Maximum likelihood estimation of the time elapsed between the ordered, censored observations of two events

Daniel Lawson[•]
University of Notre Dame, Notre Dame IN, USA
December 18, 2003

JEL Classification: C34 Econometric Methods: Multiple/Simultaneous Equation Models: Truncated and Censored Models
Keywords: Maximum Likelihood, Tobit, ordered SUR Tobit, seemingly unrelated regressions

This paper proposes a log-likelihood function to consistently estimate an ordered SUR Tobit model. This model is useful to analyze time elapsed between two events only observable in a particular order and if they have taken place by observation time.

1. Introduction

The estimation of models studying the determinants of the time elapsed between two events, one of which cannot be observed after the other, and where observation of the actual time of occurrence of one or both events may additionally be censored because it had not yet occurred at the time of observation poses intriguing challenges.

Simply analyzing the time elapsed between two events as the dependent variable fails to take advantage of all information, as cases where neither event has yet taken place or where only one of the two events has yet taken place also contain information. Instead of treating the time elapsed between the two events as the dependent variable, one should use a two equation model to examine the determinants of the time elapsed from a well-defined starting point to each event.

In this estimation, data censoring would lead to bias in ordinary least squares estimation of the two variables separately. The fact that the two events may be correlated in their error term precludes using a standard Tobit model to estimate the time elapsed until the two events separately. The ordering property precludes the use of a nested Tobit model (Lee 1992) or a standard SUR Tobit (Huang 1999). In this type of analysis, existing models fail to exploit all available information.

This paper proposes a maximum likelihood model to estimate the determinants of the time elapsed between a starting point and each of two events, where observation of both events is

[•] Send correspondence to Daniel Lawson, Department of Economics and Econometrics, University of Notre Dame, Notre Dame, IN 46556; Phone: (574) 631-7689; Fax: (574) 631-4783; E-mail dlawson1@nd.edu

censored if it has not taken place by the time of observation and observation of one of the events is censored if the other takes place first.

2. The ordered SUR Tobit model

Suppose the existence of two events of interest that occur to an individual in time: event one and event two. Let R_i^* be the time elapsed from some well-defined starting point for this individual until event one takes place for individual i and let D_i^* be the time elapsed from the same starting point until event two takes place for the same individual. D_i^* is observable if it has taken place by the time of observation. R_i^* is observable if it has taken place by the time of observation **and** if event two did not take place before event one. Define A_i to be the time elapsed from the starting point until the time of observation. Note that the starting point could be different for different individuals, but must be consistent to calculate R_i, D_i, and A_i for each individual.

Assume that the determinants of R_i^* and D_i^* can be modeled as follows:

$$R_i^* = x_i B + e^r_i \tag{1.}$$

$$D_i^* = y_i G + e^d_i \tag{2.}$$

where e^r_i and e^d_i are bivariate normally distributed with means of zero, variances of σ^2_R and σ^2_D, respectively, and correlation ρ and independent of the error terms for other individuals $j \neq i$. Given this model, one can calculate the likelihood of any set of observation of R_i, D_i, and A_i. There exist four cases that could apply to the information observed (or not observed) regarding R_i^* and D_i^*.

Case 1: R_i^* and D_i^* observed as R_i and D_i

Case 2: R_i^* observed as R_i; D_i^* not observed but greater than or equal to A_i

Case 3: D_i^* observed as D_i; R_i^* not observed but greater than D_i

Case 4: R_i^* and D_i^* both not observed, but greater than A_i

Define dummy variables W_i and L_i such that W_i is equal to zero if D_i^* is actually observed (i.e. event two happens before the time of observation) and one otherwise, while L_i is equal to zero if R_i^*

is actually observed (i.e. event one happens before both event two and the time of observation) and one otherwise. Then, the likelihood function corresponding to this model would be as follows:

$$L = \prod_{(1-W_i)(1-L_i)=1} L\left(R_i = x_i B \wedge D_i = y_i G\right) \prod_{(1-W_i)(L_i)=1} L\left(R_i = x_i B \wedge y_i G \geq A_i\right)$$
$$\prod_{(W_i)(1-L_i)=1} L\left(x_i B \geq D_i \wedge D_i = y_i G\right) \prod_{(W_i)(L_i)=1} L\left(x_i B \geq A_i \wedge y_i G \geq A_i\right) \tag{3.}$$

$$= \prod_{(1-W_i)(1-L_i)=1} L\left(R_i = x_i B \mid D_i = y_i G\right) L\left(D_i = y_i G\right)$$
$$\prod_{(1-W_i)(L_i)=1} L\left(y_i G \geq A \mid R_i = x_i B\right) L\left(R_i = x_i B\right)$$
$$\prod_{(W_i)(1-L_i)=1} L\left(x_i B \geq D_i \mid D_i = y_i G\right) L\left(D_i = y_i G\right) \tag{4.}$$
$$\prod_{(W_i)(L_i)=1} L\left(x_i B \geq A_i \wedge y_i G \geq A_i\right)$$

$$\ln L\left(B, G, \sigma_D, \sigma_R, \rho \mid \{D_i, R_i, A_i, x_i, y_i, L_i, W_i\}_{i=1\ldots n}\right) =$$
$$\sum_{i=1}^{n} \left(1-L_i\right)\left(1-W_i\right) \ln \frac{1}{\sigma_R \sigma_D \sqrt{1-\rho^2}} \phi\left(\frac{D_i - y_i G}{\sigma_D}\right) \phi\left(\frac{R_i - x_i B - \rho \frac{\sigma_R}{\sigma_D}\left(D_i - y_i G\right)}{\sigma_R \sqrt{1-\rho^2}}\right)$$
$$+ L_i \left(1-W_i\right) \ln \frac{1}{\sigma_R} \phi\left(\frac{R_i - x_i B}{\sigma_R}\right) \left(1 - \Phi\left(\frac{A_i - y_i G - \rho \frac{\sigma_D}{\sigma_R}\left(R_i - x_i B\right)}{\sigma_D \sqrt{1-\rho^2}}\right)\right)$$
$$+ \left(1-L_i\right) W_i \ln \frac{1}{\sigma_D} \phi\left(\frac{D_i - y_i G}{\sigma_D}\right) \left(1 - \Phi\left(\frac{D_i - x_i B - \rho \frac{\sigma_R}{\sigma_D}\left(D_i - y_i G\right)}{\sigma_R \sqrt{1-\rho^2}}\right)\right)$$
$$+ L_i W_i \ln\left(1 - \Lambda\left(\frac{A_i - x_i B}{\sigma_R}, \frac{A_i - y_i G}{\sigma_D}, \rho\right)\right) \tag{5.}$$

where Φ is the cumulative density function of the standard normal distribution, ϕ is the probability density function of the standard normal distribution, and Λ is the cumulative density function of the bivariate standard normal distribution with correlation ρ.

This log likelihood function would then be maximized using a Newton-Rhapson or similar algorithm to obtain a consistent Maximum Likelihood estimate, with all the standard properties thereof. Note that calculation of the last term in this expression will require double integration of the bivariate normal density function. This double integration is computationally feasible, but higher dimensions, corresponding to the evaluation of the determinants of the time elapsed until more than two events would quickly become computationally infeasible; see Huang (1999) for Monte Carlo methods of estimating this calculation for greater numbers of equations.

3. Implications and conclusion

While this model can be used to examine the determinants of time elapsed until each event, it is particularly useful for the study of the determinants of the length of time between event one and event two. One useful application of this would be the study of time spent in retirement, where event one is retirement and event two is death (Ghilarducci and Neuman, 2003). In this case, the effect of interest would be the result of a change in an independent variable on the length of time in retirement, max $(D_i^* - R_i^*, 0)$. This model takes advantage of all information available at the time of observation.

References

Ghilarducci, T. and Neuman, K., 2003, The Distribution of Early Retirement Leisure: Evidence from the HRS, unpublished

Huang, H.-C., 1999, Estimation of the SUR Tobit Model via the MCECM Algorithm, Economics Letters 64, 25-30

Lee, B.-J., 1992, A nested Tobit analysis, Economics Letters 38, 296-273