Solutions Manual

Econometric Analysis

Fifth Edition

William H. Greene

New York University

Prentice Hall, Upper Saddle River, New Jersey 07458

Contents and Notation

In the solutions, we denote:

- scalar values with italic, lower case letters, as in a or α
- column vectors with boldface lower case letters, as in **b**,
- row vectors as transposed column vectors, as in $\mathbf{b}'$,
- single population parameters with greek letters, as in β,
- sample estimates of parameters with English letters, as in **b** as an estimate of β,
- sample estimates of population parameters with a caret, as in $\hat{\alpha}$
- matrices with boldface upper case letters, as in **M** or Σ,
- cross section observations with subscript i, time series observations with subscript t.

These are consistent with the notation used in the text.

Chapter 1

Introduction

There are no exercises in Chapter 1.

Chapter 2

The Classical Multiple Linear Regression Model

There are no exercises in Chapter 2.

Chapter 3

Least Squares

1. (a) Let $X = \begin{bmatrix} 1 & x_1 \\ . & . \\ 1 & x_n \end{bmatrix}$. The normal equations are given by (3-12), $\mathbf{X}'\mathbf{e} = \mathbf{0}$, hence for each of the columns of **X**, $\mathbf{x_k}$, we know that $\mathbf{x_k}$'e=0. This implies that $\sum_i e_i = 0$ and $\sum_i x_i e_i = 0$.

(b) Use $\sum_i e_i = 0$ to conclude from the first normal equation that $a = \bar{y} - b\bar{x}$.

(c) Know that $\sum_i e_i = 0$ and $\sum_i x_i e_i = 0$. It follows then that $\sum_i (x_i - \bar{x})e_i = 0$. Further, the latter implies $\sum_i (x_i - \bar{x})(y_i - a - bx_i) = 0$ or $\sum_i (x_i - \bar{x})(y_i - \bar{y} - b(x_i - \bar{x})) = 0$ from which the result follows.

2. Suppose **b** is the least squares coefficient vector in the regression of **y** on **X** and **c** is any other *K*x1 vector. Prove that the difference in the two sums of squared residuals is

$$(\mathbf{y}-\mathbf{Xc})'(\mathbf{y}-\mathbf{Xc}) - (\mathbf{y}-\mathbf{Xb})'(\mathbf{y}-\mathbf{Xb}) = (\mathbf{c}-\mathbf{b})'\mathbf{X}'\mathbf{X}(\mathbf{c}-\mathbf{b}).$$

Prove that this difference is positive.

Write **c** as **b** + (**c** - **b**). Then, the sum of squared residuals based on **c** is

$$(\mathbf{y} - \mathbf{Xc})'(\mathbf{y} - \mathbf{Xc}) = [\mathbf{y} - \mathbf{X}(\mathbf{b} + (\mathbf{c} - \mathbf{b}))]'[\mathbf{y} - \mathbf{X}(\mathbf{b} + (\mathbf{c} - \mathbf{b}))] = [(\mathbf{y} - \mathbf{Xb}) + \mathbf{X}(\mathbf{c} - \mathbf{b})]'[(\mathbf{y} - \mathbf{Xb}) + \mathbf{X}(\mathbf{c} - \mathbf{b})]$$

$$= (\mathbf{y} - \mathbf{Xb})'(\mathbf{y} - \mathbf{Xb}) + (\mathbf{c} - \mathbf{b})'\mathbf{X}'\mathbf{X}(\mathbf{c} - \mathbf{b}) + 2(\mathbf{c} - \mathbf{b})'\mathbf{X}'(\mathbf{y} - \mathbf{Xb}).$$

But, the third term is zero, as $2(\mathbf{c} - \mathbf{b})'\mathbf{X}'(\mathbf{y} - \mathbf{Xb}) = 2(\mathbf{c} - \mathbf{b})\mathbf{X}'\mathbf{e} = \mathbf{0}$. Therefore,

$$(\mathbf{y} - \mathbf{Xc})'(\mathbf{y} - \mathbf{Xc}) = \mathbf{e}'\mathbf{e} + (\mathbf{c} - \mathbf{b})'\mathbf{X}'\mathbf{X}(\mathbf{c} - \mathbf{b})$$

or

$$(\mathbf{y} - \mathbf{Xc})'(\mathbf{y} - \mathbf{Xc}) - \mathbf{e}'\mathbf{e} = (\mathbf{c} - \mathbf{b})'\mathbf{X}'\mathbf{X}(\mathbf{c} - \mathbf{b}).$$

The right hand side can be written as $\mathbf{d}'\mathbf{d}$ where $\mathbf{d} = \mathbf{X}(\mathbf{c} - \mathbf{b})$, so it is necessarily positive. This confirms what we knew at the outset, least squares is least squares.

3. Consider the least squares regression of **y** on *K* variables (with a constant), **X**. Consider an alternative set of regressors, **Z** = **XP**, where **P** is a nonsingular matrix. Thus, each column of **Z** is a mixture of some of the columns of **X**. Prove that the residual vectors in the regressions of **y** on **X** and **y** on **Z** are identical. What relevance does this have to the question of changing the fit of a regression by changing the units of measurement of the independent variables?

The residual vector in the regression of **y** on **X** is $\mathbf{M}_X\mathbf{y} = [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{y}$. The residual vector in the regression of **y** on **Z** is

$$\begin{aligned} \mathbf{M}_Z\mathbf{y} &= [\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}']\mathbf{y} \\ &= [\mathbf{I} - \mathbf{XP}((\mathbf{XP})'(\mathbf{XP}))^{-1}(\mathbf{XP})')\mathbf{y} \\ &= [\mathbf{I} - \mathbf{XPP}^{-1}(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{P}')^{-1}\mathbf{P}'\mathbf{X}')\mathbf{y} \\ &= \mathbf{M}_X\mathbf{y} \end{aligned}$$

Since the residual vectors are identical, the fits must be as well. Changing the units of measurement of the regressors is equivalent to postmultiplying by a diagonal **P** matrix whose *k*th diagonal element is the scale factor to be applied to the *k*th variable (1 if it is to be unchanged). It follows from the result above that this will not change the fit of the regression.

4. In the least squares regression of **y** on a constant and **X**, in order to compute the regression coefficients on **X**, we can first transform **y** to deviations from the mean, $\bar{y}$, and, likewise, transform each column of **X** to deviations from the respective column means; second, regress the transformed **y** on the transformed **X** without a constant. Do we get the same result if we only transform **y**? What if we only transform **X**?

In the regression of $\mathbf{y}$ on $\mathbf{i}$ and $\mathbf{X}$, the coefficients on $\mathbf{X}$ are $\mathbf{b} = (\mathbf{X}'\mathbf{M}^0\mathbf{X})^{-1}\mathbf{X}'\mathbf{M}^0\mathbf{y}$. $\mathbf{M}^0 = \mathbf{I} - \mathbf{i}(\mathbf{i}'\mathbf{i})^{-1}\mathbf{i}'$ is the matrix which transforms observations into deviations from their column means. Since $\mathbf{M}^0$ is idempotent and symmetric we may also write the preceding as $[(\mathbf{X}'\mathbf{M}^{0\prime})(\mathbf{M}^0\mathbf{X})]^{-1}(\mathbf{X}'\mathbf{M}^{0\prime}\mathbf{M}^0\mathbf{y})$ which implies that the regression of $\mathbf{M}^0\mathbf{y}$ on $\mathbf{M}^0\mathbf{X}$ produces the least squares slopes. If only $\mathbf{X}$ is transformed to deviations, we would compute $[(\mathbf{X}'\mathbf{M}^{0\prime})(\mathbf{M}^0\mathbf{X})]^{-1}(\mathbf{X}'\mathbf{M}^{0\prime})\mathbf{y}$ but, of course, this is identical. However, if only $\mathbf{y}$ is transformed, the result is $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{M}^0\mathbf{y}$ which is likely to be quite different. We can extend the result in (6-24) to derive what is produced by this computation. In the formulation, we let $\mathbf{X}_1$ be $\mathbf{X}$ and $\mathbf{X}_2$ is the column of ones, so that $\mathbf{b}_2$ is the least squares intercept. Thus, the coefficient vector $\mathbf{b}$ defined above would be $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathrm{y} - a\mathbf{i})$. But, $a = \bar{y} - \mathbf{b}'\bar{\mathbf{x}}$ so $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{y} - \mathbf{i}(\bar{y} - \mathbf{b}'\bar{\mathbf{x}}))$. We can partition this result to produce

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{y} - \mathbf{i}\bar{y}) = \mathbf{b} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{i}(\mathbf{b}'\bar{\mathbf{x}}) = (\mathbf{I} - n(\mathbf{X}'\mathbf{X})^{-1}\bar{\mathbf{x}}\bar{\mathbf{x}}')\mathbf{b}.$$

(The last result follows from $\mathbf{X}'\mathbf{i} = n\bar{\mathbf{x}}$.) This does not provide much guidance, of course, beyond the observation that if the means of the regressors are not zero, the resulting slope vector will differ from the correct least squares coefficient vector.

5. What is the result of the matrix product $\mathbf{M}_1\mathbf{M}$ where $\mathbf{M}_1$ is defined in (3-19) and $\mathbf{M}$ is defined in (3-14)?

$$\mathbf{M}_1\mathbf{M} = (\mathbf{I} - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1')(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') = \mathbf{M} - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{M}$$

There is no need to multiply out the second term. Each column of $\mathbf{MX}_1$ is the vector of residuals in the regression of the corresponding column of $\mathbf{X}_1$ on all of the columns in $\mathbf{X}$. Since that $\mathbf{x}$ is one of the columns in $\mathbf{X}$, this regression provides a perfect fit, so the residuals are zero. Thus, $\mathbf{MX}_1$ is a matrix of zeroes which implies that $\mathbf{M}_1\mathbf{M} = \mathbf{M}$.

6. **Adding an observation.** A data set consists of n observations on $\mathbf{X}_n$ and $\mathbf{y}_n$. The least squares estimator based on these n observations is $\mathbf{b}_n = (\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{X}_n'\mathbf{y}_n$. Another observation, $\mathbf{x}_s$ and y_s, becomes available. Prove that the least squares estimator computed using this additional observation is

$$\mathbf{b}_{n,s} = \mathbf{b}_n + \frac{1}{1 + \mathbf{x}_s'(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{x}_s}(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{x}_s(y_s - \mathbf{x}_s'\mathbf{b}_n).$$

Note that the last term is e_s, the residual from the prediction of y_s using the coefficients based on $\mathbf{X}_n$ and $\mathbf{b}_n$. Conclude that the new data change the results of least squares only if the new observation on y cannot be perfectly predicted using the information already in hand.

7. A common strategy for handling a case in which an observation is missing data for one or more variables is to fill those missing variables with 0s or add a variable to the model that takes the value 1 for that one observation and 0 for all other observations. Show that this 'strategy' is equivalent to discarding the observation as regards the computation of $\mathbf{b}$ but it does have an effect on R^2. Consider the special case in which X contains only a constant and one variable. Show that replacing the missing values of $\mathbf{X}$ with the mean of the complete observations has the same effect as adding the new variable.

8. Let Y denote total expenditure on consumer durables, nondurables, and services, and E_d, E_n, and E_s are the expenditures on the three categories. As defined, $Y = E_d + E_n + E_s$. Now, consider the expenditure system

$$E_d = \alpha_d + \beta_d Y + \gamma_{dd}P_d + \gamma_{dn}P_n + \gamma_{ds}P_s + \varepsilon\gamma_d$$
$$E_n = \alpha_n + \beta_n Y + \gamma_{nd}P_d + \gamma_{nn}P_n + \gamma_{ns}P_s + \varepsilon_n$$
$$E_s = \alpha_s + \beta_s Y + \gamma_{sd}P_d + \gamma_{sn}P_n + \gamma_{ss}P_s + \varepsilon_s.$$

Prove that if all equations are estimated by ordinary least squares, then the sum of the income coefficients will be 1 and the four other column sums in the preceding model will be zero.

For convenience, reorder the variables so that $\mathbf{X} = [\mathbf{i}, \mathbf{P}_d, \mathbf{P}_n, \mathbf{P}_s, \mathbf{Y}]$. The three dependent variables are $\mathbf{E}_d$, $\mathbf{E}_n$, and $\mathbf{E}_s$, and $\mathbf{Y} = \mathbf{E}_d + \mathbf{E}_n + \mathbf{E}_s$. The coefficient vectors are

$$\mathbf{b}_d = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{E}_d,\quad \mathbf{b}_n = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{E}_n, \text{ and } \mathbf{b}_s = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{E}_s.$$

The sum of the three vectors is

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'[\mathbf{E}_d + \mathbf{E}_n + \mathbf{E}_s] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}.$$

Now, $\mathbf{Y}$ is the last column of $\mathbf{X}$, so the preceding sum is the vector of least squares coefficients in the regression of the last column of $\mathbf{X}$ on all of the columns of $\mathbf{X}$, including the last. Of course, we get a perfect fit. In addition, $\mathbf{X}'[\mathbf{E}_d + \mathbf{E}_n + \mathbf{E}_s]$ is the last column of $\mathbf{X}'\mathbf{X}$, so the matrix product is equal to the last column of an identity matrix. Thus, the sum of the coefficients on all variables except income is 0, while that on income is 1.

9. Prove that the adjusted R^2 in (3-30) rises (falls) when variable $\mathbf{x}_k$ is deleted from the regression if the square of the t ratio on $\mathbf{x}_k$ in the multiple regression is less (greater) than one.

The proof draws on the results of the previous problem. Let $\overline{R}_K^2$ denote the adjusted R^2 in the full regression on K variables including $\mathbf{x}_k$, and let $\overline{R}_1^2$ denote the adjusted R^2 in the short regression on K-1 variables when $\mathbf{x}_k$ is omitted. Let R_K^2 and R_1^2 denote their unadjusted counterparts. Then,

$$R_K^2 = 1 - \mathbf{e}'\mathbf{e}/\mathbf{y}'\mathbf{M}^0\mathbf{y}$$

$$R_1^2 = 1 - \mathbf{e}_1'\mathbf{e}_1/\mathbf{y}'\mathbf{M}^0\mathbf{y}$$

where $\mathbf{e}'\mathbf{e}$ is the sum of squared residuals in the full regression, $\mathbf{e}_1'\mathbf{e}_1$ is the (larger) sum of squared residuals in the regression which omits $\mathbf{x}_k$, and $\mathbf{y}'\mathbf{M}^0\mathbf{y} = \Sigma_i (y_i - \overline{y})^2$

Then, $$\overline{R}_K^2 = 1 - [(n-1)/(n-K)](1 - R_K^2)$$

and $$\overline{R}_1^2 = 1 - [(n-1)/(n-(K-1))](1 - R_1^2).$$

The difference is the change in the adjusted R^2 when $\mathbf{x}_k$ is added to the regression,

$$\overline{R}_K^2 - \overline{R}_1^2 = [(n-1)/(n-K+1)][\mathbf{e}_1'\mathbf{e}_1/\mathbf{y}'\mathbf{M}^0\mathbf{y}] - [(n-1)/(n-K)][\mathbf{e}'\mathbf{e}/\mathbf{y}'\mathbf{M}^0\mathbf{y}].$$

The difference is positive if and only if the ratio is greater than 1. After cancelling terms, we require for the adjusted R^2 to increase that $\mathbf{e}_1'\mathbf{e}_1/(n-K+1)]/[(n-K)/\mathbf{e}'\mathbf{e}] > 1$. From the previous problem, we have that $\mathbf{e}_1'\mathbf{e}_1 = \mathbf{e}'\mathbf{e} + b_K^2(\mathbf{x}_k'\mathbf{M}_1\mathbf{x}_k)$, where $\mathbf{M}_1$ is defined above and b_k is the least squares coefficient in the full regression of $\mathbf{y}$ on $\mathbf{X}_1$ and $\mathbf{x}_k$. Making the substitution, we require $[(\mathbf{e}'\mathbf{e} + b_K^2(\mathbf{x}_k'\mathbf{M}_1\mathbf{x}_k))(n-K)]/[(n-K)\mathbf{e}'\mathbf{e} + \mathbf{e}'\mathbf{e}] > 1$. Since $\mathbf{e}'\mathbf{e} = (n-K)s^2$, this simplifies to $[\mathbf{e}'\mathbf{e} + b_K^2(\mathbf{x}_k'\mathbf{M}_1\mathbf{x}_k)]/[\mathbf{e}'\mathbf{e} + s^2] > 1$. Since all terms are positive, the fraction is greater than one if and only $b_K^2(\mathbf{x}_k'\mathbf{M}_1\mathbf{x}_k) > s^2$ or $b_K^2(\mathbf{x}_k'\mathbf{M}_1\mathbf{x}_k/s^2) > 1$. The denominator is the estimated variance of b_k, so the result is proved.

10. Suppose you estimate a multiple regression first with then without a constant. Whether the R^2 is higher in the second case than the first will depend in part on how it is computed. Using the (relatively) standard method, $R^2 = 1 - \mathbf{e}'\mathbf{e} / \mathbf{y}'\mathbf{M}^0\mathbf{y}$, which regression will have a higher R^2?

This R^2 must be lower. The sum of squares associated with the coefficient vector which omits the constant term must be higher than the one which includes it. We can write the coefficient vector in the regression without a constant as $\mathbf{c} = (0,\mathbf{b}^*)$ where $\mathbf{b}^* = (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{y}$, with $\mathbf{W}$ being the other K-1 columns of $\mathbf{X}$. Then, the result of the previous exercise applies directly.

11. Three variables, N, D, and Y all have zero means and unit variances. A fourth variable is $C = N + D$. In the regression of C on Y, the slope is .8. In the regression of C on N, the slope is .5. In the regression of D on Y, the slope is .4. What is the sum of squared residuals in the regression of C on D? There are 21 observations and all moments are computed using $1/(n-1)$ as the divisor.

We use the notation 'Var[.]' and 'Cov[.]' to indicate the sample variances and covariances. Our information is $\text{Var}[N] = 1,\ \text{Var}[D] = 1,\ \text{Var}[Y] = 1.$

Since $C = N + D$, $\text{Var}[C] = \text{Var}[N] + \text{Var}[D] + 2\text{Cov}[N,D] = 2(1 + \text{Cov}[N,D])$.

From the regressions, we have

$\text{Cov}[C,Y]/\text{Var}[Y] = \text{Cov}[C,Y] = .8.$

But, $\text{Cov}[C,Y] = \text{Cov}[N,Y] + \text{Cov}[D,Y].$

Also, $\text{Cov}[C,N]/\text{Var}[N] = \text{Cov}[C,N] = .5,$

but, $\text{Cov}[C,N] = \text{Var}[N] + \text{Cov}[N,D] = 1 + \text{Cov}[N,D]$, so $\text{Cov}[N,D] = -.5$,

so that $\text{Var}[C] = 2(1 + -.5) = 1.$

And, $\text{Cov}[D,Y]/\text{Var}[Y] = \text{Cov}[D,Y] = .4.$

Since $\text{Cov}[C,Y] = .8 = \text{Cov}[N,Y] + \text{Cov}[D,Y]$, $\text{Cov}[N,Y] = .4.$

Finally, $\text{Cov}[C,D] = \text{Cov}[N,D] + \text{Var}[D] = -.5 + 1 = .5.$

Now, in the regression of C on D, the sum of squared residuals is $(n-1)\{\text{Var}[C] - (\text{Cov}[C,D]/\text{Var}[D])^2\text{Var}[D]\}$

based on the general regression result $\Sigma e^2 = \Sigma(y_i - \bar{y})^2 - b^2\Sigma(x_i - \bar{x})^2$. All of the necessary figures were obtained above. Inserting these and n-1 = 20 produces a sum of squared residuals of 15.

12. Using the matrices of sums of squares and cross products immediately preceding Section 3.2.3, compute the coefficients in the multiple regression of real investment on a constant, real GNP and the interest rate. Compute R^2. The relevant submatrices to be used in the calculations are

	Investment	Constant	GNP	Interest
Investment	*	3.0500	3.9926	23.521
Constant		15	19.310	111.79
GNP			25.218	148.98
Interest				943.86

The inverse of the lower right 3×3 block is $(\mathbf{X'X})^{-1}$,

$$(\mathbf{X'X})^{-1} = \begin{matrix} 7.5874 & & \\ -7.41859 & 7.84078 & \\ .27313 & -.598953 & .06254637 \end{matrix}$$

The coefficient vector is $\mathbf{b} = (\mathbf{X'X})^{-1}\mathbf{X'y} = (-.0727985, .235622, -.00364866)'$. The total sum of squares is $\mathbf{y'y}$ = .63652, so we can obtain $\mathbf{e'e} = \mathbf{y'y} - \mathbf{b'X'y}$. $\mathbf{X'y}$ is given in the top row of the matrix. Making the substitution, we obtain $\mathbf{e'e}$ = .63652 - .63291 = .00361. To compute R^2, we require $\Sigma_i (x_i - \bar{y})^2$ = .63652 - $15(3.05/15)^2$ = .01635333, so R^2 = 1 - .00361/.0163533 = .77925.

13. In the December, 1969, American Economic Review (pp. 886-896), Nathanial Leff reports the following least squares regression results for a cross section study of the effect of age composition on savings in 74 countries in 1964:

$$\log S/Y = 7.3439 + 0.1596 \log Y/N + 0.0254 \log G - 1.3520 \log D_1 - 0.3990 \log D_2 \; (R^2 = 0.57)$$

$$\log S/N = 8.7851 + 1.1486 \log Y/N + 0.0265 \log G - 1.3438 \log D_1 - 0.3966 \log D_2 \; (R^2 = 0.96)$$

where S/Y = domestic savings ratio, S/N = per capita savings, Y/N = per capita income, D_1 = percentage of the population under 15, D_2 = percentage of the population over 64, and G = growth rate of per capita income. Are these results correct? Explain.

The results cannot be correct. Since log S/N = log S/Y + log Y/N by simple, exact algebra, the same result must apply to the least squares regression results. That means that the second equation estimated must equal the first one plus log Y/N. Looking at the equations, that means that all of the coefficients would have to be identical save for the second, which would have to equal its counterpart in the first equation, plus 1. Therefore, the results cannot be correct. In an exchange between Leff and Arthur Goldberger that appeared later in the same journal, Leff argued that the difference was simple rounding error. You can see that the results in the second equation resemble those in the first, but not enough so that the explanation is credible.

Chapter 4

Finite-Sample Properties of the Least Squares Estimator

1. Suppose you have two independent unbiased estimators of the same parameter, θ, say $\hat{\theta}_1$ and $\hat{\theta}_2$, with different variances, v_1 and v_2. What linear combination, $\hat{\theta} = c_1\hat{\theta}_1 + c_2\hat{\theta}_2$ is the minimum variance unbiased estimator of θ?

Consider the optimization problem of minimizing the variance of the weighted estimator. If the estimate is to be unbiased, it must be of the form $c_1\hat{\theta}_1 + c_2\hat{\theta}_2$ where c_1 and c_2 sum to 1. Thus, $c_2 = 1 - c_1$. The function to minimize is $\text{Min}_{c_1} L_* = c_1^2 v_1 + (1 - c_1)^2 v_2$. The necessary condition is $\partial L_*/\partial c_1 = 2c_1 v_1 - 2(1 - c_1)v_2 = 0$ which implies $c_1 = v_2 / (v_1 + v_2)$. A more intuitively appealing form is obtained by dividing numerator and denominator by $v_1 v_2$ to obtain $c_1 = (1/v_1) / [1/v_1 + 1/v_2]$. Thus, the weight is proportional to the inverse of the variance. The estimator with the smaller variance gets the larger weight.

2. Consider the simple regression $y_i = \beta x_i + \varepsilon_i$.

(a) What is the minimum mean squared error linear estimator of β? [**Hint:** Let the estimator be $\hat{\beta} = \mathbf{c'y}$]. Choose **c** to minimize $\text{Var}[\hat{\beta}] + [E(\hat{\beta} - \beta)]^2$. (The answer is a function of the unknown parameters.)

(b) For the estimator in (a), show that ratio of the mean squared error of $\hat{\beta}$ to that of the ordinary least squares estimator, *b*, is $\text{MSE}[\hat{\beta}] / \text{MSE}[b] = \tau^2 / (1 + \tau^2)$ where $\tau^2 = \beta^2 / [\sigma^2/\mathbf{x'x}]$. Note that τ is the square of the population analog to the `*t* ratio' for testing the hypothesis that β = 0, which is given after (4-14). How do you interpret the behavior of this ratio as $\tau \rightarrow \infty$?

First, $\hat{\beta} = \mathbf{c'y} = \mathbf{c'x} + \mathbf{c'}\boldsymbol{\varepsilon}$. So $E[\hat{\beta}] = \beta\mathbf{c'x}$ and $\text{Var}[\hat{\beta}] = \sigma^2\mathbf{c'c}$. Therefore,

$\text{MSE}[\hat{\beta}] = \beta^2[\mathbf{c'x} - 1]^2 + \sigma^2\mathbf{c'c}$. To minimize this, we set $\partial\text{MSE}[\hat{\beta}]/\partial\mathbf{c} = 2\beta^2[\mathbf{c'x} - 1]\mathbf{x} + 2\sigma^2\mathbf{c} = \mathbf{0}$.

Collecting terms, $\beta^2(\mathbf{c'x} - 1)\mathbf{x} = -\sigma^2\mathbf{c}$

Premultiply by $\mathbf{x'}$ to obtain $\beta^2(\mathbf{c'x} - 1)\mathbf{x'x} = -\sigma^2\mathbf{x'c}$

or $\mathbf{c'x} = \beta^2\mathbf{x'x} / (\sigma^2 + \beta^2\mathbf{x'x})$.

Then, $\mathbf{c} = [(-\beta^2/\sigma^2)(\mathbf{c'x} - 1)]\mathbf{x}$,

so $\mathbf{c} = [1/(\sigma^2/\beta^2 + \mathbf{x'x})]\mathbf{x}$.

Then, $\hat{\beta} = \mathbf{c'y} = \mathbf{x'y} / (\sigma^2/\beta^2 + \mathbf{x'x})$.

The expected value of this estimator is

$$E[\hat{\beta}] = \beta\mathbf{x'x} / (\sigma^2/\beta^2 + \mathbf{x'x})$$

so

$$E[\hat{\beta}] - \beta = \beta(-\sigma^2/\beta^2) / (\sigma^2/\beta^2 + \mathbf{x'x}) = -(\sigma^2/\beta) / (\sigma^2/\beta^2 + \mathbf{x'x})$$

while its variance is $\text{Var}[\mathbf{x'}(\mathbf{x}\beta + \boldsymbol{\varepsilon}) / (\sigma^2/\beta^2 + \mathbf{x'x})] = \sigma^2\mathbf{x'x} / (\sigma^2/\beta^2 + \mathbf{x'x})^2$

The mean squared error is the variance plus the squared bias,

$$\text{MSE}[\hat{\beta}] = [\sigma^4/\beta^2 + \sigma^2\mathbf{x'x}]/[\sigma^2/\beta^2 + \mathbf{x'x}]^2.$$

The ordinary least squares estimator is, as always, unbiased, and has variance and mean squared error

$$\text{MSE}(b) = \sigma^2/\mathbf{x'x}.$$

The ratio is taken by dividing each term in the numerator

$$\frac{MSE\left[\hat{\beta}\right]}{\text{M}\Sigma\text{E}(\beta)} = \frac{(\sigma^4/\beta^2)/(\sigma^2/\mathbf{x'x}) + \sigma^2\mathbf{x'x}/(\sigma^2/\mathbf{x'x})}{\left(\sigma^2/\beta^2 + \mathbf{x'x}\right)^2}$$

$$= [\sigma^2\mathbf{x'x}/\beta^2 + (\mathbf{x'x})^2]/(\sigma^2/\beta^2 + \mathbf{x'x})^2$$
$$= \mathbf{x'x}[\sigma^2/\beta^2 + \mathbf{x'x}]/(\sigma^2/\beta^2 + \mathbf{x'x})^2$$
$$= \mathbf{x'x}/(\sigma^2/\beta^2 + \mathbf{x'x})$$

Now, multiply numerator and denominator by β^2/σ^2 to obtain

$$\text{MSE}[\hat{\beta}]/\text{MSE}[b] = \beta^2\mathbf{x'x}/\sigma^2/[1 + \beta^2\mathbf{x'x}/\sigma^2] = \tau^2/[1+\tau^2]$$

As $\tau\rightarrow\infty$, the ratio goes to one. This would follow from the result that the biased estimator and the unbiased estimator are converging to the same thing, either as σ^2 goes to zero, in which case the MMSE estimator is the same as OLS, or as $\mathbf{x'x}$ grows, in which case both estimators are consistent.

3. Suppose that the classical regression model applies, but the true value of the constant is zero. Compare the variance of the least squares slope estimator computed without a constant term to that of the estimator computed with an unnecessary constant term.

The OLS estimator fit without a constant term is $b = \mathbf{x'y} / \mathbf{x'x}$. Assuming that the constant term is, in fact, zero, the variance of this estimator is $\text{Var}[b] = \sigma^2/\mathbf{x'x}$. If a constant term is included in the regression, then,

$$b' = \sum_{i=1}^n\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right) / \sum_{i=1}^n\left(x_i - \bar{x}\right)^2$$

The appropriate variance is $\sigma^2/\sum_{i=1}^n\left(x_i - \bar{x}\right)^2$ as always. The ratio of these two is

$$\text{Var}[b]/\text{Var}[b'] = [\sigma^2/\mathbf{x'x}] / [\sigma^2/\sum_{i=1}^n\left(x_i - \bar{x}\right)^2]$$

But,

$$\sum_{i=1}^n\left(x_i - \bar{x}\right)^2 = \mathbf{x'x} + n\bar{x}^2$$

so the ratio is

$$\text{Var}[b]/\text{Var}[b'] = [\mathbf{x'x} + n\bar{x}^2]/\mathbf{x'x} = 1 - n\bar{x}^2/\mathbf{x'x} = 1 - \{n\bar{x}^2/[S_{xx} + n\bar{x}^2]\} \le 1$$

It follows that fitting the constant term when it is unnecessary inflates the variance of the least squares estimator if the mean of the regressor is not zero.

4. Suppose the regression model is $y_i = \alpha + \beta x_i + \varepsilon_i \;\; f(\varepsilon_i) = (1/\lambda)exp(-\varepsilon_i/\lambda) > 0$.
This is rather a peculiar model in that all of the disturbances are assumed to be positive. Note that the disturbances have $E[\varepsilon_i] = \lambda$. Show that the least squares constant term is unbiased but the intercept is biased.

We could write the regression as $y_i = (\alpha + \lambda) + \beta x_i + (\varepsilon_i - \lambda) = \alpha^* + \beta x_i + \varepsilon_i^*$. Then, we know that $E[\varepsilon_i^*] = 0$, and that it is independent of x_i. Therefore, the second form of the model satisfies all of our assumptions for the classical regression. Ordinary least squares will give unbiased estimators of α^* and β. As long as λ is not zero, the constant term will differ from α.

5. Prove that the least squares intercept estimator in the classical regression model is the minimum variance linear unbiased estimator.

Let the constant term be written as $a = \Sigma_i d_i y_i = \Sigma_i d_i(\alpha + \beta x_i + \varepsilon_i) = \alpha\Sigma_i d_i + \beta\Sigma_i d_i x_i + \Sigma_i d_i\varepsilon_i$. In order for a to be unbiased for all samples of x_i, we must have $\Sigma_i d_i = 1$ and $\Sigma_i d_i x_i = 0$. Consider, then, minimizing the variance of a subject to these two constraints. The Lagrangean is

$$L_* = \text{Var}[a] + \lambda_1(\Sigma_i d_i - 1) + \lambda_2\Sigma_i d_i x_i \;\text{ where }\; \text{Var}[a] = \Sigma_i\sigma^2 d_i^2.$$

Now, we minimize this with respect to d_i, λ_1, and λ_2. The $(n+2)$ necessary conditions are

$$\partial L_*/\partial d_i = 2\sigma^2 d_i + \lambda_1 + \lambda_2 x_i,\;\; \partial L_*/\partial\lambda_1 = \Sigma_i d_i - 1,\;\; \partial L_*/\partial\lambda_2 = \Sigma_i d_i x_i$$

The first equation implies that $d_i = [-1/(2\sigma^2)](\lambda_1 + \lambda_2 x_i)$.

Therefore, $\Sigma_i d_i = 1 = [-1/(2\sigma^2)][n\lambda_1 + (\Sigma_i x_i)\lambda_2]$

and $\Sigma_i d_i x_i = 0 = [-1/(2\sigma^2)][(\Sigma_i x_i)\lambda_1 + (\Sigma_i x_i^2)\lambda_2]$.

We can solve these two equations for λ_1 and λ_2 by first multiplying both equations by $-2\sigma^2$ then writing the resulting equations as $\begin{bmatrix} n & \Sigma_i x_i \\ \Sigma_i x_i & \Sigma_i x_i^2 \end{bmatrix}\begin{pmatrix}\lambda_1 \\ \lambda_2\end{pmatrix} = \begin{bmatrix} -2\sigma^2 \\ 0 \end{bmatrix}$. The solution is $\begin{pmatrix}\lambda_1 \\ \lambda_2\end{pmatrix} = \begin{bmatrix} n & \Sigma_i x_i \\ \Sigma_i x_i & \Sigma_i x_i^2 \end{bmatrix}^{-1}\begin{bmatrix} -2\sigma^2 \\ 0 \end{bmatrix}$.

Note, first, that $\Sigma_i x_i = n\bar{x}$. Thus, the determinant of the matrix is $n\Sigma_i x_i^2 - (n\bar{x})^2 = n(\Sigma_i x_i^2 - n\bar{x}^2) = nS_{xx}$ where $S_{xx} \sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$. The solution is, therefore, $\begin{pmatrix}\lambda_1 \\ \lambda_2\end{pmatrix} = \frac{1}{nS_{xx}}\begin{bmatrix} \Sigma_i x_i^2 & -n\bar{x} \\ -n\bar{x} & 0 \end{bmatrix}\begin{bmatrix} -2\sigma^2 \\ 0 \end{bmatrix}$

or $\qquad \lambda_1 = (-2\sigma^2)(\Sigma_i x_i^2/n)/S_{xx}$

$\qquad \lambda_2 = (2\sigma 2\bar{x})/S_{xx}$

Then, $\qquad d_i = [\Sigma_i x_i^2/n - \bar{x} x_i]/S_{xx}$

This simplifies if we write $\Sigma x_i^2 = S_{xx} + n\bar{x}^2$, so $\Sigma_i x_i^2/n = S_{xx}/n + \bar{x}^2$. Then,

$d_i = 1/n + \bar{x}(\bar{x} - x_i)/S_{xx}$, or, in a more familiar form, $d_i = 1/n - \bar{x}(x_i - \bar{x})/S_{xx}$.

This makes the intercept term $\Sigma_i d_i y_i = (1/n)\Sigma_i y_i - \bar{x}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)y_i / S_{xx} = \bar{y} - b\bar{x}$ which was to be shown.

6. As a profit maximizing monopolist, you face the demand curve $Q = \alpha + \beta P + \varepsilon$.
In the past, you have set the following prices and sold the accompanying quantities:

Q	3	3	7	6	10	15	16	13	9	15	9	15	12	18	21
P	18	16	17	12	15	15	4	13	11	6	8	10	7	7	7

Suppose your marginal cost is 10. Based on the least squares regression, compute a 95% confidence interval for the expected value of the profit maximizing output.

Let $q = E[Q]$. Then, $\qquad q = \alpha + \beta P$,

or $\qquad P = (-\alpha/\beta) + (1/\beta)q$.

Using a well known result, for a linear demand curve, marginal revenue is $MR = (-\alpha/\$) + (2/\beta)q$. The profit maximizing output is that at which marginal revenue equals marginal cost, or 10. Equating MR to 10 and solving for q produces $q = \alpha/2 + 5\beta$, so we require a confidence interval for this combination of the parameters.

The least squares regression results are $\hat{Q} = 20.7691 - .840583$. The estimated covariance matrix of the coefficients is $\begin{bmatrix} 7.96124 & -0.624559 \\ -0.624559 & 0.0564361 \end{bmatrix}$. The estimate of q is 6.1816. The estimate of the variance of $\hat{q}$ is (1/4)7.96124 + 25(.056436) + 5(-.0624559) or 0.278415, so the estimated standard error is 0.5276. The 95% cutoff value for a t distribution with 13 degrees of freedom is 2.161, so the confidence interval is 6.1816 - 2.161(.5276) to 6.1816 + 2.161(.5276) or 5.041 to 7.322.

7. The following sample moments were computed from 100 observations produced using a random number generator: $\mathbf{X'X} = \begin{bmatrix} 100 & 123 & 96 & 109 \\ 123 & 252 & 125 & 189 \\ 96 & 125 & 167 & 146 \\ 109 & 189 & 146 & 168 \end{bmatrix}$, $\mathbf{X'y} = \begin{bmatrix} 460 \\ 810 \\ 615 \\ 712 \end{bmatrix}$,

$$(\mathbf{X'X})^{-1} = \begin{bmatrix} 0.03767 & -0.06263 & -.06247 & 0.1003 \\ -0.06263 & 1.129 & 1.107 & -2.102 \\ -0.06247 & 1.107 & 1.110 & -2.170 \\ 0.1003 & -2.192 & -2.170 & 4.292 \end{bmatrix}, \mathbf{y'y} = 3924$$

The true model underlying these data is $y = x_1 + x_2 + x_3 + \varepsilon$.

(a) Compute the simple correlations among the regressors.

(b) Compute the ordinary least squares coefficients in the regression of y on a constant, x_1, x_2, and x_3.

(c) Compute the ordinary least squares coefficients in the regression of y on a constant, x_1, and x_2, on a constant, x_1, and x_3, and on a constant, x_2, and x_3.

(d) Compute the variance inflation factor associated with each variable).

(e) The regressors are obviously collinear. Which is the problem variable?

The sample means are (1/100) times the elements in the first column of **X'X**. The sample covariance matrix for the three regressors is obtained as $(1/99)[(\mathbf{X'X})_{ij} - 100\bar{x}_i\bar{x}_j]$.

Sample Var[**x**] = $\begin{bmatrix} 1.0127 & 0.069899 & 0.555489 \\ 0.069899 & 0.755960 & 0.417778 \\ 0.555489 & 0.417778 & 0.496969 \end{bmatrix}$ The simple correlation matrix is

$\begin{bmatrix} 1 & .07971 & .78043 \\ .07971 & 1 & .68167 \\ .78043 & .68167 & 1 \end{bmatrix}$. The vector of slopes is $(\mathbf{X'X})^{-1}\mathbf{X'y}$ = [-.4022, 6.123, 5.910, -7.525]′. For the three short regressions, the coefficient vectors are

(1) one, x_1, and x_2: [-.223, 2.28, 2.11]′
(2) one, x_1, and x_3 [-.0696, .229, 4.025]′
(3) one, x_2, and x_3: [-.0627, -.0918, 4.358]′

The magnification factors are

for x_1: $[(1/(99(1.01727)) / 1.129]^2 = .094$
for x_2: $[(1/99(.75596)) / 1.11]^2 = .109$
for x_3: $[(1/99(.496969))/ 4.292]^2 = .068$.

The problem variable appears to be x_3 since it has the lowest magnification factor. In fact, all three are highly intercorrelated. Although the simple correlations are not excessively high, the three multiple correlations are .9912 for x_1 on x_2 and x_3, .9881 for x_2 on x_1 and x_3, and .9912 for x_3 on x_1 and x_2.

8. Consider the multiple regression of **y** on K variables, **X** and an additional variable, **z**. Prove that under the assumptions A1 through A6 of the classical regression model, the true variance of the least squares estimator of the slopes on **X** is larger when **z** is included in the regression than when it is not. Does the same hold for the sample estimate of this covariance matrix? Why or why not? Assume that **X** and **z** are nonstochastic and that the coefficient on z is nonzero.

We consider two regressions. In the first, **y** is regressed on K variables, **X**. The variance of the least squares estimator, $\mathbf{b} = (\mathbf{X'X})^{-1}\mathbf{X'y}$, Var[**b**] $= \sigma^2(\mathbf{X'X})^{-1}$. In the second, **y** is regressed on **X** and an additional variable, **z**. Using result (6-18) for the partitioned regression, the coefficients on **X** when **y** is regressed on **X** and **z** are $\mathbf{b}_{.z} = (\mathbf{X'M_zX})^{-1}\mathbf{X'M_zy}$ where $\mathbf{M}_z = \mathbf{I} - \mathbf{z}(\mathbf{z'z})^{-1}\mathbf{z'}$. The true variance of $\mathbf{b}_{.z}$ is the upper left $K\times K$ matrix in Var[**b**,c] $= s^2\begin{bmatrix} \mathbf{X'X} & \mathbf{X'z} \\ \mathbf{z'X} & \mathbf{z'X} \end{bmatrix}^{-1}$. But, we have already found this above. The submatrix is Var[$\mathbf{b}_{.z}$] = $s^2(\mathbf{X'M_zX})^{-1}$. We can show that the second matrix is larger than the first by showing that its inverse is smaller. (See Section 2.8.3). Thus, as regards the true variance matrices $(\text{Var}[\mathbf{b}])^{-1} - (\text{Var}[\mathbf{b}_{.z}])^{-1} = (1/\sigma^2)\mathbf{z}(\mathbf{z'z})^{-1}\mathbf{z'}$ which is a nonnegative definite matrix. Therefore Var[**b**]$^{-1}$ is larger than Var[$\mathbf{b}_{.z}$]$^{-1}$, which implies that Var[**b**] is smaller.

Although the true variance of **b** is smaller than the true variance of $\mathbf{b}_{.z}$, it does not follow that the estimated variance will be. The estimated variances are based on s^2, not the true σ^2. The residual variance estimator based on the short regression is $s^2 = \mathbf{e'e}/(n - K)$ while that based on the regression which includes **z** is $s_z^2 = \mathbf{e}_{.z}'\mathbf{e}_{.z}/(n - K - 1)$. The numerator of the second is definitely smaller than the numerator of the first, but so is the denominator. It is uncertain which way the comparison will go. The result is derived in the previous problem. We can conclude, therefore, that if t ratio on c in the regression which includes **z** is larger than one in absolute value, then s_z^2 will be smaller than s^2. Thus, in the comparison, Est.Var[**b**] $= s^2(\mathbf{X'X})^{-1}$ is based on a smaller matrix, but a larger scale factor than Est.Var[$\mathbf{b}_{.z}$] $= s_z^2(\mathbf{X'M_zX})^{-1}$. Consequently, it is uncertain whether the estimated standard errors in the short regression will be smaller than those in the long one. Note that it is not sufficient merely for the result of the previous problem to hold, since the relative sizes of the matrices also play a role. But, to take a polar case, suppose **z** and **X** were uncorrelated. Then, $\mathbf{XNM_zX}$ equals **XNX**. Then, the estimated variance of $\mathbf{b}_{.z}$ would be less than that of **b** without **z** even though the true variance is the same (assuming the premise of the previous problem holds). Now, relax this assumption while holding the *t* ratio on c constant. The matrix in Var[$\mathbf{b}_{.z}$] is now larger, but the leading scalar is now smaller. Which way the product will go is uncertain.

9. For the classical regression model $\mathbf{y} = \mathbf{X}\beta + \varepsilon$ with no constant term and *K* regressors, assuming that the true value of β is zero, what is the exact expected value of $F[K, n\text{-}K] = (R^2/K)/[(1\text{-}R^2)/(n\text{-}K)]$?

The *F* ratio is computed as $[\mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b}/K]/[\mathbf{e}'\mathbf{e}/(n - K)]$. We substitute **e** = **M,** and $\mathbf{b} = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon$. Then, $F = [\varepsilon'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon/K]/[\varepsilon'\mathbf{M}\varepsilon/(n - K)] = [\varepsilon'(\mathbf{I} - \mathbf{M})\varepsilon/K]/[\varepsilon'\mathbf{M}\varepsilon/(n - K)]$.

The exact expectation of *F* can be found as follows: $F = [(n\text{-}K)/K][\varepsilon'(\mathbf{I} - \mathbf{M})\varepsilon]/[\varepsilon'\mathbf{M}\varepsilon]$. So, its exact expected value is $(n\text{-}K)/K$ times the expected value of the ratio. To find that, we note, first, that **M,** and **(I - M),** are independent because $\mathbf{M}(\mathbf{I} - \mathbf{M}) = \mathbf{0}$. Thus, $E\{[\varepsilon(\mathbf{I} - \mathbf{M})\varepsilon]/[\varepsilon'\mathbf{M}\varepsilon]\} = E[\varepsilon'(\mathbf{I}- \mathbf{M})\varepsilon]\times E\{1/[\varepsilon'\mathbf{M}\varepsilon]\}$. The first of these was obtained above, $E[\varepsilon'(\mathbf{I} - \mathbf{M})\varepsilon] = K\sigma^2$. The second is the expected value of the reciprocal of a chi-squared variable. The exact result for the reciprocal of a chi-squared variable is $E[1/\chi^2(n\text{-}K)] = 1/(n - K - 2)$. Combining terms, the exact expectation is $E[F] = (n - K) / (n - K - 2)$. Notice that the mean does not involve the numerator degrees of freedom. ~

10. Prove that $E[\mathbf{b}'\mathbf{b}] = \beta'\beta + \sigma^2\Sigma_k (1/\lambda_k)$ where **b** is the ordinary least squares estimator and λ_k is a characteristic root of **X′X**.

We write $\mathbf{b} = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon$, so $\mathbf{b}'\mathbf{b} = \beta'\beta + \varepsilon'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon + 2\beta'(\mathbf{X}'\mathbf{X})^{-1}X'\varepsilon$. The expected value of the last term is zero, and the first is nonstochastic. To find the expectation of the second term, use the trace, and permute $\varepsilon'\mathbf{X}$ inside the trace operator. Thus,

$$
\begin{aligned}
E[\beta'\beta] &= \beta'\beta + E[\varepsilon'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon] \\
&= \beta'\beta + E[tr\{\varepsilon'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon\}] \\
&= \beta'\beta + E[tr\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon\varepsilon'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\}] \\
&= \beta'\beta + tr[E\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon\varepsilon'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\}] \\
&= \beta'\beta + tr[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\varepsilon\varepsilon']\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] \\
&= \beta'\beta + tr[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\sigma^2\mathbf{I})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] \\
&= \beta'\beta + \sigma^2 tr[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] \\
&= \beta'\beta + \sigma^2 tr[(\mathbf{X}'\mathbf{X})^{-1}] \\
&= \beta'\beta + \sigma^2\Sigma_k (1/\lambda_k)
\end{aligned}
$$

The trace of the inverse equals the sum of the characteristic roots of the inverse, which are the reciprocals of the characteristic roots of **X′X**.

11. Data on U.S. gasoline consumption in the United States in the years 1960 to 1995 are given in Table F2.2.

(a) Compute the multiple regression of per capita consumption of gasoline, G/Pop, on all of the other explanatory variables, including the time trend, and report all results. Do the signs of the estimates agree with your expectations?

(b) Test the hypothesis that at least in regard to demand for gasoline, consumers do not differentiate between changes in the prices of new and used cars.

(c) Estimate the own price elasticity of demand, the income elasticity, and the cross price elasticity with respect to changes in the price of public transportation.

(d) Reestimate the regression in logarithms, so that the coefficients are direct estimates of the elasticities. (Do not use the log of the time trend.) How do your estimates compare to the results in the previous question? Which specification do you prefer?

(e) Notice that the price indices for the automobile market are normalized to 1967 while the aggregate price indices are anchored at 1982. Does this discrepancy affect the results? How? If you were to renormalize the indices so that they were all 1.000 in 1982, how would your results change?

Part (a) The regression results for the regression of G/Pop on all other variables are:

```
+--------------------------------------------------------------------------+
```

```
| Ordinary    least squares regression    Weighting variable = none     |
| Dep. var. = G        Mean=   100.7008114    , S.D.=   14.08790232     |
| Model size: Observations =      36, Parameters =  10, Deg.Fr.=     26 |
| Residuals:  Sum of squares= 117.5342920    , Std.Dev.=        2.12616 |
| Fit:        R-squared=  .983080, Adjusted R-squared =          .97722 |
| Model test: F[  9,     26] =  167.85,    Prob value =          .00000 |
| Diagnostic: Log-L =    -72.3796, Restricted(b=0) Log-L =    -145.8061 |
|             LogAmemiyaPrCrt.=    1.754, Akaike Info. Crt.=      4.577 |
| Autocorrel: Durbin-Watson Statistic =    .94418,   Rho =       .52791 |
+-----------------------------------------------------------------------+
```

Variable	Coefficient	Standard Error	t-ratio	P[|T|>t]	Mean of X
Constant	-1859.389661	1699.6133	-1.094	.2840	
YEAR	.9485446803	.87693228	1.082	.2893	1977.5000
PG	-12.18681017	2.6071552	-4.674	.0001	2.3166111
Y	.1110971600E-01	.32230846E-02	3.447	.0019	9232.8611
PNC	6.889686945	13.203241	.522	.6062	1.6707778
PUC	-4.121840732	2.8707832	-1.436	.1630	2.3436389
PPT	6.034560575	4.0693845	1.483	.1501	2.7448611
PN	20.50251499	16.556303	1.238	.2267	2.0851111
PD	14.18819749	17.122006	.829	.4148	1.6505636
PS	-31.48299999	12.795328	-2.461	.0208	2.3689802

The price and income coefficients are what one would expect of a demand equation (if that is what this is -- see Chapter 16 for extensive analysis). The positive coefficient on the price of new cars would seem counterintuitive. But, newer cars tend to be more fuel efficient than older ones, so a rising price of new cars reduces demand to the extent that people buy fewer cars, but increases demand if the effect is to cause people to retain old (used) cars instead of new ones and, thereby, increase the demand for gasoline. The negative coefficient on the price of used cars is consistent with this view. Since public transportation is a clear substitute for private cars, the positive coefficient is to be expected. Since automobiles are a large component of the 'durables' component, the positive coefficient on *PD* might be indicating the same effect discussed above. Of course, if the linear regression is properly specified, then the effect of PD observed above must be explained by some other means. This author had no strong prior expectation for the signs of the coefficients on *PD* and *PN*. Finally, since a large component of the services sector of the economy is businesses which service cars, if the price of these services rises, the effect will be to make it more expensive to use a car, i.e., more expensive to use the gasoline one purchases. Thus, the negative sign on PS was to be expected.

Part (b) The computer results include the following covariance matrix for the coefficients on *PNC* and *PUC* $\begin{bmatrix} 174.326 & 2.62732 \\ 2.62732 & 8.2414 \end{bmatrix}$. The test statistic for testing the hypothesis that the slopes on these two variables are equal can be computed exactly as in the first Exercise. Thus,

$t[26] = [6.889686945-(-4.121840732)]/[(174.326+8.2414-2(2.62732)]^{1/2} = 0.827$.

This is quite small, so the hypothesis is not rejected.

Part (c) The elasticities for the linear model can be computed using $\eta = b(\bar{x}/\overline{G/Pop})$ for the various *x*s. The mean of *G* is 100.701. The calculations for own price, income, and the price of public transportation are

Variable	Coefficient	Mean	Elasticity
PG	-12.18681017	2.3166111	-0.280
Y	0.011109716	9232.8611	+1.019
PPT	6.034560575	2.7448611	+0.164

Part (d) The estimates of the coefficients of the loglinear and linear equations are

```
Constant  2.276660667     -1859.389661
YEAR     -.00440933049        0.9485446803
LPG      -.5380992257       -12.18681017    (Elasticity = -0.28)
LY        1.217805741         0.01110971600 (Elasticity = +1.019)
LPNC      .09006338891        6.889686945
LPUC     -.1146769420        -4.121840732
LPPT      .1232808093         6.034560575   (Elasticity = +0.164)
LPN       1.224804198        20.50251499
LPD       .9484508600        14.18819749
LPS      -1.321253144       -31.48299999
```

The estimates are roughly similar, but not as close as one might hope. There is little prior information which would suggest which is the better model.

Part (e) We would divide P_d by .483, P_n by .375, and P_s by .353. This would have no effect on the fit of the regression or on the coefficients on the other regressors. The resulting least squares regression coefficients would be multiplied by these values.

Chapter 5

Large-Sample Properties of the Least Squares and Instrumental Variables Estimators

1. For the classical regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ with no constant term and K regressors, what is

$$\text{plim } F[K,n\text{-}K] = \text{plim } (R^2/K)/[(1\text{-}R^2)/(n\text{-}K)]$$

assuming that the true value of $\boldsymbol{\beta}$ is zero? What is the exact expected value?

The F ratio is computed as $[\mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b}/K]/[\mathbf{e}'\mathbf{e}/(n - K)]$. We substitute $\mathbf{e} = \mathbf{M},$ and $\mathbf{b} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}$. Then, $F = [\boldsymbol{\varepsilon}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}/K]/[\boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon}/(n - K)] = [\boldsymbol{\varepsilon}'(\mathbf{I} - \mathbf{M})\boldsymbol{\varepsilon}/K]/[\boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon}/(n - K)]$. The denominator converges to σ^2 as we have seen before. The numerator is an idempotent quadratic form in a normal vector. The trace of $(\mathbf{I} - \mathbf{M})$ is K regardless of the sample size, so the numerator is always distributed as σ^2 times a chi-squared variable with K degrees of freedom. Therefore, the numerator of F does not converge to a constant, it converges to σ^2/K times a chi-squared variable with K degrees of freedom. Since the denominator of F converges to a constant, σ^2, the statistic converges to a random variable, $(1/K)$ times a chi-squared variable with K degrees of freedom.

2. Let e_i be the ith residual in the ordinary least squares regression of $\mathbf{y}$ on $\mathbf{X}$ in the classical regression model and let ε_i be the corresponding true disturbance. Prove that $\text{plim}(e_i - \varepsilon_i) = 0$.

We can write e_i as $e_i = y_i - \mathbf{b}'\mathbf{x}_i = (\boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i) - \mathbf{b}'\mathbf{x}_i = \varepsilon_i + (\mathbf{b} - \boldsymbol{\beta})'\mathbf{x}_i$

We know that plim $\mathbf{b} = \boldsymbol{\beta}$, and $\mathbf{x}_i$ is unchanged as n increases, so as $n\rightarrow\infty$, e_i is arbitrarily close to ε_i.

3. For the simple regression model, $y_i = \mu + \varepsilon_i$, $\varepsilon_i \sim N(0,\sigma^2)$, prove that the sample mean is consistent and asymptotically normally distributed. Now, consider the alternative estimator $\hat{\mu} = \Sigma_i w_i y_i$, where $w_i = i/(n(n+1)/2) = i/\Sigma_i i$. Note that $\Sigma_i w_i = 1$. Prove that this is a consistent estimator of μ and obtain its asymptotic variance. [**Hint:** $\Sigma_i i^2 = n(n+1)(2n+1)/6$.]

The estimator is $\bar{y} = (1/n)\Sigma_i y_i = (1/n)\Sigma_i (\mu + \varepsilon_i) = \mu + (1/n)\Sigma_i \varepsilon_i$. Then, $E[\bar{y}]\ \mu + (1/n)\Sigma_i E[\varepsilon_i] = \mu$ and $\text{Var}[\bar{y}] = (1/n^2)\Sigma_i \Sigma_j \text{Cov}[\varepsilon_i,\varepsilon_j] = \sigma^2/n$. Since the mean equals μ and the variance vanishes as $n\rightarrow\infty$, $\bar{y}$ is consistent. In addition, since $\bar{y}$ is a linear combination of normally distributed variables, $\bar{y}$ has a normal distribution with the mean and variance given above in every sample. Suppose that ε_i were not normally distributed. Then, $\sqrt{n}(\bar{y} - \mu) = (1/\sqrt{n})(\Sigma_i \varepsilon_i)$ satisfies the requirements for the central limit theorem. Thus, the asymptotic normal distribution applies whether or not the disturbances have a normal distribution.

For the alternative estimator, $\hat{\mu} = \Sigma_i w_i y_i$, so $E[\hat{\mu}] = \Sigma_i w_i E[y_i] = \Sigma_i w_i \mu = \mu \Sigma_i w_i = \mu$ and $\text{Var}[\hat{\mu}] = \Sigma_i w_i^2 \sigma^2 = \sigma^2 \Sigma_i w_i^2$. The sum of squares of the weights is $\Sigma_i w_i^2 = \Sigma_i i^2/[\Sigma_i i]^2 = [n(n+1)(2n+1)/6]/[n(n+1)/2]^2 = [2(n^2 + 3n/2 + 1/2)]/[1.5n(n^2 + 2n + 1)]$. As $n\rightarrow\infty$, the fraction will be dominated by the term $(1/n)$ and will tend to zero. This establishes the consistency of this estimator. The last expression also provides the asymptotic variance. The large sample variance can be found as $\text{Asy.Var}[\hat{\mu}] = (1/n)\lim_{n\rightarrow\infty} \text{Var}[\sqrt{n}(\hat{\mu} - \mu)]$. For the estimator above, we can use $\text{Asy.Var}[\hat{\mu}] = (1/n)\lim_{n\rightarrow\infty} n\text{Var}[\hat{\mu} - \mu] = (1/n)\lim_{n\rightarrow\infty} \sigma^2[2(n^2 + 3n/2 + 1/2)]/[1.5(n^2 + 2n + 1)] = 1.3333\sigma^2$. Notice that this is unambiguously larger than the variance of the sample mean, which is the ordinary least squares estimator.

4. In the discussion of the instrumental variables estimator, we showed that the least squares estimator, **b**, is biased and inconsistent. Nonetheless, **b** does estimate something; plim **b** = θ = β + $\mathbf{Q}^{-1}\gamma$. Derive the asymptotic covariance matrix of **b** and show that **b** is asymptotically normally distributed.

To obtain the asymptotic distribution, write the result already in hand as $\mathbf{b} = (\beta + \mathbf{Q}^{-1}\gamma) + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon - \mathbf{Q}^{-1}\varepsilon$. We have established that plim **b** = $\beta + \mathbf{Q}^{-1}\gamma$. For convenience, let $\theta \neq \beta$ denote $\beta + \mathbf{Q}^{-1}\gamma$ = plim **b**. Write the preceding in the form $\mathbf{b} - \theta = (\mathbf{X}'\mathbf{X}/n)^{-1}(\mathbf{X}'\varepsilon/n) - \mathbf{Q}^{-1}\gamma$. Since $\text{plim}(\mathbf{X}'\mathbf{X}/n) = \mathbf{Q}$, the large sample behavior of the right hand side is the same as that of $\text{plim}(\mathbf{b} - \theta) = \mathbf{Q}^{-1}\text{plim}(\mathbf{X}'\varepsilon/n) - \mathbf{Q}^{-1}\gamma$. That is, we may replace $(\mathbf{X}'\mathbf{X}/n)$ with **Q** in our derivation. Then, we seek the asymptotic distribution of $\sqrt{n}\,(\mathbf{b} - \theta)$ which is the same as that of

$\sqrt{n}\,[\mathbf{Q}^{-1}\text{plim}(\mathbf{X}'\varepsilon/n) - \mathbf{Q}^{-1}\gamma] = \mathbf{Q}^{-1}\sqrt{n}\left\{\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\varepsilon_i - \gamma\right\}$. From this point, the derivation is exactly the same as that when γ = **0**, so there is no need to redevelop the result. We may proceed directly to the same asymptotic distribution we obtained before. The only difference is that the least squares estimator estimates θ, not β.

5. For the model in (5-25) and (5-26), prove that when only x^* is measured with error, the squared correlation between y and x is less than that between y^* and x^*. (Note the assumption that $y^* = y$.) Does the same hold true if y^* is also measured with error?

Using the notation in the text, $\text{Var}[x^*] = Q^*$ so, if $y = \beta x^* + \varepsilon$,

$$\text{Corr}^2[y,x^*] = (\beta Q^*)^2 / [(\beta^2 Q^* + \sigma_\varepsilon^2)Q^*] = \beta^2 Q^*/[\$^2 Q^* + \sigma_\varepsilon^2)]$$

In terms of the erroneously measured variables,

$$\text{Cov}[y,x] = \text{Cov}[\beta x^* + \varepsilon, x^* + u] = \beta Q^*,$$

so

$$\text{Corr}^2[y,x] = (\beta Q^*)^2/[(\beta^2 Q^* + \varepsilon_\varepsilon^2)(Q^* + \sigma_u^2)] = [Q^*/(Q^* + \sigma_u^2)]\text{Corr}^2[y,x^*]$$

If y^* is also measured with error, the attenuation in the correlation is made even worse. The numerator of the squared correlation is unchanged, but the term $(\beta^2 Q^* + \sigma_\varepsilon^2)$ in the denominator is replaced with $(\beta^2 Q^* + \sigma_\varepsilon^2 + \sigma_v^2)$ which reduces the squared correlation yet further. □

6. Christensen and Greene (1976) estimate a generalized Cobb-Douglas function of the form

$$\log(C/P_f) = \alpha + \beta\log Q + \gamma\log^2 Y + \delta_k\log(P_k/P_f) + \delta_l\log(P_l/P_f) + \varepsilon.$$

P_k, P_l, and P_f indicate unit prices of capital, labor, and fuel, respectively, Q is output and C is total cost. The purpose of the generalization was to produce a ∪-shaped average total cost curve. (See Example 7.3 for discussion of Nerlove's (1963) predecessor to this study.) We are interested in the output at which the cost curve reaches its minimum. That is the point at which $[\partial\log C/\partial\log Q]|_{Q = Q^*} = 1$, or $Q^* = 10^{(1 - \beta)/(2\gamma)}$. (You can simplify the analysis a bit by using the fact that $10^x = \exp(2.3026x)$. Thus, $Q^* = \exp(2.3026[(1-\beta)/(2\gamma)])$.

The estimated regression model using the Christensen and Greene (1970) data are as follows, where estimated standard errors are given in parentheses:

$$\ln(C/P_f) = -\underset{(0.34427)}{7.294} + \underset{(0.036988)}{0.39091}\ln Q + \underset{(0.0051548)}{0.062413}(\ln^2 Q)/2 + \underset{(0.061645)}{0.07479}\ln(P_k/P_f) + \underset{(0.068109)}{0.2608}\ln(P_l/P_f).$$

The estimated asymptotic covariance of the estimators of β and γ is –0.000187067. R^2=0.991538, **e'e** = 2.443509.

Using the estimates given in the example, compute the estimate of this ***efficient scale***. Estimate the asymptotic distribution of this estimator assuming that the estimate of the asymptotic covariance of $\hat{\beta}$ and $\hat{\gamma}$ is -.00008.

The estimate is $Q^* = \exp[2.3026(1 - .151)/(2(.117))] = 4248$. The asymptotic variance of $Q^* = \exp[2.3026(1 - \hat{\beta})/(2\hat{\gamma})$ is $[\partial Q^*/\partial\beta \;\; \partial Q^*/\partial\gamma]\,\text{Asy.Var}[\hat{\beta},\hat{\gamma}]\,[\partial Q^*/\partial\beta \;\; \partial Q^*/\partial\gamma]'$. The derivatives are

$\partial Q^*/\partial\hat{\beta}$ = Q*(-2.3026$\hat{\beta}$)/(2$\hat{\gamma}$) = -6312. $\partial Q^*/\partial\hat{\gamma}$ = Q*[-2.3026(1-$\hat{\beta}$)]/(2$\hat{\gamma}^2$) = -303326. The estimated asymptotic covariance matrix is $\begin{bmatrix} .00384 & -.00008 \\ -.00008 & .000144 \end{bmatrix}$. The estimated asymptotic variance of the estimate of Q* is thus 13,095,615. The estimate of the asymptotic standard deviation is 3619. Notice that this is quite large compared to the estimate. A confidence interval formed in the usual fashion includes negative values. This is common with highly nonlinear functions such as the one above.

7. The consumption function used in Example 5.3 is a very simple specification. One might wonder if the meager specification of the model could help explain the finding in the Hausman test. The data set used for the example are given in Table F5.1. Use these data to carry out the test in a more elaborate specification

$$c_t = \beta_1 + \beta_2 y_t + \beta_3 i_t + \beta_4 c_{t-1} + \varepsilon_t$$

where c_t is the log of real consumption, y_t is the log of real disposable income and i_t is the interest rate (90 day T bill rate).

Results of the computations are shown below. The Hausman statistic is 25.1 and the t statistic for the Wu test is -5.3. Both are larger than the table critical values by far, so the hypothesis that least squares is consistent is rejected in both cases.

```
--> samp;1-204$
--> crea;ct=log(realcons);yt=log(realdpi);it=tbilrate$
--> crea;ct1=ct[-1];yt1=yt[-1]$
--> samp;2-204$
--> name;x=one,yt,it,ct1;z=one,it,ct1,yt1$
--> regr;lhs=ct;rhs=x$
--> calc;s2=ssqrd$
--> matr;bls=b;xx=<x'x>$
--> 2sls;lhs=ct;rhs=x;inst=z$
--> matr;biv=b;xhxh=1/ssqrd*varb$
--> matr;d=biv-bls;vb=xhxh-xx$
--> matr;list;h=1/s2*d'*mpnv(vb)*d$
--> regr;lhs=yt;rhs=z;keep=ytf$
--> regr;lhs=ct;rhs=x,ytf$
```

```
+-----------------------------------------------------------------------+
| Ordinary    least squares regression    Weighting variable = none     |
| Dep. var. = CT       Mean=   7.884560181    , S.D.=    .5129509097     |
| Model size: Observations =     203, Parameters =   4, Deg.Fr.=    199 |
| Residuals:  Sum of squares= .1318216478E-01, Std.Dev.=         .00814 |
| Fit:        R-squared=  .999752, Adjusted R-squared =          .99975 |
| Model test: F[  3,    199] =********,    Prob value =          .00000 |
| Diagnostic: Log-L =    690.6283, Restricted(b=0) Log-L =    -152.0255 |
|             LogAmemiyaPrCrt.=   -9.603, Akaike Info. Crt.=     -6.765 |
| Autocorrel: Durbin-Watson Statistic =   1.90738,   Rho =       .04631 |
+-----------------------------------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |t-ratio |P[|T|>t] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant -.4413074204E-01  .12917632E-01    -3.416   .0008
 YT          .1833744954   .32943409E-01     5.566   .0000     7.9953259
 IT       -.1654147681E-02  .29350320E-03    -5.636   .0000     5.2499007
 CT1         .8216667186   .32285244E-01    25.450   .0000     7.8757433

+-----------------------------------------------------------------------+
| Two stage   least squares regression    Weighting variable = none     |
| Dep. var. = CT       Mean=   7.884560181    , S.D.=    .5129509097     |
| Model size: Observations =     203, Parameters =   4, Deg.Fr.=    199 |
| Residuals:  Sum of squares= .1344364458E-01, Std.Dev.=         .00822 |
| Fit:        R-squared=  .999742, Adjusted R-squared =          .99974 |
|             (Note:  Not using OLS.  R-squared is not bounded in [0,1] |
| Model test: F[  3,    199] =********,    Prob value =          .00000 |
```

```
| Diagnostic: Log-L =    688.6346, Restricted(b=0) Log-L =    -152.0255 |
|             LogAmemiyaPrCrt.=   -9.583, Akaike Info. Crt.=     -6.745 |
| Autocorrel: Durbin-Watson Statistic =   2.02762,   Rho =      -.01381 |
+---------------------------------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant -.2023353156E-01  .13906118E-01   -1.455   .1457
 YT        .9004120016E-01  .38219830E-01    2.356   .0185     7.9953259
 IT      -.1168585850E-02  .31214268E-03   -3.744   .0002     5.2499007
 CT1          .9130592037   .37448694E-01   24.382   .0000     7.8757433
 (Note: E+nn or E-nn means multiply by 10 to + or -nn power.)


Matrix H        has  1 rows and  1 columns.
               1
        +--------------
       1|    25.0986

+---------------------------------------------------------------------+
| Ordinary    least squares regression    Weighting variable = none    |
| Dep. var. = YT       Mean=   7.995325935    , S.D.=   .5109250627    |
| Model size: Observations =     203, Parameters =   4, Deg.Fr.=    199 |
| Residuals:  Sum of squares= .1478971099E-01, Std.Dev.=         .00862 |
| Fit:        R-squared=  .999720, Adjusted R-squared =          .99972 |
| Model test: F[  3,    199] =********,    Prob value =          .00000 |
| Diagnostic: Log-L =    678.9490, Restricted(b=0) Log-L =    -151.2222 |
|             LogAmemiyaPrCrt.=   -9.488, Akaike Info. Crt.=     -6.650 |
| Autocorrel: Durbin-Watson Statistic =   1.77592,   Rho =       .11204 |
+---------------------------------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |t-ratio |P[|T|>t] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant  .4045167318E-01  .13493797E-01    2.998   .0031
 IT        .2943892707E-03  .32000803E-03     .920   .3587     5.2499007
 CT1       .9130171904E-01  .35621085E-01    2.563   .0111     7.8757433
 YT1          .9057719332   .36310045E-01   24.945   .0000     7.9868448

+---------------------------------------------------------------------+
| Ordinary    least squares regression    Weighting variable = none    |
| Dep. var. = CT       Mean=   7.884560181    , S.D.=   .5129509097    |
| Model size: Observations =     203, Parameters =   5, Deg.Fr.=    198 |
| Residuals:  Sum of squares= .1151983043E-01, Std.Dev.=         .00763 |
| Fit:        R-squared=  .999783, Adjusted R-squared =          .99978 |
| Model test: F[  4,    198] =********,    Prob value =          .00000 |
| Diagnostic: Log-L =    704.3099, Restricted(b=0) Log-L =    -152.0255 |
|             LogAmemiyaPrCrt.=   -9.728, Akaike Info. Crt.=     -6.890 |
| Autocorrel: Durbin-Watson Statistic =   2.35530,   Rho =      -.17765 |
+---------------------------------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |t-ratio |P[|T|>t] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant -.2023559983E-01  .12905160E-01   -1.568   .1185
 YT           .4752021457   .62720658E-01    7.576   .0000     7.9953259
 IT      -.1168629424E-02  .28967486E-03   -4.034   .0001     5.2499007
 CT1          .9130504994   .34753056E-01   26.273   .0000     7.8757433
 YTF         -.3851520841   .72054899E-01   -5.345   .0000     7.9953259
```

8. Suppose we change the assumptions of the model in Section 5.3 to **AS5**: ($\mathbf{x}_i$,ε) are an independent and identically distributed sequence of random vectors such that $\mathbf{x}_i$ has a finite mean vector, $\boldsymbol{\mu}_\mathbf{x}$**,** finite positive definite covariance matrix $\boldsymbol{\Sigma}_{\mathbf{xx}}$ and finite fourth moments $E[x_j x_k x_l x_m] = \phi_{jklm}$ for all variables. How does the proof of consistency and asymptotic normality of **b** change? Are these assumptions weaker or stronger than the ones made in Section 5.2?

The assumption above is considerably stronger than the assumption AD5. Under these assumptions, the Slutsky theorem and the Lindberg Levy versions of the central limit theorem can be invoked.

9. Now, assume only finite second moments of **x**; $E[x_i^2]$ is finite. Is this sufficient to establish consistency of **b**? (Hint: the Cauchy-Schwartz inequality (Theorem D.13), $E[|xy|] \leq \{E[x^2]\}^{1/2}\{E[y^2]\}^{1/2}$ will be helpful.) Is

The assumption will provide that $(1/n)\mathbf{X}'\mathbf{X}$ converges to a finite matrix by virtue of the Cauchy-Schwartz inequality given above. If the assumptions made to ensure that plim $(1/n)\mathbf{X}'\varepsilon = 0$ continue to hold, then consistency can be established by the Slutsky Theorem.