Two-tier tobit panel data model

Ruei-Chi Lee

April 24, 2021

1 Introduction

In economics, we often encounter censored data that is truncated above zero such as labor force participation hours. Tobit models are suitable for this type data. Cragg (1971) proposed a flexible tobit model that separates outcomes into two steps. First one is to decide whether participate the market or not. Second stage is to decide how many quantities they want to buy. This model structure is allowed to handle in situations that separate participation and consumption decision.

2 The Two-Tier Tobit Panel Model

The major drawback of tobit model is that the choice of $y > 0$ and the value of y, given that $y > 0$, is determined by the same vector of parameters. As an alternative, Cragg proposed a model that is able to integrate the probit model to determine the probability of $y > 0$ and the truncated normal model for given positive values of y.

$$Participation \quad D_i^* = Z_i\alpha + \epsilon_i, \quad \epsilon_i \sim N(0,1)$$
$$D_i = 1 \quad if D_i^* > 0 \quad or \quad D_i = 0 \quad if D_i^* < 0$$

$$Consumption : Y_{it}^* = X_{it}\beta + v_{it} + u_i, \quad v_{it} \sim N(0,\sigma^2), u_i \sim N(0,\sigma_u^2)$$
$$Y_{it} = Y_{it}^* \quad if D_i = 1 \quad and \quad Y_{it}^* > 0, \quad or$$
$$Y_{it} = 0 \quad if D_i = 0 \quad or \quad Y_{it}^* \leq 0$$

$$y_{it}^* = \beta' x_{it} + \alpha_i + u_{it}, \quad u \sim N(0,\sigma_u^2), \quad \alpha \sim N(0,\sigma_\alpha^2)$$
$$y_{it} = max\{y_{it}^*, 0\}$$

In this case, we assume that v and u are independent. The likelihood for each individual for T observations, marginalized with respect to the random effect u_i is

$$(l_i|D_i = 1, u_i) = \prod_{t=1}^{T} \{1 - \Phi(\frac{X_{it}'\beta + u_i}{\sigma})\}^{I(y_{it}=0)} \{\frac{1}{\sigma}\phi(\frac{y_{it} - X_{it}'\beta - u_i}{\sigma})\}^{I(y_{it}>0)}$$

$$(l_i|D_i = 0) = 0 \quad if \sum_{t=1}^{T} y_{it} > 0 \quad or (l_i|D_i = 0) = 1 \quad if \sum_{t=1}^{T} y_{it} = 0$$

$$(l_i|u_i) = \Phi(Z_i'\alpha)(l_i|D_i = 1, u_i) + \{1 - \Phi(Z_i'\alpha)\}(l_i|D_i = 0)$$

$$l_i = \int_{-\infty}^{\infty} (l_i|u_i) f(u) du, \quad u \sim N(0, \sigma_u^2)$$

from this we see that the likelihood function for the whole sample is the product of L_i over the N individuals, and the log-likelihood is

$$lnL = \sum_{i=1}^{N} lnL_i$$

As you can see, the likelihood get more complex if time series increase. Besides, we have to tackle with a T-dimensional integral to estimate the whole model. Traditionally, we can apply Laplace approximation or simulated maximum likelihood approach to this model.

3 MCMC

The most difficult reason for this model is we can not observe y_{it}^* and D_i^* directly. However, with augmentation approach, Bayesian method can draw $y_{it}*$ and D_i^* from its posterior distribution. Moreover, Bayesian methods do not require handle high-dimensional integral problems. Due to this benefit, Bayesian method is more attractive than classical method for this model. Now, we assume prior distributions:

$$\beta \sim N(0, V_\beta)$$
$$\alpha \sim N(0, V_\alpha)$$
$$\sigma_u^2 \sim IG(a_u, b_u)$$
$$\sigma^2 \sim IG(a, b)$$

Model Posterior:

$$\pi(\beta)\pi(\alpha)\pi(\sigma_u^2)\pi(\sigma^2) \prod_i f(l_i|u_i) f(u_i)$$

Marginal Posterior:

1. draw $d_i^* \sim TN_{-\infty,0}(Z\alpha, 1)$ if $\sum_t y_{it} = 0$ or $d_i^* \sim TN_{0,\infty}(Z\alpha, 1) o.w.$
2. draw $y_{it}^{**} \sim TN_{-\infty,0}(X\beta + u, \sigma^2)$
3. draw $\alpha \sim N((Z'Z + V_\alpha^{-1})^{-1} Z'd, (Z'Z + V_\alpha^{-1})^{-1})$
4. draw $\beta \sim N((\frac{1}{\sigma^2} X'X + V_\beta^{-1})^{-1}(\frac{1}{\sigma^2} X'(y - u \otimes \iota_T)), (\frac{1}{\sigma^2} X'X + V_\beta^{-1})^{-1})$
5. draw $u \sim N(\frac{\sigma_u^2}{\sigma^2/T + \sigma_u^2} \frac{1}{T} \sum_t (y - X\beta), \frac{\sigma^2 \sigma_u^2}{\sigma^2 + T\sigma_u^2})$
6. draw $\sigma_u^2 \sim IG(a_u + \frac{n}{2}, b_u + \frac{u'u}{2})$
7. draw $\sigma^2 \sim IG(a + \frac{nT}{2}, b + \frac{1}{2}(y^{**} - X\beta - u \otimes \iota_T)'(y^{**} - X\beta - u \otimes \iota_T))$

4 SMC^2

Traditionally, sequential monte carlo is mainly used for state-space models in the x-dimension, which iterated through time series. Moreover, Chopin, 2002 proposed sequential monte carlo method (iterated batch importance sampling) for static models in the θ-dimension, such as probit model and mixture normal models. Combined the benefits of the two methods, a nested particle filter called SMC^2 was proposed by Chopin et al., 2013. This sequential monte carlo algorithm propagates and resample many particle filter in x-dimension for every θ. And then particle Markov chain Monte Carlo (PMCMC) developed by Andrieu et al., 2010 is applied for rejuvenation steps. Thus, the θ-particles target the correct intractable posterior distribution at each iteration t. Due to its strong flexibility that can approximate intractable distributions with unobserved data, I used SMC^2 for estimating the two-tier Tobit panel data model with unobserved heterogeneity and compare the results with MCMC method.

First, MCMC methods have to monitor the trace plots in order to diagnose the stationary of marginal distributions, but there still have no rigorous condition to guarantee the stationary. In practice, we often have to draw many times to make sure it. On the other hand, since sequnetial monte carlo is iterated through the observations, there is no need to monitor when to stop the algorithm. Hence, in practice, sequnetial monte carlo does not need to test the stationary distribution.

Second, through each iteration, MCMC methods have to take all of its observations for the posteriors. However, if the number of observation is too high or the model is high dimensional, it is too cumbersome for MCMC to draw from posteriors. On the other hand, sequential monte carlo iterates through its observation, so it is more suitable if the data is too big. Besides, the advantage of sequential monte carlo is capable of adopting parallel computing method to decrease the burden of computation cost. Since Bayesian method depends on simulation to assure the convergence, it requires a lot of computational work. However, MCMC can not apply parallel computation method because its posterior distributions have strong dependent property. Moreover, MCMC can not communicate each other after assigning the simulation into separate graphical processing unit (GPU). On the other hand, by adopting random walk metropolis hasting algorithm for move step, sequential monte carlo has a natural property to fit the parallel computing, since the particles can treat independently without communication.

4.1 SMC^2 Algorithm

sample $\theta^m = (\beta, \alpha, \sigma, \sigma_u)^m$ from $p(\theta)$, $m = 1, ..., N_\theta$ and set $w^m \leftarrow 1$. Then, at time $t = 1, ..., N$,

Step 1. Reweighting: For each particle θ^m, perform iteration t of the PF (particle filter). Sample independently $(u_t^{1:N_u,m}, a_{t-1}^{1:N_u,m})$ from $N(0, \sigma_u^{2,m})$ and compute incremental weight:

$$\hat{p}(y_t|\theta^m) = \frac{1}{N_u}\sum_{n=1}^{N_u} w_{t,\theta}(u_t^{n,m}) = \frac{1}{N_u}\sum_{n=1}^{N_u} p(l_t|u_t^{n,m})$$

Update the importance weight, $w^m \leftarrow w^m \times \hat{p}(y_t|\theta^m)$

Step 2. Resampling: resapmle $(\theta^m, w^m)_{m=1,...,N_\theta} \to (\theta^{r,m}, 1)_{m=1,...,N_\theta}$, according to a given selection scheme.

Step 3. Move:

(a). draw $\tilde{\theta}^m \sim K_{t+1}(\theta^{r,m}, .)$ for $m = 1, ..., N_\theta$, where K_{t+1} is a transitional kernel with stationary distribution π_{t+1}.

(b). Run a new PF for every $\tilde{\theta}^m$: sample independently $(\tilde{u}_{1:t}^{1:N_u}, \tilde{a}_{1:t-1}^{1:N_u})$ from $N(0, \tilde{\sigma_u}^{2,m})$ and compute $\hat{Z}_t(\tilde{\theta}, \tilde{u}_{1:t}^{1:N_u}, \tilde{a}_{1:t-1}^{1:N_u}) = \prod_t \{\frac{1}{N_u}\sum_{n=1}^{N_u} w_{t,\tilde{\theta}}(u_t^{n,m})\}$

(c). Accept the move with probability

$$1 \wedge \frac{p(\tilde{\theta})\hat{Z}_t(\tilde{\theta}, \tilde{u}_{1:t}^{1:N_u}, \tilde{a}_{1:t-1}^{1:N_u})K_{t+1}(\tilde{\theta}^m, .)}{p(\theta)\hat{Z}_t(\theta, u_{1:t}^{1:N_u}, a_{1:t-1}^{1:N_u})K_{t+1}(\theta^m, .)}$$

Step 4. Loop: $t \leftarrow t+1, (\theta^m, w^m)_{m=1,..,N_\theta}$ and return to Step 1.

5 Simulation

In this section, we provide DGA and prior setting for two-tier tobit panel data model. For comparison, we also consider the simulated maximum likelihood (SML) in addition to data augmentation and SMC^2. Since u can be simulated from standard normal distribution, we can implement SML easily. Here I set number of individuals $N = 1000$, time period $T = 10$, $\alpha = [-2, 4]'$, $\beta = [0.5, 0.3]'$, $\sigma_u = 1$ and $\sigma_\alpha = 1$. For SMC^2, the number of θ particles is $N_\theta =$ 10,000 and the number of u particles is $N_u = 100$.

Parameter	SML		data augmentation		SMC^2	
	Mean	Std	Mean	Std	Mean	Std
$\beta_0 = 0.5$	0.5909	0.0003	-1.2657	0.0698	0.4450	0.0116
$\beta_1 = 0.3$	0.2227	0.0004	0.2437	0.0361	0.4017	0.0190
$\alpha_0 = -2$	-1.4292	0.0282	2.9535	0.4025	-1.8850	0.0687
$\alpha_1 = 4$	3.2149	0.0523	0.8949	0.8019	3.7113	0.0662
$\sigma_u = 1$	1.0148	0.0001	1.9511	0.0454	0.0127	0.0068
$\sigma = 1$	0.0568	0.0098	1.0025	0.0083	1.0249	0.0220

Table 1: Mean and Std of estimators

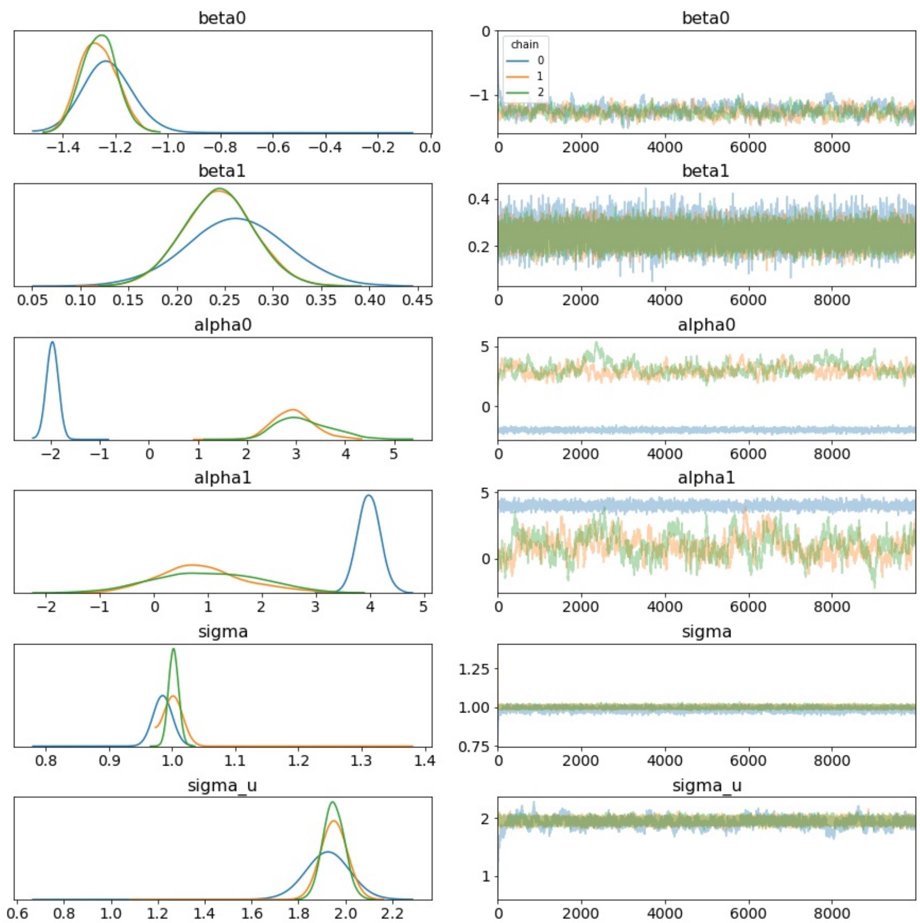

Figure 1: MCMC trace plot

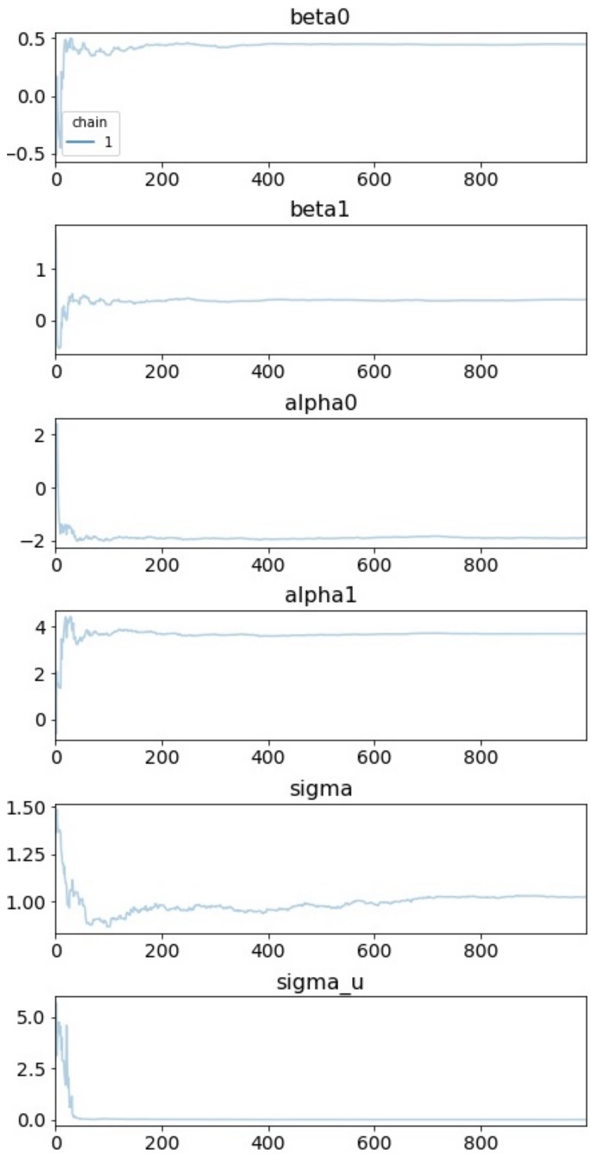

Figure 2: SMC^2 trace plot

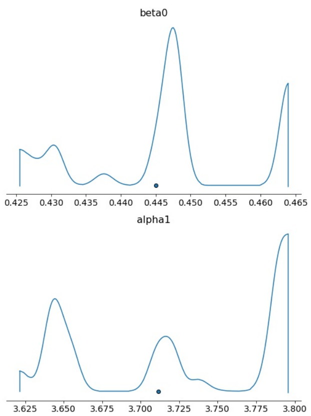

Figure 3: SMC^2 density plot

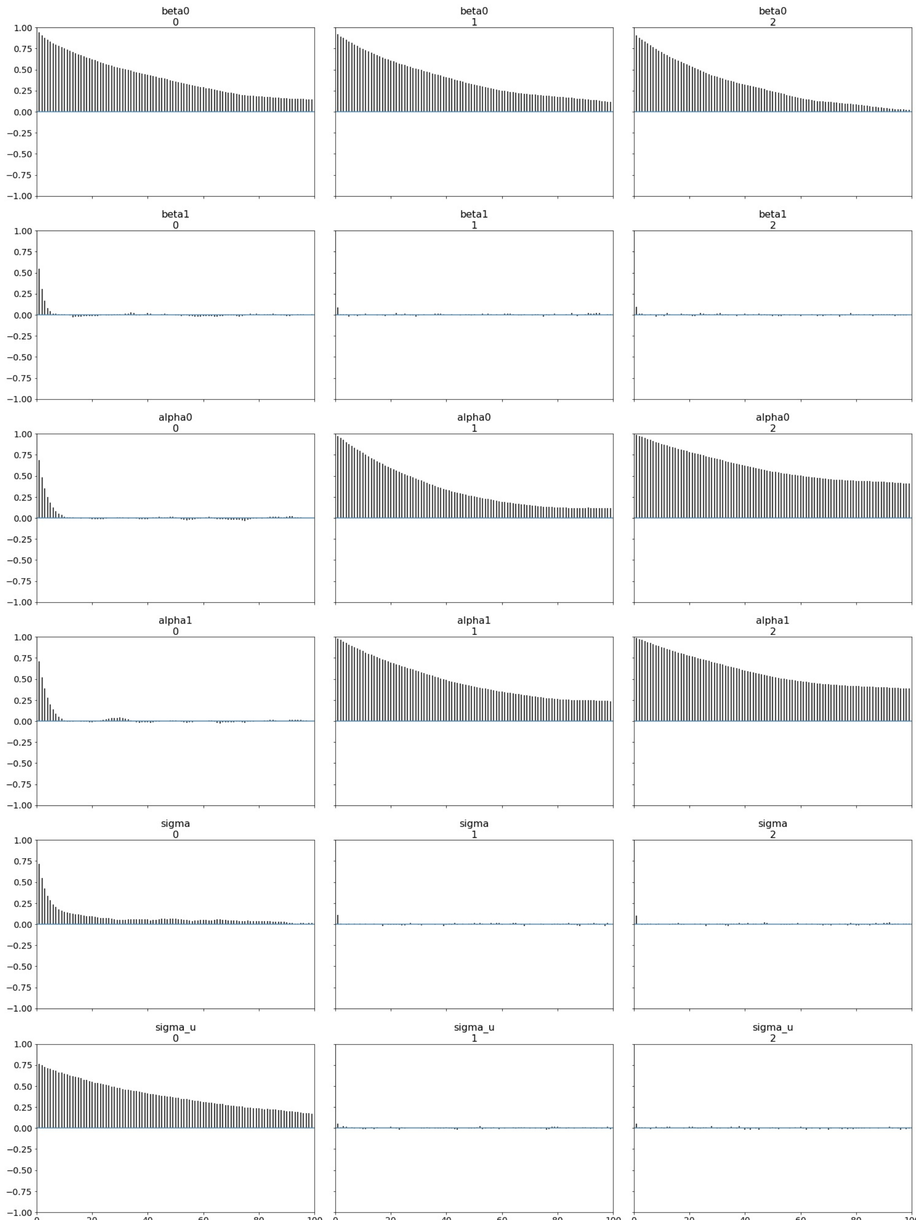

Figure 4: MCMC autocorrelation plot

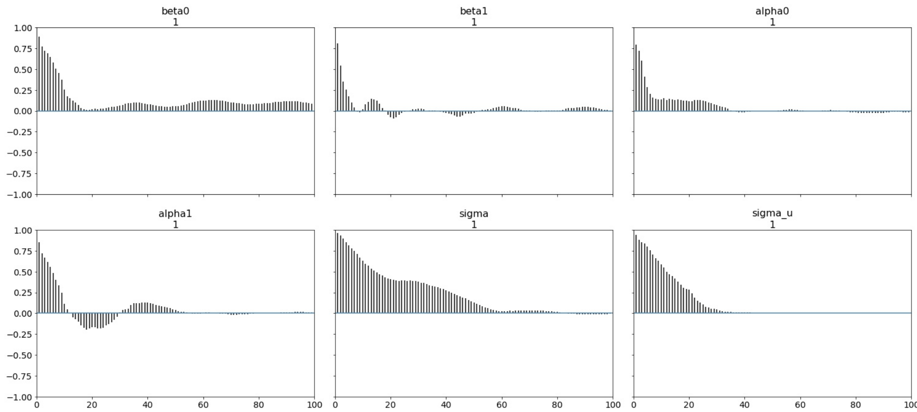

Figure 5: SMC^2 autocorrelation plot

6 Summary

From simulation results, we observe that most parameter estimators of SMC^2 is more accuracy than MCMC method except β_1 and σ_u. Specifically, SMC^2 parameters of participation stage (α_0, α_1) perform better than MCMC and SML. Besides, all SMC^2 parameter estimators of standard error is much less than MCMC. And comparison of autocorrelation plots show most parameters of SMC^2 decay faster than MCMC, which is a evidence that mixing property is improved. However, σ_u of SML is much more accurate than MCMC and SMC^2 Finally, we can observe that density plots of SMC^2 indicates this algorithms indeed could approximate nonlinear intractable posterior distribution with particles. However, according to the marginal posterior distributions of MCMC, α_0, α_1 should be uni-modal distributions rather than multi-modal.

However, it also take a lot of time to compute (1.5 weeks), because for random walk metropolis hasting part, I have to draw draw $\sigma, \sigma_u > 0$for all particles and cost a lot of time. Fortunately, I can apply parallel computation to increase the efficiency drastically. As a result, SMC is more appealing while estimating nonlinear panel data models.

References

Andrieu, C., Doucet, A., & Holenstein, R. (2010). Particle markov chain monte carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *72*(3), 269–342. https://doi.org/https://doi.org/10.1111/j.1467-9868.2009.00736.x

Chopin, N. (2002). A sequential particle filter method for static models. *Biometrika*, *89*(3), 539–552. https://doi.org/10.1093/biomet/89.3.539

Chopin, N., Jacob, P., & Papaspiliopoulos, O. (2013). Smc2: An efficient algorithm for sequential analysis of state space models. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, *75*, 397–426. https://doi.org/10.1111/j.1467-9868.2012.01046.x