Moshe Buchinsky
Department of Economics
UCLA

Economics 203C
Spring 2003

Lecture Note 13

Qualitative Response Models II

There are some applications in which the dependent variable is continuous, but the domain may be constrained. For example, expenditure on durable good, or hours of work are only in the positive range. That is, the solution to a utility maximization problem is such that there is a *corner solution*, so that the individual may not spend anything on any single durable good, or may decide it is optimal not work at all.

Let y_i^* denote the desired, *latent*, level of the variable of choice. We assume that the model is linear, so that

$$y_i^* = x_i'\beta + \varepsilon_i.$$

The observed level, denoted y_i, is observed only if it exceed a certain threshold. Without loss of generality (for as long as we have a constant in the regression function) we let this threshold equal 0. This is the *censored regression model*

$$y_i = \begin{cases} y_i^* & \text{if } y_i^* > 0, \\ 0 & \text{if } y_i^* \leq 0. \end{cases}$$

In this case applying an OLS regression will yield an inconsistent estimator for β. A regression the we might consider is for the observed y_i, that is,

$$y_i = x_i'\beta + u_i.$$

But, even if we assume that

$$E(\varepsilon_i | x_i) = 0,$$

$$E(u_i | x_i) \neq 0, \quad \text{or alternatively } E(y_i | x_i) \neq x_i'\beta.$$

The mean of y_i is larger than the mean of y_i^*, because y_i is censored from below by 0.

This is also the case if we choose to run the regression only for the observations for which $y_i > 0$, and it is independent of what is the particular distribution of ε_i.

The Tobit Model:

A model in which ε_i is assumed to have a normal distribution, i.e., $\varepsilon_i \sim N\left(0, \sigma_\varepsilon^2\right)$, is called the *Tobit model*.

Using the assumption about ε_i we can compute

$$\begin{aligned}
\Pr\left(y_i = 0|x_i\right) &= \Pr\left(x_i'\beta + \varepsilon_i \le 0|x_i\right), \\
&= \Pr\left(\varepsilon_i \le -x_i'\beta|x_i\right), \\
&= \Pr\left(\frac{\varepsilon_i}{\sigma_\varepsilon} \le \frac{-x_i'\beta}{\sigma_\varepsilon}|x_i\right), \\
&= \Phi\left(\frac{-x_i'\beta}{\sigma_\varepsilon}\right), \\
&= 1 - \Phi\left(\frac{x_i'\beta}{\sigma_\varepsilon}\right).
\end{aligned}$$

The next quantity that we need to compute is the expectation $E\left(y_i|y_i^* > 0\right)$, i.e., the expectation of the dependent variables for those who have positive value. To do this we need to know the conditional density function $f\left(y_i|y_i > 0\right)$.

Note that

$$f\left(y_0|y_i > 0\right) = \frac{d}{dy}F\left(y_0|y_i > 0\right) = \frac{d}{dy}\Pr\left(y_i \le y_0|y_i > 0\right).$$

Also, by Bayes rule,

$$\begin{aligned}
\Pr\left(y_i \le y_0|y_i > 0\right) &= \begin{cases} \frac{\Pr(y_i \le y_0 \cap y_i > 0)}{\Pr(y_i > 0)} & \text{if } y_0 > 0, \\ 0 & \text{otherwise.} \end{cases} \\
&= \frac{\int_0^{y_0} f(y)\,dy}{\Pr\left(y_i > 0\right)}.
\end{aligned}$$

Hence,

$$\begin{aligned}
f\left(y_0|y_i > 0\right) &= \frac{d}{dy}\left[\frac{\int_0^{y_0} f(y)\,dy}{\Pr\left(y_i > 0\right)}\right] \qquad (13.1) \\
&= \frac{f\left(y_0\right)}{\Pr\left(y_i > 0\right)}.
\end{aligned}$$

For a normal distribution, if $v \sim N(\mu, \sigma^2)$, then

$$f\left(v|v > a\right) = \frac{\frac{1}{\sigma}\phi\left(\frac{a-\mu}{\sigma}\right)}{1 - \Phi\left(\frac{a-\mu}{\sigma}\right)}. \qquad (13.2)$$

Also note that

$$\frac{d}{dz}e^{-z^2/2} = -ze^{-z^2/2}. \qquad (13.3)$$

Using the result in (13.1) and (13.2) we have now

$$\begin{aligned} E\left(y_i | y_i > 0, x_i\right) &= E\left(x_i'\beta + \varepsilon_i | x_i'\beta + \varepsilon_i > 0, x_i\right), \\ &= x_i'\beta + E\left(\varepsilon_i | x_i'\beta + \varepsilon_i, x_i\right), \\ &= x_i'\beta + \sigma_\varepsilon E\left(\frac{\varepsilon_i}{\sigma_\varepsilon} | \frac{\varepsilon_i}{\sigma_\varepsilon} > -\frac{x_i'\beta}{\sigma_\varepsilon}\right), \\ &= x_i'\beta + \sigma_\varepsilon \frac{\int_{-x_i'\beta/\sigma_\varepsilon}^{\infty} \frac{1}{2\pi} z e^{-z^2/2} dz}{\Phi\left(\frac{x_i'\beta}{\sigma_\varepsilon}\right)}, \\ &= x_i'\beta + \sigma_\varepsilon \frac{-\frac{1}{2\pi} e^{-z^2/2}\Big|_{-x_i'\beta/\sigma_\varepsilon}^{\infty} dz}{\Phi\left(\frac{x_i'\beta}{\sigma_\varepsilon}\right)}, \quad \text{from (13.3)}, \\ &= x_i'\beta + \sigma_\varepsilon \frac{\phi\left(\frac{x_i'\beta}{\sigma_\varepsilon}\right)}{\Phi\left(\frac{x_i'\beta}{\sigma_\varepsilon}\right)}, \\ &= x_i'\beta + \sigma_\varepsilon \lambda\left(\frac{x_i'\beta}{\sigma_\varepsilon}\right), \end{aligned} \tag{13.4}$$

where the quantity

$$\lambda(a) \equiv \phi(a)/\Phi(a)$$

is called the *inverse Mill's ratio*.

Equation (13.4) states that if we run a regression of the positive y's on x, then we should also include in the regression the term $\lambda\left(x_i'\beta/\sigma_\varepsilon\right)$. A failure to do so will result in a bias estimate of β due to omitted variable bias.

Define now

$$d_i = \begin{cases} 1 & \text{if } y_i^* > 0, \\ 0 & \text{if } y_i^* \leq 0. \end{cases} \tag{13.5}$$

Then, the likelihood for the data is given by

$$\begin{aligned} L(\beta, \sigma_\varepsilon^2) &= \prod_{i=1}^{n} \Pr\left(y_i = 0 | x_i\right)^{1-d_i} \left[f\left(y_i | y_i > 0, x_i\right) \cdot \Pr\left(y_i = 1 | x_i\right)\right]^{d_i}, \\ &= \prod_{i=1}^{n} \left[1 - \Phi\left(\frac{x_i'\beta}{\sigma_\varepsilon}\right)\right]^{1-d_i} \left[\frac{1}{\sigma_\varepsilon}\phi\left(\frac{y_i - x_i'\beta}{\sigma_\varepsilon}\right)\right]^{d_i}, \\ &= \prod_{d_i=0} \left[1 - \Phi\left(\frac{x_i'\beta}{\sigma_\varepsilon}\right)\right] \prod_{d_i=1} \frac{1}{\sigma_\varepsilon}\phi\left(\frac{y_i - x_i'\beta}{\sigma_\varepsilon}\right). \end{aligned}$$

Alternatively, the log-likelihood function is

$$l(\beta, \sigma_\varepsilon^2) = \sum_{i=1}^{n} \left\{(1 - d_i)\ln\left[1 - \Phi\left(\frac{x_i'\beta}{\sigma_\varepsilon}\right)\right] + d_i \frac{1}{\sigma_\varepsilon}\phi\left(\frac{y_i - x_i'\beta}{\sigma_\varepsilon}\right)\right\}.$$

The MLE's for β and σ_ε^2 are consistent, and, as we already established in the lecture note on MLE, are asymptotically efficient.

Note that unlike in the binary probit model, here σ_ε^2 is identified. This is because we have continuous data on y, at least for the part that exceeds the threshold 0.

Alternative two-step procedure:

Step 1: Using the definition of d_i in (13.5) estimate the probit model to obtain an estimate for $\gamma \equiv \beta/\sigma_\varepsilon$, say $\widehat{\gamma}$.

Step 2: Use the estimate $\widehat{\gamma}$ from step 1 to compute

$$\widehat{\lambda}_i = \widehat{\lambda}(x_i'\gamma) = \lambda\left(x_i'\widehat{\gamma}\right),$$

and run a regression of y_i on x_i and $\widehat{\lambda}_i$.

Note that in this procedure we identify both β and σ_ε, since σ_ε is the coefficient on $\lambda\left(x_i'\gamma\right)$ in (13.4).

Remarks:

1. While we get consistent estimators for β and σ_ε the standard error for $\widehat{\beta}$ and $\widehat{\sigma}_\varepsilon$ need to be adjusted because in the second step we plug in an estimator for $\lambda\left(x_i'\gamma\right)$, i.e., $\lambda\left(x_i'\widehat{\gamma}\right)$, rather then the true $\lambda\left(x_i'\gamma\right)$.

2. It is much easier to obtain estimators for β and σ_ε using the two-step procedure. However, these estimator are not the most efficient estimators. The asymptotically efficient estimators are obtained from the ML described above.

Interpretation:

The coefficient estimates from the tobit model can be interpreted in a number of ways, depending upon the researcher's interests.

Marginal Effects:

1. β_k represents the marginal effect of a change in x_{ki} on the expected value of the latent variable y_i^*, that is

   $$\beta_k = \frac{\partial}{\partial x_{ki}} E\left(y_i^* | x_i\right),$$

   for $k = 1, ..., K$.

2. The marginal effect of a change in x_{ki} on the expected value of y_i, of the sub-population with $y_i > 0$, is given by:

$$\begin{aligned}\frac{\partial}{\partial x_{ki}} E\left(y_i | x_i, y_i^* > 0\right) &= \beta_k + \sigma_\varepsilon \lambda' \left(x_i' \beta / \sigma_\varepsilon\right) \cdot \beta_{ki} / \sigma_\varepsilon, \\ &= \beta_k \left[1 + \lambda' \left(x_i' \beta / \sigma_\varepsilon\right)\right],\end{aligned}$$

   for $k = 1, ..., K$.

3. Note that the marginal effect of a change in x_{ki} on the expected value of the observed variable y_i is not equal to β_k. In fact, we have

$$\frac{\partial}{\partial x_{ki}} E\left(y_i | x_i\right) = \beta_k \Phi \left(\frac{x_i' \beta}{\sigma_\varepsilon}\right),$$

   for $k = 1, ..., K$.

4. The effect of a change in x_{ki} on y_i can be decomposed into two separate effects: (a) the effect of x_{ki} on the probability that the observation *will not be censored*, and (b) the effect of x_{ki} on the conditional mean of y_i^* in the *uncensored part of the distribution*. That is,

$$\begin{aligned}\frac{\partial}{\partial x_{ki}} E\left(y_i | x_i\right) &= \frac{\partial}{\partial x_{ki}} \left[E\left(y_i^* | x_i, y_i^* > 0\right) \cdot \Pr\left(y_i^* > 0 | x_i\right) + 0 \cdot \Pr\left(y_i^* \leq 0 | x_i\right)\right], \\ &= E\left(y_i^* | x_i, y_i^* > 0\right) \frac{\partial}{\partial x_{ki}} \Pr\left(y_i^* > 0 | x_i\right) + \Pr\left(y_i^* > 0 | x_i\right) \frac{\partial}{\partial x_{ki}} E\left(y_i^* | x_i, y_i^* > 0\right),\end{aligned}$$

for $k = 1, ..., K$.

Sample Section Models:

The standard tobit model imposes a structure which is often too restrictive: exactly the same variables affecting the probability of a non-zero observation determine the level of a positive observation and, moreover, with the same sign. This implies, for example, that those who are more likely to spend a positive amount are, on average, also those that spend more on a durable good. There are many examples in economics where this implication does not hold.

There are many problems in which the data we have are generated by individuals making choice of belonging to one group or another (i.e., by individuals' self-selection). This gives rise for the following model:

$$y_i = \begin{cases} y_i^* & \text{if } w_i^* > c, \\ 0 & \text{otherwise,} \end{cases}$$

where

$$\begin{aligned} w_i^* &= z_i'\gamma + v_i, \quad \text{and} \\ y_i^* &= x_i'\beta + \varepsilon_i. \end{aligned}$$

We assume that

$$\begin{aligned} E\left(\varepsilon_i | x_i, z_i\right) &= 0, \\ E\left(v_i | x_i, z_i\right) &= 0, \end{aligned}$$

and we have a random sample from a population for which the joint pdf of ε_i and v_i is $f\left(\varepsilon_i, v_i\right)$.

The discussion of the econometric implications of sample selection started in the early 1970s with the papers by Gronau (1974), Heckman (1974), and Lewis (1974). In their studies, the problem of sample selection bias is discussed in the context of the decision by women of whether or not to participate in the labor force. The distribution of the wage offers i.e., y_i^*, sampled, i.e., $y_i^* = y_i$, is truncated by the *self-selection* of women into the labor force. A woman chooses to be "in the sample" of workers if the wage offer wage exceeds her reservation wage, that is, $w_i^* > c$. Sample selection models are used in a wide variety of other applications.

Denote the decision to work by d_i and note that

$$d_i = \begin{cases} 1 & \text{if } w_i^* > c, \\ 0 & \text{otherwise.} \end{cases}$$

A simple OLS regression of the observed wages, i.e.,

$$y_i = x_i'\beta + u_i,$$

for the women who decided to work (i.e., for the women for whom $d_i = 1$), will, again, yield an inconsistent estimator for β, since, in general, we have

$$E\left(u_i | x_i, z_i, d_i = 1\right) \neq 0,$$

or alternatively

$$E\left(y_i | x_i, z_i, d_i = 1\right) \neq x_i'\beta.$$

Assume that

$$\left.\begin{pmatrix} \varepsilon_i \\ v_i \end{pmatrix}\right| x_i, z_i \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\varepsilon^2 & \sigma_{\varepsilon v} \\ \sigma_{\varepsilon v} & \sigma_v^2 \end{pmatrix}\right).$$

Recall that if

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix}\right),$$

then the conditional distribution of x, conditional on y, is given by

$$x|y \sim N\left(\mu_x + \frac{\sigma_{xy}}{\sigma_y^2}(y-\mu_y), \sigma_x^2 - \frac{\sigma_{xy}^2}{\sigma_y^2}\right).$$

Similarly to the derivation for the tobit model we get here that

$$\begin{aligned} E\left(y_i|x_i, z_i, w_i^* > 0\right) &= x_i'\beta + \rho\sigma_\varepsilon \frac{\phi\left(\frac{z_i'\gamma}{\sigma_v}\right)}{\Phi\left(\frac{z_i'\gamma}{\sigma_v}\right)}, \\ &= x_i'\beta + \rho\sigma_\varepsilon \lambda\left(\frac{z_i'\gamma}{\sigma_v}\right), \end{aligned}$$

where $\rho = \sigma_{xy}/(\sigma_x\sigma_y)$.

Note that if, in fact, $\rho = 0$, then

$$E\left(y_i|x_i, z_i, w_i^* > 0\right) = x_i'\beta,$$

and hence we can use the OLS regression of y on x to obtain a consistent estimator β. The interpretation for this results is as follows. If ε and v are uncorrelated it means that conditional on x_i and z_i, the decision of whether or not to participate in the labor force is uncorrelated with the observed outcome, i.e., y_i. Therefore the sample of with positive earnings can be viewed as a random sample from the population of all women.

Estimation:

Asymptotically the most efficient estimator is given by the MLE. To do that we have to specify the likelihood function and optimize w.r.t. β, γ, and σ_ε^2. The parameter σ_v^2 is not generally identified (it can be identified only under some very special conditions).

Alternatively, one can use Heckman's two-step procedure. This approach (as in the Tobit setting) is based on the conditional mean expressions:

$$E\left(y_i|w_i^* > 0, x_i, z_i\right) = x_i'\beta + \rho\sigma_\varepsilon \lambda\left(\frac{z_i'\gamma}{\sigma_v}\right).$$

Heckman's two-step procedure:

Step 1: For this step use **all** the observations. Run a probit model of $d_i$on z_i. This step provides a consistent estimator for (γ/σ_v), say $\widehat{(\gamma/\sigma_v)}$, although this estimator is not the most efficient estimator for (γ/σ_v), which is the MLE.

Use this estimator construct an estimator for $\lambda_i = \lambda\left(\frac{z_i'\gamma}{\sigma_v}\right)$, by

$$\widehat{\lambda}_i = \lambda\left(z_i'\left(\widehat{\frac{\gamma}{\sigma_v}}\right)\right).$$

Step 2: Run an OLS regression of y_i on x_i and the estimated inverse mill's ratio $\widehat{\lambda}_i$, that is

$$y_i = x_i'\beta + \theta\widehat{\lambda}_i + e_i.$$

This step is carried out only for the *uncensored* observations and provides consistent estimators for β and θ, say $\widehat{\beta}$ and $\widehat{\theta}$, respectively.

As for the tobit model, the standard errors for $\widehat{\beta}$ and $\widehat{\theta}$ have to be adjusted due to the fact that in the second step we use an estimate of λ_i, $\widehat{\lambda}_i$, rather than the true λ_i.

Asymptotic distribution for a two-step GMM estimator:

Let $\varphi(y, x, \theta, \delta)$ be a moment function such that

$$\varphi(y, x, \theta, \delta) = \begin{pmatrix} \varphi_1(y, x, \delta) \\ \varphi_2(y, x, \theta, \delta) \end{pmatrix},$$

that is, $\varphi_1(y, x, \delta)$ is only a function of the parameter vector δ, while $\varphi_2(y, x, \theta, \delta)$ is a function of both δ and θ. We assume here that δ and $\varphi_1(\cdot)$ are $K \times 1$ vector of parameter and $K \times 1$ vector-valued function, respectively. Similarly, θ and $\varphi_2(\cdot)$ are $L \times 1$ vector of parameter and $L \times 1$ vector-valued function, respectively. This is the case for the just-identified models (i.e., the number of equations is the same as the number of unknown parameters), such as the case for the two-step procedures discussed above for the tobit and sample selection models.

There are two ways for estimating δ and θ. One can either solve for the GMM estimator for δ and θ simultaneously using the moment function $\varphi(y, x, \theta, \delta)$. This is exactly what we covered in the lecture note on GMM.

Alternatively, we can solve for $\widehat{\delta}$ based only on $\varphi_1(y, x, \delta)$ and then plug in the estimate into $\varphi_2(y, x, \theta, \delta)$ and solve only for $\widehat{\theta}$ based on $\varphi_2(y, x, \theta, \widehat{\delta})$. This is what is done in the two-step estimation procedures of the tobit and sample selection models.

Of course, the estimators for δ and θ from the two-step procedure are not as efficient as those obtained from solving for δ and θ simultaneously, but it might be a lot easier from computational standpoint.

If one uses a two-step procedure the standard errors for $\widehat{\theta}$ in the second step need to be adjusted to take account of the fact that we plug in an estimator for δ in the second step, rather than the true value for δ. We provide here the derivation for this needed adjustment.

Note that since the first step is an MM problem (i.e., the number of unknown parameters equals the number of moment conditions) $\widehat{\delta}$ solves

$$\begin{aligned} 0 &= \frac{1}{n}\sum_{i=1}^{n}\varphi_1(y_i, x_i, \widehat{\delta}), \\ &= \frac{1}{n}\sum_{i=1}^{n}\varphi_1(y_i, x_i, \delta_0) + \frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\delta'}\varphi_1(y_i, x_i, \delta^*)(\delta_0 - \widehat{\delta}), \end{aligned} \tag{13.6}$$

where δ^* is on the line segment connecting δ_0 and $\widehat{\delta}$. Hence,

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\delta'}\varphi_1(y_i, x_i, \delta^*)\sqrt{n}(\delta_0 - \widehat{\delta}) = -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varphi_1(y_i, x_i, \delta_0). \tag{13.7}$$

Now note that (see Lecture Note 10):

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\delta'}\varphi_1(y_i, x_i, \delta^*) \xrightarrow{p} E_0\left[\frac{\partial}{\partial\delta'}\varphi_1(y, x, \delta_0)\right] \equiv \Delta_\delta, \tag{13.8}$$

a $K \times K$ non-singular matrix, and

$$-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varphi_1(y_i, x_i, \delta_0) \xrightarrow{D} N\left(0, \Lambda_{\varphi_1}\right), \tag{13.9}$$

where

$$\Lambda_{\varphi_1} \equiv E_0\left[\varphi_1(y, x, \delta_0)\varphi_1(y, x, \delta_0)'\right].$$

Consequently, substitution of the results in (13.8) and (13.9) into (13.7) gives

$$\sqrt{n}(\delta_0 - \widehat{\delta}) \xrightarrow{D} N\left(0, \Lambda_\delta\right), \tag{13.10}$$

where

$$\Lambda_\delta = \Delta_\delta \Lambda_{\varphi_1} \Delta_\delta'. \tag{13.11}$$

Note that the result obtained in (13.10) is the general result for an MM estimator (usually referred to as *M-estimator*).

In the second step we have similarly

$$\begin{aligned} 0 &= \frac{1}{n}\sum_{i=1}^{n}\varphi_2(y_i, x_i, \widehat{\theta}, \widehat{\delta}), \\ &= \frac{1}{n}\sum_{i=1}^{n}\varphi_2(y_i, x_i, \theta_0, \delta_0) + \frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\delta'}\varphi_2(y_i, x_i, \theta^*, \delta^*)(\delta_0 - \widehat{\delta}) \\ &\quad + \frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta'}\varphi_2(y_i, x_i, \theta^*, \delta^*)(\theta_0 - \widehat{\theta}), \end{aligned} \tag{13.12}$$

where δ^* and θ^* are on the line segment connecting δ_0 and θ_0 with $\widehat{\delta}$ and $\widehat{\theta}$. We can rewrite (13.12) as

$$\begin{aligned} -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta'}\varphi_2(y_i,x_i,\theta^*,\delta^*)\sqrt{n}(\theta_0-\widehat{\theta}) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varphi_2(y_i,x_i,\theta_0,\delta_0) \\ &\quad +\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\delta'}\varphi_2(y_i,x_i,\theta^*,\delta^*)\sqrt{n}(\delta_0-\widehat{\delta}). \end{aligned} \tag{13.13}$$

Now note that (see Lecture Note 10, again):

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\delta'}\varphi_2(y_i,x_i,\theta^*,\delta^*) \xrightarrow{p} E_0\left[\frac{\partial}{\partial\delta'}\varphi_2(y,x,\theta_0,\delta_0)\right] \equiv \Delta_{2\delta}, \tag{13.14}$$

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta'}\varphi_2(y_i,x_i,\theta^*,\delta^*) \xrightarrow{p} E_0\left[\frac{\partial}{\partial\theta'}\varphi_2(y,x,\theta_0,\delta_0)\right] \equiv \Delta_{2\theta}, \tag{13.15}$$

where $\Delta_{2\theta}$ is an $L \times L$ non-singular matrix, $\Delta_{2\delta}$ is an $L \times K$ matrix.

Also,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varphi_2(y_i,x_i,\theta_0,\delta_0) \xrightarrow{D} N\left(0,\Lambda_{\varphi_2}\right), \tag{13.16}$$

where

$$\Lambda_{\varphi_2} \equiv E_0\left[\varphi_2(y,x,\theta_0,\delta_0)\varphi_2(y,x,\theta_0,\delta_0)'\right].$$

Typically $\sqrt{n}(\delta_0-\widehat{\delta})$ and $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varphi_2(y_i,x_i,\theta_0,\delta_0)$ are asymptotically uncorrelated, as is the case for the tobit and two-step Heckman's procedures. That is, the asymptotic covariance of the two terms on the right-hand-side of (13.13) is 0.

Hence, substitution of the results in (13.14), (13.15), and (13.16), into (13.13), gives

$$\sqrt{n}(\theta_0-\widehat{\theta}) \xrightarrow{D} N\left(0,\Lambda_\theta\right), \tag{13.17}$$

where

$$\Lambda_\theta = \Delta_{2\theta}\left[\Lambda_{\varphi_2}+\Delta_{2\delta}\Lambda_\delta\Delta_{2\delta}'\right]\Delta_{2\theta}'. \tag{13.18}$$