University of California, Berkeley | Prof. Ken Chay
Department of Economics | Fall Semester, 2005

ECON 142

SKETCH OF SOLUTIONS FOR APPLIED EXERCISE #1

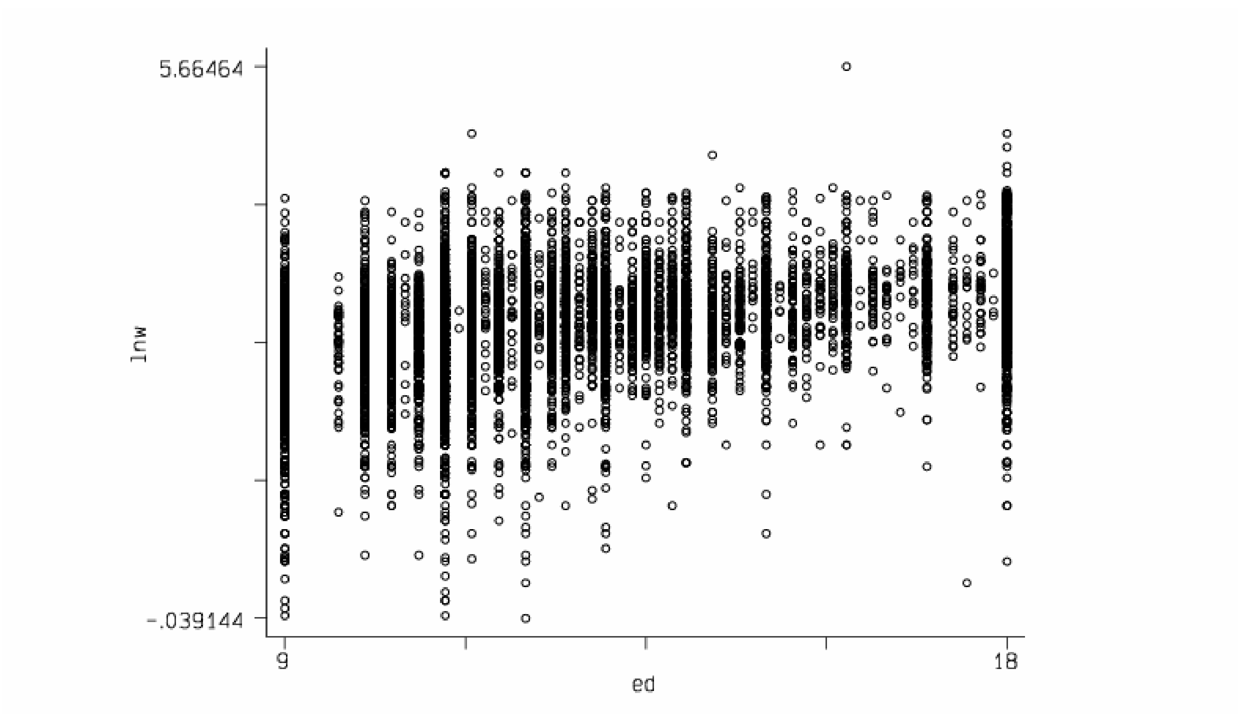
Question 1:

a.

Below is the scatter plot of log-wages on the y-axes and education on the x-axis. It is from the following STATA commands:

```
. set mem 50m
. use restricted92.dta
. graph lnw ed, saving(ps1scat, replace)
```

From the plot, it appears that log-wages has a positive covariance and correlation coefficient with educational attainment as a `best fit line' would likely be upward sloping. The resolution of the scatter plot makes it difficult to tell how many observations fall in the dark areas. Still, if your were to eyeball the sample average of lnw at each value of ed, these points may approximate a straight line. This would be consistent with a linear relationship.



b.

Define $y_i = Y_i - \overline{Y}$ and $x_i = X_i - \overline{X}$, where $\overline{Y}$ and $\overline{X}$ are the sample means of Y and X, then the formulae for the sample covariance, variance, and correlation coefficient of X and Y are:

sample var(Y) = $\frac{1}{n-1}\sum_{i=1}^{n} y_i^2$, sample var(X) = $\frac{1}{n-1}\sum_{i=1}^{n} x_i^2$

sample cov(X,Y) = $\frac{1}{n-1}\sum_{i=1}^{n} x_i y_i$

$$\text{sample corr}(X,Y) = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)}\sqrt{\text{var}(Y)}} = \frac{\text{cov}(X,Y)}{s_X s_Y}$$

The STATA command for calculating the sample covariance of log-wages with education is:

```
. corr lnw ed, covariance
(obs=20042)

         |      lnw       ed
---------+------------------
     lnw |  .212589
      ed |  .494279  6.19577
```

The diagonals give the variances and the off-diagonal gives the covariance. Covariances cannot be directly compared since covariances depend on the units of measurement and are not invariant to scale. Correlation coefficients, on the other hand, normalize the covariance in terms of standard deviation units and can therefore be compared. The appropriate STATA command is:

```
. corr lnw ed
(obs=20042)

         |      lnw       ed
---------+------------------
     lnw |   1.0000
      ed |   0.4307   1.0000
```

One example of how the correlation between education and earnings may not be causal is that individuals with higher education may also have a greater abundance of productive attributes (e.g., family background, "ability", and quality of schooling) that are positively associated with earnings potential. This possibility arises from the fact that education is a choice variable, and, according to economic theory, individuals will choose the education level that maximizes their expected lifetime earnings/utility subject to their resource constraints. In other words, education is not randomly assigned across the population, so correlation does not imply causation.

c.

The bivariate normal distribution is probably not a sensible functional form for the relationship between log-wages and education. One reason is that the bivariate normal is a continuous distribution and implies that the marginal distributions of both log-wages and education have continuous normal distributions. However, education can only take on a discrete number of values ranging from 8-18 years which implies that it does not have a continuous distribution.

If log-wages (y) and education (x) had a bivariate normal distribution, the probability density function (pdf) would be fully described by the population means of log-wages and education (μ_x, μ_y), their standard deviations (σ_x, σ_y), and the population correlation coefficient (ρ).

The population mean of log-wages conditional on education is:
$E(y|x) = \mu_{y|x} = \alpha + \beta x$, where $\alpha = \mu_y - \beta\mu_x$ and $\beta = \rho(\sigma_y/\sigma_x)$

We already have calculated the sample covariance and variance. The appropriate STATA command to get the sample means is:

```
. summ lnw ed

Variable |      Obs        Mean    Std. Dev.       Min        Max
---------+-----------------------------------------------------
```

```
      lnw |     20042     2.944305    .4610736  -.0391438    5.664639
       ed |     20042     12.17011     2.48913          9          18
```

Substituting is these sample moments, the estimated constant and slope coefficient of the conditional mean of log-wages are 1.973 and 0.080, respectively.

Under the assumption of bivariate normality, the conditional variance of log-wages is independent of education (i.e., does not vary with education). The technical term for this is homoskedasticity.

Question 2:

a.

The bivariate linear regression model of log-wages on education will result in an unbiased estimate of the causal effect of education on earnings if education is randomly assigned across the population; earnings does not "cause" education (reverse causality); education is correctly measured; and the effect of education on log-wages is linear in education. Another way to state the first two conditions is that individual schooling levels are unrelated (uncorrelated) to the unobserved/unmeasured determinants of individual earnings (a.k.a., the residuals).

The assumptions under which the least squares estimator of the bivariate model will be the best linear unbiased estimator (BLUE) are:

1. $y_i = \alpha + \beta x_i + \varepsilon_i$, for all i
2. $E(\varepsilon_i)=0$, for all i
3. $var(\varepsilon_i)=\sigma^2$, for all i
4. $cov(\varepsilon_i, \varepsilon_j)=0$, for all i, j

The last 3 assumptions are often written as $\varepsilon_i \sim$ i.i.d. $(0, \sigma^2)$ -- that is, the residuals are independently and identically distributed with zero-mean and variance σ^2. Under these assumptions, the proof of the Gauss-Markov theorem shows that the least squares estimator is the minimum variance (efficient) linear unbiased estimator of the parameters of the conditional mean of y. Condition 1 shows that the linear regression model presumes that the regression function $(\alpha + \beta x_i)$ is linear in education.

b.

The least squares estimator for the constant and slope coefficient comes from minimizing the residual sum of squares:

$(a, b) = \text{argmin} \sum e_i^2 = \text{argmin} \sum (y_i - \alpha - \beta x_i)^2$

Differentiating the above with respect to α and β and setting those equations equal to zero leads to the "first-order conditions" for the least squares estimators (a.k.a., the normal equations):

$$\frac{\partial RSS}{\partial \alpha} = -2\sum (y_i - \alpha - \beta x_i) = -2\sum e_i = 0 \text{, and}$$

$$\frac{\partial RSS}{\partial \beta} = -2\sum x_i (y_i - \alpha - \beta x_i) = -2\sum x_i e_i = 0$$

Solving these two equations with two unknowns leads to the least squares estimators of the constant and slope coefficient. The intuition of these two equations is that the estimators of the constant and slope are chosen such that sum of the estimated residuals is zero and the covariance of the regressor and the estimated residuals is zero. These "moment" conditions (a.k.a., orthogonality conditions) are implied directly by assumptions 1-4.

Three properties of the least squares line are 1) it minimizes the residuals sum of squares; 2) it passes through the point defined by the sample means of log-wages and education; and 3) the least squares residuals have 0-correlation in the sample with values of the regressor.

c.
The following runs the regression of log-wages on a constant and education in STATA:

```
. reg lnw ed

  Source |       SS       df       MS                  Number of obs =   20042
---------+------------------------------               F(  1, 20040) = 4563.60
   Model |  790.257734     1  790.257734               Prob > F      =  0.0000
Residual |   3470.2351 20040  .173165424               R-squared     =  0.1855
---------+------------------------------               Adj R-squared =  0.1854
   Total |  4260.49283 20041  .212588835               Root MSE      =  .41613

------------------------------------------------------------------------------
     lnw |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
      ed |   .0797769   .0011809     67.554   0.000       .0774622    .0820916
   _cons |   1.973412   .0146695    134.525   0.000       1.944658    2.002165
------------------------------------------------------------------------------
```

The estimates of the constant and slope of the relation between log-wages and education are virtually identical to the ones in 1c. This is to be expected since in the bivariate linear regression model, these two approaches are identical.

R-squared = ESS/TSS = 1 - (RSS/TSS),
where the ESS (model SS), RSS (residual SS), and TSS (total SS) are provided in the STATA output. Consequently, the R-squared of the regression is 0.1854.

In the bivariate model, the correlation coefficient is equal to the square-root of the R-squared. Thus, the correlation of log-wages and education is 0.431.

d.
In the log-wage regression,
$\beta = (d\ln y/dx) = (dy/dx)(1/y)$ = (% change in y)/unit change in x.
Consequently, it measures the percentage effect of one extra year of schooling on wages.

An unbiased estimate (s^2) of the variance of the residuals (σ^2) is equal to RSS/(n-2) while the RMSE is equal to sqrt(RSS/n). So $s^2 = nRMSE^2/(n-k) = 20{,}042*0.416^2/(20{,}042-2) = 0.173$. Since the estimated variance of the slope coefficient on education can be written as:

var(b) = var(residuals)/(education sum of squares) = $\dfrac{s^2}{\sum(x_i - \bar{x})^2}$,

the sample variance of education is:
$(1/n)\sum(x_i - \bar{x})^2 = (1/n)\, s^2/\mathrm{var}(b) = (1/20{,}042)(0.173/0.0012^2) = 5.99$

e.
The 95% critical value for the return to education is $t_{0.025}(n\text{-}2) = 1.96$. Consequently, the 95% confidence interval for the return to education is 0.080 ± 1.96 s.e.(b) = [0.077, 0.082]. As a result, the t-ratio testing the significance of the education coefficient (H_0: $\beta=0$):
t = b/s.e.(b) = 67.55
clearly rejects the null hypothesis that the return to education is zero. The t-statistic for testing H_0: β = 0.05 is $(b-b_0)$/s.e.(b) = (.08-.05)/.001 = 30. So this hypothesis is rejected at the 5% significance level as well. The p-value of the test of significance is virtually 0, which implies that we would reject H_0: $\beta=0$ at most levels of significance.

Formula for the F-statistic testing the significance of education is:

$$F = \frac{ESS/1}{RSS/(n-2)} = \frac{r^2(n-2)}{(1-r^2)} \sim F(1, n-2).$$

It follows that the F=4563~F(1, 20040) rejects H_0: β=0 at conventional levels of significance. In the bivariate case, the F-statistic is equal to the square of the t-ratio with (n-2) degrees of freedom. Using the above formula, this implies that the t-ratio for the education coefficient based on the R-squares of the log-wage regression is 68.

Question 3:

a.

The results from running the multivariate log-wage regression model in STATA are:

```
. reg lnw ed exp exp2 female mar computer

  Source |       SS       df       MS                  Number of obs =   20042
---------+------------------------------               F(  6, 20035) = 1669.84
   Model |  1420.31174     6  236.718624               Prob > F      =  0.0000
Residual |  2840.18109 20035  .141760973               R-squared     =  0.3334
---------+------------------------------               Adj R-squared =  0.3332
   Total |  4260.49283 20041  .212588835               Root MSE      =  .37651

------------------------------------------------------------------------------
     lnw |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
      ed |   .0703381    .001182     59.508   0.000       .0680213    .0726549
     exp |    .029843   .0010485     28.463   0.000       .0277879    .0318982
    exp2 |  -.0004571   .0000206    -22.179   0.000      -.0004975   -.0004167
  female |   -.215268   .0056023    -38.425   0.000      -.2262489   -.2042871
     mar |   .0353124   .0063153      5.592   0.000       .0229339     .047691
computer |   .1722029    .005934     29.020   0.000       .1605719     .183834
   _cons |   1.701415   .0188422     90.298   0.000       1.664482    1.738347
------------------------------------------------------------------------------
```

The regression coefficient on education gives the percentage return to an extra year of schooling and is estimated to be about 7 percent. The coefficient on the female indicator gives the log-wage gap between women and men after controlling for gender differences in observable characteristics. The estimate implies that women in this sample earn 0.22 log-points less than men with similar characteristics. The coefficient estimate on the married indicator of .04 implies that married individuals earn about 4% more on average than single people. The coefficient estimate on the computer indicator of .17 implies that individuals who use a computer at work earn about 17% more on average than those that do not.

The coefficients on exp and exp^2 imply that the life-cycle profile of earnings is quadratic and concave. In particular:

$$\frac{\partial lwage}{\partial exp} = 0.02984 + 2*(-0.00046)*exp,$$

which implies that the "return to experience" is positive until about 33 years of experience (set the expression equal to 0 and solve for *exp*). So, in this sample, the maximum wage of individuals over a lifetime appears to occur at 33 years of experience.

Including just a linear term for experience would lead to a misspecified regression model since the above regression implies that exp^2 is an important determinant of earnings, independent of experience (its estimated coefficient has a large t-ratio).

The following STATA command adds $\exp^3$ and $\exp^4$ to the regression:

```
. reg lnw ed exp exp2 exp3 exp4 female mar computer

  Source |       SS       df       MS              Number of obs =   20042
---------+------------------------------           F(  8, 20033) = 1270.60
   Model |  1434.11336     8  179.26417            Prob > F      =  0.0000
Residual |  2826.37947 20033  .141086181           R-squared     =  0.3366
---------+------------------------------           Adj R-squared =  0.3363
   Total |  4260.49283 20041  .212588835           Root MSE      =  .37561

------------------------------------------------------------------------------
     lnw |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
      ed |   .0708511   .0011812     59.984   0.000       .0685359    .0731663
     exp |   .0901068   .0066693     13.511   0.000       .0770345    .1031792
    exp2 |  -.0045452   .0004851     -9.370   0.000       -.005496   -.0035944
    exp3 |   .0001062   .0000139      7.622   0.000       .0000789    .0001335
    exp4 |  -9.32e-07   1.37e-07     -6.809   0.000      -1.20e-06   -6.64e-07
  female |  -.2136361   .0055913    -38.208   0.000      -.2245956   -.2026766
     mar |   .0315186   .0063143      4.992   0.000        .019142    .0438951
computer |   .1727878   .0059202     29.186   0.000       .1611838    .1843919
   _cons |   1.429914   .0337783     42.332   0.000       1.363706    1.496122
------------------------------------------------------------------------------
```

Compared to the model with only exp and $\exp^2$, the model that also includes $\exp^3$ and $\exp^4$ has an adjusted R-squared that is slightly higher (0.336 versus 0.333). So the fit of the regression model is improved. The F-test for whether $\exp^3$ and $\exp^4$ are important determinants of log-wages can be calculated from the RSS (or R-squareds) of the unrestricted and restricted models (imposing the null hypothesis restrictions that the coefficients on $\exp^3$ and $\exp^4$ are zero). Let the subscripts R and UR denote the restricted and unrestricted regressions, respectively. Then the F-test of the joint significance of $\exp^3$ and $\exp^4$ is:

$$F = \frac{(RSS_R - RSS_{UR})/2}{RSS_{UR}/(20042-2)} = \frac{(R^2_{UR} - R^2_R)/2}{(1-R^2_{UR})/(20042-2)} \sim F(2{,}20040)$$

The numerator degrees of freedom are 2 since we are testing whether 2 coefficients are significantly different from zero (2 restrictions implied by the null hypothesis). The denominator degrees of freedom is (n-k) = 20040. The F-statistic is equal to [(0.3366 - 0.3334)/2]/[(1 - 0.3366)/(20,042-2))=48. The 5-percent critical value for this test is about 3, which implies that we can reject the null hypothesis that $\exp^3$ and $\exp^4$ are not significant determinants of log-wages.

b.

The following STATA commands convert the education variable into dummy variables for each whole number value of education:

```
. gen edr=round(ed, 1)
. tab edr, gen(dedr)
```

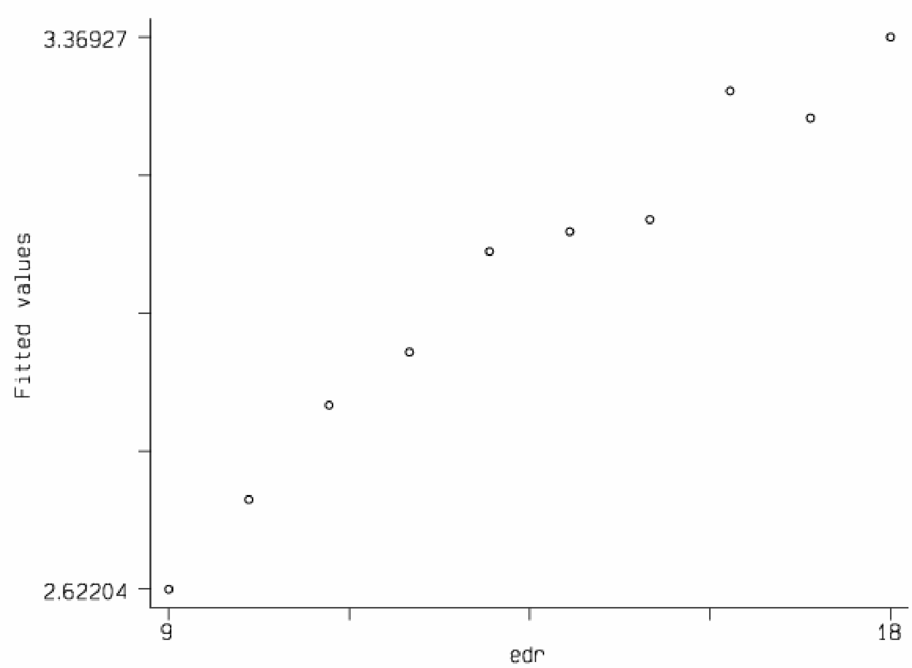
This creates 10 dummy variables called dedr1, dedr2,… dedr10 (representing the values 9, 10,… 18 for edr). When one includes all 10 of these schooling dummies in a regression, STATA drops one of the variables since the 10 schooling dummies sum up to a constant. In other words, the schooling dummies and the constant are perfectly collinear since they are a linear combination of each other. Consequently, there is no unique solution to the least squares regression. The STATA commands used to run the dummy variable regression is:

```
. reg lnw dedr*

  Source |       SS       df       MS              Number of obs =   20042
```

```
---------+------------------------------                F(  9, 20032) =  534.41
   Model |  824.893807     9  91.6548675               Prob > F      =  0.0000
Residual |  3435.59903 20032  .171505542               R-squared     =  0.1936
---------+------------------------------               Adj R-squared =  0.1933
   Total |  4260.49283 20041  .212588835               Root MSE      =  .41413

------------------------------------------------------------------------------
     lnw |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
   dedr1 |  -.6742395   .0234693    -28.729   0.000      -.7202413   -.6282376
   dedr2 |  -.5528825   .0248886    -22.214   0.000      -.6016663   -.5040987
   dedr3 |  -.4251973   .0223674    -19.010   0.000      -.4690393   -.3813553
   dedr4 |  -.3533793   .0228631    -15.456   0.000      -.3981928   -.3085659
   dedr5 |  -.2169873   .0243418     -8.914   0.000      -.2646992   -.1692754
   dedr6 |  -.1903268   .0255509     -7.449   0.000      -.2404087   -.1402448
   dedr7 |  -.1737955   .0276523     -6.285   0.000      -.2279962   -.1195947
   dedr8 |  (dropped)
   dedr9 |  -.0364896   .0324037     -1.126   0.260      -.1000036    .0270243
  dedr10 |   .0729911   .0237514      3.073   0.002       .0264363    .1195459
   _cons |   3.296283   .0218267    151.021   0.000       3.253501    3.339065
------------------------------------------------------------------------------
```

If log-wages and education are not linearly related we would expect the dummy variable model to fit the data better than the linear model. Notice that the adjusted R-squared for the dummy variable model of .1933 is higher than the adjusted R-squared of .1854 from the comparable linear model estimated in 2c. This supports that log-wages is not a linear function of education in the population. Another way to judge the linearity is to plot the coefficient estimates for the dummy variables as a function of the edr. If the relationship is linear then interpolating these points should approximate a line as the sample size gets large. The same could be said for a plot of the fitted values (recall the fitted values are equal to the coefficient estimates plus the intercept in this case.) The STATA commands used to produce this graph are:

. predict lnwhat
. sort edr
. graph lnwhat edr if edr[_n]~=edr[_n-1], saving(dedr, replace)



The graph suggests that the relationship maybe linear for values of education from 9-13, but there appears to be non-linearities for higher values of education.

To test whether education is a determinant of log-wages using the dummy variable model, we test the joint null hypothesis that all the coefficients are equal to zero. As the restricted model consists only of an intercept, the F-statistic simplifies to:

$$\frac{(TSS - RSS_{UR})/2}{RSS_{UR}/(20042-2)} = \frac{R_{UR}^2/2}{(1-R_{UR}^2)/(20042-2)} \sim F(2,20040)$$

As this formula gives a value of 534 for the F-statistic, we can reject the null hypothesis. This calculation can be performed by STATA through using the command:

```
. testparm dedr*

 ( 1)  dedr1 = 0.0
 ( 2)  dedr2 = 0.0
 ( 3)  dedr3 = 0.0
 ( 4)  dedr4 = 0.0
 ( 5)  dedr5 = 0.0
 ( 6)  dedr6 = 0.0
 ( 7)  dedr7 = 0.0
 ( 8)  dedr8 = 0.0
 ( 9)  dedr9 = 0.0
 (10)  dedr10 = 0.0
       Constraint 8 dropped

       F(  9, 20032) =  534.41
            Prob > F =    0.0000
```

The STATA command for running the dummy variable regression for log-wages including exp, exp^2, gender, marital status, and computer use controls is:

```
. reg lnw dedr* exp exp2 female mar computer

  Source |       SS       df       MS                  Number of obs =   20042
---------+------------------------------               F( 14, 20027) =  724.47
   Model |  1432.31633    14  102.308309               Prob > F      =  0.0000
Residual |   2828.1765 20027   .14121818               R-squared     =  0.3362
---------+------------------------------               Adj R-squared =  0.3357
   Total |  4260.49283 20041  .212588835               Root MSE      =  .37579

------------------------------------------------------------------------------
     lnw |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
   dedr1 |  -.5460232   .0219141    -24.916   0.000      -.5889768   -.5030697
   dedr2 |    -.39999   .0228705    -17.489   0.000       -.444818    -.355162
   dedr3 |  -.3462224   .0205074    -16.883   0.000      -.3864186   -.3060262
   dedr4 |    -.24125   .0208344    -11.579   0.000      -.2820872   -.2004127
   dedr5 |  -.1838428   .0221444     -8.302   0.000      -.2272476   -.1404379
   dedr6 |  -.1251421    .023212     -5.391   0.000      -.1706395   -.0796447
   dedr7 |  -.0746812   .0251501     -2.969   0.003      -.1239775   -.0253848
   dedr8 |  (dropped)
   dedr9 |   .0480832   .0294415      1.633   0.102      -.0096245     .105791
  dedr10 |    .120013   .0215648      5.565   0.000       .0777442    .1622819
     exp |   .0294574   .0010543     27.939   0.000       .0273908     .031524
    exp2 |  -.0004418   .0000208    -21.282   0.000      -.0004825   -.0004011
  female |   -.212833   .0056889    -37.412   0.000      -.2239836   -.2016823
     mar |   .0343005   .0063072      5.438   0.000       .0219379    .0466631
computer |   .1628591   .0060513     26.913   0.000       .1509981    .1747202
   _cons |   2.823842   .0231128    122.177   0.000       2.778539    2.869145
------------------------------------------------------------------------------
```

The resulting adjusted R-squared equals 0.3357, whereas the adjusted R-squared from the specification that only allows education to enter linearly is 0.3332 (from 3a.). Thus, it appears that allowing for nonlinearities still improves the fit of the regression model when we have controlled for other observable determinants of log-wages.

c.

The predicted <u>level</u> of wages for a woman with 12 years of education, 10 year of experience, and who uses a computer at work is:

predicted wage = exp(predicted log-wage) =

$\exp(1.7014 + 0.0703(12) + 0.0298(10) - (0.0005)(10^2) + 0.1722) = \exp(2.965) = 19.40$/hour d.

The STATA command to run the regression including the indicator variables for other work tools is:

```
. reg lnw ed exp exp2 female mar computer pencil telefon calc hammer

  Source |       SS       df       MS              Number of obs =   20042
---------+------------------------------           F( 10, 20031) = 1047.29
   Model |  1462.7508    10   146.27508            Prob > F      =  0.0000
Residual |  2797.74203 20031  .139670612           R-squared     =  0.3433
---------+------------------------------           Adj R-squared =  0.3430
   Total |  4260.49283 20041  .212588835           Root MSE      =  .37373

------------------------------------------------------------------------------
     lnw |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
      ed |   .0647356   .0012234     52.917   0.000       .0623377    .0671334
     exp |    .028837   .0010429     27.650   0.000       .0267928    .0308813
    exp2 |  -.0004385   .0000205    -21.401   0.000      -.0004786   -.0003983
  female |   -.229428   .0058712    -39.077   0.000      -.2409361     -.21792
     mar |   .0334185   .0062706      5.329   0.000       .0211276    .0457094
computer |   .1197219   .0066509     18.001   0.000       .1066856    .1327582
  pencil |   .0310925   .0080538      3.861   0.000       .0153063    .0468787
 telefon |   .0417872   .0080792      5.172   0.000       .0259512    .0576231
    calc |   .0460965   .0069809      6.603   0.000       .0324134    .0597797
  hammer |  -.0353619   .0064239     -5.505   0.000      -.0479533   -.0227704
   _cons |   1.751301   .0198236     88.344   0.000       1.712445    1.790157
------------------------------------------------------------------------------
```

Notice that all the tools have statistically significant coefficients, and that the coefficient for using a hammer is negative. These findings indicate it may be inappropriate to interpret the coefficients as the return to using the respective tool. The ability to use a pencil is not a scarce skill, so we do not expect it to be rewarded in the labor market. Alternatively, we would not expect the return to a skill to be negative, yet the coefficient on hammer is negative. This estimate suggests that the tools may be proxies for the kind of work an individual does, indicating that the kind of work an individual does may be an omitted variable in the above regression. Occupation indicator variables are included to try to control for the type of work.

The STATA command to run the demeaned regression is (recall the demeaned regression gives the same results as including the dummy variables and requires less computing power):

```
. areg lnw ed exp exp2 female mar computer pencil telefon calc hammer,
absorb(occ)

                                                  Number of obs =   20042
                                                  F( 10, 18961) =  291.05
                                                  Prob > F      =  0.0000
                                                  R-squared     =  0.4420
                                                  Adj R-squared =  0.4102
                                                  Root MSE      =   .3541

------------------------------------------------------------------------------
         lnw |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ed |   .0404327   .0017111    23.629   0.000     .0370787    .0437866
         exp |   .0261459   .0010359    25.240   0.000     .0241155    .0281764
        exp2 |  -.0003836   .0000203   -18.916   0.000    -.0004234   -.0003439
      female |  -.1604235   .0075882   -21.141   0.000     -.175297    -.14555
         mar |   .0341433   .0061256     5.574   0.000     .0221367    .0461499
    computer |   .0681671   .0072878     9.354   0.000     .0538823    .0824518
      pencil |   .0074461    .008151     0.914   0.361    -.0085307    .0234229
     telefon |   .0484365   .0083607     5.793   0.000     .0320487    .0648243
        calc |   .0223325   .0070968     3.147   0.002     .0084221     .036243
      hammer |  -.0206175   .0077834    -2.649   0.008    -.0358736   -.0053614
       _cons |   2.081236    .025542    81.483   0.000     2.031171      2.1313
------------------------------------------------------------------------------
         occ |      F(1070,18961) =     3.132   0.000         (1071 categories)
```

Including the occupation indicators led to a large change in the coefficient on computer use. This implies that tool use is probably not randomly distributed across the population, as it is not evenly distributed across occupation. This suggests that the OLS estimator for the return to computers is subject to omitted variables bias when occupation is not controlled for. Even in the regression including the occupation indicator it is unclear that omitted variables bias has been completely eliminated. For example, if occupation does not adequately control for the type of work someone does, then the estimate of 7% may still overstate the return to using a computer.