Published in final edited form as:

J Stat Plan Inference. 2012 October 1; 141(10): 3293–3303. doi:10.1016/j.jspi.2011.04.014.

On the estimation of a binary response model in a selected population

Francesco Claudio Stingo^a, Elena Stanghellini^{b,*}, and Rosa Capobianco^c

^aDipartimento di Statistica "G. Parenti", Università di Firenze, Viale Morgagni, 59, 50134, Firenze, Italy

^bDipartimento di Economia, Finanza e Statistica, Università di Perugia, Via Pascoli, 1, 06100, Perugia, Italy

^cDipartimento PE.ME.IS., Università del Sannio, P.zza Arechi, 2, 82100, Benevento, Italy

Abstract

A generalization of the Probit model is presented, with the extended skew-normal cumulative distribution as a link function, which can be used for modelling a binary response variable in the presence of selectivity bias. The estimate of the parameters via ML is addressed, and inference on the parameters expressing the degree of selection is discussed. The assumption underlying the model is that the selection mechanism influences the unmeasured factors and does not affect the explanatory variables. When this assumption is violated, but other conditional independencies hold, then the model proposed here is derived. In particular, the instrumental variable formula still applies and the model results at the second stage of the estimating procedure.

Keywords

directed acyclic graph; extended skew-normal distribution; hidden truncation; instrumental variables; self-selection; unobserved confounder

1 Introduction

We propose a generalization of the Probit model for a binary response which makes use of the extended skew-normal cumulative distribution as a link function. Since the extended skew-normal distribution may be generated by truncation on a latent variable (see e.g. Arnold and Beaver, 2000), the choice allows it to build binary response models generated by a hidden selection mechanism. Probit models for binary response in the presence of selectivity bias have been extensively used in the biomedical literature, and a review is in Bhattacharya *et al.* (2006). We here assume that selection acts as truncation, that is only the population which satisfies the selection criterion is observable. This could be e.g. the self-selected population who volunteer for an experiments, or the patients who qualify for an insurance program. A further example is in the applied section of the paper. The model contains the one proposed by Chen *et al.* (1999) as a particular case, and therefore may arise also from a different data generating mechanism.

We show how to perform ML estimate of the parameters from a sample drawn at random from the selected population. As expected, the inference on one of the parameters which express the selection mechanism, that we denote by α , is problematic. By using the results in

*Corresponding author. elena.stanghellini@stat.unipg.

Rotnitzky *et al.* (2000), we show how it is possible to test for the null hypothesis that $\alpha = 0$, corresponding to the absence of selection. Furthermore, following Copas and Li (1997), we construct a profile log likelihood for this parameter and look at the effects of small deviation of α around zero. Similar approach has been taken by Boehmke (2003) where, however, ML estimate of the parameters of the model has not been pursued.

The model proposed here may also be applied to the complete cases in situations where the explanatory variables are observable for all units in the population and the binary response variable is missing in an informative way. In this case, the parameter related to the proportion of complete cases in the population, which we denote with τ , may be assumed either as known or as belonging to an interval of real values. With this additional information, inference on the model is much simplified.

2 The Extended Skew-Probit model

Let (W, Y^*) be two unobserved random variables such that:

$$\begin{cases} W = \tau + U_W \\ Y^* = -\tilde{\beta}'x + U_{Y^*} \end{cases} \quad (1)$$

where U_W and U_{Y^*} are two jointly Gaussian random variables, with $\text{Var}(U_W) = 1$ and $\text{Var}(U_{Y^*}) = \sigma^2$ and ρ the correlation coefficient. Let x be a k -dimensional vector of independent variables with $x_1 = 1$ and $\tilde{\beta}$ is a k -dimensional vector of unknown coefficients. We now assume that the population is selected, that is, only the observations with $W > 0$ are observed. In this case,

$$(Y^* | W > 0) = -\tilde{\beta}'x + \varepsilon \quad (2)$$

where $\varepsilon = U_{Y^*} | U_W > -\tau$ has an extended skew-normal distribution (Capitanio *et al.*, 2003) with parameters $(0, \sigma, \alpha, \tau)$. The relationship between α and ρ is made explicit by the following:

$$\alpha = \frac{\rho}{\sqrt{1-\rho^2}}.$$

We now assume that a binary indicator Y is observed, such that $Y = 1$ if $Y^* > E(\varepsilon)$, with

$$E(\varepsilon) = \frac{\varphi(\tau)\alpha}{\Phi(\tau)\sqrt{1+\alpha^2}}\sigma. \text{ Then}$$

$$\mu = \text{pr}(Y=1|x) = \text{pr}(\varepsilon \leq E(\varepsilon) + \tilde{\beta}'x) = F(\tilde{\beta}'x) \quad (3)$$

where $\tilde{\beta} = (\tilde{\beta}_1 + E(\varepsilon), \tilde{\beta}_2, \dots, \tilde{\beta}_k)'$, while $F(\tilde{\beta}'x)$ is the distribution function of the residuals at $\tilde{\beta}'x$. By centering the r.v. ε , the model proposed here has been reparameterized in the fashion of Azzalini and Capitanio (1999). If we indicate with $\phi(\cdot)$ the density function of a standard univariate normal distribution and with $\Phi(\cdot)$ its integral, the density function of ε is:

$$f(\varepsilon) = \frac{\varphi(\varepsilon/\sigma)\sigma^{-1}\Phi(\alpha_0 + \alpha\sigma^{-1}\varepsilon)}{\Phi(\tau)} \quad (4)$$

with $\alpha_0 = \tau(1 + \alpha^2)^{1/2}$. The link function at $(\tilde{\beta}'x)$, is then:

$$F(\beta'x) = \Phi(\tau)^{-1} \int_{-\infty}^{\sigma^{-1}\beta'x} \int_{-\infty}^{\alpha_0 + \sigma^{-1}\alpha t} \varphi(t)\varphi(u) dt du. \quad (5)$$

The expression shows that the parameter σ^2 is not identified and therefore it is posed $\sigma^2 = 1$. The unknown parameters of the model are β , α and τ . In Fig. 1 the link function is presented. As the figure shows, the link is non-symmetric and by increasing ρ the function becomes steeper, while by increasing τ the left tail becomes heavier. This effect becomes more evident with high values of ρ . Several particular cases follow.

- a. If $\tau = 0$ then $\Phi(\tau) = 1/2$ and ε is distributed as a skew-normal with parameter α , that is $\varepsilon \sim SN(\alpha)$. This case has been considered in Chen *et al.* (1999), Bazán *et al.* (2006), Capobianco (2006). In this case some symmetries arise, due to the fact that $Z \sim SN(\alpha)$ then $-Z \sim SN(-\alpha)$. Then $F_\alpha(\beta'x) = 1 - F_{-\alpha}(-\beta'x)$ and therefore:

$$\text{pr}_\alpha(Y=1|x) = 1 - \text{pr}_{-\alpha}(Y=1|-x).$$

The last term is equal to $\text{pr}(Y=0 | -x)$ in the model with $\varepsilon \sim SN(-\alpha)$.

- b. If $\alpha = 0$ then $\rho = 0$ (and viceversa). From (4), the density function of the residuals is Gaussian and therefore the model is a Probit model. Note that in this case $E(\varepsilon) = 0$ so $\beta_1 = \beta_1$, the intercept of model (2). This situation corresponds to the absence of selection.
- c. If $\tau = +\infty$ the residuals of model (2) have a Gaussian distribution. To see this notice that, in that case, $\Phi(\tau) = 1$ and $F(\beta'x) = \Phi(\beta'x)$. Therefore, the resulting model is a Probit and also this situation corresponds to the absence of selection.

Note that a different data generating mechanism that leads to the proposed model can also be specified by allowing the introduction of $\tau = 0$ in Chen *et al.* (1999) eqs. (5) and (6). More precisely: let $Y = \delta Z + \varepsilon$ with Z truncated on $Z \geq -\tau$ and $\varepsilon \sim N(0, \sigma_\varepsilon^2)$. Then

$$Y \sim ESN(0, \sigma, \alpha, \tau) \text{ with } \alpha = \frac{\delta}{\sigma_\varepsilon} \text{ and } \sigma = \sqrt{\sigma_\varepsilon^2 + \alpha^2}. \quad (6)$$

The postulated model assumes that the x variables are not influenced by the selection mechanism, that is truncation arises after conditioning on x and it affects only the unmeasured determinants of Y^* . This assumption may be reasonable in experimental studies, where x are exogenously assigned. It is not, however, reasonable in applications outside the experimental designs, where also the explanatory variables are affected by the selection mechanism. These issues are addressed in Section 5.

3 The likelihood function

We assume to have a random sample of n observations (y_i, x_i') , drawn from model (3). Let $\theta = (\beta', \alpha, \tau)'$. The likelihood function of the sample is then:

$$L(\theta) = \prod_{i=1}^n \{F(\beta'x_i)\}^{y_i} \{1 - F(\beta'x_i)\}^{1-y_i} \quad (7)$$

which by making use of (5) becomes:

$$\ell(\theta) = -n \ln \Phi(\tau) + \sum_{i=1}^n (1-y_i) \ln \left\{ \Phi(\tau) - \int_{-\infty}^{\beta' x} \int_{-\infty}^{\alpha_0 + \alpha t} \varphi(t) \varphi(u) du dt \right\} \\ + y_i \ln \left\{ \int_{-\infty}^{\beta' x} \int_{-\infty}^{\alpha_0 + \alpha t} \varphi(t) \varphi(u) du dt \right\}. \quad (8)$$

Maximum likelihood estimate of θ is obtained by partial differentiation of the log likelihood function with respect to β , α and τ . As in the GLM, the MLE are not in closed form. The score functions are:

$$S(\beta_j) = \sum_{i=1}^n x_{ij} A_i \left\{ \frac{y_i - \mu_i}{\mu_i(1-\mu_i)} \right\} \\ S(\alpha) = \sum_{i=1}^n B_i \left\{ \frac{y_i - \mu_i}{\mu_i(1-\mu_i)} \right\} \\ S(\tau) = \sum_{i=1}^n C_i \left\{ \frac{y_i - \mu_i}{\mu_i(1-\mu_i)} \right\}$$

where

$$A_i = \frac{1}{\Phi(\tau)} \varphi(\beta' x_i) \Phi(\alpha_0 + \alpha \beta' x_i) \\ B_i = \frac{\varphi(\tau)}{\Phi(\tau)(1+\alpha^2)} \left[\frac{\alpha \tau}{2} \left\{ 2\Phi(\beta' x_i \sqrt{1+\alpha^2} + \tau \alpha) - 1 \right\} \left\{ 1 - \exp((1-\alpha_0)\alpha \beta' x_i) \right\} + \right. \\ \left. - \exp(-\alpha_0 \alpha \beta' x_i) \varphi \left(\sqrt{(\beta' x_i)^2 (1+\alpha^2) + \alpha^2 \tau^2} \right) + \frac{\varphi(\beta' x_i) \Phi(\alpha_0 + \alpha \beta' x_i)}{\Phi(\tau) \sqrt{1+\alpha^2}} \right] \\ C_i = \frac{\varphi(\tau)}{\Phi(\tau)} \left[\mu_i - \Phi(\beta' x_i \sqrt{1+\alpha^2} + \tau \alpha) + \frac{1}{2} + \frac{\alpha \varphi(\beta' x_i) \Phi(\alpha_0 + \alpha \beta' x_i)}{\Phi(\tau) \sqrt{1+\alpha^2}} \left\{ \tau + \frac{\varphi(\tau)}{\Phi(\tau)} \right\} \right].$$

The information matrix is in Appendix 1.

When a test on $\alpha = 0$ is the main object of the inference, we face two problems. It is easy to prove that for $\alpha = 0$, $B_i = 0$, so that $S(\alpha)$ vanishes and the information matrix is singular. At the same time, $C_i = \frac{\varphi(\tau)}{2\Phi(\tau)}$, therefore τ becomes a nuisance parameter which is not identifiable under the null hypothesis; see Davies (1987).

To address the problem, we first consider τ as known. The second derivative

$$\frac{\partial}{\partial \alpha} S(\alpha) = \sum_{i=1}^n D_i \left\{ \frac{y_i - \mu_i}{\mu_i(1-\mu_i)} \right\} - \sum_{i=1}^n B_i^2 \left\{ \frac{y_i(2\mu_i - 1) - (\mu_i)^2}{(\mu_i)^2(1-\mu_i)^2} \right\} \quad (9)$$

where

$$D_i = \left[\frac{\tau \varphi(\tau)}{2(1+\alpha^2)^2} \left\{ 2\Phi(\beta' x_i \sqrt{1+\alpha^2} + \tau \alpha) - 1 \right\} - \int_{-\infty}^{\beta' x_i} \left(\frac{\tau \alpha}{\sqrt{1+\alpha^2}} + t \right)^2 \varphi(t) \varphi(\alpha_0 + \alpha t) dt + \right. \\ \left. + \frac{\varphi(\tau) \varphi(\beta' x_i)}{\Phi(\tau)(1+\alpha^2)^{3/2}} \left[2 \left(\frac{\tau \alpha}{\sqrt{1+\alpha^2}} + \beta' x_i \right) \varphi(\alpha_0 + \alpha \beta' x_i) - \frac{\Phi(\alpha_0 + \alpha \beta' x_i)}{1+\alpha^2} \left\{ 3\alpha + \frac{\varphi(\tau)}{\Phi(\tau) \sqrt{1+\alpha^2}} \right\} + \frac{\alpha \varphi(\tau) \varphi(\alpha_0 + \alpha \beta' x_i)}{\Phi(\tau)(1+\alpha^2)^{3/2}} \right] \right]$$

is neither zero nor a linear combination of $(S(\beta_1), \dots, S(\beta_k))$. Since the regularity conditions (A1)–(A7) together with the assumptions (B1)–(B3) of Theorem 4 by Rotnitzky *et al.* (2000, p. 261) are satisfied, the asymptotic properties of the MLE of $(\beta_1, \dots, \beta_k, \alpha)$ can be derived. In particular, when $\theta = \theta_0 = (\beta, \alpha = 0)$:

- a. the MLE exists, it is unique with a probability tending to 1 and it is a consistent estimator of θ_0 with the asymptotic distribution as in the Appendix 1;
- b. the likelihood ratio statistics for testing the null hypothesis $H_0 : \theta = \theta_0$ converges in distribution to a mixture of a χ_{k+1}^2 and a χ_k^2 random variables with mixing probability equal to 1/2, i.e.

$$2 \left\{ L_n(\hat{\theta}) - L_n(\tilde{\theta}_0) \right\} \xrightarrow{d} \sum_{i=1}^{k+1} \tilde{Z}_i^2 I_{(\tilde{Z}_{k+1} > 0)} + \sum_{i=1}^k \tilde{Z}_i^2 I_{(\tilde{Z}_{k+1} < 0)}$$

where $\tilde{Z}_i$, for $i=1, \dots, k+1$, are independent $N(0,1)$ random variables.

When τ has to be estimated, the distribution of the stochastic component of the model does not depend on τ for $\alpha = 0$ and standard asymptotic methods are not directly applicable. In this case, for testing the null hypothesis $H_0 : \theta = \theta_0 = (\beta, \alpha = 0, \tau)$, we suggest to resort to simulation of the distribution of the test statistics under H_0 via Monte Carlo methods.

Inference on α when it takes small non-zero values is problematic. In fact, by a linear expansion of $S(\alpha)$ around zero, using (9), it is clear that in the neighbourhood of zero, the profile log likelihood depends on the distance $y_i - \mu_i$ relative to its variance and it is therefore more accurate when the sample contains observations drawn from the tails of the distribution. It is shown in Appendix 1 that the ML estimate of α is $O_p(n^{-1/4})$. The effect of truncation on the estimates of β is best clarified from eq. (10) below.

4 EM algorithm

The EM algorithm may be easily implemented by making reference to the truncated binary regression model (1). To emphasize the relationship with the bivariate Gaussian distribution, in this section we parametrize the model with ρ instead of α . For x_i given, the complete data are then $(u_{Y_i}^*, u_{W_i} | u_{W_i} > -\tau)$ which are observations from a truncated bivariate Gaussian distribution with null expected value and correlation coefficient ρ . The incomplete data are $y_i = I[u_{Y_i}^* \leq \beta' x_i | u_{W_i} > -\tau]$. The log likelihood, for a generic unit i , of the complete model is:

$$\ell(\beta, \rho, \tau; y_i^*, w_i > 0, x_i) = -\ln(2\pi) - \frac{1}{2} \ln(1 - \rho^2) - \frac{1}{2(1 - \rho^2)} \{ (y_i^* + \beta' x_i)^2 + (w_i - \tau)^2 - 2\rho(y_i^* + \beta' x_i)(w_i - \tau) \} - \ln\Phi(\tau).$$

E-step

The expected log likelihood of the complete data of a random sample of n observations is then:

$$Q(\theta, \theta_p) = -n \ln(2\pi) - n \ln\Phi(\tau) - \frac{n}{2} \ln(1 - \rho^2) - \frac{1}{2(1 - \rho^2)} \sum_{i=1}^n (v_i^{20} + v_i^{02} - 2\rho v_i^{11})$$

where $v_i^{20} = E\{(Y_i^* + \beta' x_i)^2 | \mathcal{J}_p\}$, $v_i^{02} = E\{(W_i - \tau)^2 | \mathcal{J}_p\}$ and $v_i^{11} = E\{(Y_i^* + \beta' x_i)(W_i - \tau) | \mathcal{J}_p\}$, with $\tau = \{y_i, w_i > 0, x_i, \beta_p, \rho_p, \tau_p\}$. The expressions of the expected values can be derived from Rosembaum (1961) and are reported in Appendix 2.

M-step

We perform the maximization in three steps. First, we maximize w.r. to β for fixed $\rho = \rho_p$ and $\tau = \tau_p$. Maximization w.r. to β can be performed in a closed form, that is

$$\beta_{p+1} - \beta_p = - \left(\sum_{i=1}^n x_i x_i' \right)^{-1} \sum_{i=1}^n \{E(U_{Y_i^*} | \mathcal{J}_p) - \rho E(U_{W_i} | \mathcal{J}_p)\} x_i. \quad (10)$$

Maximization w.r. to ρ for $\beta = \beta_{p+1}$ and $\tau = \tau_p$ reduces to maximization of the following function:

$$-\frac{n}{2} \ln(1 - \rho^2) - \frac{1}{2(1 - \rho^2)} \sum_{i=1}^n (v_i^{20} + v_i^{02} - 2\rho v_i^{11}),$$

while maximization w.r. to τ for $\beta = \beta_{p+1}$ and $\rho = \rho_{p+1}$, reduces to maximization of the following function:

$$-n \ln \Phi(\tau) - \frac{1}{2(1 - \rho^2)} \sum_{i=1}^n (v_i^{02} - 2\rho v_i^{11}).$$

The last two functions are easily maximized via numerical methods. If the range of possible values of τ is known, as in some applications, this information can be included in the procedure. Note that equation (10) highlights the difference between the ML estimates of the regression coefficients in a Probit and in an Extended Skew-Probit model and, for $\rho = 0$ or $\tau = +\infty$, it reduces to the M-step of a Probit model.

The EM algorithm may be easily modified to implement an estimating procedure based on the profile log likelihood for ρ or τ . As an instance, for each value of ρ , the MLE of β and τ is easily computed. Examples of profile log likelihood for ρ are in Fig. 2. An example of a profile log likelihood for τ is in Section 6.

Note that, in parallel to the skew-normal case (Azzalini, 2005), for moderate sample size we can have no finite maximum for the shape parameter α . For the skew-normal case, a solution based on the Firth's method for bias prevention of the maximum likelihood estimator is in Sartori (2006).

5 Endogenous regressors

We illustrate situations where a binary response model with extended skew-normal link function may arise. Let $V = (Y_1, \dots, Y_k)'$ be a multivariate random vector with jointly Gaussian distribution, with $\mu = E[V]$. Let $\text{cov}(V) = \Sigma$ and $\text{corr}(V) = \tilde{\Sigma}$ be, in order, the covariance and the correlation matrix. We partition $V = (Y_o', Y_s')'$. The vector μ is partitioned accordingly, as $\mu = (\mu_o', \mu_s')'$. The matrices Σ and $\tilde{\Sigma}$ are then:

$$\Sigma = \begin{pmatrix} \Sigma_{oo} & \delta \\ \cdot & \sigma_{ss} \end{pmatrix}, \tilde{\Sigma} = \begin{pmatrix} \tilde{\Sigma}_{oo} & \tilde{\delta} \\ \cdot & 1 \end{pmatrix}.$$

We assume that the observed population is selected according to $Y_s > -\tau$. Since μ_s and σ_{ss} are not identifiable, we pose $\mu_s = 0$ and $\sigma_{ss} = 1$. The observed marginal distribution of $(Y_o | Y_s > -\tau)$ is a multivariate extended skew-normal distribution with parameters $(\mu_o, \alpha, \Sigma_{oo}, \tau)$

with $\alpha = (1 - \delta' \tilde{\Sigma}_{oo}^{-1} \tilde{\delta})^{-1/2} \tilde{\Sigma}_{oo}^{-1} \tilde{\delta}$. The covariance matrix Ω_{oo} of the observable random variables is

$$\Omega_{oo} = \Sigma_{oo} + \delta [Var(Y_s | Y_s > -\tau) - 1] \delta^T.$$

Let

$$Z_2(\tau) = \frac{d^2}{d\tau^2} \log[2\Phi(\tau)] = -\frac{\varphi(\tau)}{\Phi(\tau)} \left[\tau + \frac{\varphi(\tau)}{\Phi(\tau)} \right].$$

After noting that $Z_2(\tau) = Var(Y_s | Y_s > -\tau) - 1$ (see e.g. Wooldridge, 2003, p. 522), we see that the expression is equivalent to the one in Capitanio *et al.* (2003, p. 131).

Let β_{ij} be the partial regression coefficients of Y_i in the regression of Y_i on all the other variables. If we may assume that an ordering between the variables exists, such that Y_i is a potential explanatory variable of Y_j , $i > j$, then we can write $B(V - \mu) = U$ where B is an upper triangular matrix with ones along the diagonal and $-\beta_{ij}$ as ij -th element, $i > j$. The errors U have zero means and are uncorrelated, so that $cov(U) = \Psi = \{\psi_{ij}\}$ is a diagonal matrix. In this case, the distribution is generated over a Directed Acyclic Graph (DAG); see Lauritzen (1996), Section 3.2.2. Then, $\Sigma = B^{-T} \Psi B^{-1}$ and the linear system, with some elements β_{ij} restricted to zero, is known as path analysis from the work of Wright (1923, 1934). Several interesting cases then arise. We here detail two of them.

Let $V = (Y_1^*, Y_2, Y_3, Y_4)'$, and $Y_s = Y_4$. Let Y_i^* be the response variable of interest. We here assume that truncation on Y_4 occurs after conditioning. In the saturated model, the expected values of $E(Y_1^* | Y_2, Y_3, Y_4)$ and $E(Y_1^* | Y_2, Y_3, Y_4 > -\tau)$ do not coincide, as:

$$E(Y_i^* | Y_2, Y_3, Y_4 > -\tau) = \mu_1 + \beta_{12}(Y_2 - \mu_2) + \beta_{13}(Y_3 - \mu_3) + \beta_{14} \frac{\varphi(\tau^*)}{\Phi(\tau^*)}$$

with $\tau^* = \frac{\tau + \beta_{42}^*(Y_2 - \mu_2) + \beta_{43}^*(Y_3 - \mu_3)}{Var(Y_4 | Y_2, Y_3)}$ and β_{42}^* and β_{43}^* are, in order, the coefficients of Y_2 and Y_3 in the linear regression of Y_4 against Y_2 and Y_3 only. Estimation of the parameters is difficult, due to the dependence of τ^* from the unobserved variable Y_4 and the conditioning variables Y_2 and Y_3 .

Consider the situation with Y_3 a randomized treatment, Y_1^* the response variable of interest, and assume that $\beta_{12} = \beta_{34} = 0$. The object of the inference is β_{13} , i.e. the direct effect of the treatment Y_3 on the response variable. The DAG of the model before truncation is in Fig.

3(a). Assume further that the response is observed only in subjects with $Y_4 > -\tau$. Then, from (6), the distribution of $Y_1^*|Y_3, Y_4 > -\tau$ is an extended skew-normal and the coefficient β_{13} can be estimated as the coefficient of Y_3 in the equation

$$E(Y_1^*|Y_3, Y_4 > -\tau) = \mu_1 + \beta_{13}(Y_3 - \mu_3) + \beta_{14} \frac{\varphi(\tau)}{\Phi(\tau)}.$$

If only the dichotomized version Y_1 of the response variable is observable, then the model of Y_1 against the explanatory variable Y_3 , after truncation, is an Extended Skew-Probit.

The second model is with $\beta_{13} = \beta_{34} = 0$. The DAG of the model before truncation is in Fig. 3(b). Then, $(Y_1^*|Y_3, Y_4 > -\tau)$ and $(Y_2|Y_3, Y_4 > -\tau)$ have an extended skew-normal distribution with

$$E(Y_1^*|Y_3, Y_4 > -\tau) = \mu_1 + (\beta_{12}\beta_{24} + \beta_{14}) \frac{\varphi(\tau)}{\Phi(\tau)} + \beta_{12}\beta_{23}(Y_3 - \mu_3)$$

and

$$E(Y_2|Y_3, Y_4 > -\tau) = \mu_2 + \beta_{24} \frac{\varphi(\tau)}{\Phi(\tau)} + \beta_{23}(Y_3 - \mu_3).$$

The two equations lead naturally to an instrumental variable estimator; see Bowden and Turkington (1990). When only the dichotomized version Y_1 of Y_1^* is observed, then the model for the observable binary variable Y_1 against Y_3 is an Extended Skew-Probit model. Note that Y_2 and Y_3 are allowed to be vectors, provided that the sufficient conditions of identification are satisfied, see e.g. Mardia *et al.* (1979), p. 194.

This result is in line with other results in the literature concerning the instrumental variable estimators; see e.g. Bhattacharya *et al.* (2006). Since in this model only the selected population is observed, the novelty here lays on the fact that (a) the use of an instrumental variable provides a solution to the problem of estimation also in this case and (b) the model resulting at the second stage of the estimating procedure is an Extended Skew-Probit.

Both models apply to the complete cases in experiments when the binary response variable Y_1 is defined only if Y_2 takes a value larger than a threshold; see e.g. Gilbert *et al.* (2006). In this case, typically, $\beta_{14} = \beta_{24} = 1$ and truncation on Y_4 is replaced by censoring of the kind $Y_2 > 0$. Then $\tau = \mu_2 + \beta_{23}(Y_3 - \mu_3)$ and the model is related to the Heckman's model (Heckman, 1979) with a binary outcome. For this class of model, full maximum likelihood estimating procedures exist; see e.g. Wooldridge (2003), p. 570. As there suggested, two step estimation method may also be implemented, with β_{23} estimated in the first step via a Probit regression. The derivations in this paper apply to the complete cases, and could be used in the second step of the two stage estimating procedure.

6 Health scores data

Some outcome of interest in medical research, such as mental or physical health, cannot be measured directly. A way to assess them is to ask the patients to fill a questionnaire. Each

item aims at measuring one particular aspect of health or ability, by providing a numerical evaluation, or score. Typically, scores have a finite range. The overall quality of life score is formed as a weighted sum of different scores, and is therefore bounded over an interval. For repeated measurements, later scores tend to cluster at one end of the range, usually associated with full recovery, giving raise to a censored distribution. When the aim is to compare the effect of different treatments, statistical methods have to take into account this censoring mechanism, that can bias the treatment contrasts. One possible strategy, that we here pursue, is to dichotomize the distribution of the score, considering success the full recovery event.

We here analyze data coming from a randomized controlled clinical trial for comparing four treatments (1. tubigrip, i.e. elastic bandage, 2. below-knee plaster-cast, 3. splints or 4. boot) for acute, severe ankle sprain. 565 patients were randomized into the four treatments. Ankle function was measured by the Foot and Ankle score (FAOS), a questionnaire on pain, symptoms, activities of daily living, ability to participate in sports and quality of life. Patients were asked to complete the questionnaire at baseline, and at 1, 3 and 9 months after injury. The research question is how plaster-cast, splints and boots compare with tubigrip, which is by far the cheapest of the four treatments. The primary timepoint was 3 months. See Lamb *et al.* (2009) for a full account on the trial.

We report analysis on the sport and recreation score after 3 months, which is based on 5 items. This is formed as a weighted average of each score, that is scaled to the range from 0 to 100; extreme pain or limitation scores 0, and no symptoms score 100. We pose tubigrip as the baseline treatment.

There is a quite relevant non-response rate (134 non respondent, about 24%). A qualitative study, Nakash *et al.* (2008), in which non-respondents were re-interviewed, reported that nearly half of them had made a full recovery by 3 months. Therefore, drop out cannot be assumed non informative, as the patients that stay in the trial seem to be the ones exhibiting higher difficulties in recovering. However, the drop-out seems to be equally distributed into the four treatments.

The healing process is known to depend on age and gender, which we include in the analysis (56% are males; median and mean age are, in order, 28 and 30.21). In Figure 4 the histogram (with a non parametric density estimation superimposed) of the score after 3 months (marginal with respect to treatment, age and sex) is presented. The clustering of the observations at 100 is rather extreme (about 16% of the observations). Data exhibit also a substantial negative skewness.

To take into account all the features of the data, we dichotomize the response into 0 if the score is below 100 and 1 otherwise. Other strategies have been proposed by Hutton and Stanghellini (2011). Table 1 reports the percentage of patients with score equal to 100 by the four treatments. We specify the link function as an extended skew-normal with the τ parameter expressing the proportion of complete cases in the population; the α parameter is responsible of the skewness.

As there may be some near unidentifiability problems, due to the presence of the further parameter τ that has to be estimated in conjunction with the other parameters of the model, we address the problem by first constructing the profile log-likelihood as a function of τ , in parallel to what done by Capitanio *et al.* 2003). This leads to a much more stable behavior of the optimization algorithms. We let τ vary in the wide range $(-0.4, 0.8)$, and maximize the log likelihood with respect to the other parameters. Figure 5 shows that inference based on the normalized profile matches the prior expectation about the percentage of selected

population, as the estimate of τ leads to an estimated probability of being in the selected population $\Phi(\hat{\tau}) = 0.68$.

Table 7 reports the estimates of the ESP model, for $\tau = \hat{\tau}$. The intercept corresponds to the baseline treatment, that we pose equal to tubigrip; $\beta_1, \beta_2, \beta_3$ are the parameters of, in order, the plaster cast, splints and boots treatments; β_4 is the age effect while β_5 is the gender effect (male=1). For comparison, the corresponding estimates obtained with the Probit model are also reported. Standard errors are obtained from the information matrix using the methodology for reduced models proposed by Cox and Wermuth (1990). The estimate of α (significantly different from 0) corresponds to an estimated ρ of -0.45 ($-0.453, -0.449$). Note that the negative sign confirms the expectations that the selected population tends to be formed by the patients with more difficulties in the healing process.

The estimates with the Probit do not vary significantly, a part from the lower intercept to accommodate for the symmetry. Both models indicate that plaster cast exhibits a higher probability of full recovery than the other treatments, with a p-value=0.063 in the ESP model and 0.037 in the Probit model; age has a negative but not highly significant effect, while both models indicate that gender is significantly positive improving the probability of a full recovery (p-value=0.049 in the ESP and 0.025 in the Probit).

7 Concluding remarks

We have proposed a binary response model with asymmetric link function, which comprises the Probit model as a particular case. Other models with asymmetric links exist in the literature, and a review is in Bazán *et al.* (2006). The consequences of a misspecified link on ML estimate of the regression parameters are investigated in Czado and Santner (1992). The model used in this paper can be seen as arising from a particular form of selection which can be modelled by truncation on one hidden dimension. We have shown how to perform inference based on the likelihood function of the model. When a test for $\alpha = 0$ is the object of the inference, we distinguish two cases. For τ known, the asymptotic distribution of the likelihood ratio test statistics under the null hypothesis has been derived as a mixture of χ^2 distributions. When τ has to be estimated, testing for $\alpha = 0$ is problematic and the null distribution of the likelihood ratio is not easily obtainable.

The assumption underlying the model is that selection influences the response variable, but does not affect the values of the explanatory variables. When this assumption is violated, the model complicates further. However, we have shown that if some conditional independencies can be assumed, the model presented here is then derived. Furthermore, when truncation is replaced by censoring, the model may be applied to the complete cases. In this case, a two step estimation method may be implemented. This paper details the estimation problems resulting at the second step of the procedure.

Simulation results show that the EM algorithm tends to be quite stable, especially with reference to the estimation of the parameters β . The MLE of ρ is sensitive with respect to misspecification of the covariates of the model. Simulations show that if the covariates are entered in a way that does not heavily deviate from a linear transformation, the estimate of ρ is quite stable. This holds true also if additional terms are included in the model. However, if some covariates are either removed or included in a non-linear transformed way, then the estimate of ρ becomes less reliable. For small samples, we suggest to estimate the parameters via a procedure based on maximization of the profile log likelihood. We here presented results on the maximization of the profile for α , with τ and β as nuisance parameters. In our experience, this procedure tends to be very stable. Routines can be downloaded at the web site: <http://www.stat.unipg.it/stanghellini/elena.html>. Note that the

estimates of α and τ are strongly positively correlated. This is expected as, given the data, in the region where α is positive (negative), an increase in the estimated value of τ is almost compensated by an increase (decrease) in the estimated value of α .

Besides the EM algorithm, other methods to maximize the log likelihood, such as the Fisher scoring and the Nelder-Mead algorithm, have been implemented. For τ arbitrarily fixed, simulation studies show that satisfactory results are obtained with the Nelder-Mead algorithm. We also noticed that, for high values of α , the Fisher scoring algorithm does not update the starting values, since the elements of the information matrix become too large. However, for a very small fraction of our samples (3 out of 4000 in our case), the Nelder-Mead procedure too does not update the initial estimates of β . In our experience, the EM tends to be the most reliable of the algorithms.

The practical implications of the proposed model have been shown on a randomized controlled trial on patients affected by severely ankle sprain. The patients drop-out is rather substantial and cannot be assumed to be uninformative. The analysis shows that, compared to the others, one treatment shows moderate positive effect on the probability of full recovery, also after adjusting for the asymmetry induced by the selection mechanism.

Acknowledgments

We thank Prof Sallie Lamb, Prof Matthew Cooke and the CAST team for permission to use these data, and Prof Jane Hutton for useful insights on the data analysis. The CAST trial was funded by a grant from the UK Department of Health through its Health Technology Assessment Programme, project no. 01/14/10. Part of this paper was completed when the first author was visiting the Department of Statistics, University of Warwick, the hospitality of which is gratefully acknowledged. This research was supported by grants PRIN 2007XECZ7L_003, and PRIN2006132978, from MIUR, Italy.

References

- Arnold BC, Beaver RJ. Hidden Truncation models. *Sankhya*. 2000; 62:23–35.
- Azzalini A, Capitanio A. Some properties of the multivariate skew-normal distribution. *Journal of the Royal Statistical Society B*. 1999; 61:579–602.
- Azzalini A. The Skew-normal Distribution and Related Multivariate Families. *Scandinavian Journal of Statistics*. 2005; 32:159–188.
- Bazán, JL.; Bolfarine, H.; Branco, MD. A Generalized Skew-Probit class link for binary regression. Universidade de Sao Paulo; Brasil: 2006. RT-MAE 2006-05
- Bhattacharya J, Goldman D, McCaffrey D. Estimating probit models with self-selected treatments. *Statistics in Medicine*. 2006; 25:389–413. [PubMed: 16382420]
- Boehmke FJ. Using Auxiliary Data to Estimate Selection Bias Models, with an Application to Interest Group Use of the Direct Initiative Process. *Political Analysis*. 2003; 11, 3:234–254.
- Bowden, RJ.; Turkington, DA. Instrumental variables. Cambridge University Press; 1990.
- Capitanio A, Azzalini A, Stanghellini E. Graphical models for skew-normal variates. *Scandinavian Journal of Statistics*. 2003; 30:129–144.
- Capobianco, R. Atti XLIII Riunione Scientifica SIS. Sessioni Spontanee; 2006. Discrete choice models with skewed link; p. 559-562.
- Chen MH, Dey DK, Shao QM. A New Skewed Link Model for Dichotomous Quantal Response Data. *Journal of the American Statistical Association*. 1999; 94:1172–1186.
- Cox DR, Wermuth N. An approximation to maximum likelihood estimates in reduced models. *Biometrika*. 1990; 77:747–61.
- Copas JB, Li HG. Inference for non-random samples. *Journal of the Royal Statistical Society B*. 1997; 59:55–95.
- Czado C, Santner TJ. The effect of Link Misspecification on Binary Regression Inference. *Journal of Statistical Planning and Inference*. 1992; 33:213–231.

- Davies RB. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*. 1987; 74:33–43.
- Gilbert PG, Bosch RJ, Hudgens MG. Sensitivity Analysis for the Assessment of Causal Vaccine Effects on Viral Load in HIV Vaccine Trials. *Biometrics*. 2003; 59:531–41. [PubMed: 14601754]
- Hutton JL, Stanghellini E. Modelling health scores with censored skew-normal distributions. *Statistics in Medicine*. 2011 to appear.
- Heckman JJ. Sample selection bias as a specification error. *Econometrica*. 1979; 47:153–161.
- Lamb SE, Marsh JL, Hutton JL, Nakash RA, Cooke MW. Mechanical supports for acute, severe ankle sprains: a pragmatic, multi-centre, randomised controlled trial. *The Lancet*. 2009; 373:575–581.
- Lauritzen, SL. *Graphical Models*. Oxford University Press; Oxford: 1996.
- Mardia, KV.; Kent, JY.; Bibby, JM. *Multivariate Analysis*. Academic Press; London: 1979.
- Nakash RA, Hutton JL, Lamb SE, Gates S, Fisher J. Response and non-response to postal questionnaire follow-up in a clinical trial - a qualitative study of the patient's perspective. *BMC Med Res Meth*. 2008; 14:226–235.
- Rosenbaum S. Moments of the truncated bivariate normal distribution. *Journal of the Royal Statistical Association B*. 1961; 23:405–8.
- Rotnitzky A, Cox DR, Bottai M, Robins J. Likelihood-based inference with singular information matrix. *Bernoulli*. 2000; 6:395–401.
- Sartori N. Bias prevention of maximum likelihood estimates for scalar skew normal and skew t distributions. *Journal of Statistical Planning and Inference*. 2006; 136:4259–75.
- Wooldridge, JM. *Econometrics Analysis of Cross Section and Panel data*. The MIT Press; Cambridge (MA): 2003.
- Wright S. The theory of path coefficients: a reply to Niles' criticism. *Genetics*. 1923; 8:239–55. [PubMed: 17246011]
- Wright S. The method of path coefficients. *Ann Statist*. 1934; 5:161–215.

Appendix 1

The information matrix is the following:

$$I = \begin{pmatrix} I_{\beta\beta} & I_{\tau\beta} & I_{\alpha\beta} \\ I_{\beta\tau} & I_{\tau\tau} & I_{\alpha\tau} \\ I_{\beta\alpha} & I_{\tau\alpha} & I_{\alpha\alpha} \end{pmatrix} = \begin{pmatrix} X' \Psi X & X' \Gamma & X' \Delta \\ \Gamma' X & \underline{1}' \Xi & \underline{1}' \Lambda \\ \Delta' X & \Lambda' \underline{1} & \underline{1}' \Pi \end{pmatrix}$$

where Ψ is an $n \times n$ diagonal matrix whose i -th element is equal to

$$\psi_i = A_i^2 / \{\mu_i(1-\mu_i)\},$$

$\Gamma' = (\gamma_1, \dots, \gamma_n)'$ with

$$\gamma_i = A_i C_i / \{\mu_i(1-\mu_i)\};$$

$\Delta' = (\delta_1, \dots, \delta_n)'$ with

$$\delta_i = A_i B_i / \{\mu_i(1-\mu_i)\};$$

$\underline{1}'$ is a vector of order $1 \times n$ whose elements are all equal to 1 while $\Xi' = (\xi_1, \dots, \xi_n)'$ with

$$\xi_i = C_i^2 / \{\mu_i(1-\mu_i)\};$$

$$\Lambda' = (\lambda_1, \dots, \lambda_n)' \text{ with}$$

$$\lambda_i = B_i C_i / \{\mu_i(1-\mu_i)\};$$

$$\Pi' = (\pi_1, \dots, \pi_n)' \text{ with}$$

$$\pi_i = B_i^2 / \{\mu_i(1-\mu_i)\}.$$

Under H_0 : $\theta = \theta_0$, for τ fixed we have:

$$\begin{pmatrix} n^{1/2} \left(\hat{\beta}_1 - \beta_1 - \frac{\varphi(\tau)}{\Phi(\tau)} \frac{\alpha}{\sqrt{1+\alpha^2}} \right) \\ \vdots \\ n^{1/2} (\hat{\beta}_k - \beta_k) \\ n^{1/4} \hat{\alpha} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} Z_1 \\ \vdots \\ Z_k \\ (-1)^F Z_{k+1}^{1/2} \end{pmatrix} I_{(Z_{k+1} > 0)} + \begin{pmatrix} Z_1 - (I^{(k+1)1} / I^{(k+1)(k+1)}) Z_{k+1} \\ \vdots \\ Z_k - (I^{(k+1)k} / I^{(k+1)(k+1)}) Z_{k+1} \\ 0 \end{pmatrix} I_{(Z_{k+1} < 0)},$$

where $Z = (Z_1, \dots, Z_{k+1})$ denotes a mean-zero normal random variable with variance equal to I^{-1} and F denotes a Bernoulli variable with success probability equal to 1/2 that is independent of Z , $I_{(\omega)}$ denotes the indicator function, I^{ij} denotes the ij -th entry of the inverse of the information matrix of $(S(\beta_1), \dots, S(\beta_k), \frac{\partial}{\partial \alpha} S(\alpha)/2)$,

$$I|_{\alpha=0} = \begin{pmatrix} I_{\beta\beta}^0 & I_{\alpha\beta}^0 \\ I_{\alpha\tau}^0 & I_{\alpha\alpha}^0 \end{pmatrix}$$

where $I_{\beta\beta}^0 = I_{\beta\beta}$, $I_{\alpha\beta}^0 = X' \Delta^0$, where $\Delta^0 = (\delta_1^0, \dots, \delta_n^0)'$ with

$$\delta_i^0 = \frac{1}{2} A_i \left\{ \frac{D_i \mu_i (1-\mu_i) - B_i^2 (2\mu_i - 1)}{\mu_i^2 (1-\mu_i)^2} \right\};$$

$I_{\alpha\alpha}^0 = \mathbf{1}' \Pi^0$ where $\Pi^0 = (\pi_1, \dots, \pi_n)'$ with

$$\pi_i = \frac{1}{4\mu_i(1-\mu_i)} \left\{ D_i^2 + \frac{B_i^2(8\mu_i^2 - 4\mu_i^3 - 5\mu_i + 1)}{\mu_i^2(1-\mu_i)^3} - 2 \frac{D_i B_i^2(2\mu_i - 1)}{\mu_i(1-\mu_i)} \right\}.$$

Appendix 2

The expressions of the expected values are the following:

$$\begin{aligned} v_i^{20} &= E\{(Y_i^* + \beta' x_i)^2 | I_p\} &= m_i^{20} - 2\beta'_p x_i m_i^{10} + (\beta'_p x_i)^2 + (\beta' x_i)^2 + 2\beta' x_i (m_i^{10} - \beta'_p x_i) \\ v_i^{02} &= E\{(W_i - \tau)^2 | I_p\} &= m_i^{02} + 2\tau_p m_i^{01} + \tau_p^2 + \tau^2 - 2\tau(m_i^{01} + \tau_p) \\ v_i^{11} &= E\{(Y_i^* + \beta' x_i)(W_i - \tau) | I_p\} &= -2\rho\{m_i^{11} + m_i^{10}(\tau_p - \tau) + m_i^{01}(\beta' x_i - \beta'_p x_i) - \tau_p \beta'_p x_i + \tau \beta'_p x_i + \tau_p \beta' x_i - \tau \beta' x_i\} \end{aligned}$$

where, for $y_i = 0$:

$$\begin{aligned} L_1 &= \int_{\beta'_p x_i}^{\infty} \int_{-\tau_p}^{\infty} \frac{1}{2\pi \sqrt{1-\rho_p^2}} \exp\left\{-\frac{1}{2(1-\rho_p^2)}(z_1^2 + z_2^2 - \rho_p z_1 z_2)\right\} dz_1 dz_2 \\ m_i^{10} &= \frac{1}{L_1} \left[\varphi(\beta'_p x_i) \left\{ 1 - \Phi\left(\frac{-\tau_p - \rho_p \beta'_p x_i}{\sqrt{1-\rho_p^2}}\right) \right\} + \rho_p \varphi(-\tau_p) \left\{ 1 - \Phi\left(\frac{\beta'_p x_i + \rho_p \tau_p}{\sqrt{1-\rho_p^2}}\right) \right\} \right] \\ m_i^{01} &= \frac{1}{L_1} \left[\rho_p \varphi(\beta'_p x_i) \left\{ 1 - \Phi\left(\frac{-\tau_p - \rho_p \beta'_p x_i}{\sqrt{1-\rho_p^2}}\right) \right\} + \varphi(-\tau_p) \left\{ 1 - \Phi\left(\frac{\beta'_p x_i + \rho_p \tau_p}{\sqrt{1-\rho_p^2}}\right) \right\} \right] \\ m_i^{11} &= \frac{1}{L_1} \left[\rho_p * L_1 + \{\rho_p \beta'_p x_i \varphi(\beta'_p x_i)\} \left\{ 1 - \Phi\left(\frac{-\tau_p - \rho_p \beta'_p x_i}{\sqrt{1-\rho_p^2}}\right) \right\} + \rho_p (-\tau_p) \varphi(-\tau_p) \left\{ 1 - \Phi\left(\frac{\beta'_p x_i + \rho_p \tau_p}{\sqrt{1-\rho_p^2}}\right) \right\} \right. \\ &\quad \left. + \frac{\sqrt{1-\rho_p^2}}{\sqrt{2\pi}} \varphi\left(\sqrt{\frac{(\beta'_p x_i)^2 + 2\rho_p \beta'_p x_i \tau_p + \tau_p^2}{1-\rho_p^2}}\right) \right] \\ m_i^{20} &= \frac{1}{L_1} \left[L_1 + \{\beta'_p x_i \varphi(\beta'_p x_i)\} \left\{ 1 - \Phi\left(\frac{-\tau_p - \rho_p \beta'_p x_i}{\sqrt{1-\rho_p^2}}\right) \right\} + \rho_p^2 (-\tau_p) \varphi(-\tau_p) \left\{ 1 - \Phi\left(\frac{\beta'_p x_i + \rho_p \tau_p}{\sqrt{1-\rho_p^2}}\right) \right\} \right. \\ &\quad \left. + \frac{\rho_p \sqrt{1-\rho_p^2}}{\sqrt{2\pi}} \varphi\left(\sqrt{\frac{(\beta'_p x_i)^2 + 2\rho_p \beta'_p x_i \tau_p + \tau_p^2}{1-\rho_p^2}}\right) \right] \\ m_i^{02} &= \frac{1}{L_1} \left[L_1 + \rho_p^2 (\beta'_p x_i) \varphi(\beta'_p x_i) \left\{ 1 - \Phi\left(\frac{-\tau_p - \rho_p \beta'_p x_i}{\sqrt{1-\rho_p^2}}\right) \right\} + (-\tau_p) \varphi(-\tau_p) \left\{ 1 - \Phi\left(\frac{\beta'_p x_i + \rho_p \tau_p}{\sqrt{1-\rho_p^2}}\right) \right\} \right. \\ &\quad \left. + \frac{\rho_p \sqrt{1-\rho_p^2}}{\sqrt{2\pi}} \varphi\left(\sqrt{\frac{(\beta'_p x_i)^2 + 2\rho_p \beta'_p x_i \tau_p + \tau_p^2}{1-\rho_p^2}}\right) \right] \end{aligned}$$

while for $y_i = 1$, see Rosebaum (1961).

Fig. 1.

The extended skew-normal (continuous line) versus the normal (dashed line) link functions

Fig. 2.
The profile log-likelihood for $\rho = 0.95$, $k = 3$, $n = 200$ and (a) $\tau = 0$, (b) $\tau = 0.5$

Fig. 3. Conditional independence graph before truncation of the model with (a) $\beta_{12} = \beta_{34} = 0$ and (b) $\beta_{13} = \beta_{34} = 0$

Fig. 4. Histogram of the sport and recreation score with a non parametric density estimation superimposed.

Fig. 5.
Normalized profile log likelihood for τ

Table 1

The four treatments and their percentage of success

Treatment	Percentage of success	Number of patients
Tubigrip	12.2 %	115
Below-knee plaster-cast	23.4 %	111
Splints	13.0 %	115
Boot	13.9 %	122

Table 2

ML estimates of ESP e Probit model

Parameter	ESP estimates			Probit estimates		
	MLE	S.e.	95% C.I.	MLE	S.e.	95% C.I.
Intercept	-1.4118	0.4118	(-2.22, -0.60)	-1.1746	0.2944	(-1.75, -0.60)
β_1	0.3895	0.2097	(-0.02, 0.80)	0.4210	0.2018	(0.03, 0.82)
β_2	0.0300	0.2262	(-0.41, 0.47)	0.0326	0.2139	(-0.39, 0.45)
β_3	0.0537	0.2214	(-0.38, 0.49)	0.0588	0.2096	(-0.35, 0.47)
β_4	-0.0058	0.0074	(-0.02, 0.01)	-0.0063	0.0070	(-0.02, 0.01)
β_5	0.3152	0.1604	(0.00, 0.63)	0.3407	0.1522	(0.04, 0.64)
α	-0.5562	0.0021	(-0.56, -0.55)	-	-	-