

Inference for Partial Effects in Nonlinear Panel Data Models
Summer North American Stata Users' Group Meeting
Chicago, IL, July 2008

Jeffrey M. Wooldridge
Michigan State University

1. Introduction
2. A General Setup and Quantities of Interest
3. Exogeneity Assumptions
4. Conditional Independence
5. Assumptions about the Unobserved Heterogeneity
6. Dynamic Models
7. Control Function Methods with Endogenous Covariates
8. Estimating Popular Models
9. Extensions

1. Introduction

- Consider a static, unobserved effects probit model for panel data:

$$P(y_{it} = 1|\mathbf{x}_{it}, c_i) = \Phi(\mathbf{x}_{it}\boldsymbol{\beta} + c_i), t = 1, \dots, T. \quad (1)$$

What are the quantities of interest for most purposes? Possibilities: (i) The element of $\boldsymbol{\beta}$, the β_j . These give the directions of the partial effects of the covariates on the response probability. For any two continuous covariates, the ratio of coefficients, β_j/β_h , is identical to the ratio of partial effects (and the ratio does not depend on the covariates or unobserved heterogeneity, c_i).

(ii) The magnitudes of the partial effects. These depend not only on the value of the covariates, say \mathbf{x}_t , but also on the value of the unobserved heterogeneity. In the continuous covariate case,

$$\frac{\partial P(y_t = 1 | \mathbf{x}_t, c)}{\partial x_{tj}} = \beta_j \phi(\mathbf{x}_t \boldsymbol{\beta} + c). \quad (2)$$

- Questions: (i) Assuming we can estimate $\boldsymbol{\beta}$, what should we do about the unobservable c ? (ii) If we can only estimate $\boldsymbol{\beta}$ up-to-scale, can we still learn something useful about magnitudes of partial effects? (iii) What kinds of assumptions do we need to estimate partial effects?

2. A General Setup and Quantities of Interest

- Let $\{(\mathbf{x}_{it}, y_{it}) : t = 1, \dots, T\}$ be a random draw from the cross section.

Suppose we are interested in

$$E(y_{it}|\mathbf{x}_{it}, \mathbf{c}_i) = m_t(\mathbf{x}_{it}, \mathbf{c}_i). \quad (3)$$

\mathbf{c}_i can be a vector of unobserved heterogeneity.

- Partial effects: if x_{tj} is continuous, then

$$\theta_j(\mathbf{x}_t, \mathbf{c}) \equiv \frac{\partial m_t(\mathbf{x}_t, \mathbf{c})}{\partial x_{tj}}, \quad (4)$$

or discrete changes.

- How do we account for unobserved \mathbf{c}_i ? If we know enough about the distribution of \mathbf{c}_i we can insert meaningful values for \mathbf{c} . For example, if $\boldsymbol{\mu}_c = E(\mathbf{c}_i)$, then we can compute the *partial effect at the average* (PEA),

$$PEA_j(\mathbf{x}_t) = \theta_j(\mathbf{x}_t, \boldsymbol{\mu}_c). \quad (5)$$

Of course, we need to estimate the function m_t and $\boldsymbol{\mu}_c$. If we can estimate the distribution of \mathbf{c}_i , or features in addition to its mean, we can insert different quantiles, or a certain number of standard deviations from the mean.

- Alternatively, we can obtain the *average partial effect* (APE) (or *population average effect*) by averaging across the distribution of \mathbf{c}_i :

$$APE(\mathbf{x}_t) = E_{\mathbf{c}_i}[\theta_j(\mathbf{x}_t, \mathbf{c}_i)]. \quad (6)$$

The difference between (5) and (6) can be nontrivial. In some leading cases, (6) is identified while (5) is not. (6) is closely related to the notion of the average structural function (ASF) (Blundell and Powell (2003)). The ASF is defined as

$$ASF(\mathbf{x}_t) = E_{\mathbf{c}_i}[m_t(\mathbf{x}_t, \mathbf{c}_i)]. \quad (7)$$

- Passing the derivative through the expectation in (7) gives the APE.

- How do APEs relate to parameters? Index model:

$$m_t(\mathbf{x}_t, c) = G(\mathbf{x}_t\boldsymbol{\beta} + c), \quad (8)$$

where $G(\cdot)$ is differentiable. Then

$$\theta_j(\mathbf{x}_t, c) = \beta_j g(\mathbf{x}_t\boldsymbol{\beta} + c), \quad (9)$$

where $g(\cdot)$ is the derivative of $G(\cdot)$. Even if $G(\cdot)$ is known, magnitude of effects cannot be estimated without making assumptions about the distribution of c_i

- Important: Definitions of partial effects do not depend on whether \mathbf{x}_t is correlated with \mathbf{c} . Of course, whether and how we estimate them certainly does.

3. Exogeneity Assumptions

- As in linear case, cannot get by with just specifying a model for the contemporaneous conditional distribution, $D(y_{it}|\mathbf{x}_{it}, \mathbf{c}_i)$.
- The most useful definition of strict exogeneity for nonlinear panel data models is

$$D(y_{it}|\mathbf{x}_{i1}, \dots, \mathbf{x}_{iT}, \mathbf{c}_i) = D(y_{it}|\mathbf{x}_{it}, \mathbf{c}_i). \quad (10)$$

Chamberlain (1984) labeled (10) *strict exogeneity conditional on the unobserved effects* \mathbf{c}_i . Conditional mean version:

$$E(y_{it}|\mathbf{x}_{i1}, \dots, \mathbf{x}_{iT}, \mathbf{c}_i) = E(y_{it}|\mathbf{x}_{it}, \mathbf{c}_i). \quad (11)$$

- The *sequential exogeneity* assumption is

$$D(y_{it}|\mathbf{x}_{i1}, \dots, \mathbf{x}_{it}, \mathbf{c}_i) = D(y_{it}|\mathbf{x}_{it}, \mathbf{c}_i). \quad (12)$$

Unfortunately, it is much more difficult to allow sequential exogeneity in nonlinear models. (Most progress for lagged dependent variables or specific functional forms, such as exponential.)

- Neither strict nor sequential exogeneity allows for contemporaneous endogeneity of one or more elements of \mathbf{x}_{it} , where, say, x_{itj} is correlated with unobserved, time-varying unobservables that affect y_{it} .

4. Conditional Independence

- In linear models, serial dependence of idiosyncratic shocks is easily dealt with, either by “cluster robust” inference or Generalized Least Squares extensions of Fixed Effects and First Differencing. With strictly exogenous covariates, serial correlation never results in inconsistent estimation, even if improperly modeled. The situation is different with most nonlinear models estimated by MLE.
- *Conditional independence* (CI) (under strict exogeneity):

$$D(y_{i1}, \dots, y_{iT} | \mathbf{x}_i, \mathbf{c}_i) = \prod_{t=1}^T D(y_{it} | \mathbf{x}_{it}, \mathbf{c}_i). \quad (13)$$

- In a parametric context, the CI assumption reduces our task to specifying a model for $D(y_{it}|\mathbf{x}_{it}, \mathbf{c}_i)$, and then determining how to treat the unobserved heterogeneity, \mathbf{c}_i .
- In random effects and correlated random frameworks (next section), CI plays a critical role in being able to estimate the “structural” parameters and the parameters in the distribution of \mathbf{c}_i (and therefore, in estimating PEAs). In a broad class of popular models, CI plays no essential role in estimating APEs.

5. Assumptions about the Unobserved Heterogeneity

Random Effects

- Generally stated, the key RE assumption is

$$D(\mathbf{c}_i | \mathbf{x}_{i1}, \dots, \mathbf{x}_{iT}) = D(\mathbf{c}_i). \quad (14)$$

Under (14), the APEs are actually nonparametrically identified from

$$r_t(\mathbf{x}_t) \equiv E(y_{it} | \mathbf{x}_{it} = \mathbf{x}_t). \quad (15)$$

- In some leading cases (RE probit and RE Tobit with heterogeneity normally distributed), if we want PEs for different values of \mathbf{c} , we must assume more: strict exogeneity, conditional independence, and (14) with a parametric distribution for $D(\mathbf{c}_i)$.

Correlated Random Effects

A CRE framework allows dependence between \mathbf{c}_i and \mathbf{x}_i , but restricted in some way. In a parametric setting, we specify a distribution for $D(\mathbf{c}_i|\mathbf{x}_{i1}, \dots, \mathbf{x}_{iT})$, as in Chamberlain (1980,1982), and much work since. Distributional assumptions that lead to simple estimation – homoskedastic normal with a linear conditional mean — can be restrictive.

- Possible to drop parametric assumptions with

$$D(c_i|\mathbf{x}_i) = D(c_i|\bar{\mathbf{x}}_i), \quad (16)$$

without restricting $D(c_i|\bar{\mathbf{x}}_i)$. Altonji and Matzkin (2005, *Econometrica*).

- Other functions of $\{\mathbf{x}_{it} : t = 1, \dots, T\}$ are possible.

- APEs are identified very generally. For example, under (16), a consistent estimate of the average structural function is

$$\widehat{ASF}(\mathbf{x}_t) = N^{-1} \sum_{i=1}^N m_t(\mathbf{x}_t, \bar{\mathbf{x}}_i), \quad (17)$$

where $m_t(\cdot)$ is the mean function $E(y_{it} | \mathbf{x}_{it}, \bar{\mathbf{x}}_i)$.

- Need a random sample $\{\bar{\mathbf{x}}_i : i = 1, \dots, N\}$.

Fixed Effects

- The label “fixed effects” is used in different ways by different researchers. One view: $\mathbf{c}_i, i = 1, \dots, N$ are parameters to be estimated. Usually leads to an “incidental parameters problem.”

- Second meaning of “fixed effects”: $D(\mathbf{c}_i|\mathbf{x}_i)$ is unrestricted and we look for objective functions that do not depend on \mathbf{c}_i but still identify the population parameters. Leads to “conditional MLE” if we can find “sufficient statistics” \mathbf{s}_i such that

$$D(y_{i1}, \dots, y_{iT}|\mathbf{x}_i, \mathbf{c}_i, \mathbf{s}_i) = D(y_{i1}, \dots, y_{iT}|\mathbf{x}_i, \mathbf{s}_i). \quad (18)$$

- Conditional Independence is usually maintained.
- Key point: PEAs and APEs are generally unidentified.

6. Dynamic Models

• Nonlinear models with only sequentially exogenous variables are difficult to deal with. More is known about models with lagged dependent variables and otherwise strictly exogenous variables:

$$D(\mathbf{y}_{it} | \mathbf{z}_{it}, \mathbf{y}_{i,t-1}, \dots, \mathbf{z}_{i1}, \mathbf{y}_{i0}, \mathbf{c}_i), t = 1, \dots, T, \quad (19)$$

which we assume also is $D(\mathbf{y}_{it} | \mathbf{z}_i, \mathbf{y}_{i,t-1}, \dots, \mathbf{y}_{i1}, \mathbf{y}_{i0}, \mathbf{c}_i)$. Suppose this distribution depends only on $(\mathbf{z}_{it}, \mathbf{y}_{i,t-1}, \mathbf{c}_i)$ with density

$f_t(\mathbf{y}_t | \mathbf{z}_t, \mathbf{y}_{t-1}, \mathbf{c}; \boldsymbol{\theta})$. The joint density of $(\mathbf{y}_{i1}, \dots, \mathbf{y}_{iT})$ given $(\mathbf{y}_{i0}, \mathbf{z}_i, \mathbf{c}_i)$ is

$$\prod_{t=1}^T f_t(\mathbf{y}_t | \mathbf{z}_t, \mathbf{y}_{t-1}, \mathbf{c}; \boldsymbol{\theta}). \quad (20)$$

- How do we deal with \mathbf{c}_i along with the initial condition, \mathbf{y}_{i0} ? Various approaches have been suggested. One that meshes well with Stata's built-in commands (random effects probit, Tobit, count) was proposed by Wooldridge (2005, Journal of Applied Econometrics). Idea is to model $D(\mathbf{c}_i|\mathbf{y}_{i0}, \mathbf{z}_i)$ directly. Leads to $D(\mathbf{y}_{i1}, \dots, \mathbf{y}_{iT}|\mathbf{y}_{i0}, \mathbf{z}_i)$ and MLE conditional on $(\mathbf{y}_{i0}, \mathbf{z}_i)$. This can be computationally simple for popular models, and can be made somewhat flexible.
- The APEs for the conditional mean are easy to obtain.

7. Control Function Methods with Endogenous Covariates

- General idea is to model endogeneity as an omitted (time-varying) variable. So, start with a “structural” model

$$E(y_{it1} | \mathbf{z}_i, y_{it2}, c_{i1}, v_{it1}) = E(y_{it1} | \mathbf{z}_{it1}, y_{it2}, c_{i1}, v_{it1}), \quad (21)$$

where c_{i1} is the time-constant unobserved effect and v_{it1} is a time-varying omitted factor that can be correlated with y_{it2} . (Papke and Wooldridge, 2008, Journal of Econometrics).

- Elements of \mathbf{z}_{it} are assumed strictly exogenous, and we have at least one exclusion restriction: $\mathbf{z}_{it} = (\mathbf{z}_{it1}, \mathbf{z}_{it2})$.
- APEs average out (c_{i1}, v_{it2}) .

- With a *continuous* endogenous explanatory variable, y_{it2} , can often combine the Chamberlain-Mundlak approach to unobserved effects with the control function approach (Smith-Blundell, Rivers-Vuong) to arrive at

$$D(y_{it1}|y_{it2}, \mathbf{z}_{it1}, \bar{\mathbf{z}}_i, v_{it2}) \quad (22)$$

or

$$E(y_{it1}|y_{it2}, \mathbf{z}_{it1}, \bar{\mathbf{z}}_i, v_{it2}), \quad (23)$$

where v_{it2} are reduced form errors, say,

$$y_{it2} = \psi_{t2} + \mathbf{z}_{it}\delta_2 + \bar{\mathbf{z}}_i\xi_2 + v_{it2}, t = 1, \dots, T. \quad (24)$$

- It can be shown that the APEs are generally available from

$$\widehat{ASF}(y_{t2}, \mathbf{z}_{t1}) = N^{-1} \sum_{i=1}^N \hat{m}_t(y_{t2}, \mathbf{z}_{t1}, \bar{\mathbf{z}}_i, \hat{v}_{it2}), \quad (25)$$

where $m_t(\cdot)$ is the mean function in (23). So, the time averages of the exogenous variables and the reduced form residuals get averaged out.

- Two-step pooled methods are very computationally attractive.

Usually, pooled OLS followed by pooled probit, Tobit, GLM (or even a Cragg hurdle model).

8. Estimating Popular Models

- Pooled and random effects estimation commands in Stata (for probit, Tobit, Poisson, GLM, GEE) often can be used.
- Stata egen command for generating time averages. Need leads and lags of exogenous variables, and the initial condition, for dynamic models.
- For pooled methods, use the “panel bootstrap” feature in Stata to obtain standard errors or confidence intervals.
- Computational time is an issue for dynamic models because it uses full “random effects” with lots of covariates.

8.1 Binary and Fractional Response

- Unobserved effects (UE) “probit” model:

$$E(y_{it}|\mathbf{x}_{it}, c_i) = \Phi(\mathbf{x}_{it}\boldsymbol{\beta} + c_i), t = 1, \dots, T. \quad (26)$$

Assume strict exogeneity (conditional on c_i) and Chamberlain-Mundlak device:

$$c_i = \psi + \bar{\mathbf{x}}_i\boldsymbol{\xi} + a_i, a_i|\mathbf{x}_i \sim \text{Normal}(0, \sigma_a^2). \quad (27)$$

- In binary response case under serial independence, all parameters are identified and MLE (Stata: xtprobit) can be used. Just add the time averages $\bar{\mathbf{x}}_i$ as an additional set of regressors. Then $\hat{\mu}_c = \hat{\psi} + \bar{\mathbf{x}}\hat{\boldsymbol{\xi}}$ and $\hat{\sigma}_c^2 \equiv \hat{\boldsymbol{\xi}}' \left(N^{-1} \sum_{i=1}^N \bar{\mathbf{x}}_i' \bar{\mathbf{x}}_i \right) \hat{\boldsymbol{\xi}} + \hat{\sigma}_a^2$. Can evaluate PEs at, say, $\hat{\mu}_c \pm k\hat{\sigma}_c$.

- The APEs are identified from the ASF, estimated as

$$\widehat{ASF}(\mathbf{x}_t) = N^{-1} \sum_{i=1}^N \Phi(\mathbf{x}_t \hat{\boldsymbol{\beta}}_a + \hat{\psi}_a + \bar{\mathbf{x}}_i \hat{\boldsymbol{\xi}}_a) \quad (28)$$

where, for example, $\hat{\boldsymbol{\beta}}_a = \hat{\boldsymbol{\beta}} / (1 + \hat{\sigma}_a^2)^{1/2}$.

- For binary or fractional response, APEs are identified without the conditional serial independence assumption. Use pooled Bernoulli quasi-MLE (Stata: glm) or generalized estimating equations (Stata: xtgee) to estimate scaled coefficients based on

$$E(y_{it} | \mathbf{x}_i) = \Phi(\mathbf{x}_{it} \boldsymbol{\beta}_a + \psi_a + \bar{\mathbf{x}}_i \boldsymbol{\xi}_a). \quad (29)$$

(Time dummies have been suppressed for simplicity.)

- Remember, the Bernoulli log-likelihood is in the linear exponential family (LEF). Pooled GLM or GEE easy computationally.
- Example from Papke and Wooldridge (2008, Journal of Econometrics). Effects of school spending on student performance (4th grade math pass rate). Spending might be endogenous even after controlling for unobserved heterogeneity. An IV is available, so $\mathbf{z}_{it} = (\mathbf{z}_{it1}, z_{it2})$. The reduced form for y_{t2} is

$$y_{it2} = \psi_{t2} + \mathbf{z}_{it}\delta_2 + \bar{\mathbf{z}}_i\xi_2 + v_{it2}, t = 1, \dots, T \quad (30)$$

and the estimating equation is

$$E(y_{it1}|y_{it2}, \mathbf{z}_i, v_{it2}) = \Phi(\alpha_{\kappa 1} y_{it2} + \mathbf{z}_{it1} \boldsymbol{\delta}_{\kappa 1} + \psi_{t\kappa 1} + \bar{\mathbf{z}}_i \boldsymbol{\xi}_{\kappa 1} + \eta_{\kappa 1} v_{it2}), \quad (31)$$

for scaled coefficients. In first stage, (30) is estimated by regressing $y_{it2} = \log(\textit{spending})$ on time period (year) dummies, all exogenous variables, and time averages to get residuals, \hat{v}_{it2} . These residuals are included, along with time dummies, y_{it2} , \mathbf{z}_{it1} , and $\bar{\mathbf{z}}_i$ in a pooled fractional probit.

- The estimation is relatively fast, and bootstrapping is quite feasible.
- The following code is for the strictly exogenous case.

```

use meap92_01, clear
keep if year > 1994
cap drop alavgrexp alunch alenroll
egen double alavgrexp = mean(lavgrexp), by(distid)
egen double alunch = mean(lunch), by(distid)
egen double alenroll = mean(lenroll), by(distid)

capture program drop math4boot

program math4boot, rclass
    loc x "lavgrexp alavgrexp lunch alunch lenroll alenroll"
    glm math4 `x' y96-y01 if year > 1994, fa(bin) link(probit)
    predict double x1blhat, xb
    gen double scale=normalden(x1blhat)
    gen double pe1=scale*_b[lavgrexp]
    summarize pe1, meanonly
    return scalar ape1=r(mean)
    gen double pe2=scale*_b[lunch]
    summarize pe2, meanonly
    return scalar ape2=r(mean)
    gen double pe3=scale*_b[lenroll]
    summarize pe3, meanonly
    return scalar ape3=r(mean)
    drop x1blhat scale pe1 pe2 pe3
end

loc x "lavgrexp alavgrexp lunch alunch lenroll alenroll"
glm math4 `x' y96-y01 if year > 1994, fa(bin) link(probit)
mat b=e(b)

*Bootstrap SE by resampling districts, not observations
bootstrap r(ape1) r(ape2) r(ape3), reps(500) seed(123) cluster(distid): math4boot

```

Math Pass Rates: Spending Strictly Exogenous

Model:	Linear	Fractional Probit		Fractional Probit	
Estimation Method:	Fixed Effects	Pooled QMLE		GEE	
	Coefficient	Coefficient	APE	Coefficient	APE
<i>log(arexppp)</i>	.377 (.071)	.881 (.207)	.299 (.069)	.885 (.206)	.298 (.070)
<i>lunch</i>	-.042 (.073)	-.219 (.207)	-.074 (.067)	-.237 (.209)	-.080 (.067)
<i>log(enroll)</i>	.0021 (.0488)	.089 (.138)	.030 (.044)	.088 (.139)	.029 (.045)
Working Correlation	—	—		.491	
Scale Factor	—	.337		.337	

Math Pass Rates: Spending Endogenous

Model:	Linear	Fractional Probit	
Estimation Method:	Instrumental Variables	Pooled QMLE	
	Coefficient	Coefficient	APE
$\log(\text{arexppp})$.555 (.221)	1.731 (.759)	.583 (.255)
<i>lunch</i>	-.062 (.074)	-.298 (.202)	-.100 (.068)
$\log(\text{enroll})$.046 (.070)	.286 (.209)	.096 (.070)
$\hat{\nu}_2$	-.424 (.232)	-1.378 (.811)	—

- Simple dynamic model (for binary only):

$$P(y_{it} = 1 | \mathbf{z}_{it}, y_{i,t-1}, c_i) = \Phi(\mathbf{z}_{it}\boldsymbol{\delta} + \rho y_{i,t-1} + c_i). \quad (32)$$

A simple analysis is available if we specify

$$c_i | \mathbf{z}_i, y_{i0} \sim \text{Normal}(\psi + \xi_0 y_{i0} + \mathbf{z}_i \boldsymbol{\xi}, \sigma_a^2) \quad (33)$$

Then

$$\begin{aligned} P(y_{it} = 1 | \mathbf{z}_i, y_{i,t-1}, \dots, y_{i0}, a_i) = \\ \Phi(\mathbf{z}_{it}\boldsymbol{\delta} + \rho y_{i,t-1} + \psi + \xi_0 y_{i0} + \mathbf{z}_i \boldsymbol{\xi} + a_i), \end{aligned} \quad (34)$$

where $a_i \equiv c_i - \psi - \xi_0 y_{i0} - \mathbf{z}_i \boldsymbol{\xi}$.

- Turns out we can use standard random effects probit software (Stata: xtprobit), with explanatory variables $(1, \mathbf{z}_{it}, y_{i,t-1}, y_{i0}, \mathbf{z}_i)$ in time period t . Easily get the average partial effects, too:

$$\widehat{ASF}(\mathbf{z}_t, y_{t-1}) = N^{-1} \sum_{i=1}^N \Phi(\mathbf{z}_t \hat{\boldsymbol{\delta}}_a + \hat{\rho}_a y_{t-1} + \hat{\psi}_a + \hat{\xi}_{a0} y_{i0} + \mathbf{z}_i \hat{\boldsymbol{\xi}}_a), \quad (35)$$

with coefficients scaled by $(1 + \hat{\sigma}_a^2)^{-1/2}$.

- Labor force participation example with $N = 5,663$ and $T = 5$. The APE estimated from this method is about .259. If we ignore the heterogeneity, APE is .837.

```

use lfp, clear
tsset id period
* Lagged dependent variable:
bysort id (period): gen lfp_1 = L.lfp
* Put initial condition in periods 2-5:
by id: gen lfp1 = lfp[1]
* Create kids variables for periods 2-5:
forv i=2/5 {
by id: gen kids`i' = kids[`i']
}
* Create lhinc variables for periods 2-5:
forv i=2/5 {
by id: gen lhinc`i' = lhinc[`i']
}

capture program drop reprobit_boot
program reprobit_boot, rclass
xtset newid period
xtprobit lfp lfp_1 lfp1 kids kids2-kids5 lhinc lhinc2-lhinc5 educ black age agesq per3-per5, re from(b)
predict double xdh, xb
gen double xdh0 = xdh - _b[lfp_1]*lfp_1
gen double xdh1 = xdh0 + _b[lfp_1]
replace xdh0 = xdh0/sqrt(1 + e(sigma_u)^2)
replace xdh1 = xdh1/sqrt(1 + e(sigma_u)^2)
gen double pe1 = normal(xdh1) - normal(xdh0)
summarize pe1, meanonly
return scalar ap1=r(mean)
drop xdh xdh0 xdh1 pe1
end

xtprobit lfp lfp_1 lfp1 kids kids2-kids5 lhinc lhinc2-lhinc5 educ black age agesq per3-per5, re
mat b=e(b)

* Bootstrap within women.
tsset, clear
bootstrap r(ap1), reps(200) seed(123) cluster(id) idcluster(newid): reprobit_boot
tsset id period

program drop reprobit_boot

```

8.2. Corner Solution Responses

- If y_{it} (or y_{it1}) has a corner at zero, but is unbounded, or has two corners, one might want to apply Tobit. Analysis is very similar to probit. Use Chamberlain device and then either pooled Tobit (robust, APEs) or random effects Tobit (Stata: `xttobit`).
- If y_{it2} is a continuous endogenous explanatory variable, estimate its reduced form using the Chamberlain-Mundlak device, just as in (30). Then, put these residuals, \hat{v}_{it2} , in a pooled Tobit analysis.
- In other words, Tobit of y_{it1} on $y_{it2}, \mathbf{z}_{it1}, \bar{\mathbf{z}}_i, \hat{v}_{it2}$ pooled across i and t . Simple test of the null of exogeneity of y_{it2} as robust t statistic on \hat{v}_{it2} .
- Use Tobit conditional mean expressions, average across $(\bar{\mathbf{z}}_i, \hat{v}_{it2})$.

- Simple estimation of a panel hurdle model, such as Cragg's, is more controversial. In the strictly exogenous case, assume $D(y_{it}|\mathbf{x}_{it}, \bar{\mathbf{x}}_i)$ follows Cragg's model for each t [which does *not* follow from the same assumption for $D(y_{it}|\mathbf{x}_{it}, c_i)$, which is why it is somewhat controversial]. In the case with an endogenous variable, can assume $D(y_{it1}|y_{it2}, \mathbf{z}_{it1}, \bar{\mathbf{z}}_i, v_{it2})$ follows Cragg's model. Then, for example,

$$E(y_{it1}|y_{it2}, \mathbf{z}_{it1}, \bar{\mathbf{z}}_i, v_{it2}) = \Phi(\mathbf{w}_{it1}\boldsymbol{\gamma}_{10})[\mathbf{w}_{it1}\boldsymbol{\gamma}_{11} + \sigma_{11}\lambda(\mathbf{w}_{it1}\boldsymbol{\gamma}_{11}/\sigma_{11})], \quad (36)$$

where $\mathbf{w}_{it1} = (y_{it2}, \mathbf{z}_{it1}, \bar{\mathbf{z}}_i, v_{it2})$, $\boldsymbol{\gamma}_{10}$ are the parameters in the Cragg probit, and $(\boldsymbol{\gamma}_{11}, \sigma_{11})$ are the parameters from the truncated normal distribution; $\lambda(\cdot)$ is the inverse Mills ratio.

• First step is to obtain the \hat{v}_{it2} from the pooled OLS regression y_{it2} on $1, d2_t, \dots, dT_t, \mathbf{z}_{it}, \bar{\mathbf{z}}_i$ across i and t , and obtain the residuals, \hat{v}_{it2} (see (30)). Next, estimate a pooled Cragg model of y_{it1} on $1, d2_t, \dots, dT_t, y_{it2}, \mathbf{z}_{it1}, \bar{\mathbf{z}}_i, \hat{v}_{it2}$. Given the estimates from this two-step procedure, the partial effects are obtained from the estimated average structural function,

$$\begin{aligned} \widehat{ASF}(y_{t2}, \mathbf{z}_{t1}) = N^{-1} \sum_{i=1}^N \{ & \Phi(\hat{\eta}_{t0} + \hat{\alpha}_{10}y_{t2} + \mathbf{z}_{t1}\hat{\boldsymbol{\delta}}_{10} + \bar{\mathbf{z}}_i\hat{\boldsymbol{\xi}}_{10} + \hat{\rho}_{10}v_{it2}) \\ & \cdot [(\hat{\eta}_{t1} + \hat{\alpha}_{11}y_{t2} + \mathbf{z}_{it1}\hat{\boldsymbol{\delta}}_{11} + \bar{\mathbf{z}}_i\hat{\boldsymbol{\xi}}_{11} + \hat{\rho}_{11}v_{it2}) \\ & + \hat{\sigma}_{11}\lambda((\hat{\eta}_{t1} + \hat{\alpha}_{11}y_{t2} + \mathbf{z}_{t1}\hat{\boldsymbol{\delta}}_{11} + \bar{\mathbf{z}}_i\hat{\boldsymbol{\xi}}_{11} + \hat{\rho}_{11}v_{it2})/\hat{\sigma}_{11})] \}. \end{aligned} \quad (37)$$

- Take derivatives or changes with respect to elements of $(y_{t2}, \mathbf{z}_{t1})$.

Bootstrapping is very convenient for obtaining the standard errors.

- One defense of this approach: it specifies

$D(y_{it1}|y_{it2}, \mathbf{z}_i) = D(y_{it1}|y_{it2}, \mathbf{z}_{it1}, \bar{\mathbf{z}}_i, v_{it2})$. In other words, it uses a model for a distribution conditional on observables, and this model can be tested directly. (Not all of the identification assumptions can be, of course.)

9. Extensions

- With more work (programming!) can obtain useful extensions of basic models, for example, allow nonnormality or heteroskedasticity in the heterogeneity distribution:

$$c_i | \mathbf{x}_i \sim \text{Normal}(\psi + \bar{\mathbf{x}}_i \boldsymbol{\xi}, \sigma_a^2 \exp(\bar{\mathbf{x}}_i \boldsymbol{\omega})). \quad (38)$$

- Essentially, would lead to a command in Stata such as “xthetprob,” a random effects panel extension of “hetprob.”
- Or, allow “slope” heterogeneity, such as

$$P(y_{it} = 1 | \mathbf{x}_{it}, c_i, \mathbf{b}_i) = \Phi(\mathbf{x}_{it} \mathbf{b}_i + c_i). \quad (39)$$

- Typically, can identify average partial effects.