Chapter 6

Overview of the Econometric Equations

6.1 Introduction

In this chapter we present some basic elements of the econometric approach adopted in the empirical investigation of the theoretical implications of our analysis of the bilateral FDI flows in the preceding part. Recall that a crucial feature of the formation of FDI that we emphasized in the first part is the two-fold nature of FDI decisions. There is a decision to make concerning the question whether to invent at all - captured by a "threshold" selection equation; and concerning how much to invest - captured by the flow or gravity equation.

6.2 The Heckman Selection Model

The two-fold nature of FDI decision gives rise to many cases of zero actual FDI flows. With n countries in a sample, there are potentially n(n-1) pairs of source-host (s,h) countries. In fact, the actual number of (s,h) pairs with observed flows is typically much smaller. Therefore, the selection of the actual number of (s,h) pairs, which is naturally *endogenous*, cannot be ignored; that is, this selection cannot be taken as *exogenous*, as has been often a standard practice in gravity models in the literature. This feature of FDI decisions lends itself naturally to the application of the Heckman selection model (1974, 1979).[1] This selection-bias method is adopted to jointly estimate the likelihood of surpassing a certain threshold (the selection equation) and the magnitude of the FDI flow, provided that the threshold is indeed surpassed (the flow equation).

Specify the flow equation [such as equation (5.10)] as

$$Y_{ijt} = \begin{cases} Y^*_{ijt} = X_{ijt}\beta + u_{ijt} & \text{if } \pi^*_{ijt} \geqq 0 \\ 0 & \text{if } \pi^*_{ijt} < 0 \end{cases}, \tag{6.1}$$

where Y^*_{ijt} is a latent variable denoting the flow of notional FDI from source country i to host country j in period t; X_{ijt} is a vector of explanatory variables; β is a coefficient vector; u_{ijt} is an error term; and Y_{ijt} is the actual flow of FDI. Note that Y^*_{ijt} can take both positive and negative values. Note also that the actual flow of FDI, Y_{ijt}, is zero not only when the notional flow, Y^*_{ijt}, is negative; Y_{ijt} may be zero even when Y^*_{ijt} is positive, but does not provide enough profit to surpass the threshold. π^*_{ijt} is the substance of the selection equation [see, for instance, condition (5.11)], and is specified by

$$\pi^*_{ijt} \equiv \pi^{*\prime}_{ijt}/\sigma_{\pi^{*\prime}} = (W_{ijt}\gamma - C_{ijt})/\sigma_{\pi^{*\prime}}, \tag{6.2}$$

where π^*_{ijt} indicates whether an FDI would be made or not (depending whether it is positive or negative); W_{ijt} is a vector of explanatory variables (which may overlap with the explanatory variables of X_{ijt}); C_{ijt} is the fixed cost of setting up new investment; γ is a vector of coefficients; and $\sigma_{\pi^{*\prime}}$ is the standard deviation of $\pi^{*\prime}$. The setup cost C^*_{ijt} is given by

$$C^*_{ijt} = A_{ijt}\delta + v_{ijt}, \tag{6.3}$$

where A_{ijt} is a vector of explanatory variables; δ is a vector of coefficients; and v_{ijt} is an error term. Substituting for C^*_{ijt} in equation (6.2) from equation (6.3), we get:

$$\pi^*_{ijt} = Z_{ijt}\theta + \varepsilon_{ijt}, \tag{6.4}$$

where $Z_{ijt} = (W_{ijt}, A_{ijt})$; $\theta = (\gamma/\sigma_{\pi^{*\prime}}, -\delta/\sigma_{\pi^{*\prime}},)$; and

$$\varepsilon_{ijt} = -v_{ijt}/\sigma_{\pi^{*\prime}}. \tag{6.5}$$

Assuming that u_{ijt} and v_{ijt} are normally distributed with zero means, it follows that $\varepsilon_{ijt} \sim N(0,1)$. The error terms, u_{ijt} and ε_{ijt}, are bivariate normal:

$$\begin{pmatrix} u_{ijt} \\ \varepsilon_{ijt} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{Y^*} & \rho_{Y^*\pi^*}\sigma_{Y^*} \\ \rho_{Y^*\pi^*}\sigma_{Y^*} & 1 \end{pmatrix}\right). \tag{6.6}$$

Define the following indicator function:

$$D_{ijt} = \begin{cases} 1 \text{ if } \pi^*_{ijt} \geqq 0 \\ 0 \text{ otherwise} \end{cases}. \tag{6.7}$$

The latter function indicates whether the threshold is surpassed and an FDI flow is formed or not. Note that π^* itself is a latent variable which is not observed. But we do observe D, that is we do observe whether π^* is positive or not.

The expected value of Y_{ijt}, conditional on the event that there is indeed a positive FDI flow, is given by

$$E\left(Y_{ijt}/D_{ijt}=1\right) = X_{ijt}\beta + E\left(u_{ijt}/D_{ijt}=1\right) \equiv X_{ijt}\beta + \beta_\lambda \lambda_{ijt}, \tag{6.8}$$

where

$$\beta_\lambda = \rho_{Y^*\pi^*}\sigma_{Y^*} \tag{6.9}$$

and

$$\lambda_{ijt} = \frac{\phi(Z_{ijt}\theta)}{\Phi(Z_{ijt}\theta)} \tag{6.10}$$

is the inverse Mills ratio; ϕ and Φ are the density and cumulative unit-normal distribution functions, respectively. Note again that we do not observe π^*_{ijt}, but we do observe D_{ijt}. Because $Prob(D_{ijt} = 1) = Prob(\pi^*_{ijt} \geqq 0) = \Pr ob(\varepsilon_{ijt} \geqq -Z_{ijt}\theta) = \Pr ob(\varepsilon_{ijt} \leqq Z_{ijt}\theta)$, by equation (6.4) and the symmetry of the normal distribution, it follows that

$$Prob(D_{ijt} = 1) = \Phi(Z_{ijt}\theta). \tag{6.11}$$

The maximum likelihood method is then employed to jointly estimate the flow coefficient vector β and the selection coefficient vector θ.

Note that λ_{ijt} depends on X_{ijt}. Therefore, one can see from equation (6.8) that OLS estimates of the coefficient vector β of the flow equation, confined to positive observations of Y_{ijt} (that is, discarding the zero flows), is biased because such estimates include also the effect of X_{ijt} on Y_{ijt} through the term $\beta_\lambda \lambda_{ijt}$. Figure 6.1 explains the intuition for the cause of the bias for the case where $\rho_{Y^*\pi^*} > 0$. Suppose, for instance, that x_{ijt} measures the productivity differential between the ith source country and the potential jth host country, holding all other variables constant. Our theory predicts that the parameter β_x is positive. This is shown by the upward sloping line AB. Note that the slope is an estimate of the "true" marginal effect of x_{ijt} on Y^*_{ijt}. But recall that flows could also be equal to zero, if the setup costs are sufficiently high. A threshold, which is derived from the setup costs, is shown as the curve $TT\prime$ in Figure 6.1. However, if we discard observations with zero actual FDI flows, the remaining sub-sample is no longer random. As equation (6.2) makes clear, the *selection* of country pairs into the sub-sample depends on the vector X_{ijt} (including x_{ijt}). To illustrate, suppose that for high values of x_{ijt} (say, x^H in Figure 6.1), (i, j) pair-wise FDI flows are all positive. That is, for all pairs of countries in the sub-sample the threshold is surpassed and the *observed* average of notional FDI flows for $x_{ijt} = x^H$ is also equal to the conditional population average for FDI flows, point R on line AB. However, this does not hold for low values of x_{ijt} (say, x^L). For these (i, j)-pairs, we observe positive values of $Y_{i,j,t}$ only for a subset of country pairs in the population. Point S is, for instance, excluded from the sub-sample of positive FDI flows. Consequently, for low x_{ijt}'s, we observe only flows between country pairs with low setup costs (namely, with low $v_{i,j,t}$'s). As a result, the observed average of the FDI flows is at point $M\prime$, whereas the "true" average is at point M. As seen in Figure 6.1, the OLD regression line for the sub-sample is therefore the $A\prime B\prime$ line, which underestimates the effect of productivity differentials on bilateral FDI flows.

(Figure 6.1 about here)

If we do not discard the zero FDI flow observations, the OLS estimates of β are still biased, because they are based on observations on Y rather than on Y^*.[2]

6.3 The Tobit Model

As we have seen, our theory suggests that actual FDI flows may be zero even when notional FDI flows are not. In fact, the actual FDI flow variable in a typical sample is zero for a significant fraction of the sample, but is roughly continuously distributed for positive values. In such circumstances, the Tobit model is often employed. The jist of this model is in our case as follows. Suppose that instead of equation (6.1) , we were to have:

$$Y'_{ijt} = \begin{cases} Y^*_{ijt} = X_{ijt}\beta + \mu_{ijt} & \text{if } Y^*_{ijt} \geqq 0 \\ 0 & \text{if } Y^*_{ijt} < 0 \end{cases}, \tag{6.1a}$$

where $\mu_{ijt} \sim N(0, \sigma_Y^2)$. Define now an index function

$$D'_{ijt} = \begin{cases} 1 \text{ if } Y^*_{ijt} \geqq 0 \\ 0 \text{ if } Y^*_{ijt} < 0 \end{cases}. \tag{6.7a}$$

One can show [see, for instance, Woolridge (2003)] that in this case

$$E\left(Y'_{ijt}/D'_{ijt} = 1\right) = X_{ijt}\beta + \beta'_\lambda \lambda'_{ijt}, \tag{6.8a}$$

where

$$\beta'_\lambda = \sigma_{Y^*} \tag{6.9a}$$

and

$$\lambda'_{ijt} = \frac{\phi(X_{ijt}\beta/\sigma_{Y^*})}{\Phi(X_{ijt}\beta/\sigma_{Y^*})} \tag{6.10a}$$

Comparing the Tobit equations (6.8a)-(6.10a) with the Heckman equations (6.8)-(6.10), one can see that the Tobit model is a special case of the Heckman model when Y^*_{ijt} and π^*_{ijt} are fully positively correlated, that is: $\rho_{Y^*\pi^*} = 1$. In this special case, there is a perfect correlation between the event $Y^*_{ijt} \geqq 0$ and the event $\pi^*_{ijt} \geqq 0$ (for any given X_{ijt}), and therefore the two fundamental equations (6.1) and (6.1a) are the same. Thus, if the estimated value of $\rho_{Y^*\pi^*}$ in the Heckman model is significantly below 1, then the Tobit estimates of the coefficient vector β are biased.

6.4 Conclusion

Historically, empirical studies of the flows FDI where cast in a framework of gravity equations. However, fixed setup costs create some thresholds for FDI flows. The latter are therefore positive only when a certain profitability threshold is surpassed. Thus, a typical sample of FDI flows contains a mass (a "big fraction") of zero entries. The empirical literature developed after Tinbergen (1962) has either *omitted* pairs with no FDI flows, or treated reported zero flows as *literally* indicating *zero flows*[3] and employed Tobit estimation methods.[4] In this chapter we overviewed the Heckman Selection and the Tobit methods and illustrated why the former is more appropriate than either the latter or the OLS method in case there are fixed setup costs. Unlike the Tobit and the OLS method, the Heckman method provides unbiased estimates for the determinants of the flows of FDI; it also provides unbiased estimates for the determinants of the endogenous selection of FDI flows, which selection is ignored by both the Tobit and OLS methods.