## Cornell University <br> Department of Economics

Econ 620 - Spring 2004
Instructor: Prof. Kiefer
Solution to Problem set \# 4
1)

For Model I;

$$
X^{\prime} X=\left[\begin{array}{ll}
X_{1}^{\prime} X_{1} & X_{1}^{\prime} X_{2} \\
X_{2}^{\prime} X_{1} & X_{2}^{\prime} X_{2}
\end{array}\right] \text { and } X^{\prime} y=\left[\begin{array}{c}
X_{1}^{\prime} y \\
X_{2}^{\prime} y
\end{array}\right]
$$

Therefore,

$$
\widehat{\beta}=\left[\begin{array}{c}
\widehat{\beta}_{1} \\
\widehat{\beta}_{2}
\end{array}\right]=\left(X^{\prime} X\right)^{-1} X^{\prime} y=\left[\begin{array}{ll}
X_{1}^{\prime} X_{1} & X_{1}^{\prime} X_{2} \\
X_{2}^{\prime} X_{1} & X_{2}^{\prime} X_{2}
\end{array}\right]^{-1}\left[\begin{array}{c}
X_{1}^{\prime} y \\
X_{2}^{\prime} y
\end{array}\right]
$$

We are interested only in $\widehat{\beta}_{1}$. We have to figure out two upper blocks of the inverse matrix to calculate $\widehat{\beta}_{1}$. By the formula of partitioned inverse, the upperleft corner of the inverse matrix is given by

$$
\begin{aligned}
& \left(X_{1}^{\prime} X_{1}-X_{1}^{\prime} X_{2}\left(X_{2}^{\prime} X_{2}\right)^{-1} X_{2}^{\prime} X_{1}\right)^{-1} \\
& =\left(X_{1}^{\prime}\left(I-X_{2}\left(X_{2}^{\prime} X_{2}\right)^{-1} X_{2}^{\prime}\right) X_{1}\right)^{-1}=\left(X_{1}^{\prime} M_{2} X_{1}\right)^{-1}
\end{aligned}
$$

where $M_{2}=I-X_{2}\left(X_{2}^{\prime} X_{2}\right)^{-1} X_{2}^{\prime}$. And the upper-right corner of the inverse matrix is

$$
-\left(X_{1}^{\prime} M_{2} X_{1}\right)^{-1} X_{1}^{\prime} X_{2}\left(X_{2}^{\prime} X_{2}\right)^{-1}
$$

(To see this, look at Magnus and Neudecker, Matrix differential calculus, page 11 and 12. This book is on reserve in Malott library). Hence, we have

$$
\begin{aligned}
\widehat{\beta}_{1} & =\left(X_{1}^{\prime} M_{2} X_{1}\right)^{-1} X_{1}^{\prime} y-\left(X_{1}^{\prime} M_{2} X_{1}\right)^{-1} X_{1}^{\prime} X_{2}\left(X_{2}^{\prime} X_{2}\right)^{-1} X_{2}^{\prime} y \\
& =\left(X_{1}^{\prime} M_{2} X_{1}\right)^{-1} X_{1}^{\prime}\left[I-X_{2}\left(X_{2}^{\prime} X_{2}\right)^{-1} X_{2}^{\prime}\right] y=\left(X_{1}^{\prime} M_{2} X_{1}\right)^{-1}\left(X_{1}^{\prime} M_{2} y\right)
\end{aligned}
$$

On the other hand, the least squares estimator for $\beta_{1}$ in the Model II is;

$$
\begin{aligned}
\widetilde{\beta}_{1} & =\left(X_{1}^{\prime} M_{2}^{\prime} M_{2} X_{1}\right)^{-1}\left(X_{1}^{\prime} M_{2}^{\prime} M_{2} y\right) \\
& =\left(X_{1}^{\prime} M_{2} X_{1}\right)^{-1}\left(X_{1}^{\prime} M_{2} y\right)=\widehat{\beta}_{1}
\end{aligned}
$$

Note that $M_{2}$ is symmetric and idempotent. $M_{2}$ is a projection matrix onto a subspace orthogonal to a subspace spanned by columns of $X_{2}$. Therefore, $M_{2} X_{1}$
is an $\left(n \times k_{1}\right)$ matrix whose $j^{t h}$ column consists of residuals from the regression of the $j^{\text {th }}$ column of $X_{1}$ on $X_{2}$. In the Model II, we actually regress residual vector $M_{2} y$ on the matrix of another residuals $M_{2} X_{1}$. Moreover, you can show that the residual sums of squares from the two models are identical. It is harder but doable. Why don't you try?

$$
\begin{aligned}
e_{I}^{\prime} e_{I} & =\left(y-X_{1} \widehat{\beta}_{1}-X_{2} \widehat{\beta}_{2}\right)^{\prime}\left(y-X_{1} \widehat{\beta}_{1}-X_{2} \widehat{\beta}_{2}\right) \\
& =\left(M_{2} y-M_{2} X_{1} \widetilde{\beta}_{1}\right)^{\prime}\left(M_{2} y-M_{2} X_{1} \widetilde{\beta}_{1}\right)=e_{I I}^{\prime} e_{I I}
\end{aligned}
$$

2) 

Rewrite the model as;

$$
Y=\beta_{0}+Z_{1} \beta_{1}+Z_{2} \beta_{2}+\varepsilon
$$

where $Y=\log Q, Z_{1}=\log L$, and $Z_{2}=\log K$. In mean deviation form,

$$
y=X_{1} \beta_{1}+X_{2} \beta_{2}+\varepsilon
$$

where $y=A y, X_{1}=A Z_{1}$, and $X_{2}=A Z_{2}$ with $A=\frac{1}{N} \mathbf{1 1}^{\prime}$. Instead of the true model, we run the regression;

$$
y=X_{1} \beta_{1}+\epsilon
$$

The true estimator would be;

$$
\widehat{\beta}_{1}=\left(X_{1}^{\prime} M_{2} X_{1}\right)^{-1} X_{1}^{\prime} M_{2} y
$$

The estimator from the false regression is;

$$
\widetilde{\beta}_{1}=\left(X_{1}^{\prime} X_{1}\right)^{-1} X_{1}^{\prime} y
$$

Suppose that $X_{1}$ and $X_{2}$ are orthogonal - what does it mean? - , $X_{1}^{\prime} X_{2}=0$. Then,

$$
\begin{aligned}
X_{1}^{\prime} M_{2} X_{1} & =X_{1}^{\prime}\left[I-X_{2}\left(X_{2}^{\prime} X_{2}\right)^{-1} X_{2}^{\prime}\right] X_{1} \\
& =X_{1}^{\prime} X_{1}-X_{1}^{\prime} X_{2}\left(X_{2}^{\prime} X_{2}\right)^{-1} X_{2}^{\prime} X_{1} \\
& =X_{1}^{\prime} X_{1} \\
X_{1}^{\prime} M_{2} y & =X_{1}^{\prime}\left[I-X_{2}\left(X_{2}^{\prime} X_{2}\right)^{-1} X_{2}^{\prime}\right] y \\
& =X_{1}^{\prime} y-X_{1}^{\prime} X_{2}\left(X_{2}^{\prime} X_{2}\right)^{-1} X_{2}^{\prime} y \\
& =X_{1}^{\prime} y
\end{aligned}
$$

Hence,

$$
\text { If } X_{1} \text { and } X_{2} \text { are orthogonal, } \widetilde{\beta}_{1}=\widehat{\beta}_{1}
$$

i.e. dropping the capital variable does not affect the coefficient on labor. You can verify that $\operatorname{Var}\left(\widetilde{\beta}_{1}\right)=\operatorname{Var}\left(\widehat{\beta}_{1}\right)$ if $X_{1}$ and $X_{2}$ are orthogonal.
3)
(a) Easy! -has nothing to do with econometrics. It is directly from the definition of the constant returns to scale. Use the famous Euler's theorem.
(b)

$$
\begin{gathered}
\frac{\partial \log Q}{\partial \log L}=\beta_{2}+\beta_{4} \log L+\beta_{6} \log K \\
\frac{\partial \log Q}{\partial \log K}=\beta_{3}+\beta_{5} \log K+\beta_{6} \log L
\end{gathered}
$$

Hence, constant returns to scale requires that

$$
\begin{aligned}
& \beta_{2}+\beta_{4} \log L+\beta_{6} \log K+\beta_{3}+\beta_{5} \log K+\beta_{6} \log L \\
& =\left(\beta_{2}+\beta_{3}\right)+\left(\beta_{4}+\beta_{6}\right) \log L+\left(\beta_{5}+\beta_{6}\right) \log K=1
\end{aligned}
$$

The last equality requires that

$$
\left(\beta_{2}+\beta_{3}\right)=1 \wedge\left(\beta_{4}+\beta_{6}\right)=0 \wedge\left(\beta_{5}+\beta_{6}\right)=0
$$

The above is our null hypothesis. To test the hypothesis, first of all, we run the unrestricted regression given in the question and keep the residual sum of squares and then run the restricted regression as shown below; Imposing the restrictions, we have;

$$
\begin{aligned}
\log Q & =\beta_{1}+\beta_{2} \log L+\beta_{3} \log K+\beta_{4} \frac{(\log L)^{2}}{2}+\beta_{5} \frac{(\log K)^{2}}{2}+\beta_{6} \log L \log K+\varepsilon \\
& =\beta_{1}+\beta_{2} \log L+\left(1-\beta_{2}\right) \log K \\
& +\left(1-\beta_{6}\right) \frac{(\log L)^{2}}{2}+\left(1-\beta_{6}\right) \frac{(\log K)^{2}}{2}+\beta_{6} \log L \log K+\varepsilon \\
& \Rightarrow \log Q-\log K-\frac{(\log L)^{2}}{2}-\frac{(\log K)^{2}}{2} \\
& =\beta_{1}+\beta_{2}(\log L-\log K)+\beta_{6}\left(\log L \log K-\frac{(\log L)^{2}}{2}-\frac{(\log K)^{2}}{2}\right)+\varepsilon
\end{aligned}
$$

Define

$$
\begin{aligned}
y & =\log Q-\log K-\frac{(\log L)^{2}}{2}-\frac{(\log K)^{2}}{2} \\
X_{1} & =(\log L-\log K) \\
X_{2} & =\left(\log L \log K-\frac{(\log L)^{2}}{2}-\frac{(\log K)^{2}}{2}\right)
\end{aligned}
$$

Then, we regress $y$ on a constant, $X_{1}$, and $X_{2}$ to get the restricted model. The estimation will results in the restricted residual sum of squares. Then, we form the $F$ - statistic;

$$
F=\frac{\left(e_{R}^{\prime} e_{R}-e_{U}^{\prime} e_{U}^{\prime}\right) / 3}{e_{U}^{\prime} e_{U}^{\prime} /(N-6)} \sim F(3, N-6)
$$

4) 

(a) Note that

$$
\widehat{\beta}=\left(X^{\prime} X\right)^{-1} X^{\prime} y=\beta+\left(X^{\prime} X\right)^{-1} X^{\prime} \varepsilon
$$

Hence,

$$
E(\widehat{\beta})=\beta+\left(X^{\prime} X\right)^{-1} X^{\prime} E(\varepsilon)=\beta \text { unbiased }
$$

Moreover,

$$
\begin{aligned}
E(\widetilde{\beta}) & =E\left[\widehat{\beta}+N^{-1} \mathbf{1}\right]=E(\widehat{\beta})+N^{-1} \mathbf{1}=\beta+N^{-1} \mathbf{1} \text { biased } \\
E(\bar{\beta}) & =E\left[\widehat{\beta}+N^{-\frac{1}{2}} \mathbf{1}\right]=E(\widehat{\beta})+N^{-\frac{1}{2}} \mathbf{1}=\beta+N^{-\frac{1}{2}} \mathbf{1} \text { biased }
\end{aligned}
$$

Note that biases of both $\widetilde{\beta}$ and $\bar{\beta}$ vanish as $N \rightarrow \infty$.
(b) As shown in (a) of question 4, under some proper conditions - what are they? -

$$
\widehat{\beta} \xrightarrow{p} \beta
$$

i.e. $\widehat{\beta}$ is consistent. On the other hand,

$$
\begin{aligned}
& \operatorname{plim} \widetilde{\beta}=\operatorname{plim}\left[\widehat{\beta}+N^{-1} \mathbf{1}\right]=\operatorname{plim} \widehat{\beta}+\operatorname{plim}_{N \rightarrow \infty} N^{-1} \mathbf{1}=\beta \\
& \operatorname{plim} \bar{\beta}=\operatorname{plim}\left[\widehat{\beta}+N^{-\frac{1}{2}} \mathbf{1}\right]=\operatorname{plim} \widehat{\beta}+\operatorname{plim}_{N \rightarrow \infty} N^{-\frac{1}{2}} \mathbf{1}=\beta
\end{aligned}
$$

Hence, both $\widetilde{\beta}$ and $\bar{\beta}$ are consistent.
(c) Again, under some proper conditions, we have

$$
\sqrt{N}(\widehat{\beta}-\beta) \xrightarrow{d} N\left(\mathbf{0}, Q^{-1}\right)
$$

where $Q=\operatorname{plim} \frac{X^{\prime} X}{N}$. Then,

$$
\sqrt{N}(\widetilde{\beta}-\beta)=\sqrt{N}\left(\widehat{\beta}+N^{-1} \mathbf{1}-\beta\right)=\sqrt{N}(\widehat{\beta}-\beta)+N^{-\frac{1}{2}} \mathbf{1}
$$

Note that

$$
\sqrt{N}(\widehat{\beta}-\beta) \xrightarrow{d} N\left(\mathbf{0}, Q^{-1}\right) \text { and } N^{-\frac{1}{2}} \mathbf{1} \xrightarrow{p} 0
$$

Hence,

$$
\sqrt{N}(\widetilde{\beta}-\beta) \xrightarrow{d} N\left(\mathbf{0}, Q^{-1}\right)
$$

On the other hand,

$$
\sqrt{N}(\bar{\beta}-\beta)=\sqrt{N}\left(\widehat{\beta}+N^{-\frac{1}{2}} \mathbf{1}-\beta\right)=\sqrt{N}(\widehat{\beta}-\beta)+\mathbf{1}
$$

Again,

$$
\sqrt{N}(\widehat{\beta}-\beta) \xrightarrow{d} N\left(\mathbf{0}, Q^{-1}\right) \text { and } \mathbf{1} \xrightarrow{p} \mathbf{1}
$$

Therefore,

$$
\sqrt{N}(\bar{\beta}-\beta) \xrightarrow{d} N\left(\mathbf{1}, Q^{-1}\right)
$$

5) 

(a) Since $x_{i}^{\prime} s$ are i.i.d. with the finite first moment, we can resort to the WLLN to conclude that

$$
\operatorname{plim} \frac{1}{n} \sum_{i=1}^{n} x_{i}=E\left(x_{i}\right)
$$

i.e.

$$
\operatorname{plim} \frac{1}{n} \sum_{i=1}^{n} x_{i}=\frac{1}{2}
$$

For the second quantity, it is simple again the sample average of i.i.d. random variables $x_{i}^{2}$ with the finite first moment. We have

$$
\operatorname{plim} \frac{1}{n} \sum_{i=1}^{n} x_{i}^{2}=E\left(x_{i}^{2}\right)
$$

i.e.

$$
\operatorname{plim} \frac{1}{n} \sum_{i=1}^{n} x_{i}^{2}=\frac{1}{3}
$$

By the same reason, we infer that

$$
\operatorname{plim} \frac{1}{n} \sum_{i=1}^{n} x_{i}^{3}=E\left(x_{i}^{3}\right)=\frac{1}{4}
$$

(b) Note that

$$
\begin{aligned}
\widehat{\beta} & =\frac{\sum_{i=1}^{n}\left(x_{i}-\bar{x}_{n}\right)\left(y_{i}-\bar{y}_{n}\right)}{\sum_{i=1}^{n}\left(x_{i}-\bar{x}_{n}\right)^{2}}=\frac{\sum_{i=1}^{n}\left(x_{i}-\bar{x}_{n}\right) y_{i}}{\sum_{i=1}^{n}\left(x_{i}-\bar{x}_{n}\right)^{2}} \\
& =\frac{\sum_{i=1}^{n}\left(x_{i}-\bar{x}_{n}\right)\left(x_{i}^{2}+\varepsilon_{i}\right)}{\sum_{i=1}^{n}\left(x_{i}-\bar{x}_{n}\right)^{2}}=\frac{\sum_{i=1}^{n} x_{i}^{2}\left(x_{i}-\bar{x}_{n}\right)+\sum_{i=1}^{n} \varepsilon_{i}\left(x_{i}-\bar{x}_{n}\right)}{\sum_{i=1}^{n}\left(x_{i}-\bar{x}_{n}\right)^{2}} \\
& =\frac{\frac{1}{n} \sum_{i=1}^{n} x_{i}^{3}-\frac{1}{n} \sum_{i=1}^{n} x_{i}^{2} \bar{x}_{n}+\frac{1}{n} \sum_{i=1}^{n} \varepsilon_{i} x_{i}-\frac{1}{n} \sum_{i=1}^{n} \varepsilon_{i} \bar{x}_{n}}{\frac{1}{n} \sum_{i=1}^{n}\left(x_{i}-\bar{x}_{n}\right)^{2}}
\end{aligned}
$$

We will take care of the denominator first.

$$
\begin{aligned}
\operatorname{plim} \frac{1}{n} \sum_{i=1}^{n}\left(x_{i}-\bar{x}_{n}\right)^{2} & =\operatorname{plim} \frac{1}{n}\left(\sum_{i=1}^{n} x_{i}^{2}-n \bar{x}_{n}^{2}\right)=\operatorname{plim} \frac{1}{n} \sum_{i=1}^{n} x_{i}^{2}-\operatorname{plim} \bar{x}_{n}^{2} \\
& =\operatorname{plim} \frac{1}{n} \sum_{i=1}^{n} x_{i}^{2}-\operatorname{plim}\left[\frac{1}{n} \sum_{i=1}^{n} x_{i}\right]^{2} \\
& =\operatorname{plim} \frac{1}{n} \sum_{i=1}^{n} x_{i}^{2}-\left[\operatorname{plim} \frac{1}{n} \sum_{i=1}^{n} x_{i}\right]^{2}=\frac{1}{3}-\left(\frac{1}{2}\right)^{2}=\frac{1}{12}
\end{aligned}
$$

For the first two terms in numerator,

$$
\begin{aligned}
\operatorname{plim} \frac{1}{n} \sum_{i=1}^{n} x_{i}^{3} & =\frac{1}{4} \\
\operatorname{plim} \frac{1}{n} \sum_{i=1}^{n} x_{i}^{2} \bar{x}_{n} & =\operatorname{plim} \bar{x}_{n} \frac{1}{n} \sum_{i=1}^{n} x_{i}^{2}=\operatorname{plim} \bar{x}_{n} \operatorname{plim} \frac{1}{n} \sum_{i=1}^{n} x_{i}^{2} \\
& =\operatorname{plim} \frac{1}{n} \sum_{i=1}^{n} x_{i} \operatorname{plim} \frac{1}{n} \sum_{i=1}^{n} x_{i}^{2}=\frac{1}{2} \times \frac{1}{3}=\frac{1}{6}
\end{aligned}
$$

Finally, note that $\frac{1}{n} \sum_{i=1}^{n} \varepsilon_{i} x_{i}$ is the sample average of i.i.d. random variables $\varepsilon_{i} x_{i}$ with finite expectation. Again by the WLLN,

$$
\operatorname{plim} \frac{1}{n} \sum_{i=1}^{n} \varepsilon_{i} x_{i}=E\left(\varepsilon_{i} x_{i}\right)=E\left(\varepsilon_{i}\right) E\left(x_{i}\right)=0
$$

On the other hand,

$$
\begin{aligned}
\operatorname{plim} \frac{1}{n} \sum_{i=1}^{n} \varepsilon_{i} \bar{x}_{n} & =\operatorname{plim} \bar{x}_{n} \frac{1}{n} \sum_{i=1}^{n} \varepsilon_{i}=\operatorname{plim} \bar{x}_{n} \operatorname{plim} \frac{1}{n} \sum_{i=1}^{n} \varepsilon_{i} \\
& =\operatorname{plim} \frac{1}{n} \sum_{i=1}^{n} x_{i} \operatorname{plim} \frac{1}{n} \sum_{i=1}^{n} \varepsilon_{i}=0
\end{aligned}
$$

In sum, we have

$$
\operatorname{plim} \widehat{\beta}=\frac{\left(\frac{1}{4}-\frac{1}{6}\right)}{\frac{1}{12}}=1
$$

For the intercept term,

$$
\begin{aligned}
\widehat{\alpha} & =\bar{y}_{n}-\widehat{\beta} \bar{x}_{n} \\
& =\frac{1}{n} \sum_{i=1}^{n} y_{i}-\widehat{\beta} \frac{1}{n} \sum_{i=1}^{n} x_{i}
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\operatorname{plim} \widehat{\alpha} & =\operatorname{plim}\left(\frac{1}{n} \sum_{i=1}^{n} y_{i}-\widehat{\beta} \frac{1}{n} \sum_{i=1}^{n} x_{i}\right)=\operatorname{plim} \frac{1}{n} \sum_{i=1}^{n} y_{i}-\operatorname{plim} \widehat{\beta} \frac{1}{n} \sum_{i=1}^{n} x_{i} \\
& =\operatorname{plim} \frac{1}{n} \sum_{i=1}^{n}\left(x_{i}^{2}+\varepsilon_{i}\right)-\operatorname{plim} \widehat{\beta} \operatorname{plim} \frac{1}{n} \sum_{i=1}^{n} x_{i} \\
& =\operatorname{plim} \frac{1}{n} \sum_{i=1}^{n} x_{i}^{2}+\operatorname{plim} \frac{1}{n} \sum_{i=1}^{n} \varepsilon_{i}-\operatorname{plim} \widehat{\beta} \operatorname{plim} \frac{1}{n} \sum_{i=1}^{n} x_{i} \\
& =E\left(x_{i}^{2}\right)+E\left(\varepsilon_{i}\right)-1 \times E\left(x_{i}\right)=\frac{1}{3}+0-\frac{1}{2}=-\frac{1}{6}
\end{aligned}
$$

(c) Note that

$$
E\left(\frac{d y}{d x}\right)=E\left(\frac{d}{d x}\left(x_{i}^{2}+\varepsilon_{i}\right)\right)=E\left(2 x_{i}\right)=2 \times \frac{1}{2}=1
$$

Then, $\widehat{\beta} \xrightarrow{p} E\left(\frac{d y}{d x}\right)$
(d) By the definition of the least squares estimator,

$$
\begin{aligned}
\widehat{\gamma} & =\frac{\sum_{i=1}^{n} x_{i} y_{i}}{\sum_{i=1}^{n} x_{i}^{2}}=\frac{\sum_{i=1}^{n} x_{i}\left(x_{i}^{2}+\varepsilon_{i}\right)}{\sum_{i=1}^{n} x_{i}^{2}}=\frac{\sum_{i=1}^{n} x_{i}^{3}+\sum_{i=1}^{n} x_{i} \varepsilon_{i}}{\sum_{i=1}^{n} x_{i}^{2}} \\
& =\frac{\frac{1}{n} \sum_{i=1}^{n} x_{i}^{3}+\frac{1}{n} \sum_{i=1}^{n} x_{i} \varepsilon_{i}}{\frac{1}{n} \sum_{i=1}^{n} x_{i}^{2}}
\end{aligned}
$$

Note that

$$
\begin{aligned}
\operatorname{plim} \frac{1}{n} \sum_{i=1}^{n} x_{i}^{3} & =E\left(x_{i}^{3}\right)=\frac{1}{4} \\
\operatorname{plim} \frac{1}{n} \sum_{i=1}^{n} x_{i} \varepsilon_{i} & =E\left(x_{i} \varepsilon_{i}\right)=E\left(x_{i}\right) E\left(\varepsilon_{i}\right)=0 \\
\operatorname{plim} \frac{1}{n} \sum_{i=1}^{n} x_{i}^{2} & =E\left(x_{i}^{2}\right)=\frac{1}{3}
\end{aligned}
$$

Then,

$$
\operatorname{plim} \widehat{\gamma}=\frac{\frac{1}{4}}{\frac{1}{3}}=\frac{3}{4}
$$

Since, $E\left(\frac{d y}{d x}\right)=1, \widehat{\gamma} \xrightarrow{p} E\left(\frac{d y}{d x}\right)$.
6)
(a) We have $N$ sets of observations on $\left(y_{i}, x_{i}\right)$ to estimate the parameters $\left(\beta^{\prime}, \sigma^{2}\right)$. In addition, we know the error distribution. Considering the linear relationship between $y$ and $\varepsilon$, we can immediately identify the distribution of $y$ as

$$
y \sim N\left(X \beta, \sigma^{2} I\right)
$$

Conceptually, we have to start the joint density function of $\left(y_{1}, y_{2}, \cdots, y_{N}\right)$ since that is the data we have. Let's denote the joint density as

$$
P\left(y_{1}, y_{2}, \cdots, y_{N} ; \beta, \sigma^{2}\right)
$$

What do we know about the distribution of $y_{i}^{\prime} s$ ? First of all, they are independent so that we can express the joint density function as the product of the marginal density functions;

$$
P\left(y_{1}, y_{2}, \cdots, y_{N} ; \beta, \sigma^{2}\right)=\prod_{i=1}^{N} p_{i}\left(y_{i} ; \beta, \sigma^{2}\right)
$$

where $p_{i}\left(y_{i} ; \beta, \sigma^{2}\right)$ is the marginal distribution of the $i^{\text {th }}$ observation of $y$. Can we go farther than that? Yes, we know that all $y_{i}^{\prime} s$ have identical distribution. Then,

$$
P\left(y_{1}, y_{2}, \cdots, y_{N} ; \beta, \sigma^{2}\right)=\prod_{i=1}^{N} p\left(y_{i} ; \beta, \sigma^{2}\right)
$$

The density function is a function of random variable $y$ given some parameter value. Changing the perspective, we can treat the density function as a function of unknown parameters given data $y$. It is called the likelihood function;

$$
L\left(\beta, \sigma^{2} ; y_{1}, y_{2}, \cdots, y_{N}\right)=\prod_{i=1}^{N} p\left(y_{i} ; \beta, \sigma^{2}\right)
$$

Taking logs to get the log-likelihood function;

$$
\log L\left(\beta, \sigma^{2} ; y_{1}, y_{2}, \cdots, y_{N}\right)=\sum_{i=1}^{N} \log p\left(y_{i} ; \beta, \sigma^{2}\right)
$$

In case of i.i.d. sample, the log likelihood function of the whole observation is simply the sum of individual $\log$ likelihood functions. Let's figure out the form of $\log p\left(y_{i} ; \beta, \sigma^{2}\right)$. We know that

$$
y_{i} \sim N\left(\beta^{\prime} x_{i}, \sigma^{2}\right)
$$

hence,

$$
p\left(y_{i} ; \beta, \sigma^{2}\right)=\frac{1}{\sqrt{2 \pi \sigma^{2}}} \exp \left[-\frac{1}{2 \sigma^{2}}\left(y_{i}-\beta^{\prime} x_{i}\right)^{2}\right]
$$

Taking logs,

$$
\log p\left(y_{i} ; \beta, \sigma^{2}\right)=-\frac{1}{2} \log 2 \pi-\frac{1}{2} \log \sigma^{2}-\frac{1}{2 \sigma^{2}}\left(y_{i}-\beta^{\prime} x_{i}\right)^{2}
$$

Then,

$$
\begin{aligned}
\log L\left(\beta, \sigma^{2} ; y_{1}, y_{2}, \cdots, y_{N}\right) & =\sum_{i=1}^{N}\left[-\frac{1}{2} \log 2 \pi-\frac{1}{2} \log \sigma^{2}-\frac{1}{2 \sigma^{2}}\left(y_{i}-\beta^{\prime} x_{i}\right)^{2}\right] \\
& =-\frac{N}{2} \log 2 \pi-\frac{N}{2} \log \sigma^{2}-\frac{1}{2 \sigma^{2}} \sum_{i=1}^{N}\left(y_{i}-\beta^{\prime} x_{i}\right)^{2}
\end{aligned}
$$

Yes, this is the thing we want to have. Sometimes, it is more convenient to have the log likelihood function in vector notation as;
$\ell=\log L\left(\beta, \sigma^{2} ; y_{1}, y_{2}, \cdots, y_{N}\right)=-\frac{N}{2} \log 2 \pi-\frac{N}{2} \log \sigma^{2}-\frac{1}{2 \sigma^{2}}(y-X \beta)^{\prime}(y-X \beta)$
We will play with the vector notation since it is the common practice in econometrics. To get the MLE, we differentiate the log likelihood function with respect to parameters;

$$
\begin{aligned}
\frac{\partial \ell}{\partial \beta} & =-\frac{1}{2 \sigma^{2}}\left(-2 X^{\prime} y+2 X^{\prime} X \beta\right) \\
\frac{\partial \ell}{\partial \sigma^{2}} & =-\frac{N}{2 \sigma^{2}}+\frac{1}{2 \sigma^{4}}(y-X \beta)^{\prime}(y-X \beta)
\end{aligned}
$$

Setting the derivatives equal to zero, we have MLE.

$$
\begin{aligned}
\bar{\beta} & =\left(X^{\prime} X\right)^{-1} X^{\prime} y \\
\bar{\sigma}^{2} & =\frac{(y-X \bar{\beta})^{\prime}(y-X \bar{\beta})}{N}=\frac{e^{\prime} e}{N}
\end{aligned}
$$

(b) To get the asymptotic distribution of MLE. Recall that

$$
\sqrt{N}\left(\bar{\gamma}-\gamma_{0}\right) \xrightarrow{d} N\left(0, i\left(\gamma_{0}\right)^{-1}\right)
$$

when $\bar{\gamma}$ is the MLE of the true value $\gamma_{0} . i\left(\gamma_{0}\right)$ is the (expected) information matrix defined as

$$
i\left(\gamma_{0}\right)=E\left(-\frac{\partial^{2} \log p}{\partial \gamma^{2}}\right)
$$

remember that the information matrix is defined in terms of the second derivative of individual density. Going back to our case, we can conclude that

$$
\left[\begin{array}{c}
\sqrt{N}(\bar{\beta}-\beta) \\
\sqrt{N}\left(\bar{\sigma}^{2}-\sigma^{2}\right)
\end{array}\right] \xrightarrow{d} N\left(\mathbf{0}, i\left(\beta, \sigma^{2}\right)^{-1}\right)
$$

To determine the asymptotic variance matrix, we have to calculate the second derivative of the log likelihood function. One thing you should pay attention to is that the log likelihood function itself is something involved in the joint density but we need to calculate the second derivative of the individual likelihood. The saver lies in the fact that in case of i.i.d. the log likelihood for the whole observation is the sum of individual log likelihood;

$$
\ell=\log L\left(\beta, \sigma^{2} ; y_{1}, y_{2}, \cdots, y_{N}\right)=\sum_{i=1}^{N} \log p\left(y_{i} ; \beta, \sigma^{2}\right)
$$

Hence,

$$
\frac{\partial^{2} \ell}{\partial \beta \partial \beta^{\prime}}=\sum_{i=1}^{N} \frac{\partial^{2}}{\partial \beta \partial \beta^{\prime}} \log p\left(y_{i} ; \beta, \sigma^{2}\right)
$$

Then,

$$
E\left(-\frac{\partial^{2} \ell}{\partial \beta \partial \beta^{\prime}}\right)=E\left[\sum_{i=1}^{N}-\frac{\partial^{2}}{\partial \beta \partial \beta^{\prime}} \log p\left(y_{i} ; \beta, \sigma^{2}\right)\right]=N i\left(\beta, \sigma^{2}\right)
$$

Therefore, we can work with the log likelihood function to get the second order derivatives and divide by the sample size $N$ to get the correct estimator of the asymptotic variance matrix. Taking the second order derivative to get

$$
\begin{aligned}
\frac{\partial^{2} \ell}{\partial \beta \partial \beta^{\prime}} & =-\frac{1}{\sigma^{2}}\left(X^{\prime} X\right) \\
\frac{\partial^{2} \ell}{\partial \beta \partial \sigma^{2}} & =-\frac{1}{2 \sigma^{4}}\left(X^{\prime} y-X^{\prime} X \beta\right) \\
\frac{\partial^{2} \ell}{\partial\left(\sigma^{2}\right)^{2}} & =\frac{N}{2 \sigma^{4}}-\frac{1}{\sigma^{6}}(y-X \beta)^{\prime}(y-X \beta)
\end{aligned}
$$

Assigning minus sign and then taking expectations;

$$
\begin{aligned}
E\left(-\frac{\partial^{2} \ell}{\partial \beta \partial \beta^{\prime}}\right) & =\frac{1}{\sigma^{2}} E\left(X^{\prime} X\right)=\frac{1}{\sigma^{2}}\left(X^{\prime} X\right) \\
E\left(-\frac{\partial^{2} \ell}{\partial \beta \partial \sigma^{2}}\right) & =\frac{1}{2 \sigma^{4}}\left[X^{\prime} E(y)-X^{\prime} X \beta\right]=\frac{1}{2 \sigma^{4}}\left[X^{\prime} X \beta-X^{\prime} X \beta\right]=0 \\
E\left(-\frac{\partial^{2} \ell}{\partial\left(\sigma^{2}\right)^{2}}\right) & =-\frac{N}{2 \sigma^{4}}+\frac{1}{\sigma^{6}} E\left[(y-X \beta)^{\prime}(y-X \beta)\right] \\
& =-\frac{N}{2 \sigma^{4}}+\frac{1}{\sigma^{6}} E\left(\varepsilon^{\prime} \varepsilon\right)=-\frac{N}{2 \sigma^{4}}+\frac{N \sigma^{2}}{\sigma^{6}}=\frac{N}{2 \sigma^{4}}
\end{aligned}
$$

Then, the information matrix is given by

$$
i\left(\beta, \sigma^{2}\right)=\frac{1}{N}\left[\begin{array}{cc}
\frac{1}{\sigma^{2}}\left(X^{\prime} X\right) & \mathbf{0} \\
\mathbf{0} & \frac{N}{2 \sigma^{4}}
\end{array}\right]=\left[\begin{array}{cc}
\frac{1}{\sigma^{2}}\left(\frac{X^{\prime} X}{N}\right) & \mathbf{0} \\
\mathbf{0} & \frac{1}{2 \sigma^{4}}
\end{array}\right]
$$

Taking the inverse of the information matrix to get the asymptotic variance matrix

$$
i\left(\beta, \sigma^{2}\right)^{-1}=\left[\begin{array}{cc}
\sigma^{2}\left(\frac{X^{\prime} X}{N}\right)^{-1} & \mathbf{0} \\
\mathbf{0} & 2 \sigma^{4}
\end{array}\right]
$$

Therefore,

$$
\sqrt{N}\left[\begin{array}{c}
(\bar{\beta}-\beta) \\
\left(\bar{\sigma}^{2}-\sigma^{2}\right)
\end{array}\right] \xrightarrow{d} N\left(\mathbf{0},\left[\begin{array}{cc}
\sigma^{2}\left(\frac{X^{\prime} X}{N}\right)^{-1} & \mathbf{0} \\
\mathbf{0} & 2 \sigma^{4}
\end{array}\right]\right)
$$

In other notation, we may write as;

$$
\begin{aligned}
& \bar{\beta} \xrightarrow{d} N\left(\beta, \sigma^{2}\left(X^{\prime} X\right)^{-1}\right) \\
& \bar{\sigma}^{2} \xrightarrow{d} N\left(\sigma^{2}, \frac{2 \sigma^{4}}{N}\right)
\end{aligned}
$$

What do we know about the exact distribution of $\bar{\beta}$ ? $\bar{\beta}$ is equivalent to the least squares estimator $\widehat{\beta}$. Under the normality, we know that

$$
\widehat{\beta} \sim N\left(\beta, \sigma^{2}\left(X^{\prime} X\right)^{-1}\right)
$$

The asymptotics work!
(c) From (b);

$$
E\left(-\frac{\partial^{2} \ell}{\partial \beta \partial \beta^{\prime}}\right)=\frac{1}{\sigma^{2}} E\left(X^{\prime} X\right)=\frac{1}{\sigma^{2}}\left(X^{\prime} X\right)
$$

On the other hand,

$$
\begin{aligned}
E\left(\left[\frac{\partial \ell}{\partial \beta}\right]\left[\frac{\partial \ell}{\partial \beta}\right]^{\prime}\right) & =E\left(\left[-\frac{1}{2 \sigma^{2}}\left(-2 X^{\prime} y+2 X^{\prime} X \beta\right)\right]\left[-\frac{1}{2 \sigma^{2}}\left(-2 X^{\prime} y+2 X^{\prime} X \beta\right)\right]^{\prime}\right) \\
& =\frac{1}{\sigma^{4}} E\left(\left[X^{\prime} y-X^{\prime} X \beta\right]\left[X^{\prime} y-X^{\prime} X \beta\right]^{\prime}\right) \\
& =\frac{1}{\sigma^{4}} E\left[X^{\prime} \varepsilon \varepsilon^{\prime} X\right]=\frac{1}{\sigma^{2}}\left(X^{\prime} X\right)
\end{aligned}
$$

