Tobit regression with Lasso penalty

Ahmed Alhamzawi
Department of Mathematics, College of Science, University of Al-Qadisiyah, Iraq
E-mail: ahmed.alhamzawi@qu.edu.iq

Abstract. A new Bayesian lasso tobit regression method is proposed. The new method is presented as a scale mixture of truncated normal distribution (SMTND) of Laplace distribution with exponential mixing density. The equality $e^{\beta} = e^{\alpha |\beta|} e^{(1-\alpha)|\beta|}$, is used to write the familiar laplace density as the product of two different Gibbs samplers. Simulation studies are used to illustrate the new algorithm.

1. Introduction

Tobit regression has become one of the most commonly used statistical tools utilized by researchers to describe the relationship between a non-negative response variable and a set of covariates. Since its introduction by [1], it has been routinely applied in medicine, biology, ecology, economics, and social sciences [2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14]. This model can be viewed as a linear regression model with a latent continuous response $y^*$. Consider the standard tobit regression model

$$y_i = \max\{0, y^*_i\}, \quad i = 1, \ldots, n,$$

$$y^*_i = x'_i \beta + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2),$$

where $x_i = (x_{i1}, \ldots, x_{ik})'$, $\beta = (\beta_1, \ldots, \beta_k)'$ and $\varepsilon_i$'s are residuals. The unknown vector $\beta$ can be estimated by using the maximum likelihood method [1]. Although the asymptotic theory for maximum likelihood has been well studied in tobit models [15, 16], a Bayesian method produces exact inference even if $n$ is small. [17] proposed a Bayesian approach for tobit regression model using the normal density for the residuals and generating $\beta$ from its full conditional posterior distribution using Gibbs sampler (GS).

Under the tobit model (6), it is postulated that only an unknown subset of predictors are important in the model, so that the problem of covariate selection is to select this active covariates. Different approaches have been proposed over the recent years for dealing with covariate selection in linear regression models. Among these, the least absolute shrinkage and selection operator (lasso) method [18] has received considerable attention over the years due to its susceptibility to overfitting problems. For the linear model, the lasso is obtained by minimizing

$$\min_{\beta} \sum_{i=1}^{n}(y_i - x'_i \beta)^2 + \lambda \sum_{j=1}^{k}|\beta_j|,$$

where

$$e^{\beta} = e^{\alpha |\beta|} e^{(1-\alpha)|\beta|},$$

is used to write the familiar laplace density as the product of two different Gibbs samplers. Simulation studies are used to illustrate the new algorithm.
where $\lambda \geq 0$. Many fast algorithms have been reported over the years for the computation of the lasso estimators. However, one of the main drawbacks of these algorithms is that they cannot provide valid standard errors for the zero estimated coefficients [19, 20]. Bayesian framework overcome this drawback by producing a valid measure of standard error [19, 20]. Within this framework, [18] observed that the solution of the minimization problem (2) can be presented as a Bayesian posterior mode estimate when specifying Laplace distributions on the regression parameters. Following this observation, [21] proposed a Bayesian framework for the Bayesian Lasso regression and [22] proposed a Bayesian frame work for Lasso mixed quantile regression. Since [21], different Bayesian Lasso methods have proposed over the recent years. Se fore example [23, 20, 24, 25, 26]. In this paper, we propose a new Gibbs sampler algorithm for Bayesian Lasso in tobit regression. This GS is constructed based on the approach of [27]. Specifically, along the same line as [27], we present a new hierarchical model for the Bayesian lasso method using a SMTND representation of the Laplace prior.

2. Bayesian Lasso tobit regression
Consider the standard Lasso tobit regression model

$$
\min_{\beta} \sum_{i=1}^{n} (y_i - \max\{x_i'\beta, 0\})^2 + \lambda \sum_{j=1}^{k} |\beta_j|,
$$

Rather than minimizing the above regularization problem for tobit model, we solve it by constructing a Bayesian framework.

2.1. Laplace representation
The Laplace distribution can be represented as a SMTND with exponential mixing distributions, i.e.

$$
\frac{\lambda}{a\sqrt{\sigma^2}} \exp\left\{-\frac{\lambda|\beta_j|}{\sqrt{\sigma^2}}\right\} = \frac{\lambda}{a\sqrt{\sigma^2}} \exp\left\{-\frac{\lambda(\alpha|\beta_j| + (1-\alpha)|\beta_j|)}{\sqrt{\sigma^2}}\right\}
$$

$$
= \int_{0}^{\infty} \int_{u_j > \sqrt{\frac{4\lambda^2(1-\alpha)^2}{\sigma^2}} |\beta_j|} \frac{1}{2\sigma^2\sigma^2 s_j} \exp\left\{-\frac{2\alpha^2 \sigma^2 |\beta_j|^2}{\sigma^2 s_j}\right\} \exp\left\{-\frac{u_j}{2}\right\} \frac{\lambda^2}{8} \exp\left\{-\frac{\lambda^2 s_j}{8}\right\} \ du_j d s_j
$$

$$
\propto \int_{0}^{\infty} \int_{u_j > \sqrt{\frac{4\lambda^2(1-\alpha)^2}{\sigma^2}} |\beta_j|} \ N(\beta_j; 0, \sigma^2 s_j/4\alpha^2) \text{Exponential}(u_j; \frac{1}{2}) \text{Exponential}(s_j; \frac{\lambda^2}{8}),
$$

where $\alpha \in [0, 1]$ that controls on the boundaries of the regression coefficients. Almost all the existing representations of the Bayesian Lasso can be considered as a special form of the above representation. For example, when $\alpha = 0$, it is the Baysian Lasso of [20] induced by represented the Laplace distribution as a scale mixture of uniforms (with a Gamma mixing density). When $\alpha = \frac{1}{2}$, it is the Baysian Lasso of [27] induced by represented the Laplace distribution as a SMTND (with exponential mixing densities). When $\alpha = 1$, it is the Bayesian Lasso of [21] induced by represented the Laplace distribution as a scale mixture of normals (with an exponential mixing density). To proceed a Bayesian analysis, we assign a Gamma prior for $\lambda^2$. We also assume that $\sigma^2 \sim 1/\sigma^2$.

2.2. Hierarchical representation
Here, we can formulate the hierarchical representation of our model as

$$
y_i = \max\{0, y_i^*\}, \quad i = 1, \cdots, n,
$$

$$
y^* | y, X, \beta, \sigma^2 \sim N_n(X\beta, \sigma^2 I_n),
$$
Thus, we have

\[
\beta|\sigma^2, \lambda^2, u, s \sim \prod_{j=1}^{k} N(\beta_j; 0, \sigma^2 s_j/(4\alpha^2)) I\left\{|\beta_j| < 2(1-\alpha)\sqrt{\frac{\sigma^2}{\lambda^2} u_j}\right\},
\]

\[
s|\lambda^2 \sim \prod_{j=1}^{k} \text{Exponential}\left(\frac{\lambda^2}{8}\right),
\]

\[
u \sim \prod_{j=1}^{k} \text{Exponential}\left(\frac{1}{2}\right),
\]

\[
\lambda^2 \sim (\lambda^2)^{a-1} \exp(-b\lambda^2),
\]

\[
\sigma^2 \sim 1/\alpha^2,
\]

where \(y^* = (y_1^*, \ldots, y_n^*)\), \(X = (x_1, \ldots, x_n)'\), \(u = (u_1, \ldots, u_k)\) and \(s = (s_1, \ldots, s_k)\).

3. Computation

Based on the hierarchical representation (4), the fully conditional posterior of \(\beta\) is as follows:

\[
\pi(\beta|y, y^*, X, \sigma^2, \lambda^2, u, s) \propto \pi(y|y^*, X, \beta, \sigma^2)\pi(\beta|\sigma^2, \lambda^2, u, s)
\]

\[
\propto \exp\left\{-\frac{1}{2\sigma^2}[(y^*-X\beta)'(y^*-X\beta) + \sum_{j=1}^{k} \frac{4\alpha^2\beta_j^2}{s_j}]\right\} \prod_{j=1}^{k} I\{|\beta_j| < 2(1-\alpha)\sqrt{\frac{\sigma^2}{\lambda^2} u_j}\}.
\]

Thus, we have

\[
\beta|y, y^*, X, \sigma^2, \lambda^2, u, s \sim N_k(A^{-1}X'y^*, \sigma^2 A^{-1}) \prod_{j=1}^{k} I\{|\beta_j| < 2(1-\alpha)\sqrt{\frac{\sigma^2}{\lambda^2} u_j}\},
\]

where, \(A = X'X + 4\alpha^2 S^{-1}\) and \(S = \text{diag}(s_1, \ldots, s_k)\).

The full conditional distribution of \(\sigma^2\) is as follows

\[
\pi(\sigma^2|y^*, X, \beta, \lambda^2, u, s) \propto \pi(y^*|X, \beta, \sigma^2)\pi(\beta|\lambda^2, \sigma^2, u, s)\pi(\sigma^2),
\]

\[
\propto \left(\frac{1}{\sigma^2}\right)^{\frac{n-k+1}{2}} \exp\left\{-\frac{1}{2\sigma^2}[(y^*-X\beta)'(y^*-X\beta) + \sum_{j=1}^{k} \frac{4\alpha^2\beta_j^2}{s_j}]\right\} I\{|\sigma^2 > \text{Max}_j \left(\frac{\lambda^2\beta_j^2}{4u_j^2(1-\alpha)^2}\right)\}.
\]

Thus, we have

\[
\sigma^2|y, y^*, X, \sigma^2, \lambda^2, u, s \sim
\]

\[
\text{IG}\left(\frac{n-1+k}{2} + 1, \frac{1}{2}[(y^*-X\beta)'(y^*-X\beta) + \sum_{j=1}^{k} \frac{4\alpha^2\beta_j^2}{s_j}]\right) I\left\{|\sigma^2 > \text{Max}_j \left(\frac{\lambda^2\beta_j^2}{4u_j^2(1-\alpha)^2}\right)\}\right\},
\]
The full conditional distribution of $s_j$ is as follows
\[
\pi(s_j|\beta_j, \sigma^2, \lambda^2) \propto \pi(\beta_j|\sigma^2, s_j)\pi(s_j|\lambda^2)
\]
\[
\propto \frac{1}{\sqrt{2\pi\sigma^2 s_j}} \exp\left(-\frac{2\alpha^2\beta^2}{\sigma^2 s_j}\right) \exp\left(-\frac{\lambda^2}{8}s_j\right)
\]
\[
\propto \frac{1}{\sqrt{s_j}} \exp\left\{-\frac{1}{2}\left[\frac{\lambda^2}{4}s_j + \frac{\beta_j^2}{\sigma^2 s_j} - 1\right]\right\}
\]

The full conditional distribution of $u_j$ is as follows
\[
\pi(u_j|y, y^*, X, \sigma^2, \lambda^2, s) \propto \pi(\beta|\sigma^2, \lambda^2, u)\pi(u)
\]
\[
\propto \prod \text{Exponential}\left(\frac{1}{2}\right) I\left\{u_j > \sqrt{\frac{\lambda^2}{\sigma^2} \frac{|\beta_j|}{2(1 - \alpha)}}\right\}
\]

The full conditional distribution of $\lambda^2$ is as follows
\[
\pi(\lambda^2|y, y^*, X, u, s, \beta, \sigma^2) \propto \pi(\lambda^2, \sigma^2, \lambda^2, u)\pi(s|\lambda^2)\pi(\lambda^2)
\]
\[
\propto \text{Gamma}\left(\alpha + k, b + \frac{1}{8}\sum_{j=1}^{k} s_j\right) I\left\{\lambda^2 < \text{Max}_j\left(\frac{4(1 - \alpha)^2\sigma^2 u^2_j}{\beta_j^2}\right)\right\}
\]

The full conditional distribution of $y_i^*$ is given by:
\[
y_i^*|y_i, x_i, \beta, u, s, \lambda^2, \sigma^2 \sim \begin{cases} 
\delta(y_i), & \text{if } y_i > y_0, \\
N(x_i^T\beta, \sigma^2)I(y_i^* \leq y_0), & \text{otherwise},
\end{cases}
\]
where $\delta(y_i)$ denotes a degenerate distribution with all its mass at $y_i$.

4. Simulation studies
Here, we compare the prediction accuracy of our method (referred to as “NBTLasso”) with Bayesian Lasso tobit regression (referred to as “OBTLasso”) and Bayesian elastic net regression (referred to as “BTEnet”). The prediction accuracy measured by MMAD on the test set, where \( \text{MMAD} = \text{median}(|X_{test}\hat{\beta} - X_{test}\hat{\beta}_{true}|) \). Model selection performance is evaluated based on the credible intervals for the approaches in the comparison.

4.1. Simulation 1
The data are generated from the following model
\[
y_i = \max\{0, y_i^*\}, \quad i = 1, \ldots, n,
\]
\[
y_i^* = x_i^T\beta + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2),
\]

We set $k = 11$ covariates, $\sigma^2 = 9$ and the true regression coefficients vector, including the intercept term, are $\beta = (0, 4.5, 2.5, 0, 0, 0, 0, 0, 0, 0, 0)$. We generate 11 covariates from a multivariate normal distribution $N(0, \Sigma)$, where $\Sigma = I_p$ (Isotropic design). The results are lists in Table 1. We generate a training set with 50 observations and a testing set with 250 observations. The results show that NBTLasso perform very well compare to OBTLasso and BTEnet. The proposed method produces the smallest MMAD compared to the other methods and perform very well in variable selection. It has smallest FPR and FEN. The convergence of our algorithm is shown in Figures 1 and 2.
Table 1. FPR and FNR are the false positive and negative rates for Simulation 1, respectively.

| Method     | MMAD    | FPR   | FNR   |
|------------|---------|-------|-------|
| NBTLasso   | 1.382 (0.337) | 0.180 (0.382) | 0.000 (0.000) |
| OBTLasso   | 1.789 (0.497) | 0.241 (0.519) | 0.013 (0.001) |
| BTEnet     | 1.963 (0.639) | 0.223 (0.526) | 0.082 (0.001) |

Figure 1. Histograms for Simulation 1.

4.2. Simulation 2
This simulation is the same as Simulation 1, except that we set $\beta = (0, 1, 1, 0, 0, 0, 0, 0, 0)$. The results are summarized in Table 2. Again, the results show that NBTLasso performs very well compared to OBTLasso and BTEnet. The proposed method produces the smallest MMAD compared to the other methods and performs very well in variable selection. It has smallest FPR and FNR. The convergence of our algorithm is shown in Figures 3 and 4.

Table 2. FPR and FNR are the false positive and negative rates for Simulation 2, respectively.

| Method     | MMAD    | FPR   | FNR   |
|------------|---------|-------|-------|
| NBTLasso   | 0.826 (0.042) | 0.991 (0.101) | 0.051 (0.103) |
| OBTLasso   | 1.789 (0.497) | 1.332 (0.201) | 0.562 (0.284) |
| BTEnet     | 1.963 (0.639) | 1.257 (0.119) | 0.457 (0.391) |
5. Conclusions

We have introduced a new hierarchical formulation of Bayesian lasso tobit regression using a SMTND with exponential mixing densities. The simulation results show that the proposed method perform very well compare to other methods.

References

[1] James Tobin. Estimation of relationships for limited dependent variables. Econometrica: journal of the Econometric Society, pages 24–36, 1958.
[2] John F McDonald and Robert A Moffitt. The uses of tobit analysis. The review of economics and statistics, pages 318–321, 1980.
[3] Lung-Fei Lee. Generalized econometric models with selectivity. Econometrica: Journal of the Econometric Society, pages 507–512, 1983.
[4] Takeshi Amemiya. Tobit models: A survey. Journal of econometrics, 24(1-2):3–61, 1984.
[5] Angus Deaton and Margaret Irish. Statistical models for zero expenditures in household budgets. Journal of Public Economics, 23(1-2):59–80, 1984.
[6] Hussein Hashem, Veronica Vinciotti, Rahim Alhamzawi, and Keming Yu. Quantile regression with group lasso for classification. Advances in Data Analysis and Classification, 10(3):375–390, 2016.
[7] Rahim Alhamzawi. Model selection in quantile regression models. Journal of Applied Statistics, 42(2):445–458, 2015.
[8] Daniel L McFadden. Econometric analysis of qualitative response models. Handbook of econometrics, 2:1395–1457, 1984.
[9] Whitney K Newey. Efficient estimation of limited dependent variable models with endogenous explanatory variables. Journal of Econometrics, 36(3):231–250, 1987.
[10] Gangadharrao S Madlala. Limited dependent variable models using panel data. Journal of Human resources, pages 307–338, 1987.
[11] Mary Kathryn Cowles, Bradley P Carlin, and John E Connett. Bayesian tobit modeling of longitudinal ordinal clinical trial compliance data with nonignorable missingness. Journal of the American Statistical Association, 91(433):86–98, 1996.
Figure 3. Histograms for Simulation 2.

[12] Antonie Vonk Noordegraaf, Mirjam Nielen, and Jack PC Kleijnen. Sensitivity analysis by experimental design and metamodelling: Case study on simulation in national animal disease control. *European Journal of Operational Research*, 146(3):433–443, 2003.

[13] Jeffrey M Wooldridge. Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics*, 20(1):39–54, 2005.

[14] Eleanor M Pullenayegum, Jean-Eric Tarride, Feng Xie, Ron Goeree, Hertzel C Gerstein, and Daria O’Reilly. Analysis of health utility data when some subjects attain the upper bound of 1: are tobit and clad models appropriate? *Value in Health*, 13(4):487–494, 2010.

[15] James L Powell. Symmetrically trimmed least squares estimation for tobit models. *Econometrica: Journal of the Econometric Society*, pages 1435–1460, 1986.

[16] Benedikt M Pötscher and Ingmar R Prucha. Dynamic nonlinear econometric models: Asymptotic theory. Springer Science & Business Media, 2013.

[17] Siddhartha Chib. Bayes inference in the tobit censored regression model. *Journal of Econometrics*, 51(1-2):79–99, 1992.

[18] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.

[19] Minjung Kyung, Jeff Gill, Malay Ghosh, George Casella, et al. Penalized regression, standard errors, and bayesian lassos. *Bayesian Analysis*, 5(2):369–411, 2010.

[20] Himel Mallick and Nengjun Yi. A new bayesian lasso. *Statistics and Its Interface*, 7(4):487–494, 2014.

[21] Trevor Park and George Casella. The bayesian lasso. *Journal of the American Statistical Association*, 103(482):681–686, 2008.

[22] Rahim Alhamzawi and Keming Yu. Bayesian lasso-mixed quantile regression. *Journal of Statistical Computation and Simulation*, 84(4):686–880, 2014.

[23] Chris Hans. Bayesian lasso regression. *Biometrika*, 96(4):835–845, 2009.

[24] Rahim Alhamzawi, Ahmed Alhamzawi, and Haithem Tah Mohammad Ali. New gibbs sampling methods for bayesian regularized quantile regression. *Computers in biology and medicine*, 110:52–65, 2019.

[25] Rahim Alhamzawi and Haithem Tah Mohammad Ali. The bayesian adaptive lasso regression. *Mathematical biosciences*, 303:75–82, 2018.

[26] Rahim Alhamzawi and Haithem Tah Mohammad Ali. Bayesian tobit quantile regression with penalty.
Figure 4. Trace plots for Simulation 2.

Communications in Statistics-Simulation and Computation, 47(6):1739–1750, 2018.

[27] Rahim Alhamzawi and Haithem Taha Mohammad Ali. A new gibbs sampler for bayesian lasso. Communications in Statistics-Simulation and Computation, pages 1–17, 2018.