Final Examination (Fall 1996)
Economics 573
Suggested Answer Sheets

In what follows are the final exam questions (in bold) and suggested answers:

1. (20 points) Consider the linear regression model, with:

$$y = Xb + e \qquad \textbf{(1)}$$

where $E[e|X] = 0$ and $E[ee'|X] = s^2\Omega$ **. Let:**

$$\hat{b} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y \text{ .} \qquad \textbf{(2)}$$

denote the GLS estimator of b . Prove that any other linear unbiased estimator of b , say b^* , will have a variance which exceeds that of $\hat{b}$ by a positive definite matrix.

You could approach this proof in two ways. First, since b^* is linear, we can write it in the form:

$$\begin{aligned} b^* &= Ay \\ &= \left[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1} + C\right]y \end{aligned} \qquad \text{(A.1)}$$

where

$$C \equiv (X'\Omega^{-1}X)^{-1}X'\Omega^{-1} - A \qquad \text{(A.2)}$$

Notice that the GLS estimator corresponds to $C = \underset{K\times N}{0}$, so that:

$$\begin{aligned} b^* &= \hat{b} + Cy \\ &= \hat{b} + CXb + Ce \end{aligned} \qquad \text{(A.3)}$$

Since $\hat{b}$ is unbiased, b^* will be unbiased only if:

$$CX = 0 \text{ .} \qquad \text{(A.4)}$$

Thus

$$b^* - b = \left[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1} + C\right]e \qquad \text{(A.5)}$$

Finally, then, the covariance matrix of b^* is then given by:

$$
\begin{aligned}
Var\left[b^{*}\right] &= E\left\{\left(b^{*}-b\right)\left(b^{*}-b\right)'\right\} \\
&= E\left\{\left[\left(X'\Omega^{-1}X\right)^{-1}X'\Omega^{-1}+C\right]ee'\left[\Omega^{-1}X\left(X'\Omega^{-1}X\right)^{-1}+C'\right]\right\} \\
&= \left[\left(X'\Omega^{-1}X\right)^{-1}X'\Omega^{-1}+C\right]s^{2}\Omega\left[\Omega^{-1}X\left(X'\Omega^{-1}X\right)^{-1}+C'\right] \\
&= \left(X'\Omega^{-1}X\right)^{-1}X'\Omega^{-1}s^{2}\Omega\Omega^{-1}X\left(X'\Omega^{-1}X\right)^{-1}+Cs^{2}\Omega\Omega^{-1}X\left(X'\Omega^{-1}X\right)^{-1} \\
&\quad +\left(X'\Omega^{-1}X\right)^{-1}X'\Omega^{-1}s^{2}\Omega C'+Cs^{2}\Omega C' \\
&= s^{2}\left(X'\Omega^{-1}X\right)^{-1}+s^{2}CX\left(X'\Omega^{-1}X\right)^{-1}+s^{2}\left(X'\Omega^{-1}X\right)^{-1}X'C'+s^{2}C\Omega C' \\
&= s^{2}\left(X'\Omega^{-1}X\right)^{-1}+s^{2}C\Omega C' \\
&= Var\left[\hat{b}\right]++s^{2}C\Omega C'
\end{aligned}
\qquad \text{(A.5)}
$$

where the second to the last equality follows from (A.3).

An alternative approach to this problem would have been to note that we can define P such that:

$$P'P=\Omega^{-1}. \qquad \text{(A.6)}$$

Then premulitplying both sides of equation (1) by P yields:

$$
\begin{aligned}
&Py = PXb + Pe \\
&\text{or} \\
&\tilde{y} = \tilde{X}b + \tilde{e}
\end{aligned}
\qquad \text{(A.7)}
$$

where

$$
\begin{aligned}
&E\left[\tilde{e}\right] = PE\left[e\right] = 0 \\
&Var\left[\tilde{e}\tilde{e}'\right] = s^{2}I
\end{aligned}
\qquad \text{(A.8)}
$$

so that transformed model in satisfies the classical linear regression model assumptions used in the Gauss-Markov Theorem. In this case, the OLS estimator of b using the transformed model is *best* (i.e., minimum variance) among the class of unbiased estimators. But,

$$
\begin{aligned}
\tilde{b} &= \left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}\tilde{y} \\
&= \left(X'P'PX\right)^{-1}X'P'Py \\
&= \left(X'\Omega^{-1}X\right)^{-1}X'\Omega^{-1}y \\
&= \hat{b}
\end{aligned}
\qquad \text{(A.9)}
$$

Thus, the GLS estimator of b is also *best*.

2. (25 points) Consider the linear regression model, with:

$$y_i = bx_i + e_i \quad i = 1,\dots,N \tag{3}$$

where x_i **is a scalar and** $E\left[e_i \middle| x_i\right] = 0,\ E\left[e_i^2 \middle| x_i\right] = s_i^2$, and $E\left[e_i e_j \middle| X\right] = 0$. **Under what conditions (i.e., specifications for** s_i^2**) will the GLS estimator of** b **be given by:**

i. $\hat{b} = \dfrac{\sum_{i=1}^{N} x_i^2 y_i}{\sum_{i=1}^{N} x_i^3}$

ii. $\hat{b} = \frac{1}{N}\sum_{i=1}^{N} \dfrac{y_i}{x_i}$

iii. $\hat{b} = \dfrac{\sum_{i=1}^{N} y_i}{\sum_{i=1}^{N} x_i}$

iv. $\hat{b} = \dfrac{\sum_{i=1}^{N} x_i y_i}{\sum_{i=1}^{N} x_i^2}$

The first thing to note about this problem is that the model is one with heteroskedasticity. As we discussed in class, the GLS estimator in this case is given by:

$$\begin{aligned}\hat{b} &= \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}y \\ &= \left[\sum_{i=1}^{N} \frac{x_i x_i'}{s_i^2}\right]^{-1} \sum_{i=1}^{N} \frac{x_i y_i}{s_i^2}\end{aligned} \tag{A.10}$$

For the case in which x_i is a scalar, this reduces to:

$$\hat{b} = \frac{\sum_{i=1}^{N} \frac{x_i y_i}{s_i^2}}{\sum_{i=1}^{N} \frac{x_i^2}{s_i^2}} \tag{A.11}$$

We can now answer the questions posed above.

i. For this question, if $s_i^2 = \frac{s^2}{x_i}$, then (A.11) yields

$$\hat{b} = \frac{\sum_{i=1}^{N} \frac{x_i y_i}{s^2 x_i^{-1}}}{\sum_{i=1}^{N} \frac{x_i^2}{s^2 x_i^{-1}}} = \frac{\sum_{i=1}^{N} x_i^2 y_i}{\sum_{i=1}^{N} x_i^3} \tag{A.12}$$

ii. For this question, if $s_i^2 = s^2 x_i^2$, then (A.11) yields

$$\hat{b} = \frac{\sum_{i=1}^{N} \frac{x_i y_i}{s^2 x_i^2}}{\sum_{i=1}^{N} \frac{x_i^2}{s^2 x_i^2}} = \frac{\sum_{i=1}^{N} \frac{y_i}{x_i}}{\sum_{i=1}^{N} 1} = \frac{1}{N} \sum_{i=1}^{N} \frac{y_i}{x_i} \tag{A.13}$$

iii. For this question, if $s_i^2 = s^2 x_i$, then (A.11) yields

$$\hat{b} = \frac{\sum_{i=1}^{N} \frac{x_i y_i}{s^2 x_i}}{\sum_{i=1}^{N} \frac{x_i^2}{s^2 x_i}} = \frac{\sum_{i=1}^{N} y_i}{\sum_{i=1}^{N} x_i} \tag{A.14}$$

iv. For this question, if $s_i^2 = s^2$, then (A.11) yields

$$\hat{b} = \frac{\sum_{i=1}^{N} \frac{x_i y_i}{s^2}}{\sum_{i=1}^{N} \frac{x_i^2}{s^2}} = \frac{\sum_{i=1}^{N} x_i y_i}{\sum_{i=1}^{N} x_i^2} \tag{A.15}$$

Notice that in each of these cases, s^2 cancels out of the numerator and denominator of each estimator.

3. (20 points) You are given the following two models of an individual's income. In the first model,

$$y_i = a + b_E E_i + e_i \tag{4}$$

where y_i denotes the individual's annual income and

$$E_i = \begin{cases} 3 & \text{individual } i\text{'s highest degree is a Ph.D. or M.D.} \\ 2 & \text{individual } i\text{'s highest degree is a Masters degree} \\ 1 & \text{individual } i\text{'s highest degree is a Bachelors.} \\ 0 & \text{otherwise} \end{cases} \quad \textbf{(5)}$$

indicates the highest degree achieved by the individual. In the second model, we have:

$$y_i = \alpha + \beta_{BS} D_{BS,i} + \beta_{MS} D_{MS,i} + \beta_{PhD} D_{PhD,i} + \varepsilon_i \quad \textbf{(6)}$$

where

$$D_{BS,i} = \begin{cases} 1 & \text{individual } i\text{'s highest degree is a Bachelors} \\ 0 & \text{otherwise} \end{cases} \quad \textbf{(7)}$$

$$D_{MS,i} = \begin{cases} 1 & \text{individual } i\text{'s highest degree is a Masters} \\ 0 & \text{otherwise} \end{cases} \quad \textbf{(8)}$$

$$D_{PhD,i} = \begin{cases} 1 & \text{individual } i\text{'s highest degree is a Ph.D. or M.D.} \\ 0 & \text{otherwise} \end{cases} \quad \textbf{(9)}$$

How would you test model (4) as a restriction on model (6)? In particular, specify the hypothesis that you would be making and the test statistic you would employ.

The hypothesis of interest corresponds to:

$$H_0: \beta_{PhD} = 3\beta_{BS} \text{ and } \beta_{MS} = 2\beta_{BS} \quad \text{(A.16)}$$

or, equivalently:

$$H_0: \frac{1}{3}\beta_{PhD} = \frac{1}{2}\beta_{MS} = \beta_{BS} \quad \text{(A.17)}$$

Substituting this restriction into model (6) yields:

$$\begin{aligned} y_i &= \alpha + \beta_{BS} D_{BS,i} + 2\beta_{BS} D_{MS,i} + 3\beta_{BS} D_{PhD,i} + \varepsilon_i \\ &= \alpha + \beta_{BS}\left(D_{BS,i} + 2D_{MS,i} + 3D_{PhD,i}\right) + \varepsilon_i \\ &= \alpha + \beta_{BS} E_i + \varepsilon_i \end{aligned} \quad \text{(A.18)}$$

where β_{BS} would correspond to β_E in equation (4). Note that there are only two restrictions being imposed, as (A.16) makes clear. We could test this restriction using

- the standard F-test, distributed F with [2,N-4] degrees of freedom, or

- a Wald test, distributed c_2^2.

4. **(35 points) You are interested in modeling the electricity usage of households in Iowa. The local utility can provide you with two data sets to help you in your analysis. The first data set consists of 500 households randomly selected from their *low usage* customers (i.e., with annual usage less than 6,000 kWh). The second data set consists of 500 households randomly selected from their *high usage* customers (i.e., with annual usage greater than 20,000 kWh). For each household, you are provided with data on their:**

   - **annual usage (K_i),**
   - **annual household income (I_i), and**
   - **the number of household members (M_i).**

   You hypothesize that, in the population as a whole, annual household usage follows the linear regression model:

$$K_i = a + bI_i + gM_i + e_i \qquad \textbf{(10)}$$

   a. **Detail how you would estimate your model, including a complete specification of the associated likelihood function you would employ and any additional assumptions you would need.**

There are two problems with the available data sets. First, we have no observations from the middle of the distribution of usage. Instead, the two sets combined represent a truncated data set, with the middle of the sample (i.e., $6{,}000 \le K_i \le 20{,}000$ missing. There is a second problem with this data set, namely that we are not given a random sample from the combined truncated regions (i.e., $K_i \le 6{,}000$ and $20{,}000 \le K_i$). Instead we are given 500 observations from each of the upper and lower tails of the usage distribution, respectively. For now, lets ignore this second problem and assume that the low and high usage groups are roughly the same size in the population. What we have then is a truncated sample, with observations only in the tails of the distribution. In general, we can write the truncated distribution as:

$$\begin{aligned} f_K\left(K_i \middle| K_i \le 6{,}000 \text{ or } K_i \ge 20{,}000\right) &= f_e\left(e_i \middle| e_i \le [6{,}000 - a - bI_i - gM_i] \text{ or } e_i \ge [20{,}000 - a - bI_i - gM_i]\right) \\ &= \frac{f_e(e_i)}{\Pr\left[e_i \le [6{,}000 - a - bI_i - gM_i] \text{ or } e_i \ge [20{,}000 - a - bI_i - gM_i]\right]} \qquad \text{(A.19)} \\ &= \frac{f_e(e_i)}{1 - \Pr\left([6{,}000 - a - bI_i - gM_i] < e_i < [20{,}000 - a - bI_i - gM_i]\right)} \end{aligned}$$

The first equality comes from the fact that K_i and e_i are linearly related through equation (10). In order to estimate the model in question, we will have to make some assumption about the error distribution. If we assume that $e_i \sim iidN(0, s^2)$, then

$$f_e(e_i) = \frac{1}{\sigma}\phi\left(\frac{e_i}{\sigma}\right) \tag{A.20}$$

where $\phi(\cdot)$ denotes the standard normal pdf and

$$\begin{aligned}\Pr\left(\left[6{,}000-\alpha-\beta I_i-\gamma M_i\right] < e_i < \left[20{,}000-\alpha-\beta I_i-\gamma M_i\right]\right) \\ = \Phi\left(\frac{20{,}000-\alpha-\beta I_i-\gamma M_i}{\sigma}\right) - \Phi\left(\frac{6{,}000-\alpha-\beta I_i-\gamma M_i}{\sigma}\right)\end{aligned} \tag{A.21}$$

Substituting these into equation (A.19) yields:

$$\begin{aligned} f_K\left(K_i \middle| K_i \le 6{,}000 \text{ or } K_i \ge 20{,}000\right) \\ = \frac{\frac{1}{\sigma}\phi\left(\frac{K_i-\alpha-\beta I_i-\gamma M_i}{\sigma}\right)}{1-\left[\Phi\left(\frac{20{,}000-\alpha-\beta I_i-\gamma M_i}{\sigma}\right)-\Phi\left(\frac{6{,}000-\alpha-\beta I_i-\gamma M_i}{\sigma}\right)\right]}\end{aligned} \tag{A.22}$$

The corresponding log-likelihood function becomes:

$$\begin{aligned} l(K,I,M;\alpha,\beta,\gamma,\sigma) = -\frac{N}{2}\ln(2\pi) - N\ln(\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}\left(K_i-\alpha-\beta I_i-\gamma M_i\right)^2 \\ + \sum_{i=1}^{N}\ln\left(1-\left[\Phi\left(\frac{20{,}000-\alpha-\beta I_i-\gamma M_i}{\sigma}\right)-\Phi\left(\frac{6{,}000-\alpha-\beta I_i-\gamma M_i}{\sigma}\right)\right]\right)\end{aligned} \tag{A.23}$$

where the last term in the likelihood function controls for the truncation in the data base. This last term is analogous to what we saw in class when the tails of the distribution were truncated. In this case, it is the center of the distribution that has been truncated.

The second problem that I mentioned early on is that we do not really have a random sample from the tails of the distribution of K_i. Instead, we have 500 observations from the lower tail and 500 observations from the upper tail. To see that these are not the same, suppose that in the original (untruncated) population, there were 10,000 households with annual usage below 6,000 kWh and only 1,000 households with annual usage above 20,000. Thus, there would be 11,000 households in the population meeting the conditions: $K_i \le 6{,}000$ and $20{,}000 \le K_i$. A random sample from these 11,000 households would, on average, yield 91 households from high usage group and 909 from the lower usage households. This is because the chance of randomly drawing a high usage household from the truncated sample should be:

$$P_H = \frac{\Pr\left[K_i \ge 20{,}000\right]}{\Pr\left[K_i \le 6{,}000\right]+\Pr\left[K_i \ge 20{,}000\right]} = \frac{1{,}000}{11{,}000} = .091 \tag{A.24}$$

We can account for non-random nature of the sampling to reflect the population proportions. Suppose that the first N_N observations consist of the high usage observations. The adjusted log-likelihood function would then become:

$$
\begin{aligned}
l_{adj}(K, I, M; a, b, g, s) = \frac{.5}{P_H}\Bigg\{ & -\frac{N_H}{2}\ln(2p) - N_H \ln(s) - \frac{1}{2s^2}\sum_{i=1}^{N_H}\left(K_i - a - bI_i - gM_i\right)^2 \\
& + \sum_{i=1}^{N_H}\ln\left[1 - \left\{\Phi\left(\frac{20{,}000 - a - bI_i - gM_i}{s}\right) - \Phi\left(\frac{6{,}000 - a - bI_i - gM_i}{s}\right)\right\}\right]\Bigg\} \\
+ \left(\frac{.5}{1 - P_H}\right)\Bigg\{ & -\frac{N_L}{2}\ln(2p) - N_L \ln(s) - \frac{1}{2s^2}\sum_{i=N_H+1}^{N}\left(K_i - a - bI_i - gM_i\right)^2 \\
& + \sum_{i=N_H+1}^{N}\ln\left[1 - \left\{\Phi\left(\frac{20{,}000 - a - bI_i - gM_i}{s}\right) - \Phi\left(\frac{6{,}000 - a - bI_i - gM_i}{s}\right)\right\}\right]\Bigg\}
\end{aligned}
\tag{A.23}
$$

where $N_L = N - N_H$.

Note: In grading the exam, I did not deduct points for failing to specify the above likelihood function, but I did take a couple of points off if you failed to notice the second problem at all.

b. How would you assess the fit of your model. Be specific about your measure of fit, including its advantages and limitations.

There is no perfect answer in this case. One obvious suggestion would be to use a pseudo-R^2 statistic, similar to McFadden's statistic in the case of the probit model. That is:

$$LRI = 1 - \frac{l}{l_0} \tag{10.30}$$

where l_0 denotes the log-likelihood value if we constrain $b = g = 0$ in our model. If the model fits perfectly, then all of the probabilities would become 1 and $l = 0$, with LRI=1, whereas if the model did no better than one with only the intercept included, then $\frac{l}{l_0} = 1$, so that LRI=0. The problem with this measure is that it is difficult to interpret the intermediate values of LRI. What is a "good fit?" and how much worse is LRI = .5 from an LRI = .6?