# Econometrics Qualifying Exam

May 2022

**Instructions**

This is a two-hour exam, but you have three hours to complete it. Answer all questions. You may answer out of order if you clearly label which question you are answering. For question 3, the associated Stata output is appended at the end of the exam. For all questions, rigorously show/justify all work (rule of thumb is to make sure enough is shown so that we can see how you got your answer). There are 100 points possible. Good luck!

## 1. Raising Tuition

How do increases in tuition affect undergraduate university enrollment? To answer this question, you have collected data on the total undergraduate enrollment of $i = 1, .., n$ four-year universities across $t = 1, .., T$ years, denoted by $y_{it}$, which you regress on the average cost of attending university $i$ in year $t$, denoted by $x_{it}$. The cost variable includes costs of tuition, housing, and food, but does *not* include financial aid. Specifically, you estimate

$$y_{it} = \beta_0 + \beta_1 x_{it} + e_{it}$$

where $e_{it} = y_{it} - \mathbf{E}[y_{it}|x_{i1}, ..., x_{iT}]$.

[a] Under what set of circumstances does an omitted variable bias arise?

[b] Identify two sources of omitted variable bias in the model above, *other* than the omission of financial aid. Many answers may be correct. Your answers will be evaluated based on your explanations.

For the rest of the question, suppose that you introduce control variables to address concerns about omitted variable bias.

[c] Suppose you observe a variable $z_{it}$ measuring the percentage of undergraduate students receiving financial aid, and the mean across $i$ and $t$, given by $\bar{z}$, is 40%. You add the term $\beta_2(z_{it} - \bar{z})x_{it}$ to the regression. Carefully interpret $\beta_1$ and $\beta_2$ in this regression. What do you expect of the sign on $\beta_2$?

[d] Public universities like ours charge two tuition rates. The in-state tuition rate is substantially lower than the out-of-state tuition rate. The out-of-state market is more competitive. The largest universities in the country are public universities. Give a reason why including an additive dummy/indicator/binary variable – i.e., $+\beta_3 D_{it}$ – for the type of university (public or private) would *not* be adequate to capture this distinction.

## 2. Nonlinearity

Consider two iid sequences of mean-zero random variables $(x_i)$ and $(y_i)$ such that

$$\mathbf{E}[y_i|x_i] = \beta_1 x_i - \exp(\beta_2)x_i^2$$

for $i = 1, ..., n$.

[a] Why might you estimate $-\exp(\beta_2)x_i^2$ instead of simply $\beta_3 x_i^2$ with $\beta_3 = -\exp(\beta_2)$?

[b] Show that the ordinary least squares estimator of the vector $\beta = (\beta_1, -\exp(\beta_2))$ is consistent. If you do not remember all of the formulas you need for a proof, then explain in words for partial credit.

[c] Can you estimate $\beta_2$ consistently using the consistent estimator in part [c]? Explain.

[d] Suppose you are only interested in estimating $\beta_1$, so you estimate a model with only $\beta_1 x_i$ and you omit $-\exp(\beta_2)x_i^2$. Is it possible for the ordinary least squares estimator to be consistent? Either show that consistency is impossible regardless of the properties of $(x_i)$ and $(y_i)$ or else provide a counterexample in which OLS is consistent.

### 3. Treatment Effect of Job Training Program

The following questions pertain to the Stata output generated from a subset of the observational data in Dhejia and Wahba (1999) on a job training program in the United States in the late 1970s. The outcome is *unem78*, a binary indicator if the man was unemployed in 1978 – after the training program took place in 1977. The key explanatory variable is *train*, indicating whether the man participated in the job training program.

[a] Consider the regression of *unem78* on *train*, *age*, *educ*, *black*, *hisp*, and *married*. What is the estimated effect of participating in the job training on the unemployment probability? Participation in the job training program is voluntary (not randomly assigned). Explain why this will cause endogeneity in this setting. What direction is the OLS estimate biased?

[b] Consider the proxy variables *re75* and *unem75*. State the two conditions needed for a proxy variable to be a valid proxy. Do you believe these two variables could be valid proxies? Explain. What is the estimated effect of participating in the job training on the unemployment probability when using these proxy variables?

[c] Now consider the maximum likelihood estimate of the probit model including *unem78* on *train*, *age*, *educ*, *black*, *hisp*, *married*, *re75*, and *unem75* as covariates. What is the estimated average treatment effect of participating in the job training? Compare the ATE estimates and standard errors to what you find in part [b].

[d] For the estimated model in part [c], what are the average treatment effects for individuals who are unemployed in 1975 vs the average treatment effects for individuals who are employed in 1975? Are they very similar? Very different? Explain by writing out the formula for the two average treatment effects.

[e] Now consider the probit model that also includes interactive effects between *train* and *unem75*. What are the average treatment effects for individuals who are unemployed in 1975 vs the average treatment effects for individuals who are employed in 1975? Compare to what you found in part [d] and explain.

## References

Dehejia, R.H. and Wahba, S.. "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs." Journal of the American Statistical Association 94, 1053-1062.

```
------------------------------------------------------------------------------------
      name:  <unnamed>
       log:  C:\Users\alyss\Dropbox\Teaching\Qualifying Exams\Spring2022\jtrain.log
  log type:  text
 opened on:  19 May 2022, 10:59:42

. describe

Contains data from JTRAIN3.dta
  obs:           806
 vars:            10                          19 May 2022 10:43
------------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
------------------------------------------------------------------------------------
train           byte    %9.0g                 =1 if in job training
age             byte    %9.0g                 in years, 1977
educ            byte    %9.0g                 years of schooling
black           byte    %9.0g                 =1 if black
hisp            byte    %9.0g                 =1 if Hispanic
married         byte    %9.0g                 =1 if married
re75            float   %9.0g                 '75 earnings, $1000s '82
unem75          byte    %9.0g                 =1 if unem. all of '75
re78            float   %9.0g                 '78 earnings, $1000s '82
unem78          byte    %9.0g                 =1 if unem. all of '78
------------------------------------------------------------------------------------
Sorted by:
     Note: Dataset has changed since last saved.

. summarize

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
       train |        806    .2233251    .4167331          0          1
         age |        806    32.48511    11.22372         17         55
        educ |        806    10.86725      3.0012          0         17
       black |        806    .4627792    .4989223          0          1
        hisp |        806    .0545906    .2273204          0          1
-------------+---------------------------------------------------------
     married |        806    .6488834    .4776157          0          1
        re75 |        806    3.088299    3.492438          0    9.93987
      unem75 |        806    .4466501    .4974543          0          1
        re78 |        806    7.988779    10.30933          0    121.174
      unem78 |        806    .3238213    .4682233          0          1

.
. // Part (a)
. reg unem78 train age educ black hisp married, vce(robust)

Linear regression                               Number of obs     =        806
```

```
                                                F(6, 799)         =      27.26
                                                Prob > F          =     0.0000
                                                R-squared         =     0.1552
                                                Root MSE          =     .43198

------------------------------------------------------------------------------
             |               Robust
      unem78 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       train |   .0613226   .0452368     1.36   0.176    -.0274744    .1501196
         age |    .017233   .0015084    11.42   0.000     .0142721     .020194
        educ |   .0160402   .0061042     2.63   0.009     .0040581    .0280223
       black |  -.0867674   .0349235    -2.48   0.013    -.1553201   -.0182148
        hisp |  -.1729479    .057347    -3.02   0.003    -.2855164   -.0603794
     married |  -.0805822   .0375103    -2.15   0.032    -.1542125   -.0069518
       _cons |  -.3221199   .1029239    -3.13   0.002     -.524153   -.1200868
------------------------------------------------------------------------------

.
. // Part (b)
. ivregress 2sls unem78 (train = re75 unem75) age educ black hisp married, vce(robus
> t)

Instrumental variables (2SLS) regression          Number of obs   =        806
                                                  Wald chi2(6)    =     177.47
                                                  Prob > chi2     =     0.0000
                                                  R-squared       =          .
                                                  Root MSE        =     .59373

------------------------------------------------------------------------------
             |               Robust
      unem78 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       train |   1.304458   .1744243     7.48   0.000     .9625926    1.646323
         age |   .0234272   .0020744    11.29   0.000     .0193615    .0274929
        educ |   .0217214   .0076952     2.82   0.005     .0066392    .0368037
       black |  -.3556239   .0601565    -5.91   0.000    -.4735285   -.2377193
        hisp |  -.3684556   .0955703    -3.86   0.000    -.5557699   -.1811413
     married |   .3637192   .0884255     4.11   0.000     .1904084      .53703
       _cons |  -1.015907   .1557649    -6.52   0.000    -1.321201   -.7106138
------------------------------------------------------------------------------
Instrumented:  train
Instruments:   age educ black hisp married re75 unem75

.
. regress unem78 train age educ black hisp married re75 unem75, vce(robust)

Linear regression                               Number of obs     =        806
                                                F(8, 797)         =      48.36
                                                Prob > F          =     0.0000
```

```
                                                R-squared         =     0.2768
                                                Root MSE          =     .40017

------------------------------------------------------------------------------
             |               Robust
      unem78 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       train |  -.1274731   .0493552    -2.58   0.010    -.2243547   -.0305915
         age |   .0102473   .0016449     6.23   0.000     .0070184    .0134761
        educ |   .0039438    .005904     0.67   0.504    -.0076454    .0155329
       black |  -.0447396   .0328418    -1.36   0.173    -.1092062     .019727
        hisp |  -.1305906   .0551496    -2.37   0.018    -.2388462   -.0223349
     married |  -.0593282   .0375249    -1.58   0.114    -.1329876    .0143311
        re75 |  -.0253681   .0064781    -3.92   0.000    -.0380843    -.012652
      unem75 |   .2171584   .0517864     4.19   0.000     .1155046    .3188122
       _cons |   .0242295   .1064316     0.23   0.820    -.1846899    .2331489
------------------------------------------------------------------------------

.
. // Part (c)
. probit unem78 i.train age educ i.black i.hisp i.married re75 i.unem75

Iteration 0:   log likelihood = -507.55141
Iteration 1:   log likelihood = -388.11949
Iteration 2:   log likelihood =  -386.6129
Iteration 3:   log likelihood =  -386.6086
Iteration 4:   log likelihood =  -386.6086

Probit regression                               Number of obs     =        806
                                                LR chi2(8)        =     241.89
                                                Prob > chi2       =     0.0000
Log likelihood =  -386.6086                     Pseudo R2         =     0.2383

------------------------------------------------------------------------------
      unem78 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     1.train |  -.3650487   .1662081    -2.20   0.028    -.6908105   -.0392869
         age |   .0353563   .0057181     6.18   0.000     .0241491    .0465634
        educ |   .0130961   .0192422     0.68   0.496     -.024618    .0508102
     1.black |  -.1403094   .1324467    -1.06   0.289    -.3999002    .1192815
      1.hisp |  -.6457841   .2735679    -2.36   0.018    -1.181967   -.1096008
   1.married |  -.2622174   .1370278    -1.91   0.056     -.530787    .0063521
        re75 |  -.1040066   .0277702    -3.75   0.000    -.1584352   -.0495781
    1.unem75 |   .5547259   .1748952     3.17   0.002     .2119377    .8975141
       _cons |  -1.455061   .3934711    -3.70   0.000    -2.226251   -.6838721
------------------------------------------------------------------------------

.
. margins, dydx(*)
```

```
Average marginal effects                        Number of obs     =        806
Model VCE    : OIM

Expression   : Pr(unem78), predict()
dy/dx w.r.t. : 1.train age educ 1.black 1.hisp 1.married re75 1.unem75

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     1.train |  -.0961086   .0423562    -2.27   0.023    -.1791252   -.0130919
         age |   .0095304   .0014285     6.67   0.000     .0067307    .0123301
        educ |   .0035301   .0051787     0.68   0.495      -.00662    .0136802
     1.black |  -.0381743   .0362787    -1.05   0.293    -.1092792    .0329306
      1.hisp |  -.1536766    .054956    -2.80   0.005    -.2613884   -.0459649
   1.married |  -.0700292    .036039    -1.94   0.052    -.1406643     .000606
        re75 |  -.0280354   .0073375    -3.82   0.000    -.0424166   -.0136542
    1.unem75 |     .16539   .0558689     2.96   0.003      .055889    .2748909
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.

. margins, dydx(*) atmeans

Conditional marginal effects                    Number of obs     =        806
Model VCE    : OIM

Expression   : Pr(unem78), predict()
dy/dx w.r.t. : 1.train age educ 1.black 1.hisp 1.married re75 1.unem75
at           : 0.train         =    .7766749 (mean)
               1.train         =    .2233251 (mean)
               age             =    32.48511 (mean)
               educ            =    10.86725 (mean)
               0.black         =    .5372208 (mean)
               1.black         =    .4627792 (mean)
               0.hisp          =    .9454094 (mean)
               1.hisp          =    .0545906 (mean)
               0.married       =    .3511166 (mean)
               1.married       =    .6488834 (mean)
               re75            =    3.088299 (mean)
               0.unem75        =    .5533499 (mean)
               1.unem75        =    .4466501 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     1.train |  -.1144115   .0483507    -2.37   0.018    -.2091771   -.0196458
         age |   .0118555   .0019086     6.21   0.000     .0081147    .0155963
        educ |   .0043913   .0064526     0.68   0.496    -.0082555    .0170381
     1.black |  -.0468777    .044066    -1.06   0.287    -.1332454      .03949
```

```
     1.hisp |  -.1746436   .0551029    -3.17   0.002    -.2826433   -.0666439
  1.married |  -.0897246   .0475727    -1.89   0.059    -.1829654    .0035163
       re75 |   -.034875   .0092079    -3.79   0.000    -.0529221   -.0168278
    1.unem75 |   .1875445   .0593636     3.16   0.002      .071194    .3038949
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.

. margins, atmeans

Adjusted predictions                              Number of obs     =        806
Model VCE    : OIM

Expression   : Pr(unem78), predict()
at           : 0.train         =    .7766749 (mean)
               1.train         =    .2233251 (mean)
               age             =    32.48511 (mean)
               educ            =    10.86725 (mean)
               0.black         =    .5372208 (mean)
               1.black         =    .4627792 (mean)
               0.hisp          =    .9454094 (mean)
               1.hisp          =    .0545906 (mean)
               0.married       =    .3511166 (mean)
               1.married       =    .6488834 (mean)
               re75            =    3.088299 (mean)
               0.unem75        =    .5533499 (mean)
               1.unem75        =    .4466501 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .2777679   .0181565    15.30   0.000     .2421819     .313354
------------------------------------------------------------------------------

.
.
. // Part (d)
. margins, dydx(train) at(unem75 = (0 1))

Average marginal effects                          Number of obs     =        806
Model VCE    : OIM

Expression   : Pr(unem78), predict()
dy/dx w.r.t. : 1.train

1._at        : unem75          =           0

2._at        : unem75          =           1

------------------------------------------------------------------------------
```

```
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
0.train      |  (base outcome)
-------------+----------------------------------------------------------------
1.train      |
         _at |
          1  |  -.0905206   .0387715    -2.33   0.020    -.1665115   -.0145298
          2  |  -.1176718   .0529663    -2.22   0.026    -.2214837   -.0138598
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.

.
. // Part (e)
. probit unem78 i.train#i.unem75 age educ i.black i.hisp i.married re75  i.unem75

Iteration 0:   log likelihood = -507.55141
Iteration 1:   log likelihood = -380.98751
Iteration 2:   log likelihood = -379.42299
Iteration 3:   log likelihood = -379.41727
Iteration 4:   log likelihood = -379.41727

Probit regression                               Number of obs     =        806
                                                LR chi2(9)        =     256.27
                                                Prob > chi2       =     0.0000
Log likelihood = -379.41727                     Pseudo R2         =     0.2525

------------------------------------------------------------------------------
      unem78 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
train#unem75 |
        1 0  |    .253371   .2259572     1.12   0.262     -.189497    .6962389
        1 1  |  -.7657804   .1987921    -3.85   0.000    -1.155406    -.376155
             |
         age |   .0336294   .0058061     5.79   0.000     .0222497    .0450091
        educ |   .0071471   .0195483     0.37   0.715    -.0311668    .0454611
     1.black |  -.1156055   .1337079    -0.86   0.387    -.3776682    .1464572
      1.hisp |  -.6497728   .2746377    -2.37   0.018    -1.188053   -.1114928
   1.married |  -.3171294   .1390778    -2.28   0.023    -.5897169   -.0445419
        re75 |  -.0704559   .0291213    -2.42   0.016    -.1275325   -.0133792
    1.unem75 |   .9686407   .2078686     4.66   0.000     .5612258    1.376056
       _cons |  -1.592083   .4000805    -3.98   0.000    -2.376226   -.8079397
------------------------------------------------------------------------------

.
. margins, dydx(train) at(unem75 = (0 1))

Average marginal effects                        Number of obs     =        806
Model VCE    : OIM
```

```
Expression   : Pr(unem78), predict()
dy/dx w.r.t. : 1.train

1._at        : unem75          =           0

2._at        : unem75          =           1

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
0.train      |  (base outcome)
-------------+----------------------------------------------------------------
1.train      |
         _at |
          1  |   .0696381   .0644983     1.08   0.280    -.0567762    .1960524
          2  |  -.2557733   .0655207    -3.90   0.000    -.3841915   -.1273552
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.

.
. log close
      name:  <unnamed>
       log:  C:\Users\alyss\Dropbox\Teaching\Qualifying Exams\Spring2022\jtrain.log
  log type:  text
 closed on:  19 May 2022, 10:59:44
------------------------------------------------------------------------------------
```