Szacowanie modeli dla nielosowej selekcji w pakiecie STATA

Paweł Strawiński

Uniwersytet Warszawski
Wydział Nauk Ekonomicznych

17 kwietnia 2005

1 Dane

Przedmiotem badania jest oszacowanie funkcji płacy dla mieszkańców mazowsza, oraz stwierdzenie czy z ekonomicznego punktu widzenia opłacalne jest uzyskanie wyższego wykształcenia.

Zostanie oszacowany następujący model płacy

$$ln(zarobki) = f(plec, wiek, wiek2, wwiek, wwiek2, rodzina)$$

Jest to równanie płacy bazujące na teorii kapitału ludzkiego. Jeśli współczynniki przy zmiennych `wwiek` oraz `wwiek2` okażą się większe od zera będzie oznaczało to, że wyższe wykształcenie ma pozytywny wpływ na wysokość płac oraz, że ludzie z wyższym wykształceniem szybciej zdobywają doświadczenie zawodowe.

Źródłem danych dla analizowanego przykładu jest część Badanie Budżetów Gospodarstw Domowych przeprowadzonego przez GUS w roku 2000. Z tego badania zostały wybrane pełne obserwacje dla województwa mazowieckiego.

```
Contains data from C:\stata8\DATA\hecklab.dta
  obs:        42,312                          Badanie budzetow gospodarstw
                                                domowych 2000
 vars:            14                          13 Apr 2005 20:17
 size:     1,100,112 (96.7% of memory free)
-------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
wiek            byte   %9.0g                  wiek w latach
ldonp           float  %9.0g                  logarytm dochodu osob
                                                niepracujacych
ld14            byte   %9.0g                  Liczba dzieci do 14 lat
samo            byte   %9.0g                  samozatrudnienie
```

```
wiek2           int    %9.0g                 kwadrat wieku
wyzsze          byte   %9.0g                 wyzsze wyksztalcenie
srednie         byte   %9.0g                 srednie wyksztalcenie
wwiek           byte   %9.0g                 wiek*wyzsze wyksztalcenie
wwiek2          int    %9.0g                 wiek*kwadrat wyzszego
                                               wyksztalcenia
lzarobki        float  %9.0g                 logarytm zarobkow
plec            byte   %9.0g      plec_ety   plec respondenta
rodzina         byte   %9.0g                 posiadanie rodziny
miastodu        byte   %9.0g                 duze miasto
select          byte   %9.0g
-------------------------------------------------------------------------------
```

Rozpatrywana jest próba z badania przekrojowego zawierająca informacje o 42312 osobach. Zmienna `select` jest zmienną pomocniczą, która wyróżnia obserwacje wybrane do modelu.

Dokonamy oszacowania modelu selekcji za pomocą kilku alternatywnych sposobów. Sprawdzimy jak wygląda rozkład zmiennej zależnej. Ponieważ nie wszystkie badane osoby pracują i osiągają z tego tytułu dochody, to zmienna logarytm zarobków `lzarobki` posiada braki. By się o tym przekonać wykonajmy

```
. count if lzarobki==.
31230
```

Wobec tego utworzymy zmienną pomocniczą `lz`, która posłuży do zrobienia słupkowego wykresu rozkładu wartości logarytmu dochodu

```
gen lz=lzarobki
recode lz .=0
histogram lz
```

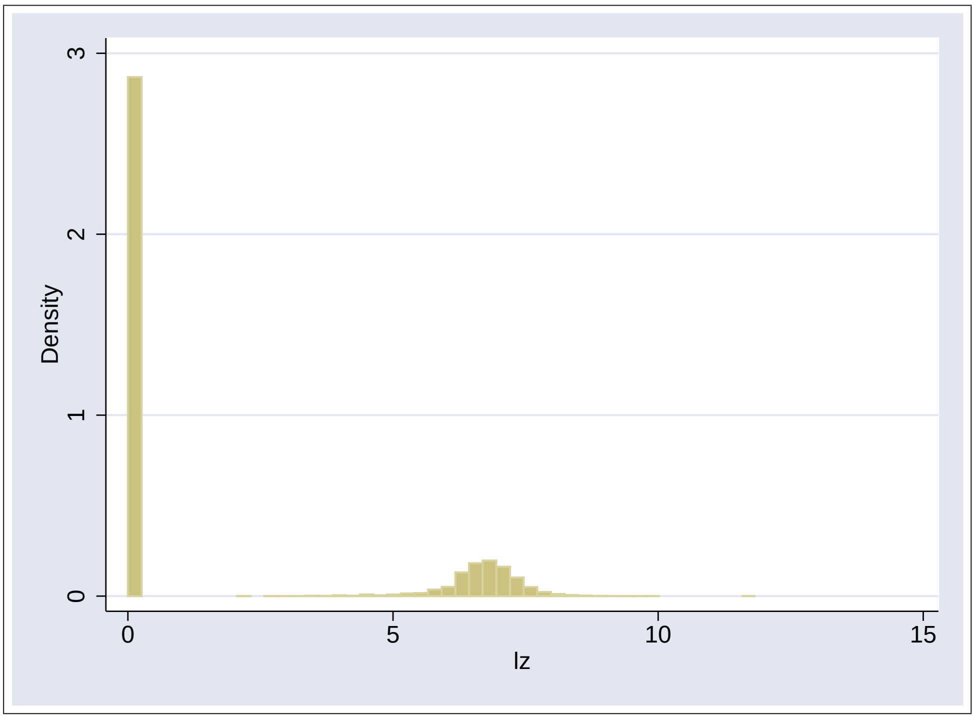
Nieobserwowane wartości zastępujemy zerami dla prostoty. Należy jednak pamiętać, że jeżeli osoba zarabia 0, to logarytm jej zarobków wynosi $-\infty$, i z tego powodu pakiet Stata traktuję te obserwacje jako brakujące.

Jak widzimy rozkład zarobków byłby logarytmo-normalny gdyby nie występowała duża liczba obserwacji zerowych.

2 Model tobitowy

Model tobitowy jest modelem regresji dla próby uciętej. Wybór tego modelu jest ekonomicznie uzasadniony, gdy uznamy że osoby nie pracujące nie podejmują pracy wskutek tego, że została zaoferowana im płaca niższa od płacy minimalnej. Jeśli tak rzeczywiście jest to powinniśmy obserwować dla każdej osoby płacę równą zero lub nie niższą niż minimalna.

W dostępnym zbiorze danych występują jednak zarobki niższe od minimalnych. Są one wynikiem pracy w niepełnym wymiarze godzinowym, lub pracy która jest dodatkiem do świadczeń socjalnych. Z uwagi na ich charakter nie biorą one udziału przy szacowaniu modelu, zmienna `tobit` przyjmuje dla nich wartość zero.



Opcja `ll(·)` ustawia punkt lewostronnego obcięcia próby zadeklarowaną wartość. Wszystkie wartości zmiennej `lz` o wartościach mniejszych bądź równych tej wartości nie są uwzględniane przy obliczaniu oszacowań parametrów modelu.

Ponieważ płaca minimalna według danych GUS w roku 2000 wynosiła przeciętnie 687,50 zł, rozsądne będzie obcięcie w modelu obserwacji o płacach niższych. Aby tego dokonać musimy najpierw obliczyć logarytm tej kwoty

```
. di ln(687.50)
 6.5330618
```

Wobec tego używamy obliczonej wartości w opcji `ll(·)`

```
. tobit lz plec wiek wiek2 wwiek wwiek2 rodzina if tobit==1,ll(6.53)

Tobit estimates                                   Number of obs   =      38679
                                                  LR chi2(6)      =   14597.80
                                                  Prob > chi2     =     0.0000
Log likelihood = -14021.496                       Pseudo R2       =     0.3423

------------------------------------------------------------------------------
          lz |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        plec |   .3577091   .0114236    31.31   0.000     .3353186    .3800996
        wiek |   .1734932   .0037761    45.95   0.000      .166092    .1808944
       wiek2 |  -.0023535   .0000464   -50.69   0.000    -.0024445   -.0022625
       wwiek |   .0210963   .0017792    11.86   0.000      .017609    .0245836
      wwiek2 |  -.0000759   .0000346    -2.20   0.028    -.0001436   -8.14e-06
     rodzina |   .2167354   .0147393    14.70   0.000      .187846    .2456248
       _cons |   2.966095   .0752042    39.44   0.000     2.818693    3.113497
```

```
-------------+----------------------------------------------------------------
        _se |   .6583414   .0060776           (Ancillary parameter)
------------------------------------------------------------------------------

  Obs. summary:      31230  left-censored observations at lz<=6.53
                      7449     uncensored observations
```

Jak widzimy aż 31230 obserwacji zostało ocenzurowanych. W rezultacie otrzymujemy dużo wyższe oszacowania nieznanych parametrów niż z modelu regresji liniowej. Wyniki interpretujemy identycznie jak w przypadku regresji liniowej. Uzyskane rezultaty śą zgodne z teorią ekonomiczną. Dodatni efekt dla mężczyzn, kwadratowa zależność płacy od wieku, i premia do zarobków jeśli się posiada rodzinę śą szeroko uzasadnione w literaturze ekonomicznej.

Nawet gdy ograniczymy regresję do niezerowych obserwacji to otrzymane estymatory nie będą zgodne.

```
. reg lz plec wiek wiek2 wwiek wwiek2 rodzina if lz>0 & tobit==1

      Source |       SS       df       MS              Number of obs =    7449
-------------+------------------------------           F(  6,  7442) =  159.66
       Model |  139.235784     6   23.205964           Prob > F      =  0.0000
    Residual |  1081.64353  7442  .145343124           R-squared     =  0.1140
-------------+------------------------------           Adj R-squared =  0.1133
       Total |  1220.87931  7448  .163920423           Root MSE      =  .38124

------------------------------------------------------------------------------
          lz |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        plec |   .1562638   .0097118    16.09   0.000      .137226    .1753016
        wiek |   .0201409   .0034538     5.83   0.000     .0133706    .0269113
       wiek2 |  -.0002226   .0000429    -5.20   0.000    -.0003066   -.0001386
       wwiek |   .0142162   .0013639    10.42   0.000     .0115427    .0168898
      wwiek2 |  -.0001428   .0000288    -4.96   0.000    -.0001992   -.0000864
     rodzina |   .0017956   .0136775     0.13   0.896     -.025016    .0286073
       _cons |   6.475064   .0685771    94.42   0.000     6.340634    6.609495
------------------------------------------------------------------------------
```

Jak widać uzyskane oszacowania są zdecydowanie niższe niż w modelu tobitowym.

2.1 Efekty krańcowe

Czasem głównym punktem zainteresowania badacza nie są same oszacowania nieznane parametry modelu, a np. wartość oczekiwana zarobków. W tym celu obliczymy wartości krańcowe funkcji prawodpodobieństwa (pochodną).

Po oszacowaniu modelu Tobitowego mogą zostać obliczone cztery rodzaje efektów krańcowych.

Po pierwsze efekt krańcowy obliczony w sposób tradycyjny

```
.mfx compute

Marginal effects after tobit
      y  = Fitted values (predict)
         =  5.2004885
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
   plec*|   .3577091      .01142   31.31   0.000   .335319  .380099    .43437
    wiek |   .1734932      .00378   45.95   0.000   .166092  .180894    53.677
   wiek2 |  -.0023535      .00005  -50.69   0.000  -.002445 -.002263   3161.98
   wwiek |   .0210963      .00178   11.86   0.000   .017609  .024583   3.56594
  wwiek2 |  -.0000759      .00003   -2.20   0.028  -.000144 -8.1e-06    203.92
 rodzina*|   .2167354      .01474   14.70   0.000   .187847  .245624   .684868
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

Jest to efekt krańcowy dla nieobserwowanej zmiennej y^*. Są to efekty cząstkowe dla zmiennej ukrytej. Nie posiadają one interpretacji ekonomicznej.

Po drugie efekt krańcowy prawdopodobieństwa nieocenzurowania, czyli:

$$\frac{\partial E(y > y_c \mid X)}{\partial x}$$

gdzie y_c jest punktem ocenzurowania. Do komendy `mfx compute` dodajemy opcję `predict(p(·,·))`. Jeśli · pojawia się po lewej stronie przecinka oznacza, że zmienna nie jest ograniczona z lewej strony, jeśli pojawia się po prawej stronie oznacza, że zmienna jest nieograniczona z góry. Jeśli zamiast kropki są wstawione liczby oznaczają one ocenzurowanie zmiennej zależnej odpowiednio z dołu lub z góry.

```
. mfx compute, predict(p(6.53,.))

Marginal effects after tobit
      y  = Pr(lz>6.53) (predict, p(6.53,.))
         =  .02171834
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
   plec*|   .0313834      .00179   17.56   0.000    .02788  .034886    .43437
    wiek |   .0136818      .00051   26.98   0.000   .012688  .014676    53.677
   wiek2 |  -.0001856      .00001  -27.52   0.000  -.000199 -.000172   3161.98
   wwiek |   .0016637      .00018    9.05   0.000   .001303  .002024   3.56594
  wwiek2 |  -5.98e-06      .00000   -2.10   0.036  -.000012 -4.0e-07    203.92
 rodzina*|   .0153136      .00121   12.61   0.000   .012933  .017694   .684868
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

Jeśli zwrócimy uwagę na lewą kolumnę zobaczymy, że przy nazwach wybranych zmiennych pojawiają się gwiazdki. Oznaczają one, że dana zmienna jest dyskretna, bądź jakościowa. W takim przypadku obliczenie pochodnej nie jest możliwe, i liczone są zmiany dyskretne.

$$\frac{E(y > y_c \mid X = 1) - E(y > y_c \mid X = 0)}{\Delta x}$$

Druga miara którą możemy obliczyć jest efekt krańcowy dla warunkowej wartości oczekiwanej pod warunkiem nieocenzurowania.

$$\frac{\partial E(y > y_c \mid X)}{\partial x}$$

```
. mfx compute, predict(e(6.53,.))

Marginal effects after tobit
      y  = E(lz|lz>6.53) (predict, e(6.53,.))
         =  6.7742446
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
   plec*|   .0413998        .0014   29.52   0.000   .038651  .044149    .43437
    wiek |   .0196273        .0003   65.56   0.000    .01904  .020214    53.677
   wiek2 |  -.0002663       .00000  -79.96   0.000  -.000273  -.00026   3161.98
   wwiek |   .0023866       .00021   11.24   0.000    .00197  .002803   3.56594
  wwiek2 |  -8.58e-06       .00000   -2.17   0.030  -.000016 -8.4e-07    203.92
 rodzina*|   .0237981        .0016   14.89   0.000   .020667   .02693   .684868
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

W takim wypadku współczynniki mają interpretację zmian wysokości logarytmu zarobków pod wpływem zmiany danego czynnika o 1 % pod warunkiem, że osoba ma pracę (zarobki > płacy minimalnej).

Ostatnią miarą są efekty krańcowe dla zmiennej ukrytej

$$\frac{\partial E(y^* \mid X)}{\partial x}$$

```
. mfx compute, predict (ys(6.53,.))

Marginal effects after tobit
      y  = E(lz*|lz>6.53) (predict, ys(6.53,.))
         =  6.5353046
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
   plec*|   .0089438       .00058   15.33   0.000     .0078  .010087    .43437
    wiek |    .003768       .00017   21.63   0.000   .003427  .004109    53.677
   wiek2 |  -.0000511       .00000  -21.87   0.000  -.000056 -.000047   3161.98
   wwiek |   .0004582       .00005    8.57   0.000   .000353  .000563   3.56594
  wwiek2 |  -1.65e-06       .00000   -2.08   0.037  -3.2e-06 -9.7e-08    203.92
 rodzina*|   .0041566       .00035   11.73   0.000   .003462  .004851   .684868
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

Wielkości oszacowań nie mają interpretacji.

Gdy mamy zainstalowany pakiet dodatkowy `dtobit`, to wywołując jego nazwę możemy uzyskać wszystkie powyższe oszacowania za pomocą jednej komendy.

```
Marginal Effects: Latent Variable
-----------------------------------------------------------------------------
variable |      dF/dx   Std. Err.       z    P>|z|     X_at    [    95% C.I.   ]
---------+-------------------------------------------------------------------
    plec*|   .3577091   .0114236    31.31   0.000   0 --> 1   .335319  .380099
    wiek |   .1734932   .0037761    45.95   0.000    53.677   .166092  .180894
   wiek2 |  -.0023535   .0000464   -50.69   0.000   3161.98  -.002445 -.002263
   wwiek |   .0210963   .0017792    11.86   0.000   3.56594   .017609  .024583
  wwiek2 |  -.0000759   .0000346    -2.20   0.028    203.92  -.000144 -8.1e-06
 rodzina*|   .2167354   .0147393    14.70   0.000   0 --> 1   .187847  .245624
   _cons |   2.966095   .0752042    39.44   0.000         1    2.8187  3.11349
-----------------------------------------------------------------------------
```

Pierwsza tabela jest równoważna wywołaniu komendy `mfx compute`. Są to efekty cząstkowe dla zmiennej ukrytej. Nie posiadają one interpretacji ekonomicznej.

```
Marginal Effects: Unconditional Expected Value
-----------------------------------------------------------------------------
variable |      dF/dx   Std. Err.       z    P>|z|     X_at    [    95% C.I.   ]
---------+-------------------------------------------------------------------
    plec*|   .0089438   .0002481    36.05   0.000   0 --> 1   .008457   .00943
    wiek |    .003768    .000082    45.95   0.000    53.677   .003607  .003929
   wiek2 |  -.0000511   1.01e-06   -50.69   0.000   3161.98  -.000053 -.000049
   wwiek |   .0004582   .0000386    11.86   0.000   3.56594   .000382  .000534
  wwiek2 |  -1.65e-06   7.51e-07    -2.20   0.028    203.92  -3.1e-06 -1.8e-07
 rodzina*|   .0041566   .0003201    12.98   0.000   0 --> 1   .003529  .004784
   _cons |   .0644187   .0016333    39.44   0.000         1   .061217   .06762
-----------------------------------------------------------------------------
```

Kolejnym wydrukiem są bezwarunkowe wartości oczekiwane. Możemy je również uzyskać komendą `. mfx compute, predict (ys(6.53,.))`. Współczynniki możemy interpretować jako oczekiwaną zmianę wartości zmiennej zależnej pod wpływem zmiany wartości przez zmienną niezależną o 1 %.

```
Marginal Effects: Conditional on being Uncensored
-----------------------------------------------------------------------------
variable |      dF/dx   Std. Err.       z    P>|z|     X_at    [    95% C.I.   ]
---------+-------------------------------------------------------------------
    plec*|   .0413998   .0012924    32.03   0.000   0 --> 1   .038867  .043933
    wiek |   .0196273   .0004272    45.95   0.000    53.677    .01879  .020465
   wiek2 |  -.0002663   5.25e-06   -50.69   0.000   3161.98  -.000277 -.000256
   wwiek |   .0023866   .0002013    11.86   0.000   3.56594   .001992  .002781
  wwiek2 |  -8.58e-06   3.91e-06    -2.20   0.028    203.92  -.000016 -9.2e-07
 rodzina*|   .0237981   .0016675    14.27   0.000   0 --> 1    .02053  .027066
   _cons |   .3355541   .0085078    39.44   0.000         1   .318879  .352229
-----------------------------------------------------------------------------
```

W trzeciej tabeli są krańcowe efekty obliczone pod warunkiem nieocenzurowania zmiennej. Obliczone efekty są identyczne jak dla komendy `.mfx compute, predict(e(6.53,.))`. Współczynniki interpretujemy jako zmianę zmiennej zależnej pod wpływem zmiany wartości zmiennej zależnej, o ile wartość tej ostatniej była obserwowana.

```
Marginal Effects: Probability Uncensored
------------------------------------------------------------------------------
variable |      dF/dx   Std. Err.      z    P>|z|     X_at   [    95% C.I.   ]
---------+--------------------------------------------------------------------
    plec*|   .0313834   .0009009    34.84   0.000   0 --> 1   .029618  .033149
    wiek |   .0136818   .0002978    45.95   0.000    53.677   .013098  .014265
   wiek2 |  -.0001856   3.66e-06   -50.69   0.000   3161.98  -.000193 -.000178
   wwiek |   .0016637   .0001403    11.86   0.000   3.56594   .001389  .001939
  wwiek2 |  -5.98e-06   2.73e-06    -2.20   0.028    203.92  -.000011 -6.4e-07
 rodzina*|   .0153136   .0011624    13.17   0.000   0 --> 1   .013035  .017592
   _cons |   .2339089   .0059307    39.44   0.000         1   .222285  .245533
------------------------------------------------------------------------------
(*) dF/dx is for discrete change of dummy variable from 0 to 1
```

Ostatnia tabela jest to efekt komendy `mfx compute, predict(p(6.53,.))`, czyli prawdopodobieństwo, że zmienna nie jest ocenzurowana.

3 Model Heckmana

3.1 Regresja liniowa

Na początku zostanie oszacowane równanie regresji liniowej. Jego zaletą jest prostota estymacji i łatwość interpretacji jego wyników.

```
. reg lz plec wiek wiek2 wwiek wwiek2 rodzina

      Source |       SS       df       MS              Number of obs =   42312
-------------+------------------------------           F(  6, 42305) = 2209.59
       Model |  88951.1878     6  14825.198            Prob > F      =  0.0000
    Residual |  283844.128 42305  6.70946999           R-squared     =  0.2386
-------------+------------------------------           Adj R-squared =  0.2385
       Total |  372795.316 42311  8.81083679           Root MSE      =  2.5903

------------------------------------------------------------------------------
          lz |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        plec |   .5765174   .0264844    21.77   0.000     .5246074    .6284274
        wiek |   .0226098    .004519     5.00   0.000     .0137524    .0314671
       wiek2 |  -.0008826   .0000437   -20.20   0.000    -.0009683    -.000797
       wwiek |   .1524604   .0043079    35.39   0.000     .1440167     .160904
      wwiek2 |  -.0021082   .0000695   -30.35   0.000    -.0022443    -.001972
     rodzina |   .4921381   .0302997    16.24   0.000     .4327501    .5515261
       _cons |   2.561404   .1063367    24.09   0.000     2.352982    2.769826
------------------------------------------------------------------------------
```

Jak widzimy mężczyźni zarabiają przeciętnie więcej, oraz rzeczywiście występuje premia za wyższe studia, ponieważ współczynnik przy zmiennej `wwiek` jest większy od zera i statystycznie istotny. Wpływ kwadratu wieku jest na tyle mały, że możemy go pominąć.

Jeśli przeprowadzimy diagnostykę modelu okaże się że występuje problem zmiennej pominiętej.

```
. ovtest, rhs
(note:  wiek2 dropped due to collinearity)
(note:  wiek2^2 dropped due to collinearity)
(note:  wwiek^2 dropped due to collinearity)
(note:  wwiek^4 dropped due to collinearity)

Ramsey RESET test using powers of the independent variables
       Ho:  model has no omitted variables
               F(9, 42297) =    819.10
                  Prob > F =      0.0000
```

Jednak gdyby problem ten nie występował, oraz składniki losowe równania doboru i równania zjawiska byłyby nieskorelowane, wówczas estymatory uzyskane metodą najmniejszych kwadratów byłyby zgodne i efektywne.

3.2 Estymacja oparta na MNK

3.2.1 Procedura dwustopniowa

Rozwiązanie problemu obciążenia estymatorów metody najmniejszych kwadratów zaproponował James Heckman. Zaproponował by potraktować obciążenie estymatora tak samo jak traktuje się model o błędnej specyfikacji.

Model szacowany jest za pomocą procedury dwustopniowej. W pierwszym kroku jest równanie uczestnictwa określające prawdopodobieństwo z jakim dana obserwacja trafia do modelu. Następnie prawdopodobieństwa są przekształcane w tzw. odwrócone ilorazy Millsa i są używane jako dodatkowa zmienna objaśniająca w równaniu zjawiska.

```
. heckman lzarobki plec wiek wiek2 wwiek wwiek2 rodzina,
select(plec wiek wiek2 wwiek wwiek2 rodzina ld14 ldonp) twostep

Heckman selection model -- two-step estimates   Number of obs      =     42312
(regression model with sample selection)        Censored obs       =     31230
                                                Uncensored obs     =     11082

                                                Wald chi2(12)      =   6296.60
                                                Prob > chi2        =    0.0000

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
lzarobki     |
        plec |   .2853284   .0139103    20.51   0.000     .2580648    .3125921
```

```
        wiek |   .0960455    .005161    18.61   0.000     .0859301     .106161
       wiek2 |  -.0012665   .0000655   -19.33   0.000    -.0013949   -.0011381
       wwiek |   .0198695   .0018766    10.59   0.000     .0161915    .0235475
      wwiek2 |  -.0001396   .0000374    -3.73   0.000    -.0002129   -.0000662
     rodzina |   .1331675   .0177234     7.51   0.000     .0984302    .1679048
       _cons |   4.607437   .1132801    40.67   0.000     4.385412    4.829462
-------------+----------------------------------------------------------------
select       |
        plec |   .4187567   .0165401    25.32   0.000     .3863387    .4511747
        wiek |   .1920273   .0042093    45.62   0.000     .1837772    .2002775
       wiek2 |  -.0024844   .0000485   -51.23   0.000    -.0025794   -.0023893
       wwiek |   .0439469   .0030818    14.26   0.000     .0379066    .0499871
      wwiek2 |  -.0004023   .0000543    -7.41   0.000    -.0005088   -.0002959
     rodzina |   .1421164   .0204784     6.94   0.000     .1019794    .1822533
        ld14 |   .0743513   .0082463     9.02   0.000     .0581888    .0905138
       ldonp |  -.3849657   .0087019   -44.24   0.000    -.4020211   -.3679103
       _cons |  -1.765745   .0971778   -18.17   0.000     -1.95621    -1.57528
-------------+----------------------------------------------------------------
mills        |
      lambda |   .1043912   .0215924     4.83   0.000      .062071    .1467115
-------------+----------------------------------------------------------------
         rho |    0.16139
       sigma |  .64684178
      lambda |  .10439121   .0215924
------------------------------------------------------------------------------
```

Na początku zwróćmy uwagę na górną część wydruku.

```
. heckman lzarobki plec wiek wiek2 wwiek wwiek2 rodzina,
select(plec wiek wiek2 wwiek wwiek2 rodzina ld14 ldonp) twostep
```

Po poleceniu `heckman` zapisane jest główne równanie modelu, następnie opcja `select` definiuje równanie doboru obserwacji do próby. To równanie ma domyślną zmienną objaśniającą `lzarobki`. W celu zapewnienia identyfikacji równania selekcji i zjawiska powinny różnić się conajmniej jedną zmienną objaśniającą. Eliminując zmienne z równań należy kierować się teorią ekonomiczną i zdrowym rozsądkiem. Należy również pamiętać jakie konsekwencje niesie nałożenie ograniczenia zerowego na daną zmienną. W równaniu zarobków nie uwzględniona została zmienna oznaczająca liczbę dzieci do lat 14 w gospodarstwie domowym `ld14`, bowiem trudno przypuszczać by wpływała ona na zarobki. Inną zmienną nie posiadającą wpływu na wysokość zarobków, a decydującą o podjęciu bądź niepodjęciu pracy jest logarytm dochodu pozapłacowego `ldonp`.

Następnie analizujemy własności statystyczne modelu

```
Heckman selection model -- two-step estimates   Number of obs      =     42312
(regression model with sample selection)        Censored obs       =     31230
                                                Uncensored obs     =     11082

                                                Wald chi2(12)      =   6296.60
                                                Prob > chi2        =    0.0000
```

Informuje nas ona, że estymacja została wykonana za pomocą procedury dwustopniowej regresji z selekcją próby. Zbiór liczył 42312 obserwacji, z czego 31230 zostało ocenzurowanych, a na podstawie pozostałych 11082 zbudowano równanie płacy. Statystyka Walda informuje nas o tym, że zmienne w równaniu płacy są łącznie istotne.

Następnie zajmijmy się dolną częścią

```
-------------+----------------------------------------------------------------
mills        |
      lambda |   .1043912   .0215924     4.83   0.000      .062071    .1467115
-------------+----------------------------------------------------------------
         rho |    0.16139
       sigma |  .64684178
      lambda |  .10439121   .0215924
------------------------------------------------------------------------------
```

Parametr lambda jest estymatorem "zmiennej pominiętej". Jest on na poziomie istotności 7.4 % różny od zera, co świadczy o tym, że chociaż zjawisko nielosowej selekcji jest słabe jednak występuje. `rho` jest oszacowaniem współczynnika korelacji między składnikami losowymi obu równań, a `sigma` jest oszacowaniem odchylenia standardowego składnika losowego w równaniu selekcji.

Wartości oszacowań parametrów probitowego równania selekcji nie mają interpretacji. Ważne jest by były łącznie istotne. Natomiast równanie zjawiska intepretujemy standardowo.

```
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
lzarobki     |
        plec |   .2853284   .0139103    20.51   0.000     .2580648    .3125921
        wiek |   .0960455    .005161    18.61   0.000     .0859301     .106161
       wiek2 |  -.0012665   .0000655   -19.33   0.000    -.0013949   -.0011381
       wwiek |   .0198695   .0018766    10.59   0.000     .0161915    .0235475
      wwiek2 |  -.0001396   .0000374    -3.73   0.000    -.0002129   -.0000662
     rodzina |   .1331675   .0177234     7.51   0.000     .0984302    .1679048
       _cons |   4.607437   .1132801    40.67   0.000     4.385412    4.829462
-------------+----------------------------------------------------------------
```

Pracujący mężczyźni zarabiają przeciętnie 30 % więcej niż kobiety ($exp(0.28) \approx 0.3$), pracownicy posiadający rodzinę zarabiają przeciętnie o 14 % ($exp(0.13) \approx 0.14$) więcej niż osoby samotne. Interpretacja współczynników przy zmiennych odnoszących się do wieku jest bardziej skomplikowana. Można jedynie powiedzieć, że płaca rośnie wraz z wiekiem pracownika do pewnego momentu potem zaczyna spadać, oraz że ludzie z wyższym wykształceniem cieszą się szybszym wzrostem płac od pozostałych pracowników.

3.2.2 Odtworzenie procedury Heckmana

Identyczne wyniki można osiągnąć posługując się instrukcją zawartą w artykule Heckmana. Na początku szacujemy probitowe równanie uczestnictwa. Tłumaczymy zmienną pomocniczą `select` za pomocą czynników, które uznaliśmy za determinujące dobór do próby.

```
. probit select plec wiek wiek2 wwiek wwiek2 rodzina ld14 ldonp

Iteration 0:   log likelihood = -24331.381
Iteration 1:   log likelihood = -17267.802
Iteration 2:   log likelihood = -16106.018
Iteration 3:   log likelihood = -15810.048
Iteration 4:   log likelihood = -15786.086
Iteration 5:   log likelihood = -15785.916
Iteration 6:   log likelihood = -15785.916

Probit estimates                                  Number of obs   =      42312
                                                  LR chi2(8)      =   17090.93
                                                  Prob > chi2     =     0.0000
Log likelihood = -15785.916                       Pseudo R2       =     0.3512

------------------------------------------------------------------------------
      select |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        plec |   .4187567   .0165401    25.32   0.000     .3863387    .4511747
        wiek |   .1920273   .0042093    45.62   0.000     .1837772    .2002775
       wiek2 |  -.0024844   .0000485   -51.23   0.000    -.0025794   -.0023893
       wwiek |   .0439469   .0030818    14.26   0.000     .0379066    .0499871
      wwiek2 |  -.0004023   .0000543    -7.41   0.000    -.0005088   -.0002959
     rodzina |   .1421164   .0204784     6.94   0.000     .1019794    .1822533
        ld14 |   .0743513   .0082463     9.02   0.000     .0581888    .0905138
       ldonp |  -.3849657   .0087019   -44.24   0.000    -.4020211   -.3679103
       _cons |  -1.765745   .0971778   -18.17   0.000     -1.95621    -1.57528
------------------------------------------------------------------------------

note: 496 failures and 0 successes completely determined.
```

Uzyskawszy oszacowania modelu probitowego generujemy wartości dopasowane

```
predict selecthat, xb
```

a następnie ilorazy Millsa

```
gen lambda = normden(selecthat) / norm(selecthat)
```

ostatnim krokiem jest dołączenie zmiennej `lambda` do równania płacy i wykonanie regresji na podstawie tej części próby dla której znamy wartość logarytmu zarobków.

```
. reg lzarobki plec wiek wiek2 wwiek wwiek2 rodzina lambda

      Source |       SS       df       MS              Number of obs =   11082
-------------+------------------------------           F(  7, 11074) =  281.34
       Model |  812.074557     7  116.010651           Prob > F      =  0.0000
    Residual |  4566.40334 11074  .412353562           R-squared     =  0.1510
-------------+------------------------------           Adj R-squared =  0.1504
       Total |   5378.4779 11081  .485378386           Root MSE      =  .64215

------------------------------------------------------------------------------
    lzarobki |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
```

```
    plec |   .2853284   .0138707    20.57   0.000     .2581394    .3125174
    wiek |   .0960455     .00515    18.65   0.000     .0859506    .1061405
   wiek2 |  -.0012665   .0000654   -19.37   0.000    -.0013947   -.0011384
   wwiek |   .0198695   .0018699    10.63   0.000     .0162043    .0235348
  wwiek2 |  -.0001396   .0000373    -3.74   0.000    -.0002127   -.0000664
 rodzina |   .1331675   .0176904     7.53   0.000     .0984911    .1678438
  lambda |   .1043912    .021504     4.85   0.000     .0622396    .1465428
   _cons |   4.607437   .1129889    40.78   0.000     4.385958    4.828915
------------------------------------------------------------------------------
```

Jak widać uzyskano dokładnie takie same oszacowania nieznanych parametrów modelu. Ewentualne niewielkie odchylenia wartości parametrów oraz ich błędów standardowych są wynikiem stosowania różnych metod szacowania macierzy wariancji-kowariancji składnika losowego. W przypadku polecenia `heckman` są one obliczone dużo bardziej skomplikowaną metodą, niż w procedurze dwukrokowej. Mimo wszystko rozbieżności są małe.

Oszacowania parametrów uzyskane metodą najmniejszych kwadratów są zgodne, ale nie są efektywne.

3.3 MNW

3.3.1 MNW odtwarzające wyniki MNK

Za pomocą metody największej wiarogodności można w prosty sposób odtworzyć wyniki otrzymane metodą największych kwadratów. Aby uzyskać ten efekt należy ograniczyć liczbę iteracji do zera dodając do polecenia `heckman` opcję `iterate(0)`

```
. heckman lzarobki plec wiek wiek2 wwiek wwiek2 rodzina,
select(plec wiek wiek2 wwiek wwiek2 rodzina ld14 ldonp) iterate(0)
Iteration 0:   log likelihood = -26606.877
convergence not achieved

Heckman selection model                         Number of obs      =     42312
(regression model with sample selection)        Censored obs       =     31230
                                                Uncensored obs     =     11082

                                                Wald chi2(6)       =   1415.55
Log likelihood = -26606.88                      Prob > chi2        =    0.0000

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
lzarobki     |
        plec |   .2853284   .0134532    21.21   0.000     .2589607    .3116962
        wiek |   .0960455   .0045514    21.10   0.000     .0871249    .1049661
       wiek2 |  -.0012665   .0000564   -22.45   0.000    -.0013771    -.001156
       wwiek |   .0198695   .0018721    10.61   0.000     .0162002    .0235388
      wwiek2 |  -.0001396   .0000374    -3.73   0.000    -.0002129   -.0000662
     rodzina |   .1331675   .0175083     7.61   0.000     .0988519     .167483
       _cons |   4.607437   .0975805    47.22   0.000     4.416182    4.798691
-------------+----------------------------------------------------------------
```

```
select       |
        plec |   .4187567   .0165362    25.32   0.000     .3863464     .451167
        wiek |   .1920273   .0042082    45.63   0.000     .1837794    .2002752
       wiek2 |  -.0024844   .0000485   -51.25   0.000    -.0025794   -.0023893
       wwiek |   .0439469   .0030805    14.27   0.000     .0379092    .0499846
      wwiek2 |  -.0004023   .0000543    -7.41   0.000    -.0005088   -.0002959
     rodzina |   .1421164    .020474     6.94   0.000      .101988    .1822447
        ld14 |   .0743513   .0082102     9.06   0.000     .0582595     .090443
       ldonp |  -.3849657   .0087291   -44.10   0.000    -.4020745    -.367857
       _cons |  -1.765745   .0972018   -18.17   0.000    -1.956257   -1.575233
-------------+----------------------------------------------------------------
     /athrho |   .1628094   .0237722     6.85   0.000     .1162167    .2094021
    /lnsigma |  -.4356536   .0071045   -61.32   0.000    -.4495782   -.4217289
-------------+----------------------------------------------------------------
         rho |    .161386   .0231531                      .1156963    .2063942
       sigma |   .6468418   .0045955                      .6378972    .6559118
      lambda |   .1043912    .015237                      .0745271    .1342553
------------------------------------------------------------------------------
LR test of indep. eqns. (rho = 0):   chi2(1) =     5.73   Prob > chi2 = 0.0167
------------------------------------------------------------------------------
```

W rezultacie otrzymaliśmy identyczne oszacowania jak w pierwszym analizowanym modelu selekcji próby. Metoda największej wiarogodności ma tę zaletę, ze jednocześnie szukane są estymatory dla parametrów obu równań. Pozwala to na przeprowadzenie metodą ilorazu wiarogodności test niezależności równań. Jak widać w ostatniej linii wydruku odrzucamy hipotezę, o tym że parametr `rho` jest równy zero, na każdym poziomie istotności. W analizowanym modelu występuje niezerowa korelacja między składnikami losowymi równania doboru i równania zjawiska. Wobec tego zastosowanie modelu Heckmana jest uzasadnione.

3.3.2 MNW

Komendę szacująca model Heckmana pełną metodą największej wiarogodności można zapisać używając dwóch równoważnych sposobów

```
.heckman lzarobki plec wiek wiek2 wwiek wwiek2 rodzina,
select(plec wiek wiek2 wwiek wwiek2 rodzinald14 ldonp)
```

albo alternatywnie

```
.heckman lzarobki plec wiek wiek2 wwiek wwiek2 rodzina,
select(select=plec wiek wiek2 wwiek wwiek2 rodzina ld14 ldonp)
```

W pierwszym zapisie zmienną zależną równania selekcji jest zmienna `lzarobki`. Gdy posiada ona wartość to obserwacja trafia do próby na podstawie której szacowany jest model. Czasami jednak braki w obserwacjach zapisane są jako 0 lub bardzo duża liczba dodatnia. Z tego powodu istnieje drugi sposób wywołania komendy `heckman`. Istnieje możliwość samodzielnego wygenerowania zmiennej zero-jedynkowej która decyduje czy na podstawie danej obserwacji ma być szacowane równanie zjawiska czy też nie. W takim przypadku używamy alternatywnego sposobu wywołania polecenia `heckman`. Ten sposób jest przydatny gdy np. chcemy sprawdzić czy jedna z obserwacji w naszym zbiorze nie jest nietypowa bez usuwania jej.

Zbiór danych do ćwiczeń został utworzony w taki sposób, że oba wywołania dają taki sam rezultat.

```
Iteration 0:   log likelihood = -26606.877
Iteration 1:   log likelihood = -26602.837
Iteration 2:   log likelihood = -26602.804
Iteration 3:   log likelihood = -26602.804

Heckman selection model                         Number of obs      =     42312
(regression model with sample selection)        Censored obs       =     31230
                                                Uncensored obs     =     11082

                                                Wald chi2(6)       =   1254.52
Log likelihood =  -26602.8                      Prob > chi2        =    0.0000

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
lzarobki     |
        plec |   .2751912   .0134784    20.42   0.000     .2487741    .3016084
        wiek |   .0890127   .0046028    19.34   0.000     .0799914    .0980339
       wiek2 |  -.0011702   .0000572   -20.46   0.000    -.0012823   -.0010581
       wwiek |    .019494   .0018688    10.43   0.000     .0158311    .0231568
      wwiek2 |   -.000142   .0000374    -3.80   0.000    -.0002153   -.0000688
     rodzina |    .125106   .0175096     7.14   0.000     .0907878    .1594242
       _cons |   4.773756   .0988922    48.27   0.000     4.579931    4.967581
-------------+----------------------------------------------------------------
select       |
        plec |   .4196785   .0165414    25.37   0.000     .3872581     .452099
        wiek |    .192081   .0042088    45.64   0.000     .1838319    .2003301
       wiek2 |  -.0024843   .0000485   -51.24   0.000    -.0025793   -.0023893
       wwiek |   .0439072    .003081    14.25   0.000     .0378686    .0499458
      wwiek2 |  -.0004016   .0000543    -7.40   0.000    -.0005081   -.0002952
     rodzina |   .1411091   .0204774     6.89   0.000     .1009742     .181244
        ld14 |    .075616   .0082343     9.18   0.000     .0594772    .0917549
       ldonp |   -.385874   .0087099   -44.30   0.000     -.402945    -.368803
       _cons |  -1.763376    .097171   -18.15   0.000    -1.953828   -1.572925
-------------+----------------------------------------------------------------
     /athrho |   .0936672   .0246531     3.80   0.000     .0453481    .1419863
    /lnsigma |  -.4400404    .006844   -64.30   0.000    -.4534544   -.4266264
-------------+----------------------------------------------------------------
         rho |   .0933942    .024438                       .045317    .1410398
       sigma |   .6440104   .0044076                      .6354294    .6527074
      lambda |   .0601469   .0158231                      .0291342    .0911596
------------------------------------------------------------------------------
LR test of indep. eqns. (rho = 0):   chi2(1) =    13.87   Prob > chi2 = 0.0002
------------------------------------------------------------------------------
```

3.4 Efekty Krańcowe

Tak jak w przypadku modelu dla zmiennej ocenzurowanej, możemy obliczyć różne efekty krańcowe dla zmiennej zależnej.

Możemy obliczyć krańcowy efekt pod warunkiem, że zmienna zależna jest obserwowana

$$\frac{\partial E(y \mid yobserwowane)}{\partial x}$$

```
. mfx compute, predict(ycond)

Marginal effects after heckman
      y  = E(lzarobki|Zg>0) (predict, ycond)
         =  6.2348473
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
    plec*|    .254484      .01311   19.41   0.000   .228788   .28018   .431556
    wiek |   .0794953      .00385   20.63   0.000   .071944  .087046   52.7701
   wiek2 |  -.0010471      .00005  -22.92   0.000  -.001137 -.000958   3062.33
   wwiek |   .0173184      .00191    9.05   0.000   .013568  .021069   3.52186
  wwiek2 |  -.0001221      .00004   -3.23   0.001  -.000196 -.000048   200.282
 rodzina*|   .1180917      .01728    6.83   0.000   .084215  .151968   .691293
    ld14 |  -.0037467      .00108   -3.47   0.001  -.005863 -.001631   .624551
   ldonp |   .0191195      .00504    3.79   0.000   .009245  .028994   6.15523
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

Współczynniki interpretujemy jako zmianą wartości oczekiwanej zmiennej zależnej pod wpływem zmiany zmiennej niezależnej o 1 % pod warunkiem, że zmienna zależna jest obserwowana.

Efekt krańcowy prawdopodobieństwa selekcji

$$\frac{\partial Pr(yobserwowane)}{\partial x}$$

uzyskujemy następująco:

```
. mfx compute, predict(psel)

Marginal effects after heckman
      y  = Pr(select) (predict, psel)
         =  .11312427
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
    plec*|   .0835625      .00358   23.36   0.000   .076552  .090573   .431556
    wiek |   .0368492      .00065   56.56   0.000   .035572  .038126   52.7701
   wiek2 |  -.0004766      .00001  -69.61   0.000   -.00049 -.000463   3062.33
   wwiek |   .0084232      .00063   13.31   0.000   .007183  .009664   3.52186
  wwiek2 |  -.0000771      .00001   -7.15   0.000  -.000098 -.000056   200.282
```

```
 rodzina*|   .0262026      .00369    7.10   0.000   .018974  .033431   .691293
    ld14 |   .0145063      .00164    8.86   0.000   .011298  .017714   .624551
   ldonp |  -.0740268      .00217  -34.13   0.000  -.078277 -.069776   6.15523
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

Współczynniki interpretujemy jako zmianą prawdopodobieństwa doboru zmiennej do próby do wpływem zmiany wartości zmiennej niezależnej o 1 %.

Ostatnim efektem jest wartość krańcowa dla prawdopodobieństwa wartości oczekiwanej zmiennej zależnej.

$$\frac{\partial E(y^* \mid z = 1)}{\partial x}$$

```
. mfx compute, predict(yexpected)

Marginal effects after heckman
      y  = E(lzarobki*|Pr(select)) (predict, yexpected)
         =  .70531254
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
    plec*|   .5539878       .0228   24.29   0.000   .509292  .598684   .431556
    wiek |   .2387417      .00413   57.79   0.000   .230644  .246839   52.7701
   wiek2 |  -.0030899      .00004  -70.70   0.000  -.003176 -.003004   3062.33
   wwiek |   .0544767      .00397   13.73   0.000     .0467  .062253   3.52186
  wwiek2 |  -.0004942      .00007   -7.33   0.000  -.000626 -.000362   200.282
 rodzina*|   .1756028      .02292    7.66   0.000   .130675  .220531   .691293
    ld14 |   .0900208      .01015    8.87   0.000   .070121   .10992   .624551
   ldonp |  -.4593827      .01353  -33.96   0.000  -.485895  -.43287   6.15523
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

Współczynniki są efektami krańcowymi dla oczekiwanego dochodu z pracy. Więc jest to sumaryczny efekt wzrost prawdopodobieństwa zatrudnienia i wzrostu oczekiwanej płacy pod wpływem zmiany zmiennej niezależnej.

4 Model efektów oddziaływania (wyłącznie dla IiE)

Model Heckmana ma zastosowanie gdy zmienna zależna jest ocenzurowana. Jednak czasem zachodzi konieczność zbadania oddziaływania jednego czynnika na inny, np. wpływu szkolenia na płace pracowników. Jeżeli jeden z tych czynników jest endogeniczny względem drugiego, np. jeżeli badamy wysokość zarobków w zależności od faktu czy pracownik należy do związku zawodowego to musimy rozpatrzyć dwa przypadki. Jeżeli wierzymy, że przynależność do związku jest sprawą czysto losową wówczas możemy zastosować zwykły model regresji, jednak gdy ufamy że pracownik sam decyduje o przynależności do związku, wówczas bycie w związku jest czynnikiem endogenicznym względem płacy. W takim przypadku używa się modelu efektów oddziaływania. Jest on matematycznie równoważny modelowi selekcji próby, ale nie wymaga by zmienna zależna była ocenzurowana. Oczywiście istnieje możliwość sztucznego ocenzurowania zmiennej, jednak prowadzi to do utraty informacji.

Zbadamy czy zarobki pracowników z wykształceniem wyższym znacząco różnią się od pensji pozostałych pracowników. Oszacujemy następujący model płacy

$$ln(zarobki) = f(plec, wiek, wiek2, rodzina)$$

Model zastał oszacowany na podstawie całej próby, ale równanie płacy wyłącznie dla osób o wykształceniu wyższym. Pierwszy model oszacowana metodą dwustopniową.

```
Treatment effects model -- two-step estimates   Number of obs      =     11082

                                                Wald chi2(9)       =   1119.09
                                                Prob > chi2        =    0.0000

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
lzarobki     |
        plec |   .3223371   .0166584    19.35   0.000     .2896872    .3549871
        wiek |   .0941707   .0046847    20.10   0.000     .0849889    .1033525
       wiek2 |  -.0012375   .0000564   -21.93   0.000    -.0013481   -.001127
     rodzina |   .1134364   .0199527     5.69   0.000     .0743297    .1525431
      wyzsze |   1.660543   .1414317    11.74   0.000     1.383342    1.937744
       _cons |   4.576325   .1026646    44.58   0.000     4.375106    4.777544
-------------+----------------------------------------------------------------
wyzsze       |
        plec |  -.2617793   .0329088    -7.95   0.000    -.3262792   -.1972793
        wiek |  -.0404883   .0090814    -4.46   0.000    -.0582875   -.0226891
       wiek2 |   .0005921   .0001032     5.74   0.000     .0003899    .0007942
    miastodu |   .5173512   .0418904    12.35   0.000     .4352476    .5994548
     rodzina |    .041806   .0421377     0.99   0.321    -.0407824    .1243944
       _cons |  -.5390671   .1939149    -2.78   0.005    -.9191333   -.1590008
-------------+----------------------------------------------------------------
hazard       |
      lambda |  -.6154013   .0756754    -8.13   0.000    -.7637222   -.4670803
-------------+----------------------------------------------------------------
         rho |   -0.82834
       sigma |  .74293573
      lambda | -.61540126   .0756754
------------------------------------------------------------------------------
```

Współczynnik `lambda` jest statystycznie istotny. Ponadto `rho` jest równe -1, co jest dowodem na występowanie dokładnej linowej zależności między składnikami resztowymi obu równań.

```
. reg lzarobki plec wiek wiek2 rodzina if wyzsze==1

      Source |       SS       df       MS              Number of obs =    1396
-------------+------------------------------           F(  4,  1391) =   55.85
       Model |  102.373682     4  25.5934206           Prob > F      =  0.0000
    Residual |   637.46997  1391   .45828179           R-squared     =  0.1384
```

```
-------------+------------------------------           Adj R-squared =  0.1359
       Total |  739.843653  1395  .530353873           Root MSE      =  .67697

------------------------------------------------------------------------------
    lzarobki |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        plec |    .341363   .0391103     8.73   0.000     .2646415    .4180844
        wiek |   .1004311    .010552     9.52   0.000     .0797314    .1211307
       wiek2 |  -.0012122   .0001155   -10.49   0.000    -.0014388   -.0009855
     rodzina |    .126696   .0467979     2.71   0.007     .0348939    .2184981
       _cons |   4.924052   .2339208    21.05   0.000     4.465177    5.382928
------------------------------------------------------------------------------
```

Jeżeli porównamy wyniki z modelem regresji dla osób o wykształceniu wyższym, to łatwo zauważymy że zignorowanie problemu endogeniczności spowoduje niedoszacowanie lub przeszacowanie efektów wyższego wykształcenia. Różnice w uzyskanych oszacowaniach są mimo wszystko nieznaczne.

Tak jak poprzednio estymatory metody dwustopniowej nie są efektywne. Lepsze rezultaty daje użycie metody największej wiarogodności.

```
. treatreg lzarobki plec wiek wiek2 rodzina,treat(wyzsze=plec wiek
wiek2 miastodu rodzina)

Iteration 0:   log likelihood = -15765.752
Iteration 1:   log likelihood = -14963.736
Iteration 2:   log likelihood = -14858.743
Iteration 3:   log likelihood = -14852.415
Iteration 4:   log likelihood = -14852.309
Iteration 5:   log likelihood = -14852.309

Treatment effects model -- MLE                  Number of obs      =     11082

                                                Wald chi2(5)       =   1354.83
Log likelihood = -14852.309                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
lzarobki     |
        plec |    .273855   .0132242    20.71   0.000     .2479359     .299774
        wiek |    .082542   .0038194    21.61   0.000      .075056    .0900279
       wiek2 |   -.001071   .0000447   -23.98   0.000    -.0011585   -.0009834
     rodzina |   .1133073   .0173585     6.53   0.000     .0792853    .1473292
      wyzsze |   .7275764   .0460687    15.79   0.000     .6372833    .8178694
       _cons |   4.898903   .0801325    61.14   0.000     4.741846     5.05596
-------------+----------------------------------------------------------------
wyzsze       |
        plec |  -.2585516   .0328213    -7.88   0.000    -.3228801    -.194223
        wiek |  -.0393215   .0090747    -4.33   0.000    -.0571077   -.0215353
       wiek2 |     .00058   .0001031     5.63   0.000     .0003779    .0007821
    miastodu |   .5575744   .0421984    13.21   0.000     .4748671    .6402817
     rodzina |   .0471117   .0421107     1.12   0.263    -.0354237    .1296472
```

```
       _cons |  -.5765116   .1936897    -2.98   0.003    -.9561365   -.1968867
-------------+----------------------------------------------------------------
     /athrho |  -.1693054   .0361993    -4.68   0.000    -.2402547   -.0983561
    /lnsigma |  -.4364315   .0071235   -61.27   0.000    -.4503933   -.4224697
-------------+----------------------------------------------------------------
         rho |  -.1677061   .0351812                     -.2357363   -.0980402
       sigma |   .6463388   .0046042                      .6373774    .6554261
      lambda |  -.1083949   .0230085                     -.1534907   -.0632992
------------------------------------------------------------------------------
LR test of indep. eqns. (rho = 0):   chi2(1) =    16.65   Prob > chi2 = 0.0000
------------------------------------------------------------------------------
```

Jak widać z testu ilorazu wiarogodności odrzucamy hipotezę o niezależności równań. Zmienne w równaniu płacy są łącznie istotne. Jak widać zmiana metody oszacowania modelu dość znacząco zmieniła wielkości oszacowań.