APPENDIX

PART 1

A. Generalized Inverse

Definition 1: Let $\mathbf{A}$ be an $m \times n$ matrix. Then a matrix $\mathbf{A}^{-}$: $n \times m$ is said to be a generalized inverse of $\mathbf{A}$ if

$$\mathbf{A}\mathbf{A}^{-}\mathbf{A} = \mathbf{A}$$

holds (see Rao and Toutenburg (1999), p.372).

A generalized inverse always exists although it is not unique in general.

Definition 2: (Moore-Penrose) A matrix $\mathbf{A}^{+}$ satisfying the following conditions is called the Moore-Penrose inverse of $\mathbf{A}$:

(i) $\mathbf{A}\mathbf{A}^{+}\mathbf{A} = \mathbf{A}$,

(ii) $\mathbf{A}^{+}\mathbf{A}\mathbf{A}^{+} = \mathbf{A}^{+}$,

(iii) $\left(\mathbf{A}^{+}\mathbf{A}\right)' = \mathbf{A}^{+}\mathbf{A}$,

(iv) $\left(\mathbf{A}\mathbf{A}^{+}\right)' = \mathbf{A}\mathbf{A}^{+}$.

$\mathbf{A}$ is unique.

B. The Augmented Model

Let the $\mathbf{X}$ matrix and the observation vector $\mathbf{Y}$ be augmented by $\sqrt{k}\mathbf{I}_p$ and $\mathbf{Y}_A$ respectively (subscript "*A*" denoting augmentation). The model will then take the form,

$$\begin{bmatrix} \mathbf{Y}_X \\ \cdots \\ \mathbf{Y}_A \end{bmatrix} = \begin{bmatrix} \mathbf{X} \\ \cdots \\ k^{1/2}\mathbf{I}_p \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \end{bmatrix} + \mathbf{U}, \tag{A.1}$$

where $\mathbf{Y}_x$ is the same as original $\mathbf{Y}$, $\mathbf{Y}_A$ is a $p \times 1$ observation vector corresponding to the augmented part, $\mathbf{I}_p$ is a $p \times p$ identity matrix, and $\mathbf{U}$ is $(n+p)\times 1$ error vector. In this augmented model, we have $E(\mathbf{Y}_X) = \mathbf{X}\boldsymbol{\beta}$ and $E(\mathbf{Y}_A) = \sqrt{k}\boldsymbol{\beta}$. The (unbiased) least squares estimates of $\boldsymbol{\beta}$ in the augmented model are given by

$$\begin{aligned} \hat{\boldsymbol{\beta}}_A &= (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}(\mathbf{X}'\mathbf{Y} + \sqrt{k}\mathbf{Y}_A) \\ &= \hat{\boldsymbol{\beta}}^* + \sqrt{k}(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{Y}_A . \end{aligned}$$

One might say that we use, in fact, the biased estimator $\hat{\boldsymbol{\beta}}^*$ in place of the unbiased estimator $\hat{\boldsymbol{\beta}}_A$, and not in place of $\hat{\boldsymbol{\beta}}$, and that in using $\hat{\boldsymbol{\beta}}^*$, the part, $\Delta = \sqrt{k}(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{Y}_A$, is omitted from the estimation procedure. The bias in estimation will therefore come from this omitted part. Thus, if an unbiased estimator was to be used at all, it would be $\hat{\boldsymbol{\beta}}_A$ and not $\hat{\boldsymbol{\beta}}$. So if $\hat{\boldsymbol{\beta}}_A$ is adopted as the unbiased estimator, the mean squared error of the biased estimator shall be compared with the variance of $\hat{\boldsymbol{\beta}}_A$.

Hoerl and Kennard have shown that the wquared bias of $\hat{\boldsymbol{\beta}}^*$ is given by

$$k^2\boldsymbol{\beta}'(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-2}\boldsymbol{\beta}. \tag{A.2}$$

On the other hand we have

$$\begin{aligned} E(\boldsymbol{\Delta}) &= E\left[\sqrt{k}(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{Y}_A\right] \\ &= k(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\boldsymbol{\beta}. \end{aligned} \tag{A.3}$$

Squaring (A.3), we have $\{E(\boldsymbol{\Delta})\}^2 = k^2\boldsymbol{\beta}'(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-2}\boldsymbol{\beta}$, which is the same as (A.2).

A more general model than (A.1) could also be considered. Thus, rather than considering the additional data $(\mathbf{Y}_A, k\mathbf{I}_p)$, we might consider the data $(\mathbf{Y}_A, \mathbf{V})$, where $\mathbf{V}'\mathbf{V} = \mathbf{K}$ is a diagonal matrix with diagonal elements k_i. However, we are considering model (A.1) in view of the following reasons:

(i) Hoerl and Kennard ultimately thought in terms of one *k*, and not in terms of k_i.

(ii) Model (A.1) will show how little of the observed **Y** (if observable) is being discarded to obtain the biased estimator.

C. Influence Analysis

The usual multiple regression model can be defined as

$\mathbf{Y} = \mathbf{1}\beta_0 + \mathbf{X}\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}$, where $\mathbf{Y}$ is an n vector of observable random variables, $\mathbf{X}$ is an $n \times r$ centred and standardized matrix of known constants, β_0 is an unknown parameter, $\boldsymbol{\beta}_1$ is an r vector of unknown parameters and $\boldsymbol{\varepsilon}$ is an n vector of unobservable disturbances.

If $\mathbf{Z} = (\mathbf{1}, \mathbf{X})$ then the LS estimator is $\mathbf{b} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}$ and the vector of fitted responses $\hat{\mathbf{Y}} = \mathbf{Z}\mathbf{b}$. The estimator of σ^2 is $s^2 = \mathbf{e}'\mathbf{e}/(n-p)$, where $\boldsymbol{e}$ is the vector of residuals.

A particularly appealing perturbation scheme is case deletion. The influence of a case can be viewed as the product of two factors, the first a function of the residual and the second a function of the position of the point in the *Z* space. The position or leverage of the *i*th point is measured by h_i, the *i*th diagonal element of the "hat" matrix $\mathbf{H} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$.

Among the most popular single-case influence measures is the difference in fit standardized (*DFFITS*), which evaluated at the *i*th case is given by

$$DFFITS(i) = z_i\,(\mathbf{b} - \mathbf{b}(i))/SE(z_i\mathbf{b}), \tag{A.4}$$

where $\mathbf{b}(i)$ is the LS estimator of $\boldsymbol{\beta}$ without the *i*th case and $SE(z_i\mathbf{b})$ is an estimator of the standard error (SE) of the fitted value.

DFFITS is the standardized change in the fitted value of a case when it is deleted. Thus it can be considered a measure of influence on individual fitted values. *DFFITS* can be written as the product of two factors, one depending on the residual and the other depending on leverage,

$$DFFITS(i) = \left[\frac{e_i}{s(i)}\right]\left[\frac{h_i^{1/2}}{(1-h_i)}\right], \tag{A.5}$$

where $s(i)$ is the LS estimator of σ when the *i*th case has been deleted, e_i is the *i*th residual, and h_i is the leverage of the point.

Another useful measure of influence is Cook's *D*, which evaluated at the *i*th case is given by

$$D_i = \frac{(\mathbf{b} - \mathbf{b}(i))' \mathbf{Z}'\mathbf{Z}(\mathbf{b} - \mathbf{b}(i))}{ps^2}. \tag{A.6}$$

D_i is a measure of the change in all of the fitted values when a case is deleted. It can also be written as

$$D_i = \frac{e_i^2}{ps^2} \frac{h_i}{(1 - h_i^2)}. \tag{A.7}$$

To determine influential cases, Cook and Weisberg suggested that D_i to be compared with an $F(p, n-p)$ distribution.

These measures are useful for detecting single cases having an unduly high influence. However, they suffer from the problem of masking- that is, the presence of cases that can disguise or mask the potential influence of other cases (Walker and Birch, 1988).

PART 2

Table 1 Longley data

PEOPLE EMPLOYED	GNP DEFLATOR	GNP	UNEMPLOYED	ARMED FORCES	POPULATION	YEAR
60,323	83.0	234,289	2,356	1,590	107,608	1947
61,122	88.5	259,426	2,325	1,456	108,632	1948
60,171	88.2	258,054	3,682	1,616	109,773	1949
61,187	89.5	284,599	3,351	1,650	110,929	1950
63,221	96.2	328,975	2,099	3,099	112,075	1951
63,639	98.1	346,999	1,932	3,594	113,270	1952
64,989	99.0	365,385	1,870	3,547	115,094	1953
63,761	100.0	363,112	3,578	3,350	116,219	1954
66,019	101.2	397,469	2,904	3,048	117,388	1955
67,857	104.6	419,180	2,822	2,857	118,734	1956
68,169	108.4	442,769	2,936	2,798	120,445	1957
66,513	110.8	444,546	4,681	2,637	121,950	1958
68,655	112.6	482,704	3,813	2,552	123,366	1959
69,564	114.2	502,601	3,931	2,514	125,368	1960
69,331	115.7	518,173	4,806	2,572	127,852	1961
70,551	116.9	554,894	4,007	2,827	130,081	1962

Source: J. Longley (1967) "An Appraisal of Least Squares Programs for the Electronic Computer from the Point of View of the User", *Journal of the American Statistical Association*, vol. 62. September, pp. 819-841

TABLE 2: RESULTS OF THE SIMULATION

CASE 1 α and α_* equal to 0.99

	$\boldsymbol{\beta} = \boldsymbol{\beta}_S$				$\boldsymbol{\beta} = \boldsymbol{\beta}_L$			
Signal-to-noise ratio, ρ	100	4	1	0.04	100	4	1	0.04
HK								
Ratio of total mean square errors	*0.999*	*0.991*	*0.973*	*0.916*	*0.866*	*1.069*	*0.976*	*0.916*
k values	0.0004	0.0070	0.0108	0.0146	0.0004	0.0160	0.0132	0.0146
St.deviation of k	(0.0001)	(0.0033)	(0.0106)	(0.0212)	(0.0002)	(0.0533)	(0.0287)	(0.0212)
HKB								
Ratio of total mean square errors	*0.999*	*0.987*	*0.958*	*0.859*	*0.599*	*1.163*	*0.990*	*0.860*
k values	0.0016	0.0180	0.0270	0.0423	0.0019	0.0410	0.0366	0.0438
St.deviation of k	(0.0005)	(0.0095)	(0.0277)	(0.0556)	(0.0012)	(0.1140)	(0.0735)	(0.0593)
LW								
Ratio of total mean square errors	*0.999*	*0.987*	*0.946*	*0.804*	*3.472*	*1.444*	*1.110*	*0.796*
k values	0.0004	0.0098	0.0368	0.7048	0.1816	1.869	1.4367	1.5727
St.deviation of k	(0.0011)	(0.0027)	(0.0142)	(0.6996)	(0.0811)	(5.7995)	(2.3857)	(1.9047)

CASE 2 α equal to 0.99, α_* equal to 0.10

	$\boldsymbol{\beta} = \boldsymbol{\beta}_S$				$\boldsymbol{\beta} = \boldsymbol{\beta}_L$			
Signal-to-noise ratio, ρ	100	4	1	0.04	100	4	1	0.04
HK								
Ratio of total mean square errors	*1.000*	*0.997*	*0.986*	*0.953*	*0.939*	*1.202*	*1.053*	*1.002*
k values	0.0003	0.0074	0.0191	0.0408	0.0003	0.0199	0.0358	0.0416
St.deviation of k	(0.0001)	(0.0027)	(0.0134)	(0.0656)	(0.0001)	(0.0495)	(0.0721)	(0.0799)

HKB								
Ratio of total mean square errors	*1.000*	*0.996*	*0.979*	*0.923*	*0.941*	*1.481*	*1.126*	*1.013*
k values	0.0015	0.0263	0.0584	0.1278	0.0017	0.0731	0.1224	0.1332
St.deviation of k	(0.0004)	(0.0116)	(0.0446)	(0.1731)	(0.0007)	(0.1758)	(0.2127)	(0.1909)
LW								
Ratio of total mean square errors	*1.000*	*0.996*	*0.9741*	*0.908*	*22.316*	*2.078*	*1.282*	*1.025*
k values	0.0005	0.0127	0.0527	0.9082	0.0851	1.0525	1.5282	1.5738
St.deviation of k	(0.0001)	(0.0041)	(0.0194)	(0.7026)	(0.0363)	(0.9687)	(2.0644)	(1.5073)

CASE 3 α equal to 0.70, α_* equal to 0.30

	$\boldsymbol{\beta} = \boldsymbol{\beta}_S$				$\boldsymbol{\beta} = \boldsymbol{\beta}_L$			
Signal-to-noise ratio, ρ	100	4	1	0.04	100	4	1	0.04
HK								
Ratio of total mean square errors	*1.000*	*1.000*	*1.002*	*1.004*	*1.000*	*1.020*	*1.040*	*1.013*
k values	0.0003	0.0073	0.0278	0.3219	0.0003	0.0081	0.0503	0.4112
St.deviation of k	(0.0001)	(0.0022)	(0.0098)	(0.2878)	(0.0001)	(0.0033)	(0.1256)	(0.5522)
HKB								
Ratio of total mean square errors	*1.000*	*1.004*	*1.011*	*1.010*	*1.000*	*1.088*	*1.125*	*1.026*
k values	0.0014	0.0348	0.1149	0.8411	0.0014	0.0388	0.1892	1.0878
St.deviation of k	(0.0004)	(0.0104)	(0.0412)	(0.7317)	(0.0004)	(0.0158)	(0.3746)	(1.3966)
LW								
Ratio of total mean square errors	*1.000*	*1.002*	*1.006*	*1.010*	*1.022*	*1.242*	*1.236*	*1.035*
k values	0.0007	0.0180	0.0655	0.9934	0.0057	0.1420	0.4779	1.4743
St.deviation of k	(0.0002)	(0.0053)	(0.0229)	(1.2711)	(0.0016)	(0.0555)	(0.5317)	(1.5955)