MODELS FOR FRACTIONAL RESPONSES

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

1. Introduction
2. Possible Approaches to Fractional Responses
3. Fractional Logit and Probit
4. Endogenous Explanatory Variables
5. Two-Part Models
6. Panel Data

1. INTRODUCTION

- Suppose y is a fractional response, that is, $0 \leq y \leq 1$.
- Allow the possibility that y is a corner solution at zero, one, or both. It could also be an essentially continuous variable strictly between zero and one in the population.
- y can be a proportion computed from the fraction of events occuring in a given number of trials. [For example, it could be the fraction of workers participating in a 401(k) pension plan.] But it could also be fundamentally continuous, such as the proportion of county land zoned for agriculture.

- For now, no problem of missing data, so we avoid the phrase "censored at zero" or "censored at one."
- If a variable is initially measured as a percentage, divide it by 100 to turn it into a proportion.
- Makes sense to start with linear models; at a minimum, estimated partial effects can be compared with those from more complicated nonlinear models.
- Remember a general rule: issues such as endogenous explanatory variables and unobserved heterogeneity are more easily handled with linear models. To allow nonlinear functional forms, we will impose extra assumptions.

2. POSSIBLE APPROACHES TO FRACTIONAL RESPONSES

- In the case where y has corners at zero *and* one, a two-limit Tobit model is logically consistent. But it uses a full set of distributional assumptions [which, of course, has the benefit of allowing us to estimate any feature of $D(y|\mathbf{x})$]. Plus, it is logically inconsistent if we have only one corner.

- Can use other logically consistent distributions. If y_i is a continuous on $(0,1)$, a conditional Beta distribution makes sense.

- The Beta distribution is not in the LEF, so, like the Tobit approach, MLE using the Beta distribution is inconsistent for the parameters in a correctly specified conditional mean.

• We focus mainly on models for estimating the conditonal mean. Later, discuss two-part models.

• Linear model has essentially same drawbacks as for binary response:

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

can hold over all potential values of $\mathbf{x}$ only in rare circumstances (such as mutually exclusive and exhaustive dummy variables).

- As with other limited dependent variables, we should view the linear model as the best linear approximation to $E(y|\mathbf{x})$ (which we can potentially improve by using quadratics, interactions, and other functional forms such as logarithms).
- As always, the OLS estimators are consistent for the linear projection parameters, which approximate (we hope) average partial effects.

• A common approach when $0 < y < 1$ is to use the so-called **log-odds transformation** of y, $\log[y/(1 - y)]$, in a linear regression. Define $w = \log[y/(1 - y)]$ and assume

$$E(w|\mathbf{x}) = \mathbf{x}\boldsymbol{\beta}.$$

• The log-odds approach is simple and, because w can range over all real values, the linear conditional mean is attractive.

• Drawbacks to the log-odds approach: First, it cannot be applied to corner solution responses unless we make some arbitrary adjustments. Because $\log[y/(1-y)] \rightarrow -\infty$ as $y \rightarrow 0$ and $\log[y/(1-y)] \rightarrow \infty$ as $y \rightarrow 1$, our estimates might be sensitive to the adjustments at the endpoints.

• Second, even if y is strictly in the unit interval, $\boldsymbol{\beta}$ is difficult to interpret: without further assumptions, it is not possible to estimate $E(y|\mathbf{x})$ from a model for $E\{\log[y/(1-y)]|\mathbf{x}\}$.

• One possibility is to assume the log-odds transformation yields a linear model with an additive error *independent* of $\mathbf{x}$:

$$\log[y/(1-y)] = \mathbf{x}\boldsymbol{\beta} + e, \ D(e|\mathbf{x}) = D(e),$$

where we take $E(e) = 0$ (and assume that $x_1 = 1$). Then, we can write

$$y = \exp(\mathbf{x}\boldsymbol{\beta} + e)/[1 + \exp(\mathbf{x}\boldsymbol{\beta} + e)].$$

• If e and $\mathbf{x}$ are independent,

$$E(y|\mathbf{x}) = \int \exp(\mathbf{x}\boldsymbol{\beta} + e)/[1 + \exp(\mathbf{x}\boldsymbol{\beta} + e)]dF(e),$$

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

• Duan's (*JASA*, 1983) "smearing estimate" can be used without specifying $D(e)$:

$$\hat{E}(y|\mathbf{x}) = N^{-1}\sum_{i=1}^{N}\exp(\mathbf{x}\hat{\boldsymbol{\beta}} + \hat{e}_i)/[1 + \exp(\mathbf{x}\hat{\boldsymbol{\beta}} + \hat{e}_i)],$$

where $\hat{\boldsymbol{\beta}}$ is the OLS estimator from w_i on $\mathbf{x}_i$ and $\hat{e}_i = w_i - \mathbf{x}_i\hat{\boldsymbol{\beta}}$ are the OLS residuals.

- Estimated partial effects are obtained by taking derivatives with respect to the x_j, or discrete differences.
- A similar analysis applies if we replace the log-odds transformation with $\Phi^{-1}(y)$, where $\Phi^{-1}(\cdot)$ is the inverse function of the standard normal cdf, in which case we average $\Phi(\mathbf{x}\hat{\boldsymbol{\beta}} + \hat{e}_i)$ across i to estimate $E(y|\mathbf{x})$.
- Can use the delta method for standard errors, or the bootstrap.
- Question: If we are mainly interested in $E(y|\mathbf{x})$, why not just model it directly?

3. FRACTIONAL LOGIT AND PROBIT

- Let y be a response in $[0,1]$, possibly including the endpoints. We can model its mean as

$$E(y|\mathbf{x}) = \exp(\mathbf{x}\boldsymbol{\beta})/[1+\exp(\mathbf{x}\boldsymbol{\beta})],$$

or as a probit function,

$$E(y|\mathbf{x}) = \Phi(\mathbf{x}\boldsymbol{\beta}).$$

- In each case the fitted values will be in $(0,1)$ and each allows y to take on any values in $[0,1]$, including the endpoints zero and one.

- Partial effects are obtained just as in standard logit and probit, but these are on the mean and not the response probability.
- The above functional forms do not, of course, exhaust the possibilities. For example,

$$E(y|\mathbf{x}) = \exp[-\exp(\mathbf{x}\boldsymbol{\beta})]$$

allows a different shape. (The function $G(z) = \exp(-\exp(z))$ is the cumulative distribution function of an asymmetric random variable.)

- Generally, let the mean function be $G(\mathbf{x}\boldsymbol{\beta})$. We could estimate $\boldsymbol{\beta}$ by nonlinear least squares. NLS is consistent and inference is straightforward, provided we use the fully robust sandwich variance matrix estimator that does not restrict $Var(y|\mathbf{x})$.
- As in estimating models of conditional means for unbounded, nonnegative responses, NLS is unlikely to be efficient for fractional responses because common distributions for a fractional response imply heteroskedasticity.
- Could use a two-step weighted NLS if we model $Var(y|\mathbf{x})$.

• A simpler, one-step strategy is to use a QMLE approach. We know the Bernoulli log likelihood is in the linear exponential family. Therefore, the QMLE that solves

$$\max_{\mathbf{b}} \sum_{i=1}^{N} \{(1-y_i)\log[1-G(\mathbf{x}_i\mathbf{b})] + y_i \log[G(\mathbf{x}_i\mathbf{b})]\}$$

is consistent for $\boldsymbol{\beta}$ whenever the conditional mean is correctly specified.

• Notice that the quasi-LLF is well defined for any y_i in $[0,1]$ and functions $0 < G(\cdot) < 1$. Plus, it is a standard estimation problem because it is identical to estimating binary response models.

• Call the QMLE **fractional logit regression** or **fractional probit regression**.

• These are just as robust as the NLS estimators.

• Fully robust inference is straightforward for QMLE. When the mean is correctly specified, estimate the asymptotic variance of $\hat{\boldsymbol{\beta}}$ as

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

where

$$\hat{u}_i = y_i - G(\mathbf{x}_i\hat{\boldsymbol{\beta}}).$$

- If we allow the mean to be misspecified, we replace the outer part of the sandwich with the estimated Hessian, not expected Hessian conditional on $\mathbf{x}_i$.
- The **Bernoulli GLM variance assumption** is

$$Var(y|\mathbf{x}) = \sigma^2 E(y|\mathbf{x})[1 - E(y|\mathbf{x})].$$

- When this assumption holds it is often with $\sigma^2 < 1$; in this case inference based on the usual binary response statistics will be too conservative – often, much too conservative.

- As we discussed in the general case, the Bernoulli QMLE has an attractive efficiency property. If it turns out that the GLM variance assumption holds, then the QMLE is efficient in the class of all estimators that use only $E(y|\mathbf{x}) = G(\mathbf{x}\boldsymbol{\beta})$ for consistency. In particular, the QMLE is more efficient than NLS.
- Of course, if, say, a Beta distribution were correct, and we use the correct MLE, this would be more efficient than the QMLE. But the MLE uses more assumptions for consistency.

• If the GLM variance assumption holds, the asymptotic variance matrix estimator simplies to

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

with

$$\begin{aligned}\hat{\sigma}^2 &= (N-K)^{-1}\sum_{i=1}^{N}(\hat{u}_i^2/\hat{v}_i),\\ \hat{v}_i &= G(\mathbf{x}_i\hat{\boldsymbol{\beta}})[1-G(\mathbf{x}_i\hat{\boldsymbol{\beta}})].\end{aligned}$$

- One case where Bernoulli GLM assumption holds. Suppose $y_i = s_i/n_i$, where s_i is the number of "successes" in n_i Bernoulli draws. Suppose that s_i given $(n_i, \mathbf{x}_i)$ follows a $Binomial[n_i, G(\mathbf{x}_i\boldsymbol{\beta})]$ distribution.

- Then $E(y_i|n_i, \mathbf{x}_i) = G(\mathbf{x}_i\boldsymbol{\beta})$ and $Var(y_i|n_i, \mathbf{x}_i) = n_i^{-1}G(\mathbf{x}_i\boldsymbol{\beta})[1 - G(\mathbf{x}_i\boldsymbol{\beta})]$. If n_i is independent of $\mathbf{x}_i$,

$$\begin{aligned} Var(y_i|\mathbf{x}_i) &= Var[E(y_i|n_i, \mathbf{x}_i)|\mathbf{x}_i] + E[Var(y_i|n_i, \mathbf{x}_i)|\mathbf{x}_i] \\ &= 0 + E(n_i^{-1}|\mathbf{x}_i)G(\mathbf{x}_i\boldsymbol{\beta})[1 - G(\mathbf{x}_i\boldsymbol{\beta})] \\ &\equiv \sigma^2 G(\mathbf{x}_i\boldsymbol{\beta})[1 - G(\mathbf{x}_i\boldsymbol{\beta})], \end{aligned}$$

where $\sigma^2 \equiv E(n_i^{-1}) \leq 1$ (with strict inequality unless $n_i = 1$ with probability one).

- In practice, it is unlikely that n_i and $\mathbf{x}_i$ are independent, so fully robust inference should be used.
- Further, within-group correlation – that is, if we write $s_i = \sum_{r=1}^{n_i} w_{ir}$ for binary responses w_{ir}, the $\{w_{ir} : r = 1, \ldots, n_i\}$ are correlated conditional on $(n_i, \mathbf{x}_i)$ – generally invalidates the GLM variance assumption, as in the Binomial case.

• If we are given data on proportions but do not know n_i, it makes sense to use a fractional logit or probit analysis. If we observe the n_i, we might use binomial regression instead (which is fully robust provided $E(s_i|n_i, \mathbf{x}_i) = n_i G(\mathbf{x}_i \boldsymbol{\beta})$).

• If we maintain $E(s_i|n_i, \mathbf{x}_i) = n_i G(\mathbf{x}_i \boldsymbol{\beta})$ and $y_i = s_i/n_i$,

$$E(y_i|n_i, \mathbf{x}_i) = E(s_i|n_i, \mathbf{x}_i)/n_i = G(\mathbf{x}_i \boldsymbol{\beta}) = E(y_i|\mathbf{x}_i)$$

This means that binomial regression using the counts s_i and fractional regression using y_i should yield similar estimates of $\boldsymbol{\beta}$.

- If the binomial distributional assumption is true, MLE using (s_i, n_i) is asymptotically more efficient than fractional regression. But the variance in binomial regression often has overdispersion. The fractional regression can actually be more efficient. (And, it is often more resilient to outliers.)

- Can compare the APEs to OLS estimates of a linear model. For a continuous variable x_j,

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

- If x_j is binary,

$$\widehat{APE}_j = N^{-1} \sum_{i=1}^{N} [G(\mathbf{x}_{ij}^{(1)} \hat{\boldsymbol{\beta}}) - G(\mathbf{x}_{ij}^{(0)} \hat{\boldsymbol{\beta}})]$$

where $\mathbf{x}_{ij}^{(1)}$ has $x_{ij} = 1$ and $\mathbf{x}_{ij}^{(0)}$ has $x_{ij} = 0$.

• Whether (say) x_K is discrete or continuous, we can obtain an estimate of APE_K when x_K changes from, say, $a_K^{(0)}$ to $a_K^{(1)}$, without using a calculus approximation, as in the previous equation but where

$$\begin{aligned}\mathbf{x}_{iK}^{(0)} &= \hat{\beta}_1 + \hat{\beta}_2 x_{i2} + \ldots + \hat{\beta}_{K-1} x_{i,K-1} + \hat{\beta}_K a_K^{(0)} \\ \mathbf{x}_{iK}^{(1)} &= \hat{\beta}_1 + \hat{\beta}_2 x_{i2} + \ldots + \hat{\beta}_{K-1} x_{i,K-1} + \hat{\beta}_K a_K^{(1)}\end{aligned}$$

• If, say, x_K is the key variable, and it is continuous, might plot the response as a function of x_K, inserting mean values (say) of the other variables or averaging them out:

$$G(\hat{\beta}_1 + \hat{\beta}_2\bar{x}_2 + \ldots + \hat{\beta}_{K-1}\bar{x}_{K-1} + \hat{\beta}_K x_K)$$

or

$$\widehat{ASF}_K(x_K) = N^{-1}\sum_{i=1}^{N} G(\hat{\beta}_1 + \hat{\beta}_2 x_{i2} + \ldots + \hat{\beta}_{K-1}x_{i,K-1} + \hat{\beta}_K x_K)$$

• Can compare these response functions with linear model.

• Can put the usual functional forms in the index; makes partial effects more difficult to compute.

• Simple functional form test: After estimation of $\hat{\boldsymbol{\beta}}$, add powers of $\mathbf{x}_i\hat{\boldsymbol{\beta}}$, such as $(\mathbf{x}_i\hat{\boldsymbol{\beta}})^2$, $(\mathbf{x}_i\hat{\boldsymbol{\beta}})^3$, use fractional QMLE on the expanded "model"

$$G(\mathbf{x}_i\boldsymbol{\beta} + \alpha_1(\mathbf{x}_i\hat{\boldsymbol{\beta}})^2 + \alpha_2(\mathbf{x}_i\hat{\boldsymbol{\beta}})^3),$$

and use a robust Wald test of joint significance for α_1, α_2. (This test, an extension of RESET for linear models, can be applied to any index context, including count regression with an exponential mean.)

• This is an example of a variable addition test, which is essentially a score test but slightly easier to implement.

- For goodness-of-fit of the mean, can compute an R-squared as

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}$$

where $\hat{y}_i = G(\mathbf{x}_i\hat{\boldsymbol{\beta}})$.

- Another possibility is the squared correlation between y_i and $\hat{y}_i$.
- Unlike OLS estimation of a linear model, these are not algebraically the same.

- GLM in Stata:

```
glm y x1 ... xK, fam(bin) link(logit) robust
glm y x1 ... xK, fam(bin) link(probit) sca(x2)
glm y x1 ... xK, fam(bin) link(loglog) robust
```

- Best to make inference fully robust, but the GLM variance assumption often gives similar standard errors.

- The usual MLE standard errors are too conservative, often very conservative.

- The "loglog" link implements the model $E(y|\mathbf{x}) = \exp[-\exp(\mathbf{x}\boldsymbol{\beta})]$.

- After any of the commands, fitted values are easy to get:

```
predict yhat
```

- To get the estimated indices, $\mathbf{x}_i\hat{\boldsymbol{\beta}}$, and powers of them:

```
predict xbhat, xb
gen xbhatsq = xbhat^2
gen xbhatcu = xbhat^3
```

EXAMPLE: Participation rates in 401(k) pension plans.

```
. use 401k

. des

Contains data from \swbook1_4e\statafiles\401k.dta
  obs:         1,534
 vars:             8                          9 Jun 1998 08:20
 size:        46,020 (99.9% of memory free)
-------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
prate           float  %7.0g                  participation rate, percent
mrate           float  %7.0g                  401k plan match rate
totpart         float  %7.0g                  total 401k participants
totelg          float  %7.0g                  total eligible for 401k plan
age             byte   %7.0g                  age of 401k plan
totemp          float  %7.0g                  total number of firm employees
sole            byte   %7.0g                  = 1 if 401k is firm's sole plan
ltotemp         float  %9.0g                  log of totemp
-------------------------------------------------------------------------------
Sorted by:
```

```
. sum

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       prate |      1534    87.36291    16.71654          3        100
       mrate |      1534    .7315124    .7795393        .01       4.91
     totpart |      1534    1354.231    4629.265         50      58811
      totelg |      1534    1628.535    5370.719         51      70429
         age |      1534    13.18123    9.171114          4         51
-------------+--------------------------------------------------------
      totemp |      1534    3568.495    11217.94         58     144387
        sole |      1534    .4876141    .5000096          0          1
     ltotemp |      1534    6.686034    1.453375   4.060443   11.88025

. count if mrate > 1
  292

. count if mrate > 2
  101

. replace prate = prate/100
(1534 real changes made)
```

```
. reg prate mrate age ltotemp sole, robust

Linear regression                                      Number of obs =    1534
                                                       F(  4,  1529) =   73.36
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.1474
                                                       Root MSE      =  .15456

------------------------------------------------------------------------------
             |               Robust
       prate |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       mrate |   .0485354   .0043289    11.21   0.000     .0400443    .0570266
         age |   .0031704   .0004032     7.86   0.000     .0023795    .0039613
     ltotemp |  -.0240487   .0031777    -7.57   0.000    -.0302818   -.0178156
        sole |   .0217378   .0086932     2.50   0.013      .004686    .0387896
       _cons |   .9465254   .0218303    43.36   0.000     .9037049    .9893458
------------------------------------------------------------------------------

. * The nonrobust standard errors are similar, actually slightly larger.
```

```
. glm prate mrate age ltotemp sole, fam(bin) link(logit) robust
note: prate has non-integer values

Generalized linear models                          No. of obs      =      1534
Optimization     : ML                              Residual df     =      1529
                                                   Scale parameter =         1
Deviance         =   314.528326                    (1/df) Deviance =  .2057085
Pearson          =  367.9839977                    (1/df) Pearson  =  .2406697

Variance function: V(u) = u*(1-u/1)                [Binomial]
Link function    : g(u) = ln(u/(1-u))              [Logit]

                                                   AIC             =  .5589556
Log pseudolikelihood = -423.7189416                BIC             = -10901.66

------------------------------------------------------------------------------
             |               Robust
       prate |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       mrate |   .9167158    .134119     6.84   0.000     .6538474    1.179584
         age |   .0322364   .0049561     6.50   0.000     .0225226    .0419502
     ltotemp |  -.2080024   .0258256    -8.05   0.000    -.2586195   -.1573852
        sole |   .1676861   .0846774     1.98   0.048     .0017215    .3336507
       _cons |   2.370495   .1921688    12.34   0.000     1.993851    2.747139
------------------------------------------------------------------------------
```

```
. margeff

Average partial effects after glm
      y  = Pr(prate)

------------------------------------------------------------------------------
    variable |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       mrate |   .0969144   .0140539     6.90   0.000     .0693694    .1244595
         age |   .0034081   .0005304     6.43   0.000     .0023686    .0044477
     ltotemp |  -.0219898   .0027723    -7.93   0.000    -.0274233   -.0165562
        sole |   .0176176   .0083374     2.11   0.035     .0012766    .0339586
------------------------------------------------------------------------------

. * The APE for mrate is about double the linear model estimate.
```

```
. predict prateh_l
(option mu assumed; predicted mean prate)

. corr prate prateh_l
(obs=1534)

             |    prate prateh_l
-------------+------------------
       prate |   1.0000
    prateh_l |   0.4263   1.0000

. di .4263^2
.18173169
```

```
. * The nonrobust standard errors are too large:

. glm prate mrate age ltotemp sole, fam(bin) link(logit)
note: prate has non-integer values

Generalized linear models                          No. of obs      =       1534
Optimization     : ML                              Residual df     =       1529
                                                   Scale parameter =          1
Deviance         =  314.528326                     (1/df) Deviance =   .2057085
Pearson          =  367.9839977                    (1/df) Pearson  =   .2406697

                                                   AIC             =   .5589556
Log likelihood   = -423.7189416                    BIC             =  -10901.66

------------------------------------------------------------------------------
             |                 OIM
       prate |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       mrate |   .9167158   .2059862     4.45   0.000     .5129902    1.320441
         age |   .0322364    .010257     3.14   0.002      .012133    .0523398
     ltotemp |  -.2080024   .0551219    -3.77   0.000    -.3160393   -.0999654
        sole |   .1676861   .1716409     0.98   0.329    -.1687239    .5040961
       _cons |   2.370495   .4263752     5.56   0.000     1.534815    3.206175
------------------------------------------------------------------------------
```

```
. gen mratesq = mrate^2
. gen agesq = age^2
. gen ltotempsq = ltotemp^2

. reg prate mrate mratesq age agesq ltotemp ltotempsq sole, robust

Linear regression                                      Number of obs =    1534
                                                       F(  7,  1526) =   56.14
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.1883
                                                       Root MSE      =  .15095

------------------------------------------------------------------------------
             |               Robust
       prate |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       mrate |    .137551   .0124891    11.01   0.000     .1130534    .1620487
     mratesq |  -.0255695   .0029956    -8.54   0.000    -.0314454   -.0196936
         age |   .0076809   .0015391     4.99   0.000     .0046619    .0106999
       agesq |   -.000129   .0000371    -3.48   0.001    -.0002017   -.0000563
     ltotemp |   -.113806   .0218575    -5.21   0.000    -.1566799    -.070932
   ltotempsq |   .0061188   .0014904     4.11   0.000     .0031953    .0090423
        sole |   .0119101   .0087466     1.36   0.173    -.0052465    .0290667
       _cons |     1.2029   .0788964    15.25   0.000     1.048143    1.357657
------------------------------------------------------------------------------
```

```
. * Now compute the APE for mrate:

. gen mrate_me_lin = _b[mrate] + 2*_b[mratesq]*mrate

. sum mrate_me_lin

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
mrate_me_lin |      1534    .1001422    .0398649  -.1135418   .1370397

. * Obtain RESET using the square and cube:

. predict xbh_sq_lin
(option xb assumed; fitted values)

. gen xbh_sq_linsq = xbh_sq_lin^2

. gen xbh_sq_lincu = xbh_sq_lin^3
```

```
. reg prate mrate mratesq age agesq ltotemp ltotempsq sole xbh_sq_linsq
      xbh_sq_lincu, robust

------------------------------------------------------------------------------
             |               Robust
       prate |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       mrate |   5.612875    2.25341     2.49   0.013     1.192761    10.03299
     mratesq |  -1.042382   .4188927    -2.49   0.013    -1.864049   -.2207152
         age |   .3121919    .125702     2.48   0.013     .0656248    .5587591
       agesq |  -.0052452   .0021126    -2.48   0.013    -.0093891   -.0011012
     ltotemp |  -4.634524     1.8645    -2.49   0.013    -8.291782   -.9772653
   ltotempsq |   .2492836   .1002737     2.49   0.013     .0525946    .4459725
        sole |   .4824868   .1956819     2.47   0.014     .0986524    .8663211
xbh_sq_linsq |  -40.86979   17.93412    -2.28   0.023    -76.04795   -5.691631
xbh_sq_lincu |   13.81167   6.515698     2.12   0.034     1.030992    26.59236
       _cons |   36.28292   14.74269     2.46   0.014     7.364812    65.20104
------------------------------------------------------------------------------

. test xbh_sq_linsq xbh_sq_lincu

 ( 1)  xbh_sq_linsq = 0
 ( 2)  xbh_sq_lincu = 0

       F(  2,  1524) =   21.68
            Prob > F =    0.0000

. * A strong statistical rejection of the linear model even with quadratics.
```

```
. glm prate mrate mratesq age agesq ltotemp ltotempsq sole, fam(bin)
      link(logit) robust
note: prate has non-integer values

Generalized linear models                          No. of obs      =      1534
Optimization     : ML                              Residual df     =      1526
                                                   Scale parameter =         1
Deviance         =  301.5717563                    (1/df) Deviance =  .1976224
Pearson          =  318.2190643                    (1/df) Pearson  =  .2085315

                                                   AIC             =  .5544207
Log pseudolikelihood = -417.2406567                BIC             = -10892.61

------------------------------------------------------------------------------
             |               Robust
       prate |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       mrate |   1.614381   .1675732     9.63   0.000     1.285943    1.942818
     mratesq |  -.2753789   .0435835    -6.32   0.000     -.360801   -.1899567
         age |   .0764414   .0158978     4.81   0.000     .0452823    .1076006
       agesq |  -.0012815    .000386    -3.32   0.001     -.002038    -.000525
     ltotemp |  -1.199122   .2209129    -5.43   0.000    -1.632103   -.7661407
   ltotempsq |   .0650906   .0145924     4.46   0.000       .03649    .0936912
        sole |   .1015973   .0837603     1.21   0.225    -.0625698    .2657644
       _cons |   5.535748    .833326     6.64   0.000     3.902459    7.169037
------------------------------------------------------------------------------
```

```
. predict prateh_l2

. corr prate prateh_l2
(obs=1534)

             |    prate prateh~2
-------------+------------------
       prate |   1.0000
   prateh_l2 |   0.4602   1.0000

. di .4602^2
.21178404

. * Fits better than linear model with quadratics (R-squared = .188).

. di 1.614/(2*.275)
2.9345455

. count if mrate > 2.93
   52

. * So only 52 out of 1,534 observations are to the right of the turning
. * point.
```

```
. * Using margeff with the quadratics doesn’t make much sense.
. * Compute APE "by hand."

. predict xbh_l2, xb

. gen scale = exp(xbh_l2)/(1 + exp(xbh_l2))^2

. sum scale

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       scale |      1534     .104565    .0540043   .0071288   .2364579

. gen mrate_me = (_b[mrate] + 2*_b[mratesq]*mrate)*scale

. sum mrate_me

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
    mrate_me |      1534    .1414986    .0902254  -.0718435   .3778262

. * About 40% higher than the linear model estimated APE, .100.
```

```
. predict xbh_sq_log, xb

. gen xbh_sq_logsq =  xbh_sq_log^2

. gen xbh_sq_logcu =  xbh_sq_log^3

. glm prate mrate mratesq age agesq ltotemp ltotempsq sole xbh_sq_logsq
     xbh_sq_logcu, fam(bin)  link(logit) robust
note: prate has non-integer values

------------------------------------------------------------------------------
             |               Robust
       prate |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       mrate |   4.065386   1.356546     3.00   0.003     1.406605    6.724167
     mratesq |   -.697588   .2330727    -2.99   0.003    -1.154402   -.2407738
         age |   .1928297   .0648958     2.97   0.003     .0656361    .3200232
       agesq |  -.0032323   .0011257    -2.87   0.004    -.0054387   -.0010259
     ltotemp |  -3.050577   1.057474    -2.88   0.004    -5.123189   -.9779661
   ltotempsq |   .1659176   .0585484     2.83   0.005     .0511647    .2806704
        sole |   .2623277   .1261936     2.08   0.038     .0149928    .5096626
xbh_sq_logsq |  -.8103757   .4117121    -1.97   0.049    -1.617317   -.0034348
xbh_sq_logcu |    .129453   .0625514     2.07   0.038     .0068545    .2520515
       _cons |   13.20982   4.311299     3.06   0.002     4.759834    21.65981
------------------------------------------------------------------------------
```

```
. test  xbh_sq_logsq xbh_sq_logcu

 ( 1)  [prate]xbh_sq_logsq = 0
 ( 2)  [prate]xbh_sq_logcu = 0

           chi2(  2) =    4.51
         Prob > chi2 =    0.1048

. * So we do not reject the fractional logit at the 10% significance level.

. * Can plot the mean function as a function of mrate, with other
. * variables fixed at specific values.
```

4. ENDOGENOUS EXPLANATORY VARIABLES

- The fractional probit model can easily handle certain kinds of continuous endogenous explanatory variables.

- As before, model endogeneity as an omitted variable:

$$E(y_1|\mathbf{z},y_2,c_1) = E(y_1|\mathbf{z}_1,y_2,c_1) = \Phi(\mathbf{z}_1\boldsymbol{\delta}_1 + \gamma_1 y_2 + c_1)$$
$$y_2 = \mathbf{z}\boldsymbol{\pi}_2 + v_2 = \mathbf{z}_1\boldsymbol{\pi}_{21} + \mathbf{z}_2\boldsymbol{\pi}_{22} + v_2,$$

where c_1 is an omitted factor thought to be correlated with y_2 but independent of the exogenous variables $\mathbf{z}$.

- Ideally, could assume the linear equation for y_2 represents a linear projection. But we need to assume more.
- Sufficient is

$$c_1 = \rho_1 v_2 + e_1,\ e_1|\mathbf{z}, v_2 \sim Normal(0, \sigma_{e1}^2),$$

where a sufficient, though not necessary, condition is that (c_1, v_2) is bivariate normal and independent of **z**.

• Then

$$E(y_1|\mathbf{z},y_2) = E(y_1|\mathbf{z},y_2,v_2) = \Phi(\mathbf{z}_1\boldsymbol{\delta}_{e1} + \gamma_{e1}y_2 + \rho_{e1}v_2),$$

where the "e" subscript denotes multiplication by the scale factor $1/(1+\sigma_{e_1}^2)^{1/2}$.

• Fortunately, the scaled coefficients index the average partial effects.

• Two-step method: (1) Obtain the OLS residuals $\hat{v}_{i2}$ from the regression y_{i2} on $\mathbf{z}_i$. Next, use fractional probit of y_{i1} on $\mathbf{z}_{i1}, y_{i2}, \hat{v}_{i2}$ to estimate the scaled coefficients.

• Simple test of the null hypothesis that y_2 is exogenous is the fully robust t statistic on $\hat{v}_{i2}$; the first-step estimation can be ignored under the null.

• If $\rho_1 \neq 0$, then the robust sandwich variance matrix estimator of the scaled coefficients is not valid because it does not account for the first step estimation. Can adjust for the two-step M-estimation results or use the bootstrap.

• The average structural function is consistently estimated as

$$\widehat{ASF}(\mathbf{z}_1, y_2) = N^{-1} \sum_{i=1}^{N} \Phi(\mathbf{z}_1 \hat{\boldsymbol{\delta}}_{e1} + \hat{\gamma}_{e1} y_2 + \hat{\rho}_{e1} \hat{v}_{i2}),$$

and this can be used to obtain APEs with respect to y_2 or $\mathbf{z}_1$.

• Bootstrapping the standard errors and test statistics is a sensible way to proceed with inference.

• Basic model can be extended in many ways. For example, can replace $y_2 = \mathbf{z}\boldsymbol{\pi}_2 + v_2$ with

$$h(y_2) = \mathbf{z}\boldsymbol{\pi}_2 + v_2$$

where $h(\bullet)$ is strictly monotonic. (This is for the case where we want y_2 in the structural model yet it is unlikely to have a linear reduced form with additive, independent error.)

• If $y_2 > 0$ then $h_2(y_2) = \log(y_2)$ is natural; if $0 < y_2 < 1$, might use the log-odds transformation, $h_2(y_2) = \log[y_2/(1 - y_2)]$.

• Unfortunately, if y_2 has a mass point – such as a binary response, or corner response, or count variable – a transformation yielding an additive, independent error probably does not exist.

• Allowing flexible functional forms for y_2 is easy. For example, if the structural model contains y_2^2 and interactions, say $y_2\mathbf{z}_1$, the estimating equation could look like

$$E(y_1|\mathbf{z},y_2,v_2) = \Phi(\mathbf{z}_1\boldsymbol{\delta}_{e1} + \gamma_{e1}y_2 + \psi_{e1}y_2^2 + y_2\mathbf{z}_1\boldsymbol{\alpha}_{e1} + \rho_{e1}v_2),$$

so that a single control function, v_2, corrects the endogeneity of y_2.

- After the two-step QMLE, the ASF is estimated as

$$\widehat{ASF}(\mathbf{z}_1, y_2) = N^{-1} \sum_{i=1}^{N} \Phi(\mathbf{z}_1 \hat{\boldsymbol{\delta}}_{e1} + \hat{\gamma}_{e1} y_2 + \hat{\psi}_{e1} y_2^2 + y_2 \mathbf{z}_1 \hat{\boldsymbol{\alpha}}_{e1} + \hat{\rho}_{e1} \hat{v}_{i2}),$$

and now derivatives or changes with respect to $(\mathbf{z}_1, y_2)$ can be obtained.

- Further, we might allow $D(c_1|v_2)$ to be more flexible, such as

$$c_1 = \rho_{11} v_2 + \rho_{12} v_2^2 + \rho_{13} v_2^3 + e_1,\ e_1|\mathbf{z}, v_2 \sim Normal(0, \sigma_{e1}^2).$$

- Notice that c_1 cannot have an unconditional normal distribution, particular if v_2 is normal. This bothers some people.

• In the second stage, we would add a cubic in $\hat{v}_2$ to the fractional probit. In the model just above,

$$\widehat{ASF}(\mathbf{z}_1, y_2) = N^{-1}\sum_{i=1}^{N}\Phi(\mathbf{z}_1\hat{\boldsymbol{\delta}}_{e1} + \hat{\gamma}_{e1}y_2 + \hat{\psi}_{e1}y_2^2 + y_2\mathbf{z}_1\hat{\boldsymbol{\alpha}}_{e1} + \hat{\rho}_{e11}\hat{v}_{i2} + \hat{\rho}_{e12}\hat{v}_{i2}^2 + \hat{\rho}_{e13}\hat{v}_{i2}^3),$$

that is, we again just average out the control function. The bootstrap would be very convenient for standard errors.

- Recent work by Blundell and Powell (2004, *Review of Economic Studies*) goes even further. Just allow $E(y_1|\mathbf{z}_1, y_2, v_2)$ to be a flexible function of its arguments, say

$$E(y_1|\mathbf{z}_1, y_2, v_2) = g_1(\mathbf{z}_1, y_2, v_2)$$

- To obtain the control function, assume

$$y_2 = g_2(\mathbf{z}) + v_2,\ v_2 \text{ independent of } \mathbf{z},$$

where $g_2(\bullet)$ is only assumed to be a smooth function. Estimate $g_2(\bullet)$ nonparametrically, obtain $\hat{v}_{i2} = y_{i2} - \hat{g}_2(\mathbf{z}_i)$. Then use nonparametric regression of y_{i1} on $\mathbf{z}_{i1}, y_{i2}, \hat{v}_{i2}$

- The ASF is consistently estimated as

$$\widehat{ASF}(\mathbf{z}_1, y_2) = N^{-1} \sum_{i=1}^{N} \hat{g}_1(\mathbf{z}_1, y_2, \hat{v}_{i2}).$$

- Can approximate this approach by using flexible parametric models.

- We can accomodate multiple continuous endogenous varaibles. Let $\mathbf{x}_1 = \mathbf{k}_1(\mathbf{z}_1, \mathbf{y}_2)$ for a vector of functions $\mathbf{k}_1(\cdot, \cdot)$, and allow a set of reduced forms for strictly monotonic functions $h_{2g}(y_{2g})$, $g = 1, \ldots, G_1$, where G_1 is the dimension of $\mathbf{y}_2$.
- See Wooldridge (2005, "Unobserved Heterogeneity and Estimation of Average Partial Effects," *Rothenberg Festschrift*)

• Recent (unpublished) work: If y_2 is binary and follows a probit model, can have y_1 fractional with a probit conditional mean and apply "bivariate probit" to (y_1, y_2), even though y_1 is not binary. (Not currently allowed in Stata.)

5. TWO-PART MODELS

- If have corners at $y = 0$ or $y = 1$ (or, occasionally at both values), might want to use a two-part (hurdle) model.
- For concreteness, assume $P(y = 0) > 0$ but y is continuous in $(0,1)$, so there is no pile-up at one.
- In addition to modeling $P(y = 0|\mathbf{x})$, could model $D(y|\mathbf{x}, y > 0)$. But more robust to model $E(y|\mathbf{x}, y > 0)$ using fractional response.

- Let

$$P(y = 0|\mathbf{x}) = 1 - F(\mathbf{x}\boldsymbol{\alpha})$$
$$E(y|\mathbf{x}, y > 0) = G(\mathbf{x}\boldsymbol{\beta})$$

- Let $w = 1[y > 0]$, so that $P(w = 1|\mathbf{x}) = F(\mathbf{x}\boldsymbol{\alpha})$.

- Now the "unconditional" expectation is

$$E(y|\mathbf{x}) = F(\mathbf{x}\boldsymbol{\alpha})G(\mathbf{x}\boldsymbol{\beta}),$$

which complicates partial effects.

• Estimation is straightforward. (1) Estimate $\boldsymbol{\alpha}$ by binary response (say, logit or probit) of w_i on $\mathbf{x}_i$. (2) Use QMLE (fractional logit or probit, or some other functional form) of y_i on $\mathbf{x}_i$ using data for $y_i > 0$ to estimate $\boldsymbol{\beta}$.

• Can compute an R-squared for the overall mean (and the mean conditional on $y > 0$) to compare with one-part models. Can test the functional forms of the two parts, too, using RESET and other tests (such as for "heteroskedasticity").

• Open (?) question: How to combine two-part models and endogeneity?

6. PANEL DATA METHODS

- If no interest in explicitly including unobserved heterogeneity, can use pooled versions of methods discussed. Of course, should allow for arbitrary serial dependence in inference as well as variance misspecification in the LEF distribution.

- Might have dynamic completeness in the mean if lagged dependent variables have been included. (What is the best functional form for doing so?) As usual, if $E(y_{it}|\mathbf{z}_{it}, y_{i,t-1}, \mathbf{z}_{i,t-1}, \ldots)$ has been properly specified, then serial correlation cannot be an issue.

- In Stata, we just use the “glm” command with a clustering option:

```
glm y x1 ... xK, fam(bin) link(logit)
cluster(id)
```

- With complete dynamics in the mean:

```
glm y x1 ... xK, fam(bin) link(logit) robust
```

or replace the logit link with another.

Models and Partial Effects with Heterogeneity

• Consider, for $0 \leq y_{it} \leq 1$,

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

• As with an endogenous explanatory variable in a cross section setting, with unobserved heterogeneity the fractional probit approach has advantages over other functional forms.

• Elements of $\boldsymbol{\beta}$ give the directions of the partial effects. For example, if x_{tj} is continuous, then

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

For discrete changes, we compute

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

for two different settings of the covariates, $\mathbf{x}_t^{(1)}$ and $\mathbf{x}_t^{(0)}$.

- Partial effects depend on $\mathbf{x}_t$ and c. What should we plug in for c?
- Instead, focus on the average partial effects (APEs):

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

which depends on $\mathbf{x}_t$ (and $\boldsymbol{\beta}$) but not on c. (Or discrete differences.) As before, essentially the same as the "average structural function," $ASF(\mathbf{x}_t) = E_{c_i}[\Phi(\mathbf{x}_t\boldsymbol{\beta} + c_i)]$.

• When are the APEs identified? More generally than the parameters, but more assumptions are needed.

• Strict exogeneity conditional on c_i: if $\mathbf{x}_i \equiv (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT})$,

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

As always, rules out lagged dependent variables, feedback, and contemporaneous endogeneity.

• Need to restrict $D(c_i|\mathbf{x}_i)$. Enough would be, say,

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

where $\bar{\mathbf{x}}_i$ is the time average.

- Altonji and Matzkin (2005, *Econometrica*) use general exchangeability. Could allow the distribution to depend on other features of $\{\mathbf{x}_{it} : t = 1, \ldots, T\}$, such as time trends or average growth rates.

- We assume more:

$$c_i|(\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT}) \sim Normal(\psi + \bar{\mathbf{x}}_i\xi, \sigma_a^2).$$

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

- Do not impose additional distributional assumptions on $D(y_{it}|\mathbf{x}_i, c_i)$. Leave the serial dependence in $\{y_{it}\}$ across time unrestricted.

• $\boldsymbol{\beta}$ is identified up to a positive scale factor, and the APEs are identified:

$$E(y_{it}|\mathbf{x}_i, a_i) = \Phi(\psi + \mathbf{x}_{it}\boldsymbol{\beta} + \bar{\mathbf{x}}_i\boldsymbol{\xi} + a_i)$$

and so

$$\begin{aligned} E(y_{it}|\mathbf{x}_i) &= E[\Phi(\psi + \mathbf{x}_{it}\boldsymbol{\beta} + \bar{\mathbf{x}}_i\boldsymbol{\xi} + a_i)|\mathbf{x}_i] \\ &= \Phi[(\psi + \mathbf{x}_{it}\boldsymbol{\beta} + \bar{\mathbf{x}}_i\boldsymbol{\xi})/(1 + \sigma_a^2)^{1/2}] \\ &\equiv \Phi(\psi_a + \mathbf{x}_{it}\boldsymbol{\beta}_a + \bar{\mathbf{x}}_i\boldsymbol{\xi}_a), \end{aligned}$$

where the "a" subscript denotes division by $(1 + \sigma_a^2)^{1/2}$.

• $\psi_a, \boldsymbol{\beta}_a,$ and $\boldsymbol{\xi}_a$ are identified if there is time variation in $\mathbf{x}_{it}$. Chamberlain device: replace $\bar{\mathbf{x}}_i$ with $\mathbf{x}_i$.

• Coneniently, the APEs can be obtained by differentiating or differencing

$$E_{\bar{x}_i}[\Phi(\psi_a + \mathbf{x}_t\boldsymbol{\beta}_a + \bar{\mathbf{x}}_i\boldsymbol{\xi}_a)]$$

with respect to the elements of x_t. The average structural function is consistently estimated by

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

where $\hat{\psi}_a$, $\hat{\boldsymbol{\beta}}_a$, $\hat{\boldsymbol{\xi}}_a$ are consistent estimators.

- As usual, APEs for continuous and discrete variables can be obtained from $\widehat{ASF}(\mathbf{x}_t)$.

- In practice, we would have time dummies, which we could just indicate with $\hat{\psi}_{at}$.

- We can always include time constant variables, say $\mathbf{r}_i$, along with $\bar{\mathbf{x}}_i$. It is then up to us to interpret the partial effects with respect to $\mathbf{r}_i$.

Estimation Methods Under Strict Exogeneity

• Many consistent estimators of the scaled parameters. Define $\mathbf{w}_{it} \equiv (1, \mathbf{x}_{it}, \bar{\mathbf{x}}_i)$ (or with time dummies and time constant variables) and $\boldsymbol{\theta} \equiv (\psi_a, \boldsymbol{\beta}_a', \boldsymbol{\xi}_a')'$. Then $\boldsymbol{\theta}$ can be estimated using pooled nonlinear least squares (NLS), with regression function $\Phi(\mathbf{w}_{it}\boldsymbol{\theta})$.

- Pooled NLS estimator is consistent and $\sqrt{N}$-asymptotically normal (with fixed T), but is likely to be inefficient.
- First, it ignores the serial dependence in the y_{it}, which is likely to be substantial even after conditioning on $\mathbf{x}_i$. Second, $Var(y_{it}|\mathbf{x}_i)$ is very unlikely to be homoskedastic. Could ignore serial correlation, model $Var(y_{it}|\mathbf{x}_i)$, and use weighted least squares.

• We already know that we can used pooled fractional probit (or logit, for that matter), with explanatory variables $(1, \mathbf{x}_{it}, \bar{\mathbf{x}}_i)$ (and, likely, year dummies).

• The "working variance" assumption for pooled FP is

$$Var(y_{it}|\mathbf{x}_i) = \tau^2\Phi(\mathbf{w}_{it}\boldsymbol{\theta})[1 - \Phi(\mathbf{w}_{it}\boldsymbol{\theta})],$$

where $0 < \tau^2 \leq 1$.

• Still need to cluster to obtain standard errors robust to serial correlation, even if the variance function is correct.

- In Stata, with year dummies explicit:

```
glm y x1 ... xK x1bar ... xKbar d2 ... dT,
fam(bin) link(probit) cluster(id)
margeff
```

- The “margeff” command gives APEs and appropriate standard errors.
- Can add time-constant variables to the list of explanatory variables.

- Random effects approaches – that is, that attempt to obtain a joint distribution $D(y_{i1},\ldots,y_{iT}|\mathbf{x}_i)$ by modeling and then integrating out unobserved heterogeneity – would require additional distributional assumptions while being computationally demanding. A nice middle ground is the GEE approach. We already have the working variance assumption for fractional probit.
- We need to specify a "working" correlation matrix, too. Define the errors as

$$u_{it} \equiv y_{it} - E(y_{it}|\mathbf{x}_i) = y_{it} - \Phi(\mathbf{w}_{it}\boldsymbol{\theta}) = \Phi(\psi_a + \mathbf{x}_{it}\boldsymbol{\beta}_a + \bar{\mathbf{x}}_i\boldsymbol{\xi}_a), t = 1,\ldots,T.$$

• Define standardized errors as

$$e_{it} \equiv u_{it}/\sqrt{\Phi(\mathbf{w}_{it}\boldsymbol{\theta})[1 - \Phi(\mathbf{w}_{it}\boldsymbol{\theta})]};$$

under, $Var(e_{it}|\mathbf{x}_i) = \tau^2$. Exchangeability is that the pairwise correlations between pairs of standardized errors is contant, say ρ.

• To estimate a common correlation parameter, let $\check{\boldsymbol{\theta}}$ be a preliminary estimator of $\boldsymbol{\theta}$ – probably the pooled Bernoulli QMLE.

• Define residuals $\check{u}_{it} \equiv y_{it} - \Phi(\mathbf{w}_{it}\check{\boldsymbol{\theta}})$ and the standardized (Pearson) residuals $\check{e}_{it} \equiv \check{u}_{it}/\sqrt{\Phi(\mathbf{w}_{it}\check{\boldsymbol{\theta}})[1 - \Phi(\mathbf{w}_{it}\check{\boldsymbol{\theta}})]}$. Then,

$$\hat{\rho} = [NT(T-1)]^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s \neq t} \check{e}_{it}\check{e}_{is}.$$

• Given the estimated $T \times T$ working correlation matrix, $\mathbf{R}(\hat{\rho})$, which has unity down its diagonal and $\tilde{\rho}$ everywhere else, we can construct the estimated "working" variance matrix:

$$\mathbf{V}(\mathbf{x}_i, \check{\boldsymbol{\theta}})^{1/2}\mathbf{R}(\hat{\rho})\mathbf{V}(\mathbf{x}_i, \check{\boldsymbol{\theta}})^{1/2},$$

where $\mathbf{V}(\mathbf{x}_i, \check{\boldsymbol{\theta}})$ is the $T \times T$ diagonal matrix with $\Phi(\mathbf{w}_{it}\check{\boldsymbol{\theta}})[1 - \Phi(\mathbf{w}_{it}\check{\boldsymbol{\theta}})]$ down its diagonal.

- Now apply multivariate WNLS, which is asymptotically the same as GEE. Naturally, use a fully robust variance matrix estimator.
- Can allow an “unstructured” correlation matrix, too, but the correlations never depend on $\mathbf{x}_i$.

```
xtgee y x1 ... xK x1bar ... xKbar, fam(bin)
link(probit) corr(exch) robust
xtgee y x1 ... xK x1bar ... xKbar, fam(bin)
link(probit) corr(uns) robust
```

• Can apply the "margeff" command in Stata to get APEs averaged across the cross section and time. For continuous explanatory variables, the common scale factor is

$$(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\phi(\hat{\psi}_a + \mathbf{x}_{it}\hat{\boldsymbol{\beta}}_a + \bar{\mathbf{x}}_i\hat{\boldsymbol{\xi}}_a).$$

• Can compare APEs with linear model estimated by fixed effects.

• As with previous models, can add $\mathbf{x}_{i,t+1}$ (or a subset of variables) as a test of strict exogeneity. Estimation can be pooled QMLE or GEE.

Models with Endogenous Explanatory Variables

- Represent endogeneity as an omitted, time-varying variable, in addition to unobserved heterogeneity:

$$\begin{aligned} E(y_{it1}|\mathbf{z}_i, y_{it2}, c_{i1}, v_{it1}) &= E(y_{it1}|\mathbf{z}_{it1}, y_{it2}, c_{i1}, v_{it1}) \\ &= \Phi(\mathbf{z}_{it1}\boldsymbol{\delta}_1 + \alpha_1 y_{it2} + c_{i1} + v_{it1}), \end{aligned}$$

where c_{i1} is the time-constant unobserved effect and v_{it1} is a time-varying omitted factor that can be correlated with y_{it2}.

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

• Use a Chamberlain-Mundlak approach, but only relating the heterogeneity to all strictly exogenous variables:

$$c_{i1} = \psi_1 + \bar{\mathbf{z}}_i \boldsymbol{\xi}_1 + a_{i1},\ D(a_{i1}|\mathbf{z}_i) = D(a_{i1}).$$

• Even before we specify $D(a_{i1})$, this is restrictive because it assumes, in particular, $E(c_i|\mathbf{z}_i)$ is linear in $\bar{\mathbf{z}}_i$ and that $Var(c_i|\mathbf{z}_i)$ is constant. More recent work has shown how we can get by with less, such as $D(c_{i1}|\mathbf{z}_i) = D(c_{i1}|\bar{\mathbf{z}}_i)$.

- Need to obtain an estimating equation. First, note that

$$
\begin{aligned}
E(y_{it1}|\mathbf{z}_i, y_{it2}, a_{i1}, v_{it1}) &= \Phi(\mathbf{z}_{it1}\boldsymbol{\delta}_1 + \alpha_1 y_{it2} + \psi_1 + \bar{\mathbf{z}}_i\boldsymbol{\xi}_1 + a_{i1} + v_{it1}) \\
&\equiv \Phi(\mathbf{z}_{it1}\boldsymbol{\delta}_1 + \alpha_1 y_{it2} + \psi_1 + \bar{\mathbf{z}}_i\boldsymbol{\xi}_1 + r_{it1}).
\end{aligned}
$$

- Assume a linear reduced form for y_{it2}:

$$
\begin{aligned}
y_{it2} &= \psi_2 + \mathbf{z}_{it}\delta_2 + \bar{\mathbf{z}}_i\boldsymbol{\xi}_2 + v_{it2}, t = 1,\ldots,T \\
D(v_{it2}|\mathbf{z}_i) &= D(v_{it2})
\end{aligned}
$$

(and we might allow for time-varying coefficients).

• Rather than assume (a_{i1}, v_{it2}) is independent of $\mathbf{z}_i$, we can get by with a weaker assumption, but we imposed normality:

$$r_{it1}|(\mathbf{z}_i, v_{it2}) \sim Normal(\eta_1 v_{it2}, \kappa_1^2), t = 1, \ldots, T.$$

[Easy to allow η_1 to change over time.]

• Either way, the assumptions effectively rule out discreteness in y_{it2}.

- Write

$$r_{it1} = \eta_1 v_{it2} + e_{it1}$$

where e_{it1} is independent of $(\mathbf{z}_i, v_{it2})$ (and, therefore, of y_{it2}) and normally distributed. Again, using a standard mixing property of the normal distribution,

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

where the "κ" denotes division by $(1 + \kappa_1^2)^{1/2}$.

- Identification comes off of the exclusion of the time-varying exogenous variables $\mathbf{z}_{it2}$.

- Two step procedure:

(1) Estimate the reduced form for y_{it2} (pooled or for each t separately). Obtain the residuals, $\hat{v}_{it2}$.

(2) Use the probit QMLE to estimate $\boldsymbol{\delta}_{\kappa 1}, \alpha_{\kappa 1}, \psi_{\kappa 1}, \boldsymbol{\xi}_{\kappa 1}$ and $\eta_{\kappa 1}$.

(GEE would require strict exogeneity of $\{v_{it2}\}$!)

- How do we interpret the scaled estimates? They give directions of effects. Conveniently, they also index the APEs. For given $\mathbf{z}_1$ and y_2, average out $\bar{\mathbf{z}}_i$ and $\hat{v}_{it2}$ (for each t):

$$\hat{\alpha}_{\kappa 1} \cdot \left[N^{-1} \sum_{i=1}^{N} \phi(\mathbf{z}_{t1}\hat{\boldsymbol{\delta}}_{\kappa 1} + \hat{\alpha}_{\kappa 1} y_{t2} + \hat{\psi}_{\kappa 1} + \bar{\mathbf{z}}_i \hat{\boldsymbol{\xi}}_{\kappa 1} + \hat{\eta}_{\kappa 1} \hat{v}_{it2}) \right].$$

- Applying “margeff” in the second stage consistently estimates the APEs averaged across t, but the standard errors do not account for the two-step estimation. Use panel bootstrap for standard errors to allow for serial dependence and the two-step estimation.
- Of course, we can also compute discrete changes for any of the elements of $(\mathbf{z}_{t1}, y_{t2})$.

EXAMPLE: **Effects of Spending on Test Pass Rates**

- Reform occurs between 1993/94 and 1994/95 school year; its passage was a surprise to almost everyone.
- Since 1994/95, each district receives a foundation allowance, based on revenues in 1993/94.
- Intially, all districts were brought up to a minimum allowance – $4,200 in the first year. The goal was to eventually give each district a basic allowance ($5,000 in the first year).
- Districts divided into three groups in 1994/95 for purposes of initial foundation allowance. Subsequent grants determined by statewide School Aid Fund.

- Catch-up formula for districts receiving below the basic. Initially, more than half of the districts received less than the basic allowance. By 1998/99, it was down to about 36%. In 1999/00, all districts began receiving the basic allowance, which was then $5,700. Two-thirds of all districts now receive the basic allowance.
- From 1991/92 to 2003/04, in the 10th percentile, expenditures rose from $4,616 (2004 dollars) to $7,125, a 54 percent increase. In the 50th percentile, it was a 48 percent increase. In the 90th percentile, per pupil expenditures rose from $7,132 in 1992/93 to $9,529, a 34 percent increase.

• Response variable: $math4$, the fraction of fourth graders passing the MEAP math test at a school.

• Spending variable is $\log(avgrexppp)$, where the average is over the current and previous three years.

• The linear model is

$$math4_{it} = \theta_t + \beta_1 \log(avgrexp_{it}) + \beta_2 lunch_{it} + \beta_3 \log(enroll_{it}) + c_{i1} + u_{it1}$$

Estimating this model by fixed effects is identical to adding the time averages of the three explanatory variables and using pooled OLS.

• The "fractional probit" model:

$$E(math4_{it}|\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT}) = \Phi(\theta_{at} + \mathbf{x}_{it}\boldsymbol{\beta}_a + \bar{\mathbf{x}}_i\boldsymbol{\xi}_a).$$

- Allowing spending to be endogenous. Controlling for 1993/94 spending, foundation grant should be exogenous. Exploit nonsmoothness in the grant as a function of initial spending.

$$math4_{it} = \theta_t + \beta_1 \log(avgrexp_{it}) + \beta_2 lunch_{it} + \beta_3 \log(enroll_{it})$$

$$+ \beta_{4t} \log(rexppp_{i,1994}) + \xi_1 \overline{lunch_i} + \xi_2 \overline{\log(enroll_i)} + v_{it1}$$

- And, fractional probit version of this.

```
. use meap92_01

. xtset distid year
       panel variable:  distid (strongly balanced)
        time variable:  year, 1992 to 2001
                delta:  1 unit

. des math4 avgrexp lunch enroll found

              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
math4           double %9.0g                  fraction satisfactory, 4th
                                                grade math
avgrexp         float  %9.0g                  (rexppp + rexppp_1 + rexppp_2 +
                                                rexppp_3)/4
lunch           float  %9.0g                  fraction eligible for free lunch
enroll          float  %9.0g                  district enrollment
found           int    %9.0g                  foundation grant, $: 1995-2001

. sum math4 rexppp lunch

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       math4 |      5010    .6149834    .1912023       .059          1
      rexppp |      5010     6331.99    1168.198   3553.361   15191.49
       lunch |      5010    .2802852    .1571325      .0087   .9126999
```

```
. xtreg math4 lavgrexp lunch lenroll y96-y01, fe cluster(distid)

Fixed-effects (within) regression               Number of obs      =      3507
Group variable: distid                          Number of groups   =       501

R-sq:  within  = 0.4713                         Obs per group: min =         7
       between = 0.0219                                        avg =       7.0
       overall = 0.2049                                        max =         7

                                                F(9,500)           =    171.93
corr(u_i, Xb)  = -0.1787                        Prob > F           =    0.0000

                               (Std. Err. adjusted for 501 clusters in distid)
------------------------------------------------------------------------------
             |               Robust
       math4 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lavgrexp |   .3770929   .0705668     5.34   0.000     .2384489    .5157369
       lunch |  -.0419467   .0731611    -0.57   0.567    -.1856877    .1017944
     lenroll |   .0020568   .0488107     0.04   0.966    -.0938426    .0979561
         y96 |  -.0155968   .0063937    -2.44   0.015    -.0281587   -.003035
         y97 |  -.0589732   .0095232    -6.19   0.000    -.0776837   -.0402628
         y98 |   .0781686   .0112949     6.92   0.000     .0559772    .1003599
         y99 |   .0642748   .0123103     5.22   0.000     .0400884    .0884612
         y00 |   .0895688   .0133223     6.72   0.000     .0633942    .1157434
         y01 |   .0630091    .014717     4.28   0.000     .0340943    .0919239
       _cons |  -2.640402   .8161357    -3.24   0.001     -4.24388   -1.036924
```

```
-------------+----------------------------------------------------------------
     sigma_u |   .1130256
     sigma_e |  .08314135
         rho |  .64888558   (fraction of variance due to u_i)
------------------------------------------------------------------------------

. des alavgrexp alunch alenroll

              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
alavgrexp       float  %9.0g                  time average lavgrexp, 1995-2001
alunch          float  %9.0g                  time average lunch, 1995-2001
alenroll        float  %9.0g                  time average lenroll, 1995-2001
```

```
. reg math4 lavgrexp alavgrexp lunch alunch lenroll alenroll y96-y01,
      cluster(distid)

Linear regression                                      Number of obs =    3507
                                                       F( 12,   500) =  161.09
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4218
                                                       Root MSE      =  .11542

                               (Std. Err. adjusted for 501 clusters in distid)
------------------------------------------------------------------------------
             |               Robust
       math4 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lavgrexp |    .377092   .0705971     5.34   0.000     .2383884    .5157956
   alavgrexp |   -.286541   .0731797    -3.92   0.000    -.4303185   -.1427635
       lunch |  -.0419466   .0731925    -0.57   0.567    -.1857494    .1018562
      alunch |  -.3770088   .0766141    -4.92   0.000    -.5275341   -.2264835
     lenroll |   .0020566   .0488317     0.04   0.966     -.093884    .0979972
   alenroll |  -.0031646   .0491534    -0.06   0.949    -.0997373    .0934082
         y96 |  -.0155968   .0063965    -2.44   0.015    -.0281641   -.0030295
         y97 |  -.0589731   .0095273    -6.19   0.000    -.0776916   -.0402546
         y98 |   .0781687   .0112998     6.92   0.000     .0559678    .1003696
         y99 |    .064275   .0123156     5.22   0.000     .0400782    .0884717
         y00 |    .089569    .013328     6.72   0.000     .0633831    .1157548
         y01 |   .0630093   .0147233     4.28   0.000     .0340821    .0919365
       _cons |  -.0006233   .2450239    -0.00   0.998    -.4820268    .4807801
------------------------------------------------------------------------------
```

```
. * Now use fractional probit.

. glm math4 lavgrexp alavgrexp lunch alunch lenroll alenroll y96-y01,
      fa(bin) link(probit) cluster(distid)
note: math4 has non-integer values

Generalized linear models                          No. of obs      =      3507
Optimization     : ML                              Residual df     =      3494
                                                   Scale parameter =         1
Deviance         =   237.643665                    (1/df) Deviance =  .0680148
Pearson          =  225.1094075                    (1/df) Pearson  =  .0644274

                               (Std. Err. adjusted for 501 clusters in distid)
------------------------------------------------------------------------------
             |               Robust
       math4 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lavgrexp |   .8810302   .2068026     4.26   0.000     .4757045    1.286356
   alavgrexp |  -.5814474   .2229411    -2.61   0.009    -1.018404   -.1444909
       lunch |  -.2189714   .2071544    -1.06   0.290    -.6249865    .1870437
      alunch |  -.9966635   .2155739    -4.62   0.000    -1.419181   -.5741465
     lenroll |   .0887804   .1382077     0.64   0.521    -.1821017    .3596626
    alenroll |  -.0893612   .1387674    -0.64   0.520    -.3613404    .1826181
         y96 |  -.0362309   .0178481    -2.03   0.042    -.0712125   -.0012493
         y97 |  -.1467327   .0273205    -5.37   0.000      -.20028   -.0931855
         y98 |   .2520084   .0337706     7.46   0.000     .1858192    .3181975
         y99 |   .2152507   .0367226     5.86   0.000     .1432757    .2872257
         y00 |   .3049632   .0399409     7.64   0.000     .2266805    .3832459
         y01 |   .2257321   .0439608     5.13   0.000     .1395705    .3118938
       _cons |  -1.855832   .7556621    -2.46   0.014    -3.336902   -.3747616
------------------------------------------------------------------------------
```

```
. margeff

Average partial effects after glm
      y  = Pr(math4)

------------------------------------------------------------------------------
    variable |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lavgrexp |   .2968496   .0695326     4.27   0.000     .1605682     .433131
   alavgrexp |  -.1959097   .0750686    -2.61   0.009    -.3430414   -.0487781
       lunch |  -.0737791   .0698318    -1.06   0.291    -.2106469    .0630887
      alunch |  -.3358104   .0723725    -4.64   0.000    -.4776579   -.1939629
     lenroll |   .0299132   .0465622     0.64   0.521     -.061347    .1211734
    alenroll |  -.0301089   .0467477    -0.64   0.520    -.1217326    .0615149
         y96 |  -.0122924   .0061107    -2.01   0.044    -.0242692   -.0003156
         y97 |  -.0508008   .0097646    -5.20   0.000     -.069939   -.0316625
         y98 |   .0809879   .0100272     8.08   0.000     .0613349    .1006408
         y99 |   .0696954   .0111375     6.26   0.000     .0478662    .0915245
         y00 |   .0970224   .0115066     8.43   0.000     .0744698     .119575
         y01 |   .0729829   .0132849     5.49   0.000      .046945    .0990208
------------------------------------------------------------------------------

. * These standard errors are very close to bootstrapped standard errors.
```

```
. xtgee math4 lavgrexp alavgrexp lunch alunch lenroll alenroll y96-y01,
        fa(bin) link(probit) corr(exch) robust

GEE population-averaged model                   Number of obs      =      3507
Group variable:                     distid      Number of groups   =       501
Link:                               probit      Obs per group: min =         7
Family:                           binomial                     avg =       7.0
Correlation:                  exchangeable                     max =         7
                                                Wald chi2(12)      =   1815.43
Scale parameter:                         1      Prob > chi2        =    0.0000

                                 (Std. Err. adjusted for clustering on distid)
------------------------------------------------------------------------------
             |             Semi-robust
       math4 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lavgrexp |    .884564   .2060662     4.29   0.000     .4806817    1.288446
   alavgrexp |  -.5835138   .2236705    -2.61   0.009      -1.0219   -.1451277
       lunch |  -.2372942   .2091221    -1.13   0.256    -.6471659    .1725775
      alunch |  -.9754696   .2170624    -4.49   0.000    -1.400904   -.5500351
     lenroll |   .0875629   .1387427     0.63   0.528    -.1843677    .3594935
    alenroll |  -.0820307   .1393712    -0.59   0.556    -.3551933    .1911318
         y96 |  -.0364771   .0178529    -2.04   0.041    -.0714681    -.001486
         y97 |  -.1471389   .0273264    -5.38   0.000    -.2006976   -.0935801
         y98 |   .2515377   .0337018     7.46   0.000     .1854833     .317592
         y99 |   .2148552   .0366599     5.86   0.000      .143003    .2867073
         y00 |   .3046286   .0399143     7.63   0.000     .2263981    .3828591
         y01 |   .2256619   .0438877     5.14   0.000     .1396437    .3116801
       _cons |  -1.914975   .7528262    -2.54   0.011    -3.390487   -.4394628
------------------------------------------------------------------------------
```

```
. margeff

Average partial effects after xtgee
      y  = Pr(math4)

------------------------------------------------------------------------------
    variable |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lavgrexp |   .2979576   .0692519     4.30   0.000     .1622263    .4336889
   alavgrexp |  -.1965515   .0752801    -2.61   0.009    -.3440978   -.0490052
       lunch |  -.0799305   .0704803    -1.13   0.257    -.2180693    .0582082
      alunch |  -.3285784   .0728656    -4.51   0.000    -.4713924   -.1857644
     lenroll |   .0294948   .0467283     0.63   0.528    -.0620909    .1210805
    alenroll |  -.0276313   .0469381    -0.59   0.556    -.1196283    .0643656
         y96 |   -.012373   .0061106    -2.02   0.043    -.0243497   -.0003964
         y97 |  -.0509306   .0097618    -5.22   0.000    -.0700633   -.0317979
         y98 |   .0808226    .010009     8.08   0.000     .0612054    .1004399
         y99 |   .0695541   .0111192     6.26   0.000     .0477609    .0913472
         y00 |   .0968972   .0115004     8.43   0.000     .0743568    .1194376
         y01 |   .0729416   .0132624     5.50   0.000     .0469478    .0989353
------------------------------------------------------------------------------
```

```
. * Now allow spending to be endogenous. Use foundation allowance, and
. * interactions, as IVs.
. * First, linear model:

. ivreg math4 lunch alunch lenroll alenroll y96-y01 lexppp94 le94y96-le94y01
        (lavgrexp = lfound lfndy96-lfndy01), cluster(distid)

Instrumental variables (2SLS) regression               Number of obs =    3507
                                                       F( 18,   500) =  107.05
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4134
                                                       Root MSE      =  .11635

                               (Std. Err. adjusted for 501 clusters in distid)
------------------------------------------------------------------------------
             |               Robust
       math4 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lavgrexp |   .5545247   .2205466     2.51   0.012     .1212123     .987837
       lunch |  -.0621991   .0742948    -0.84   0.403    -.2081675    .0837693
      alunch |  -.4207815   .0758344    -5.55   0.000    -.5697749   -.2717882
     lenroll |   .0463616   .0696215     0.67   0.506    -.0904253    .1831484
    alenroll |   -.049052    .070249    -0.70   0.485    -.1870716    .0889676
         y96 |  -1.085453   .2736479    -3.97   0.000    -1.623095   -.5478119
         y97 |  -1.049922    .376541    -2.79   0.005     -1.78972   -.3101244
         y98 |  -.4548311   .4958826    -0.92   0.359    -1.429102    .5194394
         y99 |  -.4360973   .5893671    -0.74   0.460    -1.594038    .7218439
         y00 |  -.3559283   .6509999    -0.55   0.585    -1.634961     .923104
         y01 |   -.704579   .7310773    -0.96   0.336    -2.140941    .7317831
    lexppp94 |  -.4343213   .2189488    -1.98   0.048    -.8644944   -.0041482
     le94y96 |   .1253255   .0318181     3.94   0.000     .0628119    .1878392
     le94y97 |     .11487   .0425422     2.70   0.007     .0312865    .1984534
     le94y98 |   .0599439   .0554377     1.08   0.280    -.0489757    .1688636
     le94y99 |   .0557854   .0661784     0.84   0.400    -.0742367    .1858075
     le94y00 |    .048899   .0727172     0.67   0.502    -.0939699    .1917678
```

```
     le94y01 |   .0865874   .0816732     1.06   0.290    -.0738776    .2470524
       _cons |   -.334823   .2593105    -1.29   0.197    -.8442955    .1746496
------------------------------------------------------------------------------
Instrumented:  lavgrexp
Instruments:   lunch alunch lenroll alenroll y96 y97 y98 y99 y00 y01
               lexppp94 le94y96 le94y97 le94y98 le94y99 le94y00 le94y01
               lfound lfndy96 lfndy97 lfndy98 lfndy99 lfndy00 lfndy01
------------------------------------------------------------------------------

. * Estimate is substantially larger than when spending is treated as exogenous.
```

```
. * Get reduced form residuals for fractional probit:

. reg lavgrexp lfound lfndy96-lfndy01 lunch alunch lenroll alenroll y96-y01
       lexppp94 le94y96-le94y01, cluster(distid)

Linear regression                                      Number of obs =    3507
                                                       F( 24,   500) = 1174.57
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.9327
                                                       Root MSE      =  .03987

                               (Std. Err. adjusted for 501 clusters in distid)
------------------------------------------------------------------------------
             |               Robust
    lavgrexp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lfound |   .2447063   .0417034     5.87   0.000     .1627709    .3266417
     lfndy96 |   .0053951   .0254713     0.21   0.832     -.044649    .0554391
     lfndy97 |  -.0059551   .0401705    -0.15   0.882    -.0848789    .0729687
     lfndy98 |   .0045356   .0510673     0.09   0.929    -.0957972    .1048685
     lfndy99 |   .0920788   .0493854     1.86   0.063    -.0049497    .1891074
     lfndy00 |   .1364484   .0490355     2.78   0.006     .0401074    .2327894
     lfndy01 |   .2364039   .0555885     4.25   0.000      .127188    .3456198
       ...
       _cons |   .1632959   .0996687     1.64   0.102    -.0325251     .359117
------------------------------------------------------------------------------

. predict v2hat, resid
(1503 missing values generated)
```

```
. glm math4 lavgrexp v2hat lunch alunch lenroll alenroll y96-y01 lexppp94
      le94y96-le94y01, fa(bin) link(probit) cluster(distid)
note: math4 has non-integer values

Generalized linear models                          No. of obs      =      3507
Optimization     : ML                              Residual df     =      3487
                                                   Scale parameter =         1
Deviance         =  236.0659249                    (1/df) Deviance =  .0676989
Pearson          =  223.3709371                    (1/df) Pearson  =  .0640582

Variance function: V(u) = u*(1-u/1)                [Binomial]
Link function    : g(u) = invnorm(u)               [Probit]

                               (Std. Err. adjusted for 501 clusters in distid)
------------------------------------------------------------------------------
             |               Robust
       math4 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lavgrexp |   1.731039   .6541194     2.65   0.008      .4489886    3.013089
       v2hat |  -1.378126    .720843    -1.91   0.056     -2.790952    .0347007
       lunch |  -.2980214   .2125498    -1.40   0.161     -.7146114    .1185686
      alunch |  -1.114775   .2188037    -5.09   0.000     -1.543623    -.685928
     lenroll |   .2856761    .197511     1.45   0.148     -.1014383    .6727905
    alenroll |  -.2909903   .1988745    -1.46   0.143     -.6807771    .0987966
         ...
       _cons |  -2.455592   .7329693    -3.35   0.001     -3.892185   -1.018998
------------------------------------------------------------------------------
```

```
. margeff

Average partial effects after glm
      y  = Pr(math4)

------------------------------------------------------------------------------
    variable |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lavgrexp |   .5830163   .2203345     2.65   0.008     .1511686    1.014864
       v2hat |  -.4641533    .242971    -1.91   0.056    -.9403678    .0120611
       lunch |  -.1003741   .0716361    -1.40   0.161    -.2407782      .04003
      alunch |  -.3754579   .0734083    -5.11   0.000    -.5193355   -.2315803
     lenroll |   .0962161   .0665257     1.45   0.148    -.0341719    .2266041
    alenroll |  -.0980059   .0669786    -1.46   0.143    -.2292817    .0332698
         ...
------------------------------------------------------------------------------

. * These standard errors do not account for the first-stage estimation. Should
. * use the panel bootstrap accounting for both stages.
. * Only marginal evidence that spending is endogenous, but the negative sign
. * fits the story that districts increase spending when performance is
. * (expected to be) worse, based on unobservables (to us).
```