ECON 620 Spring 2001 Midterm Solution

Professor Nicholas M. Kiefer, TA Nigar Hashimzade

Problem 1

Model:

$$y_i = x_i\beta + \epsilon_i$$

(no intercept), $i = 1, 2$, $\beta = 1$ is the true value, $Pr(\epsilon_i = -1) = Pr(\epsilon_i = 1) = 1/2$, ϵ_i are independent, $x_1 = 1$, $x_2 = 2$.

(a) OLS is BLUE if

- the model is correctly specified
- X's are non-stochastic
- ϵ_i are homoskedastic with mean zero.

Let us check the last requirement.

Mean:

$$E(\epsilon_i) = \frac{1}{2} \times (-1) + \frac{1}{2} \times 1 = 0 \text{ for } i = 1, 2$$

Variance:

$$V(\epsilon_i) = E\left[(\epsilon_i - E(\epsilon_i))^2\right] = \frac{1}{2} \times (-1)^2 + \frac{1}{2} \times 1^2 = 1 \text{ for } i = 1, 2$$

Covariance:

$$Cov(\epsilon_i, \epsilon_j) = E(\epsilon_i \epsilon_j) - E(\epsilon_i)E(\epsilon_j) = 0 \text{ for } i \neq j,\ i, j = 1, 2$$

(here we used independence of ϵ_i's). We showed that the error term is homoskedastic with mean zero. Hence, if X's are non-stochastic and the model is correctly specified, by Gauss-Markov theorem OLS is BLUE.

(b) For the simple model without intercept the OLS estimator of the slope:

$$\begin{aligned}\hat{\beta} &= \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\sum x_i(x_i\beta + \epsilon_i)}{\sum x_i^2} = \\ &= \beta + \frac{\sum x_i \epsilon_i}{\sum x_i^2} = 1 + \frac{1 \times \epsilon_1 + 2 \times \epsilon_2}{1^2 + 2^2}\end{aligned}$$

With the given distribution of ϵ_i we have the following exact distribution of $\hat{\beta}$:

Pr	1/4	1/4	1/4	1/4
$\hat{\beta}$	2/5	4/5	6/5	8/5

(c) The alternative estimator

$$\begin{aligned}\beta^* &= \frac{\sum y_i}{\sum x_i} = \frac{\sum(x_i\beta + \epsilon_i)}{\sum x_i} = \\ &= \beta + \frac{\sum \epsilon_i}{\sum x_i} = 1 + \frac{\epsilon_1 + \epsilon_2}{1 + 2}\end{aligned}$$

Hence, the exact distribution of β^* is

Pr	1/4	1/2	1/4
β^*	1/3	1	5/3

It is easy to see that β^* is unbiased:

$$E(\beta^*) = E\left(\beta + \frac{\sum \epsilon_i}{\sum x_i}\right) = \beta + \frac{\sum E(\epsilon_i)}{\sum x_i} = \beta$$

(d) The exact variances of the two estimators:

$$\begin{aligned}V(\hat{\beta}) &= V(\frac{\epsilon_1 + 2\,\epsilon_2}{5}) = \frac{1}{25}\,(V(\epsilon_1) + 4\,V(\epsilon_2)) = \frac{1}{25}(1+5) = \frac{1}{5} \\ V(\beta^*) &= V(\frac{\epsilon_1 + \epsilon_2}{3}) = \frac{1}{9}\,(V(\epsilon_1) + V(\epsilon_2)) = \frac{1}{9}(1+1) = \frac{2}{9}\end{aligned}$$

Hence, $V(\beta^*) > V(\hat{\beta})$, which is consistent with $\hat{\beta}$ being BLUE.

Problem 2

Model:

$$Y = X\beta + \epsilon$$

K regressors, N observations, all variables are measured in deviations from mean (no intercept), error term has normal distribution with mean zero and covariance matrix $\sigma^2 I$. We want to test $H_0 : \beta = 0$.

(a) To calculate the score statistic S, the likelihood ratio statistic LR, and the Wald statistic W for the hypothesis H_0 we need the likelihood function of the sample (x_i, y_i). Given the distribution of the error term, we have

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

Therefore, Y_i's are independent, and the joint density of the sample is the product of marginal densities of each observation:

$$\begin{aligned} f_{X,Y}(x, y|\beta, \sigma^2) &= \prod_{i=1}^{N} f_{X_i,Y_i}(x_i, y_i|\beta, \sigma^2) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp - \frac{(y_i - x_i'\beta)^2}{2\ \sigma^2} = \\ &= \left[\frac{1}{\sqrt{2\pi\sigma^2}}\right]^N \exp - \frac{\sum_{i=1}^{N}(y_i - x_i'\beta)^2}{2\ \sigma^2} \end{aligned}$$

Hence, the log-likelihood function of the sample is

$$\mathcal{L} = -\frac{N}{2} \ln 2\pi - \frac{N}{2}\sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{N} (y_i - x_i'\beta)^2$$

Without restriction, $\mathcal{L} = \mathcal{L}^U(\theta)$, where $\theta = (\beta,\ \sigma^2)'$ is $(K+1)$-vector of parameters. We will need the first and the second derivative of the log-likelihood function with respect to θ.

$$\frac{\partial \mathcal{L}^U}{\partial \theta} = \begin{pmatrix} \frac{\partial \mathcal{L}^U}{\partial \beta} \\ \frac{\partial \mathcal{L}^U}{\partial \sigma^2} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma^2} \sum x_i(y_i - x_i'\beta) \\ -\frac{N}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4} \sum (y_i - x_i'\beta)^2 \end{pmatrix}$$

$$\frac{\partial^2 \mathcal{L}^U}{\partial\theta\partial\theta'} = \begin{pmatrix} \frac{\partial^2\mathcal{L}^U}{\partial\beta\partial\beta'} & \frac{\partial^2\mathcal{L}^U}{\partial\beta\partial\sigma^2} \\ \frac{\partial^2\mathcal{L}^U}{\partial\beta'\partial\sigma^2} & \frac{\partial^2\mathcal{L}^U}{\partial(\sigma^2)^2} \end{pmatrix} = \begin{pmatrix} -\frac{1}{\sigma^2}\sum x_i x_i' & -\frac{1}{2\sigma^4}\sum x_i(y_i - x_i'\beta) \\ -\frac{1}{2\sigma^4}\sum x_i(y_i - x_i'\beta) & \frac{N}{2\sigma^4} - \frac{1}{\sigma^6}\sum(y_i - x_i'\beta)^2 \end{pmatrix}$$

In matrix notations,

$$\frac{\partial \mathcal{L}^U}{\partial\theta} = \begin{pmatrix} \frac{1}{\sigma^2}\left(X'Y - X'X\beta\right) \\ -\frac{N}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\left(Y - X\beta\right)\left(Y - X\beta\right)' \end{pmatrix} \tag{1}$$

$$\frac{\partial^2 \mathcal{L}^U}{\partial\theta\partial\theta'} = \begin{pmatrix} -\frac{1}{\sigma^2}X'X & -\frac{1}{2\sigma^4}\left(X'Y - X'X\beta\right) \\ -\frac{1}{2\sigma^4}\left(X'Y - X'X\beta\right) & \frac{N}{2\sigma^4} - \frac{1}{\sigma^6}\left(Y - X\beta\right)\left(Y - X\beta\right)' \end{pmatrix} \tag{2}$$

To find the maximum likelihood estimator $\hat{\theta}^U \equiv (\hat{\beta},\ \hat{\sigma}_U^2)'$ for θ in the unrestricted model we set $\frac{\partial \mathcal{L}^U}{\partial\theta} = 0$. From (1) we find

$$\begin{aligned} \hat{\beta} &= \left(X'X\right)^{-1} X'Y \\ \hat{\sigma}_U^2 &= \frac{1}{N} Y'MY \end{aligned}$$

where $M = I - X\left(X'X\right)^{-1} X'$ is a symmetric idempotent matrix.

To find the information matrix we compute $E\left[-\frac{\partial^2 \mathcal{L}^U}{\partial\theta\partial\theta'}\right]$. From (2), assuming X's are non-stochastic, we find:

$$\mathcal{I}(\theta) = \begin{pmatrix} \frac{1}{\sigma^2}X'X & 0 \\ 0 & \frac{N}{2\sigma^4} \end{pmatrix}$$

Here we used $Y - X\beta = \epsilon$ and $E(\epsilon) = 0,\ E(\epsilon\epsilon') = \sigma^2$. MLE for the information matrix in the unrestricted model:

$$\mathcal{I}\left(\hat{\theta}^U\right) = \begin{pmatrix} N\frac{X'X}{Y'MY} & 0 \\ 0 & \frac{N^3}{2(Y'MY)^2} \end{pmatrix}$$

Now we need the same for the restricted model. Under $H_0 : \beta = 0$ the parameter vector is $\theta = (0\ \sigma^2)$. MLE of σ^2 in the restricted model satisfies

$$\frac{\partial \mathcal{L}^R}{\partial\sigma^2} = -\frac{N}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\sum y_i^2 = -\frac{N}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}Y'Y = 0$$

Hence,

$$\begin{aligned}
\hat{\theta}^R &= \begin{pmatrix} 0 \\ \hat{\sigma}_R^2 \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{1}{N} Y'Y \end{pmatrix} \\
\mathcal{I}\left(\hat{\theta}^R\right) &= \begin{pmatrix} N\frac{X'X}{Y'Y} & 0 \\ 0 & \frac{N^3}{2(Y'Y)^2} \end{pmatrix}
\end{aligned}$$

Restriction $\beta = 0$ is equivalent to $g(\theta) = 0$, where $g(\theta) \equiv \beta$ is a K-vector, and, therefore, $\dfrac{\partial g}{\partial \theta} = (I_{(K\times K)}\ 0_{(K\times 1)})'$ is a $(K+1)\times K$ matrix.

Score statistic:

$$\begin{aligned}
S &= \frac{\partial \mathcal{L}^R}{\partial \theta'} \mathcal{I}^{-1}\left(\hat{\theta}^R\right) \frac{\partial \mathcal{L}^R}{\partial \theta} = \\
&= \left(N\frac{Y'X}{Y'Y}\ 0\right) \begin{pmatrix} \left(N\frac{X'X}{Y'Y}\right)^{-1} & 0 \\ 0 & \left(\frac{N^3}{2(Y'Y)^2}\right)^{-1} \end{pmatrix} \begin{pmatrix} N\frac{X'Y}{Y'Y} \\ 0 \end{pmatrix} = \\
&= N\frac{Y'X}{Y'Y} Y'Y \frac{(X'X)^{-1}}{N} N\frac{X'Y}{Y'Y} = \frac{N\, Y'(I-M)Y}{Y'Y}
\end{aligned}$$

Wald statistic:

$$\begin{aligned}
W &= g'(\hat{\theta}^U) \left[\frac{\partial g(\hat{\theta}^U)}{\partial \theta'} \mathcal{I}^{-1}\left(\hat{\theta}^U\right) \frac{\partial g'(\hat{\theta}^U)}{\partial \theta}\right]^{-1} g(\hat{\theta}^U) = \\
&= \hat{\beta}' \left[(I_{(K\times K)}\ 0_{(K\times 1)}) \begin{pmatrix} \left(N\frac{X'X}{Y'MY}\right)^{-1} & 0 \\ 0 & \left(\frac{N^3}{2(Y'MY)^2}\right)^{-1} \end{pmatrix} \begin{pmatrix} I_{(K\times K)} \\ 0_{(1\times K)}) \end{pmatrix}\right]^{-1} \hat{\beta} = \\
&= N\, \hat{\beta}' \frac{X'X}{Y'MY} \hat{\beta} = \frac{N\, Y'(I-M)Y}{Y'MY}
\end{aligned}$$

Likelihood ratio statistic:

$$\begin{aligned}
LR &= 2\left(\mathcal{L}(\hat{\theta}^U) - \mathcal{L}(\hat{\theta}^R)\right) = \\
&= 2\left[-\frac{N}{2}\ln 2\pi - \frac{N}{2}\ln \hat{\sigma}_U^2 - \frac{1}{2\hat{\sigma}_U^2}\left(Y - X\hat{\beta}\right)'\left(Y - X\hat{\beta}\right) + \right.
\end{aligned}$$

$$
\begin{aligned}
& + \frac{N}{2}\ln 2\pi + \frac{N}{2}\ln \hat{\sigma}_R^2 + \frac{1}{2\hat{\sigma}_R^2}Y'Y \Big] \\
= \; & N \ln \frac{\hat{\sigma}_R^2}{\hat{\sigma}_U^2} - \frac{Y'MY}{\hat{\sigma}_U^2} + \frac{Y'Y}{\hat{\sigma}_R^2} = N \ln \frac{Y'Y}{Y'MY} - N + N \\
= \; & N \ln \frac{Y'Y}{Y'MY}
\end{aligned}
$$

(b) Let us recall the definition of R^2:

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{Y'MY}{Y'Y}$$

when Y is measured in deviations from mean. It is straightforward to show that

$$
\begin{aligned}
S &= N\, R^2 \\
W &= N\, \frac{R^2}{1-R^2} \\
LR &= N\, \ln \frac{1}{1-R^2}
\end{aligned}
$$

(c) R^2 takes values between 0 and 1. Note that all three statistics are monotonically increasing functions of R^2 on the interval (0,1). Define

$$
\begin{aligned}
S(x) &= N\, x \\
LR(x) &= N\, \ln \frac{1}{1-x} \\
W(x) &= N\, \frac{x}{1-x}
\end{aligned}
$$

Then,

$$
\begin{aligned}
S'(x) &= N \\
LR'(x) &= N\, \frac{1}{1-x} \\
W'(x) &= N\, \frac{1}{(1-x)^2}
\end{aligned}
$$

Clearly, $S(x) = LR(x) = W(x)$ and $S'(x) = LR'(x) = W'(x)$ at $x = 0$, and $S'(x) <$

$LR'(x) < W'(x)$ for all $x \in (0,1)$. This implies that these three curves do not cross in the interval $(0,1)$, and the curve $S(x)$ is always below $LR(x)$, which, in its turn, is always below $W(x)$. Therefore,

$$S \leq LR \leq W$$

(d) Let us drop the normality assumption but still assume that observations are i.i.d. Let $f(y_i|x_i,\theta)$ be the density function of Y_i conditioned on $X_i = x_i$ and θ, a K-vector of unknown parameters of the distribution. By i.i.d. assumption, the log-likelihood function of the sample is

$$\mathcal{L}(\theta) = \sum_{i=1}^{N} \ln f(y_i|x_i,\theta)$$

Define the score function and information matrix as

$$\begin{aligned} \tilde{\mathcal{S}}(\theta) &= \frac{\partial \mathcal{L}}{\partial \theta} \\ \tilde{\mathcal{I}}(\theta) &= E\left[-\frac{\partial^2 \ln f}{\partial \theta \partial \theta'}\right] = E\left[\left(\frac{\partial \ln f}{\partial \theta}\frac{\partial \ln f}{\partial \theta'}\right)^2\right] \end{aligned}$$

Then, by Central Limit Theorem,

$$\frac{1}{\sqrt{N}}\tilde{\mathcal{S}}(\theta) \longrightarrow^d \mathcal{N}\left(0, \tilde{\mathcal{I}}(\theta)\right)$$

and

$$\tilde{\mathcal{S}}(\theta) \sim^a \mathcal{N}\left(0, N\tilde{\mathcal{I}}(\theta)\right)$$

(here, $\sim^a$ stands for "asymptotically distributed"). Therefore,

$$\frac{1}{N}\tilde{\mathcal{S}}'(\theta)\tilde{\mathcal{I}}^{-1}(\theta)\tilde{\mathcal{S}}(\theta) \sim^a \chi^2(K)$$

By the consistency of the maximum likelihood estimator ($p\lim\hat{\theta} = \theta$) and Slutsky's theorem, we have

$$\frac{1}{N}\tilde{\mathcal{S}}'(\hat{\theta}^R)\tilde{\mathcal{I}}^{-1}(\hat{\theta}^R)\tilde{\mathcal{S}}(\hat{\theta}^R) \sim^a \chi^2(K)$$

under $H_0 : \theta = \theta^R$. But the left-hand side is exactly the score test statistic S, since $\mathcal{I}(\theta) = N\tilde{\mathcal{I}}(\theta)$. Hence, score statistic is asymptotically distributed as $\chi^2(K)$. LR and W - statistics are asymptotically equivalent to S and also have asymptotic distribution of $\chi^2(K)$ under H_0.

(e) Note that there is no difference among these three statistics in terms of computation, since all of them can be expressed through R^2 from the unrestricted regression. Also, we can derive exact finite sample distribution under H_0 for each of these statistics after some trivial transformations, because under $H_0 : \beta = 0$, $Y \sim \mathcal{N}\left(0, \sigma^2\, I\right)$. Therefore,

$$\begin{aligned} \frac{1}{\sigma^2} Y'Y &\sim \chi^2(N) \\ \frac{1}{\sigma^2} Y'MY &\sim \chi^2(N-K) \\ \frac{1}{\sigma^2} Y'(I-M)Y &\sim \chi^2(K) \end{aligned}$$

under H_0, and we can further use the fact that $\dfrac{\chi^2(p)/p}{\chi^2(q)/q} \sim F_{p,q}$. Hence, in terms of finite sample inference neither of these three statistics is inferior to the other ones. If we drop the normality assumption, we cannot use the finite sample distribution, but, still, all three test statictics under H_0 are asymptotically distributed as $\chi^2(K)$, and, hence, are asymptotically equivalent. In this sense it is hard to justify why one of these tests should be preferred to others.