MODULE FOUR, PART TWO: SAMPLE SELECTION

IN ECONOMIC EDUCATION RESEARCH USING LIMDEP (NLOGIT)

Part Two of Module Four provides a cookbook-type demonstration of the steps required to use LIMDEP (NLOGIT) in situations involving estimation problems associated with sample selection. Users of this model need to have completed Module One, Parts One and Two, but not necessarily Modules Two and Three. From Module One users are assumed to know how to get data into LIMDEP, recode and create variables within LIMDEP, and run and interpret regression results. Module Four, Parts Three and Four demonstrate in STATA and SAS what is done here in LIMDEP.

THE CASE, DATA, AND ROUTINE FOR EARLY HECKMAN ADJUSTMENT

The change score or difference in difference model is used extensively in education research. Yet, before Becker and Walstad (1990), little if any attention was given to the consequence of missing student records that result from: 1) "data cleaning" done by those collecting the data, 2) student unwillingness to provide data, or 3) students self-selecting into or out of the study. The implications of these types of sample selection are shown in the work of Becker and Powers (2001) where the relationship between class size and student learning was explored using the third edition of the Test of Understanding in College Economics (TUCE), which was produced by Saunders (1994) for the National Council on Economic Education (NCEE), since renamed the Council for Economic Education.

Module One, Part Two showed how to get the Becker and Powers data set “beck8WO.csv” into LIMDEP (NLOGIT). As a brief review this was done with the read command:

```
READ; NREC=2837; NVAR=64; FILE=k:\beck8WO.csv; Names=
A1,A2,X3, C,AL,AM,AN,CA,CB,CC,CH,CI,CJ,CK,CL,CM,CN,CO,CS,CT,
CU,CV,CW,DB,DD,DI,DJ,DK,DL,DM,DN,DQ,DR,DS,DY,DZ,EA,EB,EE,EF,
EI,EJ,EP,EQ,ER,ET,EY,EZ,FF,FN,FX,FY,FZ,GE,GH,GM,GN,GQ,GR,HB,
HC,HD,HE,HF $
```

where

A1: term, where 1= fall, 2 = spring
A2: school code, where 100/199 = doctorate,
200/299 = comprehensive,
300/399 = lib arts,
400/499 = 2 year
hb: initial class size (number taking preTUCE)

hc: final class size (number taking postTUCE)
dm: experience, as measured by number of years teaching
dj: teacher's highest degree, where Bachelors=1, Masters=2, PhD=3
cc: postTUCE score (0 to 30)
an: preTUCE score (0 to 30)
ge: Student evaluation measured interest
gh: Student evaluation measured textbook quality
gm: Student evaluation measured regular instructor's English ability
gq: Student evaluation measured overall teaching effectiveness
ci: Instructor sex (Male = 1, Female = 2)
ck: English is native language of instructor (Yes = 1, No = 0)
cs: PostTUCE score counts toward course grade (Yes = 1, No = 0)
ff: GPA*100
fn: Student had high school economics (Yes = 1, No = 0)
ey: Student's sex (Male = 1, Female = 2)
fx: Student working in a job (Yes = 1, No = 0)

In Module One, Part Two the procedure for changing the size of the work space in earlier versions of LIMDEP and NLOGIT was shown but that is no longer required for the 9th version of LIMDEP and the 4th version of NLOGIT. Starting with LIMDEP version 9 and NLOGIT version 4 the required work space is automatically determined by the "Read" command and increased as needed with subsequent "Create" commands.

Separate dummy variables need to be created for each type of school (A2), which is done with the following code:

```
recode; a2; 100/199 = 1; 200/299 = 2; 300/399 = 3; 400/499 =4$
create; doc=a2=1; comp=a2=2; lib=a2=3; twoyr=a2=4$
```

To create a dummy variable for whether the instructor had a PhD we use

```
Create; phd=dj=3$
```

To create a dummy variable for whether the student took the postTUCE we use

```
final=cc>0;
```

To create a dummy variable for whether a student did (noeval = 0) or did not (noeval = 1) complete a student evaluation of the instructor we use

```
Create evalsum=ge+gh+gm+gq; noeval=evalsum=-36$
```

"Noeval" reflects whether the student was around toward the end of the term, attending classes, and sufficiently motivated to complete an evaluation of the instructor. In the Saunder's data set evaluation questions with no answer where coded -9; thus, these four questions summing to -36 indicates that no questions were answered.

And the change score is created with

```
Create; change=cc-an$
```

Finally, there was a correction for the term in which student record 2216 was incorrectly recorded:

```
recode; hb; 90=89$
```

All of these recoding and create commands are entered into LIMDEP command file as follows:

```
recode; a2; 100/199 = 1; 200/299 = 2; 300/399 = 3; 400/499 =4$
create; doc=a2=1; comp=a2=2; lib=a2=3; twoyr=a2=4; phd=dj=3;final=cc>0;
evalsum=ge+gh+gm+gq; noeval=evalsum=-36$
Create; change=cc-an$
recode; hb; 90=89$ #2216 counted in term 2, but in term 1 with no posttest
```

To remove records with missing data the following is entered:

```
Reject; AN=-9$
```

```
Reject; HB=-9$
Reject; ci=-9$
Reject; ck=-9$
Reject; cs=0$
Reject; cs=-9$
Reject; a2=-9$
Reject; phd=-9$
```

The use of these data entry and management commands will appear in the LIMDEP (NLOGIT) output file for the equations to be estimated in the next section.

THE PROPENSITY TO TAKE THE POSTTEST AND THE CHANGE SCORE EQUATION

To address attrition-type sample selection problems in change score studies, Becker and Powers first add observations that were dropped during the early stage of assembling data for TUCE III. Becker and Powers do not have any data on students before they enrolled in the course and thus cannot address selection into the course, but to examine the effects of attrition (course withdrawal) they introduce three measures of class size (beginning, ending, and average) and argue that initial or beginning class size is the critical measure for assessing learning over the entire length of the course.[i] To show the effects of initial class size on attrition (as discussed in Module Four, Part One) they employ what is now the simplest and most restrictive of sample correction methods, which can be traced to James Heckman (1979), recipient of the 2000 Nobel Prize in Economics.

From Module Four, Part One, we have the data generating process for the difference between post and preTUCE scores for the i^{th} student (Δy_i):

$$\Delta y_i = \mathbf{X}_i\boldsymbol{\beta} + \varepsilon_i = \beta_1 + \sum_{j=2}^{k} \beta_j x_{ji} + \varepsilon_i \qquad (1)$$

where the data set of explanatory variables is matrix **X**, where $\mathbf{X}_i$ is the row of x_{ji} values for the relevant variables believed to explain the i^{th} student's pretest and posttest scores, the β_j 's are the associated slope coefficients in the vector $\boldsymbol{\beta}$, and ε_i is the individual random shock (caused, for example, by unobservable attributes, events or environmental factors) that affect the i^{th} student's test scores. Sample selection associated with students' unwillingness to take the postteest (dropping the course) results in population error term and regressor correlation that biases and makes coefficient estimators in this change score model inconsistent.

The data generating process for the i^{th} student's propensity to take the posttest is:

$$T_i^* = \mathbf{H}_i\boldsymbol{\alpha} + \omega_i \qquad (2)$$

where

$T_i = 1$, if $T_i^* > 0$, and student i has a posttest score, and

$T_i = 0$, if $T_i^* \leq 0$, and student i does not have a posttest score.

$\mathbf{T}^*$ is the vector of all students' propensities to take a posttest.

$\mathbf{H}$ is the matrix of explanatory variables that are believed to drive these propensities.

$\boldsymbol{\alpha}$ is the vector of slope coefficients corresponding to these observable variables.

$\boldsymbol{\omega}$ is the vector of unobservable random shocks that affect each student's propensity.

The effect of attrition between the pretest and posttest, as reflected in the absence of a posttest score for the i^{th} student $(T_i = 0)$ and a Heckman adjustment for the resulting bias caused by excluding those students from the change-score regression requires estimation of equation (2) and the calculation of an inverse Mill's ratio for each student who has a pretest. This inverse Mill's ratio is then added to the change-score regression (1) as another explanatory variable. In essence, this inverse Mill's ratio adjusts the error term for the missing students.

For the Heckman adjustment for sample selection each disturbance in vector $\boldsymbol{\varepsilon}$, equation (1), is assumed to be distributed bivariate normal with the corresponding disturbance term in the $\boldsymbol{\omega}$ vector of the selection equation (2). Thus, for the i^{th} student we have:

$$(\varepsilon_i, \omega_i) \sim \text{ bivariate normal}(0, 0, \sigma_\varepsilon, 1, \rho) \tag{3}$$

and for all perturbations in the two-equation system we have:

$$E(\boldsymbol{\varepsilon}) = E(\boldsymbol{\omega}) = 0,\ E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma_\varepsilon^2\mathbf{I},\ E(\boldsymbol{\omega}\boldsymbol{\omega}') = \mathbf{I}, \text{ and } E(\varepsilon\omega') = \rho\sigma_\varepsilon\mathbf{I}. \tag{4}$$

That is, the disturbances have zero means, unit variance, and no covariance among students, but there is covariance between selection in getting a posttest score and the measurement of the change score.

The regression for this censored sample of $n_{T=1}$ students who took the posttest is now:

$$E(\Delta y_i \mid \mathbf{X}_i, T_i = 1) = \mathbf{X}_i\boldsymbol{\beta} + E(\varepsilon_i \mid T_i^* > 0);\ i = 1, 2, \ldots n_{T=1} \text{ , for } n_{T=1} < N \tag{5}$$

which suggests the Heckman adjusted regression to be estimated:

$$E(\Delta y_i \mid \mathbf{X}_i, T_i = 1) = \mathbf{X}_i\boldsymbol{\beta} + (\rho\sigma_\varepsilon)\lambda_i;\ i = 1, 2, \ldots n_{T=1} \tag{6}$$

where λ_i is the inverse Mill's ratio (or hazard) such that $\lambda_i = f(-T_i^*)/[1-F(-T_i^*)]$, and $f(.)$ and $F(.)$ are the normal density and distribution functions. λ_i is the standardized mean of the disturbance term ω_i, for the i^{th} student who took the posttest; it is close to zero only for those well above the $T = 1$ threshold. The values of λ are generated from the estimated probit selection equation (2) for all students.

The probit command for the selection equation to be estimated in LIMDEP (NLOGIT) is

```
probit;lhs=final;rhs=one,an,hb,doc,comp,lib,ci,ck,phd,noeval;hold results$
```

where the "hold results" extension tells LIMDEP to hold the results for the change equation to be estimated by least squares with the inverse Mill's ratio used as regressor.

The command for estimating the adjusted change equation using both the inverse Mills ratio as a regressor and maximum likelihood estimation of the ρ and σ_ε is written

```
selection;lhs=change;rhs=one,hb,doc,comp,lib,ci,ck,phd,noeval;mle$
```

where the extension "mle" tells LIMDEP (NLOGIT) to use maximum likelihood estimation.

As described in Module One, Part Two, entering all of these commands into the command file in LIMDEP (NLOGIT), highlighting the bunch and pressing the GO button yields the following output file:

```
Initializing NLOGIT Version 4.0.7

--> READ; NREC=2837; NVAR=64; FILE=k:\beck8WO.csv; Names=
    A1,A2,X3, C,AL,AM,AN,CA,CB,CC,CH,CI,CJ,CK,CL,CM,CN,CO,CS,CT,
    CU,CV,CW,DB,DD,DI,DJ,DK,DL,DM,DN,DQ,DR,DS,DY,DZ,EA,EB,EE,EF,
    EI,EJ,EP,EQ,ER,ET,EY,EZ,FF,FN,FX,FY,FZ,GE,GH,GM,GN,GQ,GR,HB,
    HC,HD,HE,HF $
--> recode; a2; 100/199 = 1; 200/299 = 2; 300/399 = 3; 400/499 =4$
--> recode; hb; 90=89$ #2216 counted in term 2, but in term 1 with no posttest
--> create; doc=a2=1; comp=a2=2; lib=a2=3; twoyr=a2=4; phd=dj=3; final=cc>0;
    evalsum=ge+gh+gm+gq; noeval=evalsum=-36$
--> Create; change=cc-an$
--> Reject; AN=-9$
--> Reject; HB=-9$
--> Reject; ci=-9$
--> Reject; ck=-9$
--> Reject; cs=0$
--> Reject; cs=-9$
--> Reject; a2=-9$
--> Reject; phd=-9$

--> probit;lhs=final;rhs=one,an,hb,doc,comp,lib,ci,ck,phd,noeval;hold results$
```

```
Normal exit:   6 iterations. Status=0. F=    822.7411

+---------------------------------------------+
| Binomial Probit Model                       |
| Dependent variable                FINAL     |
| Log likelihood function       -822.7411     |
| Restricted log likelihood     -1284.216     |
| Chi squared [   9 d.f.]       922.95007     |
| Significance level             .0000000     |
| McFadden Pseudo R-squared      .3593438     |
| Estimation based on N =   2587, K =  10     |
| AIC =      .6438  Bayes IC =      .6664     |
| AICf.s. =      .6438  HQIC =      .6520     |
| Model estimated: Dec 08, 2009, 12:12:49     |
| Results retained for SELECTION model.       |
| Hosmer-Lemeshow chi-squared =  26.06658     |
| P-value=  .00102 with deg.fr. =       8     |
|                                             |

+--------+-------------+--------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+-------------+--------------+--------+--------+----------+
+--------+Index function for probability                            |
|Constant|     .99535***       .24326        4.092   .0000           |
|AN      |     .02204**        .00948        2.326   .0200    10.5968|
|HB      |    -.00488**        .00192       -2.538   .0112    55.5589|
|DOC     |     .97571***       .14636        6.666   .0000     .31774|
|COMP    |     .40649***       .13927        2.919   .0035     .41786|
|LIB     |     .52144***       .17665        2.952   .0032     .13568|
|CI      |     .19873**        .09169        2.168   .0302    1.23116|
|CK      |     .08779          .13429         .654   .5133     .91998|
|PHD     |    -.13351          .10303       -1.296   .1951     .68612|
|NOEVAL  |   -1.93052***       .07239      -26.668   .0000     .29068|
+--------+---------------------------------------------------------+
| Note: ***, **, * = Significance at 1%, 5%, 10% level.            |
+------------------------------------------------------------------+

+----------------------------------------+
| Fit Measures for Binomial Choice Model |
| Probit   model for variable FINAL      |
+----------------------------------------+
|                Y=0       Y=1     Total|
| Proportions  .19714    .80286   1.00000|
| Sample Size     510      2077      2587|
+----------------------------------------+
| Log Likelihood Functions for BC Model  |
|              P=0.50    P=N1/N   P=Model|
| LogL =     -1793.17  -1284.22   -822.74|
+----------------------------------------+
| Fit Measures based on Log Likelihood   |
| McFadden = 1-(L/L0)            =   .35934|
| Estrella = 1-(L/L0)^(-2L0/n) =   .35729|
| R-squared (ML)                 =   .30006|
| Akaike Information Crit.       =   .64379|
| Schwartz Information Crit.     =   .66643|
+----------------------------------------+
| Fit Measures Based on Model Predictions|
| Efron                          =   .39635|
| Ben Akiva and Lerman           =   .80562|
| Veall and Zimmerman            =   .52781|
| Cramer                         =   .38789|
+----------------------------------------+
+-----------------------------------------------------------+
```

```
|Predictions for Binary Choice Model.  Predicted value is |
|1 when probability is greater than  .500000, 0 otherwise.|
|Note, column or row total percentages may not sum to     |
|100% because of rounding. Percentages are of full sample.|
+------+---------------------------------+----------------+
|Actual|         Predicted Value         |                |
|Value |       0                1        | Total Actual   |
+------+----------------+----------------+----------------+
|  0   |     342 ( 13.2%)|    168 (  6.5%)|    510 ( 19.7%)|
|  1   |     197 (  7.6%)|   1880 ( 72.7%)|   2077 ( 80.3%)|
+------+----------------+----------------+----------------+
|Total |     539 ( 20.8%)|   2048 ( 79.2%)|   2587 (100.0%)|
+------+----------------+----------------+----------------+
+---------------------------------------------------------+

|Crosstab for Binary Choice Model.  Predicted probability |
|vs. actual outcome. Entry = Sum[Y(i,j)*Prob(i,m)] 0,1.   |
|Note, column or row total percentages may not sum to     |
|100% because of rounding. Percentages are of full sample.|
+------+---------------------------------+----------------+
|Actual|      Predicted Probability      |                |
|Value |    Prob(y=0)        Prob(y=1)   | Total Actual   |
+------+----------------+----------------+----------------+
| y=0  |     259 ( 10.0%)|    250 (  9.7%)|    510 ( 19.7%)|
| y=1  |     252 (  9.7%)|   1824 ( 70.5%)|   2077 ( 80.2%)|
+------+----------------+----------------+----------------+
|Total |     512 ( 19.8%)|   2074 ( 80.2%)|   2587 ( 99.9%)|
+------+----------------+----------------+----------------+

=======================================================================
Analysis of Binary Choice Model Predictions Based on Threshold =  .5000
-----------------------------------------------------------------------
Prediction Success
-----------------------------------------------------------------------
Sensitivity = actual 1s correctly predicted                     87.819%
Specificity = actual 0s correctly predicted                     50.784%
Positive predictive value = predicted 1s that were actual 1s    87.946%
Negative predictive value = predicted 0s that were actual 0s    50.586%
Correct prediction = actual 1s and 0s correctly predicted       80.518%
-----------------------------------------------------------------------
Prediction Failure
-----------------------------------------------------------------------
False pos. for true neg. = actual 0s predicted as 1s            49.020%
False neg. for true pos. = actual 1s predicted as 0s            12.133%
False pos. for predicted pos. = predicted 1s actual 0s          12.054%
False neg. for predicted neg. = predicted 0s actual 1s          49.219%
False predictions = actual 1s and 0s incorrectly predicted      19.405%
=======================================================================


--> selection;lhs=change;rhs=one,hb,doc,comp,lib,ci,ck,phd,noeval;mle$


+----------------------------------------------------------+
| Sample Selection Model                                   |
| Probit selection equation based on FINAL                 |
| Selection rule is: Observations with FINAL    =  1       |
| Results of selection:                                    |
|                  Data points     Sum of weights          |
| Data set             2587            2587.0              |
| Selected sample      2077            2077.0              |
+----------------------------------------------------------+
```

```
+----------------------------------------------------+
| Sample Selection Model                             |
| Two step    least squares regression               |
| LHS=CHANGE   Mean                 =   5.456909     |
|              Standard deviation   =   4.582964     |
|              Number of observs.   =       2077     |
| Model size   Parameters           =         10     |
|              Degrees of freedom   =       2067     |
| Residuals    Sum of squares       =   39226.14     |
|              Standard error of e  =   4.356298     |
| Fit          R-squared            =   .0960355     |
|              Adjusted R-squared   =   .0920996     |
| Model test   F[  9,  2067] (prob) =  24.40 (.0000) |
| Diagnostic   Log likelihood       =  -5998.683     |
|              Restricted(b=0)      =  -6108.548     |
|              Chi-sq [  9]  (prob) = 219.73 (.0000) |
| Info criter. LogAmemiya Prd. Crt. =   2.948048     |
|              Akaike Info. Criter. =   2.948048     |
|              Bayes Info. Criter.  =   2.975196     |
| Not using OLS or no constant. Rsqd & F may be < 0. |
| Model was estimated Dec 08, 2009 at 00:12:49PM     |
| Standard error corrected for selection..   4.36303 |
| Correlation of disturbance in regression           |
| and Selection Criterion (Rho)...........    .11132 |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
|Constant|    6.74123***       .75107        8.976   .0000            |
|HB      |    -.01022*         .00563       -1.815   .0695     55.7429|
|DOC     |    2.07968***       .57645        3.608   .0003      .33558|
|COMP    |    -.32946          .44269        -.744   .4567      .40924|
|LIB     |    2.27448***       .53733        4.233   .0000      .14011|
|CI      |     .40823          .25929        1.574   .1154     1.22773|
|CK      |  -2.73074***       .37755       -7.233   .0000      .91815|
|PHD     |     .63345**        .29104        2.177   .0295      .69957|
|NOEVAL  |    -.88434         1.27223        -.695   .4870      .15744|
|LAMBDA  |     .48567         1.59683         .304   .7610      .21796|
+--------+------------------------------------------------------------+
| Note: ***, **, * = Significance at 1%, 5%, 10% level.               |
+---------------------------------------------------------------------+

Normal exit:  25 iterations. Status=0. F=    6826.467


-----------------------------------------------------------------
ML Estimates of Selection Model
Dependent variable                CHANGE
Log likelihood function      -6826.46734
Estimation based on N =   2587, K =  21
Information Criteria: Normalization=1/N
              Normalized   Unnormalized
AIC              5.29375    13694.93469
Fin.Smpl.AIC     5.29389    13695.29492
Bayes IC         5.34131    13817.95802
Hannan Quinn     5.31099    13739.52039
Model estimated: Mar 31, 2010, 15:17:41
FIRST 10 estimates are probit equation.
--------+--------------------------------------------------------
        |                  Standard           Prob.
  CHANGE| Coefficient        Error       z    z>|Z|
```

```
--------+----------------------------------------------------------
        |Selection (probit) equation for FINAL
Constant|     .99018***      .24020     4.12  .0000
      AN|     .02278**       .00940     2.42  .0153
      HB|    -.00489**       .00206    -2.37  .0178
     DOC|     .97154***      .15076     6.44  .0000
    COMP|     .40431***      .14433     2.80  .0051
     LIB|     .51505***      .19086     2.70  .0070
      CI|     .19927**       .09054     2.20  .0277
      CK|     .08590         .11902      .72  .4705
     PHD|    -.13208         .09787    -1.35  .1772
  NOEVAL|   -1.92902***      .07138   -27.03  .0000
        |Corrected regression, Regime 1
Constant|    6.81754***      .72389     9.42  .0000
      HB|    -.00978*        .00559    -1.75  .0803
     DOC|    1.99729***      .55348     3.61  .0003
    COMP|    -.36198         .43327     -.84  .4034
     LIB|    2.23154***      .50534     4.42  .0000
      CI|     .39401         .25339     1.55  .1199
      CK|   -2.74337***      .38031    -7.21  .0000
     PHD|     .64209**       .28964     2.22  .0266
  NOEVAL|    -.63201        1.26902     -.50  .6185
SIGMA(1)|    4.35713***      .07012    62.14  .0000
RHO(1,2)|     .03706         .35739      .10  .9174
--------+----------------------------------------------------------
```

The estimated probit model (as found on page 7) is

$$\begin{aligned}
&\textit{Estimated propensity to take the posttest} \;=\; 0.995 \;+\; 0.022\textit{(preTUCE score)} \\
&\quad -0.005\textit{(initial class size)} + 0.976\textit{(Doctoral Institution)} \\
&\quad +\; 0.406\;\textit{(Comprehensive Institution)} \;+\; 0.521\textit{(Liberal Arts Institution)} \\
&\quad +0.199\;\textit{(Male instructor)} \;+0.0878\textit{(English Instructor Native Language)} \\
&\quad -0.134\textit{(Instructor has PhD)} - \;1.930\textit{(No Evaluation of Instructor)}
\end{aligned}$$

The beginning or initial class size is negatively and highly significantly related to the propensity to take the posttest, with a one-tail p value of 0.0056.

The corresponding change-score equation employing the inverse Mills ratio is on page 9:

$$\begin{aligned}
&\text{Predicted Change} = \;6.741 \;-0.010\text{(initial class size)} + \;2.080\text{(Doctoral Institution)} \\
&\quad -\;0.329\;\text{(Comprehensive Institution)} \;+\; 2.274\;\text{Liberal Arts Institution)} \\
&\quad +\;.408\text{(Male instructor)} \;-\; 2.731\text{(English Instructor Native Language)} \\
&\quad +\;0.633\text{(Instructor has PhD)} \;-\; 0.88434\text{(No Evaluation of Instructor)} \;+0.486\,\lambda
\end{aligned}$$

The change score is negatively and significantly related to the class size, with a one-tail p value of 0.0347, but it takes an additional 100 students to lower the change score by a point.

Page 10 provides maximum likelihood estimation of both the probit equation and the change score equation with separate estimation of ρ and σ_ε. The top panel provides the probit coefficients for the propensity equation, where it is shown that initial class size is negatively and significantly related to the propensity to take the posttest with a one-tail p value of 0.009. The second panel gives the change score results, where initial class size is negatively and significantly related to the change score with a one-tail p value of 0.040. Again, it takes approximately 100 students to move the change score in the opposite direction by a point.

As a closing comment on the estimation of the Heckit model, it is worth pointing out that there is no unique way to estimate the standard errors via maximum likelihood computer routines. Historically, LIMDEP used the conventional second derivatives matrix to compute standard errors for the maximum likelihood estimation of the two-equation Heckit model. In the process of preparing this module, differences in standard errors produced by LIMDEP and STATA suggested that STATA was using the alternative outer products of the first derivatives. To achieve consistency, Bill Greene modified the LIMDEP routine in April 2010 so that it also now uses the outer products of the first derivatives.

AN APPLICATION OF PROPENSITY SCORE MATCHING

Unfortunately, we are not aware of a study in economic education for which propensity score matching has been used. Thus, we looked outside economic education and elected to redo the example reported in Becker and Ichino (2002). This application and data are derived from Dehejia and Wahba (1999), whose study, in turn was based on LaLonde (1986). The data set consists of observed samples of treatments and controls from the National Supported Work demonstration. Some of the institutional features of the data set are given by Becker and Ichino. The data were downloaded from the website http://www.nber.org/~rdehejia/nswdata.html. The data set used here is in the original text form, contained in the data file "matchingdata.txt." They have been assembled from the several parts in the NBER archive.

Becker and Ichino report that they were unable to replicate Dehejia and Wahba's results, though they did obtain similar results. (They indicate that they did not have the original authors' specifications of the number of blocks used in the partitioning of the range of propensity scores, significance levels, or exact procedures for testing the balancing property.) In turn, we could not precisely replicate Becker and Ichino's results – we can identify the reason, as discussed below. Likewise, however, we obtain similar results.

There are 2,675 observations in the data set, 2490 controls (with t = 0) and 185 treated observations (with t = 1). The variables in the raw data set are

t = treatment dummy variable
age = age in years
educ = education in years
black = dummy variable for black
hisp = dummy variable for Hispanic
marr = dummy variable for married
nodegree = dummy for no degree (not used)
*re*74 = real earnings in 1974
*re*75 = real earnings in 1975
*re*78 = real earnings in 1978 – the outcome variable

We will analyze these data following Becker and Ichino's line of analysis. We assume that you have completed Module One, Part Two, and thus are familiar with placing commands in the text editor and using the GO button to submit commands, and where results are found in the output window. In what follows, we will simply show the commands you need to enter into LIMDEP (NLOGIT) to produce the results that we will discuss.

To start, the data are imported by using the command (where the data file is on the C drive but your data could be placed wherever):

READ ; file=C:\matchingdata.txt;
names=t,age,educ,black,hisp,marr,nodegree,re74,re75,re78;nvar=10;nobs=2675$

Transformed variables added to the equation are

age2 = age squared
educ2 = educ squared
re742 = re74 squared
re752 = re75 squared
blacku74 = black times 1(re74 = 0)

In order to improve the readability of some of the reported results, we have divided the income variables by 10,000. (This is also an important adjustment that accommodates a numerical problem with the original data set. This is discussed below.) The outcome variable is re78.

The data are set up and described first. The transformations used to create the transformed variables are

CREATE ; age2 = age^2 ; educ2 = educ^2 $
CREATE ; re74 = re74/10000 ; re75 = re75/10000 ; re78 = re78/10000 $
CREATE ; re742 = re74^2 ; re752 = re75^2 $
CREATE ; blacku74 = black * (re74 = 0) $

The data are described with the following statistics:

DSTAT ; Rhs = * $

```
Descriptive Statistics
All results based on nonmissing observations.
==============================================================================
Variable     Mean       Std.Dev.     Minimum      Maximum        Cases Missing
==============================================================================
All observations in current sample
--------+---------------------------------------------------------------------
       T|  .691589E-01  .253772      .000000      1.00000         2675       0
     AGE|  34.2258      10.4998      17.0000      55.0000         2675       0
    EDUC|  11.9944      3.05356      .000000      17.0000         2675       0
   BLACK|  .291589      .454579      .000000      1.00000         2675       0
    HISP|  .343925E-01  .182269      .000000      1.00000         2675       0
    MARR|  .819439      .384726      .000000      1.00000         2675       0
NODEGREE|  .333084      .471404      .000000      1.00000         2675       0
    RE74|  1.82300      1.37223      .000000      13.7149         2675       0
    RE75|  1.78509      1.38778      .000000      15.6653         2675       0
    RE78|  2.05024      1.56325      .000000      12.1174         2675       0
    AGE2|  1281.61      766.842      289.000      3025.00         2675       0
   EDUC2|  153.186      70.6223      .000000      289.000         2675       0
```

```
   RE742|  5.20563      8.46589      .000000      188.098         2675       0
   RE752|  5.11175      8.90808      .000000      245.402         2675       0
BLACKU74|  .549533E-01  .227932      .000000      1.00000         2675       0
```

We next fit the logit model for the propensity scores. An immediate problem arises with the data set as used by Becker and Ichino. The income data are in raw dollar terms – the mean of re74, for example is $18,230.00. The square of it, which is on the order of 300,000,000, as well as the square of re75 which is similar, is included in the logit equation with a dummy variable for Hispanic which is zero for 96.5% of the observations and the blacku74 dummy variable which is zero for 94.5% of the observations. Because of the extreme difference in magnitudes, estimation of the logit model in this form is next to impossible. But rescaling the data by dividing the income variables by 10,000 addresses the instability problem.[ii] These transformations are shown in the second CREATE command above. This has no impact on the results produced with the data, other than stabilizing the estimation of the logit equation. We are now quite able to replicate the Becker and Ichino results except for an occasional very low order digit.

The logit model from which the propensity scores are obtained is fit using

NAMELIST ; X = age,age2,educ,educ2,marr,black,hisp,
re74,re75,re742,re752,blacku74,one $
LOGIT ; Lhs = t ; Rhs = x ; Hold $

(Note: Becker and Ichino's coefficients on re74 and re75 are multiplied by 10,000, and coefficients on re742 and re752 are multiplied by 100,000,000. Some additional logit results from LIMDEP are omitted. Becker and Ichino's results are included in the results for comparison.)

```
-----------------------------------------------------------------
Binary Logit Model for Binary Choice
Dependent variable                     T      Becker/Ichino
Log likelihood function      -204.97536       (-204.97537)
Restricted log likelihood    -672.64954       (identical)
Chi squared [  12 d.f.]       935.34837
Significance level               .00000
McFadden Pseudo R-squared      .6952717
Estimation based on N =   2675, K =  13
Information Criteria: Normalization=1/N
              Normalized   Unnormalized
AIC               .16297      435.95071
Fin.Smpl.AIC      .16302      436.08750
Bayes IC          .19160      512.54287
Hannan Quinn      .17333      463.66183
Hosmer-Lemeshow chi-squared =  12.77381
P-value=  .11987 with deg.fr. =       8
--------+--------------------------------------------------------
        |                Standard          Prob.       Mean
       T| Coefficient       Error       z    z>|Z|       of X
--------+--------------------------------------------------------   Becker/Ichino
        |Characteristics in numerator of Prob[Y = 1]                Coeff.    |t|
     AGE|     .33169***      .12033     2.76  .0058    34.2258   .3316904  (2.76)
```

```
    AGE2|    -.00637***      .00186    -3.43  .0006    1281.61  -.0063668  (3.43)
    EDUC|     .84927**       .34771     2.44  .0146    11.9944   .8492683  (2.44)
   EDUC2|    -.05062***      .01725    -2.93  .0033    153.186  -.0506202  (2.93)
    MARR|  -1.88554***       .29933    -6.30  .0000     .81944  -1.885542  (6.30)
   BLACK|   1.13597***       .35179     3.23  .0012     .29159   1.135973  (3.23)
    HISP|   1.96902***       .56686     3.47  .0005     .03439   1.969020  (3.47)
    RE74|  -1.05896***       .35252    -3.00  .0027    1.82300  -.1059000  (3.00)
    RE75|  -2.16854***       .41423    -5.24  .0000    1.78509  -.2169000  (5.24)
   RE742|     .23892***      .06429     3.72  .0002    5.20563   .2390000  (3.72)
   RE752|     .01359         .06654      .20  .8381    5.11175   .0136000  (0.21)
BLACKU74|    2.14413***      .42682     5.02  .0000     .05495   2.144129  (5.02)
Constant|  -7.47474***     2.44351    -3.06  .0022              -7.474742  (3.06)
--------+-------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------
+-----------------------------------------------------------+
|Predictions for Binary Choice Model.  Predicted value is |
|1 when probability is greater than  .500000, 0 otherwise.|
|Note, column or row total percentages may not sum to     |
|100% because of rounding. Percentages are of full sample.|
+------+---------------------------------+----------------+
|Actual|         Predicted Value         |                |
|Value |       0               1         | Total Actual   |
+------+----------------+----------------+----------------+
|  0   |   2463 ( 92.1%)|     27 (  1.0%)|   2490 ( 93.1%)|
|  1   |     51 (  1.9%)|    134 (  5.0%)|    185 (  6.9%)|
+------+----------------+----------------+----------------+
|Total |   2514 ( 94.0%)|    161 (  6.0%)|   2675 (100.0%)|
+------+----------------+----------------+----------------+
```

The first set of matching results uses the kernel estimator for the neighbors, lists the intermediate results, and uses only the observations in the common support.[iii]

MATCH ; Lhs = re78 ; Kernel ; List ; Common Support $

The estimated propensity score function is echoed first. This merely reports the earlier estimated binary choice model for the treatment assignment. The treatment assignment model is not reestimated. (The ;Hold in the LOGIT or PROBIT command stores the estimated model for this use.)

```
+----------------------------------------------------+
| ******* Propensity Score Matching Analysis ****** |
| Treatment variable = T       , Outcome = RE78     |
| Sample In Use                                      |
| Total number of observations     =   2675          |
| Number of valid (complete) obs.  =   2675          |
| Number used (in common support)  =   1342          |
| Sample Partitioning of Data In Use                 |
|                     Treated   Controls     Total  |
| Observations            185       1157      1342  |
| Sample Proportion    13.79%     86.21%   100.00%  |
+----------------------------------------------------+

+---------------------------------------------------------------+
```

```
| Propensity Score Function = Logit  based on T            |
| Variable   Coefficient  Standard Error  t statistic      |
| AGE             .33169       .12032986       2.757       |
| AGE2          -.00637       .00185539      -3.432       |
| EDUC            .84927       .34770583       2.442       |
| EDUC2         -.05062       .01724929      -2.935       |
| MARR         -1.88554       .29933086      -6.299       |
| BLACK         1.13597       .35178542       3.229       |
| HISP          1.96902       .56685941       3.474       |
| RE74         -1.05896       .35251776      -3.004       |
| RE75         -2.16854       .41423244      -5.235       |
| RE742           .23892       .06429271       3.716       |
| RE752           .01359       .06653758        .204       |
| BLACKU74      2.14413       .42681518       5.024       |
| ONE          -7.47474      2.44351058      -3.059       |
| Note:Estimation sample may not be the sample analyzed here. |
| Observations analyzed are restricted to the common support =|
| only controls with propensity in the range of the treated.  |
+-------------------------------------------------------------+
```

The note in the reported logit results reports how the common support is defined, that is, as the range of variation of the scores for the treated observations.

The next set of results reports the iterations that partition the range of estimated probabilities. The report includes the results of the *F* tests within the partitions as well as the details of the full partition itself. The balancing hypothesis is rejected when the p value is less than 0.01 within the cell. Becker and Ichino do not report the results of this search for their data, but do report that they ultimately found seven blocks, as we did. They do not report the means by which the test of equality is carried out within the blocks or the critical value used.

```
Partitioning the range of propensity scores
================================================================================
Iteration  1. Partitioning range of propensity scores into  5 intervals.
================================================================================
    Range                 Controls                Treatment
                  # Obs. Mean PS S.D. PS   # obs. Mean PS S.D. PS     F     Prob
----------------  ----------------------   ----------------------  -------------
  .00061  .19554    1081  .02111  .03337       17  .07358  .05835   13.68  .0020
  .19554  .39047      41  .28538  .05956       26  .30732  .05917    2.18  .1460
  .39047  .58540      15  .49681  .05098       20  .49273  .06228     .05  .8327
  .58540  .78033      13  .68950  .04660       19  .64573  .04769    6.68  .0157
  .78033  .97525       7  .96240  .00713      103  .93022  .05405   29.05  .0000
Iteration  1  Mean scores are not equal in at least one cell
================================================================================

Iteration  2. Partitioning range of propensity scores into  6 intervals.
================================================================================
    Range                 Controls                Treatment
                  # Obs. Mean PS S.D. PS   # obs. Mean PS S.D. PS     F     Prob
----------------  ----------------------   ----------------------  -------------
  .00061  .09807    1026  .01522  .02121       11  .03636  .03246    4.64  .0566
  .09807  .19554      55  .13104  .02762        6  .14183  .02272    1.16  .3163
  .19554  .39047      41  .28538  .05956       26  .30732  .05917    2.18  .1460
  .39047  .58540      15  .49681  .05098       20  .49273  .06228     .05  .8327
  .58540  .78033      13  .68950  .04660       19  .64573  .04769    6.68  .0157
```

```
  .78033  .97525       7  .96240  .00713      103  .93022  .05405   29.05  .0000
Iteration  2  Mean scores are not equal in at least one cell
================================================================================
Iteration  3. Partitioning range of propensity scores into  7 intervals.
================================================================================
    Range                Controls              Treatment
                  # Obs. Mean PS S.D. PS   # obs. Mean PS S.D. PS     F     Prob
----------------  ----------------------   ----------------------  -------------
  .00061  .09807    1026  .01522  .02121       11  .03636  .03246    4.64  .0566
  .09807  .19554      55  .13104  .02762        6  .14183  .02272    1.16  .3163
  .19554  .39047      41  .28538  .05956       26  .30732  .05917    2.18  .1460
  .39047  .58540      15  .49681  .05098       20  .49273  .06228     .05  .8327
  .58540  .78033      13  .68950  .04660       19  .64573  .04769    6.68  .0157
  .78033  .87779       0  .00000  .00000       17  .81736  .02800     .00 1.0000
  .87779  .97525       7  .96240  .00713       86  .95253  .01813    8.77  .0103
Mean PSCORES are tested equal within the blocks listed below
```

After partitioning the range of the propensity scores, we report the empirical distribution of the propensity scores and the boundaries of the blocks estimated above. The values below show the percentiles that are also reported by Becker and Ichino. The reported search algorithm notwithstanding, the block boundaries shown by Becker and Ichino shown below are roughly the same.

```
+-------------------------------------------------------------+
| Empirical Distribution of Propensity Scores in Sample Used  | Becker/Ichino
|   Percent      Lower   Upper   Sample size =   1342         | Percentiles (lower)
|   0% -   5%   .000611  .000801 Average score  .137746       | .0006426
|   5% -  10%   .000802  .001088 Std.Dev score  .274560       | .0008025
|  10% -  15%   .001093  .001378 Variance       .075383       | .0010932
|  15% -  20%   .001380  .001809 Blocks used to test balance  |
|  20% -  25%   .001815  .002355      Lower    Upper   # obs  |
|  25% -  30%   .002355  .003022  1  .000611  .098075   1037  | .0023546
|  30% -  35%   .003046  .004094  2  .098075  .195539     61  |
|  35% -  40%   .004097  .005299  3  .195539  .390468     67  |
|  40% -  45%   .005315  .007631  4  .390468  .585397     35  |
|  45% -  50%   .007632  .010652  5  .585397  .780325     32  |
|  50% -  55%   .010682  .015103  6  .780325  .877790     17  | .0106667
|  55% -  60%   .015105  .022858  7  .877790  .975254     93  |
|  60% -  65%   .022888  .035187                              |
|  65% -  70%   .035316  .051474                              |
|  70% -  75%   .051488  .075104                              |
|  75% -  80%   .075712  .135218                              | .0757115
|  80% -  85%   .135644  .322967                              |
|  85% -  90%   .335230  .616205                              |
|  90% -  95%   .625082  .949302                              | .6250832
|  95% - 100%   .949302  .975254                              | .949382 to .970598
+-------------------------------------------------------------+
```

The blocks used for the balancing hypothesis are shown at the right in the table above. Becker and Ichino report that they used the following blocks and sample sizes:

	Lower	Upper	Observations
1	0.0006	0.05	931
2	0.05	0.10	106
3	0.10	0.20	3
4	0.20	0.40	69
5	0.40	0.60	35

6	0.60	0.80	33
7	0.80	1.00	105

At this point, our results begin to differ somewhat from those of Becker and Ichino because they are using a different (cruder) blocking arrangement for the ranges of the propensity scores. This should not affect the ultimate estimation of the ATE; it is an intermediate step in the analysis that is a check on the reliability of the procedure.

The next set of results reports the analysis of the balancing property for the independent variables. A test is reported for each variable in each block as listed in the table above. The lines marked (by the program) with "*" show cells in which one or the other group had no observations, so the *F* test could not be carried out. This was treated as a "success" in each analysis. Lines marked with an "o" note where the balancing property failed. There are only four of these, but those we do find are not borderline. Becker and Ichino report their finding that the balancing property is satisfied. Note that our finding does not prevent the further analysis. It merely suggests to the analyst that they might want to consider a richer specification of the propensity function model.

```
Examining exogenous variables for balancing hypothesis
* Indicates no observations, treatment and/or controls, for test.
o Indicates means of treated and controls differ significantly.
=================================================================
Variable  Interval Mean Control   Mean Treated    F     Prob
--------  -------- ------------   ------------  ------  -----
AGE          1       31.459064      30.363636     .41  .5369
AGE          2       27.727273      26.500000     .10  .7587
AGE          3       28.170732      28.769231     .07  .7892
AGE          4       26.800000      25.050000     .44  .5096
AGE          5       24.846154      24.210526     .10  .7544
AGE          6         .000000      30.823529     .00 1.0000 *
AGE          7       23.285714      23.837209     .55  .4653
AGE2         1     1081.180312     953.454545    1.43  .2576
AGE2         2      822.200000     783.833333     .02  .8856
AGE2         3      873.341463     906.076923     .05  .8202
AGE2         4      774.400000     690.350000     .25  .6193
AGE2         5      644.230769     623.789474     .03  .8568
AGE2         6         .000000    1003.058824     .00 1.0000 *
AGE2         7      543.857143     596.023256    1.99  .1666
EDUC         1       11.208577      11.545455     .37  .5575
EDUC         2       10.636364      10.166667     .40  .5463
EDUC         3       10.414634      10.076923     .31  .5819
EDUC         4       10.200000      10.150000     .01 1.0000
EDUC         5       10.230769      11.000000    1.03  .3218
EDUC         6         .000000      11.058824     .00 1.0000 *
EDUC         7       10.571429      10.046512     .86  .3799
EDUC2        1      132.446394     136.636364     .11  .7420
EDUC2        2      117.618182     106.166667     .60  .4624
EDUC2        3      113.878049     107.769231     .31  .5829
EDUC2        4      108.066667     107.650000     .00 1.0000
EDUC2        5      109.923077     124.263158     .83  .3703
EDUC2        6         .000000     124.705882     .00 1.0000 *
EDUC2        7      113.714286     104.302326     .70  .4275
MARR         1         .832359        .818182     .01  .9056
MARR         2         .563636        .833333    2.63  .1433
MARR         3         .268293        .269231     .00 1.0000
MARR         4         .200000        .050000    1.73  .2032
```

```
MARR          5         .153846        .210526     .17  .6821
MARR          6         .000000        .529412     .00 1.0000 *
MARR          7         .000000        .000000     .00 1.0000
BLACK         1         .358674        .636364    3.63  .0833
BLACK         2         .600000        .500000     .22  .6553
BLACK         3         .780488        .769231     .01  .9150
BLACK         4         .866667        .500000    6.65  .0145
BLACK         5         .846154        .947368     .81  .3792
BLACK         6         .000000        .941176     .00 1.0000 *
BLACK         7        1.000000        .953488     .00 1.0000 *
HISP          1         .048733        .000000   52.46  .0000 o
HISP          2         .072727        .333333    1.77  .2311
HISP          3         .048780        .000000    2.10  .1547
HISP          4         .066667        .150000     .66  .4224
HISP          5         .153846        .052632     .81  .3792
HISP          6         .000000        .058824     .00 1.0000 *
HISP          7         .000000        .046512    4.19  .0436
RE74          1        1.230846       1.214261     .00 1.0000
RE74          2         .592119        .237027   10.63  .0041 o
RE74          3         .584965        .547003     .06  .8074
RE74          4         .253634        .298130     .16  .6875
RE74          5         .154631        .197888     .44  .5108
RE74          6         .000000        .002619     .00 1.0000 *
RE74          7         .000000        .000000     .00 1.0000
RE75          1        1.044680        .896447     .41  .5343
RE75          2         .413079        .379168     .09  .7653
RE75          3         .276234        .279825     .00 1.0000
RE75          4         .286058        .169340    2.39  .1319
RE75          5         .137276        .139118     .00 1.0000
RE75          6         .000000        .061722     .00 1.0000 *
RE75          7         .012788        .021539     .37  .5509
RE742         1        2.391922       2.335453     .00 1.0000
RE742         2         .672950        .092200    9.28  .0035 o
RE742         3         .638937        .734157     .09  .7625
RE742         4         .127254        .245461    1.14  .2936
RE742         5         .040070        .095745    1.31  .2647
RE742         6         .000000        .000117     .00 1.0000 *
RE742         7         .000000        .000000     .00 1.0000
RE752         1        1.779930       1.383457     .43  .5207
RE752         2         .313295        .201080    1.48  .2466
RE752         3         .151139        .135407     .14  .7133
RE752         4         .128831        .079975     .97  .3308
RE752         5         .088541        .037465     .51  .4894
RE752         6         .000000        .037719     .00 1.0000 *
RE752         7         .001145        .005973    2.57  .1124
BLACKU74      1         .014620        .000000   15.12  .0001 o
BLACKU74      2         .054545        .000000    3.17  .0804
BLACKU74      3         .121951        .192308     .58  .4515
BLACKU74      4         .200000        .100000     .66  .4242
BLACKU74      5         .230769        .315789     .29  .5952
BLACKU74      6         .000000        .941176     .00 1.0000 *
BLACKU74      7        1.000000        .953488     .00 1.0000 *
Variable BLACKU74 is unbalanced in block  1
Other variables may also be unbalanced
You might want to respecify the index function for the P-scores
```

This part of the analysis ends with a recommendation that the analyst reexamine the specification of the propensity score model. Because this is not a numerical problem, the analysis continues with estimation of the average treatment effect on the treated.

The first example below shows estimation using the kernel estimator to define the counterpart observation from the controls and using only the subsample in the common support. This stage consists of *nboot* + 1 iterations. In order to be able to replicate the results, we set the seed of the random number generator before computing the results.

CALC ; Ran(1234579) $
MATCH ; Lhs = re78 ; Kernel ; List ; Common Support $

The first result is the actual estimation, which is reported in the intermediate results. Then the *nboot* repetitions are reported. (These will be omitted if **; List** is not included in the command.) Recall, we divided the income values by 10,000. The value of .156255 reported below thus corresponds to \$1,562.55. Becker and Ichino report a value (see their section 6.4) of \$1537.94. Using the bootstrap replications, we have estimated the asymptotic standard error to be \$1042.04. A 95% confidence interval for the treatment effect is computed using \$1537.94 ± 1.96(1042.04) = (-\$325.41,\$3474.11).

```
+----------------------------------------------------------------------+
| Estimated Average Treatment Effect (T       )  Outcome is RE78       |
| Kernel            Using Epanechnikov kernel with bandwidth =  .0600  |
| Note, controls may be reused in defining matches.                    |
| Number of bootstrap replications used to obtain variance    =     25 |
+----------------------------------------------------------------------+
  Estimated average treatment effect =       .156255
  Begin bootstrap iterations *******************************************
  Boootstrap estimate   1           =       .099594
  Boootstrap estimate   2           =       .109812
  Boootstrap estimate   3           =       .152911
  Boootstrap estimate   4           =       .168743
  Boootstrap estimate   5           =      -.015677
  Boootstrap estimate   6           =       .052938
  Boootstrap estimate   7           =      -.003275
  Boootstrap estimate   8           =       .212767
  Boootstrap estimate   9           =      -.042274
  Boootstrap estimate  10           =       .053342
  Boootstrap estimate  11           =       .351122
  Boootstrap estimate  12           =       .117883
  Boootstrap estimate  13           =       .181123
  Boootstrap estimate  14           =       .111917
  Boootstrap estimate  15           =       .181256
  Boootstrap estimate  16           =      -.012129
  Boootstrap estimate  17           =       .240363
  Boootstrap estimate  18           =       .201321
  Boootstrap estimate  19           =       .169463
  Boootstrap estimate  20           =       .238131
  Boootstrap estimate  21           =       .358050
  Boootstrap estimate  22           =       .199020
  Boootstrap estimate  23           =       .083503
  Boootstrap estimate  24           =       .146215
  Boootstrap estimate  25           =       .266303
  End bootstrap iterations   *******************************************
+----------------------------------------------------------------------+
| Number of Treated observations =    185  Number of controls =   1157 |
| Estimated Average Treatment Effect   =        .156255                | (.153794)
| Estimated Asymptotic Standard Error  =        .104204                | (.101687)
| t statistic (ATT/Est.S.E.)           =       1.499510                |
```

```
| Confidence Interval for ATT = (      -.047985  to        .360496) 95% |
| Average Bootstrap estimate of ATT    =        .144897                 |
| ATT - Average bootstrap estimate     =        .011358                 |
+-----------------------------------------------------------------------+
```

Note that the estimated asymptotic standard error is somewhat different. As we noted earlier, because of differences in random number generators, the bootstrap replications will differ across programs. It will generally not be possible to exactly replicate results generated with different computer programs. With a specific computer program, replication is obtained by setting the seed of the random number generator. (The specific seed chosen is immaterial, so long as the same seed is used each time.)

The next set of estimates is based on all of the program defaults. The single nearest neighbor is used for the counterpart observation; 25 bootstrap replications are used to compute the standard deviation, and the full range of propensity scores (rather than the common support) is used. Intermediate output is also suppressed. Once again, we set the seed for the random number generator before estimation.

CALC ; Ran(1234579) $
MATCH ; Rhs = re78 $

```
Partitioning the range of propensity scores
Iteration  1  Mean scores are not equal in at least one cell
Iteration  2  Mean scores are not equal in at least one cell
Mean PSCORES are tested equal within the blocks listed below.

+-------------------------------------------------------------+
| Empirical Distribution of Propensity Scores in Sample Used  |
|   Percent      Lower   Upper   Sample size =   2675         |
|   0% -   5%   .000000  .000000 Average score  .069159       |
|   5% -  10%   .000000  .000002 Std.Dev score  .206287       |
|  10% -  15%   .000002  .000006 Variance       .042555       |
|  15% -  20%   .000007  .000015 Blocks used to test balance  |
|  20% -  25%   .000016  .000032      Lower    Upper   # obs  |
|  25% -  30%   .000032  .000064  1  .000000  .097525   2370  |
|  30% -  35%   .000064  .000121  2  .097525  .195051     60  |
|  35% -  40%   .000121  .000204  3  .195051  .390102     68  |
|  40% -  45%   .000204  .000368  4  .390102  .585152     35  |
|  45% -  50%   .000368  .000618  5  .585152  .780203     32  |
|  50% -  55%   .000618  .001110  6  .780203  .877729     17  |
|  55% -  60%   .001123  .001851  7  .877729  .975254     93  |
|  60% -  65%   .001854  .003047                              |
|  65% -  70%   .003057  .005451                              |
|  70% -  75%   .005451  .010756                              |
|  75% -  80%   .010877  .023117                              |
|  80% -  85%   .023149  .051488                              |
|  85% -  90%   .051703  .135644                              |
|  90% -  95%   .136043  .625082                              |
|  95% - 100%   .625269  .975254                              |
+-------------------------------------------------------------+
Examining exogenous variables for balancing hypothesis
Variable BLACKU74 is unbalanced in block  1
Other variables may also be unbalanced
You might want to respecify the index function for the P-scores

+-----------------------------------------------------------------------+
| Estimated Average Treatment Effect (T       )  Outcome is RE78        |
```

```
| Nearest Neighbor  Using average of  1 closest neighbors               |
| Note, controls may be reused in defining matches.                     |
| Number of bootstrap replications used to obtain variance    =     25 |
+---------------------------------------------------------------------+
  Estimated average treatment effect =       .169094
  Begin bootstrap iterations *****************************************
  End bootstrap iterations   *****************************************
+---------------------------------------------------------------------+
| Number of Treated observations =    185  Number of controls =     54 |
| Estimated Average Treatment Effect   =        .169094               |
| Estimated Asymptotic Standard Error  =        .102433               |
| t statistic (ATT/Est.S.E.)           =       1.650772               |
| Confidence Interval for ATT = (     -.031675  to        .369864) 95% |
| Average Bootstrap estimate of ATT    =        .171674               |
| ATT - Average bootstrap estimate     =       -.002579               |
+---------------------------------------------------------------------+
```

Using the full sample in this fashion produces an estimate of $1,690.94 for the treatment effect with an estimated standard error of $1,093.29. Note that from the results above, we find that only 54 of the 2490 control observations were used as nearest neighbors for the 185 treated observations. In comparison, using the 1,342 observations in their estimated common support, and the same 185 treateds, Becker and Ichino reported estimates of $1,667.64 and $2,113.59 for the effect and the standard error, respectively and use 57 of the 1,342 controls as nearest neighbors.

The next set of results uses the caliper form of matching and again restricts attention to the estimates in the common support.

CALC ; Ran(1234579) $
MATCH ; Rhs = re78 ; Range = .0001 ; Common Support $
CALC ; Ran(1234579) $
MATCH ; Rhs = re78 ; Range = .01 ; Common Support $

The estimated treatment effects are now very different. We see that only 23 of the 185 treated observations had a neighbor within a range (radius in the terminology of Becker and Ichino) of 0.0001. The treatment effect is estimated to be only $321.95 with a standard error of $307.95. In contrast, using this procedure, and this radius, Becker and Ichino report a nonsense result of -$5,546.10 with a standard error of $2,388.72. They state that this illustrates the sensitivity of the estimator to the choice of radius, which is certainly the case. To examine this aspect, we recomputed the estimator using a range of 0.01 instead of 0.0001. This produces the expected effect, as seen in the second set of results below. The estimated treatment effect rises to $1433.54 which is comparable to the other results already obtained

```
+---------------------------------------------------------------------+
| Estimated Average Treatment Effect (T       )  Outcome is RE78       |
| Caliper          Using distance of  .00010 to locate matches        |
| Note, controls may be reused in defining matches.                   |
| Number of bootstrap replications used to obtain variance    =     25 |
+---------------------------------------------------------------------+
  Estimated average treatment effect =       .032195
```

```
  Begin bootstrap iterations *******************************************
  End bootstrap iterations   *******************************************
+----------------------------------------------------------------------+
| Number of Treated observations =     23  Number of controls =     66 |
| Estimated Average Treatment Effect   =        .032195                |
| Estimated Asymptotic Standard Error  =        .030795                |
| t statistic (ATT/Est.S.E.)           =       1.045454                |
| Confidence Interval for ATT = (     -.028163  to        .092553) 95% |
| Average Bootstrap estimate of ATT    =        .018996                |
| ATT - Average bootstrap estimate     =        .013199                |
+----------------------------------------------------------------------+


+----------------------------------------------------------------------+
| Estimated Average Treatment Effect (T       )  Outcome is RE78       |
| Caliper           Using distance of  .01000 to locate matches        |
| Note, controls may be reused in defining matches.                    |
| Number of bootstrap replications used to obtain variance    =     25 |
+----------------------------------------------------------------------+
  Estimated average treatment effect =       .143354
  Begin bootstrap iterations *******************************************
  End bootstrap iterations   *******************************************
+----------------------------------------------------------------------+
| Number of Treated observations =    146  Number of controls =   1111 |
| Estimated Average Treatment Effect   =        .143354                |
| Estimated Asymptotic Standard Error  =        .078378                |
| t statistic (ATT/Est.S.E.)           =       1.829010                |
| Confidence Interval for ATT = (     -.010267  to        .296974) 95% |
| Average Bootstrap estimate of ATT    =        .127641                |
| ATT - Average bootstrap estimate     =        .015713                |
+----------------------------------------------------------------------+
```

CONCLUDING COMMENTS

Results obtained from the two equation system advanced by Heckman over 30 years ago are sensitive to the correctness of the equations and their identification. On the other hand, methods such as the propensity score matching depend on the validity of the logit or probit functions estimated along with the methods of getting smoothness in the kernel density estimator. Someone using Heckman's original selection adjustment method can easily have their results replicated in LIMDEP, STATA and SAS, although standard error estimates may differ somewhat because of the difference in routines used. Such is not the case with propensity score matching. Propensity score matching results are highly sensitive to the computer program employed while Heckman's original sample selection adjustment method can be relied on to give comparable coefficient estimates across programs.

REFERENCES

Becker, William and William Walstad. "Data Loss From Pretest to Posttest As a Sample Selection Problem," *Review of Economics and Statistics*, Vol. 72, February 1990: 184-188, Becker, William and John Powers. “Student Performance, Attrition, and Class Size Given Missing Student Data,” *Economics of Education Review*, Vol. 20, August 2001: 377-388.

Becker, S. and A. Ichino, “Estimation of Average Treatment Effects Based on Propensity Scores,” *The Stata Journal*, Vol. 2, November 2002: 358-377.

Deheija, R. and S. Wahba “Causal Effects in Nonexperimental Studies: Reevaluation of the Evaluation of Training Programs,” *Journal of the American Statistical Association*, Vol. 94, 1999: 1052-1062.

Heckman, James. Sample Bias as a Specific Error. *Econometrica,* Vol. 47, 1979: 153-162.

Huynh, Kim, David Jacho-Chavez, and James K. Self.“The Efficacy of Collaborative Learning Recitation Sessions on Student Outcomes?” *American Economic Review*, (Forthcoming May 2010).

LaLonde, R., “Evaluating the Econometric Evaluations of Training Programs with Experimental Data,” American Economic Review, Vol. 76, 4, 1986, 604-620.

Saunders, Phillip. The TUCE III Data Set: Background information and file codes (documentation, summary tables, and five 3.5-inch double-sided, high density disks in ASCII format). New York: National Council on Economic Education, 1994.

ENDNOTES

[i] Huynh, Jacho-Chavez, and Self (2010) have a data set that enables them to account for selection into, out of and between collaborative learning sections of a large principles course in their change-score modeling.

[ii] An attempt to compute a linear regression of the original RE78 on the original unscaled other variables is successful, but produces a warning that the condition number of the X matrix is 6.5 times 10^9. When the data are scaled as done above, no warning about multicollinearity is given.

[iii] The Kernel density estimator is a *nonparametric* estimator. Unlike a parametric estimator (which is an equation), a non-parametric estimator has no fixed structure and is based on a histogram of all the data. Histograms are bar charts, which are not smooth, and whose shape depends on the width of the bin into which the data are divided. In essence, with a fixed bin width, the kernel estimator smoothes out the histogram by centering each of the bins at each data point rather than fixing the end points of the bin. The optimum bin width is a subject of debate and well beyond the technical level of this module.