The University of British Columbia
Department of Economics

Economics 560: Labour Economics

Professor Nicole M. Fortin Winter 2006

Empirical Assignment #3

Due Date: April 18th, before 12:00noon

The purpose of this assignment is to give you practical knowledge of the Oaxaca-Blinder decomposition methodology for means and of the DFL decomposition methodology for wage distributions. As always, feel free to work cooperatively and in groups. However, each student must hand in his/her own problem set using his/her own interpretation of the results. You are encouraged to use STATA which is available on the computers on the 11th floor. Do not hand in your computer output. You can paste parts of your log files into your assignment but you should sift through your output to include only the relevant parts. The data (and programs where applicable) for the assignment are available in the assignment section of the course web site. The estimations in both papers below were performed using the SAS software, so that you may not be able to get the exact same numbers. Note that the Stata commands below use `;` as delimiter, you need to begin your program with the command

```
#delimit ;  */the delimit command replace the cr as end of line*/
```

A package to do the usual oaxaca decomposition can be install from the STATA web site. Open Stata, Click Help, click Search, click All, type oaxaca8 in the box, click OK and follow the instructions to install.

1. In this exercise, you are asked to reproduce some of the results of the paper *O'Neill, J. and D. O'Neill, "What Do Wage Differentials Tell Us about Labor Market Discrimination?" NBER Working Paper 11240 (April 2005). (available on the NBER web site.)*
   The NLSY data used by the O'Neills can be downloaded from on the course web site as **nlsy00.dta.** Refer to the article for explanations of the variables.
   First rescale the AFTQ scores

```
replace afqtp89=afqtp89/100.0;
```

   a) Begin by running the OLS regressions to measure the log wage differentials as partial regression coefficients of the dummies variables for *black* and *hispanic* in samples of men. Your goal is to reproduce the 2 columns "Total" in Table 3 and in Table 6 (no need to reproduce the other columns). Note that `sch10_12` is the omitted category for education. For example the regression corresponding to row 4) column 1 of Table 3 is:

```
reg lropc00 black age00 msa ctrlcity north_central south00 west sch_10
    diploma_hs ged_hs smcol bachelor_col master_col doctor_col afqtp89
    wkswk_18 yrsmil78_00 if female==0 & hispanic==0;
```

and the regression corresponding to row 5) column 1 of Table 6 is:

```
reg lropc00 black age00 msa ctrlcity north_central south00 west sch_10
      diploma_hs ged_hs smcol bachelor_col master_col doctor_col
      afqtp89 age1stb30 wkswk_18 yrsmil78_00 famrspb pcntpt_22
      if female==1 & hispanic==0;
```

Begin to gather your results in 4 columns in one table and include the standard errors in a row underneath the point estimates. Comment on the different role (by gender) that AFTQ scores plays in accounting for racial and ethnic wage differentials.

b) Next you are asked to reproduce the decomposition results of Table 5 for the model M1. Begin by obtaining means of the explanatory variables for the 3 racial/ethnic groups,
for example for whites, run the command

```
sum age00 msa ctrlcity north_central south00 west sch_10  sch10_12 diploma_hs
      ged_hs smcol bachelor_col master_col doctor_col afqtp89
      if female==0 & white==1;
```

Can you predict which variables are likely to have more explanatory which respect to racial/ethnic wage differentials?

c) Use the following commands to perform an Oaxaca-Blinder decomposition (after being sure that the `oaxaca` package has been installed on your computer) of the white-black wage differential among men. Use in turn the white coefficients, the black coefficients and a weigthed (wgt=0.651) average of the two.

```
reg lropc00 age00 msa ctrlcity north_central south00 west
      sch_10 diploma_hs ged_hs smcol bachelor_col master_col
      doctor_col  afqtp89
       if female==0 & white==1;
estimates store white;
reg lropc00 age00 msa ctrlcity north_central south00 west
      sch_10 diploma_hs ged_hs smcol bachelor_col master_col
      doctor_col  afqtp89
      if female==0 & black==1;
estimates store black;
oaxaca8 white black, weight(1 0 0.651) detail notf;
```

Does the unexplained part in this last weighted average decomposition correspond to the coefficient of the black dummy estimated in a) row 3)? Would you conclude that there remains no significant explained racial wage gap once racial differences in schooling and AFTQ scores have been controlled for? Are your predictions in b) confirmed? Explain. What portion of the racial wage gap is accounted for by racial differences in AFTQ scores?

d) Now perform the corresponding estimation for the white-Hispanic wage differential, which weight do you have to use in the third decomposition for it to correspond to your result in a). Explain what the purpose of the weighted decomposition is.

e) One problem with the Oaxaca-Blinder decomposition is that it is not-invariant to the choice of omitted category in the case of categorical variables. Redo the black-white decomposition above but this time, choose `south00` as the omitted region, you will need to construct the dummy variable `east`

```
gen east=0;
replace east=1 if west==0 & north_central==0 & south00==0;
```

Does anything change? Explain what is going on.

2. In this exercise, you are asked to reproduce some of the results of the paper *DiNardo, J., N. Fortin, and T. Lemieux, "Labor Market Institutions and the Distribution of Wages,1973-1992: A Semiparametric Approach," Econometrica, Vol. 64 (Sept. 1996): 1001-44. (available on JSTOR)*

The CPS data for 1979 and 1988 for male workers used in DFL can be downloaded from on the course web site as **men7988_cell.dta.** Refer to the article for explanations of the variables.

a) Begin by illustrating the density of male log wages in 1979 and 1988, that is reproduce Figure 4a.

```
set mem 50m;
use men7988_cell;
replace wage=wage*65.8/104.3 if year88==1; /*transform to 1979 dollars*/
gen lwage=log(wage);
*to simplify the kernel density estimation, estimated it only at 200 values;
sum lwage, detail;
gen xstep=(r(max)-r(min))/200;
*kwage will be the wage at which the density is estimated;
gen kwage=r(min)+(_n-1)*xstep if _n<=200;
gen hweight=eweight*uhrswk;                  /*hours weighted*/
kdensity lwage [aweight=hweight] if year88==1 , at(kwage) gauss width(0.065)
     generate(w88 fd88) nograph ;
kdensity lwage [aweight=hweight] if year88==0 , at(kwage) gauss width(0.065)
     generate(w79 fd79) nograph ;
label var fd88 "Men 1988";
label var fd79 "Men 1979";
label var kwage "Log(Wage)";
graph twoway (connected fd88 kwage if kwage>=0 & kwage<=3.91, msymbol(i)
      clwidth(medium) ) (connected fd79 kwage if kwage>=0 & kwage<=3.91,
      msymbol(i) lpattern(dash) clwidth(medium) ),
      xlabel(.69 1.61 2.3 3.22) xline(0.748 1.065) scheme(sj)
      saving(dflfig4a,replace);
```

b) Now perform the simulation of the change in individual attributes described in the second part of section 3.2 (p.1014). First estimate a probit for the year effect

```
***probit for year effect;
probit year88 ee1-ee15 exper exper2 exper3 exper4 edex educ reg1-reg3
          ind1-ind18 occ1 occ2 nonwhite partt married smsa [pweight=eweight];
predict py88, p;
summ year88 [weight=eweight] ;
gen pbar=r(mean);
gen phix=eweight*((1-py88)/py88)*(pbar/(1-pbar)) if year88==1;
replace phix=phix*hweight;  */hours weight*/
```

Then obtain the corresponding kernel density estimate

```
kdensity lwage [aweight=phix] if year88==1 , at(kwage) gauss width(0.065)
      generate(w79 fd88x79) nograph ;
```

and plot it as in the same graph as the original estimate `fd88` obtained in a)

c) Compute the standard deviation of log wages along with the 90-10, 90-50 and 50-10 wage differentials comparing for the 1988, 1979 and 1988 as if individual attributes were as in 1979,
for example for 1988

```
 ***Standard deviation of log wages;
gen m1=w88*fd88;
integ m1 w88, generate(mint);
summ mint;
gen mrw=_result(6);
di mrw;
gen v1=((w88-mrw)^2)*fd88;
integ v1 w88, generate(vint);
summ vint;
gen var=_result(6);
di var;
gen std=sqrt(var);
di std;
***wage differentials;
integ fd88 w88, generate(cint);
gen cent10=w88 if cint>.1 & cint[_n-1]<.1 & cint~=.;
sum cent10;
gen d10=r(mean);
gen cent90=w88 if cint>.9 & cint[_n-1]<.9 & cint~=.;
sum cent90;
gen d90=r(mean);
gen cent50=w88 if cint>.5 & cint[_n-1]<.5 & cint~=.;
sum cent50;
gen d50=r(mean);
gen d9010=d90-d10;
gen d9050=d90-d50;
gen d5010=d50-d10;
di d9010;
di d9050;
di d5010;
```

Put your numbers in 3 columns, first compare them to those reported in Table A1, and then compute the relevant changes. How much of the 1988-79 difference is explained by the counterfactual experiment performed in b). Compare your results to those of Table III and V. Explain the differences taking into account the fact that the order of the decomposition matters.