AN EVALUATION OF ESTIMATORS FOR CENSORED SYSTEMS OF EQUATIONS USING MONTE CARLO SIMULATION

Yunfei Zhao*, Thomas L. Marsh and Huixin Li

School of Economic Sciences
Washington State University
PO Box 646210
Pullman, WA 99164 – 6210

*Corresponding author. Email: yunfei.zhao@email.wsu.edu

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association's 2012 AAEA Annual Meeting, Seattle, Washington, August 12-14, 2012.

An Evaluation of Estimators for Censored Systems of Equations Using Monte Carlo Simulation

Yunfei Zhao, Thomas L. Marsh and Huixin Li

School of Economic Sciences, Washington State University

Introduction

Micro-level data are more accessible, and its expanded use gives rise to more applications of estimators for censored equations. Provided that underlying data generation process deviates from assumption, how sensitive and robust is an estimator to estimate population parameters in censored equations?

This study empirically examines performance of proposed estimators for censored equations:

- ❖ Conduct a Monte Carlo experiment for censored equations under specific sample selection rules.
- ❖ examine performance of estimators for ordinary censoring problem.
- ❖ examine performance of estimators for sample selection problem.
- ❖ examine the impact of experiment design on estimator performance.

Monte Carlo Experiment Implementation

The data generation process is a bivariate limited dependent model with a double sample selection rule, and it is similar to previous studies (Flood and Grasjo 2001; Shonkwiler and Yen 1999; Tauchmann 2005).

Latent structural equations:

$y^*_{1t} = \beta_{01} + \beta_{11} z_{11t} + \beta_{21} x_{21t} + \varepsilon_{1t}$ (1.1)

$y^*_{2t} = \beta_{02} + \beta_{12} z_{12t} + \beta_{22} x_{21t} + \varepsilon_{2t}$ (1.2)

Latent index equations:

$d^*_{1t} = \alpha_{01} + \alpha_{11} z_{11t} + \alpha_{21} z_{21t} + v_{1t}$ (1.3)

$d^*_{2t} = \alpha_{02} + \alpha_{12} z_{12t} + \alpha_{22} z_{21t} + v_{2t}$ (1.4)

Observed index equations:

$d_{1t} = \begin{cases} 1\; if d^*_{1t} > 0 \\ 0\; if d^*_{1t} \le 0 \end{cases}$ (2.1); $d_{2t} = \begin{cases} 1\; if d^*_{2t} > 0 \\ 0\; if d^*_{2t} \le 0 \end{cases}$ (2.2)

Observed structural equations:

$y_{1t} = \begin{cases} y^*_{1t}\; if d_{1t} = 1\; and y^*_{1t} > 0 \\ 0\; else \end{cases}$ (2.3); $y_{2t} = \begin{cases} y^*_{2t}\; if d_{2t} = 1\; and y^*_{2t} > 0 \\ 0\; else \end{cases}$ (2.4)

The distributions of independent variables and error vector are given as:

$z_{11}, z_{21}, z_{12}, x_{21} \sim IIDN(0,1)$

$$\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ v_1 \\ v_2 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1.50 & -1.00 & 0.75 & 0.00 \\ -1.00 & 2.00 & 0.00 & 0.75 \\ 0.75 & 0.00 & 1.00 & 0.00 \\ 0.00 & 0.75 & 0.00 & 1.00 \end{bmatrix} \right)$$

This experiment has some features worth our attention:

- ❖ it is a generalized double hurdle model where individuals must pass two separate hurdles before a positive level of measurement is observed.
- ❖ an ordinary censoring problem occurs when relations embedded in equations (1.3), (1.4), (2.1) and (2.2) are not discovered.
- ❖ a sample selection problem occurs when the role of Tobit censoring in equations (2.3) and (2.4) is not discovered.
- ❖ it is a special case for a mixed censoring problem where some observations are censored as an ordinary censoring problem, some observations are censored as a sample selection problem, and some observations are censored as a double hurdle model.

Estimators and Empirical Results

In order to correct for censoring problem within a multiple equation framework, a number of estimation procedures have been developed.

Following estimators are examined in an ordinary censoring problem:

- ❖ Method of Simulated Scores (QLIM) (Hajivassiliou and McFadden 1998)
- ❖ Bayesian Estimation (BAES) (Huang 2001)
- ❖ Expectation Maximization (EXMA) (Zhao 2011)

Following estimators are examined in a sample selection problem:

- ❖ Multivariate Heckman Two-Step Method (MHTS) (Tauchmann 2005)
- ❖ Shonkwiler - Yen Two-Step Method (SHTS) (Shonkwiler and Yen 1999)

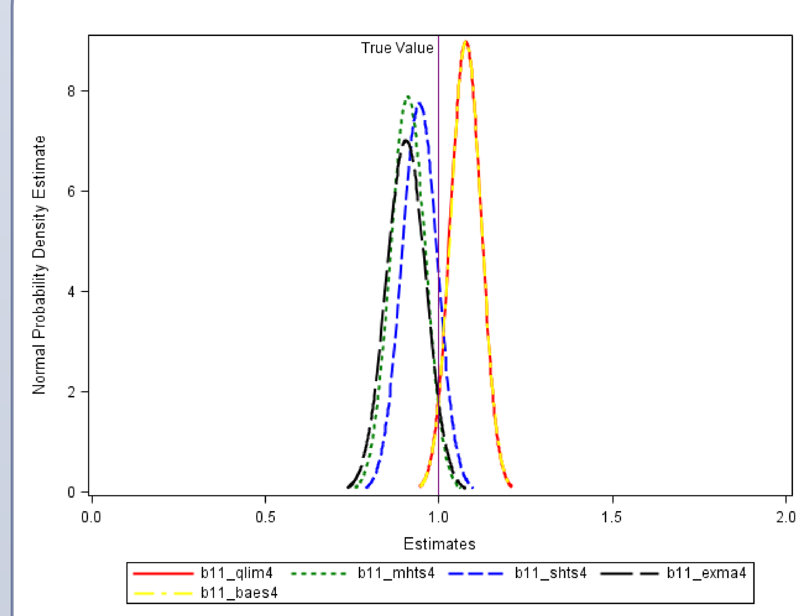
Comparison of Parameter Estimates

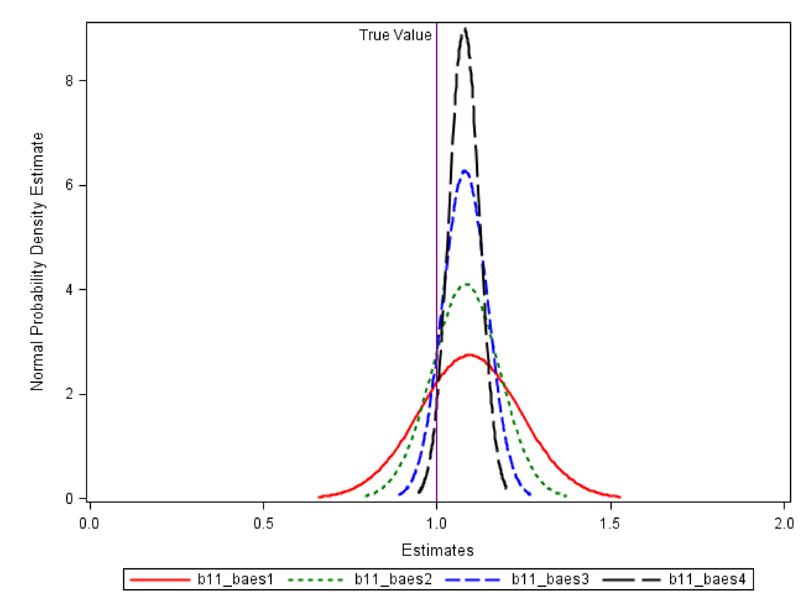
Estimator	Compare	β_{01}	β_{11}	β_{21}	β_{02}	β_{12}	β_{22}
	True	2.50	1.00	1.00	1.50	1.00	1.00
qlim	Estimate	2.36	1.08	0.95	1.28	1.10	0.95
	Variance	0.00	0.00	0.00	0.00	0.00	0.00
	Bias	-0.14	0.08	-0.05	-0.22	0.10	-0.05
	MSE	0.02	0.01	0.00	0.05	0.01	0.01
mhts	Estimate	2.64	0.91	0.90	1.88	0.80	0.78
	Variance	0.00	0.00	0.00	0.01	0.00	0.00
	Bias	0.14	-0.09	-0.10	0.38	-0.20	-0.22
	MSE	0.02	0.01	0.01	0.15	0.04	0.05
shts	Estimate	2.56	0.94	0.93	1.72	0.85	0.83
	Variance	0.00	0.00	0.00	0.01	0.00	0.00
	Bias	0.06	-0.06	-0.07	0.22	-0.15	-0.17
	MSE	0.01	0.01	0.01	0.05	0.03	0.03
exma	Estimate	2.57	0.90	0.81	1.88	0.76	0.64
	Variance	0.00	0.00	0.00	0.00	0.00	0.01
	Bias	0.07	-0.10	-0.19	0.38	-0.24	-0.36
	MSE	0.01	0.01	0.04	0.15	0.06	0.13
baes	Estimate	2.36	1.08	0.95	1.28	1.10	0.95
	Variance	0.00	0.00	0.00	0.00	0.00	0.00
	Bias	-0.14	0.08	-0.05	-0.22	0.10	-0.05
	MSE	0.02	0.01	0.00	0.05	0.01	0.01

Comparison of Single Statistics

Estimator	R-square for y^*_{1t}	R-square for y^*_{2t}	MSPE for y^*_{1t}	MSPE for y^*_{2t}
qlim	0.56	0.48	1.46	1.90
mhts	0.56	0.44	1.50	2.17
shts	0.57	0.48	1.47	2.01
exma	0.56	0.43	1.50	2.30
baes	0.56	0.48	1.46	1.90

Note: Sample Size 1,000, replicated samples 500, censoring proportions: 0.05 (d^*_{1t}), 0.10(d^*_{2t}), 0.10(y^*_{1t}) and 0.25(y^*_{2t}) .





Note: Estimated Normal Distribution of $\hat{\beta}_{11}$. Following abbreviations are applicable regarding to sample size: b11_qlim1, b11_qlim2, b11_qlim3, and b11_qlim4 respectively have sample sizes 100, 200, 500, and 1,000.

Discussions and Conclusions

All estimates are highly significant, suggesting that all estimators successfully detect the impact of each independent variable, and corresponding signs of parameters.

The variance of the estimator becomes smaller as the sample size grows, indicating that an estimator becomes more stable.

The results indicate that all estimators converged to points that are reasonably close to estimated parameters.

Smaller variance, bias, and mean square error is found in estimators for parameters in y^*_{1t} which has fewer censored observations.

Larger bias is found in estimates for β_{11} and β_{12} which are respectively associated with z_{11t} and z_{12t} appearing in both structural equations and index equations.

QLIM and BAES perform better in terms of variance and mean square error.

QLIM, BAES and SHTS perform better in terms of within-sample R-square and out-of-sample mean square prediction error.

There is generally a substantial difference in the performance of estimators.

References

Hajivassiliou, V. A. and D. L. McFadden (1998). "The Method of Simulated Scores for the Estimation of LDV Models." Econometrica, Vol. 66, No. 4 (Jul.), pp. 863-896.

Huang, H.-C. R. (2001). "Bayesian Analysis of the SUR Tobit Model." Applied Economics Letters, 2001, 8, 617-622.

Zhao, Y. (2011). "Empirical Analysis of the Fresh Tree Fruit Market." Doctor Dissertation, Washington State University.

Tauchmann, H. (2005). "Efficiency of Two-Step Estimators for Censored Systems of Equations: Shonkwiler and Yen Reconsidered." Applied Economics 37(4): 367-374.

Shonkwiler, J. S. and S. T. Yen (1999). "Two-Step Estimation of a Censored System of Equations." American Journal of Agricultural Economics 81(4): 972-982.

Flood, L. and U. Grasjo (2001). "A Monte Carlo Simulation Study of Tobit Models." Applied Economics Letters, 8, 581-584.