Topics in Microeconometrics

William Greene
Department of Economics
Stern School of Business
Part 7: Sample Selection in Nonlinear and Panel Models
Samples and Populations

• Consistent estimation
  • The sample is randomly drawn from the population
  • Sample statistics converge to their population counterparts

• A presumption: The ‘population’ is the population of interest.

• Implication: If the sample is randomly drawn from a specific subpopulation, statistics converge to the characteristics of that subpopulation. These may not be the same as the full population

• Can one generalize from the subpopulation to the full population?
Nonrandom Sampling

- Simple nonrandom samples: Average incomes of airport travelers → mean income in the population as a whole?
- Self-selection:
  - Labor supply models
  - Shere Hite’s (1976) “The Hite Report” ‘survey’ of sexual habits of Americans. “While her books are ground-breaking and important, they are based on flawed statistical methods and one must view their results with skepticism.”
Heckman’s Canonical Model

A behavioral model:
Offered wage \( o^* = \beta + v \) (\( x = \text{age, experience, educ} \ldots \))
Reservation wage \( r^* = \delta + u \) (\( z = \text{age, kids, family stuff} \))
Labor force participation:
\[
\text{LFP} = 1 \text{ if } o^* \geq r^*, \ 0 \text{ otherwise}
\]
\[
\text{Prob}(\text{LFP}=1) = \Phi\left( (\beta - \delta) / \sqrt{\sigma_v^2 + \sigma_u^2} \right)
\]
Desired Hours \( = H^* = \gamma'w + \varepsilon \)
Actual Hours \( = H^* \) if LFP = 1
\( \text{unobserved if LFP} = 0 \)
\( \varepsilon \) and \( u \) are correlated. \( \varepsilon \) and \( v \) might be correlated.
What is \( \mathbb{E}[H^* | w, \text{LFP} = 1]? \) Not \( \gamma'w \).
Dueling Selection Biases – From two emails, same day.

• “I am trying to find methods which can deal with data that is non-randomised and suffers from selection bias.”

• “I explain the probability of answering questions using, among other independent variables, a variable which measures knowledge breadth. Knowledge breadth can be constructed only for those individuals that fill in a skill description in the company intranet. This is where the selection bias comes from.”
Sample Selection Observations

- The selection ‘problem’ is caused by the correlation of the unobservables
  - Selection on observables is often manageable within the conventional model.
  - Selection on unobservables often requires a more detailed specification of the model – where does the effect come from?
- The ‘bias’ relates to the inconsistency of familiar estimators such as least squares
- The data are not biased; the (an) estimator is biased.
Standard Sample Selection Model

\[ d_i^* = \alpha' z_i + u_i \]
\[ d_i = 1(d_i^* > 0) \]
\[ y_i^* = \beta' x_i + \varepsilon_i \]
\[ y_i = y_i^* \text{ when } d_i = 1, \text{ unobserved otherwise} \]

\[(u_i, v_i) \sim \text{Bivariate Normal}[(0,0),(1, \rho \sigma, \sigma^2)]\]

\[ E[y_i | y_i \text{ is observed}] = E[y_i | d_i = 1] \]
\[ = \beta' x_i + E[\varepsilon_i | d_i = 1] \]
\[ = \beta' x_i + E[\varepsilon_i | u_i > -\alpha' z_i] \]
\[ = \beta' x_i + (\rho \sigma) \frac{\phi(\alpha' z_i)}{\Phi(\alpha' z_i)} \]
\[ = \beta' x_i + \theta \lambda_i \]
Incidental Truncation

\( u_1, u_2 \sim N[(0,0),(1,.71,1)] \)

Unconditional distribution of \( u_2 \sim N[0,1] \)

Conditional distribution of \( u_2|u_1 > 0 \). No longer \( \sim N[0,1] \)
Selection as a Specification Error

- \( \mathbb{E}[y_i|x_i, y_i \text{ observed}] = \beta'x_i + \theta \lambda_i \)
- Regression of \( y_i \) on \( x_i \) omits \( \lambda_i \).
  - \( \lambda_i \) will generally be correlated with \( x_i \) if \( z_i \) is.
  - \( z_i \) and \( x_i \) often have variables in common.
  - There is no specification error if \( \theta = 0 \iff \rho = 0 \)
- The “selection bias” is \( \text{plim} (\beta - \beta) \)
Estimation of the Selection Model

- **Two step least squares**
  - Inefficient
  - Simple – exists in current software
  - Simple to understand and widely used

- **Full information maximum likelihood**
  - Efficient. Not more or less robust
  - Simple to do – exists in current software
  - Not so simple to understand – widely misunderstood
Estimation

Heckman’s two step procedure

• (1) Estimate the probit model and compute $\lambda_i$ for each observation using the estimated parameters.

• (2) a. Linearly regress $y_i$ on $x_i$ and $\lambda_i$ using the observed data

        b. Correct the estimated asymptotic covariance matrix for the use of the estimated $\lambda_i$. (An application of Murphy and Topel (1984) – Heckman was 1979).
Mroz Application – Labor Supply

MROZ labor supply data. Cross section, 753 observations
Use LFP for binary choice, KIDS for count models.
LFP  = labor force participation, 0 if no, 1 if yes.
WHRS = wife's hours worked. 0 if LFP=0
KL6  = number of kids less than 6
K618 = kids 6 to 18
WA   = wife's age
WE   = wife's education
WW   = wife's wage, 0 if LFP=0.
RPWG = Wife's reported wage at the time of the interview
HHRS = husband's hours
HA   = husband's age
HE   = husband's education
HW   = husband's wage
FAMINC = family income
MTR  = marginal tax rate
WMED = wife's mother's education
WFED = wife's father's education
UN   = unemployment rate in county of residence
CIT  = dummy for urban residence
AX   = actual years of wife's previous labor market experience
AGE  = Age
AGESQ = Age squared
EARNINGS= WW * WHRS
LOGE = Log of EARNINGS
KIDS = 1 if kids < 18 in the home.
Labor Supply Model

NAMELIST ; Z = One,KL6,K618,WA,WE,HA,HE $
NAMELIST ; X = One,KL6,K618,Age,Agesq,WE,Faminc $
PROBIT ; Lhs = LFP ; Rhs = Z ; Hold(IMR=Lambda) $
SELECT ; Lhs = WHRS ; Rhs = X $
REGRESS ; Lhs = WHRS ; Rhs = X, Lambda $
REJECT ; LFP = 0 $
REGRESS ; Lhs = WHRS ; Rhs = X $
### Participation Equation

+---------------------------------------------+
| Binomial Probit Model                        |
| Dependent variable                           LFP |
| Weighting variable                           None |
| Number of observations                       753  |
+---------------------------------------------+

| Variable | Coefficient | Standard Error | b/St.Er. | P[|Z|>z] | Mean of X |
|----------|-------------|----------------|----------|--------|-----------|
| Index function for probability               |             |               |          |        |           |
| Constant | 1.00264501  | 0.49994379     | 2.006    | .0449  |           |
| KL6      | -.90399802  | 0.11434394     | -7.906   | .0000  | .23771580 |
| K618     | -.05452607  | 0.04021041     | -1.356   | .1751  | 1.35325365 |
| WA       | -.02602427  | 0.01332588     | -1.953   | .0508  | 42.5378486 |
| WE       | 0.16038929  | 0.02773622     | 5.783    | .0000  | 12.2868526 |
| HA       | -.01642514  | 0.01329110     | -1.236   | .2165  | 45.1208499 |
| HE       | -.05191039  | 0.02040378     | -2.544   | .0110  | 12.4913679 |
## Hours Equation

Sample Selection Model

<table>
<thead>
<tr>
<th>Two stage least squares regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>LHS=WHRS  Mean  =  1302.930</td>
</tr>
<tr>
<td>Standard deviation  =  776.2744</td>
</tr>
<tr>
<td>WTS=none  Number of observs.  =  428</td>
</tr>
<tr>
<td>Model size  Parameters  =  8</td>
</tr>
<tr>
<td>Degrees of freedom  =  420</td>
</tr>
<tr>
<td>Residuals  Sum of squares  =  2.267214E+09</td>
</tr>
<tr>
<td>Standard error of e  =  734.7195</td>
</tr>
<tr>
<td>Correlation of disturbance in regression and Selection Criterion (Rho) ...........  -0.84541</td>
</tr>
</tbody>
</table>

| Variable | Coefficient | Standard Error | b/St.Er. | P[|Z|>z] | Mean of X |
|----------|-------------|----------------|---------|----------|-----------|
| Constant | 2442.26665  | 1202.11143     | 2.032   | 0.0422   |           |
| KL6      | 115.109657  | 282.008565     | 0.408   | 0.6831   | 0.14018692 |
| K618     | -101.720762 | 38.2833942     | -2.657  | 0.0079   | 1.35046729 |
| AGE      | 14.6359451  | 53.1916591     | 0.275   | 0.7832   | 41.9719626 |
| AGESQ    | -0.10078602 | 0.61856252     | -0.163  | 0.8706   | 1821.12150 |
| WE       | -102.203059 | 39.4096323     | -2.593  | 0.0095   | 12.6588785 |
| FAMINC   | 0.01379467  | 0.00345041     | 3.998   | 0.0001   | 24130.4229 |
| LAMBDA   | -793.857053 | 494.541008     | -1.605  | 0.1084   | 61466207   |
## Selection “Bias” of OLS

| Variable | Coefficient | Standard Error | b/St.Er. | P[|Z|>z] | Mean of X |
|----------|-------------|----------------|---------|---------|-----------|
| Constant | 2442.26665  | 1202.11143     | 2.032   | .0422   |           |
| KL6      | 115.109657  | 282.008565     | .408    | .6831   | .14018692 |
| K618     | -101.720762 | 38.2833942     | -2.657  | .0079   | 1.35046729 |
| AGE      | 14.6359451  | 53.1916591     | .275    | .7832   | 41.9719626 |
| AGESQ    | -.10078602  | .61856252      | -.163   | .8706   | 1821.12150 |
| WE       | -102.203059 | 39.4096323     | -2.593  | .0095   | 12.6588785 |
| FAMINC   | .01379467   | .00345041      | 3.998   | .0001   | 24130.4229 |
| LAMBDA   | -793.857053 | 494.541008     | -1.605  | .1084   | .61466207 |

| Variable | Coefficient | Standard Error | t-ratio | P[|T|>t] | Mean of X |
|----------|-------------|----------------|---------|---------|-----------|
| Constant | 1812.12538  | 1144.33342     | 1.584   | .1140   |           |
| KL6      | -299.128041 | 100.033124     | -2.990  | .0030   | .14018692 |
| K618     | -126.399697 | 30.8728451     | -4.094  | .0001   | 1.35046729 |
| AGE      | 11.2795338  | 53.8442084     | -.209   | .8342   | 41.9719626 |
| AGESQ    | -.26103541  | .62632815      | -.417   | .6771   | 1821.12150 |
| WE       | -47.3271780 | 17.2968137     | -2.736  | .0065   | 12.6588785 |
| FAMINC   | .01261889   | .00338906      | 3.723   | .0002   | 24130.4229 |
Maximum Likelihood Estimation

\[ \log L = \sum_{d=1} \log \left[ \frac{\exp\left( -\frac{1}{2} (\varepsilon_i / \sigma)^2 \right)}{\sigma \sqrt{2\pi}} \Phi \left( \frac{\rho (\varepsilon_i / \sigma) + \alpha' z_i}{\sqrt{1 - \rho^2}} \right) \right] + \sum_{d=0} \log \left[ 1 - \Phi(\alpha' z_i) \right] \]

Reparameterize this: let \( q_i = \alpha' z_i \)

(1) \( \theta = 1/\sigma \)

(2) \( \gamma = \beta/\sigma \) (Olsen transformation)

(3) \( \tau = \rho / \sqrt{1 - \rho^2} \)

(4) Constrain \( \rho \) to be in (-1,1) by using

\( \psi = \frac{1}{2} \ln \left( \frac{1+\rho}{1-\rho} \right) = \tanh \rho \), so \( \rho = \tanh^{-1}(\psi) = \frac{\exp(2\psi) - 1}{\exp(2\psi) + 1} \)

\[ \log L = \sum_{d=0} \log \Phi(-q_i) + \sum_{d=1} \log \theta - \frac{1}{2} \log 2\pi - \frac{1}{2} (\theta y_i - \gamma' x_i)^2 + \log \Phi[\tau(\theta y_i - \gamma' x_i) + q_i \sqrt{1 + \tau^2}] \]
| ML Estimates of Selection Model | |
| Maximum Likelihood Estimates | |
| Number of observations | 753 |
| Iterations completed | 47 |
| Log likelihood function | -3894.471 |
| Number of parameters | 16 |
| FIRST 7 estimates are probit equation. |

| Variable | Coefficient | Standard Error | b/St.Er. | P[|Z|>z] |
|---|---|---|---|---|
| Constant | 1.01350651 | .54823177 | 1.849 | .0645 |
| KL6 | -.90129694 | .11081111 | -8.134 | .0000 |
| K618 | -.05292375 | .04137216 | -1.279 | .2008 |
| WA | -.02491779 | .01428642 | -1.744 | .0811 |
| WE | .16396194 | .02911763 | 5.631 | .0000 |
| HA | -.01763340 | .01431873 | -1.231 | .2181 |
| HE | -.0559671 | .02133647 | -2.623 | .0087 |

Selection (probit) equation for LFP

Corrected regression, Regime 1

| Constant | 1946.84517 | 1167.56008 | 1.667 | .0954 |
| KL6 | -209.024866 | 222.027462 | -.941 | .3465 |
| K618 | -120.969192 | 35.4425577 | -3.413 | .0006 |
| AGE | 12.0375636 | 51.9850307 | .232 | .8169 |
| AGESQ | -.22652298 | .59912775 | -.378 | .7054 |
| WE | -59.2166488 | 33.3802882 | -1.774 | .0761 |
| FAMINC | .01289491 | .00332219 | 3.881 | .0001 |
| SIGMA(1) | 748.131644 | 59.7508375 | 12.521 | .0000 |
| RHO(1,2) | -.22965163 | .50082203 | -.459 | .6466 |
# MLE vs. Two Step

## Two Step

<table>
<thead>
<tr>
<th></th>
<th>Parameter</th>
<th>Value</th>
<th>Standard Error</th>
<th>T-Value</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td></td>
<td>2442.26665</td>
<td>1202.11143</td>
<td>2.032</td>
<td>.0422</td>
</tr>
<tr>
<td>KL6</td>
<td></td>
<td>115.109657</td>
<td>282.008565</td>
<td>.408</td>
<td>.6831</td>
</tr>
<tr>
<td>K618</td>
<td></td>
<td>-101.720762</td>
<td>38.2833942</td>
<td>-2.657</td>
<td>.0079</td>
</tr>
<tr>
<td>AGE</td>
<td></td>
<td>14.6359451</td>
<td>53.1916591</td>
<td>.275</td>
<td>.7832</td>
</tr>
<tr>
<td>AGESQ</td>
<td></td>
<td>-.10078602</td>
<td>.61856252</td>
<td>-.163</td>
<td>.8706</td>
</tr>
<tr>
<td>WE</td>
<td></td>
<td>-102.203059</td>
<td>39.4096323</td>
<td>-2.593</td>
<td>.0095</td>
</tr>
<tr>
<td>FAMINC</td>
<td></td>
<td>.01379467</td>
<td>.00345041</td>
<td>3.998</td>
<td>.0001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Standard error of e</th>
<th>734.7195</th>
</tr>
</thead>
<tbody>
<tr>
<td>RHO(1,2)</td>
<td>Correlation of disturbance in regression</td>
<td>-.84541</td>
</tr>
</tbody>
</table>

## MLE

<table>
<thead>
<tr>
<th></th>
<th>Parameter</th>
<th>Value</th>
<th>Standard Error</th>
<th>T-Value</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td></td>
<td>1946.84517</td>
<td>1167.56008</td>
<td>1.667</td>
<td>.0954</td>
</tr>
<tr>
<td>KL6</td>
<td></td>
<td>-209.024866</td>
<td>222.027462</td>
<td>-.941</td>
<td>.3465</td>
</tr>
<tr>
<td>K618</td>
<td></td>
<td>-120.969192</td>
<td>35.4425577</td>
<td>-3.413</td>
<td>.0006</td>
</tr>
<tr>
<td>AGE</td>
<td></td>
<td>12.0375636</td>
<td>51.9850307</td>
<td>.232</td>
<td>.8169</td>
</tr>
<tr>
<td>AGESQ</td>
<td></td>
<td>-.22652298</td>
<td>.59912775</td>
<td>-.378</td>
<td>.7054</td>
</tr>
<tr>
<td>WE</td>
<td></td>
<td>-59.2166488</td>
<td>33.3802882</td>
<td>-1.774</td>
<td>.0761</td>
</tr>
<tr>
<td>FAMINC</td>
<td></td>
<td>.01289491</td>
<td>.00332219</td>
<td>3.881</td>
<td>.0001</td>
</tr>
</tbody>
</table>

|        | SIGMA(1)    | 748.131644 | 59.7508375     | 12.521  | .0000   |
| RHO(1,2)| -.22965163  | .50082203  | -.459          | .6466   |
Extension – Treatment Effect

What is the value of an elite college education?

\[ d_i^* = z_i' y + u_i; d_i = 1[d_i^* > 0] \text{ (probit)} \]

\[ y_i^* = x_\beta + \delta d_i + \epsilon_i \text{ observed for everyone} \]

\[ [\epsilon_i, u_i] \sim \text{Bivariate Normal}[0, 0, \sigma^2, \rho, 1] \]

\[ E[y_i^* | x_i, d_i = 1] = x_\beta + \delta + E[\epsilon_i | x_i, d_i = 1] \]

\[ = x_\beta + \delta + E[\epsilon_i | x_i, u_i > -z_i' y] \]

\[ = x_\beta + \delta + \rho \sigma \left( \frac{\phi(z_i' y)}{\Phi(z_i' y)} \right) \]

\[ = x_\beta + \delta + \rho \sigma \lambda_i \]

\[ E[y_i^* | x_i, d_i = 0] = x_\beta + E[\epsilon_i | x_i, d_i = 0] \]

\[ = x_\beta + \rho \sigma \left( \frac{-\phi(-z_i' y)}{\Phi(-z_i' y)} \right) \]

Least squares is still biased and inconsistent. Left out variable
Treatment Effect

\[ E[y_{i^*}|x_i,d_i=1] - E[y_{i^*}|x_i,d_i=0] \]

\[ = x\beta + \delta + \rho \sigma \left( \frac{\phi(z_{iY})}{\Phi(z_{iY})} \right) \]

\[ - x\beta - \rho \sigma \left( \frac{-\phi(-z_{iY})}{\Phi(-z_{iY})} \right) \]

\[ = \delta + \rho \sigma \left[ \left( \frac{\phi(z_{iY})}{\Phi(z_{iY})} \right) - \left( \frac{-\phi(-z_{iY})}{\Phi(-z_{iY})} \right) \right] \]

\[ = \text{Treatment} + \text{Selection Effect} \]
Sample Selection in Exponential Regression

An approach modeled on Heckman's model

Regression Equation: \( \text{Prob}[y=j|x,u]=P(\lambda); \)
\[ \lambda=\exp(x\beta + \theta u) \]

Selection Equation: \( d=1[u+z\delta >0] \) (The usual probit)

\([u,\varepsilon] \sim n[0,0,1,1,\rho] \) (Var\([u]\) is absorbed in \(\theta\))


\[ E[y|x,d=1]=\exp(x\beta + \theta u) \frac{\Phi(z\delta + \rho)}{\Phi(z\delta)} \]
Panel Data and Selection

Selection equation with time invariant individual effect
\[ d_{it} = 1[z_{it} + \theta_i + \eta_{it} > 0] \]
Observation mechanism: \((y_{it}, x_{it})\) observed when \(d_{it} = 1\)

Primary equation of interest
Common effects linear regression model
\[ y_{it} \mid (d_{it} = 1) = x_{it} \beta + \alpha_i + \varepsilon_{it} \]
"Selectivity" as usual arises as a problem when the unobservables are correlated; \(\text{Corr}(\varepsilon_{it}, \eta_{it}) \neq 0\).
The common effects, \(\theta_i\) and \(\alpha_i\) make matters worse.
Panel Data and Sample Selection Models: A Nonlinear Time Series


II. 1995 and 2005: Model Identification through Conditional Mean Assumptions

III. 1997-2005: Semiparametric Approaches based on Differences and Kernel Weights

IV. 2007: Return to Conventional Estimators, with Bias Corrections
Panel Data Sample Selection Models


\( d_{it} = 1[ z_i' y_t + w_i + \eta_{it} > 0 ] \) (Random effects probit)

\( y_{it} | (d_{it} = 1) = x_i' \beta + \alpha_i + \varepsilon_{it} \) (Fixed effects regression)

Proposed "marginal likelihood" based on joint normality

\[
\log L_i = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{t=1}^{T_i} \Phi \left[ (2d_{it} - 1) \frac{z_i'y_t + \Delta_{it} + u_{i,1} + d_{it}u_{i,2}}{\sqrt{\sigma_n^2 (1 - d_{it} \rho^2)}} \right] f(u_{i,1}, u_{i,2}) du_{i,1} du_{i,2}
\]

\( \Delta_{it} = (\rho / \sigma_{\varepsilon}) d_{it} \left[ (y_{it} - \bar{y}_i) - (x_{it} \bar{\beta} \bar{x}_i)' \right] \)

(Integrate out the random effects; difference out the fixed effects.)

\( u_{i,1}, u_{i,2} \) are time invariant uncorrelated standard normal variables

How to do the integration? Natural candidate for simulation.

(Not mentioned in the paper. Too early.)

Zabel – Economics Letters

• Inappropriate to have a mix of FE and RE models
• Two part solution
  • Treat both effects as “fixed”
  • Project both effects onto the group means of the variables in the equations (Mundlak approach)
  • Resulting model is two random effects equations
• Use both random effects
Selection with Fixed Effects

\[ y_{it}^* = \eta_i + x_{it}' \beta + \varepsilon_{it}, \quad \eta_i = \bar{x}_i' \pi + \tau \omega_i, \quad \omega_i \sim N[0,1] \]
\[ d_{it}^* = \theta_i + z_{it}' \alpha + u_{it}, \quad \theta_i = \bar{z}_i' \delta + \omega \nu_i, \quad \nu_i \sim N[0,1] \]
\[ (\varepsilon_{it}, u_{it}) \sim N_2[(0, 0), (\sigma^2, 1, \rho \sigma)]. \]

\[ L_i = \int_{-\infty}^{\infty} \prod_{d_{it}=0} \Phi \left[ -z_{it}' \alpha - \bar{z}_i' \delta - \omega \nu_i \right] \phi(\nu_i) d\nu_i \]
\[ \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{d_{it}=1} \left( \Phi \left[ \frac{z_{it}' \alpha + \bar{z}_i' \delta + \omega \nu_i + (\rho / \sigma) \varepsilon_{it}}{\sqrt{1-\rho^2}} \right] \right) d\nu_i d\omega_i \]
\[ \varepsilon_{it} = y_{it} - x_{it}' \beta - \bar{x}_i' \pi - \tau \omega_i \]
Practical Complications

The bivariate normal integration is actually the product of two univariate normals, because in the specification above, $v_i$ and $w_i$ are assumed to be uncorrelated. Vella notes, however, “… given the computational demands of estimating by maximum likelihood induced by the requirement to evaluate multiple integrals, we consider the applicability of available simple, or two step procedures.”
Simulation

The first line in the log likelihood is of the form \( E_v[\prod_{d=0} \Phi(\ldots)] \) and the second line is of the form \( E_w[E_v[\Phi(\ldots)\phi(\ldots)/\sigma]] \). Using simulation instead, the simulated likelihood is

\[
L_i^S = \frac{1}{R} \sum_{r=1}^{R} \prod_{d_{it}=0} \Phi \left[ -z_i^' \alpha - \bar{z}_i^' \delta - \omega v_{i,r} \right]
\]

\[
\times \frac{1}{R} \sum_{r=1}^{R} \prod_{d_{it}=1} \Phi \left[ z_i^' \alpha + \bar{z}_i^' \delta + \omega v_{i,r} + (\rho / \sigma) \varepsilon_{i,r} \right] \frac{1}{\sigma} \phi \left( \frac{\varepsilon_{i,r}}{\sigma} \right)
\]

\[
\varepsilon_{i,r} = y_{it} - x_i^' \beta - \bar{x}_i^' \pi - \tau w_{i,r}
\]
Correlated Effects

Suppose that $w_i$ and $v_i$ are bivariate standard normal with correlation $\rho_{vw}$. We can project $w_i$ on $v_i$ and write

$$w_i = \rho_{vw}v_i + (1-\rho_{vw}^2)^{1/2}h_i$$

where $h_i$ has a standard normal distribution. To allow the correlation, we now simply substitute this expression for $w_i$ in the simulated (or original) log likelihood, and add $\rho_{vw}$ to the list of parameters to be estimated. The simulation is then over still independent normal variates, $v_i$ and $h_i$. 
Conditional Means

Wooldridge (1995) proposes an estimator that can be based on straightforward applications of conventional, everyday methods.

\[ y_{it}^{*} = \eta_{i} + x_{it}' \beta + \varepsilon_{it}, \]

\[ d_{it}^{*} = \theta_{i} + z_{it}' \alpha + u_{it}, \]

\[ (\varepsilon_{it}, u_{it}) \sim N_{2}[(0, 0), (\sigma^{2}, 1, \rho\sigma)]. \]

Under the mean independence assumption

\[ E[\varepsilon_{it} | \eta_{i}, \theta_{i}, z_{i1}, \ldots, z_{iT}, v_{i1}, \ldots, v_{iT}, d_{i1}, \ldots, d_{iT}] = \rho u_{it}, \quad v_{it} = \theta_{i} + u_{it} \]

\[ E[y_{it} | x_{i1}, \ldots, x_{iT}, \eta_{i}, \theta_{i}, z_{i1}, \ldots, z_{iT}, v_{i1}, \ldots, v_{iT}, d_{i1}, \ldots, d_{iT}] = \eta_{i} + x_{it}' \beta + \rho u_{it}. \]

This suggests an approach to estimating the model parameters, however it requires computation of \( u_{it} \). That would require estimation of \( \theta_{i} \) which cannot be done, at least not consistently – and that precludes simple estimation of \( u_{it} \).
A Feasible Estimator

To escape the dilemma, Wooldridge suggests Chamberlain’s approach to the fixed effects model,

\[ \Theta_i = \beta_0 + z_{i1}' \beta_1 + z_{i2}' \beta_2 + \ldots + z_{iT}' \beta_T + \epsilon_i. \]

With this substitution,

\[ d_{it}^* = z_{it}' \alpha + \bar{\epsilon}_0 + z_{i1}' \bar{\epsilon}_1 + z_{i2}' \bar{\epsilon}_2 + \ldots + z_{iT}' \bar{\epsilon}_T + \bar{\epsilon}_i + u_{it} = z_{it}' \alpha + \bar{\epsilon}_0 + z_{i1}' \bar{\epsilon}_1 + z_{i2}' \bar{\epsilon}_2 + \ldots + z_{iT}' \bar{\epsilon}_T + w_{it} \]

where \( w_{it} \) is independent of \( z_{it}, t = 1, \ldots, T \). This now implies that

\[
E[y_{it}|x_{i1}, \ldots, x_{iT}, \eta_i, \Theta_i, z_{i1}, \ldots, z_{iT}, v_{i1}, \ldots, v_{iT}, d_{i1}, \ldots, d_{iT}] = \eta_i + x_{it}' \beta + \rho(w_{it} - h_i) = (\eta_i - \rho h_i) + x_{it}' \beta + \rho w_{it}.
\]
Estimation

To complete the estimation procedure, we now compute $T$ cross sectional probit models (reestimating $f_0, f_1, \ldots$ each time), and compute $\hat{\lambda}_{it}$ from each one. The resulting equation,

$$y_{it} = a_i + x_{it}'\beta + \rho \hat{\lambda}_{it} + v_{it}$$

now forms the basis for estimation of $\beta$ and $\rho$ by using a conventional fixed effects linear regression with the observed data.
Kyriazidou - Semiparametrics

Assume 2 periods
Estimate selection equation by FE logit
Use first differences and weighted least squares:

\[ \hat{\beta} = \left[ \sum_{i=1}^{N} d_{i1} d_{i2} \Psi_i \Delta x_i \Delta x_i' \right]^{-1} \left[ \sum_{i=1}^{N} d_{i1} d_{i2} \hat{\Psi}_i \Delta x_i \Delta y_i \right] \]

\[ \hat{\Psi}_i = \frac{1}{h} K \left( \frac{\Delta w'_i \hat{\alpha}}{h} \right) \text{ kernel function.} \]

Use with longer panels - any pairwise differences
Extensions based on pairwise differences by Rochina-Barrachina and Dustman/Rochina-Barrachina (1999)
Bias Corrections

- Val and Vella, 2007 (Working paper)
- Assume fixed effects
  - Bias corrected probit estimator at the first step
  - Use fixed probit model to set up second step Heckman style regression treatment.
Postscript

• What selection process is at work?
  • All of the work examined here (and in the literature) assumes the selection operates anew in each period
  • An alternative scenario: Selection into the panel, once, at baseline.
  • Alternative: Sequential selection = endogenous attrition (Wooldridge 2002, inverse probability weighting)

• Why aren’t the time invariant components correlated?

• Other models
  • All of the work on panel data selection assumes the main equation is a linear model.
  • Any others? Discrete choice? Counts?
Attrition

- In a panel, \( t=1, \ldots, T \) individual \( I \) leaves the sample at time \( K_i \) and does not return.
- If the determinants of attrition (especially the unobservables) are correlated with the variables in the equation of interest, then the now familiar problem of sample selection arises.
Dealing with Attrition in a QOL Study

• The attrition issue: Appearance for the second interview was low for people with initial low QOL (death or depression) or with initial high QOL (don’t need the treatment). Thus, missing data at exit were clearly related to values of the dependent variable.

• Solutions to the attrition problem
  • Heckman selection model (used in the study)
    □ Prob[Present at exit|covariates] = \( \Phi(z'\theta) \) (Probit model)
    □ Additional variable added to difference model \( \lambda_i = \Phi(z_i'\theta)/\Phi(z_i'\theta) \)
  • The FDA solution: fill with zeros. (!)
An Early Attrition Model


A two period model:
Structural response model (Random Effects Regression)
\[ y_{i1} = x_{i1}' \beta_1 + \varepsilon_{i1} + u_i \]
\[ y_{i2} = x_{i2}' \beta_2 + \varepsilon_{i2} + u_i \]

Attrition model for observation in the second period (Probit)
\[ z_{i2}^* = \delta y_{i2} + x_{i2}' \theta + w_{i2} + v_{i2} \]
\[ z_{i2} = 1(z_{i2}^* > 0) \]

Endogeneity "problem"
\[ \rho_{12} = \text{Corr}[\varepsilon_{i1} + u_i, \varepsilon_{i2} + u_i] = \sigma_u^2 / (\sigma_{\varepsilon}^2 + \sigma_u^2) \]
\[ \tau = \text{Corr}[v_{i2}, \varepsilon_{i2} + u_i] = \text{Corr}[v_{i2} + \delta(\varepsilon_{i2} + u_i), \varepsilon_{i2} + u_i] \]
Methods of Estimating the Attrition Model

- Heckman style “selection” model
- Two step maximum likelihood
- Full information maximum likelihood
- Two step method of moments estimators
- Weighting schemes that account for the “survivor bias”
Selection Model

Reduced form probit model for second period observation equation

\[ z_{i2}^* = x'_{i2} (\theta + \delta \beta) + w_i \alpha + \delta (\varepsilon_{i2} + u_i + v_i) \]
\[ = r'_{i2} \gamma + h_{i2} \]
\[ z_{i2} = 1(z_{i2}^* > 0) \]

Conditional means for observations observed in the second period

\[ E[y_{i2} | x_{i2}, z_{i2} = 1] = x'_{i2} \beta + (\rho_{12} \sigma_{\varepsilon}) \frac{\phi(r'_{i2} \gamma)}{\Phi(r'_{i2} \gamma)} \]

First period conditional means for observations observed in the second period

\[ E[y_{i1} | x_{i1}, z_{i2} = 1] = x'_{i1} \beta + (\rho_{12} \tau \sigma_{\varepsilon}) \frac{\phi(r'_{i2} \gamma)}{\Phi(r'_{i2} \gamma)} \]

(1) Estimate probit equation
(2) Combine these two equations with a period dummy variable, use OLS with a constructed regressor in the second period

THE TWO DISTURBANCES ARE CORRELATED.
TREAT THIS IS A SUR MODEL. (EQUIVALENT TO MDE)
Maximum Likelihood

$$\text{LogL}_i = \frac{-\log 2\pi}{2} - \log \sigma_\varepsilon - \frac{(y_{i1} - x_{i1}'\beta)^2}{2\sigma_\varepsilon^2}$$

$$\begin{bmatrix}
\log \sigma_\varepsilon + \log \sqrt{1 - \rho_{12}^2} - \frac{[(y_{i2} - \rho_{12}y_{i1}) - (x_{12} - \rho_{12}x_{i1})'\beta]_2}{2\sigma_\varepsilon^2(1 - \rho_{12}^2)} \\
+ z_{i2} \left[ \log \Phi \left( \frac{r_{i2}'\gamma + (\tau / \sigma_\varepsilon)(y_{i2} - x_{i2}'\beta)}{\sqrt{1 - \tau^2}} \right) \right] \\
+ (1 - z_{i2}) \left[ \log \Phi \left( - \frac{r_{i2}'\gamma + (\rho_{12}\tau / \sigma_\varepsilon)(y_{i1} - x_{i1}'\beta)}{\sqrt{1 - \rho_{12}^2\tau^2}} \right) \right]
\end{bmatrix}$$

(1) See H&W for FIML estimation
(2) Use the invariance principle to reparameterize
(3) Estimate $\gamma$ separately and use a two step ML with Murphy and Topel correction of asymptotic covariance matrix.
### Table IVa

**Parameter Estimates of the Earnings Function Structural Model With and Without a Correction for Attrition**

<table>
<thead>
<tr>
<th>Variables</th>
<th>With attrition correction: maximum likelihood estimates (standard errors)</th>
<th>Without attrition correction: generalized least squares estimates (standard errors)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Earnings function parameters</td>
<td>Attrition parameters</td>
</tr>
<tr>
<td>Constant</td>
<td>5.8539 (.0903)</td>
<td>-.6347 (.3351)</td>
</tr>
<tr>
<td>Experimental effect</td>
<td>-.0822 (.0402)</td>
<td>.2414 (.1211)</td>
</tr>
<tr>
<td>Time effect</td>
<td>.0940 (.0520)</td>
<td>—</td>
</tr>
<tr>
<td>Education</td>
<td>.0209 (.0052)</td>
<td>-.0204 (.0061)</td>
</tr>
<tr>
<td>Experience</td>
<td>.0037 (.0013)</td>
<td>-.0038 (.0061)</td>
</tr>
<tr>
<td>Income</td>
<td>-.0131 (.0050)</td>
<td>.1752 (.0470)</td>
</tr>
<tr>
<td>Union</td>
<td>.2159 (.0362)</td>
<td>1.4290 (.1252)</td>
</tr>
<tr>
<td>Poor health</td>
<td>-.0601 (.0330)</td>
<td>.2480 (.1237)</td>
</tr>
</tbody>
</table>

\[ \hat{\sigma}_{\eta}^2 = .1832, \quad l^* = 64.35, \quad \hat{\sigma}_{\eta}^2 = .1236 \]
\[ \hat{\rho}_{12} = .2596, \quad \rho_{23} = -.1089, \quad \hat{\rho}_{12} = .2003 \]
A Model of Attrition

- Nijman and Verbeek, Journal of Applied Econometrics, 1992
- Consumption survey (Holland, 1984 – 1986)
  - Exogenous selection for participation (rotating panel)
  - Voluntary participation (missing not at random – attrition)
Attrition Model

The main equation

\[ y_{i,t} = \beta_0 + x'_{i,t} \beta + \tilde{\alpha}_i + \varepsilon_{i,t}, \]

Random effects consumption function

\[ \tilde{\alpha}_i = \bar{x}' \theta + u_i, \]

Mundlak device; \( u_i \) uncorrelated with \( X_i \)

\[ y_{i,t} = \beta_0 + x'_{i,t} \beta + \bar{x}' \theta + u_i + \varepsilon_{i,t}, \]

Reduced form random effects model

The selection mechanism

\( a_{it} = 1 \) [individual \( i \) asked to participate in period \( t \)] Purely exogenous

\( a_{it} \) may depend on observables, but does not depend on unobservables

\( r_{it} = 1 \) [individual \( i \) chooses to participate if asked] Endogenous.

\( r_{it} \) is the endogenous participation dummy variable

\[ a_{it} = 0 \Rightarrow r_{it} = 0 \]

\[ a_{it} = 1 \Rightarrow \) the selection mechanism operates
Selection Equation

The main equation
\[ y_{i,t} = \beta_0 + x'_{i,t}\theta + \theta'\theta + u_i + \varepsilon_{i,t}, \]
Reduced form random effects model

The selection mechanism
\[ r_{i,t} = 1[i\text{ndividual }i\text{ chooses to participate if asked}] \quad \text{Endogenous.} \]
\[ r_{i,t} \text{ is the endogenous participation dummy variable} \]
\[ a_{i,t} = 0 \Rightarrow r_{i,t} = 0 \]
\[ a_{i,t} = 1 \Rightarrow \text{the selection mechanism operates} \]
\[ r_{i,t} = 1[\gamma_0 + x'_{i,t}\gamma + \theta'\theta + z'_{i,t}\delta + v_i + w_{i,t} > 0] \quad \text{all observed if } a_{i,t} = 1 \]
State dependence: \( z \) may include \( r_{i,t-1} \)

Latent persistent unobserved heterogeneity: \( \sigma_v^2 > 0 \).

"Selection" arises if \( \text{Cov}[\varepsilon_{i,t}, w_{i,t}] \neq 0 \) or \( \text{Cov}[u_i, v_i] \neq 0 \)
Estimation Using One Wave

• Use any single wave as a cross section with observed lagged values.
• Advantage: Familiar sample selection model
• Disadvantages
  • Loss of efficiency
  • “One can no longer distinguish between state dependence and unobserved heterogeneity.”
One Wave Model

A standard sample selection model.

\[ y_{it} = \beta_0 + x_{it}'\beta + \bar{x}_{i}'\theta + (u_i + \varepsilon_{it}) \]

\[ r_{it} = 1[\gamma_0 + x_{it}'\gamma + \bar{x}_{i}'\mu + \delta_{1}r_{i,t-1} + \delta_{2}a_{i,t-1} + (v_i + w_{it}) > 0] \]

With only one period of data and \( r_{i,t-1} \) exogenous, this is the Heckman sample selection model. If \( \beta > 0 \), then \( r_{i,t-1} \) is correlated with \( v_i \) and the Heckman approach fails.

An assumption is required:
(1) Include \( r_{i,t-1} \) and assume no unobserved heterogeneity
(2) Exclude \( r_{i,t-1} \) and assume there is no state dependence.

In either case, now if \( \text{Cov}[(u_i + \varepsilon_{it}), (v_i + w_{it})] \) we can use OLS. Otherwise, use the maximum likelihood estimator.
Maximum Likelihood Estimation

• Because numerical integration is required in one or two dimensions for every individual in the sample at each iteration of a high dimensional numerical optimization problem, this is, though feasible, not computationally attractive.
• The dimensionality of the optimization is irrelevant.
• This is much easier in 2008 than it was in 1992 (especially with simulation) The authors did the computations with Hermite quadrature.
Testing for Selection?

- Selectivity is parameterized in these models – coefficients are correlations or covariances.
- Maximum Likelihood Results
  - Covariances were highly insignificant.
  - LR statistic=0.46.
- Two step results produced the same conclusion based on a Hausman test
- ML Estimation results looked like the two step results.
Selectivity in Nonlinear Models

• The ‘Mills Ratio’ approach – just add a ‘lambda’ to whatever model is being estimated?
  • The Heckman model applies to a probit model with a linear regression.
  • The conditional mean in a nonlinear model is not something “+lambda”
• The model can sometimes be built up from first principles
A Bivariate Probit Model

Labor Force Participation Equation
\[ d^* = \alpha'z + u \]
\[ d = 1(d^* > 0) \]

Full Time or Part Time?
\[ f^* = \beta'x + \varepsilon \]
\[ f = 1(f^* > 0) \]

Probability Model:
Nonparticipant: \( \text{Prob}[d=0] = \Phi(-\alpha'z) \)

Participant and Full Time
\[ \text{Prob}[f=1,d=1] = \text{Prob}[f=1|d=1]\text{Prob}[d=1] \]
\[ = \text{Bivariate Normal}(\beta'x, \alpha'z, \rho) \]

Participant and Part Time
\[ \text{Prob}[f=0,d=1] = \text{Prob}[f=0|d=1]\text{Prob}[d=1] \]
\[ = \text{Bivariate Normal}(\beta'x, -\alpha'z, -\rho) \]
## FT/PT Selection Model

<table>
<thead>
<tr>
<th>FIML Estimates of Bivariate Probit Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent variable                       FULLFP</td>
</tr>
<tr>
<td>Weighting variable                       None</td>
</tr>
<tr>
<td>Number of observations                   753</td>
</tr>
<tr>
<td>Log likelihood function                  -723.9798</td>
</tr>
<tr>
<td>Number of parameters                     16</td>
</tr>
<tr>
<td>Selection model based on LFP</td>
</tr>
</tbody>
</table>

| Variable | Coefficient | Standard Error | b/St.Er. | P[|Z|>z] | Mean of X |
|----------|-------------|----------------|----------|----------|-----------|
| Index equation for FULLTIME |
| Constant | .94532822   | 1.61674948    | .585     | .5587    |
| WW      | -.02764944  | .01941006     | -1.424   | .1543    | 4.17768154 |
| KL6     | .04098432   | .26250878     | .156     | .8759    | .14018692  |
| K618    | -.13640024  | .05930081     | -2.300   | .0214    | 1.35046729 |
| AGE     | .03543435   | .07530788     | .471     | .6380    | 41.9719626 |
| AGESQ   | -.00043848  | .00088406     | -.496    | .6199    | 1821.12150 |
| WE      | -.08622974  | .02808185     | -3.071   | .0021    | 12.6588785 |
| FAMINC  | .210971D-04 | .503746D-05   | 4.188    | .0000    | 24130.4229 |

| Index equation for LFP |
| Constant | .98337341   | .50679582     | 1.940    | .0523    |
| KL6     | -.88485756  | .1251971      | -7.864   | .0000    | .23771580  |
| K618    | -.04101187  | .04020437     | -1.020   | .3077    | 1.35325365 |
| WA      | -.02462108  | .01308154     | -1.882   | .0598    | 42.5378486 |
| WE      | .16636047   | .02738447     | 6.075    | .0000    | 12.2868526 |
| HA      | -.01652335  | .01287662     | -1.283   | .1994    | 45.1208499 |
| HE      | -.0627670   | .01912877     | -3.281   | .0010    | 12.4913679 |

Disturbance correlation

| RHO(1,2) | -.84102682 | .25122229 | -3.348 | .0008 |

Full Time = Hours > 1000
Building a Likelihood for a Poisson Regression Model with Selection

Poisson Probability Functions
\[ P(y_i | x_i) = \exp(-\lambda_i) \lambda_i^y / y_i! \]

Covariates and Unobserved Heterogeneity
\[ \lambda(x_i, \beta) = \exp(x_i' + \varepsilon_i) \]

Conditional Contribution to the Log Likelihood
\[ \log L_i | \varepsilon_i = -\lambda(x_i, \varepsilon_i) + y_i \log \lambda(x_i, \varepsilon_i) - \log y_i! \]

Probit Selection Mechanism
\[ d_i^* = z_i' \gamma + u_i, \quad d_i = 1[d_i^* > 0] \]
\[ [\varepsilon_i, u_i] \sim \text{BVN} \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho \sigma \\ \rho \sigma & 1 \end{pmatrix} \]
\[ y_i, x_i \text{ observed only when } d_i = 1. \]
Building the Likelihood

The Conditional Probit Probability

\[ u_i \mid \varepsilon_i \sim N[(\rho / \sigma)\varepsilon_i, (1 - \rho^2)] \]

\[ \text{Prob}[d_i = 1 \mid \mathbf{z}_i, \varepsilon_i] = \Phi \left[ \frac{\mathbf{z}_i' \gamma + (\rho / \sigma)\varepsilon_i}{\sqrt{1 - \rho^2}} \right] \]

\[ \text{Prob}[d_i = 0 \mid \mathbf{z}_i, \varepsilon_i] = \Phi \left[ \frac{-\mathbf{z}_i' \gamma - (\rho / \sigma)\varepsilon_i}{\sqrt{1 - \rho^2}} \right] \]

Conditional Contribution to Likelihood

\[ L_i(y_i, d_i = 1) \mid \varepsilon_i, = [f(y_i \mid \mathbf{x}_i, \varepsilon_i, d_i = 1) \text{Prob}[d_i = 1 \mid \mathbf{z}_i, \varepsilon_i] \]

\[ L_i(d_i = 0) = \text{Prob}[d_i = 0 \mid \mathbf{z}_i, \varepsilon_i] \]
Conditional Likelihood

Conditional Density (not the log)
\[ f(y_i, d_i = 1 \mid \varepsilon_i) = [f(y_i \mid \varepsilon_i, d_i = 1)] \text{Prob}[d_i = 1 \mid \varepsilon_i] \]
\[ f(y_i, d_i = 0 \mid \varepsilon_i) = \text{Prob}[d_i = 0 \mid \varepsilon_i] \]

Unconditional Densities
\[ f(y_i, d_i = 1) = \int_{-\infty}^{\infty} [f(y_i \mid \varepsilon_i, d_i = 1)] \text{Prob}[d_i = 1 \mid \varepsilon_i] \frac{1}{\sigma} \phi \left( \frac{\varepsilon}{\sigma} \right) d\varepsilon \]
\[ f(y_i, d_i = 0) = \int_{-\infty}^{\infty} \text{Prob}[d_i = 0 \mid \varepsilon_i] \frac{1}{\sigma} \phi \left( \frac{\varepsilon}{\sigma} \right) d\varepsilon \]

Log Likelihoods
\[ \log L_i = \log f(y_i, d_i) \]
Poisson Model with Selection

- Strategy:
  - Hermite quadrature or maximum simulated likelihood.
  - Not by throwing a ‘lambda’ into the unconditional likelihood.

- Could this be done without joint normality?
  - How robust is the model?
  - Is there any other approach available?
  - Not easily. The subject of ongoing research.
Poisson Model with Sample Selection.

Dependent variable: DOCVIS
Log likelihood function: -3592.42064
Restricted log likelihood: -6076.83457
Chi-squared [2 d.f.]: 4968.82786
Significance level: .00000
McFadden Pseudo R-squared: .4088336
Estimation based on N = 27326, K = 12
Inf.Cr.AIC = 7208.8 AIC/N = .264
Mean of LHS Variable = 3.12451
Restr. Log-L is Poisson+Probit (indep).
LogL for initial probit = -2442.4091
LogL for initial Poisson= -3634.4254
Means for Psn/Neg.Bin. use selected data.
Means for Probit based on all observations.

<table>
<thead>
<tr>
<th>DOCVIS</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>z</th>
<th>Prob.</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Parameters of Poisson/Neg. Binomial Probability</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>1.22286</td>
<td>1.03286</td>
<td>1.18</td>
<td>.2364</td>
<td>- .80151</td>
</tr>
<tr>
<td>AGE</td>
<td>.01528**</td>
<td>.00654</td>
<td>2.34</td>
<td>.0194</td>
<td>.00247</td>
</tr>
<tr>
<td>EDUC</td>
<td>-.06984**</td>
<td>.02745</td>
<td>-2.54</td>
<td>.0110</td>
<td>-.12364</td>
</tr>
<tr>
<td>HHNINC</td>
<td>-.38472</td>
<td>.36241</td>
<td>-1.06</td>
<td>.2884</td>
<td>-1.09504</td>
</tr>
<tr>
<td>MARRIED</td>
<td>-.22330</td>
<td>.15519</td>
<td>-1.44</td>
<td>.1502</td>
<td>-.52748</td>
</tr>
<tr>
<td>HHKIDS</td>
<td>-.18695</td>
<td>.14693</td>
<td>-1.27</td>
<td>.2032</td>
<td>-.47493</td>
</tr>
<tr>
<td></td>
<td>Parameters of Probit Selection Model</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-3.40612***</td>
<td>.12282</td>
<td>-27.73</td>
<td>.0000</td>
<td>-3.64684</td>
</tr>
<tr>
<td>AGE</td>
<td>.00634***</td>
<td>.00178</td>
<td>3.57</td>
<td>.0004</td>
<td>.00286</td>
</tr>
<tr>
<td>EDUC</td>
<td>.06403***</td>
<td>.00757</td>
<td>8.45</td>
<td>.0000</td>
<td>.04919</td>
</tr>
<tr>
<td>HHNINC</td>
<td>.75095***</td>
<td>.09452</td>
<td>7.95</td>
<td>.0000</td>
<td>.56570</td>
</tr>
<tr>
<td>Sigma</td>
<td>1.07226***</td>
<td>.05985</td>
<td>17.92</td>
<td>.0000</td>
<td>.95495</td>
</tr>
<tr>
<td>Rho</td>
<td>-.02597</td>
<td>.25188</td>
<td>-.10</td>
<td>.9179</td>
<td>-.51964</td>
</tr>
</tbody>
</table>

Note: ***, **, * ==> Significance at 1%, 5%, 10% level.
### Poisson Regression

Dependent variable: DOCVIS

- Log likelihood function: $-1659.65094$
- Restricted log likelihood: $-1751.09275$
- Chi squared [6 d.f.]: $182.88363$
- Significance level: $0.0000$
- McFadden Pseudo R-squared: $0.0522199$

Estimation based on $N = 27326$, $K = 7$

- Inf. Cr. AIC = $3333.3$, $AIC/N = 0.122$
- Chi-squared = $2887.14468$, $RsqP = 0.1766$
- G-squared = $2219.26013$, $RsqD = 0.0761$

Overdispersion tests: $g = \mu(i): 5.121$

Overdispersion tests: $g = \mu(i)^2: 5.728$

Cov. matrix corrected for 2 step estimation

| DOCVIS | Coefficient | Standard Error | z  | Prob. $|z| > Z^*$ | 95% Confidence Interval |
|--------|-------------|----------------|----|----------------|-------------------------|
| Constant | 1.13394     | 4082.032       | 0.00 | 0.9998       | -7999.50166 8001.76954 |
| AGE    | 0.01440     | 6.37321        | 0.00 | 0.9982       | -12.47686 12.50566    |
| EDUC   | -0.05828    | 63.44852       | 0.00 | 0.9993       | -124.41510 124.29854  |
| HHNINC | -0.19541    | 742.8324       | 0.00 | 0.9998       | -1456.12012 1455.72930|
| MARRIED| -0.37640    | 62847          | -0.60 | 0.5492     | -1.60818 1.60818  .85539 |
| HHKIDS | -0.19118    | 27342          | -0.70 | 0.4844     | -0.72707 0.34472    |
| MlsRatios | 0.21226   | 1130.177       | 0.00 | 0.9999       | -2214.89378 2215.31831|

Note: ***, **, * ==> Significance at 1%, 5%, 10% level.
Stochastic Frontier Model: ML

\[ y_i = \beta' x_i + v_i - u_i \]

where \( u_i = |\sigma_u U_i| = \sigma_u |U_i|, U_i \sim N[0,1^2], \)

\[ v_i = \sigma_v V_i, V_i \sim N[0,1^2]. \]

\[
\log L(\beta, \sigma, \lambda) = \sum_{i=1}^{N} \left[ \frac{1}{2} \log\left(\frac{2}{\pi}\right) - \log \sigma - \frac{1}{2} \left( \frac{\varepsilon_i}{\sigma} \right)^2 + \log \Phi\left( -\lambda \frac{\varepsilon_i}{\sigma} \right) \right]
\]

where \( \varepsilon_i = y_i - \beta' x_i = v_i - u_i, \)

\( \lambda = \frac{\sigma_u}{\sigma_v}, \)

\( \sigma = \sqrt{\sigma_v^2 + \sigma_u^2} \)
Sample Selected SF Model

\[ d_i = 1[\alpha'z_i + w_i > 0], \ w_i \sim N[0, 1^2] \]
\[ y_i = \beta'x_i + \varepsilon_i, \ \varepsilon_i \sim N[0, \sigma_{\varepsilon}^2] \]
(yᵢ, xᵢ) observed only when \( d_i = 1 \).
\[ \varepsilon_i = v_i - u_i \]
\[ u_i = |\sigma_u U_i| = \sigma_u |U_i| \text{ where } U_i \sim N[0, 1^2] \]
\[ v_i = \sigma_v V_i \text{ where } V_i \sim N[0, 1^2]. \]
\[ (w_i, v_i) \sim N_2[(0,1), (1, \rho \sigma_v, \sigma_v^2)] \]

\[
f(y_i \mid x_i, U_i, d_i, z_i) = \begin{cases} 
   d_i \frac{\exp\left(-\frac{1}{2} \left(y_i - \beta'x_i + \sigma_u |U_i| / \sigma_v^2 \right)^2 \right)}{\sigma_v \sqrt{2\pi}} \Phi \left( \frac{\rho(y_i - \beta'x_i + \sigma_u |U_i| / \sigma_v)}{\sqrt{1 - \rho^2}} + \alpha'z_i \right) & d_i = 1 \\
   (1 - d_i) \Phi(-\alpha'z_i) & d_i = 0
\end{cases}
\]
Sample Selection in a Nonlinear Model

\[ d_i = 1(\alpha'z_i + w_i > 0) \quad w_i \sim N[0,1], \]

\[ g|\varepsilon_i = g(\beta'x_i, \sigma_{\varepsilon_i}) \quad \varepsilon_i \sim N[0,1] \]

\[ y_i | x_i, \varepsilon_i \sim f[y_i | g(\beta'x_i, \sigma_{\varepsilon_i})] \]

\[ [w_i, \varepsilon_i] \sim N[(0,1),(1,\rho,1)] \]

\[ y_i, x_i \text{ are observed only when } z_i = 1. \]

\[
\begin{align*}
    f(y_i, d_i | x_i, z_i) &= \int_{-\infty}^{\infty} \{ (1-d_i) + d_i f[y_i | g(\beta'x_i, \sigma_{\varepsilon_i})] \} \times \\
    &\quad \Phi \left( (2d_i - 1)[\alpha'z_i + \rho\varepsilon_i] / \sqrt{1 - \rho^2} \right) \phi(\varepsilon_i) d\varepsilon_i,
\end{align*}
\]
Simulated Log Likelihood for a Simpler Model

\[
\log L_s(\beta, \sigma_u, \sigma_v, \alpha, \rho) = \sum_{i=1}^{N} \log \frac{1}{R} \sum_{r=1}^{R} \left[ d_i \left[ \frac{\exp\left(-\frac{1}{2}(y_i - \beta'x_i + \sigma_u |U_{ir}|) / \sigma_v^2\right)}{\sigma_v \sqrt{2\pi}} \right] \times \right.

\left. \Phi \left( \frac{\rho(y_i - \beta'x_i + \sigma_u |U_{ir}|) / \sigma_\varepsilon + \alpha'z_i}{\sqrt{1 - \rho^2}} \right) \right]

\left. + (1 - d_i) \Phi(-\alpha'z_i) \right]
\]
A 2 Step MSL Approach

\[
\log L_{SC}(\beta, \sigma_u, \sigma_v, \rho) = \sum_{i=1}^{N} \log \frac{1}{R} \sum_{r=1}^{R} \frac{\exp(-\frac{1}{2}(y_i - \beta'x_i + \sigma_u |U_{ir}|)^2 / \sigma_v^2)}{\sigma_v \sqrt{2\pi}} \times \Phi\left( \frac{\rho(y_i - \beta'x_i + \sigma_u |U_{ir}|) / \sigma_{\epsilon} + a_i}{\sqrt{1 - \rho^2}} \right) + (1 - d_i)\Phi(-a_i)
\]

where \(a_i = \hat{\alpha}'z_i\)
Simulated ML for the SF Model

\[ f(y_i \mid x_i, |U_i|) = \frac{\exp[-\frac{1}{2} (y_i - \beta' x_i + \sigma_u |U_i|)^2 / \sigma_v^2]}{\sigma_v \sqrt{2\pi}} \]

\[ f(y_i \mid x_i) = \int_{|U_i|} \frac{\exp[-\frac{1}{2} (y_i - \beta' x_i + \sigma_u |U_i|)^2 / \sigma_v^2]}{\sigma_v \sqrt{2\pi}} p(|U_i|) d|U_i| \]

\[ p(|U_i|) = \frac{2 \exp[-\frac{1}{2} |U_i|^2]}{\sqrt{2\pi}}, |U_i| \geq 0. \]

\[ f(y \mid x_i) \approx \frac{1}{R} \sum_{r=1}^{R} \frac{\exp[-\frac{1}{2} (y_i - \beta' x_i + \sigma_u |U_{ir}|)^2 / \sigma_v^2]}{\sigma_v \sqrt{2\pi}} \]

\[ \log L_S(\beta, \sigma_u, \sigma_v) = \sum_{i=1}^{N} \log \left\{ \frac{1}{R} \sum_{r=1}^{R} \frac{\exp[-\frac{1}{2} (y_i - \beta' x_i + \sigma_u |U_{ir}|)^2 / \sigma_v^2]}{\sigma_v \sqrt{2\pi}} \right\} \]

This is simply a linear regression with a random constant term, \( \alpha_i = \alpha - \sigma_u |U_i| \)
Nonnormality Issue

- How robust is the Heckman model to nonnormality of the unobserved effects?
- Are there other techniques
  - Parametric: Copula methods
  - Semiparametric: Klein/Spady and Series methods
- Other forms of the selection equation – e.g., multinomial logit
- Other forms of the primary model: e.g., as above.
A Study of Health Status in the Presence of Attrition
Model for Self Assessed Health

- British Household Panel Survey (BHPS)
  - Self assessed health on 0,1,2,3,4 scale
  - Sociological and demographic covariates
  - Dynamics – inertia in reporting of top scale
- Dynamic ordered probit model
  - Balanced panel – analyze dynamics
  - Unbalanced panel – examine attrition
Dynamic Ordered Probit Model

Latent Regression - Random Utility

\[ h_{it}^* = \beta'x_{it} + \gamma' H_{i,t-1} + \alpha_i + \varepsilon_{it} \]

\( x_{it} \) = relevant covariates and control variables

\( H_{i,t-1} = 0/1 \) indicators of reported health status in previous period

\( H_{i,t-1}(j) = 1[\text{Individual i reported } h_{it} = j \text{ in previous period}], j=0,\ldots,4 \)

Ordered Choice Observation Mechanism

\( h_{it} = j \text{ if } \mu_{j-1} < h_{it}^* \leq \mu_j, j = 0,1,2,3,4 \)

Ordered Probit Model - \( \varepsilon_{it} \sim N[0,1] \)

Random Effects with Mundlak Correction and Initial Conditions

\[ \alpha_i = \alpha_0 + \alpha_1' H_{i,1} + \alpha_2' \bar{X}_i + \upsilon_i, \upsilon_i \sim N[0,\sigma^2] \]

It would not be appropriate to include \( h_{i,t-1} \) itself in the model as this is a label, not a measure.
**Random Effects Dynamic Ordered Probit Model**

Random Effects Dynamic Ordered Probit Model

\[ h_{i,t}^* = x_{i,t}' \beta + \sum_{j=1}^{J} \gamma_j h_{i,t-1}(j) + \alpha_i + \epsilon_{i,t} \]

\[ h_{i,t} = j \text{ if } \mu_{j-1} < h_{i,t}^* < \mu_j \]

\[ h_{i,t}(j) = 1 \text{ if } h_{i,t} = j \]

\[ P_{i,t,j} = P[h_{i,t} = j] = \Phi(\mu_j - x_{i,t}' \beta - \sum_{j=1}^{J} \gamma_j h_{i,t-1}(j) - \alpha_i) \]

\[ - \Phi(\mu_{j-1} - x_{i,t}' \beta - \sum_{j=1}^{J} \gamma_j h_{i,t-1}(j) - \alpha_i) \]

Parameterize Random Effects

\[ \alpha_i = \alpha_0 + \sum_{j=1}^{J} \alpha_{i,j} h_{i,1}(j) + \alpha' \bar{x}_i + u_i \]

Simulation or Quadrature Based Estimation

\[ \ln L = \sum_{i=1}^{N} \ln \int_{\alpha_i} \prod_{t=1}^{T_i} P_{i,t,j}(\alpha_j) d\alpha_j \]
# Data

## Table I. Variable definitions

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAH</td>
<td><strong>Self-Assessed Health</strong>: 5 if excellent, 4 if good, 3 if fair, 2 if poor, 1 if very poor</td>
</tr>
<tr>
<td>WIDOW</td>
<td>1 if widowed, 0 otherwise</td>
</tr>
<tr>
<td>SINGLE</td>
<td>1 if never married, 0 otherwise</td>
</tr>
<tr>
<td>DIV/SEP</td>
<td>1 if divorced or separated, 0 otherwise</td>
</tr>
<tr>
<td>NON-WHITE</td>
<td>1 if a member of ethnic group other than white, 0 otherwise</td>
</tr>
<tr>
<td>DEGREE</td>
<td>1 if highest academic qualification is a degree or higher degree, 0 otherwise</td>
</tr>
<tr>
<td>HND/A</td>
<td>1 if highest academic qualification is HND or A level, 0 otherwise</td>
</tr>
<tr>
<td>O/CSE</td>
<td>1 if highest academic qualification is O level or CSE, 0 otherwise</td>
</tr>
<tr>
<td>HHSIZE</td>
<td>Number of people in household including respondent</td>
</tr>
<tr>
<td>NCHO4</td>
<td>Number of children in household aged 0–4</td>
</tr>
<tr>
<td>NCH511</td>
<td>Number of children in household aged 5–11</td>
</tr>
<tr>
<td>NCH1218</td>
<td>Number of children in household aged 12–18</td>
</tr>
<tr>
<td>INCOME</td>
<td>Equivalized annual real household income in pounds</td>
</tr>
<tr>
<td>AGE</td>
<td>Age in years at 1st December of current wave</td>
</tr>
</tbody>
</table>
Variable of Interest

![Bar chart showing self-assessed health status by wave for women.](image)

Figure 1. Self-assessed health status by wave
# Dynamics

**Table II. Transition matrices, balanced panel**

**(a) Men**

<table>
<thead>
<tr>
<th>SAH</th>
<th>EX</th>
<th>GOOD</th>
<th>FAIR</th>
<th>POOR</th>
<th>VERY POOR</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>EX</td>
<td>0.600</td>
<td>0.342</td>
<td>0.046</td>
<td>0.010</td>
<td>0.002</td>
<td>5485</td>
</tr>
<tr>
<td>GOOD</td>
<td>0.184</td>
<td>0.651</td>
<td>0.142</td>
<td>0.019</td>
<td>0.004</td>
<td>9263</td>
</tr>
<tr>
<td>FAIR</td>
<td>0.055</td>
<td>0.361</td>
<td>0.471</td>
<td>0.100</td>
<td>0.012</td>
<td>3433</td>
</tr>
<tr>
<td>POOR</td>
<td>0.029</td>
<td>0.120</td>
<td>0.340</td>
<td>0.418</td>
<td>0.093</td>
<td>1031</td>
</tr>
<tr>
<td>VERY POOR</td>
<td>0.032</td>
<td>0.073</td>
<td>0.133</td>
<td>0.423</td>
<td>0.339</td>
<td>248</td>
</tr>
<tr>
<td>N</td>
<td>5231</td>
<td>9287</td>
<td>3565</td>
<td>1111</td>
<td>266</td>
<td>19460</td>
</tr>
</tbody>
</table>

**(b) Women**

<table>
<thead>
<tr>
<th>SAH</th>
<th>EX</th>
<th>GOOD</th>
<th>FAIR</th>
<th>POOR</th>
<th>VERY POOR</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>EX</td>
<td>0.572</td>
<td>0.353</td>
<td>0.059</td>
<td>0.013</td>
<td>0.004</td>
<td>5164</td>
</tr>
<tr>
<td>GOOD</td>
<td>0.150</td>
<td>0.657</td>
<td>0.162</td>
<td>0.026</td>
<td>0.005</td>
<td>11306</td>
</tr>
<tr>
<td>FAIR</td>
<td>0.040</td>
<td>0.362</td>
<td>0.465</td>
<td>0.116</td>
<td>0.017</td>
<td>4928</td>
</tr>
<tr>
<td>POOR</td>
<td>0.021</td>
<td>0.156</td>
<td>0.360</td>
<td>0.365</td>
<td>0.098</td>
<td>1587</td>
</tr>
<tr>
<td>VERY POOR</td>
<td>0.014</td>
<td>0.106</td>
<td>0.192</td>
<td>0.326</td>
<td>0.362</td>
<td>423</td>
</tr>
<tr>
<td>N</td>
<td>4884</td>
<td>11329</td>
<td>5082</td>
<td>1649</td>
<td>464</td>
<td>23408</td>
</tr>
</tbody>
</table>
### Attrition

#### Table V. Sample size, drop-outs and attrition rates by wave

(a) All data

<table>
<thead>
<tr>
<th>Wave</th>
<th>No. individuals</th>
<th>Survival rate</th>
<th>Drop-outs</th>
<th>Attrition rate</th>
<th>EX at $t - 1$</th>
<th>Attrition rate</th>
<th>GOOD at $t - 1$</th>
<th>Attrition rate</th>
<th>FAIR at $t - 1$</th>
<th>Attrition rate</th>
<th>POOR at $t - 1$</th>
<th>Attrition rate</th>
<th>VPOOR at $t - 1$</th>
<th>Attrition rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10,256</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>8,927</td>
<td>87.33%</td>
<td>1,299</td>
<td>12.67%</td>
<td>11.54%</td>
<td>12.57%</td>
<td>13.01%</td>
<td>13.73%</td>
<td>23.74%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>8,162</td>
<td>79.58%</td>
<td>795</td>
<td>8.88%</td>
<td>8.08%</td>
<td>8.13%</td>
<td>9.65%</td>
<td>12.62%</td>
<td>19.46%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>7,825</td>
<td>76.30%</td>
<td>337</td>
<td>4.13%</td>
<td>6.67%</td>
<td>6.54%</td>
<td>6.73%</td>
<td>10.35%</td>
<td>14.74%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>7,430</td>
<td>72.45%</td>
<td>395</td>
<td>5.05%</td>
<td>6.21%</td>
<td>6.18%</td>
<td>7.87%</td>
<td>9.11%</td>
<td>16.34%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>7,238</td>
<td>70.57%</td>
<td>192</td>
<td>2.58%</td>
<td>3.11%</td>
<td>3.24%</td>
<td>5.06%</td>
<td>10.47%</td>
<td>18.83%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>7,102</td>
<td>69.25%</td>
<td>136</td>
<td>1.88%</td>
<td>3.15%</td>
<td>3.85%</td>
<td>4.79%</td>
<td>8.83%</td>
<td>8.75%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>6,839</td>
<td>66.68%</td>
<td>263</td>
<td>3.70%</td>
<td>3.43%</td>
<td>3.82%</td>
<td>5.30%</td>
<td>5.88%</td>
<td>17.01%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Testing for Attrition Bias

Table 9: Verbeek and Nijman tests for attrition: based on dynamic ordered probit models with Wooldridge specification of correlated effects and initial conditions

<table>
<thead>
<tr>
<th></th>
<th>MEN</th>
<th>WOMEN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>β</td>
<td>Std.err.</td>
</tr>
<tr>
<td>NEXT WAVE</td>
<td>.199</td>
<td>.035</td>
</tr>
<tr>
<td>ALL WAVES</td>
<td>.139</td>
<td>.031</td>
</tr>
<tr>
<td>NUMBER OF WAVES</td>
<td>.031</td>
<td>.009</td>
</tr>
</tbody>
</table>

Three dummy variables added to full model with unbalanced panel suggest presence of attrition effects.
Probability Weighting Estimators

• A Patch for Attrition
  (1) Fit a participation probit equation for each wave.
  (2) Compute \( p(i,t) = \text{predictions of participation for each individual in each period}. \)
  • Special assumptions needed to make this work
  • Ignore common effects and fit a weighted pooled log likelihood: \( \sum_i \sum_t [d_{it}/p(i,t)] \log LP_{it}. \)
Attributes Model with IP Weights

Assumes (1) \(\text{Prob}(\text{attrition}|\text{all data}) = \text{Prob}(\text{attrition}|\text{selected variables})\) (ignorability)

(2) Attrition is an ‘absorbing state.’ No reentry.

Obviously not true for the GSOEP data above.

Can deal with point (2) by isolating a subsample of those present at wave 1 and the monotonically shrinking subsample as the waves progress.
Inverse Probability Weighting

Panel is based on those present at WAVE 1, N1 individuals.
Attrition is an absorbing state. No reentry, so N1 ≥ N2 ≥ ... ≥ N8.
Sample is restricted at each wave to individuals who were present at the previous wave.

\[ d_{it} = 1 \text{[Individual is present at wave t]} \]

\[ d_{i1} = 1 \quad \forall \quad i, d_{it} = 0 \quad \Rightarrow \quad d_{i,t+1} = 0. \]

\[ \tilde{x}_{i1} = \text{covariates observed for all i at entry that relate to likelihood of being present at subsequent waves.} \]

(health problems, disability, psychological well being, self employment, unemployment, maternity leave, student, caring for family member, ...)

Probit model for \( d_{it} = I[\delta' \tilde{x}_{i1} + w_{it}] \), \( t = 2,\ldots,8 \). \( \hat{\pi}_{it} = \text{fitted probability.} \)

Assuming attrition decisions are independent, \( \hat{P}_{it} = \prod_{s=1}^{t} \hat{\pi}_{is} \)

Inverse probability weight \( \hat{W}_{it} = \frac{d_{it}}{\hat{P}_{it}} \)

Weighted log likelihood \( \log L_W = \sum_{i=1}^{N} \sum_{t=1}^{8} \log L_{it} \) (No common effects.)
# Estimated Partial Effects by Model

**Table 12: Average partial effects on probability of reporting excellent health for selected variables**

<table>
<thead>
<tr>
<th>a) Men</th>
<th>(1) Pooled model, balanced sample</th>
<th>(2) Pooled model, unbalanced sample</th>
<th>(3) Pooled model, IPW-1</th>
<th>(4) Pooled model, IPW-2</th>
<th>(5) Random effects, balanced sample</th>
<th>(6) Random effects, unbalanced sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ln(INCOME)</td>
<td>.009 (.004)</td>
<td>.009 (.004)</td>
<td>.009 (.004)</td>
<td>.011 (.005)</td>
<td>.013 (.006)</td>
<td>.012 (.005)</td>
</tr>
<tr>
<td>Mean Ln(INCOME)</td>
<td>.049 (.024)</td>
<td>.043 (.022)</td>
<td>.042 (.021)</td>
<td>.045 (.022)</td>
<td>.066 (.028)</td>
<td>.056 (.025)</td>
</tr>
<tr>
<td>DEGREE</td>
<td>.010 (.005)</td>
<td>.017 (.009)</td>
<td>.018 (.009)</td>
<td>.018 (.009)</td>
<td>.015 (.006)</td>
<td>.015 (.006)</td>
</tr>
<tr>
<td>HND/A</td>
<td>.019 (.009)</td>
<td>.021 (.011)</td>
<td>.021 (.010)</td>
<td>.022 (.011)</td>
<td>.028 (.011)</td>
<td>.030 (.013)</td>
</tr>
<tr>
<td>O/CSE</td>
<td>.016 (.008)</td>
<td>.020 (.010)</td>
<td>.020 (.010)</td>
<td>.020 (.010)</td>
<td>.024 (.010)</td>
<td>.028 (.012)</td>
</tr>
<tr>
<td>SAHEX(t-1)</td>
<td>.234 (.087)</td>
<td>.231 (.090)</td>
<td>.231 (.090)</td>
<td>.230 (.089)</td>
<td>.082 (.031)</td>
<td>.085 (.035)</td>
</tr>
<tr>
<td>SAHFAIR(t-1)</td>
<td>-.170 (.085)</td>
<td>-.163 (.084)</td>
<td>-.162 (.084)</td>
<td>-.162 (.083)</td>
<td>-.080 (.034)</td>
<td>-.077 (.036)</td>
</tr>
<tr>
<td>SAHPPOORT(t-1)</td>
<td>-.242 (.167)</td>
<td>-.233 (.163)</td>
<td>-.232 (.162)</td>
<td>-.232 (.162)</td>
<td>-.151 (.077)</td>
<td>-.145 (.078)</td>
</tr>
<tr>
<td>SAHVPOORT(t-1)</td>
<td>-.260 (.198)</td>
<td>-.253 (.197)</td>
<td>-.255 (.199)</td>
<td>-.255 (.200)</td>
<td>-.184 (.104)</td>
<td>-.179 (.106)</td>
</tr>
</tbody>
</table>
Partial Effect for a Category

These are 4 dummy variables for state in the previous period. Using first differences, the 0.234 estimated for SAHEX means transition from EXCELLENT in the previous period to GOOD in the previous period, where GOOD is the omitted category. Likewise for the other 3 previous state variables. The margin from ‘POOR’ to ‘GOOD’ was not interesting in the paper. The better margin would have been from EXCELLENT to POOR, which would have (EX,POOR) change from (1,0) to (0,1).