Double-Hurdle Models with Dependent Errors and Heteroscedasticity

Julian Fennema and Mathias Sinning

Heriot-Watt University and RWI-Essen

Essen, April 2nd 2007

Tobit model

Model

Censoring of the dependent variable is traditionally dealt with using the Tobit model

$$y_i = y_i^* = x_i\beta + \varepsilon_1 \qquad \text{if} \qquad y_i^* > 0 \qquad \text{otherwise} \qquad y_i = 0$$

Likelihood function

$$L = \prod_0 \left(1 - \Phi\left(\frac{x_i\beta}{\sigma_{\varepsilon_1}}\right)\right) + \prod_+ \left(\frac{1}{\sigma_{\varepsilon_1}}\phi\left(\frac{y_i - x_i\beta}{\sigma_{\varepsilon_1}}\right)\right)$$

Cragg's formulation

Model

Cragg (1971) proposed the extension that the probability of a zero realisation, $1 - \Phi(.)$, is not directly to the density for a continuous realisation, $\phi(.)$, but instead governed by some other process

$$
\begin{aligned}
y_i &= y_i^* = x_i\beta + \varepsilon_{1i} && \text{if} \quad x_i\beta + \varepsilon_{1i} > 0 \text{ and } z_i\alpha + \varepsilon_{2i} > 0 \\
&= 0 && \text{if} \quad x_i\beta + \varepsilon_{1i} \leq 0 \text{ and } z_i\alpha + \varepsilon_{2i} > 0 \\
& && \text{or} \quad x_i\beta + \varepsilon_{1i} > 0 \text{ and } z_i\alpha + \varepsilon_{2i} \leq 0 \\
& && \text{or} \quad x_i\beta + \varepsilon_{1i} \leq 0 \text{ and } z_i\alpha + \varepsilon_{2i} \leq 0
\end{aligned}
$$

Cragg's formulation - errors

Independent

The original model made the assumption that the two error terms were jointly normal,

$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} \sim N(0, \Sigma), \text{ and uncorrelated, } \Sigma = \begin{pmatrix} \sigma^2_{\varepsilon_1} & 0 \\ 0 & 1 \end{pmatrix}$$

Likelihood function

$$L = \prod_0 \left(1 - \Phi\left(\frac{x_i\beta}{\sigma_{\varepsilon_1}}\right) \cdot 1 - \Phi(z_i\alpha)\right) \prod_+ \left(\Phi(z_i\alpha)\frac{1}{\sigma_{\varepsilon_1}}\phi\left(\frac{y_i - x_i\beta}{\sigma_{\varepsilon_1}}\right)\right)$$

Estimation

Separability

Similar to that demonstrated by McDowell (2003) for count models in Stata, the separability of the likelihood function permits the use of a combination of Stata command to estimate this model: `truncreg` and `probit`

Jones' extension

Correlated errors

This assumption has been relaxed in later work, e.g. Jones (1992), where $\Sigma = \begin{pmatrix} \sigma^2_{\varepsilon_1} & \sigma_{\varepsilon_1}\rho \\ \sigma_{\varepsilon_1}\rho & 1 \end{pmatrix}$

Likelihood function

$$L = \prod_0 \left[1 - F_2\left(z_i\alpha, x_i\beta/\sigma, \rho\right)\right] \prod_+ \Phi\left(\frac{z_i\alpha + \frac{\rho}{\sigma}(y - x_i\beta)}{\sqrt{1-\rho^2}}\right) \frac{1}{\sigma}\phi\left(\frac{(y - x_i\beta)}{\sigma}\right)$$

Estimation

Non-separable

Both parts of this likelihood function must, however, be maximised simultaneously; there is no two-step equivalent. This has been available in Stata, on an *ad hoc* basis since 2004 using the `dhurdle` command written for Stata 7.

Syntax

```
dhurdle y x1 x2, sel(d x1 t1)
```

Comparison to Flood and Gråsjö

		Stata		Gauss	
	True value	Bias(%)	RMSE	Bias(%)	RMSE
β_0	-0.2	-23.3	.665	-18.0	.836
β_1	0.2	-17.6	.152	3.7	.085
β_2	1	.34	.081	2.7	.182
α_0	0.7	49.0	1.25	175	8.065
α_1	0.2	24.6	.881	-44.7	1.325
α_2	-0.2	27.3	.392	-80.7	.570
σ	2	1.38	.256	-.2	.227
ρ	-0.5	-26.3	.261	31.2	.447

Assumptions on the error terms

There is, by now, a wide variety of literature demonstrating that if the assumption of homoscedatic, normally-distributed, errors is violated then ML parameter estimates are inconsistent.

Solutions

Two extensions

1. Heteroscdastic errors
2. Non-normal errors

`dhurdle` now can incorporate variance dependent on a set of independent variables.

Syntax

```
dhurdle y x1 x2, sel(d x1 t1) het(.)
```

Non-normality

Non-normality

Robinson (1982) showed that ML estimation of LDV models leads to inconsistent parameter estimates if the assumption of normally distributed errors does not hold.

Syntax

```
dhurdle y x1 x2, sel(d x1 t1) ihs
```

Pipeline

Work in progress

The final steps to complete the estimation package that are currently underway are

1. Finalise `predict` options for the double hurdle.
2. Code a series of LR tests to test model specification post-estimation.

Testing 1, 2, 3

Using the IHS as the general form, the imposition of the following restrictions is feasible

1. If $\gamma = 0$ then conventional formulation without transformation.
2. If σ is constant then homoscedastic errors.
3. If $\rho = 0$ then independent double hurdle.
4. If $\prod_{+} \Phi\left(z_i \alpha\right)$ then no censoring present and model simplifies to a Heckman.
5. If $\Phi\left(z_i \alpha\right)$ and $\rho = 0$ then no censoring or selection present and model simplifies to a Tobit.