Colombian government carried out profound reforms to the health care system in the first half of the 1990s. These reforms were due to the enormous financial and coverage problems that the public system, whose target was the poorest segment of the population, had. As a consequence, two parallel regimes were created: the contributory social insurance program financed by mandatory taxes, and a subsidized program aimed at covering the lowest income group and the most vulnerable of the population. The main objectives of these new systems were to increase coverage and solidarity through cross subsidies from the contributory to the subsidized program, improve efficiency by allowing competition in the industry and changing the supply side subsidies to demand side subsidies.

To guarantee health services, the contributory program required that employees and employers pay 4% and 8% of the employee's income. A percentage of this contribution is used to support the subsidized program, which is also financed with transfers from the central government and resources owned by the local governments and regional entities. Beneficiaries of the subsidized program are identified by the municipalities where they live, and have to pay a coinsurance rate (Ramírez Hassan et al., 2013).

Unfortunately, the new Colombian health care system has not been successful in achieving its objectives. There are targeting problems due to imperfect monitoring, corruption, and administrative problems. In particular, there are high income families that obtain benefits, whereas some poor families do not. In addition, there are coverage issues due to the high levels of informal employment, which in turn implies a low level of employees in the contributory system, and as a consequence, a scarcity of resources to finance the subsidized program. This situation is exacerbated by government transfer problems at the national and regional levels.

Under these financial constraints, it is necessary to perform a program evaluation to identify better ways to allocate the financial resources. In particular, we will focus on preventive health care utilization, since it is less expensive than hospitalization or curative health treatment (Weinstein, 1990; Cohen et al., 2008). So, effective programs of preventive health care utilization

would imply enormous savings that would help to achieve a better service.

4 Econometric Results

The subsidized health care program is targeted to poor households that are classified in levels one or two in the SISBEN[2], strata one, two or three[3], and do not belong to the contributory program (Ramírez Hassan et al., 2013).

We use the Medellín Living Standards Survey of 2007 to evaluate preventive health care utilization in this city. We found in this dataset that out of a sample of 12,975 individuals that can be in the subsidized program, 9,893 of them are in the program, that is 76.24% of the sample. This coverage rate should be 100%, so the lower rate is due to targeting problems. We can see in Table 1 that the unconditional mean of the number of preventive health care visits in this year was 2.45 and 2.49 for uncovered and covered individuals by the subsidized program. There is no statistical difference between these figures, due to the small differences and high variability. In addition, we found that subsidized program users compared to non-users are older, with a lower level of education, belong to strata one or two, most of them are women, and their perception of their health status is better (Ramírez Hassan et al., 2013). Table 2 shows that 14.91% of the individuals did not visit a doctor for preventive purposes. Additionally, the cumulative relative frequency until four visits is 98.4%. The highest frequency is observed at four visits, whose percentage is 36.31%. As a consequence of this evidence, we estimate ordered models using six categories, ranging from no preventive health care utilization to more than four visits in a year. In particular, our treatment variable is being in the subsidized program, and the response variable is the range in preventive health care utilization. Our prior assumption is that there are selection issues due to unobservable variables, and observable interaction effects, so we should use an ordered endogenous switching model (Maddala, 1983). In particular, selection is present if unobservable characteristics that lead an individual to be in the subsidized program, like ex-

[2]SISBEN is a survey whose main objective is to identify and classify households that can obtain some governmental benefits.

[3]The Colombian government classifies households into six strata to implement cross subsidies programs. These ranges from low-low to high, where low-low comprised the poorest households.

pectations of future utilization, also alter her preventive health care utilization. In addition, interaction effects due to observable variables are associated with observable individual characteristics that cause differences in preventive health care demand between the treatment group and control group. For instance, the effect on health utilization associated with age or education can be different in these groups. However, we additionally estimate ordered self-selection and exogenous models to compare the differences in treatments effects between the three specifications. We should take into account that in our setting, which is nonlinear, identification is achieved by non-linearity in functional forms, even if all the variables in the treatment equation are included in the outcome equation. However, we identify exclusion restrictions using a sequential approach where we eliminate statistically non-significant variables in both equations (treatment and outcome). This procedure guarantees a more robust identification of the causal effects, as well as more efficient estimates (Munkin and Trivedi, 2003).

We estimate our models using standard Markov chain Monte Carlo (MCMC) procedures (Chib, 1995; Robert and Casella, 2004), specifically the 'Metropolis-within-Gibbs' procedure based on data augmentation (Tanner and Wong, 1987; Albert and Chib, 1993; Li and Tobias, 2008). We implement the Markov chain Monte Carlo algorithms using 300,000, 37,000 and 37,000 iterations, a burn-in of 30,000, 10,000 and 10,000, and a thinning equal to 100, 10 and 10 for the endogenous switching, self-selection and exogenous ordered models, respectively. This procedure gives an effective posterior sample size equal to 2,700 in our three specifications. We computed several diagnostics to assess the convergence and stationarity of the chains. In particular, we employed the method due to Heidelberger and Welch (1983), the mean difference test proposed by Geweke (1992), and the diagnostic from Raftery et al. (1992). In general, the chains seem stable, and the diagnostics indicate that the chains converge to stationary distributions, except for some cut-points (see Tables 11, 12, 13 and 14 in the Appendix).

In the first part of Tables 3 and 5 there can be seen the results associated with the Bayesian estimation of the treatment equation of the endogenous switching and self-selection models, that is, a binary variable indicating subsidized regime status (covered or uncovered). We observe that older women, living in stratum 2 (stratum 1 is the reference) with a low level of education (No education is the reference), and whose own perceived health status is not Bad

(Bad health status is the reference) have a higher probability of being covered by the subsidized health system regime. We can see in the second part of Table 3 the coefficients of the outcome equations (number of preventive health care visits in a year), we obtain the same qualitative results in both equations (covered and uncovered), that is, older individuals (at decreasing rate) from stratum 2 with no education and fair to excellent own perceived health status have a higher probability of more visits to a physician for preventive care. Despite the fact that the qualitative characteristics of these two outcome equations are the same, we observe that the magnitude of these coefficients are different, where some have non-overlapping credible intervals. This fact highlights the presence of interaction effects with observable characteristics.

We can see in Table 4 the covariance and cut-point estimates for the outcome equations. In particular, we observe that a positive shock in the treatment equation increases the probability of being in the subsidized regime, and also increases the probability of visiting more frequently a physician, for individuals in this group ($E(\sigma_{12}) = 0.3684$). There is a statistically significant positive self-selection process in this group (95% credible interval equal to (0.2944, 0.4378)). On the other hand, a negative shock in the treatment equation decreases the probability of being in the subsidized regime, and also decreases the probability of visiting a physician, for individuals that are not in the subsidized regime ($E(\sigma_{13}) = 0.9717$), that is, the unobserved variables that drive the decision to not be in the subsidized regime are positively associated with those unobserved variables that drive the decision process to decrease the number of preventive health care visits in the control group. There is a 95% significant selectivity bias (credible interval equal to (0.9646, 0.9775)). In addition, we found that $E(\sigma_{13}) > E(\sigma_{12})$. This implies that *ceteris paribus* a positive shock in the treatment equation implies that the conditional probability of visiting a physician, for an individual in the subsidized regime, given that they are in this group, is higher than the conditional probability of visiting a physician, for individuals that are not in the treatment group, if they were in the subsidized regime (potential outcome). Finally, we can observe that our Bayesian approach allows estimating σ_{23}, despite the fact that y_2 and y_3 are not jointly observed, and as a consequence, is not identifiable from sample information. However, there is Bayesian learning associated with this parameter, and we can estimate significant policy measures such as the probability of a positive treatment effect $P(y_2 > y_3)$ or the probability of a positive effect on the treated $P(y_2 > y_3|y_1 = 1)$, which depends on σ_{23}. Observe that

$E(\sigma_{23}) = 0.3630$ achieves the positive definiteness constraint of the covariance matrix, that is, $\sigma_{23} \in (0.3401, 0, 3758)$.[4] This parameter space is very informative despite the fact that our prior distribution of the covariance matrix has as scale matrix an identity matrix.

In Table 6 and the second part of Table 5 can be seen the coefficient estimates associated with the outcome equation in the exogenous and self-selection ordered probit models. In general these results are qualitatively similar to the results obtained in the endogenous switching model, except for the estimates related to the education level. Regarding the estimates of the treatment, the means are 0.2718 and 0.0593 for the self-selection and ordered probit. One paradoxical result is the covariance estimate in the self-selection model (mean equal to -0.1271): this implies that positive unobserved shocks that increase the probability of being in the subsidized regime are negatively associated with unobserved shocks that increase the probability of visiting more frequently a physician for preventive care.

We can see in Figures 1 and 2 the posterior distributions of the treatment effect for a representative woman who is 32 years old, lives in stratum 2, with primary education level, and good own perceived health status (see the convergence diagnostics in Table 15 in Appendix). In particular, we can see in these figures a remarkable contrast between the results of the endogenous switching model and those of the self-selection and exogenous ordered probit models when comparing the average treatment effects (Equation 21). The 95% credible interval of the first model is (-0.1112, -0.0815), whereas for the last two models, they are (0.0195, 0.0900) and (0.0003, 0.0198), respectively. These outcomes again highlight significant differences due to the treatment of the interaction effects with observable characteristics between the representative woman in the treatment and control groups. We can see in these figures, and Table 7, the point estimates of the Average Treatment Effect. The endogenous switching model suggests that the probability of the representative woman in the subsidized regime of visiting a physician for preventive health care is on average 9.62 p.p. lower than that of her counterpart that is not in the subsidized regime. This evidence is in contrast with the outcomes that are obtained using the self-selection and exogenous ordered probit models, which suggest that on average the representative woman in the subsidized regime has a higher probability of visiting a physician,

[4] $\sigma_{23} \in (\sigma_{12} \times \sigma_{13} - ((1-\sigma_{12})^2 \times (1-\sigma_{13})^2)^{1/2}, \sigma_{12} \times \sigma_{13} + ((1-\sigma_{12})^2 \times (1-\sigma_{13})^2)^{1/2})$.

5.15 p.p. and 1.11 p.p.

In addition we estimate the posterior distributions of the treatment effect for the representative man who is 40 years old, lives in stratum 2, has the educational level high school, and whose own perceived health status is good (see the convergence diagnostics in Table 16 in the Appendix). These outcomes can be seen in Figures 3 and 4, and Table 8. We obtain again contrasting outcomes between the endogenous switching model and the self-selection and exogenous ordered probit models regarding the Average Treatment Effect (Equation 21). The point estimate of the ATE for the first model indicates that the representative man from the subsidized regime has on average 33.26 p.p. lower probability of visiting a physician for preventive health care than a representative man who is not in the subsidized regime. The 95% credible interval is (-0.3950, -0.2635). On the other hand, the point estimates of the self-selection and exogenous ordered probit models are 10.25 p.p. and 2.26 p.p., with 95% credible intervals equal to (0.0415, 0.1652) and (0.0056, 0.0392), respectively. This implies that omitting the observable interaction effects may have enormous consequences for the ATE estimates.

Although the ATE is a good reference point for evaluating the impact of the program, we should keep in mind that the previous formulation does not take into account explicitly the selection effect. For this reason we estimate the selection corrected average treatment effect (Equation 22), whose posterior distributions can be seen in Figures 1 and 2 for the representative woman, and Figures 3 and 4 for the representative man. The 95% credible interval for the representative woman using the endogenous switching model is (0.1235, 0.1823), whereas this criterion is not significant using the self-selection model, and is equal to the ATE using the exogenous model. The SCATE for the representative man is not significant using the endogenous switching model, whereas the 95% credible interval using the self-selection model is (0.0066, 0.0372). These outcomes suggest different implications: first, the significance of the positive selection effect, which indicates that individuals (the representative woman and man) self-select into the subsidized regime depending on healthcare utilization (ATE vs SCATE). Second, interaction effects are very significant, and this fact is illustrated by the differences between the treatment effects obtained using the different methodologies (endogenous switching model vs self-selection model). And third, there are significant differences between the representative

woman and the representative man. The representative woman in the subsidized regime has, on average, a statistically significant 15.29 p.p. higher probability of visiting a physician for preventive healthcare in a year than a representative woman who is not in this regimen, whereas the selection corrected average treatment effect for the representative man is not significant (see Tables 7 and 8). It seems that the selection and interaction effects are more remarkable for the representative woman than the representative man (woman vs man).

In our effort to find a good measure of the treatment effects of the subsidized regime on the utilization of preventive healthcare, we should take into account that the subsidized regime has specific selection criteria, which means that this program does not have universal applicability, and, as a consequence, the relevant measure of the impact of the program is the Average Treatment Effect on the Treated (Equation 23). This measure takes into account simultaneously interaction and selection effects, and eligibility criteria. We can see the posterior distributions of the average treatment effect on the treated in Figures 1 and 2 for the representative woman, and in Figures 3 and 4 for the representative man. In addition, Tables 7 and 8 show the point estimates from using the different models. Using the endogenous switching model, we found that the probability of visiting a physician for preventive reasons is on average 8.93 p.p. and 34.34 p.p. lower for the representative woman and man that are involved in the subsidized regime compared to what their counterparts that are not in the subsidized regime would do if they were in it (the counter-factual scenario). The 95% credible intervals are (-0.09309, -0.07538) and (-0.40897, -0.29772), respectively.

Finally, our Bayesian methodological approach allows finding the posterior distribution of estimands that involve σ_{23}. So, we estimate the probability of a positive effect on the treated due to the subsidized program (Equation 24) using the joint probabilities for the representative individuals (Equation 26) as if she/he would have simultaneously been observed at both states, covered and uncovered by the subsidized regime (see Tables 9 and 10). We can see in Figures 1 and 3 the posterior distributions of the PPET for the representative woman and man. As we can see in these figures, and Tables 7 and 8, the probability of a positive effect due to the subsidized regime for the representative woman and man is on average 0.41% and 3.37%. In addition, the probability of a neutral effect on treated ($P(y_2 = y_3|y_1 = 1)$) is approximately

10.63% and 3.84% for the representative woman and man, and the probabilities of a negative effect ($P(y_2 < y_3|y_1 = 1)$) are 89.43% and 92.76%, respectively.[5]

These results show that apparently there are moral hazard problems, which in turn may generate financial problems for the health system. Despite the fact that the statistical inference suggests positive self-selection effects (see the discussion about the covariance estimates associated with the endogenous switching model in the previous paragraphs), which can be based on expectations about future utilization of preventive health care services, once individuals are covered by the subsidized regime, they visit the doctor less frequently than would uncovered individuals if they were in the regime. It seems that once individuals ensure their status in the health system, they are guaranteed future health service, preventive and curative, and they do not care about the former as would their counterparts that are not covered if they were in the subsidized regime.

5 Concluding Remarks

The Colombian healthcare system has tight financial constraints, which motivates the efficient use of the available resources. In addition, there is evidence that shows that preventive healthcare services involve lower costs than curative healthcare services. As a consequence, the good use of preventive healthcare may help to allocate more efficiently the financial resources of the subsidized regime, and as a result, to improve this service.

Statistical inference suggests that in the case of Medellín (Colombia) there are moral hazard problems in the case of preventive healthcare utilization by the representative individuals. Specifically, the representative woman and man has on average a 8.93 p.p. and 34.34 p.p. lower probability of visiting a physician than would their counterparts that are not involved in the subsidized regime if they were involved in it. In addition, the probability of a positive effect on

[5]Observe that our Bayesian approach allows obtaining the joint distribution ($P(y_2 = i, y_3 = j|y_1 = 1)$). So, we can marginalize this distribution to obtain the marginal distributions ($y_2 = i|y = 1$ and $y_3 = j|y_1 = 1$), which in turn are used to estimate the Average Treatment Effects on the treated. To check the consistency of our procedure we compare the outcomes using both procedure, and we found similar results.

the treated is on average just 0.41% and 3.37%, respectively. These outcomes are due to enormous interaction effects on eligible individuals that reflect behavioural changes when individuals are enrolled in the subsidized regime. Once they are involved in the subsidized system, they do not care about preventive healt care.

We found that the flexibility of the endogenous switching model allows us to identify selection and interaction effects that are very important in our application, whereas the self-selection and exogenous ordered models can give misleading results due to their implicit restrictions regarding these effects. In addition, our Bayesian approach has two good characteristics. First, it allows obtaining the posterior distributions of the treatment effects, which are complicated non-linear functions of the parameter estimates, and as a consequence, we can perform statistical inferences regarding these measures. And second, it allows performing a statistical inference of the covariance between the outcomes associated with two different states, despite the fact that we never observe an individual in both states at the same time. In our application, we found that the parameter space of this covariance is very narrow despite the fact that the scale matrix of our prior distribution is an identity matrix.

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.

Cagatay, K. (2005). Health-specific moral hazard effects. *Southern Economic Journal*, 72(1):98–118.

Cameron, C. and Pravin, K. (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press.

Cardon, J. and Hendel, I. (2001). Asymmetric information in health insurance: Evidence from the national medical expenditure survey. *The RAND Journal of Economics*, 32:408–427.

Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*, 90(432):1313–1321.

Cohen, J. T., Neumann, P. J., and Weinstein, M. C. (2008). Does preventive care save money? health economics and the presidential candidates. *New England Journal of Medicine*, 358(7):661–663.

Deb, M. and P, T. (2006). Private insurance, selection, and health care use: A Bayesian analysis of a Roy-type model. *Journal of Business and Economic Statistics*, 24(4):403–415.

Deb, P. and Trivedi, P. (2006). Specification and simulated likelihood estimation of a non-normal treatment-outcome model with selection: Application to health care utilization. *Econometrics Journal*, 9:307–331.

Frakt, A. B., Hanchate, A., and Pizer, S. D. (2015). The effect of Medicaid expansions on demand for care from the Veterans Health Administration. *Healthcare*, 3(3):123–128.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M., editors, *Bayesian Statistics 4: Proceedings of the Fourth Valencia International Meeting*, pages 169–193. Oxford University Press.

Goldman, D. (1995). Managed care as a public cost–cotainment mechanism. *The RAND Journal of Economics*, 26:277–295.

Heidelberger, P. and Welch, P. D. (1983). Simulation run length control in the presence of an initial transient. *Operations Research*, 31(6):1109–1144.

Holland, P. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81(396):945–970.

Karve, Panish, e. a. (2012). Health care utilization and costs among medicaid-enrolled patients with schizophrenia experiencing multiple psychiatric relapses. *Health Outcomes Research in Medicine*, 3(4):el83–e194.

Li, M. and Tobias, J. L. (2008). Bayesian analysis of treatment effects in an ordered potential outcomes model. *Advances in Econometrics*, 21:57–91.

Maddala, G. (1983). *Limited-Dependent and Qualitative Variables in Economics*. Cambridge University Press.

Melo, M., Stearns, S., and Norton, E. (2002). Do Medicare HMOs still reduce health services use after controlling for slection bias? *Health Economics*, 11:323–340.

Munkin, M. and Trivedi, P. (2003). Bayesian analysis of a self-selection model with multiple outcomes using simulation-based estimation: An application to the demand for healthcare. *Journal of Econometrics*, 114:197–220.

Munkin, M. and Trivedi, P. (2008). Bayesian analysis of the ordered probit model with endogenous selection. *Journal of Econometrics*, 143:334–348.

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.

Nobile, A. (2000). Comment: Bayesian multinomial probit models with a normalization constraint. *Journal of Econometrics*, 99(2):335–345.

Prieger, J. (2002). A flexible parametric selection model for non-normal data with application to health care usage. *Journal of Applied Econometrics*, 17:367–392.

Raftery, A. E., Lewis, S., et al. (1992). How many iterations in the Gibbs sampler? *Bayesian Statistics*, 4(2):763–773.

Ramírez Hassan, A., Cardona, J., and Cadavid, R. (2013). The impact of subsidized health insurance on the poor in Colombia: Evaluating the case of Medellín. *Economia Aplicada*, 17(4):543–556.

Riphahn, W. and Million (2003). Incentive effects in the demand for health care: A bivariate panel count data estimation. *Journal of Applied Economics*, 18:387–405.

Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods, 2 Edition*. Springer-Verlag, Berlin.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

Rossi, P. E., Allenby, G. M., and McCulloch, R. (2012). *Bayesian Statistics and Marketing*. John Wiley & Sons.

Roy, A. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers*, 3:135–146.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701.

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

Smith, M. D. (2003). Modelling sample selection using Archimedean copulas. *The Econometrics Journal*, 6(1):99–123.

Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82(398):528–540.

Weinstein, M. C. (1990). The costs of prevention. *Journal of General Internal Medicine*, 5:S89–S92.

Zimmer, D. and Trivedi, P. (2006). Using trivariate copulas to model sample selection and treatment effects: Application to family health care demand. *Journal of Business & Economic Statistics*, 24(1):63–76.

6 Appendix

Tables

Table 1: Descriptive statistics.

Variable	No Subsidized Regimen	Subsidized Regimen	Mean Hypothesis Test[1]
Number Preventive Visits	2.45	2.49	-0.99
	(2.17)	(1.82)	
Age	25.96	32.37	-16.85^{*}
	(17.43)	(21.29)	
Female	0.45	0.59	-13.56^{*}
	(0.50)	(0.49)	
No Education	0.31	0.39	-7.94^{*}
	(0.46)	(0.49)	
Primary	0.41	0.40	1.04
	(0.49)	(0.49)	
High School	0.25	0.20	5.82^{*}
	(0.43)	(0.40)	
Vocational	0.01	0.01	2.49^{*}
	(0.12)	(0.09)	
University	0.01	0.00	4.23^{*}
	(0.06)	(0.11)	
Bad Health Status	0.01	0.02	-5.46^{*}
	(0.08)	(0.13)	
Fair Health Status	0.07	0.13	-9.49^{*}
	(0.26)	(0.33)	
Good Health Status	0.83	0.77	6.90^{*}
	(0.38)	(0.42)	
Excellent Health Status	0.09	0.08	1.64
	(0.29)	(0.28)	
Stratum 1	0.24	0.26	-2.95^{*}
	(0.43)	(0.44)	
Stratum 2	0.52	0.66	-13.80^{*}
	(0.50)	(0.47)	
Stratum 3	0.24	0.08	20.41^{*}
	(0.43)	(0.26)	

Standard Deviation in parenthesis.

[1] Null hypothesis: means are equal. Critical value at 5% level of significance is 1.96

* Rejection of null hypothesis at 5% significance level

Source: Authors' estimations

Table 2: Preventive Health Care Visits: Observed Frequencies.

Medical appointments per year	Percent	Frequency
0	14.91	1,935
1	21.87	2,837
2	9.56	1,241
3	15.75	2,043
4	36.31	4,711
more than 4	1.60	208

Source: Authors' calculations

Table 3: Endogenous Switching: Preventive Health Care Utilization

Treatment: Subsidized Regime				
Parameter	Mean	Median	95% Credible Intervals	
			Lower	Upper
Age×Female	0.0076	0.0076	0.0067	0.0085
Stratum 2	0.1275	0.1274	0.0828	0.1719
Stratum 3	-0.7515	-0.7516	-0.8166	-0.6857
Primary	-0.1121	-0.1126	-0.1587	-0.0674
High School	-0.2211	-0.2217	-0.2753	-0.1677
Vocational	-0.3406	-0.3436	-0.5310	-0.1512
University	-0.7003	-0.7041	-0.9247	-0.4652
Fair Health Status	0.9145	0.9152	0.8355	0.9948
Good Health Status	0.6805	0.6807	0.6336	0.7275
Excellent Health Status	0.7299	0.7316	0.6497	0.8057
Outcome: Preventive Health Care Utilization covered by subsidized regime				
Parameter	Mean	Median	95% Credible Intervals	
			Lower	Upper
Age	0.0199	0.0199	0.0166	0.0230
Age^2	-0.0002	-0.0002	-0.0003	-0.0002
Stratum 2	0.1319	0.1324	0.0920	0.1693
Stratum 3	-0.2851	-0.2852	-0.3617	-0.2079
Primary	-0.0068	-0.0063	-0.0507	0.0370
High School	-0.2058	-0.2063	-0.2592	-0.1524
Vocational	-0.0944	-0.0954	-0.2952	0.0955
University	-0.5201	-0.5125	-0.8139	-0.2545
Fair Health Status	0.5762	0.5769	0.4971	0.6542
Good Health Status	0.6523	0.6533	0.5881	0.7149
Excellent Health Status	0.9489	0.9476	0.8618	1.0367
Outcome: Preventive Health Care Utilization uncovered by subsidized regime				
Parameter	Mean	Median	95% Credible Intervals	
			Lower	Upper
Age	0.0110	0.0110	0.0081	0.0141
Age^2	-0.0001	-0.0001	-0.0002	-0.0001
Stratum 2	0.1429	0.1425	0.0925	0.1926
Stratum 3	-0.6118	-0.6120	-0.6771	-0.5450
Primary	-0.0558	-0.0559	-0.1104	-0.0046
High School	-0.2414	-0.2409	-0.3040	-0.1793
Vocational	0.0043	0.0088	-0.1958	0.1991
University	-0.4900	-0.4935	-0.7142	-0.2616
Fair Health Status	1.8893	1.8879	1.7980	1.9806
Good Health Status	1.6784	1.6778	1.6217	1.7391
Excellent Health Status	1.7323	1.7315	1.6461	1.8201

Table 4: Endogenous Switching (covariances and cut-points): Preventive Health Care Utilization

Covariances				
Parameter	Mean	Median	95% Credible Intervals	
			Lower	Upper
σ_{12}	0.3684	0.3686	0.2944	0.4378
σ_{13}	0.9717	0.9720	0.9646	0.9775
σ_{23}	0.3630	0.3649	0.2647	0.4546
Cut-points: Preventive Health Care Utilization covered by subsidized regime				
Parameter	Mean	Median	95% Credible Intervals	
			Lower	Upper
α_{23}	0.3037	0.3036	0.2916	0.3155
α_{24}	0.6706	0.6703	0.6439	0.6966
α_{25}	0.9290	0.9286	0.8920	0.9650
α_{26}	1.1910	1.1905	1.1436	1.2372
Cut-points: Preventive Health Care Utilization uncovered by subsidized regime				
Parameter	Mean	Median	95% Credible Intervals	
			Lower	Upper
α_{33}	0.7775	0.7779	0.7231	0.8284
α_{34}	1.0479	1.0479	0.9969	1.0978
α_{35}	1.5382	1.5368	1.4997	1.5775
α_{36}	2.9051	2.9070	2.8382	2.9672

Table 5: Self-Selection: Preventive Health Care Utilization

Treatment: Subsidized Regime				
Parameter	Mean	Median	95% Credible Intervals	
			Lower	Upper
Age×Female	0.0141	0.0141	0.0130	0.0152
Stratum 2	0.1264	0.1257	0.0802	0.1785
Stratum 3	-0.7653	-0.7661	-0.8334	-0.6973
Primary	-0.1646	-0.1640	-0.2157	-0.1170
High School	-0.2549	-0.2552	-0.3113	-0.1982
Vocational	-0.4187	-0.4205	-0.6168	-0.2179
Universit	-0.8234	-0.8214	-1.0573	-0.5837
Fair Health Status	0.8066	0.8077	0.7220	0.8868
Good Health Status	0.6392	0.6386	0.5913	0.6870
Excellent Health Status	0.6525	0.6536	0.5693	0.7319
Outcome: Preventive Health Care Utilization				
Parameter	Mean	Median	95% Credible Intervals	
			Lower	Upper
Age	0.0179	0.0179	0.0149	0.0209
Age^2	-0.0002	-0.0002	-0.0003	-0.0002
Stratum 2	0.0897	0.0893	0.0548	0.1254
Stratum 3	-0.0935	-0.0927	-0.1597	-0.0273
Primary	0.0398	0.0398	0.0011	0.0785
High School	-0.1662	-0.1663	-0.2141	-0.1173
Vocational	0.1586	0.1622	0.0037	0.3159
University	-0.1673	-0.1658	-0.3596	0.0224
Fair Health Status	0.5168	0.5183	0.3912	0.6400
Good Health Status	0.6269	0.6281	0.5157	0.7389
Excellent Health Status	0.8789	0.8798	0.7538	1.0023
Treatment: Subsidized Regime	0.2718	0.2725	0.1288	0.4146
Covariance				
Parameter	Mean	Median	95% Credible Intervals	
			Lower	Upper
σ_{12}	-0.1271	-0.1281	-0.2147	-0.0417
Cut-points				
Parameter	Mean	Median	95% Credible Intervals	
			Lower	Upper
α_3	0.7757	0.7729	0.7282	0.8279
α_4	1.0242	1.0247	0.9701	1.0717
α_5	1.5352	1.5334	1.4991	1.5710
α_6	2.9161	2.9170	2.8598	2.9696

Table 6: Ordered Probit: Preventive Health Care Utilization

Categories				
Parameter	Mean	Median	95% Credible Intervals	
			Lower	Upper
Age	0.0187	0.0188	0.0158	0.0218
Age^2	-0.0002	-0.0002	-0.0003	-0.0002
Stratum 2	0.0928	0.0932	0.0562	0.1282
Stratum 3	-0.1408	-0.1406	-0.1995	-0.0859
Primary	0.0300	0.0308	-0.0075	0.0683
High School	-0.1807	-0.1802	-0.2283	-0.1333
Vocational	0.1385	0.1379	-0.0155	0.2997
University	-0.2256	-0.2259	-0.4171	-0.0283
Fair Health Status	0.6252	0.6252	0.5534	0.6962
Good Health Status	0.7159	0.7162	0.6624	0.7705
Excellent Health Status	0.9629	0.9630	0.8892	1.0360
Treatment: Subsidized Regime	0.0593	0.0589	0.0237	0.0966
Cut-points				
Parameter	Mean	Median	95% Credible Intervals	
			Lower	Upper
α_3	0.7058	0.7066	0.6857	0.7240
α_4	0.9557	0.9566	0.9332	0.9756
α_5	1.3605	1.3609	1.3372	1.3826
α_6	3.1943	3.1940	3.1461	3.2459

Table 7: Treatment Effects: Representative Woman

	ATE		SCATE		ATET		PPET	
Variable	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Endogenous Switching	-0.0962	0.0074	0.1529	0.0148	-0.0893	0.0046	0.0041	0.0049
Self Selection	0.0515	0.0185	0.0033	0.0021	0.0324	0.0158	Na	Na
Ordered Probit	0.0111	0.0042	0.0111	0.0042	0.0111	0.0042	Na	Na

Table 8: Treatment Effects: Representative Man

	ATE		SCATE		ATET		PPET	
Variable	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Endogenous Switching	-0.3326	0.0319	-0.0115	0.0527	-0.3434	0.0289	0.0337	0.0369
Self Selection	0.1025	0.0332	0.0207	0.0079	0.1017	0.0340	Na	Na
Ordered Probit	0.0226	0.0084	0.0226	0.0084	0.0226	0.0084	Na	Na

Table 9: Conditional joint distribution categorical variables: Representative Woman

Conditional joint distribution						
	y_3					
y_2	1	2	3	4	5	6
1	$3.0437e{-10}$	$1.1584e{-07}$	$3.3703e{-05}$	0.0006	0.0044	0.0843
2	$8.5117e{-11}$	$4.0492e{-08}$	$1.5674e{-05}$	0.0003	0.0032	0.1738
3	$3.0157e{-11}$	$1.4574e{-08}$	$5.6892e{-06}$	0.0001	0.0012	0.0861
4	$5.7108e{-11}$	$2.6384e{-08}$	$1.0164e{-05}$	0.0002	0.0021	0.1724
5	$1.6688e{-10}$	$7.0282e{-08}$	$2.3891e{-05}$	0.0005	0.0042	0.3658
6	$3.2217e{-10}$	$1.0477e{-07}$	$2.6303e{-05}$	0.0004	0.0031	0.0969

Table 10: Conditional joint distribution categorical variables: Representative Man

Conditional joint distribution						
	y_3					
y_2	1	2	3	4	5	6
1	$3.6861e{-}05$	0.0011	0.0162	0.0354	0.0506	0.2401
2	$1.1295e{-}05$	0.0004	0.0067	0.0171	0.0287	0.2098
3	$3.6322e{-}06$	0.0002	0.0019	0.0049	0.0084	0.0696
4	$6.0334e{-}06$	0.0002	0.0029	0.0072	0.0118	0.1042
5	$1.1854e{-}05$	0.0003	0.0048	0.0105	0.0157	0.1230
6	$1.0247e{-}05$	0.0003	0.0028	0.0052	0.0065	0.0132

6.1 Appendix 1: Tables

Table 11: Diagnostics Endogenous Switching: Preventive Health Care Utilization

Treatment: Subsidized Regime			
Parameter	Heidelberger[a]	Geweke[b]	Raftery[c]
Age×Female	0.9780	-0.0618	2.8100
Stratum 2	0.3450	-0.2926	1.0800
Stratum 3	0.7690	0.2656	1.0800
Primary	0.7750	-0.1948	1.0800
High School	0.7420	-0.2421	0.9150
Vocational	0.3650	-1.2729	1.0500
University	0.9120	-0.7055	0.9090
Fair Health Status	0.5560	0.1019	1.0600
Good Health Status	1.0000	0.0410	1.0800
Excellent Health Status	0.7090	-0.1415	1.1500
Outcome: Preventive Health Care Utilization covered by subsidized regime			
Parameter	Heidelberger[a]	Geweke[b]	Raftery[c]
Age	0.9692	0.3066	0.9170
Age^2	0.9442	-0.2574	0.9470
Stratum 2	0.3418	-0.4398	0.9470
Stratum 3	0.0585	-1.0497	1.1700
Primary	0.4422	1.1200	0.9820
High School	0.9913	0.4621	0.9910
Vocational	0.3887	1.3623	1.1000
University	0.4002	0.0955	1.0100
Fair Health Status	0.6654	-1.1452	1.1000
Good Health Status	0.7869	-0.9119	1.2800
Excellent Health Status	0.8527	-0.8631	1.1700
Outcome: Preventive Health Care Utilization uncovered by subsidized regime			
Parameter	Heidelberger[a]	Geweke[b]	Raftery[c]
Age	0.5530	-0.5752	1.0600
Age^2	0.6950	1.0455	1.0900
Stratum 2	0.9210	0.4687	1.0000
Stratum 3	0.7160	-0.0955	1.1400
Primary	0.5130	0.0577	0.9970
High School	0.4290	1.0535	0.9860
Vocational	0.1190	-1.7087	1.1500
University	0.8090	-1.1425	0.9470
Fair Health Status	0.8730	-0.4173	1.0800
Good Health Status	0.7700	-0.9418	1.0700
Excellent Health Status	0.2690	-1.2166	1.2000

Notes: [a] Null hypothesis is stationarity of the chain (p-values), [b] Mean difference test z-score, and [c] Dependence factor (threshold of 5)

Table 12: Diagnostics Endogenous Switching (covariances and cut-points): Preventive Health Care Utilization

Sigma			
Parameter	Heidelberger[a]	Geweke[b]	Raftery[c]
σ_{12}	0.9230	0.4543	2.3400
σ_{13}	0.9860	1.12258	3.5400
σ_{23}	0.9490	-0.0725	3.9000
Cut-points: Preventive Health Care Utilization covered by subsidized regime			
Parameter	Heidelberger[a]	Geweke[b]	Raftery[c]
α_{23}	0.3867	-2.0252	1.5600
α_{24}	0.0619	-2.6283	1.6600
α_{25}	0.3432	-1.0671	1.1600
α_{26}	0.8751	-0.1939	1.2700
Cut-points: Preventive Health Care Utilization uncovered by subsidized regime			
Parameter	Heidelberger[a]	Geweke[b]	Raftery[c]
α_{33}	0.9760	-0.8516	6.0400
α_{34}	0.9760	-0.8516	6.0400
α_{35}	0.9760	-0.8516	6.0400
α_{36}	0.9760	-0.8516	6.0400

Notes: [a] Null hypothesis is stationarity of the chain (p-values), [b] Mean difference test z-score, and [c] Dependence factor (threshold of 5)

Table 13: Diagnostics Self-Selection: Preventive Health Care Utilization

Treatment: Subsidized Regime			
Parameter	Heidelberger[a]	Geweke[b]	Raftery[c]
Age×Female	0.5804	1.1237	1.1000
Stratum 2	0.1949	1.7120	0.9940
Stratum 3	0.0775	0.1524	0.9360
Primary	0.3304	-1.4419	0.9940
High School	0.1180	0.8106	1.0100
Vocational	0.8393	-0.6487	1.0400
University	0.5973	-0.6910	0.9580
Fair Health Status	0.0914	1.7501	1.0700
Good Health Status	0.1464	-1.4246	0.9510
Excellent Health Status	0.4603	1.0516	0.9710
Preventive Health Care Utilization			
Parameter	Heidelberger[a]	Geweke[b]	Raftery[c]
Age	0.1040	-1.6781	1.1000
Age^2	0.1230	2.1574	1.0900
Stratum 2	0.7360	0.3764	1.0200
Stratum 3	0.7810	-0.6527	1.3700
Primary	0.8890	0.2889	0.9760
High School	0.4680	0.1996	0.9620
Vocational	0.9600	-0.0095	0.9820
University	0.5770	-0.8568	1.0300
Fair Health Status	0.5560	0.5323	3.7900
Good Health Status	0.6080	0.5174	3.9900
Excellent Health Status	0.8640	0.3875	3.4700
Subsidized Regime	0.6360	-0.8591	4.5700
σ_{12}	0.6310	0.7708	4.8900
α_3	0.6960	-0.9792	4.6000
α_4	0.6200	-1.2165	6.3800
α_5	0.4930	-1.5654	1.5900
α_6	0.9880	0.8876	2.6900

Notes: [a] Null hypothesis is stationarity of the chain (p-values), [b] Mean difference test z-score, and [c] Dependence factor (threshold of 5)

Table 14: Diagnostics Ordered Probit: Preventive Health Care Utilization

Preventive Health Care Utilization			
Parameter	Heidelberger[a]	Geweke[b]	Raftery[c]
Age	0.6400	1.4597	0.9920
Age^2	0.8350	-1.2895	0.9980
Stratum 2	0.7210	-0.9999	1.0400
Stratum 3	0.7750	0.7941	1.0300
Primary	0.5100	-0.6153	0.9490
High School	0.2710	-1.1521	0.9900
Vocational	0.4780	2.0290	1.0600
University	0.8770	-0.6624	1.0400
Fair Health Status	0.5200	-1.3764	1.1400
Good Health Status	0.4990	-0.1016	1.0800
Excellent Health Status	0.2400	0.3571	1.0900
Subsidized Regime	0.1810	0.7480	1.0400
α_3	0.8340	1.0651	10.7000
α_4	0.9410	0.7161	9.3200
α_5	0.8200	0.9393	7.4900
α_6	0.4850	0.0198	3.4600

Notes: [a] Null hypothesis is stationarity of the chain (p-valus), [b] Mean difference test z-score, and [c] Dependence factor (threshold of 5)

Table 15: Diagnostics Representative Woman: Treatment Effects

Diagnostics Representative Woman: Treatment Effects			
Parameter	Heidelberger[a]	Geweke[b]	Raftery[c]
Endogenous Switching(ATE)	0.900	-0.4088	1.4100
Endogenous Switching(ATET)	0.812	0.0599	1.0400
Endogenous Switching(SCATE)	0.984	-0.0530	1.1500
Endogenous Switching(PPET)	0.996	0.3717	3.0500
Self Selection(ATE)	0.741	-0.7349	4.4400
Self Selection(ATET)	0.896	-0.3553	4.4300
Self Selection(SCATE)	0.267	0.7129	1.1400
Ordered Probit(ATE)	0.163	-0.5612	1.1100

Notes: [a] Null hypothesis is stationarity of the chain (p-value), [b] Mean difference test z-score, and [c] Dependence factor (threshold of 5)

Table 16: Diagnostics Representative Man: Treatment Effects

Diagnostics Representative Man: Treatment Effects			
Parameter	Heidelberger[a]	Geweke[b]	Raftery[c]
Endogenous Switching(ATE)	0.894	-0.9980	1.0500
Endogenous Switching(ATET)	0.999	-0.4265	1.0400
Endogenous Switching(SCATE)	0.744	-0.7944	1.1200
Endogenous Switching(PPET)	0.955	0.3731	2.9200
Self Selection(ATE)	0.616	-0.8844	4.5900
Self Selection(ATET)	0.677	-0.5864	4.5400
Self Selection(SCATE)	0.380	-0.8705	1.0500
Ordered Probit(ATE)	0.110	-0.9033	1.1000

Notes: [a] Null hypothesis is stationarity of the chain (p-value), [b] Mean difference test z-score, [c] Dependence factor (threshold of 5)

Figures

Figure 1: Treatment Effects Endogenous Switching: Representative Woman

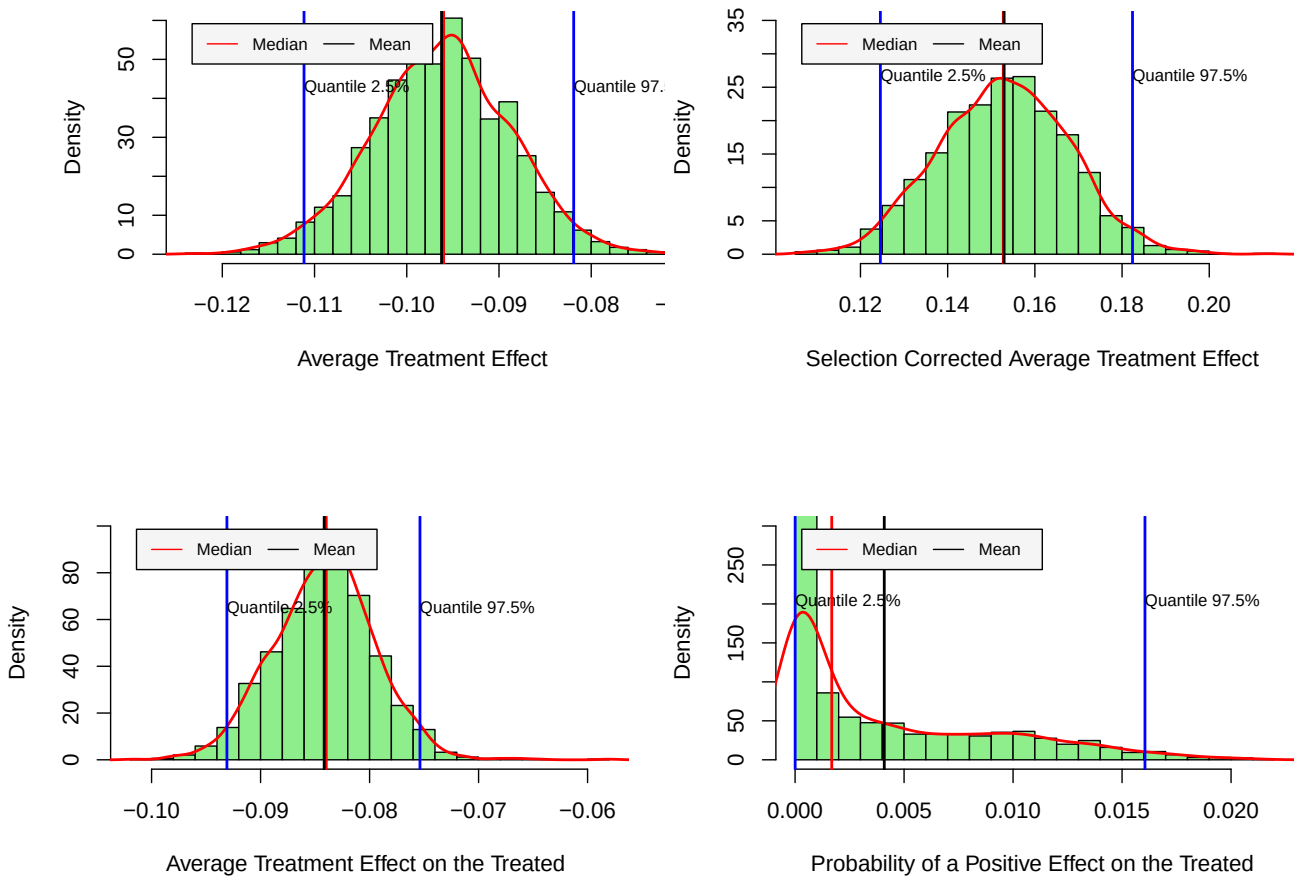
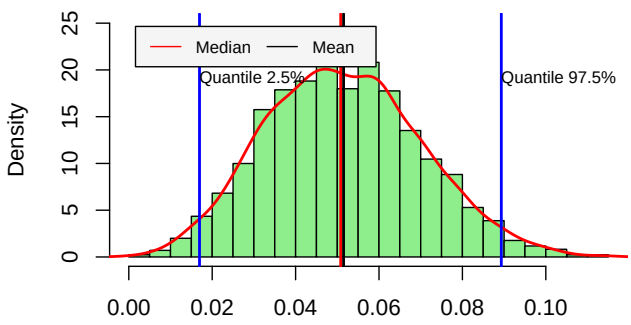
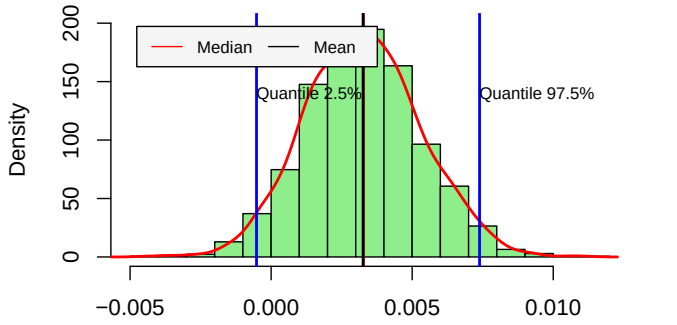
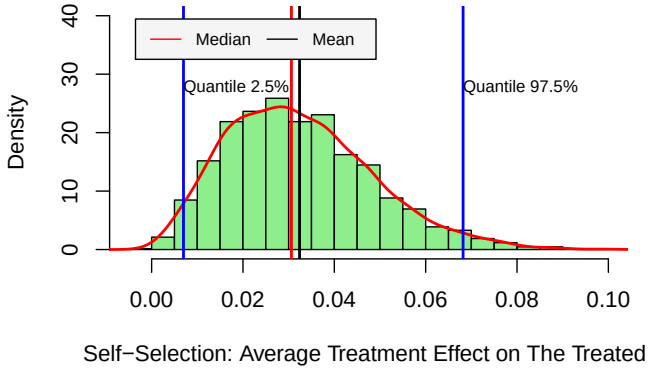


Figure 2: Treatment Effects Self-Selection and Exogenous Ordered Probit: Representative Woman







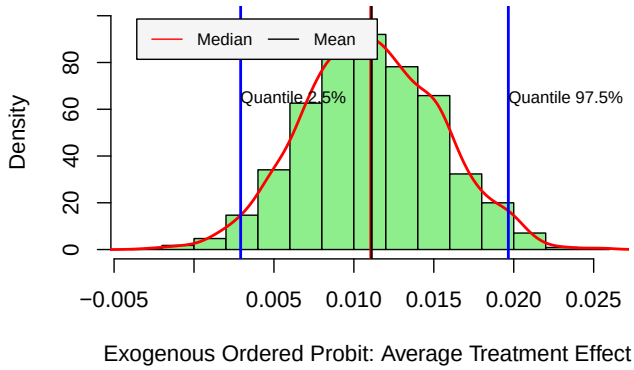


Figure 3: Treatment Effects Endogenous Switching: Representative Man

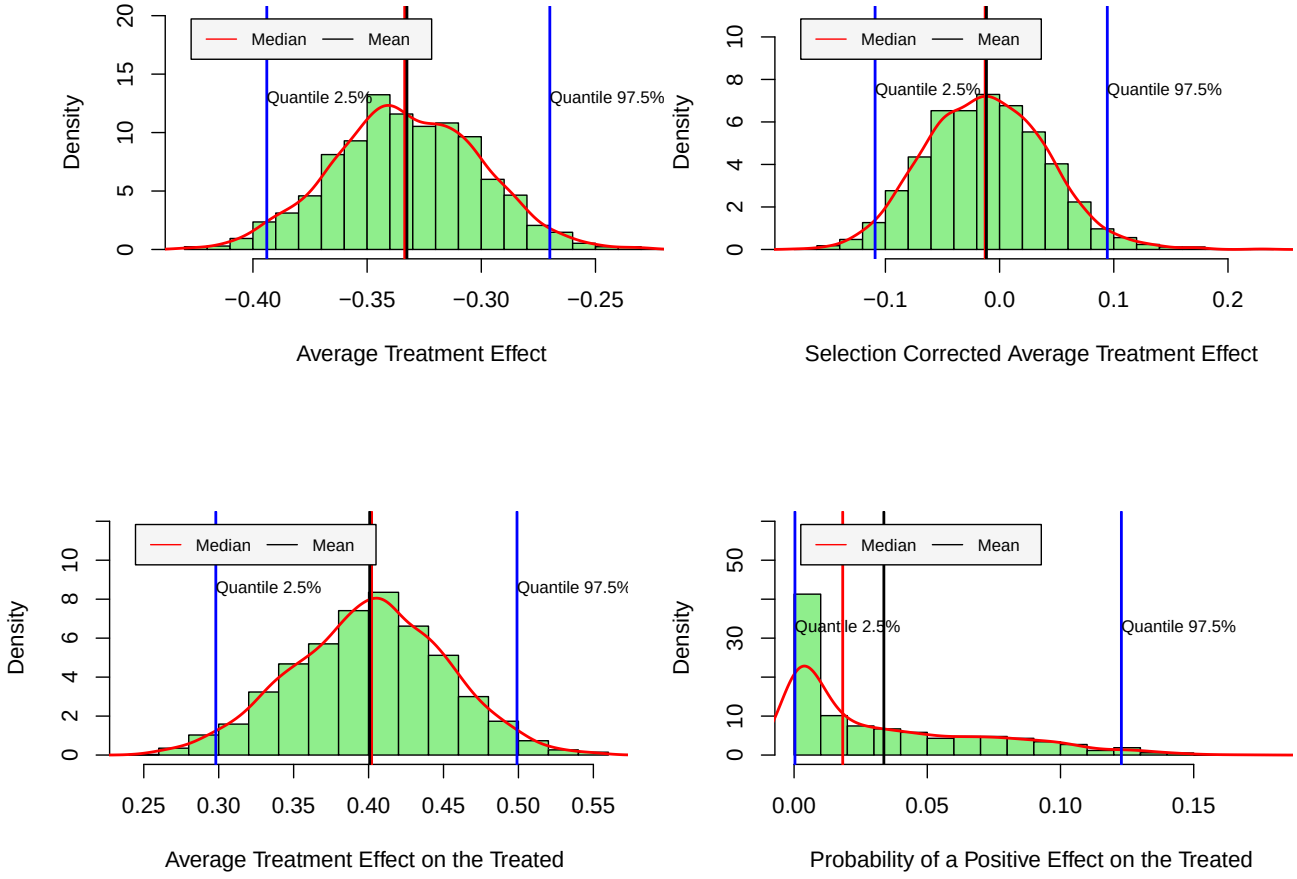


Figure 4: Treatment Effects Self-Selection and Exogenous Ordered Probit: Representative Man