BOSTON COLLEGE
Department of Economics
EC 228 01 Econometric Methods
Fall 2008, Prof. Baum, Ms. Phillips (tutor), Mr. Dmitriev (grader)
Problem Set 2
Due at classtime, Thursday 2 Oct 2008

2.4

(i)(5 marks) When cigs = 0, predicted birth weight is 119.77 ounces. When cigs = 20, .bwght = 109.49. This is about an 8.6 percent drop.

(ii) (5 marks) Not necessarily. There are many other factors that can affect birth weight, particularly overall health of the mother and quality of prenatal care. These could be correlated with cigarette smoking during birth. Also, something such as caffeine consumption can affect birth weight, and might also be correlated with cigarette smoking.

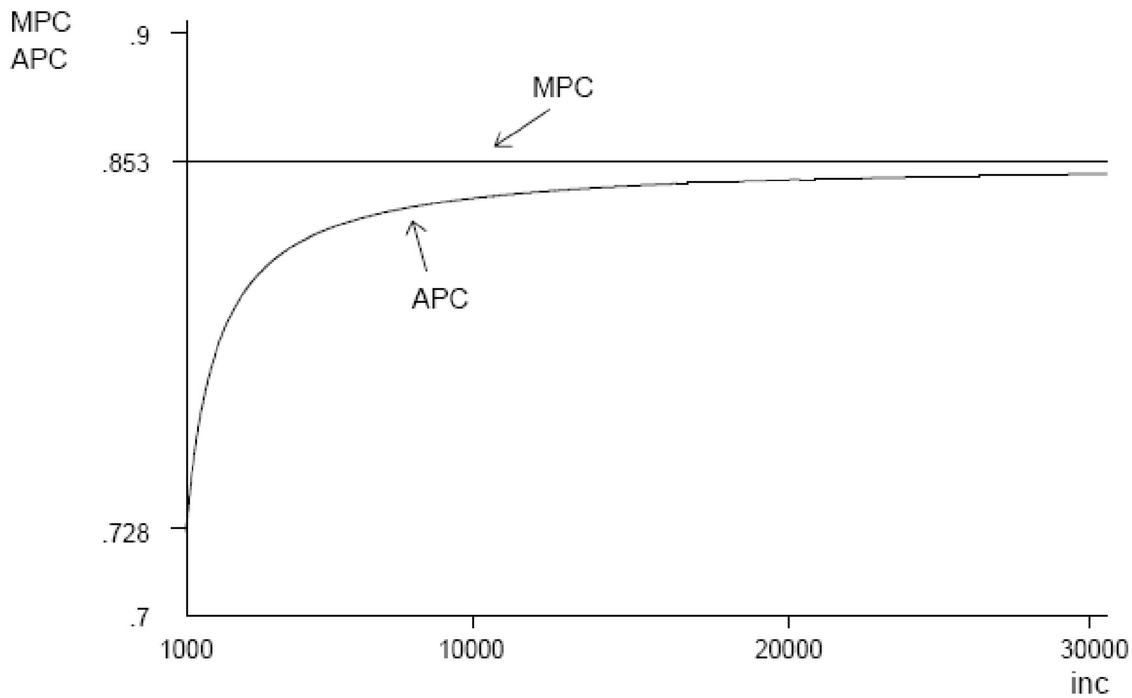
(iii) (10 marks)If we want a predicted *bwght* of 125, then $cigs = (125119.77)/(.524) \approx$ 10.18, or about 10 cigarettes! This is nonsense, of course, and it shows what happens when we are trying to predict something as complicated as birth weight with only a single explanatory variable. The largest predicted birth weight is necessarily 119.77. Yet almost 700 of the births in the sample had a birth weight higher than 119.77.

(iv) (5 marks)1,176 out of 1,388 women did not smoke while pregnant, or about 84.7 percent. Because we are using only cigs to explain birth weight, we have only one predicted birth weight at $cigs = 0$. The predicted birth weight is necessarily roughly in the middle of the observed birth weights at $cigs = 0$, and so we will under predict high birth rates.

2.5

(i) (10 marks)The intercept implies that when inc = 0, cons is predicted to be negative 124.84 dollars. This, of course, cannot be true, and reflects that fact that this consumption function might be a poor predictor of consumption at very low-income levels. On the other hand, on an annual basis, 124.84 dollars is not so far from zero.

(ii) (5 marks) Just plug 30,000 into the equation: = 124.84 + .853(30,000) = 25,465.16 dollars. .cons (iii)(5 marks) The MPC and the APC are shown in the following graph. Even though the intercept is negative, the smallest APC in the sample is positive. The graph starts at an annual income level of 1,000 (in 1970 dollars).



2.6

(i)(5 marks) Yes. If living closer to an incinerator depresses housing prices, then being farther away increases housing prices.

(ii)(5 marks) If the city chose to locate the incinerator in an area away from more expensive neighborhoods, then $log(dist)$ is positively correlated with housing quality. This would violate SLR.4, and OLS estimation is biased.

(iii)(5 marks) Size of the house, number of bathrooms, size of the lot, age of the home, and quality of the neighborhood (including school quality), are just a handful of factors. As mentioned in part (ii), these could certainly be correlated with dist [and $log(dist)$].

3.1

(i)(5 marks) *hsperc* is defined so that the smaller it is, the lower the students standing in high school. Everything else equal, the worse the students standing in high school, the lower is his/her expected college GPA.

(ii) (5 marks)Just plug these values into the equation: $\widehat{colgpa} = 1.392 - .0135(20) + .00148(1050) = 2.676$.

(iii)(5 marks) The difference between A and B is simply 140 times the coefficient on sat, because hsperc is the same for both students. So A is predicted to have a score $.00148(140) \approx .207 higher$.

(iv) (10 marks)With *hsperc* fixed, $\triangle \widehat{colgpa} = .00148.\text{sat}$. Now, we want to find $\triangle sat$ such that $\triangle \widehat{colgpa} = .5$, so $.5 = .00148(\triangle sat)$ or $\triangle sat = .5/(.00148) = 338$. Perhaps not surprisingly, a large ceteris paribus difference in SAT score almost two and one-half standard deviations is needed to obtain a predicted difference in college GPA or a half a point.

3.3

(i)(5 marks) If adults trade off sleep for work, more work implies less sleep (other things equal), so $\beta_1 < 0$.

(ii)(5 marks) The signs of β_2 and β_3 are not obvious. One could argue that more educated people like to get more out of life, and so, other things equal, they sleep less ($\beta_2 < 0$). The relationship between sleeping and age is more complicated than this model suggests, and economists are not in the best position to judge such things.

(iii)(5 marks) Since *totwrk* is in minutes, we must convert five hours into minutes: $\triangle totwrk = 5(60) = 300$. Then sleep is predicted to fall by $.148(300) = 44.4$ minutes. For a week, 45 minutes less sleep is not an overwhelming change.

(iv)(5 marks) More education implies less predicted time sleeping, but the effect is quite small. If we assume the difference between college and high school is four years, the college graduate sleeps about 45 minutes less per week, other things equal.

(v)(10 marks) Not surprisingly, the three explanatory variables explain only about 11.3 percent of the variation in *sleep*. One important factor in the error term is general health. Another is marital status, and whether the person has children. Health (however we measure that), marital status, and number and ages of children would generally be correlated with *totwrk*. (For example, less healthy people would tend to work less.)

3.4

(i)(5 marks) A larger rank for a law school means that the school has less prestige; this lowers starting salaries. For example, a rank of 100 means there are 99 schools thought to be better.

(ii)(10 marks) $\beta_1 > 0$, $\beta_2 > 0$. Both LSAT and GPA are measures of the quality of the entering class. No matter where better students attend law school, we expect them to earn more, on average. $\beta_3 > 0$, $\beta_4 > 0$. The number of volumes in the law library and the tuition cost are both measures of the school quality. (Cost is less obvious than library volumes, but should reflect quality of the faculty, physical plant, and so on.)

(iii) (5 marks)This is just the coefficient on GPA, multiplied by 100: 24.8 percent.

(iv)(5 marks) This is an elasticity: a one percent increase in library volumes implies a .095 percent increase in predicted median starting salary, other things equal. (v) It is definitely better to attend a law school with a lower rank. If law school A has a ranking 20 less than law school B, the predicted difference in starting salary is 100(.0033)(20) = 6.6 percent higher for law school A.

C2.1

(i) (5 marks)

```
The average participation rate is 86.88214, the average match rate is .7510169.
. summ prate

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       prate |       767    86.88214    16.96393       20.1        100

. summ mrate

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       mrate |       767    .7510169    .7829485        .02       4.91
```

(ii) (marks)

```
. regress prate mrate
```

```
      Source |       SS       df       MS              Number of obs =     767
-------------+------------------------------           F(  1,   765) =   75.21
       Model |   19731.386     1   19731.386           Prob > F      =  0.0000
    Residual |   200704.26   765   262.35851           R-squared     =  0.0895
-------------+------------------------------           Adj R-squared =  0.0883
       Total |  220435.646   766  287.774995           Root MSE      =  16.197

------------------------------------------------------------------------------
       prate |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       mrate |   6.482331   .7474807     8.67   0.000     5.014974    7.949688
       _cons |    82.0138   .8106757   101.17   0.000     80.42238    83.60521
------------------------------------------------------------------------------
```

Answer: so $\widehat{prate} = 82.0138 + 6.482331 mrate$. Sample size is 767, and R^2 is 0.0895.

(iii) (10 marks) if $mrate = 0$, the predicted participation rate is 82.0138 percent. Coefficient in *mrate* implies that a one dollar increase in the match rate, fairly large increase is estimated to increase prate by 6.482331 percentage points. This assumes, of course, that this change prate is possible (if, say, prate is already at 98, this interpretation makes no sense).

(iv) (5 marks)If we plug 3.5 in the equation, we get $prate = 82.0138 + 3.5 * 6.482331 = 104.702$.This is impossible, as we can have at most a 100 percent participation rate. This illustrates that, especially when dependent variables are bounded, a simple regression model can give strange predictions for extreme values of the independent variable. (In the sample of 765 firms, only 15 have $mrate > 3.5$.)

(v) (5 marks) *mrate* explains 8.95 percent of the variation. This is not much, and many other factors may affect participation rate.

C2.4

(i) (5 marks)

```
. summ IQ wage

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

```
          IQ |       935    101.2824    15.05264         50        145
        wage |       935    957.9455    404.3608        115       3078
```

Answer: Average salary is about 957.95 dollars and average IQ is about 101.28. The sample standard deviation of IQ is about 15.05, which is pretty close to the population value of 15.

(ii) (10 marks)

```
. regress wage IQ

      Source |       SS       df       MS              Number of obs =     935
-------------+------------------------------           F(  1,   933) =   98.55
       Model |  14589782.6     1  14589782.6           Prob > F      =  0.0000
    Residual |   138126386   933  148045.429           R-squared     =  0.0955
-------------+------------------------------           Adj R-squared =  0.0946
       Total |   152716168   934  163507.675           Root MSE      =  384.77

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          IQ |   8.303064   .8363951     9.93   0.000     6.661631    9.944498
       _cons |   116.9916   85.64153     1.37   0.172    -51.08078    285.0639
------------------------------------------------------------------------------
```

Answer: $\widehat{wage} = 116.99 + 8.30IQ$, n=935, $R^2 = 0.096$. An increase in IQ of 15 increases predicted monthly salary by 8.30(15) = 124.50 (in 1980 dollars). IQ score does not even explain 10 percent of the variation in wage.

(iii) (10 marks)

```
. regress lwage IQ

      Source |       SS       df       MS              Number of obs =     935
-------------+------------------------------           F(  1,   933) =  102.62
       Model |  16.4150981     1  16.4150981           Prob > F      =  0.0000
    Residual |  149.241196   933   .15995841           R-squared     =  0.0991
-------------+------------------------------           Adj R-squared =  0.0981
       Total |  165.656294   934  .177362199           Root MSE      =  .39995
```

```
------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          IQ |   .0088072   .0008694    10.13   0.000      .007101    .0105134
       _cons |   5.886994   .0890206    66.13   0.000      5.71229    6.061698
------------------------------------------------------------------------------
```

Answer: $\widehat{log(wage)} = 5.89 + 0.0088IQ$, n=935, $R^2 = 99$. If $\triangle IQ = 15$ then $\triangle \widehat{log(wage)} = 0.0088(15) = 0.132$, which is (approximate) proportinate change in predicted wage. The percentage increase is therefore approximately 13.2.

C3.1

(i) (5 marks) Probably $\beta_2 > 0$, as more income typically means better nutrition for the mother and better prenatal care.

(ii) (10 marks) On the one hand, an increase in income generally increases the consumption of a good, and *cigs* and $famcin$ could be positively correlated. On the other hand, family incomes are also higher for families with more education, and more education and cigarette smoking tend to be negatively correlated. The sample correlation between *cigs* and $faminc$ is about -0.173, indicating a negative correlation.

(iii) (10 marks)

```
. regress  bwght cigs

      Source |       SS       df       MS              Number of obs =     694
-------------+------------------------------           F(  1,   692) =   25.33
       Model |  10394.4794     1  10394.4794           Prob > F      =  0.0000
    Residual |  283941.338   692  410.319852           R-squared     =  0.0353
-------------+------------------------------           Adj R-squared =  0.0339
       Total |  294335.817   693  424.727009           Root MSE      =  20.256

------------------------------------------------------------------------------
       bwght |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        cigs |   -.601789    .119565    -5.03   0.000    -.8365427   -.3670353
```

```
       _cons |   120.3839    .821228   146.59   0.000     118.7715    121.9963
------------------------------------------------------------------------------

. regress bwght cigs faminc

      Source |       SS       df       MS              Number of obs =     694
-------------+------------------------------           F(  2,   691) =   14.21
       Model |   11626.062     2  5813.03102           Prob > F      =  0.0000
    Residual |  282709.755   691  409.131339           R-squared     =  0.0395
-------------+------------------------------           Adj R-squared =  0.0367
       Total |  294335.817   693  424.727009           Root MSE      =  20.227

------------------------------------------------------------------------------
       bwght |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        cigs |  -.5632265   .1214429    -4.64   0.000     -.801668   -.3247851
      faminc |    .073165   .0421699     1.74   0.083    -.0096316    .1559616
       _cons |   118.1664   1.518518    77.82   0.000      115.185    121.1479
------------------------------------------------------------------------------
```

$\widehat{bwght} = 120.3839 - .601789 cigs$, n=694, $R^2 = 0.0353$, another equation with $faminc$ $\widehat{bwght} = 118.1664 - -.5632265 cigs + .073165 faminc$, n=694, $R^2 = 0.0395$ The effect of cigarette smoking is slightly smaller when $faminc$ is added to the regression, but the difference is not great. This is due to the fact that *cigs* and $faminc$ are not very correlated, and the coefficient on $faminc$ is practically small.(The variable $faminc$ is measured in thousands, so 10000 more dollars in 1988 inome increases predicted weight by only 0.93 ounces.)