ECNS 562 –Econometrics II
Eric Belasco
Homework 1
Due Thursday, February 2

1. Consider the simple linear regression model:

$$y_i = \beta_0 + \beta_1 x_i + u_i,$$

where $E(u_i|x_i) = 0$. Define a new variable, z_i , which is a transformation of the independent variable: $\hat{z}_i = a + b\log(x_i)$, such that $\hat{z}_i$ and u_i are independent. Define a new slope estimator as:

$$\tilde{\beta}_1 = \frac{S_{ZY}}{S_{ZX}} = \frac{\sum_{i=1}^{n}(\hat{z}_i - \bar{\hat{z}})y_i}{\sum_{i=1}^{n}(\hat{z}_i - \bar{\hat{z}})x_i}$$

(a) Show that $\tilde{\beta}_1$ is unbiased
(b) Adding the assumption that $var(u_i) = \sigma^2$ for all i, and that the u_i are mutually uncorrelated, what is $Var(\hat{\beta}_1)$?
(c) What does the Gauss-Markov Theorem have to say about the relationship between $Var(\tilde{\beta}_1)$ and $(\hat{\beta}_1)$, where $\hat{\beta}_1$ is the least squares slope estimator?
(d) Suppose that n=10 and that the variable x_i takes on the values 1,2,…,10, while $a = 1$ and $b = 3$. What is the ratio of $Var(\hat{\beta}_1)$ to $Var(\tilde{\beta}_1)$? (That is, what is the relative efficiency of the least squares estimator?) Does this support your statement in part (c)?

2. You are asked by the USDA to conduct a survey of the grizzly bear encroaching on livestock rangeland in Montana. You are asked to scout out one particular rangeland area and count the number of grizzly bear sightings per day. You assume the number of grizzly bears sighted (y_i) on day i can be distributed as a Poisson distribution such that:

$$f(y_i|\theta) = \frac{e^{-\theta}\theta^{y_i}}{y_i!}$$

After 10 days, you submit a report with observations from each day, which are y = (5,0,1,1,0,3,2,3,4,1).

(a) (4 points) Derive the log-likelihood function necessary for computing the Maximum Likelihood Estimate of θ.

(b) (6 points) Using the observations over the first 10 days, derive the optimal solution for θ (Show your work). (*Hint: The optimal solution for θ is where* $\frac{dLL}{d\theta} = 0$)

(c) Use R to minimize the $-LL$ derived in part (a) with respect to θ. Is your answer the same as in part (b)? Explain why this is the case.

3. Download the dataset labeled "mroz87" in order to address the following questions. The following command will be used to download the data.

```
mroz87 =  read.table( "http://www.montana.edu/ebelasco/ecns562/homework/mroz87.dat", header = T)
```

(a) Create a subset of mroz87 that only includes positive wage earners by using the following command, '**posWage=subset(mroz87, wage>0)**.' Notice that posWage is the new dataset that will be used in this analysis. Next, create a new variable by taking the log of wage ($logWage$).

(b) Fit the following model using ordinary least squares.

$$\log Wage_i = \beta_0 + \beta_1 educ_i + \beta_2 exper_i + \beta_3 exper_i^2 + \epsilon_i$$

Given that an important omitted variable, $ability$, is correlated with $educ$, describe how parameter estimates are unreliable. Note: Assume that $exper$ and $exper^2$ are exogenous.

(c) Two variables that can be used as instruments for $educ$ include $mothereduc$ and $fathereduc$. Show what the first stage regression is and what can be learned from its results.

(d) Use the results from part (c) to derive the 2SLS parameter estimates for $\beta_0, \beta_1, \beta_2, \& \beta_3$. Compare these results with the OLS results obtained in part (b). Explain why or why not it appears that 2SLS results are consistent, based on the difference between the OLS and 2SLS results.

(e) Use the tsls() function in the SEM package in order to obtain 2SLS estimates. Compare the 2SLS results with that of the results derived in part (d). Explain the difference between the standard errors reported under the two methods and describe which is more accurate.

(f) Interpret the marginal impact of an additional year of $educ$ on $logWage$. Also, interpret the marginal impact of an additional year of $exper$ on $logWage$, at the mean level of experience.