BINARY RESPONSE MODELS

Econometric Analysis of Cross Section and Panel Data, 2e
MIT Press
Jeffrey M. Wooldridge

1. Introduction
2. Linear Probability Model
3. Index Models: Probit and Logit
4. Endogenous Explanatory Variables
5. Panel Data Models
6. Multivariate Probit

1. INTRODUCTION

- The most common application of binary response models is when we are interested in "explaining" a binary outcome in terms of some explanatory variables. Thus, we are interested in a conditional probability.
- A less common application is when we have a linear model of an underlying quantitative variable, but the data collection scheme censors the data. For example, we have a linear model for willingness to pay for a project or product. However, because it is difficult to elicit WTP, each individual may be presented with a cost of the project; we then only observe whether they are in favor of the project at that cost.

• We treat data censoring problems later. For now, we focus on the first situation. So, y is a binary (zero-one) variable. For example, $y = employed$ or $y = arrested$. Given a set of (exogenous) covariates $\mathbf{x}$, we are interested in

$$P(y = 1|\mathbf{x}) = p(\mathbf{x}),$$

which is called the *response probability*. It is the probability of a "success," that is, $y = 1$.

• As in regression, we are interested in the partial effects of the x_j on $p(\mathbf{x})$. For continuous x_j, these are usually

$$\frac{\partial p(\mathbf{x})}{\partial x_j}.$$

- For discrete x_j, look at changes in the response probability (usually holding other variables fixed). For example, if $x_K = train$ (job training indicator) and y is an employment indicator,

$$p(x_1,\ldots,x_{K-1},1) - p(x_1,\ldots,x_{K-1},0)$$

is the effect of job training on the employment probability, at given values for the other covariates.

- In nonlinear models generally, and binary response models specifically, it is often useful to have a single number to summarize the relationship between $P(y = 1|\mathbf{x})$ and x_j. In a linear model that is simply the coefficient.

• Generally, we might report an estimated *average partial effect (APE)*. The APE for a continuous x_j is

$$E_{\mathbf{x}}\left[\frac{\partial p(\mathbf{x})}{\partial x_j}\right],$$

which means we average the partial effect across the population distribution of $\mathbf{x}$. This is a weighted average of the partial effects at each outcome $\mathbf{x}$.

• Suppose x_K is a binary variable. Then its APE is

$$E_{\mathbf{x}_{(K)}}[p(\mathbf{x}_{(K)},1) - p(\mathbf{x}_{(K)},0)]$$

where $\mathbf{x}_{(K)}$ is the $1 \times K$ vector with x_K excluded.

• Another partial effect that has been reported in empirical work is the *partial effect at the average (PEA)*. For a continous variable x_j,

$$\frac{\partial p(\boldsymbol{\mu}_{\mathbf{x}})}{\partial x_j}.$$

• In nonlinear models, the APE and PEA can be very different: the expected value does not pass through nonlinear functions.

• Because $\boldsymbol{\mu}_{\mathbf{x}}$ might not even represent a population unit – for example, if $\mathbf{x}$ includes discrete variables, such as dummy variables – the PEA might not be especially interesting.

• Some simple, useful facts about Bernoulli (zero-one) random variables are

$$
\begin{aligned}
E(y|\mathbf{x}) &= P(y = 1|\mathbf{x}) = p(\mathbf{x}) \\
Var(y|\mathbf{x}) &= p(\mathbf{x})[1 - p(\mathbf{x})]
\end{aligned}
$$

• So a binary variable has natural heteroskedasticity except in the special case where $p(\mathbf{x})$ does not depend on $\mathbf{x}$.

• Unlike variables that take on more than two values, there is a necessary link between the mean and the variance. It is not possible for $E(y|\mathbf{x}) = p(\mathbf{x})$ while $Var(y|\mathbf{x}) \neq p(\mathbf{x})[1 - p(\mathbf{x})]$. (If, say, y is a takes values in $\{0, 1, 2, \ldots\}$, $Var(y|\mathbf{x})$ need not be related to $E(y|\mathbf{x})$, even though that is true for popular distributions such as the Poisson.)

2. THE LINEAR PROBABILITY MODEL

• The linear probability model (LPM) models the response probability as a function linear in parameters. Absorbing an intercept into **x**, if we take the model literally we are assuming

$$P(y = 1|\mathbf{x}) = \beta_1 + \beta_2 x_2 + \ldots + \beta_K x_K \equiv \mathbf{x}\boldsymbol{\beta}.$$

Because this is also $E(y|\mathbf{x})$, we can use OLS to consistently estimate $\boldsymbol{\beta}$. In fact, if the conditional mean is truly $\mathbf{x}\boldsymbol{\beta}$, the OLS estimator is unbiased.

- Because $Var(y|\mathbf{x}) = \mathbf{x}\boldsymbol{\beta}(1 - \mathbf{x}\boldsymbol{\beta})$ – a rare case where we know the functional form of heteroskedasticity – inference for OLS should be made robust to heteroskedasticity. As we know, this is easy to do.
- Because y is binary, we must rely on large-sample properties for inference; clearly normality of $D(y|\mathbf{x})$ does not hold.
- The LPM is always a good starting point when y is the variable we hope to explain. The estimated coefficients give direct estimates of the effects of each x_j on the response probability. (Of course, as with any regression framework, we can include various functional forms in $\mathbf{x}$, such as quadratics, interactions, and dummy variables.)

- The LPM is simple to estimate and interpret. The often cited drawbacks of the LPM include

(1) Nothing guarantees the OLS fitted values, $\hat{y}_i = \mathbf{x}_i\hat{\boldsymbol{\beta}}$, are in the unit interval. As these are estimates of the $p(\mathbf{x}_i)$, one might worry about estimated probabilites above one or negative. (In practice, this is a minor issue.)

(2) While we can use various functional forms in $\mathbf{x}$, it is difficult to impose, in a simple way, diminishing effects of the x_j on the $p(\mathbf{x})$. For example, if $\beta_j > 0$, increasing x_j in increases $p(\mathbf{x}) = \mathbf{x}\boldsymbol{\beta}$ by β_j, no matter the values of x_j or the other elements of $\mathbf{x}$. Logically, the effect must diminish at some point.

(3) Heteroskedasticity. This has asymptotic efficiency implications *if* we assume that $p(\mathbf{x}) = \mathbf{x}\boldsymbol{\beta}$. That is, in principle we can improve efficiency by weighted least squares, but $\mathbf{x}_i\hat{\boldsymbol{\beta}}$ not strict between zero and one for all i causes problems because the efficient weights are supposed to be $1/[\mathbf{x}_i\hat{\boldsymbol{\beta}}(1 - \mathbf{x}_i\hat{\boldsymbol{\beta}})]$.

- WLS hardly seems worth it because we can use the usual heteroskedasticity-robust inference for OLS without worrying about adjusting the fitted values.

- As a practical matter, it makes more sense to think of the LPM as the best linear approximation (in a mean squared error sense) to the true response probability, $p(\mathbf{x})$. That is,

$$
\begin{aligned}
y &= \mathbf{x}\boldsymbol{\beta} + u \\
E(\mathbf{x}'u) &= \mathbf{0}
\end{aligned}
$$

is all we are willing to assume. If so, then $E(u^2|\mathbf{x})$ generally depends on $p(\mathbf{x})$ in addition to $\mathbf{x}\boldsymbol{\beta}$, but the heteroskedasticity-robust variance matrix estimator is still valid (because it is valid for heteroskedasticity of unknown form).

- A carefully chosen linear model can yield good estimates of the APEs defined earlier. In other words, the LPM often yields good estimates of *average* effects.
- A leading reason for going from the LPM to nonlinear models of $p(\mathbf{x})$ is to allow the partial effects to vary across different values of $\mathbf{x}$.
- When we view the LPM as a linear projection, weighted least squares – even if all fitted values are in $(0,1)$ – is not even consistent for the parameters of the linear projection $L(y|\mathbf{x}) = \mathbf{x}\boldsymbol{\beta}$. (The parameters identified by WLS are necessarily less interesting than those in the linear projection, but they are different.)

3. INDEX MODELS: PROBIT AND LOGIT

- A general index model has the form

$$P(y = 1|\mathbf{x}) = G(\mathbf{x}\boldsymbol{\beta})$$

for some $G : \mathbb{R} \to (0,1)$. That is, $0 < G(\bullet) < 1$. In most cases, $G(\bullet)$ is actually a cumulative distribution function for a continuous random variable with density $g(\bullet)$. Then, $G(\bullet)$ is strictly increasing, and the estimates are easier to interpret.

- The leading cases are $G(z) = \Phi(z)$ (probit) and $G(z) = \exp(z)/[1 + \exp(z)]$ (logit).

- MLE is straightforward. The general log likelihood for random draw i is

$$\ell_i(\boldsymbol{\beta}) = (1 - y_i)\log[1 - G(\mathbf{x}_i\boldsymbol{\beta})] + y_i\log[G(\mathbf{x}_i\boldsymbol{\beta})].$$

- Asymptotic variance has the same form as for probit:

$$\left(\sum_{i=1}^{N}\frac{[g(\mathbf{x}_i\hat{\boldsymbol{\beta}})]^2\mathbf{x}_i'\mathbf{x}_i}{G(\mathbf{x}_i\hat{\boldsymbol{\beta}})[1 - G(\mathbf{x}_i\hat{\boldsymbol{\beta}})]}\right)^{-1},$$

where

$$g(z) = \phi(z) \text{ for probit}$$
$$g(z) = \exp(z)/[1 + \exp(z)]^2 \text{ for logit}$$

- Testing multiple hypotheses about $\boldsymbol{\beta}$ (we drop the "o" subscript for simplicity) – usually joint exclusion restrictions – is most easily done with the Wald and LR statistics. The former is commonly used in canned packages (in Stata, it is computed with the "test" command), and the LR statistic is easily obtained because the value of the log likelihood is reported routinely.
- The score statistics is convenient for testing the standard index models against more complicated alternatives (below).

Estimating Partial Effects

- More interesting is: What do we do with the estimates? Let x_j be continuous. Then

$$\frac{\partial p(\mathbf{x})}{\partial x_j} = \beta_j g(\mathbf{x}\boldsymbol{\beta})$$

and, because $g(z) > 0$ (assume it is a continuous density), β_j gives the direction of the partial effect. But its magnitude depends on $g(\mathbf{x}\boldsymbol{\beta})$.

- For probit, the largest value of the scale factor is about $.4 = g(0)$. For logit, it is $.25$.

• For two continous covariates, the ratio of the coefficients give the ratio of the partial effects, independent of $\mathbf{x}$.

$$\frac{\partial p(\mathbf{x})/\partial x_j}{\partial p(\mathbf{x})/\partial x_h} = \frac{\beta_j g(\mathbf{x}\boldsymbol{\beta})}{\beta_h g(\mathbf{x}\boldsymbol{\beta})} = \beta_j/\beta_h.$$

• No simple relationship exists for discrete variables or changes.

• In any case, we would like the magnitude of the effect.

• Two common summary measures are the estimated PEAs and APEs. The estimated PEA for a continuous variable is

$$\widehat{PEA}_j = \hat{\beta}_j g(\bar{\mathbf{x}}\hat{\boldsymbol{\beta}})$$

• As discussed earlier, putting in averages for discrete covariates might not be especially interesting.

• When $\mathbf{x}$ includes nonlinear functions, such as age^2, probably makes more sense to use $(\overline{age})^2$ rather than average age_i^2.

• Delta method or bootstrapping can be used to get a standard error for $\widehat{PEA}_j$.

• The APE has more appeal, as we are averaging partial effects for actual units:

$$\widehat{APE}_j = \hat{\beta}_j \left[N^{-1} \sum_{i=1}^{N} g(\mathbf{x}_i \hat{\boldsymbol{\beta}}) \right]$$

• To use the delta method, must account for randomness in $\mathbf{x}_i$, too. Bootstrap makes that easy.

• Whether we use the PEA or APE, the scale factor multiplying $\hat{\beta}_j$ is below one, and sometimes well below one.

- It makes no sense to compare magnitudes of coefficients across probit, logit, LPM. Comparing APEs is preferred.
- In particular, if $\hat{\gamma}_j$ is the linear regression coefficient on x_j from estimating an LPM, it can be compared with $\widehat{APE}_j$ (provided no other function of x_j appears in the regressors).

• Suppose x_K is a binary variable. Then its APE is estimated as

$$\widehat{APE}_K = N^{-1} \sum_{i=1}^{N} [G(\mathbf{x}_{i(K)} \hat{\boldsymbol{\beta}}_{(K)} + \hat{\beta}_K) - G(\mathbf{x}_{i(K)} \hat{\boldsymbol{\beta}}_{(K)})],$$

where $\mathbf{x}_{i(K)}$ is $\mathbf{x}_i$ but without x_{iK}.

• The APE has a nice counterfactual interpretation that is especially useful in policy analysis. Called the *average treatment effect (ATE)* in the treatment effect literature with a binary outcome. (The "treatment," x_K, is binary.)

• Can average the individual treatment effects across subgroups, too, or insert fixed values for some of the other covariates.

• Stata, with its “margins” (marginal effects) command can report at PEA or APE. For a discrete x_K, the estimated PEA is

$$\widehat{PEA}_K = G(\bar{\mathbf{x}}_{(K)}\hat{\boldsymbol{\beta}}_{(K)} + \hat{\beta}_K) - G(\bar{\mathbf{x}}_{(K)}\hat{\boldsymbol{\beta}}_{(K)})$$

Again, this might correspond to a weird population unit, or might not be representative of the population.

• To obtain standard errors of APEs and PEAs, we can use the delta method or bootstrap.

• Stata uses the delta method to obtain standard errors.

• Complicated functional forms are, in principle, easily handled within the index structure. For example, suppose

$$P(y = 1|\mathbf{z}) = G[\beta_0 + \beta_1 z_1 + \beta_2 z_1^2 + \beta_3 \log(z_2) + \beta_4 z_3] \equiv G(\mathbf{x}\boldsymbol{\beta})$$

Then

$$\frac{\partial P(y = 1|\mathbf{z})}{\partial z_1} = (\beta_1 + 2\beta_2 z_1) g(\mathbf{x}\boldsymbol{\beta})$$

$$\frac{\partial P(y = 1|\mathbf{z})}{\partial z_2} = (\beta_3 / z_2) g(\mathbf{x}\boldsymbol{\beta})$$

$$\frac{\partial \log P(y = 1|\mathbf{z})}{\partial \log z_2} = \beta_3 g(\mathbf{x}\boldsymbol{\beta}) / G(\mathbf{x}\boldsymbol{\beta})$$

• The signs of the coefficients are informative, but the partial effects are somewhat complicated. Need to evaluate them at interesting values or average across the distribution of **x** similar to the usual APE calculation.

• For example, the average elasticity of $P(y = 1|\mathbf{z})$ with respect to z_2 is

$$\hat{\beta}_3 \left[N^{-1} \sum_{i=1}^{N} g(\mathbf{x}_i \hat{\boldsymbol{\beta}}) / G(\mathbf{x}_i \hat{\boldsymbol{\beta}}) \right].$$

Goodness of Fit

- In addition to reporting coefficients, standard errors, partial effects, and their standard errors, some additional goodness-of-fit measures are sometimes reported.
- Define, for each i, a binary predictor

$$\begin{aligned} \tilde{y}_i &= 1 \text{ if } G(\mathbf{x}_i\hat{\boldsymbol{\beta}}) \geq .5 \\ &= 0 \text{ if } G(\mathbf{x}_i\hat{\boldsymbol{\beta}}) < .5 \end{aligned}$$

- We make a correct prediction if $y_i = 0$ and $\tilde{y}_i = 0$ or $\tilde{y}_i = 1$ and $y_i = 1$. Let N_0 be the number of observations with $y_i = 0$ and N_1 the number with $y_i = 1$, so that $N = N_0 + N_1$.

• We can compute the percent correctly predicted for each of the outcomes, and the overall percent correctly predicted. If N_{00} is the number of observations with $y_i = 0$ and $\tilde{y}_i = 0$ and N_{11} is the number of observations with $\tilde{y}_i = 1$ and $y_i = 1$, then the proportions correctly predicted are

$$q_0 = \frac{N_{00}}{N_0}, q_1 = \frac{N_{11}}{N_1}.$$

• If one of q_0 or q_1 seems "too small," the prediction threshold can be chosen to be different from $.5$.

• For example, some suggest using the fraction of "successes," $\bar{y}$, as the threshold. With random sampling, $\bar{y}$ is a consistent estimator of the unconditional probability of success, $P(y_i = 1)$.

• So, the idea is to predict one if the estimated conditional probability of success exceeds the unconditional probability. (Of course, changing the threshold increases the proportion correctly predicted for one outcome but generally decreases the proportion for the other outcome.)

• The overall proportion correctly predicted is

$$q = \frac{(N_{00} + N_{11})}{N} = \left(\frac{N_0}{N}\right)q_0 + \left(\frac{N_1}{N}\right)q_1,$$

which is a weighted average of the two.

• Whether we use an R-squared or the percent correctly predicted to summarize goodness of fit, it is not necessary to have a "good" fit in order for the estimated partial effects to be useful. For example, we might be able to get a good estimate of the average effect of job training on the probability of employment even though we cannot predict with much accuracy whether a particular person in an at-risk group becomes employed.

- Because the Kullback-Leibler information criterion is maximized for the true density, the values of the log likelihoods can be used to choose among different nonnested models. In practice, it might be difficult to choose between, say, logit and probit. (Often the differences are practically unimportant, although they can be when fitted values at the extreme tails are important.)

Specification Issues and Testing

• There is much confusion about specification issues in probit, logit, and other models, because sometimes inappropriate parallels are made with linear models.

• Probit is easiest to discuss because analytical results are available.

Omitted Variable Independent of Covariates

• Consider first the problem of an omitted variable independent of **x**, call it c:

$$P(y = 1|\mathbf{x}, c) = \Phi(\mathbf{x}\boldsymbol{\beta} + c)$$
$$c|\mathbf{x} \sim Normal(0, \sigma_c^2)$$

where **x** includes unity so $E(c) = 0$ is without loss of generality.

- Write the underlying latent variable as $y^* = \mathbf{x}\boldsymbol{\beta} + c + e$, $(c + e)|\mathbf{x} \sim Normal(0, \sigma_c^2 + 1)$. So

$$P(y = 1|\mathbf{x}) = \Phi[\mathbf{x}\boldsymbol{\beta}/(1 + \sigma_c^2)^{1/2}].$$

- It follows immediately that probit of y_i on $\mathbf{x}_i$ consistently estimates $\boldsymbol{\beta}_c \equiv \boldsymbol{\beta}/(1 + \sigma_c^2)^{1/2}$.
- That $\boldsymbol{\beta}_c$ is attenuated toward zero has been called "attenuation bias." This would not happen in a linear model. Question: Is it truly a "problem"?

• Answer: Not really. The scaled coefficients give directions of effects and relative effects just as well as the original parameters.

• For magnitudes, the β_j index the PEAs at the average value of c, $E(c) = 0$:

$$\frac{\partial P(y|\mathbf{x}, c = 0)}{\partial x_j} = \beta_j \phi(\mathbf{x}\boldsymbol{\beta}).$$

So the PEAs at $c = 0$ (or any other value of c) are not identified.

• But the APE is identified. Can show that

$$E_c\left[\frac{\partial P(y|\mathbf{x}, c)}{\partial x_j}\right] = \beta_{cj}\phi(\mathbf{x}\boldsymbol{\beta}_c)$$

- So, in fact, the scaled coefficients – which we consistently estimate – index a quantity that is of significant interest.
- More generally, in any model, if c is independent of $\mathbf{x}$, just estimating $P(y = 1|\mathbf{x})$ consistently estimates the APEs (as a function of $\mathbf{x}$). But, of course, we could not estimate the heterogeneity distribution.
- Of course, if c is correlated with $\mathbf{x}$, a much different story (later).

Heteroskedasticity in the Latent Variable Model

- Again suppose y is the variable of interest, and now we allow heteroskedasticity in the error e in

$$y = 1[\mathbf{x}\boldsymbol{\beta} + e > 0].$$

- Suppose we assume

$$e|\mathbf{x} \sim Normal(0, \exp(2\mathbf{x}_1\boldsymbol{\delta})),$$

where $\mathbf{x}_1$ is a subset of $\mathbf{x}$ (and does not include a constant). So homoskedasticity is $\boldsymbol{\delta} = 0$ and then e has unit variance.

• Clearly the introduction of heteroskedasticity in e changes the response probability, $P(y = 1|\mathbf{x})$. In fact,

$$
\begin{aligned}
P(y = 1|\mathbf{x}) &= P(e > -\mathbf{x}\boldsymbol{\beta}|\mathbf{x}) = P[\exp(-\mathbf{x}_1\boldsymbol{\delta})e > -\exp(-\mathbf{x}_1\boldsymbol{\delta})\mathbf{x}\boldsymbol{\beta}|\mathbf{x}] \\
&= 1 - \Phi[-\exp(-\mathbf{x}_1\boldsymbol{\delta})\mathbf{x}\boldsymbol{\beta}] = \Phi[\exp(-\mathbf{x}_1\boldsymbol{\delta})\mathbf{x}\boldsymbol{\beta}].
\end{aligned}
$$

• Estimation by Bernoulli MLE, as before.

• Now, the derivatives and changes in $P(y = 1|\mathbf{x})$ are much more complicated, and need not have the same sign as the relevant coefficient.

• If we view $P(y = 1|\mathbf{x}) = \Phi[\exp(-\mathbf{x}_1\boldsymbol{\delta})\mathbf{x}\boldsymbol{\beta}]$ as just a way to generalize functional form, partial effects should be computed.

- Of course, it may be sufficient to include covariates in a flexible way in probit and logit.
- After estimation of, say, probit with squares and interactions, it is legitimate to compare log likelihood with the heteroskedastic probit log likelihood.
- Generally, heteroskedastic probit and probit with flexible polynomials are nonnested. Can use Vuong's (1989, *Econometrica*) model selection test.

• If we truly believe the index structure with e heteroskedastic, there is a different way to proceed. Define the *average structural function* as a function of $\mathbf{x}$:

$$ASF(\mathbf{x}) = E_e\{1[\mathbf{x}\boldsymbol{\beta} + e > 0]\} = 1 - F(-\mathbf{x}\boldsymbol{\beta})$$

where $F(\bullet)$ is the unconditional distribution of e.

• Let $\mathbf{x}_{i1}$ denote the random quantity. Then we can use the law of iterated expectations to show

$$ASF(\mathbf{x}) = E_{\mathbf{x}_{i1}}\{\Phi[\exp(-\mathbf{x}_{i1}\boldsymbol{\delta})\mathbf{x}\boldsymbol{\beta}]\}$$

and a consistent estimator is

$$\widehat{ASF}(\mathbf{x}) = N^{-1}\sum_{i=1}^{N}\Phi[\exp(-\mathbf{x}_{i1}\hat{\boldsymbol{\delta}})\mathbf{x}\hat{\boldsymbol{\beta}}].$$

• The estimated average partial effect, for a continuous x_j, is

$$\widehat{APE_j}(\mathbf{x}) = \hat{\beta}_j \left\{ N^{-1} \sum_{i=1}^{N} \exp(-\mathbf{x}_{i1}\hat{\boldsymbol{\delta}})\phi[\exp(-\mathbf{x}_{i1}\hat{\boldsymbol{\delta}})\mathbf{x}\hat{\boldsymbol{\beta}}] \right\}$$

which is the same sign as $\hat{\beta}_j$ because the term in $\{\bullet\}$ is strictly positive.

• Of course, ignoring heteroskedasticity in e does generally lead to inconsistent estimators of the β_j, but that is largely beside the point. The important question is: how far off are estimated partial effects?

- Possible point of confusion: using the "robust" option with probit does not mean the probit estimators of $\boldsymbol{\beta}$ somehow robust to heteroskedasticity in the latent error. In fact, $\boldsymbol{\beta}$ will be inconsistently estimated (but the MLE is still of value, providing the "best" approximation).
- Using "robust" means that a sandwich estimator is used for the asymptotic variance of the quasi-MLE (that is, the usual probit estimator).

• Remember, allowing heteroskedasticity in e with $y = 1[\mathbf{x}\boldsymbol{\beta} + e > 0]$ changes $P(y = 1|\mathbf{x})$, which completely describes $D(y|\mathbf{x})$. This is not like other regression applications where we can have $E(y|\mathbf{x}) = \mathbf{x}\boldsymbol{\beta}$ and separately talk about heteroskedasticity in $Var(y|\mathbf{x})$.

• Testing the probit model against a heteroskedastic alternative is a good functional form test. The score test is convenient because it only requires estimation of the probit model. A variable addition test is convenient, too. After the initial probit to get the estimated linear indices, $\mathbf{x}_i\hat{\boldsymbol{\beta}}$, do probit of

$$y_{i1} \text{ on } \mathbf{x}_i, (\mathbf{x}_i\hat{\boldsymbol{\beta}})^2\mathbf{x}_{i1}$$

and use a joint test Wald test on $(\mathbf{x}_i\hat{\boldsymbol{\beta}})^2\mathbf{x}_{i1}$. The degrees-of-freedom in the χ^2 distribution equals the dimension of $\mathbf{x}_{i1} \subseteq \mathbf{x}_i$.

• Another functional form test is like the RESET from regression. For example, after probit, do an expanded probit of

$$y_i \text{ on } \mathbf{x}_i, (\mathbf{x}_i\hat{\boldsymbol{\beta}})^2, (\mathbf{x}_i\hat{\boldsymbol{\beta}})^3$$

and test the last two terms for joint significance using a Wald test. (Some think it is best to add $(\mathbf{x}_i\hat{\boldsymbol{\beta}})^4$, but the expanded test need not have more power.)

• If you want to proceed with the heteroskedastic probit model, the command is "hetprob" in Stata.

Nonnormality in the Latent Variable Model

- Again, consider

$$y = 1[\mathbf{x}\boldsymbol{\beta} + e > 0]$$

where e is independent of $\mathbf{x}$ but not normally distributed. What if we apply probit? Not surprisingly, the probit MLE is not consistent for $\boldsymbol{\beta}$ if e is not normal. But the partial effects are often very close, at least over the range of $\mathbf{x}$ where we can have some confidence in the estimated partial effects. (For example, logit and probit can give similar partial effects except in the extreme tails of the distribution.)

- The key is that we should focus on partial effects and not just parameters.

The Linear Probability Model, Revisited

- Now write an index model with the "intercept" shown explicitly,

$$P(y = 1|\mathbf{x}) = G(\alpha + \mathbf{x}\boldsymbol{\beta})$$

where $\mathbf{x}$ is a continuous random vector. Define the APE for x_j as

$$\beta_j E[g(\alpha + \mathbf{x}\boldsymbol{\beta})].$$

- Let η and $\boldsymbol{\gamma}$ be the linear projection parameters,

$$L(y|1, \mathbf{x}) = \eta + \mathbf{x}\boldsymbol{\gamma}$$

• Can show that if $\mathbf{x}$ is multivariate normal then

$$\gamma_j = \beta_j E[g(\alpha + \mathbf{x}\boldsymbol{\beta})], j = 1, \ldots, K.$$

In other words, estimating an LPM consistently estimates the APEs.

• Multivariate normality is restrictive, but suggests that OLS on the LPM might get close to the APEs more generally.

• Of course, we miss out on some of the richness of nonlinear binary response models by focusing only on the APEs.

4. ENDOGENOUS EXPLANATORY VARIABLES

- For nonlinear binary response models, the nature of the endogenous explanatory variable(s) plays a role in estimation. In principle, one can use joint maximum likelihood. But specifying the joint distribution can be tricky, and the methods generally require the distributional assumptions to hold for consistency.
- In some cases, control function (CF) methods are available. CF methods are useful for testing, too.
- Sometimes plug-in methods produce consistent estimators of scaled coefficients, but in many cases they do not. With random samples, CF methods are usually preferred to plugging in fitted values.

• Because of the limitations of nonlinear models, some have proposed using linear models and applying standard IV estimation methods. Recall the linear model

$$
\begin{aligned}
y_1 &= \alpha_1 y_2 + \mathbf{z}_1 \boldsymbol{\delta}_1 + u_1 \\
L(y_2|\mathbf{z}) &= \mathbf{z}\boldsymbol{\delta}_2 = \mathbf{z}_1 \boldsymbol{\delta}_{21} + \mathbf{z}_2 \boldsymbol{\delta}_{22}
\end{aligned}
$$

with $\boldsymbol{\delta}_{22} \neq \mathbf{0}$. We can apply this with y_1 binary as an approximation. No special restrictions are needed on y_2 to apply 2SLS. y_2 can be continuous, binary, count, and so on.

• Some simulations show that the average partial effects can be estimated pretty well by 2SLS.

Continuous EEV

- If we want to allow nonconstant partial effects, we need to turn to nonlinear models.

- With a single EEV (for simplicity), consider the model

$$y_1 = 1[\alpha_1 y_2 + \mathbf{z}_1\boldsymbol{\delta}_1 + u_1 > 0]$$
$$u_1|\mathbf{z} \sim Normal(0,1)$$

where $\mathbf{z}$ is the vector of all endogenous variables. Analysis goes through if we replace $(\mathbf{z}_1, y_2)$ with any known function $\mathbf{x}_1 \equiv \mathbf{g}_1(\mathbf{z}_1, y_2)$.

- The parameters $(\alpha_1, \boldsymbol{\delta}_1)$ index the average structural function, and so they index the APEs, too.

• The Rivers-Vuong (1988) approach is to make a homoskedastic-normal assumption on the reduced form for y_2,

$$y_2 = \mathbf{z}\boldsymbol{\delta}_2 + v_2 = \mathbf{z}_1\boldsymbol{\delta}_{21} + \mathbf{z}_2\boldsymbol{\delta}_{22} + v_2,\ \boldsymbol{\delta}_{22} \neq \mathbf{0}$$

$$v_2|\mathbf{z} \sim Normal(0, \tau_2^2)$$

• Can relax normality in two-step methods. In fact, sufficient is

$$u_1 = \theta_1 v_2 + e_1$$

$$e_1|v_2, \mathbf{z} \sim Normal(0, 1 - \theta_1^2\tau_2^2)$$

- The CF approach is a two-step method. Write

$$y_1 = 1[\alpha_1 y_2 + \mathbf{z}_1\boldsymbol{\delta}_1 + \theta_1 v_2 + e_1 > 0]$$

so that

$$P(y_1 = 1|y_2, \mathbf{z}) = \Phi(\alpha_{\rho 1} y_2 + \mathbf{z}_1\boldsymbol{\delta}_{\rho 1} + \theta_{\rho 1} v_2),$$

where each coefficient is multiplied by $(1 - \rho_1^2)^{-1/2}$ and $\rho_1 = \theta_1\tau_2 = Corr(v_2, u_1)$. The scaled coefficients are identified because we effectively observe $v_2 = y_2 - \mathbf{z}\boldsymbol{\delta}_2$.

- The RV two-step approach is

(i) OLS of y_2 on $\mathbf{z}$, to obtain the residuals, $\hat{v}_2$.

(ii) Probit of y_1 on $\mathbf{z}_1, y_2, \hat{v}_2$ to estimate the scaled coefficients. A simple t test on $\hat{v}_2$ is valid to test $H_0 : \theta_1 = 0$.

- The original coefficients, which appear in the partial effects, are easily obtained from the set of two-step estimates:

$$\hat{\boldsymbol{\beta}}_1 = \hat{\boldsymbol{\beta}}_{\rho 1}/(1 + \hat{\theta}_{\rho 1}^2 \hat{\tau}_2^2)^{1/2}$$

- Notice that the two-step estimates are larger than the unscaled coefficients.
- Bootstrapping is convenient for standard errors; also for APEs, such as

$$\hat{\alpha}_1 \phi(\hat{\alpha}_1 y_2 + \mathbf{z}_1 \hat{\boldsymbol{\delta}}_1)$$

- The APE for y_2 across the entire population is then estimated as

$$\hat{\alpha}_1 \left[N^{-1} \sum_{i=1}^{N} \phi(\hat{\alpha}_1 y_{i2} + \mathbf{z}_{i1} \hat{\boldsymbol{\delta}}_1) \right]$$

• Alternatively, we average out the reduced form residuals using the scaled coefficients:

$$\widehat{ASF}(\mathbf{z}_1, y_2) = N^{-1} \sum_{i=1}^{N} \Phi(\mathbf{x}_1 \hat{\boldsymbol{\beta}}_{\rho 1} + \hat{\theta}_{\rho 1} \hat{v}_{i2})$$

and take derivatives or changes with respect to the elements $(y_2, \mathbf{z}_1)$, even if $\mathbf{x}_1$ is nonlinear functions of them. This formulation is useful for more complicated models.

• If instead of adding RF residuals we replace y_2 with $\hat{y}_2$, the two-step procedure consistently estimates a different set of scaled parameters in the basic model. With random sampling, it has little to offer over the CF, and does not work if, say, y_2^2 or $y_2\mathbf{z}$ appear in the model.

- If we make the stronger assumption

$$(u_1, v_2)|\mathbf{z} \sim BivariateNormal$$

with $\rho_1 = Corr(u_1, v_2)$, then we can proceed with MLE based on $f(y_1, y_2|\mathbf{z}) = f(y_1|y_2, \mathbf{z})f(y_2|\mathbf{z})$.

- The distribution $f(y_2|\mathbf{z})$ is straightforward because it is homoskedastic normal with a linear conditional mean.

• For $f(y_1|y_2,\mathbf{z})$ we have, for example,

$$P(y_1 = 1|y_2,\mathbf{z}) = \Phi\left[\frac{\alpha_1 y_2 + \mathbf{z}_1\boldsymbol{\delta}_1 + (\rho_1/\tau_2)(y_2 - \mathbf{z}\boldsymbol{\delta}_2)}{(1-\rho_1^2)^{1/2}}\right]$$

and then $P(y_1 = 0|y_2,\mathbf{z})$ is immediate. Then, all parameters – $\alpha_1, \boldsymbol{\delta}_1, \rho_1, \boldsymbol{\delta}_2, \tau_2$ are estimated jointly by MLE conditional on $\mathbf{z}$.

• The Stata command is “ivprobit.” The same sorts of goodness-of-fit measures and partial effects are available, of course. For APEs, still might want to bootstrap the standard errors, confidence intervals.

Binary EEV

- What if y_2 is not continuous? No generally useful two-step methods are available when discreteness in y_2 is important. The CF approach above – and even more recent nonparametric approaches by Blundell and Powell – hinges on being able to write

$$y_2 = g_2(\mathbf{z}) + v_2$$

where

$$v_2 \text{ is independent of } \mathbf{z}.$$

• If, say, y_2 is binary, this representation does not exist. Generally, the natural choice for $g_2(\mathbf{z})$ is $E(y_2|\mathbf{z})$. But when y_2 is discrete, $v_2 = y - E(y_2|\mathbf{z})$ usually depends on $\mathbf{z}$ in higher moments, such as the variance.

• When y_2 is binary, the support of v_2 conditional on $\mathbf{z}$ is just the two points $\{-g_2(\mathbf{z}), 1 - g_2(\mathbf{z})\}$, and so v_2 and $\mathbf{z}$ are clearly not independent.

- Somewhat radical suggestion: First, standardize y_2 as

$$r_2 = \frac{[y_2 - E(y_2|\mathbf{z})]}{sd(y_2|\mathbf{z})},$$

so that $E(r_2|\mathbf{z}) = 0$, $Var(r_2|\mathbf{z}) = 1$. Then, just *assume* that

$$D(u_1|y_2, \mathbf{z}) = D(u_1|r_2)$$

that is, $D(u_1|y_2, \mathbf{z})$ depends on $(y_2, \mathbf{z})$ only through the standardized error r_2.

• Could use standardized residuals $\hat{r}_{i2} = [y_{i2} - \hat{E}(y_{i2}|\mathbf{z}_i)]/\widehat{sd}(y_{i2}|\mathbf{z}_i)$ in a control function approach.

• This is “radical” because it does not follow from standard assumptions, such as joint normality of (u_1, v_2).

• Some methods exist for estimating parameters up to an uknown (but common) scale, but they often require special assumptions and do not deliver magnitudes of effect.

- Generally available approach: MLE. Assume (y_1, y_2) are generated as

$$\begin{aligned} y_1 &= 1[\alpha_1 y_2 + \mathbf{z}_1 \boldsymbol{\delta}_1 + u_1 > 0] \\ y_2 &= 1[\mathbf{z} \boldsymbol{\delta}_2 + v_2 > 0], \end{aligned}$$

where (u_1, v_2) is independent of $\mathbf{z}$ and

$$\begin{pmatrix} u_1 \\ v_2 \end{pmatrix} \sim Normal\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{pmatrix}\right]$$

- Distribution $D(y_2|\mathbf{z})$ is straightforward: probit.
- $D(y_1|y_2,\mathbf{z})$ is more complicated, but tractable. For example,

$$P(y_1 = 1|y_2 = 1,\mathbf{z}) = \frac{1}{\Phi(\mathbf{z}\boldsymbol{\delta}_2)}\int_{-\mathbf{z}\boldsymbol{\delta}_2}^{\infty}\Phi[(\alpha_1 y_2 + \mathbf{z}_1\boldsymbol{\delta}_1 + \rho_1 \mathfrak{v}_2)/(1-\rho_1^2)^{1/2}] \cdot \phi(\mathfrak{v}_2)d\mathfrak{v}_2$$

- The other three conditional probabilities are similar. Combine these with the probit for $D(y_2|\mathbf{z})$ to obtain the MLE (conditional on $\mathbf{z}$).

• In Stata, can get "biprobit" to estimate this model. Get all parameter estimates directly.

• Much harder is to allow true simultaneity between y_1 and y_2. In fact, the model does not make logical sense for all values of parameters. Most applications are not truly simultaneous in nature.

• Because we are working with $D(y_1|y_2, \mathbf{z})$, it is straightforward to replace the linear function of $(y_2, \mathbf{z})$ with other functions, such as interactions between y_2 and elements of $\mathbf{z}$.

• You should **not** try to emulate "two stage least squares" as follows. (1) Run probit of y_{i2} on $\mathbf{z}_i$ and obtain the fitted probabilities, $\hat{\Phi}_{i2}$. (2) Run probit of y_{i1} on $\hat{\Phi}_{i2}$, $\mathbf{z}_{i1}$. The coefficients are usually much larger than other coefficients because $\hat{\Phi}_{i2}$ has a smaller range than y_{i2}.

• As far as we know, this "forbidden regression" estimates nothing interesting, although APEs have not been studied (I think).

• The problem is trying to take the expected value through the indicator function: If

$$y_1 = 1[\alpha_1 y_2 + \mathbf{z}_1\boldsymbol{\delta}_1 + u_1 > 0]$$

does it follow that

$$P(y_1 = 1|\mathbf{z}) = \Phi[\alpha_1\Phi(\mathbf{z}\boldsymbol{\delta}_2) + \mathbf{z}_1\boldsymbol{\delta}_1]?$$

• No. To see why, write $y_2 = \Phi(\mathbf{z}\boldsymbol{\delta}_2) + r_2$. Then

$$P(y_1 = 1|\mathbf{z}) = P[\alpha_1\Phi(\mathbf{z}\boldsymbol{\delta}_2) + \mathbf{z}_1\boldsymbol{\delta}_1 + \alpha_1 r_2 + u_1 > 0|\mathbf{z}].$$

But $\alpha_1 r_2 + u_1$ is not independent of $\mathbf{z}$ and is clearly not normally distributed.

5. PANEL DATA MODELS

Pooled Methods

• Useful to start with methods that do not explicitly introduce unobserved heterogeneity. Assume a balanced panel, $\{(\mathbf{x}_{it}, y_{it}) : t = 1, \ldots, T\}$ and N cross section observations.

• An index model for $P(y_{it} = 1|\mathbf{x}_{it})$ is

$$P(y_{it} = 1|\mathbf{x}_{it}) = G(\mathbf{x}_{it}\boldsymbol{\beta}), \; t = 1, \ldots, T,$$

where $\mathbf{x}_{it}$ generally includes a constant, time dummies, explanatory variables that do not change across i, and those that do.

• $\mathbf{x}_{it}$ can contain lagged dependent variables and lags of other variables.

- Pooled (partial) MLE is very attractive, as it is simple and require no further modeling. For each (i, t), the log likelihood is

$$\ell_{it}(\boldsymbol{\beta}) = (1 - y_{it}) \log[1 - G(\mathbf{x}_{it}\boldsymbol{\beta})] + y_{it} \log[G(\mathbf{x}_{it}\boldsymbol{\beta})]$$

- Consistency follows from general pooled MLE results. Generally, we need a sandwich etimator to account for serial correlation.

- For each t, the APE is estimated as

$$\left[N^{-1} \sum_{i=1}^{N} g(\mathbf{x}_{it}\hat{\boldsymbol{\beta}}) \right] \hat{\beta}_j$$

and these can be further averaged across t if desired to get a single scale factor.

- With small T and large N (our setting), apply the "panel bootstrap," where cross section units are resampled. That is, we sample from the integers $\{1, 2, \ldots, N\}$ and keep all time periods for each unit drawn. We do not resample time periods within cross section units.
- If the model is "dynamically complete" in the sense that

$$P(y_{it} = 1|\mathbf{x}_{it}, y_{i,t-1}, \mathbf{x}_{i,t-1}, \ldots, y_{i1}, \mathbf{x}_{i1}) = P(y_{it} = 1|\mathbf{x}_{it})$$

then we can used the usual standard errors reported with the pooled MLE. In addition, all of the standard tests, include the "likelihood ratio" test, are valid.

• As usual, this condition is unlikely to hold unless $\mathbf{x}_{it}$ contains one or more lagged dependent variables.

• How might we test for dynamic completeness? Lots of possibilities, but here is one. Compute residuals as $\hat{u}_{it} = y_{it} - G(\mathbf{x}_{it}\hat{\boldsymbol{\beta}})$, and then estimate the probit or logit "model" of y_{it} on $\mathbf{x}_{it}, \hat{u}_{i,t-1}$, $t = 2,\ldots,T, i = 1,\ldots,N$ and use the usual t statistic on $\hat{u}_{i,t-1}$.

• Dynamic models can be useful for prediction and controlling for endogeneity of policy interventions – just as in linear regression.

Models with Heterogeneity and Strictly Exogenous Regressors

- It does not hurt to start with a linear model

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

and also assume the strict exogeneity assumption,

$$P(y_{it} = 1|\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT}, c_i) = P(y_{it} = 1|\mathbf{x}_{it}, c_i), t = 1, \ldots, T$$

- Assuming the elements of $\mathbf{x}_{it}$ are time-varying (for at least some individuals), $\boldsymbol{\beta}$ can be consistently estimated by the usual fixed effects estimator applied to a linear model:

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}, E(u_{it}|\mathbf{x}_i, c_i) = 0, t = 1,\ldots,T.$$

- We should not take the LPM literal, because we must have

$$0 \leq \mathbf{x}_{it}\boldsymbol{\beta} + c_i \leq 1, \text{ all } \mathbf{x}_{it}$$

which puts strange restrictions on the heterogeneity distribution.

- But FE estimation of the linear model does not restrict $D(c_i|\mathbf{x}_i)$. Easy to make inference robust to serial correlation in u_{it} and heteroskedasticity.
- The FE coefficients can give reasonable estimates of average partial effects. In particular, they can be compared with APEs from nonlinear models

• Unobserved effects logit and probit models are popular nonlinear models. The probit model is given as

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

• Logit replaces $\Phi(\bullet)$ with $\Lambda(\bullet)$.

• Before introducing any additional assumptions, we can ask: What are the quantities of interest for most purposes? Usually, partial effects. For a continuous x_{tj},

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

• Depends on unobserved c, but sign is given by β_j.

• Can look at discrete changes:

$$\Phi(\mathbf{x}_t^{(1)}\boldsymbol{\beta} + c) - \Phi(\mathbf{x}_t^{(0)}\boldsymbol{\beta} + c)$$

Again, this depends on c.

• 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).

• But we often want 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.

• 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 a common scale, can we still learn something useful about magnitudes of partial effects? (iii) What kinds of assumptions do we need to estimate partial effects?

• Helpful to have a general setup. Let $\{(\mathbf{x}_{it}, y_{it}) : t = 1, \ldots, T\}$ be a random draw from the cross section. Suppose we are interested in

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

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

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

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

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 $\boldsymbol{\mu}_{\mathbf{c}} = E(\mathbf{c}_i)$, then we can compute the *partial effect at the average (PEA)*,

$$PEA_j(\mathbf{x}_t) = \theta_j(\mathbf{x}_t, \boldsymbol{\mu}_{\mathbf{c}}).$$

Of course, we need to estimate the function m_t and $\boldsymbol{\mu}_{\mathbf{c}}$. 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 $\mathbf{c}_i$:

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

- The APE is closely related to the notion of the *average structural function (ASF)* (Blundell and Powell (2003)). The ASF is defined as a function of $\mathbf{x}_t$:

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

- Passing the derivative through the expectation in the ASF gives an APE.

- How do APEs relate to parameters? Index model:

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

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

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

where $g(\cdot)$ is the derivative of $G(\cdot)$.

• The APE as a function of $\mathbf{x}_t$ "integrates out" c_i:

$$APE(\mathbf{x}_t) = \beta_j E_{c_i}[g(\mathbf{x}_t\boldsymbol{\beta} + c_i)]$$

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 $\mathbf{x}_{it}$ is correlated with $\mathbf{c}_i$. Of course, whether we can estimate the APEs, and how, certainly does.

Exogeneity Assumptions

• As in linear case, cannot get by with just specifying a model for the contemporaneous conditional distribution, $D(\mathbf{y}_{it}|\mathbf{x}_{it},\mathbf{c}_i)$.

• The most useful definition of strict exogeneity for nonlinear panel data models is

$$D(\mathbf{y}_{it}|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT},\mathbf{c}_i) = D(\mathbf{y}_{it}|\mathbf{x}_{it},\mathbf{c}_i).$$

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

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

• The *sequential exogeneity* assumption is

$$D(\mathbf{y}_{it}|\mathbf{x}_{i1},\ldots,\mathbf{x}_{it},\mathbf{c}_i) = D(\mathbf{y}_{it}|\mathbf{x}_{it},\mathbf{c}_i).$$

Unfortunately, it is much more difficult to allow sequential exogeneity in 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 $\mathbf{y}_{it}$.

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):

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

- In a parametric context, the CI assumption reduces our task to specifying a model for $D(\mathbf{y}_{it}|\mathbf{x}_{it}, \mathbf{c}_i)$, 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.

Assumptions about the Unobserved Heterogeneity

Random Effects

- Generally stated, the key RE assumption is

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

and then the unconditional distribution of $\mathbf{c}_i$ is modeled. This is very restrictive. It implies that all APEs can be obtained by just estimating $E(y_{it}|\mathbf{x}_{it} = \mathbf{x}_t)$.

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 $D(\mathbf{c}_i|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT})$, 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

$$D(\mathbf{c}_i|\mathbf{x}_i) = D(\mathbf{c}_i|\bar{\mathbf{x}}_i),$$

without restricting $D(\mathbf{c}_i|\bar{\mathbf{x}}_i)$.

- We will use parametric assumptions for $D(\mathbf{c}_i|\bar{\mathbf{x}}_i)$, such as normality (other possibilities exist), but some general arguments do not rely on a specific form for $D(\mathbf{c}_i|\bar{\mathbf{x}}_i)$.

- In particular, we can show that the APEs are identified very generally. By the LIE, we can always write

$$\begin{aligned} ASF(\mathbf{x}_t) &= E_{\mathbf{c}_i}[m_t(\mathbf{x}_t, \mathbf{c}_i)] \;= E_{\mathbf{x}_i}\{E[m_t(\mathbf{x}_t, \mathbf{c}_i)|\mathbf{x}_i]\} \\ &\equiv E_{\mathbf{x}_i}[r_t(\mathbf{x}_t, \bar{\mathbf{x}}_i)] \end{aligned}$$

where

$$r_t(\mathbf{x}_t, \bar{\mathbf{x}}_i) \equiv E[m_t(\mathbf{x}_t, \mathbf{c}_i)|\bar{\mathbf{x}}_i].$$

- Notice how $\mathbf{x}_t$ acts as a fixed argument; we will insert values later.

- Importantly, under strict exogeneity conditional conditional on $\mathbf{c}_i$ and the assumption $D(c_i|\mathbf{x}_i) = D(c_i|\bar{\mathbf{x}}_i)$, we have

$$E(y_{it}|\mathbf{x}_i) = E[E(y_{it}|\mathbf{x}_i, \mathbf{c}_i)|\mathbf{x}_i] = E[m_t(\mathbf{x}_{it}, \mathbf{c}_i)|\mathbf{x}_i] = \int m_t(\mathbf{x}_{it}, \mathbf{c})dF(\mathbf{c}|\mathbf{x}_i)$$
$$= \int m_t(\mathbf{x}_{it}, \mathbf{c})dF(\mathbf{c}|\bar{\mathbf{x}}_i) = r_t(\mathbf{x}_{it}, \bar{\mathbf{x}}_i).$$

- Because $E(y_{it}|\mathbf{x}_i)$ depends only on $(\mathbf{x}_{it}, \bar{\mathbf{x}}_i)$, we must have

$$E(y_{it}|\mathbf{x}_{it}, \bar{\mathbf{x}}_i) = r_t(\mathbf{x}_{it}, \bar{\mathbf{x}}_i).$$

• Therefore, once we have consistently estimated $r_t(\cdot, \cdot)$, a consistent estimator of the average structural function is

$$\widehat{ASF}(\mathbf{x}_t) = N^{-1} \sum_{i=1}^{N} \hat{r}_t(\mathbf{x}_t, \bar{\mathbf{x}}_i).$$

• We will obtain $\hat{r}_t(\cdot, \cdot)$ from parametric models, but flexible nonparametric approaches can be used because the mean function $E(y_{it}|\mathbf{x}_{it}, \bar{\mathbf{x}}_i)$ is identified generally.

Fixed Effects

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

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

- Conditional Independence is usually maintained in the approach based on finding sufficient statistics.
- Key point: PEAs and APEs are generally unidentified by methods that use conditioning to eliminate $\mathbf{c}_i$, essentially by construction.

Correlated Random Effects Probit

- Specify the model:

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

- Strict exogeneity conditional on c_i:

$$P(y_{it} = 1|\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}, c_i) = P(y_{it} = 1|\mathbf{x}_{it}, c_i),\ t = 1, \ldots, T.$$

- Conditional independence (where we condition on $\mathbf{x}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT})$ and c_i) :

$$D(y_{i1}, \ldots, y_{iT}|\mathbf{x}_i, c_i) = D(y_{i1}|\mathbf{x}_i, c_i) \cdots D(y_{iT}|\mathbf{x}_i, c_i)$$

- Model for $D(c_i|\mathbf{x}_i)$ (Mundlak special case of Chamberlain approach):

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

- Can obtain the first three assumptions from a latent variable model:

$$\begin{aligned} y_{it} &= 1[\mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it} > 0] \\ & u_{it}|(\mathbf{x}_{it},c_i) \sim Normal(0,1) \\ D(u_{it}|\mathbf{x}_i,c_i) &= D(u_{it}|\mathbf{x}_{it},c_i) \\ \{u_{it} &: t = 1,\ldots,T\} \text{ independent across } t \end{aligned}$$

• Can include time dummies in $\mathbf{x}_{it}$ but omit from $\bar{\mathbf{x}}_i$. Can also include time-constant elements, say $\mathbf{z}_i$:

$$c_i = \psi + \bar{\mathbf{x}}_i\boldsymbol{\xi} + \mathbf{z}_i\boldsymbol{\zeta} + a_i$$

(Up to you to intepret $\boldsymbol{\zeta}$)

• If $\boldsymbol{\xi} = \mathbf{0}$, get the traditional random effects probit model. Adding $\bar{\mathbf{x}}_i\boldsymbol{\xi}$ allows a specific form of correlation between c_i and $(\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT})$.

• MLE (conditional on $\mathbf{x}_i$) is relatively straightforward but it can be computationally demanding. It is based on the joint distribution $D(y_{i1},\ldots,y_{iT}|\mathbf{x}_i)$. For simplicity, omit $\mathbf{z}_i$.

$$\ell_i(\boldsymbol{\beta},\psi,\boldsymbol{\xi},\sigma_a^2) = \log\left[\int_{-\infty}^{\infty}\left(\prod_{t=1}^{T} f(y_{it}|\mathbf{x}_{it},c;\boldsymbol{\beta})\right)h(c|\bar{\mathbf{x}}_i;\psi,\boldsymbol{\xi},\sigma_a^2)dc\right]$$

• Here, $f(y_t|\mathbf{x}_t,c;\boldsymbol{\beta}) = [1-\Phi(\mathbf{x}_t\boldsymbol{\beta}+c)]^{(1-y_t)}[\Phi(\mathbf{x}_t\boldsymbol{\beta}+c)]^{y_t}$ and $h(c|\bar{\mathbf{x}}_i;\psi,\boldsymbol{\xi},\sigma_a^2)$ is the normal distributio with mean $\psi+\bar{\mathbf{x}}_i\boldsymbol{\xi}$ and variance σ_a^2.

- Requires numerical integration, but is programmed in lots of packages.
- All parameters, including are identified; inference is standard.
- In Stata, “xtprobit” with an “re” qualifier. Need to generate and include the time averages.
- Generally, including a set of time dummies is a good idea, and time constant variables can be included directly.
- Simple to compute a Wald test of whether the time averages are needed. $H_0 : \boldsymbol{\xi} = \mathbf{0}$.

```
egen x1bar = mean(x1), by(id)

egen x2bar = mean(x2), by(id)

egen xKbar = mean(xK), by(id)

xtprobit y d2 ... dTx1 x2 ... xK x1bar ... xKbar
z1 ... zJ, re

test x1bar x2bar ... xKbar
```

- Can estimate features of the unconditional distribution of c_i.
- For example, $c_i = \psi + \bar{\mathbf{x}}_i \boldsymbol{\xi} + a_i$ and so

$$\mu_c = E(c_i) = \psi + E(\bar{\mathbf{x}}_i)\boldsymbol{\xi}$$

A consistent estimator of μ_c is

$$\hat{\mu}_c = \hat{\psi} + \bar{\mathbf{x}}\hat{\boldsymbol{\xi}}$$

where $\bar{\mathbf{x}}$ is the sample average of $\bar{\mathbf{x}}_i$:

$$\bar{\mathbf{x}} = N^{-1} \sum_{i=1}^{N} \bar{\mathbf{x}}_i = (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{x}_{it}$$

- We also have

$$\sigma_c^2 = \boldsymbol{\xi}' Var(\bar{\mathbf{x}}_i)\boldsymbol{\xi} + \sigma_a^2,$$

and so

$$\hat{\sigma}_c^2 \equiv \hat{\boldsymbol{\xi}}' \left(N^{-1} \sum_{i=1}^{N} (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})' (\bar{\mathbf{x}}_i - \bar{\mathbf{x}}) \right) \hat{\boldsymbol{\xi}} + \hat{\sigma}_a^2$$

Can evaluate PEs at, say, the estimated mean value, say $\hat{\mu}_c$, or look at $\hat{\mu}_c \pm k\hat{\sigma}_c$ for various k.

- The APEs are gotten, as usual, from the ASF:

$$\begin{aligned}
ASF(\mathbf{x}_t) &= E_{c_i}[\Phi(\mathbf{x}_t\boldsymbol{\beta} + c_i)] \;\; = E_{\bar{\mathbf{x}}_i}\{E[\Phi(\mathbf{x}_t\boldsymbol{\beta} + c_i)|\bar{\mathbf{x}}_i]\} \\
&= E_{\bar{\mathbf{x}}_i}\{E[\Phi(\mathbf{x}_t\boldsymbol{\beta} + \psi + \bar{\mathbf{x}}_i\boldsymbol{\xi} + a_i)|\bar{\mathbf{x}}_i]\} \\
&= E_{\bar{\mathbf{x}}_i}\{\Phi[(\mathbf{x}_t\boldsymbol{\beta} + \psi + \bar{\mathbf{x}}_i\boldsymbol{\xi})/(1+\sigma_a^2)^{1/2}]\} \\
&\equiv E_{\bar{\mathbf{x}}_i}[\Phi(\mathbf{x}_t\boldsymbol{\beta}_a + \psi_a + \bar{\mathbf{x}}_i\boldsymbol{\xi}_a)]
\end{aligned}$$

where, for example, $\boldsymbol{\beta}_a = \boldsymbol{\beta}/(1+\sigma_a^2)^{1/2}$ are scaled coefficients.

• Because we have consistent estimators of all parameters, we can estimate $ASF(\mathbf{x}_t)$ consistently as

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

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

• Note where the averaging out occurs: across the sample of $\bar{\mathbf{x}}_i$.

• Take derivatives and changes with respect to $\mathbf{x}_t$. Can then average out across $\mathbf{x}_{it}$ to get a single APE.

• Conditional independence is very strong, and the usual RE estimator not known to be robust to its violation (unlike RE in linear model).

- If we focus on APEs, can just use a pooled method because

$$
\begin{aligned}
P(y_{it} = 1|\mathbf{x}_i) &= P(\mathbf{x}_{it}\boldsymbol{\beta} + \psi + \bar{\mathbf{x}}_i\boldsymbol{\xi} + a_i + u_{it} > 0|\mathbf{x}_i) \\
&= P[a_i + u_{it} > -(\mathbf{x}_{it}\boldsymbol{\beta} + \psi + \bar{\mathbf{x}}_i\boldsymbol{\xi})|\mathbf{x}_i] \\
&= \Phi(\mathbf{x}_{it}\boldsymbol{\beta}_a + \psi_a + \bar{\mathbf{x}}_i\boldsymbol{\xi}_a).
\end{aligned}
$$

- To estimate $\boldsymbol{\beta}_a, \psi_a$, and $\boldsymbol{\xi}_a$, just used pooled probit with $\bar{\mathbf{x}}_i$ as an additional set of explanatory variables. Cannot identify $\boldsymbol{\beta}$ and σ_a^2 separately, but do not need to for APEs.
- Pooled probit inefficient. Can use GMM or "generalized estimating equations" (essentially, multivariate nonlinear least squares) to enhance efficiency without sacrificing consistency.

• Using either the full random effects assumptions or pooled probit, it is easy to test the strict exogeneity assumption conditional on c_i, provided $T \geq 3$. Let $\mathbf{w}_{it}$ be a subset of $\mathbf{x}_{it}$ that possibly is not strictly exogenous. Then, along with time dummies, $\mathbf{x}_{it}$, $\bar{\mathbf{x}}_i$, and $\mathbf{z}_i$ (time-constant variables), include $\mathbf{w}_{i,t+1}$ and test joint significance. Lose the last time period.

- What is dubbed “fixed effects” probit is an inconsistent method (for fixed T) that treats c_i as N parameters to estimate. Suffers from incidental parameters problem.
- Some recent work shows that perhaps the APEs are well estimated without “too much” heterogeneity if T is not “too small.” Also, some corrections to the bias caused have been offered and studied.

Fixed Effects Logit

• If we replace the probit function by the logit function and maintain conditional independence, we can estimate $\boldsymbol{\beta}$ without restricting $D(c_i|\mathbf{x}_i)$. Often called "fixed effects logit," but it is really a conditional MLE were we condition on $(n_i, \mathbf{x}_i)$, where

$$n_i = \sum_{r=1}^{T} y_{ir}$$

is the total number of successes for unit i.

• Can show $D(y_{i1}, \ldots, y_{iT}|n_i, \mathbf{x}_i, c_i)$ does not depend on c_i, but does depend on $\boldsymbol{\beta}$, provided there is time variation in $\mathbf{x}_{it}$.

• Generally, $n_i = 0$ and $n_i = T$ observations are uninformative. So, when $T = 2$, only $n_i = 1$ observations contain information on $\boldsymbol{\beta}$:

$$P(y_{i2} = 1|n_i = 1, \mathbf{x}_i) = \Lambda[(\mathbf{x}_{i2} - \mathbf{x}_{i1})\boldsymbol{\beta}]$$
$$P(y_{i1} = 1|n_i = 1, \mathbf{x}_i) = 1 - \Lambda[(\mathbf{x}_{i2} - \mathbf{x}_{i1})\boldsymbol{\beta}]$$

Let $w_i = (1 - y_{i1})y_{i2}$. Then $D(w_i|\Delta\mathbf{x}_i)$ follows a standard logit model, where $\Delta\mathbf{x}_i = \mathbf{x}_{i2} - \mathbf{x}_{i1}$.

• Generally, not known to be consistent without condition independence. So it does not strictly relax assumptions for CRE probit when the latter is estimated using pooled probit, or some other robust method, such as GEE.

- PEAs and APEs not identified by FE logit (because the distribution of c_i is unspecified).

- In Stata, "xtlogit" with "fe" option.

```
xtlogit y d2 ... dT x2 ... xK, fe
```

- There is a CRE version of logit, but it is computationally hard and more difficult to work (no closed forms for APEs, for example) than CRE probit.

- Can show with $T = 2$ that, if treat c_i as parameters to estimate along with $\boldsymbol{\beta}$, the plim of the estimator is $2\boldsymbol{\beta}$.

Dynamic Models

- Difficult to specify and estimate models with heterogeneity if we do not assume strict exogeneity. Completely specified dynamic models can be estimated under certain assumptions.
- A linear model, estimated using the Arellano and Bond approach (and extensions), is a good starting point. Coefficients can be compared with partial effects from nonlinear models.

• Here we study a simple dynamic model: There is one lag of the dependent variable and all other explanatory variables are strictly exogenous:

$$P(y_{it} = 1|\mathbf{z}_i, y_{i,t-1}, \ldots, y_{i0}, c_i) = P(y_{it} = 1|\mathbf{z}_{it}, y_{i,t-1}, c_i),$$
$$t = 1, \ldots, T.$$

This also assumes that we have the dynamics correctly specified.

• Why is this specification of interest? Allows us to assess the relative important of "state dependence" – that is, whether being in a certain state last period affects the probability of being in that state this period – and unobserved heterogeneity. For example, if we control for different attributes in c_i, is welfare participation persistent? How persistent? Just seeing correlation over time, even conditional on $\mathbf{z}_{it}$, does not tell us that the previous state matters; we must also control for c_i.

• We study the dynamic probit model primarily for computational reasons; logit is more difficult:

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

which, as we will see, allows us to estimate the parameters and APEs very easily (under a distributional assumption for the heterogeneity).

• Treating the c_i as parameters to estimate causes inconsistency in $\boldsymbol{\delta}$ and ρ. Somewhat open question is how it affects bias in APEs. It is computationally intensive.

- Several different approaches to handling the "initial conditions" problem. (i) Treat the c_i as parameters to estimate (incidental parameters problem and computationally intensive). (ii) Try to estimate the parameters $\boldsymbol{\delta}$ and ρ without specifying conditional or unconditional distributions for c_i (available in some special cases). Generally, cannot estimate partial effects.). (iii) Approximate $D(y_{i0}|c_i, \mathbf{z}_i)$ and then model $D(c_i|\mathbf{z}_i)$. Leads to $D(y_{i0}, y_{i1}, \ldots, y_{iT}|\mathbf{z}_i)$ and MLE conditional on $\mathbf{z}_i$. (iv) Model $D(c_i|y_{i0}, \mathbf{z}_i)$. Leads to $D(y_{i1}, \ldots, y_{iT}|y_{i0}, \mathbf{z}_i)$ and MLE conditional on $(y_{i0}, \mathbf{z}_i)$. Wooldridge (2005b, Journal of Applied Econometrics) shows this can be computationally simple for popular models.

• Using the last approach for the probit model, a simple analysis is obtained from

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

Then

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

where $a_i \equiv c_i - \psi - \xi_0 y_{i0} - \mathbf{z}_i\boldsymbol{\xi}$. This allows us to characterize $D(y_{i1}, \ldots, y_{iT}|\mathbf{z}_i, y_{i0})$ after "integrating out" c_i.

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

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

and take differences or derivatives with respect to elements of $(\mathbf{z}_t, y_{t-1})$. As before, the coefficients are multiplied by $(1 + \hat{\sigma}_a^2)^{-1/2}$.

- Let $\mathbf{x}_{i0} \equiv (y_{i0}, \mathbf{z}_i)$. Then the first two moments of c_i are easily estimated:

$$\hat{\mu}_c = \hat{\psi} + \hat{\xi}_0 \bar{y}_0 + \bar{\mathbf{z}}\hat{\boldsymbol{\xi}}$$

$$\hat{\sigma}_c^2 = \hat{\boldsymbol{\lambda}}' \left(N^{-1} \sum_{i=1}^{N} (\mathbf{x}_{i0} - \bar{\mathbf{x}}_0)'(\mathbf{x}_{i0} - \bar{\mathbf{x}}_0) \right) \hat{\boldsymbol{\lambda}} + \hat{\sigma}_a^2$$

where $\hat{\boldsymbol{\lambda}} = (\hat{\xi}_0, \hat{\boldsymbol{\xi}}')'$.

6. MULTIVARIATE PROBIT

- Sometimes we have two or more binary responses to model. Call them y_g, $g = 1, \ldots, G$, each a binary response. No restriction such as $y_1 + y_2 + \ldots + y_G = 1$. In other words, any combination of zeros and ones is possible.
- Example: $G = 2$, y_1 indicates when a worker has employer-sponsored health insurance, y_2 indicates having an employer-sponsored pension plan.

• The marginal distributions (but conditional on **x**, as always) are assumed to follow probits:

$$P(y_g = 1|\mathbf{x}) = \Phi(\mathbf{x}_g\boldsymbol{\beta}_g), g = 1,\ldots,G.$$

• Multivariate probit is like seemingly unrelated regressions for binary response. Can be obtained from

$$\begin{aligned} y_{i1}^* &= \mathbf{x}_{i1}\boldsymbol{\beta}_1 + e_{i1} \\ y_{i2}^* &= \mathbf{x}_{i2}\boldsymbol{\beta}_2 + e_{i2} \\ &\vdots \\ y_{iG}^* &= \mathbf{x}_{iG}\boldsymbol{\beta}_G + e_{iG}, \end{aligned}$$

with $\mathbf{e}_i|\mathbf{x}_i \sim Normal(\mathbf{0},\boldsymbol{\Omega})$ with unit variances.

- Can be computationally hard with large G. Stata has the bivariate version programmed ("biprobit").
- Important difference with the linear case: if the joint distribution underlying multivariate probit is incorrect, but the probit marginals are correct, the joint MLE is (evidently) inconsistent. In the linear case,

$$y_{ig} = \mathbf{x}_{ig}\boldsymbol{\beta}_g + u_{ig}, g = 1, \ldots, G,$$

if every equation is correctly specified in the sense that $E(\mathbf{x}_i' u_{ig}) = \mathbf{0}$ for all g, the FGLS estimator is consistent even if, say, $E(\mathbf{u}_i \mathbf{u}_i' | \mathbf{x}_i)$ is heteroskedastic.

- And, of course, if $P(y_1 = 1|\mathbf{x}) = \Phi(\mathbf{x}_1\boldsymbol{\beta}_1)$ is correct but the probit model for equation two is incorrect, the joint procedure has no robustness properties.
- The reason to use multivariate probit is to enhance efficiency; how much it does is an empirical issue.
- Unlike in the linear case, there are no algebraic equivalences from having the same covariates in every equation.

7. EXAMPLES

LPM, Probit, and Logit with Exogenous Explanatory Variables

- Married women's labor force participation, using data from Mroz (1987)

- Dependent variable is *inlf*, "in the labor force."

```
. use mroz

. tab inlf

   =1 if in |
  lab frce, |
       1975 |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |        325       43.16       43.16
          1 |        428       56.84      100.00
------------+-----------------------------------
      Total |        753      100.00
```

```
. sum nwifeinc educ exper expersq age kidslt6 kidsge6

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
    nwifeinc |       753    20.12896     11.6348  -.0290575         96
        educ |       753    12.28685    2.280246          5         17
       exper |       753    10.63081     8.06913          0         45
     expersq |       753    178.0385    249.6308          0       2025
         age |       753    42.53785    8.072574         30         60
-------------+--------------------------------------------------------
     kidslt6 |       753    .2377158     .523959          0          3
     kidsge6 |       753    1.353254    1.319874          0          8
```

```
. * Estimate LPM by OLS.

. reg inlf nwifeinc educ exper expersq age kidslt6 kidsge6, robust

Linear regression                                      Number of obs =     753
                                                       F(  7,   745) =   62.48
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.2642
                                                       Root MSE      =  .42713

------------------------------------------------------------------------------
             |               Robust
        inlf |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc |  -.0034052   .0015249    -2.23   0.026    -.0063988   -.0004115
        educ |   .0379953    .007266     5.23   0.000      .023731    .0522596
       exper |   .0394924     .00581     6.80   0.000     .0280864    .0508983
     expersq |  -.0005963     .00019    -3.14   0.002    -.0009693   -.0002233
         age |  -.0160908    .002399    -6.71   0.000    -.0208004   -.0113812
     kidslt6 |  -.2618105   .0317832    -8.24   0.000    -.3242058   -.1994152
     kidsge6 |   .0130122   .0135329     0.96   0.337     -.013555    .0395795
       _cons |   .5855192   .1522599     3.85   0.000     .2866098    .8844287
------------------------------------------------------------------------------
```

```
. probit inlf nwifeinc educ exper expersq age kidslt6 kidsge6

Probit regression                                 Number of obs   =        753
                                                  LR chi2(7)      =     227.14
                                                  Prob > chi2     =     0.0000
Log likelihood = -401.30219                       Pseudo R2       =     0.2206

------------------------------------------------------------------------------
        inlf |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc |  -.0120237   .0048398    -2.48   0.013    -.0215096   -.0025378
        educ |   .1309047   .0252542     5.18   0.000     .0814074     .180402
       exper |   .1233476   .0187164     6.59   0.000     .0866641    .1600311
     expersq |  -.0018871      .0006    -3.15   0.002     -.003063   -.0007111
         age |  -.0528527   .0084772    -6.23   0.000    -.0694678   -.0362376
     kidslt6 |  -.8683285   .1185223    -7.33   0.000    -1.100628    -.636029
     kidsge6 |    .036005   .0434768     0.83   0.408     -.049208    .1212179
       _cons |   .2700768    .508593     0.53   0.595    -.7267473    1.266901
------------------------------------------------------------------------------
```

```
. * Compute partial effects at the averages.

. mfx

Marginal effects after probit
      y  = Pr(inlf) (predict)
         =  .58154201
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
nwifeinc |  -.0046962      .00189   -2.48   0.013  -.008401 -.000991    20.129
    educ |   .0511287      .00986    5.19   0.000   .031805  .070452   12.2869
   exper |   .0481771      .00733    6.57   0.000   .033815  .062539   10.6308
 expersq |  -.0007371      .00023   -3.14   0.002  -.001197 -.000277   178.039
     age |  -.0206432      .00331   -6.24   0.000  -.027127  -.01416   42.5378
 kidslt6 |  -.3391514      .04636   -7.32   0.000  -.430012 -.248291   .237716
 kidsge6 |   .0140628      .01699    0.83   0.408  -.019228  .047353   1.35325
------------------------------------------------------------------------------
```

```
. * Now the APEs. Not meaningful for the experience variables.

. margeff

Average partial effects after probit
      y  = Pr(inlf)

------------------------------------------------------------------------------
    variable |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc |  -.0036162   .0014414    -2.51   0.012    -.0064413   -.0007911
        educ |   .0393088   .0071877     5.47   0.000     .0252212    .0533964
       exper |    .037046    .005131     7.22   0.000     .0269893    .0471026
     expersq |  -.0005675   .0001771    -3.20   0.001    -.0009146   -.0002204
         age |  -.0158917   .0023569    -6.74   0.000     -.020511   -.0112723
     kidslt6 |  -.2441788   .0258995    -9.43   0.000    -.2949409   -.1934167
     kidsge6 |   .0108274   .0130538     0.83   0.407    -.0147576    .0364124
------------------------------------------------------------------------------
```

```
. logit inlf nwifeinc educ exper expersq age kidslt6 kidsge6

Logistic regression                               Number of obs   =        753
                                                  LR chi2(7)      =     226.22
                                                  Prob > chi2     =     0.0000
Log likelihood = -401.76515                       Pseudo R2       =     0.2197

------------------------------------------------------------------------------
        inlf |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc |  -.0213452   .0084214    -2.53   0.011    -.0378509   -.0048394
        educ |   .2211704   .0434396     5.09   0.000     .1360303    .3063105
       exper |   .2058695   .0320569     6.42   0.000     .1430391    .2686999
     expersq |  -.0031541   .0010161    -3.10   0.002    -.0051456   -.0011626
         age |  -.0880244    .014573    -6.04   0.000     -.116587   -.0594618
     kidslt6 |  -1.443354   .2035849    -7.09   0.000    -1.842373   -1.044335
     kidsge6 |   .0601122   .0747897     0.80   0.422     -.086473    .2066974
       _cons |   .4254524   .8603696     0.49   0.621    -1.260841    2.111746
------------------------------------------------------------------------------
```

```
. margeff

Average partial effects after logit
      y  = Pr(inlf)

------------------------------------------------------------------------------
    variable |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc |  -.0038118   .0014824    -2.57   0.010    -.0067172   -.0009064
        educ |   .0394323   .0072593     5.43   0.000     .0252044    .0536602
       exper |   .0367123   .0051289     7.16   0.000     .0266598    .0467648
     expersq |  -.0005633   .0001774    -3.18   0.001    -.0009109   -.0002156
         age |  -.0157153   .0023789    -6.61   0.000    -.0203779   -.0110527
     kidslt6 |   -.240805   .0259425    -9.28   0.000    -.2916515   -.1899585
     kidsge6 |   .0107335   .0133282     0.81   0.421    -.0153893    .0368564
------------------------------------------------------------------------------
```

Other Sources of Income Endogenous

```
. ivreg inlf educ exper expersq age kidslt6 kidsge6 (nwifeinc = huseduc)

Instrumental variables (2SLS) regression

      Source |       SS       df       MS              Number of obs =     753
-------------+------------------------------           F(  7,   745) =   36.41
       Model |  42.5996438     7  6.08566339           Prob > F      =  0.0000
    Residual |  142.128112   745  .190775989           R-squared     =  0.2306
-------------+------------------------------           Adj R-squared =  0.2234
       Total |  184.727756   752  .245648611           Root MSE      =  .43678

------------------------------------------------------------------------------
        inlf |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc |  -.0118549   .0057181    -2.07   0.038    -.0230804   -.0006294
        educ |   .0516295   .0116751     4.42   0.000     .0287096    .0745495
       exper |   .0370652   .0060138     6.16   0.000     .0252592    .0488713
     expersq |  -.0006144   .0001893    -3.25   0.001    -.0009861   -.0002428
         age |  -.0133932   .0030927    -4.33   0.000    -.0194645   -.0073218
     kidslt6 |  -.2527052   .0347755    -7.27   0.000    -.3209749   -.1844356
     kidsge6 |   .0168261   .0137223     1.23   0.221    -.0101129    .0437651
       _cons |   .4950353   .1683877     2.94   0.003     .1644645    .8256062
------------------------------------------------------------------------------
Instrumented:  nwifeinc
Instruments:   educ exper expersq age kidslt6 kidsge6 huseduc
------------------------------------------------------------------------------
```

```
. * Now Rivers-Vuong. Need first-stage residuals.

. reg nwifeinc huseduc educ exper expersq age kidslt6 kidsge6

      Source |       SS       df       MS              Number of obs =     753
-------------+------------------------------           F(  7,   745) =   27.13
       Model |  20676.7705     7  2953.82436           Prob > F      =  0.0000
    Residual |  81120.3451   745  108.886369           R-squared     =  0.2031
-------------+------------------------------           Adj R-squared =  0.1956
       Total |  101797.116   752  135.368505           Root MSE      =  10.435

------------------------------------------------------------------------------
    nwifeinc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     huseduc |   1.178155   .1609449     7.32   0.000     .8621956    1.494115
        educ |   .6746951   .2136829     3.16   0.002     .2552029    1.094187
       exper |  -.3129877   .1382549    -2.26   0.024    -.5844034   -.0415721
     expersq |  -.0004776   .0045196    -0.11   0.916    -.0093501     .008395
         age |   .3401521   .0597084     5.70   0.000     .2229354    .4573687
     kidslt6 |   .8262719   .8183785     1.01   0.313    -.7803305    2.432874
     kidsge6 |   .4355289   .3219888     1.35   0.177    -.1965845    1.067642
       _cons |  -14.72048   3.787326    -3.89   0.000    -22.15559   -7.285383
------------------------------------------------------------------------------

. predict v2hat, resid
```

```
. probit inlf nwifeinc educ exper expersq age kidslt6 kidsge6 v2hat

Probit regression                                 Number of obs   =        753
                                                  LR chi2(8)      =     229.14
                                                  Prob > chi2     =     0.0000
Log likelihood = -400.30301                       Pseudo R2       =     0.2225

------------------------------------------------------------------------------
        inlf |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc |  -.0368641   .0182706    -2.02   0.044    -.0726738   -.0010543
        educ |   .1702153   .0376718     4.52   0.000     .0963798    .2440507
       exper |   .1163123   .0193312     6.02   0.000     .0784239    .1542007
     expersq |  -.0019459   .0006009    -3.24   0.001    -.0031235   -.0007682
         age |   -.044953   .0101367    -4.43   0.000    -.0648206   -.0250855
     kidslt6 |  -.8444363   .1198154    -7.05   0.000     -1.07927   -.6096025
     kidsge6 |   .0477905   .0443204     1.08   0.281    -.0390758    .1346568
       v2hat |   .0267093   .0189352     1.41   0.158    -.0104031    .0638217
       _cons |   .0171187   .5392914     0.03   0.975    -1.039873     1.07411
------------------------------------------------------------------------------

. * Some evidence of endogeneity; p-value = .158.
```

```
. * Can still use the margeff option:

. margeff

Average partial effects after probit
      y  = Pr(inlf)

------------------------------------------------------------------------------
    variable |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc |  -.0110576   .0054418    -2.03   0.042    -.0217234   -.0003918
        educ |   .0509234   .0107908     4.72   0.000     .0297738     .072073
       exper |   .0348459   .0053706     6.49   0.000     .0243198    .0453721
     expersq |  -.0005837   .0001766    -3.30   0.001    -.0009299   -.0002375
         age |  -.0134815   .0029258    -4.61   0.000     -.019216    -.007747
     kidslt6 |  -.2377707   .0266742    -8.91   0.000    -.2900512   -.1854903
     kidsge6 |   .0143321   .0132573     1.08   0.280    -.0116518     .040316
       v2hat |   .0080116     .00566     1.42   0.157    -.0030817     .019105
------------------------------------------------------------------------------

. * Note how close the APEs are to the linear IV estimates.
```

Binary Endogenous Variable

- Binary endogenous explanatory variable is a dummy for having more than two children. Population is women with at least two children.
- Start with Linear IV. The binary variable *samesex* is the IV for *morekids*.

```
. reg morekids samesex nonmomi educ age agesq black hispan, robust

Linear regression                                      Number of obs =   31857
                                                       F(  7, 31849) =  398.53
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0717
                                                       Root MSE      =  .48174

------------------------------------------------------------------------------
             |               Robust
    morekids |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     samesex |   .0549983    .005398    10.19   0.000      .044418    .0655786
     nonmomi |  -.0010177     .00014    -7.27   0.000    -.0012921   -.0007432
        educ |  -.0337452   .0008836   -38.19   0.000    -.0354772   -.0320133
         age |   .0439758   .0113819     3.86   0.000     .0216668    .0662848
       agesq |  -.0003719   .0001958    -1.90   0.058    -.0007556    .0000119
       black |  -.0102972   .0343039    -0.30   0.764    -.0775342    .0569399
      hispan |  -.0257407   .0343662    -0.75   0.454    -.0930998    .0416183
       _cons |  -.0875206   .1668783    -0.52   0.600    -.4146085    .2395673
------------------------------------------------------------------------------
```

```
. ivreg worked nonmomi educ age agesq black hispan (morekids = samesex), robust

Instrumental variables (2SLS) regression               Number of obs =   31857
                                                       F(  7, 31849) =  374.59
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0737
                                                       Root MSE      =  .47347

------------------------------------------------------------------------------
             |               Robust
      worked |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    morekids |   -.200832   .0964728    -2.08   0.037    -.3899224   -.0117417
     nonmomi |    -.00126   .0001698    -7.42   0.000    -.0015928   -.0009271
        educ |   .0175522   .0033777     5.20   0.000     .0109318    .0241726
         age |   .0603517    .012166     4.96   0.000     .0365059    .0841974
       agesq |  -.0008178   .0001989    -4.11   0.000    -.0012076   -.0004281
       black |   .0168118   .0351723     0.48   0.633    -.0521271    .0857508
      hispan |  -.1308112   .0352456    -3.71   0.000     -.199894   -.0617284
       _cons |   -.454969   .1678432    -2.71   0.007     -.783948   -.1259899
------------------------------------------------------------------------------
Instrumented:  morekids
Instruments:   nonmomi educ age agesq black hispan samesex
------------------------------------------------------------------------------

. * So morekids has a large effect on labor force participation and is
. * marginally statistically significant.
```

```
. biprobit (worked = morekids nonmomi educ age agesq black hispan)
  (morekids = samesex nonmomi educ age agesq black hispan)

Seemingly unrelated bivariate probit              Number of obs   =      31857
                                                  Wald chi2(14)   =    5124.29
Log likelihood = -41106.422                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
worked       |
    morekids |  -.7025719    .204014    -3.44   0.001    -1.102432   -.3027119
     nonmomi |  -.0034903    .000395    -8.84   0.000    -.0042645   -.0027161
        educ |   .0405621   .0085385     4.75   0.000     .0238271    .0572972
         age |   .1632256   .0312412     5.22   0.000     .1019939    .2244573
       agesq |  -.0021524   .0005277    -4.08   0.000    -.0031867    -.001118
       black |   .0367322   .0909997     0.40   0.686    -.1416239    .2150883
      hispan |  -.3614826   .0912096    -3.96   0.000    -.5402502    -.182715
       _cons |  -2.475317   .4496294    -5.51   0.000    -3.356575    -1.59406
-------------+----------------------------------------------------------------
```

```
morekids     |
     samesex |   .1446566   .0144319    10.02   0.000     .1163705    .1729427
     nonmomi |  -.0027063   .0003685    -7.34   0.000    -.0034285   -.0019841
        educ |  -.0907148   .0024968   -36.33   0.000    -.0956083   -.0858212
         age |   .1190243   .0307613     3.87   0.000     .0587333    .1793154
       agesq |   -.001028   .0005284    -1.95   0.052    -.0020636    7.54e-06
       black |  -.0277804   .0921479    -0.30   0.763     -.208387    .1528263
      hispan |  -.0690523   .0922843    -0.75   0.454    -.2499262    .1118217
       _cons |  -1.572557   .4514335    -3.48   0.000    -2.457351   -.6877639
-------------+----------------------------------------------------------------
     /athrho |   .2599507   .1396201     1.86   0.063    -.0136996     .533601
-------------+----------------------------------------------------------------
         rho |   .2542495   .1305946                     -.0136987    .4881289
------------------------------------------------------------------------------
Likelihood-ratio test of rho=0:     chi2(1) =  3.33969    Prob > chi2 = 0.0676
```

```
. * Compute APE of morekids:

. predict xdh, xb

. gen xd0 = xdh - _b[morekids]*morekids

. gen xd1 = xd0 + _b[morekids]

. gen pe1 = norm(xd1) - norm(xd0)

. sum pe1

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         pe1 |     31857   -.2559131    .0208093  -.2746262  -.1606505

. * The APE, -.26, is somewhat larger than the IV estimate, -.20.
```

```
. * Now use the forbidden method of inserting fitted probit values from
. * a first-stage probit.

. probit morekids samesex nonmomi educ age agesq black hispan

Probit regression                                 Number of obs   =      31857
                                                  LR chi2(7)      =    2372.91
                                                  Prob > chi2     =     0.0000
Log likelihood = -20889.981                       Pseudo R2       =     0.0537

------------------------------------------------------------------------------
    morekids |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     samesex |   .1460784   .0143653    10.17   0.000     .1179229    .1742339
     nonmomi |  -.0026941   .0003681    -7.32   0.000    -.0034155   -.0019726
        educ |  -.0905486    .002495   -36.29   0.000    -.0954388   -.0856584
         age |   .1189666   .0307773     3.87   0.000     .0586441    .1792891
       agesq |  -.0010266   .0005286    -1.94   0.052    -.0020627    9.40e-06
       black |  -.0270085       .092    -0.29   0.769    -.2073252    .1533081
      hispan |  -.0683493   .0921359    -0.74   0.458    -.2489323    .1122337
       _cons |  -1.576492   .4516805    -3.49   0.000    -2.461769   -.6912142
------------------------------------------------------------------------------

. predict PHI2hat
(option pr assumed; Pr(morekids))
```

```
. probit worked PHI2hat nonmomi educ age agesq black hispan

Probit regression                                 Number of obs   =      31857
                                                  LR chi2(7)      =    2310.07
                                                  Prob > chi2     =     0.0000
Log likelihood = -20410.056                       Pseudo R2       =     0.0536

------------------------------------------------------------------------------
      worked |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     PHI2hat |  -.8426923   .2554568    -3.30   0.001    -1.343378   -.3420062
     nonmomi |  -.0036757     .00045    -8.17   0.000    -.0045576   -.0027938
        educ |   .0368082   .0088861     4.14   0.000     .0193919    .0542246
         age |   .1693934   .0327489     5.17   0.000     .1052067      .23358
       agesq |  -.0022009   .0005374    -4.10   0.000    -.0032541   -.0011476
       black |    .037665   .0915228     0.41   0.681    -.1417163    .2170463
      hispan |  -.3651419   .0919233    -3.97   0.000    -.5453083   -.1849755
       _cons |  -2.495462   .4504235    -5.54   0.000    -3.378276   -1.612649
------------------------------------------------------------------------------

. * The coefficient on PHI2hat is quite a bit larger in magnitude than the
. * bivariate MLE.
```

Static Panel Data Model

- Married Women's Labor Force Participation, LFP.DTA

```
. use lfp

. des lfp kids hinc

              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
lfp             byte   %9.0g                  =1 if in labor force
kids            byte   %9.0g                  number children < 18
hinc            float  %9.0g                  husband's monthly income, $

. tab period

  1 through |
  5, each 4 |
months long |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |      5,663       20.00       20.00
          2 |      5,663       20.00       40.00
          3 |      5,663       20.00       60.00
          4 |      5,663       20.00       80.00
          5 |      5,663       20.00      100.00
------------+-----------------------------------
      Total |     28,315      100.00

. egen kidsbar = mean(kids), by(id)

. egen lhincbar = mean(lhinc), by(id)
```

```
. * Linear model by FE:

. xtreg lfp kids lhinc per2-per5, fe cluster(id)

Fixed-effects (within) regression               Number of obs      =     28315
Group variable (i): id                          Number of groups   =      5663


                                  (Std. Err. adjusted for 5663 clusters in id)
------------------------------------------------------------------------------
             |               Robust
         lfp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        kids |  -.0388976   .0091682    -4.24   0.000    -.0568708   -.0209244
       lhinc |  -.0089439   .0045947    -1.95   0.052    -.0179513    .0000635
        per2 |  -.0042799    .003401    -1.26   0.208    -.0109472    .0023875
        per3 |  -.0108953   .0041859    -2.60   0.009    -.0191012   -.0026894
        per4 |  -.0123002   .0044918    -2.74   0.006    -.0211058   -.0034945
        per5 |  -.0176797   .0048541    -3.64   0.000    -.0271957   -.0081637
       _cons |   .8090216   .0375234    21.56   0.000     .7354614    .8825818
-------------+----------------------------------------------------------------
     sigma_u |  .42247488
     sigma_e |  .21363541
         rho |  .79636335   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

```
. * Fixed Effects Logit:

. xtlogit lfp kids lhinc per2-per5, fe
note: multiple positive outcomes within groups encountered.
note: 4608 groups (23040 obs) dropped because of all positive or
      all negative outcomes.

Conditional fixed-effects logistic regression   Number of obs      =      5275
Group variable: id                              Number of groups   =      1055

                                                Obs per group: min =         5
                                                               avg =       5.0
                                                               max =         5

                                                LR chi2(6)         =     57.27
Log likelihood  = -2003.4184                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         lfp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        kids |  -.6438386   .1247828    -5.16   0.000    -.8884084   -.3992688
       lhinc |  -.1842911   .0826019    -2.23   0.026    -.3461878   -.0223943
        per2 |  -.0928039   .0889937    -1.04   0.297    -.2672283    .0816205
        per3 |  -.2247989   .0887976    -2.53   0.011     -.398839   -.0507587
        per4 |  -.2479323   .0888953    -2.79   0.005     -.422164   -.0737006
        per5 |  -.3563745   .0888354    -4.01   0.000    -.5304886   -.1822604
------------------------------------------------------------------------------

. di 644/184
3.5

. di 389/89
4.3707865
```

```
. * CRE probit:

. xtprobit lfp kids lhinc kidsbar lhincbar educ black age agesq per2-per5, re

Random-effects probit regression                Number of obs      =     28315
Group variable (i): id                          Number of groups   =      5663

                                                Wald chi2(12)      =    824.11
Log likelihood  = -8990.0898                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         lfp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        kids |  -.3174051     .06203    -5.12   0.000    -.4389816   -.1958287
       lhinc |  -.0777949   .0414033    -1.88   0.060    -.1589439    .0033541
     kidsbar |  -.2098409   .0708676    -2.96   0.003    -.3487389   -.0709429
    lhincbar |  -.6463674   .0792719    -8.15   0.000    -.8017374   -.4909974
        educ |    .221596   .0147891    14.98   0.000     .1926099    .2505821
       black |   .5226558   .1502331     3.48   0.001     .2282042    .8171073
         age |   .4036543   .0287538    14.04   0.000     .3472979    .4600107
       agesq |  -.0054898   .0003536   -15.52   0.000    -.0061829   -.0047966
        per2 |   -.034359   .0438562    -0.78   0.433    -.1203156    .0515976
        per3 |  -.0954482   .0439688    -2.17   0.030    -.1816253    -.009271
        per4 |  -.1046944   .0439108    -2.38   0.017    -.1907581   -.0186308
        per5 |  -.1559446   .0435241    -3.58   0.000    -.2412502   -.0706389
       _cons |  -2.080352   .6567295    -3.17   0.002    -3.367518   -.7931854
-------------+----------------------------------------------------------------
    /lnsig2u |    1.73677   .0266277                      1.684581     1.78896
-------------+----------------------------------------------------------------
     sigma_u |   2.383059   .0317277                      2.321679    2.446063
         rho |   .8502764   .0033899                      .8435102    .8567997
------------------------------------------------------------------------------
Likelihood-ratio test of rho=0: chibar2(01) =  1.5e+04 Prob >= chibar2 = 0.000
```

```
. predict xdhat, xb

. gen xdhata = xdhat/sqrt(1 + 2.383059^2)

. di 1/sqrt(1 + 2.383059^2)
.38694144

. * Scaled coefficients to compare with pooled probit:

. di (1/sqrt(1 + 2.383059^2))*_b[kids]
-.1228172

. di (1/sqrt(1 + 2.383059^2))*_b[lhinc]
-.03010209
```

```
. probit lfp kids lhinc kidsbar lhincbar educ black age agesq per2-per5,
         cluster(id)

Probit regression                                 Number of obs   =      28315
                                                  Wald chi2(12)   =     538.09
                                                  Prob > chi2     =     0.0000
Log pseudolikelihood = -16516.436                 Pseudo R2       =     0.0673

                                  (Std. Err. adjusted for 5663 clusters in id)
------------------------------------------------------------------------------
             |               Robust
         lfp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        kids |  -.1173749   .0269743    -4.35   0.000    -.1702435   -.0645064
       lhinc |  -.0288098    .014344    -2.01   0.045    -.0569234   -.0006961
     kidsbar |  -.0856913   .0311857    -2.75   0.006     -.146814   -.0245685
    lhincbar |  -.2501781   .0352907    -7.09   0.000    -.3193466   -.1810097
        educ |   .0841338   .0067302    12.50   0.000     .0709428    .0973248
       black |   .2030668   .0663945     3.06   0.002     .0729359    .3331976
         age |   .1516424   .0124831    12.15   0.000      .127176    .1761089
       agesq |  -.0020672   .0001553   -13.31   0.000    -.0023717   -.0017628
        per2 |  -.0135701   .0103752    -1.31   0.191    -.0339051    .0067648
        per3 |  -.0331991   .0127197    -2.61   0.009    -.0581293    -.008269
        per4 |  -.0390317   .0136244    -2.86   0.004    -.0657351   -.0123284
        per5 |  -.0552425   .0146067    -3.78   0.000    -.0838711   -.0266139
       _cons |  -.7260562   .2836985    -2.56   0.010    -1.282095   -.1700173
------------------------------------------------------------------------------
```

```
. drop xdhat xdhata

. predict xdhat, xb

. gen scale = normden(xdhat)

. sum scale

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       scale |     28315    .3310079     .057301   .0694435   .3989423

. di .331*(-.117375)
-.03885113

. di .331*(-.02881)
-.00953611
```

```
. margeff

Average marginal effects on Prob(lfp==1) after probit

------------------------------------------------------------------------------
         lfp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        kids |   -.038852   .0089243    -4.35   0.000    -.0563433   -.0213608
       lhinc |  -.0095363   .0047482    -2.01   0.045    -.0188426     -.00023
     kidsbar |  -.0283645   .0102895    -2.76   0.006    -.0485315   -.0081974
    lhincbar |  -.0828109   .0115471    -7.17   0.000    -.1054428    -.060179
        educ |    .027849   .0021588    12.90   0.000     .0236178    .0320801
       black |   .0643443   .0200207     3.21   0.001     .0251043    .1035842
         age |   .0501948   .0039822    12.60   0.000     .0423898    .0579998
       agesq |  -.0006843   .0000493   -13.88   0.000    -.0007809   -.0005876
        per2 |  -.0044999   .0034482    -1.30   0.192    -.0112583    .0022585
        per3 |  -.0110375   .0042512    -2.60   0.009    -.0193698   -.0027052
        per4 |  -.0129865   .0045606    -2.85   0.004    -.0219252   -.0040479
        per5 |  -.0184197   .0049076    -3.75   0.000    -.0280385    -.008801
------------------------------------------------------------------------------
```

```
. probit lfp kids lhinc educ black age agesq per2-per5, cluster(id)

Probit regression                                 Number of obs   =      28315
                                                  Wald chi2(10)   =     537.36
                                                  Prob > chi2     =     0.0000
Log pseudolikelihood = -16556.671                 Pseudo R2       =     0.0651

                                  (Std. Err. adjusted for 5663 clusters in id)
------------------------------------------------------------------------------
             |               Robust
         lfp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        kids |  -.1989144   .0153153   -12.99   0.000    -.2289319   -.1688969
       lhinc |  -.2110739   .0242901    -8.69   0.000    -.2586816   -.1634661
        educ |   .0796863   .0065453    12.17   0.000     .0668577    .0925149
       black |   .2209396   .0659041     3.35   0.001       .09177    .3501093
         age |   .1449159   .0122179    11.86   0.000     .1209693    .1688624
       agesq |  -.0019912   .0001522   -13.08   0.000    -.0022895   -.0016928
        per2 |  -.0124245   .0104551    -1.19   0.235    -.0329162    .0080672
        per3 |  -.0325178   .0127431    -2.55   0.011    -.0574938   -.0075418
        per4 |   -.046097   .0136286    -3.38   0.001    -.0728087   -.0193853
        per5 |  -.0577767    .014632    -3.95   0.000    -.0864548   -.0290985
       _cons |  -1.064449    .261872    -4.06   0.000    -1.577709   -.5511895
------------------------------------------------------------------------------
```

```
. margeff

Average marginal effects on Prob(lfp==1) after probit

------------------------------------------------------------------------------
         lfp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        kids |  -.0660184   .0049233   -13.41   0.000    -.0756678    -.056369
       lhinc |   -.070054   .0079819    -8.78   0.000    -.0856981   -.0544099
        educ |   .0264473   .0021119    12.52   0.000     .0223082    .0305865
       black |   .0698835   .0197251     3.54   0.000      .031223     .108544
         age |   .0480966   .0039216    12.26   0.000     .0404105    .0557828
       agesq |  -.0006609   .0000486   -13.60   0.000    -.0007561   -.0005656
        per2 |  -.0041304   .0034828    -1.19   0.236    -.0109565    .0026957
        per3 |   -.010839   .0042694    -2.54   0.011    -.0192069   -.0024712
        per4 |  -.0153921   .0045809    -3.36   0.001    -.0243705   -.0064137
        per5 |  -.0193224   .0049309    -3.92   0.000    -.0289867   -.0096581
------------------------------------------------------------------------------

. * So, without accounting for heterogeneity through the time averages,
. * the effects are much larger.
```

```
. do ex15_5_boot1

. version 9

. capture program drop probit_boot

. program probit_boot, rclass
  1.
. probit lfp kids lhinc kidsbar lhincbar educ black age agesq per2-per5,
         cluster(id)
  2.
. predict xdhat, xb
  3. gen scale=normden(xdhat)
  4. gen pe1=scale*_b[kids]
  5. summarize pe1
  6. return scalar ape1=r(mean)
  7. gen pe2=scale*_b[lhinc]
  8. summarize pe2
  9. return scalar ape2=r(mean)
 10.
.
. drop xdhat scale pe1 pe2
 11. end
.
. bootstrap r(ape1) r(ape2), reps(500) seed(123) cluster(id) idcluster
            (newid): probit_boot
(running probit_boot on estimation sample)

Bootstrap replications (500)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50

..................................................   500

Bootstrap results                               Number of obs      =     28315
```

```
                                                 Number of clusters =      5663
                                                 Replications       =       500

      command:  probit_boot
        _bs_1:  r(ape1)
        _bs_2:  r(ape2)

------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Normal-based
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _bs_1 |   -.038852   .0085179    -4.56   0.000    -.0555469   -.0221572
       _bs_2 |  -.0095363     .00482    -1.98   0.048    -.0189833   -.0000893
------------------------------------------------------------------------------

. program drop probit_boot

end of do-file

. do ex15_5_boot2

. capture program drop probit_boot

. program probit_boot, rclass
  1.
. probit lfp kids lhinc educ black age agesq per2-per5, cluster(id)
  2.
. predict xdhat, xb
  3. gen scale=normden(xdhat)
  4. gen pe1=scale*_b[kids]
  5. summarize pe1
  6. return scalar ape1=r(mean)
  7. gen pe2=scale*_b[lhinc]
  8. summarize pe2
  9. return scalar ape2=r(mean)
```

```
 10.
.
. drop xdhat scale pe1 pe2
 11. end

. bootstrap r(ape1) r(ape2), reps(500) seed(123) cluster(id) idcluster(newid):
            probit_boot
(running probit_boot on estimation sample)

Bootstrap replications (500)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50

..................................................   500

Bootstrap results                               Number of obs      =     28315
                                                Number of clusters =      5663
                                                Replications       =       500

      command:  probit_boot
        _bs_1:  r(ape1)
        _bs_2:  r(ape2)

------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Normal-based
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _bs_1 |  -.0660184   .0047824   -13.80   0.000    -.0753916   -.0566451
       _bs_2 |   -.070054   .0078839    -8.89   0.000    -.0855061   -.0546019
------------------------------------------------------------------------------

. program drop probit_boot

end of do-file
```

Dynamic Model of Women's LFP

```
. * Start with a linear model estimated by Arellano and Bond:

. xtabond lfp kids lhinc per3 per4 per5

Arellano-Bond dynamic panel-data estimation  Number of obs         =     16989
Group variable: id                           Number of groups      =      5663
Time variable: period
                                             Obs per group:    min =         3
                                                               avg =         3
                                                               max =         3

Number of instruments =     12               Wald chi2(6)          =    378.77
                                             Prob > chi2           =    0.0000
One-step results
------------------------------------------------------------------------------
         lfp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lfp |
         L1. |   .3818295   .0201399    18.96   0.000     .3423559    .4213031
             |
        kids |  -.0130903   .0091827    -1.43   0.154     -.031088    .0049075
       lhinc |  -.0058375   .0053704    -1.09   0.277    -.0163633    .0046882
        per3 |  -.0053284   .0039777    -1.34   0.180    -.0131245    .0024677
        per4 |  -.0038833   .0039916    -0.97   0.331    -.0117067      .00394
        per5 |  -.0090286   .0039853    -2.27   0.023    -.0168396   -.0012176
       _cons |   .4848731   .0458581    10.57   0.000      .394993    .5747533
------------------------------------------------------------------------------
Instruments for differenced equation
        GMM-type: L(2/.).lfp
        Standard: D.kids D.lhinc D.per3 D.per4 D.per5
Instruments for level equation
        Standard: _cons
```

```
. * Accounting for heterogeneity is important, even in the linear
. * approximation. Without heterogeneity, the estimated state dependence is
. * much higher:

. reg lfp l.lfp kids lhinc per3 per4 per5, robust

Linear regression                                      Number of obs =   22652
                                                       F(  6, 22645) = 7938.78
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.7207
                                                       Root MSE      =  .24664

------------------------------------------------------------------------------
             |               Robust
         lfp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lfp |
         L1. |   .8510015   .0039478   215.57   0.000     .8432637    .8587394
             |
        kids |  -.0021431   .0014379    -1.49   0.136    -.0049615    .0006754
       lhinc |  -.0071892   .0025648    -2.80   0.005    -.0122164   -.0021619
        per3 |  -.0036044   .0047215    -0.76   0.445    -.0128588      .00565
        per4 |   .0010464   .0046287     0.23   0.821    -.0080262     .010119
        per5 |  -.0036555   .0045471    -0.80   0.421    -.0125681    .0052571
       _cons |    .157911   .0210127     7.52   0.000     .1167247    .1990972
------------------------------------------------------------------------------
```

```
. * Generate variables needed for dynamic probit.

. sort id period

. gen lfp_1 = lfp[_n-1] if period > 1
(5663 missing values generated)

. * Put initial condition in periods 2-5:
. gen lfp1 = lfp[_n-1] if per2
(22652 missing values generated)

. replace lfp1 = lfp[_n-2] if per3
(5663 real changes made)

. replace lfp1 = lfp[_n-3] if per4
(5663 real changes made)

. replace lfp1 = lfp[_n-4] if per5
(5663 real changes made)
```

```
. * Put all kids variables in periods 2-5:
. gen kids2 = kids if per2
(22652 missing values generated)

. replace kids2 = kids[_n-1] if per3
(5663 real changes made)

. replace kids2 = kids[_n-2] if per4
(5663 real changes made)

. replace kids2 = kids[_n-3] if per5
(5663 real changes made)

. gen kids3 = kids[_n+1] if per2
(22652 missing values generated)

. replace kids3 = kids if per3
(5663 real changes made)

. replace kids3 = kids[_n-1] if per4
(5663 real changes made)

. replace kids3 = kids[_n-2] if per5
(5663 real changes made)
```

```
. gen kids4 = kids[_n+2] if per2
(22652 missing values generated)

. replace kids4 = kids[_n+1] if per3
(5663 real changes made)

. replace kids4 = kids if per4
(5663 real changes made)

. replace kids4 = kids[_n-1] if per5
(5663 real changes made)

. gen kids5 = kids[_n+3] if per2
(22652 missing values generated)

. replace kids5 = kids[_n+2] if per3
(5663 real changes made)

. replace kids5 = kids[_n+1] if per4
(5663 real changes made)

. replace kids5 = kids if per5
(5663 real changes made)
```

```
. * Put all lhinc variables in periods 2-5:
. gen lhinc2 = lhinc if per2
(22652 missing values generated)

. replace lhinc2 = lhinc[_n-1] if per3
(5663 real changes made)

. replace lhinc2 = lhinc[_n-2] if per4
(5663 real changes made)

. replace lhinc2 = lhinc[_n-3] if per5
(5663 real changes made)

. gen lhinc3 = lhinc[_n+1] if per2
(22652 missing values generated)

. replace lhinc3 = lhinc if per3
(5663 real changes made)

. replace lhinc3 = lhinc[_n-1] if per4
(5663 real changes made)

. replace lhinc3 = lhinc[_n-2] if per5
(5663 real changes made)
```

```
. gen lhinc4 = lhinc[_n+2] if per2
(22652 missing values generated)

. replace lhinc4 = lhinc[_n+1] if per3
(5663 real changes made)

. replace lhinc4 = lhinc if per4
(5663 real changes made)

. replace lhinc4 = lhinc[_n-1] if per5
(5663 real changes made)

. gen lhinc5 = lhinc[_n+3] if per2
(22652 missing values generated)

. replace lhinc5 = lhinc[_n+2] if per3
(5663 real changes made)

. replace lhinc5 = lhinc[_n+1] if per4
(5663 real changes made)

. replace lhinc5 = lhinc if per5
(5663 real changes made)
```

```
. * Now include initial condition, leads and lags, and other
. * time-constant variables in RE probit
.
. xtprobit lfp lfp_1 lfp1 kids kids2-kids5 lhinc lhinc2-lhinc5 educ
            black age agesq per3-per5, re


Random-effects probit regression                Number of obs      =     22652
Group variable (i): id                          Number of groups   =      5663

Random effects u_i ~Gaussian                   Obs per group: min =          4
                                                               avg =        4.0
                                                               max =          4

                                                Wald chi2(19)      =   4091.17
Log likelihood  = -5028.9785                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         lfp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lfp_1 |   1.541288    .066803    23.07   0.000     1.410357     1.67222
        lfp1 |   2.530053   .1565322    16.16   0.000     2.223256    2.836851
        kids |  -.1455379   .0787386    -1.85   0.065    -.2998626    .0087868
       kids2 |   .3236282   .0968499     3.34   0.001      .133806    .5134504
       kids3 |   .1072842   .1235197     0.87   0.385    -.1348099    .3493784
       kids4 |     .01792   .1275595     0.14   0.888    -.2320921    .2679322
       kids5 |  -.3912412   .1058482    -3.70   0.000    -.5986998   -.1837825
       lhinc |  -.0748846   .0508406    -1.47   0.141    -.1745304    .0247612
      lhinc2 |  -.0232267   .0590167    -0.39   0.694    -.1388973    .0924438
      lhinc3 |   -.083386   .0626056    -1.33   0.183    -.2060908    .0393188
      lhinc4 |  -.0862979    .060961    -1.42   0.157    -.2057793    .0331835
      lhinc5 |   .0627793   .0592742     1.06   0.290     -.053396    .1789547
        educ |    .049906   .0100314     4.97   0.000     .0302447    .0695672
       black |   .1316009   .0982941     1.34   0.181     -.061052    .3242539
         age |   .1278946   .0193999     6.59   0.000     .0898715    .1659177
```

```
       agesq |  -.0016882     .00024    -7.03   0.000    -.0021586   -.0012177
        per3 |  -.0560723   .0458349    -1.22   0.221    -.1459071    .0337625
        per4 |   -.029532   .0463746    -0.64   0.524    -.1204245    .0613605
        per5 |  -.0784793   .0464923    -1.69   0.091    -.1696025     .012644
       _cons |  -2.946082   .4367068    -6.75   0.000    -3.802011   -2.090152
-------------+----------------------------------------------------------------
    /lnsig2u |   .0982792   .1225532                     -.1419206     .338479
-------------+----------------------------------------------------------------
     sigma_u |   1.050367   .0643629                      .9314989    1.184404
         rho |     .52455   .0305644                      .4645793     .583821
------------------------------------------------------------------------------
Likelihood-ratio test of rho=0: chibar2(01) =   160.73 Prob >= chibar2 = 0.000
```

```
. predict xdh, xb
(5663 missing values generated)

. gen xd0 = xdh - _b[lfp_1]*lfp_1
(5663 missing values generated)

. gen xd1 = xd0 + _b[lfp_1]
(5663 missing values generated)

. gen xd0a = xd0/sqrt(1 + (1.050367)^2)
(5663 missing values generated)

. gen xd1a = xd1/sqrt(1 + (1.050367)^2)
(5663 missing values generated)

. gen PHI0 = norm(xd0a)
(5663 missing values generated)

. gen PHI1 = norm(xd1a)
(5663 missing values generated)

. gen pelfp_1 = PHI1 - PHI0
(5663 missing values generated)
```

```
. sum pelfp_1

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
     pelfp_1 |     22652    .2591284    .0551711   .0675151   .4047995

. * .259 is the average probability of being in the labor force in
. * period t, given participation in t-1. This is somewhat lower than
. * the linear model estimate, .382.\pagebreak
. * A nonlinear model without heterogeneity gives a much larger
. * estimate:
```

```
. probit lfp lfp_1 kids lhinc educ black age agesq per3-per5

Probit regression                                 Number of obs   =      22652
                                                  LR chi2(10)     =   17744.22
                                                  Prob > chi2     =     0.0000
Log likelihood = -5332.5289                       Pseudo R2       =     0.6246

------------------------------------------------------------------------------
         lfp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lfp_1 |   2.875679   .0269811   106.58   0.000     2.822797    2.928561
        kids |   -.060792    .012217    -4.98   0.000    -.0847368   -.0368472
       lhinc |  -.1143176   .0211668    -5.40   0.000    -.1558037   -.0728315
        educ |   .0291868   .0052362     5.57   0.000     .0189241    .0394495
       black |   .0792495   .0536694     1.48   0.140    -.0259406    .1844395
         age |    .084403   .0099983     8.44   0.000     .0648067    .1039993
       agesq |  -.0010991   .0001236    -8.90   0.000    -.0013413    -.000857
        per3 |  -.0340795   .0369385    -0.92   0.356    -.1064777    .0383187
        per4 |   .0022816   .0371729     0.06   0.951    -.0705759    .0751391
        per5 |  -.0304156   .0371518    -0.82   0.413    -.1032318    .0424006
       _cons |  -2.170796   .2219074    -9.78   0.000    -2.605727   -1.735866
------------------------------------------------------------------------------
```

```
. predict xdph, xb
(5663 missing values generated)

. gen xdp0 = xdph - _b[lfp_1]*lfp_1
(5663 missing values generated)

. gen xdp1 = xdp0 + _b[lfp_1]
(5663 missing values generated)

. gen PHI0p = norm(xdp0)
(5663 missing values generated)

. gen PHI1p = norm(xdp1)
(5663 missing values generated)

. gen pelfp_1p = PHI1p - PHI0p
(5663 missing values generated)
```

```
. sum pelfp_1p

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
    pelfp_1p |     22652    .8373056     .012207   .6019558   .8495204

. * Without accounting for heterogeneity, the average state dependence
. * is much larger: .837 versus .259.

. * The .837 estimate is pretty close to the dynamic linear model without
. * heterogeneity, .851.
```