

Discrete choice models with skewed link

Modelli a scelta discreta con funzione link asimmetrica

Rosa Capobianco

Dipartimento PE.ME.IS, Università degli Studi del Sannio
e-mail: capobianco@unisannio.it

Riassunto: In questo lavoro si utilizza la funzione di ripartizione della variabile casuale normale asimmetrica come funzione link in un modello a scelta discreta. Si dimostra che se il parametro di asimmetria λ è nullo le funzioni score sono linearmente dipendenti e pertanto la matrice di informazione è singolare. Per ovviare a questo problema si propone una riparametrizzazione del modello che consente di derivare gli stimatori di massima verosimiglianza anche nel caso in $\lambda = 0$ ed assicura l'esistenza delle proprietà asintotiche degli stimatori.

Keywords: skew normal, reparameterization, discrete choice model.

1. Introduction

In discrete choice models the outcome Y_i , for $i=1, \dots, n$, depends on a linear combination of explanatory variables $(x_1, \dots, x_k)$ through a link function F , i.e. $P(Y_i = 1) = 1 - F(-x_i' \beta)$, where $(\beta_1, \dots, \beta_k)'$ is a vector of unknown parameters. Generally F is the distribution function of a symmetric random variable (r.v.) such as the normal distribution (probit model) or the logistic r.v. (logit model). Symmetric links, however, do not always provide the best fit to the data when the probability of a given binary response approaches to zero or one at different rates. In order to overcome this inconvenient Aranda-Ordaz (1981) proposes a power transformation to model symmetric and asymmetric departures from the logistic model while Chen (2004) suggests to adopt flexible classes of distributions as link functions such that 1) the most common links can be considered as special or limiting cases, 2) the skewness is regulated by a parameter so that a test on this parameter can be carried out.

Flexible classes of distributions are characterized by a density function defined as $p(x) = 2f(x)G\{w(x)\}$, for $-\infty < x < \infty$ (Azzalini, 1985), where f is a density function symmetric around zero, G is a differentiable function which takes value in $(0,1)$ and is such that $G(-x) = 1 - G(x)$ and $w(-x) = w(x)$. When $w(x) = \lambda x$ and f and G are the density and the distribution function of the normal r.v., φ and Φ respectively, then $p(x) = 2\varphi(x)\Phi(\lambda x)$ is the density function of the skew normal (SN) r.v.. The distribution properties of $p(x)$ have been extensively studied in the statistical literature (Genton, 2004). It has been shown that the information matrix is always singular when $\lambda = 0$ if $G''(0) = g'(0) = 0$, where $G'(\cdot) = g(\cdot)$, so that a reparameterization is required (Pewsey, 2000).

In this paper we consider a discrete choice model where the link function is the SN distribution function. Also in this case the information matrix is singular when $\lambda = 0$ (Chiogna, 2005). In order to overcome this problem we propose a reparameterization, that resembles the one proposed by Azzalini (1985), which yields a non singular information

matrix and is such that Rotnitzky *et al.* (2000) results can be applied to derive the asymptotic properties of the maximum likelihood (ML) estimators.

2. The skew normal link

Let $(x_1, x_2, \dots, x_k)$ be a set of independent variables, with $x_1 = 1$, and let us suppose that the link function is the skew normal distribution function: $P(Y_i = 1) = 1 - \Phi(-x'_i\beta, \lambda) = 1 - 2 \int_{-\infty}^{-x'_i\beta} \int_{-\infty}^{\lambda t} \varphi(t)\varphi(u)du dt$ (Azzalini, 1985). By estimating $\theta = (\beta_1, \dots, \beta_k, \lambda)'$ through the maximum likelihood method we get the following score functions

$$S_{\beta_j} = 2 \sum_{i=1}^n x_{ij} \varphi(x'_i\beta) \Phi(-x'_i\beta\lambda) (y_i - \mu_y) / [(1 - \mu_y)\mu_y]$$

for $j=1, \dots, k$, with $\mu_y = 1 - \Phi(-x'_1\beta, \lambda)$, and

$$S_{\lambda} = (2/\pi)^{1/2} (1 + \lambda^2)^{-1} \sum_{i=1}^n \varphi(x'_i\beta\sqrt{1 + \lambda^2}) (y_i - \mu_y) / [(1 - \mu_y)\mu_y].$$

When $\lambda=0$ the derivatives $S_{\lambda}|_{\lambda=0}$ and $S_{\beta_1}|_{\lambda=0}$ are linearly dependent: $S_{\beta_1}|_{\lambda=0} = \sqrt{\frac{\pi}{2}} S_{\lambda}|_{\lambda=0}$. As a consequence, the information matrix is singular.

In order to have a non singular information matrix for the vector of estimators when $\lambda = 0$ we need to reparameterize the model. By following Azzalini's proposal (1985) and Chiogna (2005) we consider the following reparameterization: $x'_i\tilde{\beta} = x'_i\beta + \tilde{\lambda} \left[2/(\pi(1 + \tilde{\lambda}^2)) \right]^{1/2}$, where $\tilde{\beta}_1 = \beta_1 + \tilde{\lambda} \left[2/(\pi(1 + \tilde{\lambda}^2)) \right]^{1/2}$ while $\tilde{\lambda} = \lambda$ and $\tilde{\beta}_j = \beta_j$ for $j = 2, \dots, k$. After the reparameterization the score functions for $\tilde{\theta} = (\tilde{\beta}_1, \dots, \tilde{\beta}_k, \tilde{\lambda})'$ are equal to

$$S_{\tilde{\beta}_j} = 2 \sum_{i=1}^n x_{ij} \varphi(x'_i\tilde{\beta}) \Phi(-x'_i\tilde{\beta}\tilde{\lambda}) \left[\frac{y_i - \mu_y}{(1 - \mu_y)\mu_y} \right],$$

for $j=1, \dots, k$, with $\tilde{\mu}_y = 1 - \Phi(-x'_1\tilde{\beta}, \tilde{\lambda})$, and

$$S_{\tilde{\lambda}} = \sum_{i=1}^n \left[(2/\pi)^{1/2} (1 + \tilde{\lambda}^2)^{-1} \varphi(x'_i\tilde{\beta}\sqrt{1 + \tilde{\lambda}^2}) - 2M_1 \varphi(x'_i\tilde{\beta}) \right] \left[\frac{y_i - \tilde{\mu}_y}{(1 - \tilde{\mu}_y)\tilde{\mu}_y} \right],$$

with $M_1 = \sqrt{\frac{2}{\pi(1 + \tilde{\lambda}^2)}} - \tilde{\lambda}^2 \pi \sqrt{\frac{2}{\pi^3(1 + \tilde{\lambda}^2)^3}}$.

When $\tilde{\lambda} = 0$ $S_{\tilde{\lambda}}$ vanishes while $\frac{\partial}{\partial \tilde{\lambda}} S_{\tilde{\lambda}}$ neither is zero nor is a linear combination of $S_{\tilde{\beta}_1}, \dots, S_{\tilde{\beta}_k}$, the conditions (B1) and (B3) of theorem 4 by Rotnitzky *et al.* (2000, p. 261) are satisfied. Consequently we have the following results

a) the MLE of $\tilde{\theta} = (\tilde{\beta}_1, \dots, \tilde{\beta}_k, \tilde{\lambda})$ exists when $\tilde{\theta} = \tilde{\theta}_0$, where $\tilde{\theta}_0 = (\tilde{\beta}_1, \dots, \tilde{\beta}_k, \tilde{\lambda} = 0)$, it is unique with a probability tending to 1 and it is a consistent estimator of $\tilde{\theta}$ when $\tilde{\theta} = \tilde{\theta}_0$;

b) by setting $\hat{\theta} = (\hat{\beta}_b, \dots, \hat{\beta}_k, n^{-1/4}\hat{\lambda})'$, then

$$\sqrt{n}(\hat{\theta} - \tilde{\theta}_0) \longrightarrow \begin{bmatrix} Z_1 \\ \vdots \\ Z_k \\ (-1)^B Z_{k+1}^{1/2} \end{bmatrix} I_{(Z_{k+1} > 0)} + \begin{bmatrix} Z_1 - \frac{I^{(k+1)1}}{I^{(k+1)(k+1)}} Z_{k+1} \\ \vdots \\ Z_k - \frac{I^{(k+1)k}}{I^{(k+1)(k+1)}} Z_{k+1} \\ 0 \end{bmatrix} I_{(Z_{k+1} < 0)},$$

where $\tilde{I}$ denotes the covariance matrix of $(S_{\tilde{\beta}_1}, \dots, S_{\tilde{\beta}_k}, \frac{\partial}{\partial \tilde{\lambda}} S_{\tilde{\lambda}}/2)$, i.e.

$$\tilde{I} = \begin{bmatrix} \sum_{i=1}^n x_{i1}^2 \tilde{\psi}_i & \sum_{i=1}^n x_{i1} x_{i2} \tilde{\psi}_i & \dots & \sum_{i=1}^n x_{i1} x_{ik} \tilde{\psi}_i & - \sum_{i=1}^n x_{i1} \tilde{\gamma}_i \\ \sum_{i=1}^n x_{i1} x_{i2} \tilde{\psi}_i & \sum_{i=1}^n x_{i2}^2 \tilde{\psi}_i & \dots & \sum_{i=1}^n x_{i2} x_{ik} \tilde{\psi}_i & - \sum_{i=1}^n x_{i2} \tilde{\gamma}_i \\ \dots & \dots & \dots & \dots & \dots \\ \sum_{i=1}^n x_{i1} x_{ik} \tilde{\psi}_i & \sum_{i=1}^n x_{i2} x_{ik} \tilde{\psi}_i & \dots & \sum_{i=1}^n x_{ik}^2 \tilde{\psi}_i & - \sum_{i=1}^n x_{ik} \tilde{\gamma}_i \\ - \sum_{i=1}^n x_{i1} \tilde{\gamma}_i & - \sum_{i=1}^n x_{i2} \tilde{\gamma}_i & \dots & - \sum_{i=1}^n x_{ik} \tilde{\gamma}_i & \sum_{i=1}^n \tilde{\xi}_i \end{bmatrix}$$

$$\tilde{\psi}_i = \frac{4[\varphi(x'_i \tilde{\beta}) \Phi(-\tilde{\lambda} x'_i \tilde{\beta})]^2}{(1-\tilde{\mu}_y) \tilde{\mu}_y}; \tilde{\gamma}_i = \frac{\varphi(x'_i \tilde{\beta}) \Phi(-x'_i \tilde{\beta} \tilde{\lambda}) B(\tilde{\theta}, x_i)}{(1-\tilde{\mu}_y) \tilde{\mu}_y}, \tilde{\xi}_i = \frac{1}{4} \left\{ \frac{B^2(\tilde{\theta}, x_i)}{(1-\tilde{\mu}_y) \tilde{\mu}_y} + \frac{C^2(\tilde{\theta}, x_i) (2\tilde{\mu}_y - 1)^2}{(1-\tilde{\mu}_y)^3 \tilde{\mu}_y^3} \right\},$$

for $i = 1, \dots, n$; where

$$\begin{aligned} B(\tilde{\theta}, x_i) &= -\frac{\varphi(x'_i \tilde{\beta} \sqrt{1 + \tilde{\lambda}^2})}{\sqrt{2\pi}(1 + \tilde{\lambda}^2)^2} \left[(x'_i \tilde{\beta})^2 (1 + \tilde{\lambda}^2) + 2 \right] \\ &\quad + 2M_1 (x'_i \tilde{\beta}) \varphi(x'_i \tilde{\beta}) \varphi(\tilde{\lambda} x'_i \tilde{\beta}) - 2M_2 \varphi(x'_i \tilde{\beta}) \Phi(-\tilde{\lambda} x'_i \tilde{\beta}) \\ &\quad + 2M_1 \varphi(x'_i \tilde{\beta}) \left[(x'_i \tilde{\beta}) M_1 \Phi(-\tilde{\lambda} x'_i \tilde{\beta}) + (x'_i \tilde{\beta} - \tilde{\lambda} M_1) \varphi(\tilde{\lambda} x'_i \tilde{\beta}) \right]. \\ C(\tilde{\theta}, x_i) &= 4 \left\{ -\frac{1}{\sqrt{2\pi}} \frac{1}{1 + \tilde{\lambda}^2} \varphi(x'_i \tilde{\beta} \sqrt{1 + \tilde{\lambda}^2}) + M_1 \varphi(x'_i \tilde{\beta}) \Phi(-\tilde{\lambda} x'_i \tilde{\beta}) \right\}^2. \\ M_2 &= -3\pi \tilde{\lambda} \sqrt{\frac{2}{\pi^3 (1 + \tilde{\lambda}^2)^3}} + 3\tilde{\lambda}^3 \pi^2 \sqrt{\frac{2}{\pi^5 (1 + \tilde{\lambda}^2)^5}}. \end{aligned}$$

$\tilde{I}^{ij}$, for $j, l = 1, \dots, k+1$, denotes the jl -th entry of $\tilde{I}^{-1}$, $Z = (Z_1, \dots, Z_{k+1})$ denotes a mean-zero normal r.v. with variance equal to $\tilde{I}^{-1}$ and B denotes a Bernoulli variable with parameter 1/2 that is independent of Z ; $I_{(\omega)}$ denotes the indicator function;

c) the likelihood ratio statistics for testing the simple null hypothesis $H_0 : \tilde{\theta} = \tilde{\theta}_0$ converges in distribution to a mixture of a χ_{k+1}^2 and a χ_k^2 r.v. with mixing probability equal to 1/2, i.e.

$$2 \left\{ L_n(\hat{\theta}) - L_n(\theta_0) \right\} \longrightarrow \sum_{i=1}^{k+1} Z_i^{*2} I_{(Z_{k+1}^* > 0)} + \sum_{i=1}^k Z_i^{*2} I_{(Z_{k+1}^* < 0)}$$

where Z_i^* , for $i = 1, \dots, k+1$, are independent $N(0, 1)$ r.v..

In order to investigate the finite-sample properties of the estimators when $\lambda = 0$, a simulation study has been conducted. It has been considered a model with an explicative variable, whose values have been generated from a Uniform r.v. defined on $[-1,1]$, and parameters $\beta_1 = 0.5$ and $\beta_2 = 2$. The outcome of the model, y_i , has been put equal to 1 when $(\beta_1 + \beta_2 x_i + \epsilon_i) > 0$, ϵ_i is $N(0,1)$ distributed. 10000 samples of sizes 250, 500 and 1000 have been generated. The parameters have been estimated by considering the proposed reparameterization. The biases and root mean squared errors (RMSEs) of θ estimator are presented in table 1 where the last column reports the number of samples for which the ML estimate of λ occurs on the boundary. In order to handle this problem we followed the procedure implemented by DiCiccio and Monti (2004): $\hat{\lambda}$ was taken to be the smallest value λ_0 such that $H_0 : \lambda = \lambda_0$ is not rejected at the 5% level by a test that assumes that the likelihood ratio statistic has the χ_1^2 distribution. The slow convergence of $\hat{\lambda}$ to λ can be explained by the adopted reparameterization. $\hat{\beta}_2$ is generally reliable, while $\hat{\beta}_1$ depends on $\hat{\lambda}$ so that a large sample size is required for the estimate.

Table 1: Biases and RMSEs of $\tilde{\theta}$ MLEs: $\lambda = 0$, $\beta_1 = 0.5$, $\beta_2 = 2$.

n	$\hat{\lambda}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\%(\hat{\lambda} > 10)$
250	-1.04 (1.97)	-0.37 (0.58)	-0.29 (0.50)	729
500	-0.91 (1.54)	-0.39 (0.56)	-0.26 (0.43)	90
1000	-0.65 (1.01)	-0.05 (0.09)	-0.21 (0.35)	0

References

- Aranda-Ordaz F.J. (1985) On two families of transformations to additivity for binary response data *Biometrika*, 68, 357–363.
- Azzalini A. (1985) A class of distributions that includes the normal ones *Scandinavian Journal of Statistics*, 12, 171–178.
- Chen M. (2004) Skewed link models for categorical response data, in: *Skew-elliptical distributions and their applications. A journey beyond normality*, Genton M.G. (Eds.), Chapman and Hall, 131–152.
- Chiogna A. (2005) A note on the asymptotic distribution of the maximum likelihood estimator for the scalar skew-normal distribution *Statistical Methods & Applications*, 14, 331–341.
- DiCiccio T.J., Monti A.C. (2004) Inferential aspects of the skew exponential power distribution, *Journal of the American Statistical Association*, 99, 439–450.
- Genton M.G. (2004) *Skew-elliptical distributions and their applications. A journey beyond normality*, Chapman and Hall, London.
- Pewsey. A. (2000) Problems of inference for Azzalini’s skew-normal distribution, *Journal of Applied Statistics*, 27, 859–870.
- Rotnitzky A., Cox D.R., Bottai M., Robins J. (2000) Likelihood-based inference with singular information matrix, *Bernoulli*, 6, 243–284.