```
********** OUTPUT FILE: CPS DATA**********

 * STATA 10.0 CODE
 * copyright C 2008 by Tito Boeri & Jan van Ours
 * "LABOR MARKETS, POLICIES, AND INSTITUTIONS"
 * by Tito Boeri & Jan van Ours (2008)
 * Princeton University Press

 *Chapter 3 Unions and Collective Bargaining
 *BOX 3.4 Measuring and Decomposing the Effects of Unions on the Wage
Distribution, pages 67-69


 *VARIABLE DESCRIPTION

*tip       0=no tips, 1=tipped
*eweight   earnings wt for all
*uhrswk    usual hrs last week
*paidhr    hrly rated 1=y, 2=n
*topcode   weekly earnings topcoded
*wage      hourly wage rate
*waget     wage rate including tips
*detopcod  replace 999 by unedited earnings
*nind      21 industries classification
*nocc      3 occupations classification
*married   married spouse present
*smsa      lives in smsa
*partt     part-time worker
*class     class of worker
*region    four regions classification
*state
*nonwhite  0=white, 1=non-white
*educ      years of education
*exper     potential experience (a-e-5)
*month     calendar month
*union     member of a union
*covered   covered by union contract
*female    1 if female



/* Analyze the difference between the wage distribution of union and
non-union workers
using the Oaxaca and DiNardo-Fortin-Lemieux decomposition */
log using "C:\\Libro\Di Nardo Fortin Lemieux\DFLoaxaca.log", replace
clear all
set mem 500m
*Open the file March CPS 1988

 use "C:\dfl.dta", clear

* dataset description
des



* Generate the logarithm of wage and experience^2
gen lwage = ln(wage)
gen exper2 = exper^2



* Oaxaca decomposition
* wage regression for unionized workers
```

```
reg lwage exper exper2 educ nonwhite  partt smsa nocc nind  female
married if union==1

*store gammaU (the coefficients of the regression)
matrix gammaU=e(b)
matrix list gammaU

*store the average values of the regressors

gen cons=1
sum cons if union==1
scalar cons_mu=r(mean)
sum exper if union==1
scalar exper_mu=r(mean)
sum exper2 if union==1
scalar exper2_mu=r(mean)
sum educ if union==1
scalar educ_mu=r(mean)
sum nonwhite if union==1
scalar nonwhite_mu=r(mean)
sum partt if union==1
scalar partt_mu=r(mean)
sum smsa if union==1
scalar smsa_mu=r(mean)
sum nocc if union==1
scalar nocc_mu=r(mean)
sum nind if union==1
scalar nind_mu=r(mean)
sum female if union==1
scalar female_mu=r(mean)
sum married if union==1
scalar married_mu=r(mean)


matrix XU=( exper_mu, exper2_mu, educ_mu, nonwhite_mu, partt_mu,
smsa_mu, nocc_mu, nind_mu, female_mu,married_mu, cons_mu)
matrix list XU

* wage regression for non unionized workers
reg lwage exper exper2 educ nonwhite  partt smsa nocc nind  female
married if union==0

*store gammaN (the coefficients of the regression)

matrix gammaN=e(b)
matrix list gammaN

*store the average values of the regressors


sum cons if union==0
scalar cons_mn=r(mean)
sum exper if union==0
scalar exper_mn=r(mean)
sum exper2 if union==0
scalar exper2_mn=r(mean)
sum educ if union==0
scalar educ_mn=r(mean)
sum nonwhite if union==0
scalar nonwhite_mn=r(mean)
sum partt if union==0
scalar partt_mn=r(mean)
sum smsa if union==0
scalar smsa_mn=r(mean)
```

```
sum nocc if union==0
scalar nocc_mn=r(mean)
sum nind if union==0
scalar nind_mn=r(mean)
sum female if union==0
scalar female_mn=r(mean)
sum married if union==0
scalar married_mn=r(mean)
matrix XN=( exper_mn, exper2_mn, educ_mn, nonwhite_mn, partt_mn,
smsa_mn, nocc_mn, nind_mn, female_mn, married_mn, cons_mn)
matrix list XN

*group weight is equal to 1/2

*"explained" component
matrix EXP=(XU-XN)*0.5*(gammaU+gammaN)'

matrix list EXP

*"unexplained" component

matrix UNEXP=(gammaU-gammaN)*0.5*(XU+XN)'

matrix list UNEXP

*wage differential

matrix DIFF=EXP+UNEXP

matrix list DIFF




*Di Nardo-Fortin-Lemieux decomposition

*create the propensity score as in DFL (page 1012)

*run the probit
probit union exper exper2 educ nonwhite  partt smsa nocc nind
predict double phat
gen double wgt = (1-phat)/phat


* create 200 equidistant points between the minimum and the maximum
of lnwage

scalar  nbins = 200
summarize lwage
generate x = r(min) if _n==1
replace x = (r(max) -r(min))/(nbins-1) if _n>1
replace x = sum(x)
replace x = . if _n > nbins



* Estimate the kdensity

kdensity lwage if union== 1 , generate(unfactual) at(x) nograph
kdensity lwage if union== 0 , generate(factual) at(x) nograph
```

```
kdensity lwage if union == 1 [aw = wgt], generate(cfactual) at(x)
nograph

generate diff = unfactual - cfactual

* Label the variables
label variable x "logarithm of wage"
label variable cfactual "weighted union=0"
label variable factual "union=0"

label variable unfactual "union=1"
label variable diff "difference in densities"

*Graphs

line factual cfactual x
more
 line unfactual cfactual x
more
 line diff x
more


log close
```