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 Kx1 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 kth diagonal element is the scale factor to be applied to the kth 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 $\mathbf{M}\mathbf{X}_1$ 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, $\mathbf{M}\mathbf{X}_1$ 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, $$\overline{R}_K^2 = 1 - [(n\text{-}1)/(n\text{-}K)](1 - R_K^2)$$

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

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

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

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 $\mathbf{e}_1'\mathbf{e}_1/(n\text{-}K\text{+}1)]/[(n\text{-}K)/\mathbf{e}'\mathbf{e}] > 1$. 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 $[(\mathbf{e}'\mathbf{e} + b_K^2(\mathbf{x}_k'\mathbf{M}_1\mathbf{x}_k))(n\text{-}K)]/[(n\text{-}K)\mathbf{e}'\mathbf{e} + \mathbf{e}'\mathbf{e}] > 1$. Since $\mathbf{e}'\mathbf{e} = (n\text{-}K)s^2$, 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 $1/(n\text{-}1)$ 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 $(n\text{-}1)\{\text{Var}[C] - (\text{Cov}[C,D]/\text{Var}[D])^2\text{Var}[D]\}$

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 $(\mathbf{X}'\mathbf{X})^{-1}$,

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

The coefficient vector is $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (-.0727985, .235622, -.00364866)'$. The total sum of squares is $\mathbf{y}'\mathbf{y} = .63652$, so we can obtain $\mathbf{e}'\mathbf{e} = \mathbf{y}'\mathbf{y} - \mathbf{b}'\mathbf{X}'\mathbf{y}$. $\mathbf{X}'\mathbf{y}$ is given in the top row of the matrix. Making the substitution, we obtain $\mathbf{e}'\mathbf{e} = .63652 - .63291 = .00361$. 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 $\partial L_*/\partial c_1 = 2c_1v_1 - 2(1 - c_1)v_2 = 0$ which implies $c_1 = v_2 / (v_1 + v_2)$. A more intuitively appealing form is obtained by dividing numerator and denominator by v_1v_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 $\hat{\beta} = \mathbf{c}'\mathbf{y}$]. 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 $\tau^2 = \beta^2 / [\sigma^2/\mathbf{x}'\mathbf{x}]$. 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, $\hat{\beta} = \mathbf{c}'\mathbf{y} = \mathbf{c}'\mathbf{x} + \mathbf{c}'\boldsymbol{\varepsilon}$. So $E[\hat{\beta}] = \beta\mathbf{c}'\mathbf{x}$ and $\text{Var}[\hat{\beta}] = \sigma^2\mathbf{c}'\mathbf{c}$. Therefore,

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

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

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

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

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

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

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

The expected value of this estimator is

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

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

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

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

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

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

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

The ratio is taken by dividing each term in the numerator

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

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

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

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

As $\tau\to\infty$, 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 $\mathrm{Var}[b] = \sigma^2/\mathbf{x'x}$. 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

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

But,

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

so the ratio is

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

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 $y_i = \alpha + \beta x_i + \varepsilon_i \quad f(\varepsilon_i) = (1/\lambda)exp(-\varepsilon_i/\lambda) > 0$.
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

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

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

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

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 $(1/99)[(\mathbf{X}'\mathbf{X})_{ij} - 100\,\bar{x}_i\bar{x}_j\,]$.

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

$$\begin{bmatrix} 1 & .07971 & .78043 \\ .07971 & 1 & .68167 \\ .78043 & .68167 & 1 \end{bmatrix}.$$ The vector of slopes is $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = [-.4022, 6.123, 5.910, -7.525]'$. 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, $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, $\text{Var}[\mathbf{b}] = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$. 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 $\mathbf{b}_{.z} = (\mathbf{X}'\mathbf{M}_z\mathbf{X})^{-1}\mathbf{X}'\mathbf{M}_z\mathbf{y}$ where $\mathbf{M}_z = \mathbf{I} - \mathbf{z}(\mathbf{z}'\mathbf{z})^{-1}\mathbf{z}'$. The true variance of $\mathbf{b}_{.z}$ is the upper left $K\times K$ matrix in $\text{Var}[\mathbf{b},c] = s^2\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{z} \\ \mathbf{z}'\mathbf{X} & \mathbf{z}'\mathbf{X} \end{bmatrix}^{-1}$. But, we have already found this above. The submatrix is $\text{Var}[\mathbf{b}_{.z}] = s^2(\mathbf{X}'\mathbf{M}_z\mathbf{X})^{-1}$. 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 $(\text{Var}[\mathbf{b}])^{-1} - (\text{Var}[\mathbf{b}_{.z}])^{-1} = (1/\sigma^2)\mathbf{z}(\mathbf{z}'\mathbf{z})^{-1}\mathbf{z}'$ which is a nonnegative definite matrix. Therefore $\text{Var}[\mathbf{b}]^{-1}$ is larger than $\text{Var}[\mathbf{b}_{.z}]^{-1}$, which implies that $\text{Var}[\mathbf{b}]$ 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 $s^2 = \mathbf{e}'\mathbf{e}/(n - K)$ 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, $\text{Est.Var}[\mathbf{b}] = s^2(\mathbf{X}'\mathbf{X})^{-1}$ is based on a smaller matrix, but a larger scale factor than $\text{Est.Var}[\mathbf{b}_{.z}] = s_z^2(\mathbf{X}'\mathbf{M}_z\mathbf{X})^{-1}$. 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, $\mathbf{XNM}_z\mathbf{X}$ 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 $[\mathbf{b'X'Xb}/K]/[\mathbf{e'e}/(n - K)]$. We substitute $\mathbf{e} = \mathbf{M}$, and $\mathbf{b} = \beta + (\mathbf{X'X})^{-1}\mathbf{X'}\varepsilon = (\mathbf{X'X})^{-1}\mathbf{X'}\varepsilon$. Then, $F = [\varepsilon'\mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'X}(\mathbf{X'X})^{-1}\mathbf{X'}\varepsilon/K]/[\varepsilon'\mathbf{M}\varepsilon/(n - K)] = [\varepsilon'(\mathbf{I - M})\varepsilon/K]/[\varepsilon'\mathbf{M}\varepsilon/(n - K)]$.

The exact expectation of F can be found as follows: $F = [(n\text{-}K)/K][\varepsilon'(\mathbf{I - M})\varepsilon]/[\varepsilon'\mathbf{M}\varepsilon]$. 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 $\mathbf{M}(\mathbf{I - M}) = \mathbf{0}$. Thus, $E\{[\varepsilon(\mathbf{I - M})\varepsilon]/[\varepsilon'\mathbf{M}\varepsilon]\} = E[\varepsilon'(\mathbf{I- M})\varepsilon]\times E\{1/[\varepsilon'\mathbf{M}\varepsilon]\}$. The first of these was obtained above, $E[\varepsilon'(\mathbf{I - M})\varepsilon] = K\sigma^2$. 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 $E[\mathbf{b'b}] = \beta'\beta + \sigma^2\Sigma_k (1/\lambda_k)$ where **b** is the ordinary least squares estimator and λ_k is a characteristic root of **X′X**.

We write $\mathbf{b} = \beta + (\mathbf{X'X})^{-1}\mathbf{X'}\varepsilon$, so $\mathbf{b'b} = \beta'\beta + \varepsilon'\mathbf{X}(\mathbf{X'X})^{-1}(\mathbf{X'X})^{-1}\mathbf{X'}\varepsilon + 2\beta'(\mathbf{X'X})^{-1}\mathrm{X}'\varepsilon$. 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,

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

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 $n\to\infty$, 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 $n\to\infty$, $\bar{y}$ 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 $n\to\infty$, 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 $\text{Asy.Var}[\hat{\mu}] = (1/n)\lim_{n\to\infty} \text{Var}[\sqrt{n}(\hat{\mu} - \mu)]$. For the estimator above, we can use $\text{Asy.Var}[\hat{\mu}] = (1/n)\lim_{n\to\infty} n\text{Var}[\hat{\mu} - \mu] = (1/n)\lim_{n\to\infty} \sigma^2[2(n^2 + 3n/2 + 1/2)]/[1.5(n^2 + 2n + 1)] = 1.3333\sigma^2$. 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** = $\theta = \beta + \mathbf{Q}^{-1}\gamma$. 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 $\mathbf{b} = \beta + \mathbf{Q}^{-1}\gamma$. 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 $\gamma = \mathbf{0}$, 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 $[\partial Q^*/\partial\beta \;\; \partial Q^*/\partial\gamma]$ 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**: ($\mathbf{x}_i,\varepsilon$) 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 $E[x_jx_kx_lx_m] = \phi_{jklm}$ 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 $(1/n)\mathbf{X}'\boldsymbol{\varepsilon} = 0$ continue to hold, then consistency can be established by the Slutsky Theorem.

Chapter 6

Inference and Prediction

1. A multiple regression of y on a constant, x_1, and x_2 produces the results below:

$y = 4 + .4x_1 + .9x_2$, $R^2 = 8/60$, $\mathbf{e'e} = 520$, $n = 29$, $\mathbf{X'X} = \begin{bmatrix} 29 & 0 & 0 \\ 0 & 50 & 10 \\ 0 & 10 & 80 \end{bmatrix}$. Test the hypothesis that the two slopes sum to 1.

The estimated covariance matrix for the least squares estimates is

$s^2(\mathbf{X'X})^{-1} = \dfrac{20}{3900}\begin{bmatrix} 3900/29 & 0 & 0 \\ 0 & 80 & -10 \\ 0 & -10 & 80 \end{bmatrix} = \begin{bmatrix} .69 & 0 & 0 \\ 0 & .40 & -.051 \\ 0 & -.051 & .256 \end{bmatrix}$ where $s^2 = 520/(29\text{-}3) = 20$. Then, the test may be based on $t = (.4 + .9 - 1)/[.410 + .256 - 2(.051)]^{1/2} = .399$. This is smaller than the critical value of 2.056, so we would not reject the hypothesis.

2. . Using the results in Exercise 1, test the hypothesis that the slope on x_1 is zero by running the restricted regression and comparing the two sums of squared deviations.

In order to compute the regression, we must recover the original sums of squares and cross products for y. These are$\mathbf{X'y} = \mathbf{X'Xb} = [116, 29, 76]'$. The total sum of squares is found using $R^2 = 1 - \mathbf{e'e}/\mathbf{y'M}^0\mathbf{y}$, so $\mathbf{y'M}^0\mathbf{y} = 520 / (52/60) = 600$. The means are $\bar{x}_1 = 0$, $\bar{x}_2 = 0$, $\bar{y} = 4$, so, $\mathbf{y'y} = 600 + 29(4^2) = 1064$. The slope in the regression of $\mathbf{y}$ on $\mathbf{x}_2$ alone is $b_2 = 76/80$, so the regression sum of squares is $b_2^2(80) = 72.2$, and the residual sum of squares is $600 - 72.2 = 527.8$. The test based on the residual sum of squares is $F = [(527.8 - 520)/1]/[520/26] = .390$. In the regression of the previous problem, the t-ratio for testing the same hypothesis would be $t = .4/(.410)^{1/2} = .624$ which is the square root of .39.

3. The regression model to be analyzed is $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 +$ **,** where $\mathbf{X}_1$ and $\mathbf{X}_2$ have K_1 and K_2 columns, respectively. The restriction is $\boldsymbol{\beta}_2 = \mathbf{0}$.

(a) Using (6-14), prove that the restricted estimator is simply $[\mathbf{b}_1',\mathbf{0}']'$ where $\mathbf{b}_1$ is the least squares coefficient vector in the regression of $\mathbf{y}$ on $\mathbf{X}_1$.

(b) Prove that if the restriction is $\boldsymbol{\beta}_2 = \boldsymbol{\beta}_2^0$ for a nonzero$\boldsymbol{\beta}_2^0$, the restricted estimator of $\boldsymbol{\beta}_1$ is $\mathbf{b}_{1*} = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'(\mathbf{y} - \mathbf{X}_2\boldsymbol{\beta})$.

For the current problem, $\mathbf{R} = [\mathbf{0},\mathbf{I}]$ where $\mathbf{I}$ is the last K_2 columns. Therefore, $\mathbf{R}(\mathbf{X'X})^{-1}\mathbf{R}$N is the lower right $K_2{\times}K_2$ block of $(\mathbf{X'X})^{-1}$. As we have seen before, this is $(\mathbf{X}_2'\mathbf{M}_1\mathbf{X}2)^{-1}$. Also, $(\mathbf{X'X})^{-1}\mathbf{R}'$ is the last K_2 columns of $(\mathbf{X'X})^{-1}$. These are $(\mathbf{X'X})^{-1}\mathbf{R}' = \begin{bmatrix} -(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2(\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2)^{-1} \\ (\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2)^{-1} \end{bmatrix}$ [See (2-74).] Finally, since $\mathbf{q} = \mathbf{0}$, $\mathbf{Rb} - \mathbf{q} = (\mathbf{0b}_1 + \mathbf{Ib}_2) - \mathbf{0} = \mathbf{b}_2$. Therefore, the constrained estimator is

$\mathbf{b}_* = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} - \begin{bmatrix} -(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2(\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2)^{-1} \\ (\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2)^{-1} \end{bmatrix}(\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2)\mathbf{b}_2$, where $\mathbf{b}_1$ and $\mathbf{b}_2$ are the multiple regression coefficients in the regression of $\mathbf{y}$ on both $\mathbf{X}_1$ and $\mathbf{X}_2$. (See Section 6.4.3 on partitioned regression.) Collecting terms, this produces $\mathbf{b}_* = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} - \begin{bmatrix} -(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\mathbf{b}_2 \\ \mathbf{b}_2 \end{bmatrix}$. But, we have from Section 6.3.4 that $\mathbf{b}_1 = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y} - (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\mathbf{b}_2$ so the preceding reduces to $\mathbf{b}_* = \begin{bmatrix} (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y} \\ \mathbf{0} \end{bmatrix}$ which was to be shown.

If, instead, the restriction is $\beta_2 = \beta_2^0$ then the preceding is changed by replacing $\mathbf{R}\beta - \mathbf{q} = \mathbf{0}$ with $\mathbf{R}\beta - \beta_2^0 = \mathbf{0}$. Thus, $\mathbf{Rb} - \mathbf{q} = \mathbf{b}_2 - \beta_2^0$. Then, the constrained estimator is

$$\mathbf{b}_* = \begin{bmatrix}\mathbf{b_1}\\ \mathbf{b_2}\end{bmatrix} - \begin{bmatrix}-(\mathbf{X_1}'\mathbf{X_1})^{-1}\mathbf{X_1}'\mathbf{X_2}(\mathbf{X_2}'\mathbf{M_1}\mathbf{X_2})^{-1}\\ (\mathbf{X_2}'\mathbf{M_1}\mathbf{X_2})^{-1}\end{bmatrix}(\mathbf{X_2}'\mathbf{M_1}\mathbf{X_2})(\mathbf{b_2} - \beta_2^0)$$

or

$$\mathbf{b}_* = \begin{bmatrix}\mathbf{b_1}\\ \mathbf{b_2}\end{bmatrix} + \begin{bmatrix}(\mathbf{X_1}'\mathbf{X_1})^{-1}\mathbf{X_1}'\mathbf{X_2}(\mathbf{b_2} - \beta_2^0)\\ (\beta_2^0 - \mathbf{b_2})\end{bmatrix}$$

Using the result of the previous paragraph, we can rewrite the first part as

$$\mathbf{b}_{1*} = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y} - (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\beta_2^0 = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'(\mathbf{y} - \mathbf{X}_2\beta_2^0)$$

which was to be shown.

4. The expression for the restricted coefficient vector in (6-14) may be written in the form $\mathbf{b}_* = [\mathbf{I} - \mathbf{CR}]\mathbf{b} + \mathbf{w}$, where $\mathbf{w}$ does not involve $\mathbf{b}$. What is $\mathbf{C}$? Show that covariance matrix of the restricted least squares estimator is $\sigma^2(\mathbf{X}'\mathbf{X})^{-1} - \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}$ and that this matrix may be written as $\text{Var}[\mathbf{b}]\{[\text{Var}(\mathbf{b})]^{-1} - \mathbf{R}'[\text{Var}(\mathbf{Rb})]^{-1}\mathbf{R}\}\text{Var}[\mathbf{b}]$

By factoring the result in (6-14), we obtain $\mathbf{b}_* = [\mathbf{I} - \mathbf{CR}]\mathbf{b} + \mathbf{w}$ where $\mathbf{C} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}$ and $\mathbf{w} = \mathbf{Cq}$. The covariance matrix of the least squares estimator is

$$\begin{aligned}\text{Var}[\mathbf{b}_*] &= [\mathbf{I} - \mathbf{CR}]\sigma^2(\mathbf{X}'\mathbf{X})^{-1}[\mathbf{I} - \mathbf{CR}]'\\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1} + \sigma^2\mathbf{CR}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\mathbf{C}' - \sigma^2\mathbf{CR}(\mathbf{X}'\mathbf{X})^{-1} - \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\mathbf{C}'.\end{aligned}$$

By multiplying it out, we find that $\mathbf{CR}(\mathbf{X}'\mathbf{X})^{-1} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')^{-1}\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1} = \mathbf{CR}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\mathbf{C}'$

so $\text{Var}[\mathbf{b}_*] = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} - \sigma^2\mathbf{CR}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\mathbf{C}' = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} - \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}$

This may also be written as

$$\begin{aligned}\text{Var}[\mathbf{b}_*] &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\{\mathbf{I} - \mathbf{R}'(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')^{-1}\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\}\\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\{[\sigma^2(\mathbf{X}'\mathbf{X})^{-1}]^{-1} - \mathbf{R}'[\mathbf{R}\sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}\mathbf{R}\}\sigma^2(\mathbf{X}'\mathbf{X})^{-1}\end{aligned}$$

Since $\text{Var}[\mathbf{Rb}] = \mathbf{R}\sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'$ this is the answer we seek.

5. Prove the result that the restricted least squares estimator never has a larger variance matrix than the unrestricted least squares estimator.

The variance of the restricted least squares estimator is given in the second equation in the previous exercise. We know that this matrix is positive definite, since it is derived in the form $\mathbf{B}'\sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{B}'$, and $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ is positive definite. Therefore, it remains to show only that the matrix subtracted from $\text{Var}[\mathbf{b}]$ to obtain $\text{Var}[\mathbf{b}_*]$ is positive definite. Consider, then, a quadratic form in $\text{Var}[\mathbf{b}_*]$

$$\begin{aligned}\mathbf{z}'\text{Var}[\mathbf{b}_*]\mathbf{z} &= \mathbf{z}'\text{Var}[\mathbf{b}]\mathbf{z} - \sigma^2\mathbf{z}'(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}\mathbf{R})(\mathbf{X}'\mathbf{X})^{-1}\mathbf{z}\\ &= \mathbf{z}'\text{Var}[\mathbf{b}]\mathbf{z} - \mathbf{w}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}\mathbf{w} \quad \text{where } \mathbf{w} = \sigma\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{z}.\end{aligned}$$

It remains to show, therefore, that the inverse matrix in brackets is positive definite. This is obvious since its inverse is positive definite. This shows that every quadratic form in $\text{Var}[\mathbf{b}_*]$ is less than a quadratic form in $\text{Var}[\mathbf{b}]$ in the same vector.

6. Prove the result that the R^2 associated with a restricted least squares estimator is never larger than that associated with the unrestricted least squares estimator. Conclude that imposing restrictions never improves the fit of the regression.

The result follows immediately from the result which precedes (6-19). Since the sum of squared residuals must be at least as large, the coefficient of determination, $COD = 1 - \text{sum of squares} / \Sigma_i (y_i - \bar{y})^2$, must be no larger.

7. The Lagrange multiplier test of the hypothesis $\mathbf{R}\beta\text{-}\mathbf{q}=\mathbf{0}$ is equivalent to a Wald test of the hypothesis that $\lambda = \mathbf{0}$, where λ is defined in (6-14). Prove that $\chi^2 = \lambda'\{\text{Est.Var}[\lambda]\}^{-1}\lambda = (n\text{-}K)[\mathbf{e}_*'\mathbf{e}_*/\mathbf{e}'\mathbf{e} - 1]$. Note that the fraction in brackets is the ratio of two estimators of σ^2. By virtue of (6-15) and the preceding section, we know that this is greater than 1. Finally, prove that the Lagrange multiplier statistic is simply JF, where J is the number of restrictions being tested and F is the conventional F statistic given in (6-20).

For convenience, let $\mathbf{F} = [\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}$. Then, $\lambda = \mathbf{F}(\mathbf{Rb} - \mathbf{q})$ and the variance of the vector of Lagrange multipliers is Var[8] $= \mathbf{F}\mathbf{R}\sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\mathbf{F} = \sigma^2\mathbf{F}$. The estimated variance is obtained by replacing σ^2 with s^2. Therefore, the chi-squared statistic is

$$\chi^2 = (\mathbf{Rb} - \mathbf{q})'\mathbf{F}'(s^2\mathbf{F})^{-1}\mathbf{F}(\mathbf{Rb} - \mathbf{q}) = (\mathbf{Rb} - \mathbf{q})'[(1/s^2)\mathbf{F}](\mathbf{Rb} - \mathbf{q})$$
$$= (\mathbf{Rb} - \mathbf{q})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{Rb} - \mathbf{q})/[\mathbf{e}'\mathbf{e}/(n - K)]$$

This is exactly J times the F statistic defined in (6-19) and (6-20). Finally, J times the F statistic in (6-20) equals the expression given above.

8. Use the Lagrange multiplier test to test the hypothesis in Exercise 1.

We use (6-19) to find the new sum of squares. The change in the sum of squares is

$$\mathbf{e}_*'\mathbf{e}_* - \mathbf{e}'\mathbf{e} = (\mathbf{Rb} - \mathbf{q})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{Rb} - \mathbf{q})$$

For this problem, $(\mathbf{Rb} - \mathbf{q}) = b_2 + b_3 - 1 = .3$. The matrix inside the brackets is the sum of the 4 elements in the lower right block of $(\mathbf{X}'\mathbf{X})^{-1}$. These are given in Exercise 1, multiplied by $s^2 = 20$. Therefore, the required sum is $[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'] = (1/20)(.410 + .256 - 2(.051)) = .028$. Then, the change in the sum of squares is $.3^2 / .028 = 3.215$. Thus, $\mathbf{e}'\mathbf{e} = 520$, $\mathbf{e}_*'\mathbf{e}_* = 523.215$, and the chi-squared statistic is $26[523.215/520 - 1] = .16$. This is quite small, and would not lead to rejection of the hypothesis. Note that for a single restriction, the Lagrange multiplier statistic is equal to the F statistic which equals, in turn, the square of the t statistic used to test the restriction. Thus, we could have obtained this quantity by squaring the .399 found in the first problem (apart from some rounding error).

9. Using the data and model of Example 2.3, carry out a test of the hypothesis that the three aggregate price indices are not significant determinants of the demand for gasoline.

The sums of squared residuals for the two regressions are 207.644 when the aggregate price indices are included and 586.596 when they are excluded. The F statistic is $F = [(586.596 - 207.644)/3]/[207.644/17] = 10.342$. The critical value from the F table is 3.20, so we would reject the hypothesis.

10. The model of Example 2.3 may be written in logarithmic terms as

$$\ln G/Pop = \alpha + \beta_p\ln P_g + \beta_y\ln Y + \gamma_{nc}\ln P_{nc} + \gamma_{uc}\ln P_{uc} + \gamma_{pt}\ln P_{pt} + \beta_t year + \delta_d\ln P_d + \delta_n\ln P_n + \delta_s\ln P_s + \varepsilon.$$

Consider the hypothesis that the micro elasticities are a constant proportion of the elasticity with respect to their corresponding aggregate. Thus, for some positive 2 (presumably between 0 and 1),

$$\gamma_{nc} = 2\delta_d, \gamma_{uc} = 2\delta_d, \gamma_{pt} = 2\delta_s.$$

The first two imply the simple linear restriction $\gamma_{nc} = \gamma_{uc}$. Taking ratios, the first (or second) and third imply the nonlinear restriction $\gamma_{nc}/\gamma_{pt} = \delta_d/\delta_s$.

(a) Describe in detail how you would test the validity of the restriction.

(b) Using the gasoline market data in Table F2.2, test the restrictions separately and jointly.

Since the restricted model is quite nonlinear, it would be quite cumbersome to estimate and examine the loss in fit. We can test the restriction using the unrestricted model. For this problem,

$$\mathbf{f} = [\gamma_{nc} - \gamma_{uc}, \gamma_{nc}\delta_s - \gamma_{pt}\delta_d]'$$

The matrix of derivatives, using the order given above and " to represent the entire parameter vector, is

$$\mathbf{G} = \begin{bmatrix} \partial f_1 / \partial\alpha \\ \partial f_2 / \partial\alpha \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \delta_s & 0 & -\delta_d & 0 & -\gamma_{pt} & 0 & \gamma_{nc} \end{bmatrix}.$$ The parameter estimates are

$a = 18.5454$, $c_{uc} = -.201536$, $d_d = 1.50607$, $b_p = -.581437$, $c_{pt} = .0805074$, $d_n = .999474$, $b_y = 1.39438$, $b_t = -.0125129$, $d_s = -.817896$, $c_{nc} = -.294769$.

Thus, $\mathbf{f} = [-.092322, .119841]'$. The covariance matrix to use for the tests is

$$\mathbf{G}s^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{G}' = \begin{bmatrix} .053285 & -.0362998 \\ -.0362998 & .0342649 \end{bmatrix}$$

The statistic for the joint test is $\chi^2 = \mathbf{f}'[\mathbf{G}s^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{G}']^{-1}\mathbf{f} = .5789$. This is less than the critical value for a chi-squared with two degrees of freedom, so we would not reject the joint hypothesis. For the individual hypotheses, we need only compute the equivalent of a t ratio for each element of $\mathbf{f}$. Thus,

$$z_1 = -.092322/(.053285)^2 = .3999$$

and

$$z_2 = .119841/(.0342649)^2 = .6474.$$

Neither is large, so neither hypothesis would be rejected. (Given the earlier result, this was to be expected.).

11. Prove that under the hypothesis that $\mathbf{R}\boldsymbol{\beta} = \mathbf{q}$, the estimator $s = (\mathbf{y} - \mathbf{X}\mathbf{b}_*)'(\mathbf{y} - \mathbf{X}\mathbf{b}_*)/(n - K + J)$, where J is the number of restrictions, is unbiased for σ^2.

First, use (6-19) to write $\mathbf{e}_*'\mathbf{e}_* = \mathbf{e}'\mathbf{e} + (\mathbf{R}\mathbf{b} - \mathbf{q})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\mathbf{b} - \mathbf{q})$. Now, the result that $E[\mathbf{e}'\mathbf{e}] = (n - K)\sigma^2$ obtained in Chapter 6 must hold here, so $E[\mathbf{e}_*'\mathbf{e}_*] = (n - K)\sigma^2 + E[(\mathbf{R}\mathbf{b} - \mathbf{q})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\mathbf{b} - \mathbf{q})]$. Now, $\mathbf{b} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}$, so $\mathbf{R}\mathbf{b} - \mathbf{q} = \mathbf{R}\boldsymbol{\beta} - \mathbf{q} + \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}$. But, $\mathbf{R}\boldsymbol{\beta} - \mathbf{q} = \mathbf{0}$, so under the hypothesis, $\mathbf{R}\mathbf{b} - \mathbf{q} = \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}$. Insert this in the result above to obtain
$E[\mathbf{e}_*'\mathbf{e}_*] = (n-K)\sigma^2 + E[\boldsymbol{\varepsilon}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}]$. The quantity in square brackets is a scalar, so it is equal to its trace. Permute $\boldsymbol{\varepsilon}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'$ in the trace to obtain

$$E[\mathbf{e}_*'\mathbf{e}_*] = (n - K)\sigma^2 + E[tr\{[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']\}$$

We may now carry the expectation inside the trace and use $E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] = \sigma^2\mathbf{I}$ to obtain

$$E[\mathbf{e}_*'\mathbf{e}_*] = (n - K)\sigma^2 + \text{tr}\{[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{I}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']\}$$

Carry the σ^2 outside the trace operator, and after cancellation of the products of matrices times their inverses, we obtain $\quad E[\mathbf{e}_*'\mathbf{e}_*] = (n - K)\sigma^2 + \sigma^2\text{tr}[\mathbf{I}_J] = (n - K + J)\sigma^2$.

12. Show that in the multiple regression of $\mathbf{y}$ on a constant, $\mathbf{x}_1$, and $\mathbf{x}_2$, while imposing the restriction $\beta_1 + \beta_2 = 1$ leads to the regression of $\mathbf{y} - \mathbf{x}_1$ on a constant and $\mathbf{x}_2 - \mathbf{x}_1$.

For convenience, we put the constant term last instead of first in the parameter vector. The constraint is $\mathbf{R}\mathbf{b} - \mathbf{q} = \mathbf{0}$ where $\mathbf{R} = [1\ \ 1\ \ 0]$ so $\mathbf{R}_1 = [1]$ and $\mathbf{R}_2 = [1,0]$. Then, $\beta_1 = [1]^{-1}[1 - \beta_2] = 1 - \beta_2$. Thus, $\mathbf{y} = (1 - \beta_2)\mathbf{x}_1 + \beta_2\mathbf{x}_2 + \alpha\mathbf{i} + \boldsymbol{\varepsilon}$ or $\mathbf{y} - \mathbf{x}_1 = \beta_2(\mathbf{x}_2 - \mathbf{x}_1) + \alpha\mathbf{i} + \boldsymbol{\varepsilon}$.

Chapter 7

Functional Form and Structural Change

1. In Solow's classic (1957) study of technical change in the U.S. Economy, he suggests the following aggregate production function: $q(t) = A(t)f[k(t)]$ where $q(t)$ is aggregate output per manhour, $k(t)$ is the aggregate capital labor ratio, and $A(t)$ is the technology index. Solow considered four static models,

$$q/A = \alpha + \beta \ln k,\quad q/A = \alpha - \beta/k,\quad \ln(q/A) = \alpha + \beta \ln k,\quad \ln(q/A) = \alpha - \beta/k.$$

(He also estimated a dynamic model, $q(t)/A(t) - q(t-1)/A(t-1) = \alpha + \beta k$.)

(a) Sketch the four functions.

(b) Solow's data for the years 1909 to 1949 are listed in Table A8.1: (Op. cit., page 314. Several variables are omitted.) Use these data to estimate the α and β of the four functions listed above. (Note, your results will not quite match Solow's. See the next problem for resolution of the discrepancy.) Sketch the functions using your particular estimates of the parameters.

The least squares estimates of the four models are

$$q/A = .45237 + .23815\ln k$$
$$q/A = .91967 - .61863/k$$
$$\ln(q/A) = -.72274 + .35160\ln k$$
$$\ln(q/A) = -.032194 - .91496/k$$

At these parameter values, the four functions are nearly identical. A plot of the four sets of predictions from the regressions and the actual values appears below.

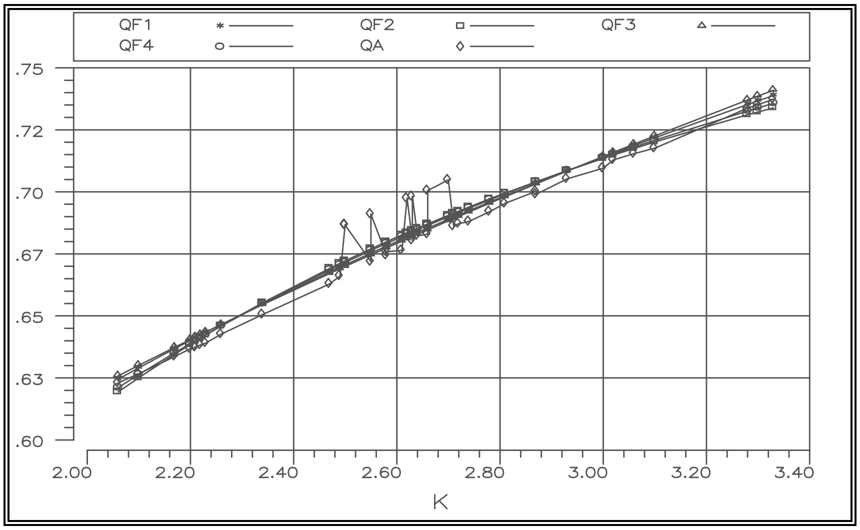
2. In the aforementioned study, Solow states

"A scatter of q/A against k is shown in Chart 4. Considering the amount of a priori doctoring which the raw figures have undergone, the fit is remarkably tight. Except, that is, for the layer of points which are obviously too high. These maverick observations relate to the seven last years of the period, 1943-1949. From the way they lie almost exactly parallel to the main scatter, one is tempted to conclude that in 1943 the aggregate production function simply shifted.

(a) Draw a scatter diagram of q/A against k. [Or, obtain Solow's original study and examine his. An alternative source of the original paper is the volume edited by A. Zellner (1968).]

(b) Estimate the four models you estimated in the previous problem including a dummy variable for the years 1943 to 1949. How do your results change? (Note, these results match those reported by Solow, though he did not report the coefficient on the dummy variable.)

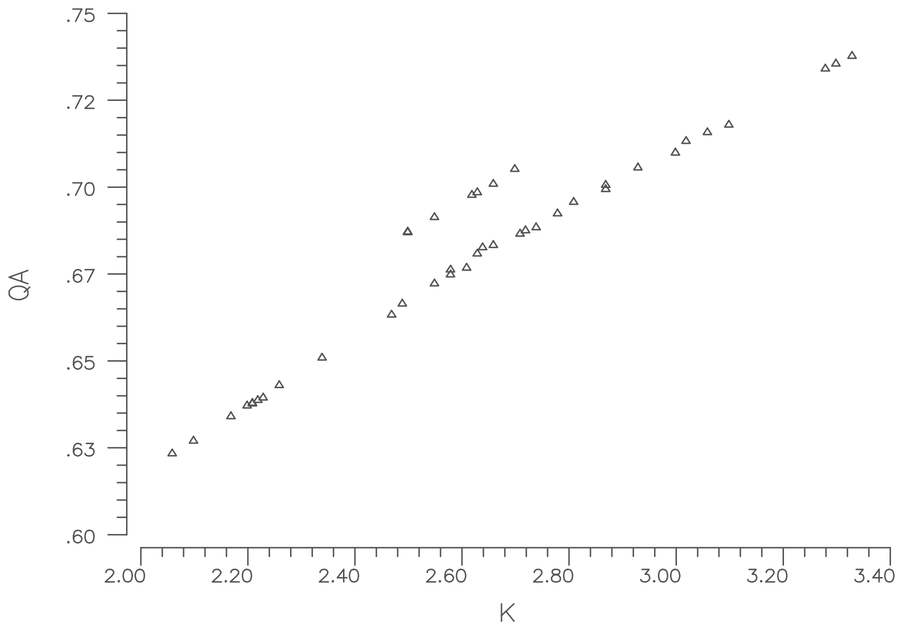
(c) Solow went on to surmise that, in fact, the data were fundamentally different in the years before 1943 than during and after. If so, one would guess that the regression should be as well (though whether the change is merely in the data or in the underlying production function is not settled). Use a Chow test to examine the difference in the two subperiods using your four functional forms. Note that with the dummy variable, you can do the test by introducing an interaction term between the dummy and whichever function of k appears in the regression. Use an F test to test the hypothesis.



The regression results for the various models are listed below. (d is the dummy variable equal to 1 for the last seven years of the data set. Standard errors for parameter estimates are given in parentheses.)

α	β	γ	δ	R^2	$\mathbf{e'e}$
Model 1: $q/A = \alpha + \beta \ln k + \gamma d + \delta(d \ln k) + \varepsilon$					
.4524	.2381			.94355	.00213
(.00903)	(.00932)				
.4477	.2396	.01900		.99914	.000032
(.00113)	(.00117)	(.000384)			
.4476	.2397	.02746	-.08883	.99915	.000032
(.00115)	(.00118)	(.0119)	(.0126)		
Model 2: $q/A = \alpha - \beta(1/k) + \gamma d + \delta(d/k) + \varepsilon$					
.9168	.6186			.94915	.001915
(.00891)	(.0229)				
.9167	.6185	.01961		.99321	.000256
(.00331)	(.00849)	(.00108)			
.9168	.6187	.008651	.02140	.99322	.000255
(.00336)	(.00863)	(.0354)	(.0917)		
Model 3: $\ln(q/A) = \alpha + \beta \ln k + \gamma d + \delta(d \ln k) + \varepsilon$					
-.7227	.3516			.94069	.004882
(.0137)	(.0141)				
-.7298	.3538	.002881		.99918	.000068
(.00164)	(.00169)	(.000554)			
-.7300	.3540	.04961	-.02182	.99921	.000065
(.00164)	(.00148)	(.0171)	(.0179)		
Model 4: $\ln(q/A) = \alpha - \beta(1/k) + \gamma d + \delta(d/k) + \varepsilon$					
-.03219	.9150			.94964	.004146
(.0131)	(.0337)				
-.03665	.9148	.02572		.99629	.000305
(.00361)	(.00928)	(.00118)			
-.03646	.9153	.004290	.05556	.99632	.000303
(.00366)	(.00941)	(.0386)	(.0999)		

The scatter diagram is shown below.



The last seven years of the data set show clearly the effect observed by Solow.

For the four models, the F test of the third specification against the first is equivalent to the Chow-test. The statistics are:

Model 1: F = (.002126 - .000032)/2 / (.000032/37) = 1210.6
Model 2: F = = 120.43
Model 3: F = = 1371.0
Model 4: F = = 234.64

The critical value from the F table for 2 and 37 degrees of freedom is 3.26, so all of these are statistically significant. The hypothesis that the same model applies in both subperiods must be rejected.

3. A regression model with $K = 16$ independent variables is fit using a panel of 7 years of data. The sums of squares for the seven separate regressions and the pooled regression are shown below. The model with the pooled data allows a separate constant for each year. Test the hypothesis that the same coefficients apply in every year.

	1954	***1955***	***1956***	***1957***	***1958***	***1959***	***1960***	***All***
Observations	65	55	87	95	103	87	78	570
$\mathbf{e}'\mathbf{e}$	104	88	206	144	199	308	211	1425

The F statistic could be computed as

F = {[1425 - (104 + 88 + ... + 211)] / (70 - 16)}/[(104 + 88 + ... + 211) / (570 - 70)] = 1.343

The 95% critical value for the F distribution with 54 and 500 degrees of freedom is 1.363.

4. *Reverse Regression* A common method of analyzing statistical data to detect discrimination in the workplace is to fit the following regression:

(1) $$y = \alpha + \boldsymbol{\beta}'\mathbf{x} + \gamma d + \varepsilon,$$

where y is the wage rate and d is a dummy variable indicating either membership (d=1) or nonmembership (d=0) in the class toward which it is suggested the discrimination is directed. The regressors, $\mathbf{x}$, include factors specific to the particular type of job as well as indicators of the qualifications of the individual. The hypothesis of interest is H_0: $\gamma < 0$ vs. H_1: $\gamma = 0$. The regression seeks to answer the question "in a given job, are individuals in the class (d=1) paid less than equally qualified individuals not in the class (d=0)?" Consider, however, the alternative possibility. Do individuals in the class in the same job as others, and receiving the same wage, uniformly have higher qualifications? If so, this might also be viewed as a form of discrimination. To analyze this question, Conway and Roberts (1983) suggested the following procedure:

(a) Fit (1) by ordinary least squares. Denote the estimates a, $\mathbf{b}$, and c.

(b) Compute the set of *qualification indices*,

(2) $$\mathbf{q} = a\mathbf{i} + \mathbf{Xb}.$$

Note the omission of $c\mathbf{d}$ from the fitted value.

(c) Regress $\mathbf{q}$ on a constant, $\mathbf{y}$, and $\mathbf{d}$. The equation is

(3) $$\mathbf{q} = \alpha_* + \beta_*\mathbf{y} + \gamma_*\mathbf{d} + \varepsilon_*.$$

The analysis suggests that if $\gamma < 0$, $\gamma_* > 0$.

(1) Prove that the theory notwithstanding, the least squares estimates, c and c_* are related by

(4) $$c_* = \frac{(\bar{y}_1 - \bar{y})(1-R^2)}{(1-P)(1-r_{yd}^2)} - c$$

where $\bar{y}_1$ is the mean of y for observations with $d = 1$,
$\bar{y}$ is the mean of y for all observations,
P is the mean of d,
R^2 is the coefficient of determination for (1)
and r_{yd}^2 is the squared correlation between y and d.

[**Hint:** The model contains a constant term. Thus, to simplify the algebra, assume that all variables are measured as deviations from the overall sample means and use a partitioned regression to compute the coefficients in (3). Second, in (2), use the fact that based on the least squares results,

$$\mathbf{y} = a\mathbf{i} + \mathbf{Xb} + c\mathbf{d} + \mathbf{e},$$

so $$\mathbf{q} = \mathbf{y} - c\mathbf{d} - \mathbf{e}.$$

From here on, we drop the constant term.] Thus, in the regression in (c), you are regressing [$\mathbf{y}$ - $c\mathbf{d}$ - $\mathbf{e}$] on $\mathbf{y}$ and $\mathbf{d}$. Remember, all variables are in deviation form.

(2) Will the sample evidence necessarily be consistent with the theory? [**Hint:** suppose $c = 0$?]

Using the hint, we seek the c_* which is the slope on $\mathbf{d}$ in the regression of $\mathbf{q} = \mathbf{y} - c\mathbf{d} - \mathbf{e}$ on $\mathbf{y}$ and $\mathbf{d}$. The regression coefficients are $\begin{bmatrix}\mathbf{y'y} & \mathbf{y'd}\\ \mathbf{d'y} & \mathbf{d'd}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{y'}(\mathbf{y}-c\mathbf{d}-\mathbf{e})\\ \mathbf{d'}(\mathbf{y}-c\mathbf{d}-\mathbf{e})\end{bmatrix} = \begin{bmatrix}\mathbf{y'y} & \mathbf{y'd}\\ \mathbf{d'y} & \mathbf{d'd}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{y'y}-c\mathbf{y'd}-\mathbf{y'e}\\ \mathbf{d'y}-c\mathbf{d'd}-\mathbf{d'e}\end{bmatrix}$. In the preceding, note that $(\mathbf{y'y},\mathbf{d'y})'$ is the first column of the matrix being inverted while $c(\mathbf{y'd},\mathbf{d'd})'$ is c times the second. An inverse matrix times the first column of the original matrix is the first column of an identity matrix, and likewise for the second. Also, since $\mathbf{d}$ was one of the original regressors in (1), $\mathbf{d'e} = 0$, and, of course, $\mathbf{y'e} = \mathbf{e'e}$. If we combine all of these, the coefficient vector is

$$-\begin{pmatrix}1\\0\end{pmatrix} - c\begin{pmatrix}0\\1\end{pmatrix} - \begin{bmatrix}\mathbf{y'y} & \mathbf{y'd}\\ \mathbf{d'y} & \mathbf{d'd}\end{bmatrix}^{-1}\begin{pmatrix}\mathbf{e'e}\\0\end{pmatrix} = -\begin{pmatrix}1\\0\end{pmatrix} - c\begin{pmatrix}0\\1\end{pmatrix} - \begin{bmatrix}\mathbf{y'y} & \mathbf{y'd}\\ \mathbf{d'y} & \mathbf{d'd}\end{bmatrix}^{-1}\begin{pmatrix}1\\0\end{pmatrix}\mathbf{e'e}.$$

We are interested in the second (lower) of the two coefficients. The matrix product at the end is $\mathbf{e'e}$ times the first column of the inverse matrix, and we wish to find its second (bottom) element. Therefore, collecting what we have thus far, the desired coefficient is $c_* = -c - \mathbf{e'e}$ times the off diagonal element in the inverse matrix. The off diagonal element is

$$-\mathbf{d'y} / [(\mathbf{y'y})(\mathbf{d'd}) - (\mathbf{y'd})^2] = -\mathbf{d'y} / \{[(\mathbf{y'y})(\mathbf{d'd})][1 - (\mathbf{y'd})^2/[(\mathbf{y'y})(\mathbf{d'd})]]\}$$
$$= -\mathbf{d'y} / [(\mathbf{y'y})(\mathbf{d'd})(1 - r_{yd}^2)].$$

Therefore, $$c_* = [(\mathbf{e'e})(\mathbf{d'y})] / [(\mathbf{y'y})(\mathbf{d'd})(1 - r_{yd}^2)] - c$$

(The two negative signs cancel.) This can be further reduced. Since all variables are in deviation form, $\mathbf{e'e}/\mathbf{y'y}$ is $(1 - R^2)$ in the full regression. By multiplying it out, you can show that $\bar{d} = P$ so that

$$\mathbf{d'd} = \Sigma_i(d_i - P)^2 = nP(1-P)$$

and $$\mathbf{d'y} = \Sigma_i(d_i - P)(y_i - \bar{y}) = \Sigma_i(d_i - P)y_i = n_1(\bar{y}_1 - \bar{y})$$

where n_1 is the number of observations which have $d_i = 1$. Combining terms once again, we have

$$c_* = \{[n_1(\bar{y}_1 - \bar{y})(1 - R^2)\} / \{nP(1-P)(1 - r_{yd}^2)\} - c$$

Finally, since $P = n_1/n$, this further simplifies to the result claimed in the problem,

$$c_* = \{(\bar{y}_1 - \bar{y})(1 - R^2)\} / \{(1\text{-}P)(1 - r_{yd}^2)\} - c$$

The problem this creates for the theory is that in the present setting, if, indeed, c is negative, $(\bar{y}_1 - \bar{y})$ will almost surely be also. Therefore, the sign of c_* is ambiguous.

5. *Reverse Regression*. This and the next exercise continue the analysis of Exercise 10, Chapter 8. In the earlier exercise, interest centered on a particular dummy variable in which the regressors were accurately measured. Here, we consider the case in which the crucial regressor in the model is measured with error. The paper by Kamlich and Polachek (1982) is directed toward this issue.

Consider the simple errors in variables model, $y = \alpha + \beta x^* + \varepsilon$, $x = x^* + u$, where u and ε are uncorrelated, and x is the erroneously measured, observed counterpart to x^*.

(a) Assume that x^*, u, and ε are all normally distributed with means μ^*, 0, and 0, variances σ_*^2, σ_u^2, and σ_ε^2 and zero covariances. Obtain the probability limits of the least squares estimates of α and β.

(b) As an alternative, consider regressing x on a constant and y, then computing the reciprocal of the estimate. Obtain the probability limit of this estimate.

(c) Do the `direct' and `reverse' estimators bound the true coefficient?

We first find the joint distribution of the observed variables. $\begin{pmatrix} y \\ x \end{pmatrix} = \begin{pmatrix} \alpha \\ 0 \end{pmatrix} + \begin{bmatrix} \beta & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}\begin{pmatrix} x^* \\ \varepsilon \\ u \end{pmatrix}$ so $[y,x]$ have a joint normal distribution with mean vector $E\begin{pmatrix} y \\ x \end{pmatrix} = \begin{pmatrix} \alpha \\ 0 \end{pmatrix} + \begin{bmatrix} \beta & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}\begin{pmatrix} \mu^* \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \alpha + \beta\mu^* \\ \mu^* \end{pmatrix}$ and covariance matrix $Var\begin{pmatrix} y \\ x \end{pmatrix} = \begin{bmatrix} \beta & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}\begin{bmatrix} \sigma_*^2 & 0 & 0 \\ 0 & \sigma_\varepsilon^2 & 0 \\ 0 & 0 & \sigma_u^2 \end{bmatrix}\begin{bmatrix} \beta & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \beta^2\sigma_*^2 + \sigma_\varepsilon^2 & \beta\sigma_*^2 \\ \beta\sigma_*^2 & \sigma_*^2 + \sigma_u^2 \end{bmatrix}$, The probability limit of the slope in the linear regression of y on x is, as usual,

$$\text{plim } b = \text{Cov}[y,x]/\text{Var}[x] = \beta/(1 + \sigma_u^2/\sigma_*^2) < \beta.$$

The probability limit of the intercept is plim

$$a = E[y] - (\text{plim } b)E[x] = \alpha + \beta\mu^* - \beta\mu^*/(1 + \sigma_u^2/\sigma_*^2)$$
$$= \alpha + \beta[\mu^*\sigma_u / (\sigma_*^2 + \sigma_u^2)] > \alpha \quad (\text{assuming } \beta > 0).$$

If x is regressed on y instead, the slope will estimate $\text{plim}[b'] = \text{Cov}[y,x]/\text{Var}[y] = \beta\sigma_*^2/(\beta^2\sigma_*^2 + \sigma_\varepsilon^2)$. Then,$\text{plim}[1/b'] = \beta + \sigma_\varepsilon^2/\beta^2\sigma_*^2 > \beta$. Therefore, b and b' will bracket the true parameter (at least in their probability limits). Unfortunately, without more information about σ_u^2, we have no idea how wide this bracket is. Of course, if the sample is large and the estimated bracket is narrow, the results will be strongly suggestive.

6. *Reverse Regression - Continued:* Suppose that the model in Exercise 5 is extended to

$$y = \beta x^* + \gamma d + \varepsilon,\ x = x^* + u.$$

For convenience, we drop the constant term. Assume that x^*, ε, and u are independent normally distributed with zero means. Suppose that d is a random variable which takes the values one and zero with probabilities π and $1\text{-}\pi$ in the population, and is independent of all other variables in the model. To put this in context, the preceding model (and variants of it) have appeared in the literature on discrimination. We view y as a "wage" variable, x^* as "qualifications" and x as some imperfect measure such as education. The dummy variable, d, is membership (d=1) or nonmembership (d=0) in some protected class. The hypothesis of discrimination turns on $\gamma<0$ versus $\gamma=0$.

(a) What is the probability limit of c, the least squares estimator of (, in the least squares regression of y on x and d? [**Hints:** The independence of x^* and d is important. Also, plim $\mathbf{d}'\mathbf{d}/n = \text{Var}[d] + E^2[d] = \pi(1\text{-}\pi) + \pi^2 = \pi$. This minor modification does not effect the model substantively, but greatly simplifies the algebra.] Now, suppose that x^* and d are not independent. In particular, suppose $E[x^*|d{=}1] = \mu^1$ and $E[x^*|d{=}0] = \mu^0$. Then, $\text{plim}[\mathbf{x}^{*\prime}\mathbf{d}/n]$ will equal $\pi\mu^1$. Repeat the derivation with this assumption.

(b) Consider, instead, a regression of x on y and d. What is the probability limit of the coefficient on d in this regression? Assume that x^* and d are independent.
(c) Suppose that x^* and d are not independent, but γ is, in fact, less than zero. Assuming that both preceding equations still hold, what is estimated by $\bar{y}|d=1 - \bar{y}|d=0$? What does this quantity estimate if γ does equal zero?

In the regression of **y** on **x** and **d**, if **d** and **x** are independent, we can invoke the familiar result for least squares regression. The results are the same as those obtained by two simple regressions. It is instructive to verify this. $plim\begin{bmatrix}\mathbf{x'x}/n & \mathbf{x'd}/n \\ \mathbf{d'x}/n & \mathbf{d'd}/n\end{bmatrix}^{-1}\begin{pmatrix}\mathbf{x'y}/n \\ \mathbf{d'y}/n\end{pmatrix} = \begin{bmatrix}\sigma_*^2+\sigma_u^2 & 0 \\ 0 & \pi\end{bmatrix}^{-1}\begin{pmatrix}\beta\sigma_*^2 \\ \gamma\pi\end{pmatrix} = \begin{pmatrix}\beta/\left(1+\sigma_u^2/\sigma_*^2\right) \\ \gamma\end{pmatrix}$. Therefore, although the coefficient on **x** is distorted, the effect of interest, namely, γ, is correctly measured. Now consider what happens if x^* and d are not independent. With the second assumption, we must replace the off diagonal zero above with plim($\mathbf{x'd}/n$). Since u and d are still uncorrelated, this equals Cov[x^*,d]. This is

$$\text{Cov}[x^*,d] = E[x^*d] = \pi E[x^*d|d=1] + (1-\pi)E[x^*d|d=0] = \pi\mu^1.$$

Also, plim[$\mathbf{y'd}/n$] is now βCov[x^*,d] + γplim($\mathbf{d'd}/n$) = $\beta\pi\mu^1 + \gamma\pi$ and plim[$\mathbf{y'x}^*/n$] equals βplim[$\mathbf{x}^{*\prime}\mathbf{x}^*/n$] + γplim[$\mathbf{x}^{*\prime}\mathbf{d}/n$] = $\beta\sigma_*^2 + \gamma\pi\mu^1$. Then, the probability limits of the least squares coefficient estimators is

$$plim\begin{bmatrix}\mathbf{x'x}/n & \mathbf{x'd}/n \\ \mathbf{d'x}/n & \mathbf{d'd}/n\end{bmatrix}^{-1}\begin{pmatrix}\mathbf{x'y}/n \\ \mathbf{d'y}/n\end{pmatrix} = \begin{bmatrix}\sigma_*^2+\sigma_u^2 & \pi\mu^1 \\ \pi\mu^1 & \pi\end{bmatrix}^{-1}\begin{pmatrix}\beta\sigma_*^2+\gamma\pi\mu^1 \\ \beta\pi\mu^1+\gamma\pi\end{pmatrix} = \begin{pmatrix}\beta/\left(1+\sigma_u^2/\sigma_*^2\right) \\ \gamma\end{pmatrix}$$

$$= \frac{1}{\pi(\sigma_*^2+\sigma_u^2)+\pi^2(\mu^1)^2}\begin{bmatrix}\pi & -\pi\mu^1 \\ -\pi\mu^1 & \sigma_*^2+\sigma_u^2\end{bmatrix}\begin{pmatrix}\beta\sigma_*^2+\gamma\pi\mu^1 \\ \beta\pi\mu^1+\gamma\pi\end{pmatrix}$$

$$= \frac{1}{\pi(\sigma_*^2+\sigma_u^2)+\pi^2(\mu^1)^2}\begin{pmatrix}\beta(\pi\sigma_*^2+\pi^2(\mu^1)^2) \\ \gamma(\pi(\sigma_*^2+\sigma_u^2)+\pi^2(\mu^1)^2)+\beta\pi\sigma_u^2\end{pmatrix}.$$

The second expression does reduce to plim $c = \gamma + \beta\pi\mu^1\sigma_u^2/[\pi(\sigma_*^2 + \sigma_u^2) - \pi^2(\mu^1)^2]$, but the upshot is that in the presence of measurement error, the two estimators become an unredeemable hash of the underlying parameters. Note that both expressions reduce to the true parameters if σ_u^2 equals zero.

Finally, the two means are estimators of

$$E[y|d=1] = \beta E[x^*|d=1] + \gamma = \beta\mu^1 + \gamma$$

and

$$E[y|d=0] = \beta E[x^*|d=0] = \beta\mu^0,$$

so the difference is $\beta(\mu^1 - \mu^0) + \gamma$, which is a mixture of two effects. Which one will be larger is entirely indeterminate, so it is reasonable to conclude that this is *not* a good way to analyze the problem. If γ equals zero, this difference will merely reflect the differences in the values of x^*, which may be entirely unrelated to the issue under examination here. (This is, unfortunately, what is usually reported in the popular press.)

7. Data on the number of incidents of damage to a sample of ships, with the type of ship and the period when it was constructed, are given in Table 7.8 below There are five types of ships and four different periods of construction. Use F tests and dummy variable regressions to test the hypothesis that there is no significant "ship type effect" in the expected number of incidents. Now, use the same procedure to test whether there is a significant "period effect."

TABLE 7.8 Ship Damage Incidents

Ship	*Period Constructed*			
Type	*1960–1964*	*1965–1969*	*1970–1974*	*1975–1979*
A	0	4	18	11
B	29	53	44	18
C	1	1	2	1
D	0	0	11	4
E	0	7	12	1

Source: Data from McCullagh and Nelder (1983, p. 137).

According to the full model, the expected number of incidents for a ship of the base type A built in the base period 1960 to 1964, is 3.4. The other 19 predicted values follow from the previous results and are left as an exercise. The relevant test statistics for differences across ship type and year are as follows:

$$\text{type: } F[4,12] = \frac{(3925.2 - 660.9)/4}{660.9/12} = 14.82, \quad \text{year: } F[3,12] = \frac{(1090.3 - 660.9)/3}{660.9/12} = 2.60.$$

The 5 percent critical values from the *F* table with these degrees of freedom are 3.26 and 3.49, respectively, so we would conclude that the average number of incidents varies significantly across ship types but not across years.

Chapter 8

Specification Analysis and Model Selection

1. Suppose the true regression model is given by (8-2). The result in (8-4) shows that if either $\mathbf{P}_{1.2}$ and β_2 are nonzero, then regression of y on $\mathbf{X}_1$ alone produces a biased and inconsistent estimator of β_1. Suppose the objective is to forecast y, not to estimate the parameters. Consider regression of $\mathbf{y}$ on $\mathbf{X}_1$ alone to estimate β_1 with $\mathbf{b}_1$ (which is biased). Is the forecast of computed using $\mathbf{X}_1\mathbf{b}_1$ also biased? Assume that $E[\mathbf{X}_2|\mathbf{X}_1]$ is a linear function of $\mathbf{X}_1$. Discuss your findings generally. What are the implications for prediction when variables are omitted from a regression?

The result cited is $\text{E}[\mathbf{b}_1] = \beta_1 + \mathbf{P}_{1.2}\beta_2$ where $\mathbf{P}_{1.2} = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2$, so the coefficient estimator is biased. If the conditional mean function $E[\mathbf{X}_2|\mathbf{X}_1]$ is a linear function of $\mathbf{X}_1$, then the sample estimator $\text{P}_{1.2}$ actually is an unbiased estimator of the slopes of that function. (That result is Theorem B.3, equation (B-68), in another form). Now, write the model in the form

$$\mathbf{y} = \mathbf{X}_1\beta_1 + \text{E}[\mathbf{X}_2|\mathbf{X}_1]\beta_2 + \varepsilon + (\mathbf{X}_2 - \text{E}[\mathbf{X}_2|\mathbf{X}_1])\beta_2$$

So, when we regress $\mathbf{y}$ on $\mathbf{X}_1$ alone and compute the predictions, we are computing an estimator of $\mathbf{X}_1(\beta_1 + \mathbf{P}_{1.2}\beta_2) = \mathbf{X}_1\beta_1 + \text{E}[\mathbf{X}_2|\mathbf{X}_1]\beta_2$. Both parts of the compound disturbance in this regression ε and $(\mathbf{X}_2 - \text{E}[\mathbf{X}_2|\mathbf{X}_1])\beta_2$ have mean zero and are uncorrelated with $\mathbf{X}_1$ and $\text{E}[\mathbf{X}_2|\mathbf{X}_1]$, so the prediction error has mean zero. The implication is that the forecast is unbiased. Note that this is not true if $\text{E}[\mathbf{X}_2|\mathbf{X}_1]$ is nonlinear, since $\mathbf{P}_{1.2}$ does not estimate the slopes of the conditional mean in that instance. The generality is that leaving out variables wil bias the coefficients, but need not bias the forecasts. It depends on the relationship between the conditional mean function $\text{E}[\mathbf{X}_2|\mathbf{X}_1]$ and $\mathbf{X}_1\mathbf{P}_{1.2}$.

2. Compare the mean squared errors of $\mathbf{b}_1$ and $\mathbf{b}_{1.2}$ in Section 8.2.2. (Hint, the comparison depends on the data and the model parameters, but you can devise a compact expression for the two quantities.)

The "long" estimator, $\mathbf{b}_{1.2}$ is unbiased, so its mean squared error equals its variance, $\sigma^2(\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1)^{-1}$

The short estimator, $\mathbf{b}_1$ is biased; $\text{E}[\mathbf{b}_1] = \beta_1 + \mathbf{P}_{1.2}\beta_2$. It's variance is $\sigma^2(\mathbf{X}_1'\mathbf{X}_1)^{-1}$. It's easy to show that this latter variance is smaller. You can do that by comparing the inverses of the two matrices. The inverse of the first matrix equals the inverse of the second one minus a positive definite matrix, which makes the inverse smaller hence the original matrix is larger - $\text{Var}[\mathbf{b}_{1.2}] \geq \text{Var}[\mathbf{b}_1]$. But, since $\mathbf{b}_1$ is biased, the variance is not its mean squared error. The mean squared error of $\mathbf{b}_1$ is $\text{Var}[\mathbf{b}_1] + \textbf{bias}\times\textbf{bias}'$. The second term is $\mathbf{P}_{1.2}\beta_2\beta_2'\mathbf{P}_{1.2}'$. When this is added to the variance, the sum may be larger or smaller than $\text{Var}[\mathbf{b}_{1.2}]$; it depends on the data and on the parameters, β_2. The important point is that the mean squared error of the biased estimator *may* be smaller than that of the unbiased estimator.

3. The J test in Example is carried out using over 50 years of data. It is optimistic to hope that the underlying structure of the economy did not change in 50 years. Does the result of the test carried out in Example 8.2 persist if it is based on data only from 1980 to 2000? Repeat the computation with this subset of the data.

The regressions are based on real consumption and real disposable income. Results for 1950 to 2000 are given in the text. Repeating the exercise for 1980 to 2000 produces: for the first regression, the estimate of α is 1.03 with a t ratio of 23.27 and for the second, the estimate is -1.24 with a t ratio of -3.062. Thus, as before, both models are rejected. This is qualitatively the same results obtained with the full 51 year data set.

4. The Cox test in Example 8.3 has the same difficulty as the J test in Example 8.2. The sample period might be too long for the test not to have been affected by underlying structural change. Repeat the computations using the 1980 to 2000 data.

Repeating the computations in Example 8.3 using the shorter data set produces q_{01} = -383.10 compared to -15,304 using the full data set. Though this is much smaller, the qualitative result is very much the same, since the critical value is -1.96. Reversing the roles of the competing hypotheses, we obtain q_{10} = 2.121 compared to the earlier value of 3.489. Though this result is close to borderline, the result is, again, the same.

Chapter 9

Nonlinear Regression Models

1. Describe how to obtain nonlinear least squares estimates of the parameters of the model $y = \alpha x^{\beta} + \varepsilon$.

We cannot simply take logs of both sides of the equation as the disturbance is additive rather than multiplicative. So, we must treat the model as a nonlinear regression. The linearized equation is

$$y \approx \alpha^0 x^{\beta^0} + x^{\beta^0}(\alpha - \alpha^0) + \alpha^0 (\log x) x^{\beta^0}(\beta - \beta^0)$$

where α^0 and β^0 are the expansion point. For given values of α^0 and β^0, the estimating equation would be

$$y - \alpha^0 x^{\beta^0} + \alpha^0 x^{\beta^0} + \alpha^0 (\log x) x^{\beta^0} = \alpha\left(x^{\beta^0}\right) + \beta\left(\alpha^0 (\log x) x^{\beta^0}\right) + \varepsilon^*$$

or

$$y + \alpha^0 (\log x) x^{\beta^0} = \alpha\left(x^{\beta^0}\right) + \beta\left(\alpha^0 (\log x) x^{\beta^0}\right) + \varepsilon^*.$$

Estimates of α and β are obtained by applying ordinary least squares to this equation. The process is repeated with the new estimates in the role of α^0 and β^0. The iteration could be continued until convergence. Starting values are always a problem. If one has no particular values in mind, one candidate would be $\alpha^0 = \bar{y}$ and $\beta^0 = 0$ or $\beta^0 = 1$ and α^0 either $\mathbf{x'y}/\mathbf{x'x}$ or $\bar{y}/\bar{x}$. Alternatively, one could search directly for the α and β to minimize the sum of squares, $S(\alpha,\beta) = \Sigma_i (y_i - \alpha x^{\beta})^2 = \Sigma_i \varepsilon_i^2$. The first order conditions for minimization are

$$\partial S(\alpha,\beta)/\partial\alpha = -2\Sigma_i (y_i - \alpha x^{\beta})x^{\beta} = 0 \quad \text{and} \quad \partial S(\alpha,\beta)/\partial\beta = -2\Sigma_i (y_i - \alpha x^{\beta})\alpha(\ln x)x^{\beta} = 0.$$

Methods for solving nonlinear equations such as these are discussed in Chapter 5.

2. Use Mackinnon, et. al's P_E test to determine whether a linear or log-linear production model is more appropriate for the data in Table F6.1. (The test is described in Section 9.4.3 and Example 9.8.)

First, the two simple regressions produce

	Linear	**Log-linear**
Constant	114.338	1.17064
	(173.4)	(.3268)
Labor	2.33814	.602999
	(1.039)	(.1260)
Capital	.471043	.37571
	(.1124)	(.08535)
R^2	.9598	.9435
Standard Error	469.86	.1884

In the regression of Y on 1, K, L, and the predicted values from the loglinear equation minus the predictions from the linear equation, the coefficient on α is -587.349 with an estimated standard error of 3135. Since this is not significantly different from zero, this evidence favors the linear model. In the regression of ln$Y$ on 1, ln$K$, ln$L$ and the predictions from the linear model minus the exponent of the predictions from the loglinear model, the estimate of α is .000355 with a standard error of .000275. Therefore, this contradicts the preceding result and favors the loglinear model. An alternative approach is to fit the Box-Cox model in the fashion of Exercise 4. The maximum likelihood estimate of λ is about -.12, which is much closer to the log-linear model than the lonear one. The log-likelihoods are -192.5107 at the MLE, -192.6266 at λ=0 and -202.837 at $\lambda = 1$. Thus, the hypothesis that $\lambda = 0$ (the log-linear model) would not be rejected but the hypothesis that $\lambda = 1$ (the linear model) would be rejected using the Box-Cox model as a framework.

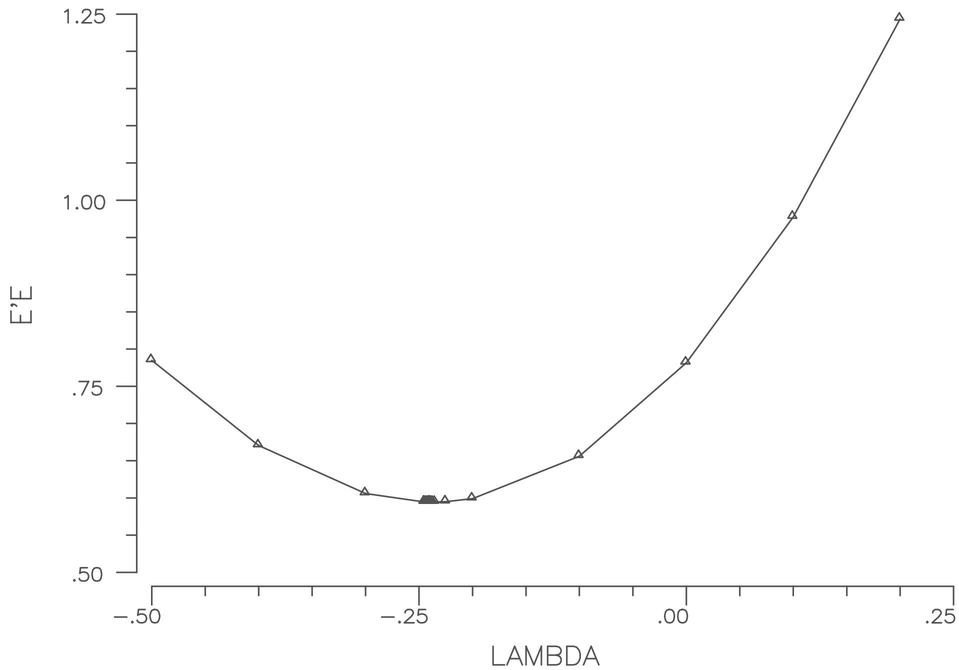
3. Using the Box-Cox transformation, we may specify an alternative to the Cobb-Douglas model as

$$\ln Y = \alpha + \beta_k(K^{\lambda} - 1)/\lambda + \beta_l(L^{\lambda} - 1)/\lambda + \varepsilon.$$

Using Zellner and Revankar's data in Table A9.1, estimate α, β_k, β_l, and λ by using the scanning method suggested in Section F9.2. (Do not forget to scale Y, K, and L by the number of establishments.) Use (9-16), (9-12) and (9-13) to compute the appropriate asymptotic standard errors for your estimates. Compute the two output elasticities, $\partial\ln Y/\partial\ln K$ and $\partial\ln Y/\partial\ln L$ at the sample means of K and L. [**Hint:** $\partial\ln Y/\partial\ln K = K\partial\ln Y/\partial K$.] How do these estimates compare to the values given in Example 10.5?

The search for the minimum sum of squares produced the following results:

λ	$e'e$
-.500	.78477
-.400	.67033
-.300	.60587
-.250	.59479
-.245	.59451
-.244	.59447
-.243	.59444
-.242	.59441
-.241	.59439
-.240	.59438
-.239	.59437
-.238	.59436
-.237	.59437
-.235	.59440
-.225	.59492
-.200	.59897
-.100	.65598
0.000	.78143
.100	.97742
.200	1.24354



The sum of squared residuals is minimized at $\lambda = -.238$. At this value, the regression results are as follows:

Parameter	**Estimate**	**OLS Std.Error**	**Correct Std.Error**
α	2.06092	.07718	.09723
β_k	.178232	.04638	.04378
β_l	.737988	.06996	.12560
λ	-.238	----	.07710

Estimated Asymptotic Covariance Matrix

	α	β_k	β_l	λ
α	.00945			
β_k	.00262	.00192		
β_l	.00511	-.00199	.01578	
λ	.00500	.00037	.00825	.00594

The output elasticities for this function evaluated at the sample means are

$$\partial\ln Y/\partial\ln K = \beta_k K^{\lambda} = (.178232).175905^{-.238} = .2695$$

$$\partial\ln Y/\partial\ln L = \beta_l L^{\lambda} = (.443954).737988^{-.238} = .7740.$$

The estimates found for Zellner and Revankar's model were .254 and .882, respectively, so these are quite similar. For the simple log-linear model, the corresponding values are .2790 and .927.

4. For the model in Exercise 3, test the hypothesis that $\lambda = 0$ using a Wald test, a likelihood ratio test, and a Lagrange multiplier test. Note, the restricted model is the Cobb-Douglas, log-linear model.

The Wald test is based on the unrestricted model. The statistic is the square of the usual t-ratio, W = $(-.232 / .0771)^2$ = 9.0546. The critical value from the chi-squared distribution is 3.84, so the hypothesis that $\lambda = 0$ can be rejected. The likelihood ratio statistic is based on both models. The sum of squared residuals for both unrestricted and restricted models is given above. The log-likelihood is $\ln L = -(n/2)[1 + \ln(2\pi) + \ln(\mathbf{e}'\mathbf{e}/n)]$, so the likelihood ratio statistic is

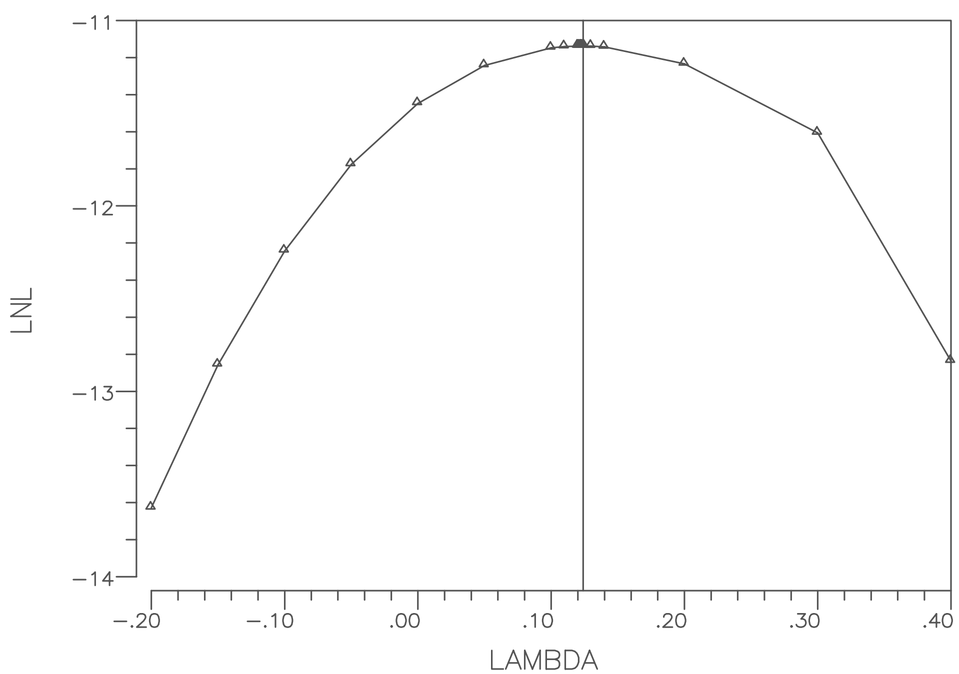
$$LR = n[\ln(\mathbf{e}'\mathbf{e}/n)|\lambda{=}0 - \ln(\mathbf{e}'\mathbf{e}/n)|\lambda{=}-.238] = n\ln[(\mathbf{e}'\mathbf{e}|\lambda{=}0) / (\mathbf{e}'\mathbf{e}|\lambda{=}-.238)$$
$$= 25\ln(.78143/.54369) = 6.8406.$$

Finally, to compute the Lagrange Multiplier statistic, we regress the residuals from the log-linear regression on a constant, $\ln K$, $\ln L$, and $(1/2)(b_k\ln^2 K + b_l\ln^2 L)$ where the coefficients are those from the log-linear model (.27898 and .92731). The R^2 in this regression is .23001, so the Lagrange multiplier statistic is $LM = nR^2$ = 25(.23001) = 5.7503. All three statistics suggest the same conclusion, the hypothesis should be rejected.

5. To extend Zellner and Revankar's model in a fashion similar to theirs, we can use the Box-Cox transformation for the dependent variable as well. Use the method of Section 10.5.2 (with $\theta = \lambda$) to repeat the study of the previous two exercises. How do your results change?

Instead of minimizing the sum of squared deviations, we now maximize the concentrated log-likelihood function, $\ln L = -(n/2)\ln(1+\ln(2\pi)) + (\lambda - 1)\Sigma_i \ln Y_i - (n/2)\ln(\varepsilon'\varepsilon/n)$.
The search for the maximum of $\ln L$ produced the following results:

λ	$\ln L$
-.200	-13.6284
-.150	-12.8568
-.100	-12.2423
-.050	-11.7764
0.000	-11.4476
.050	-11.2427
.100	-11.1480
.110	-11.1410
.120	-11.1378
.121	-11.1377
.122	-11.1376
.123	-11.1376
.124	-11.1375
.125	-11.1376
.130	-11.1383
.140	-11.1423
.200	-11.2344
.300	-11.6064
.400	-12.8371



The log-likelihood is maximized at $\lambda = .124$. At this value, the regression results are as follows:

Parameter	Estimate	OLS Std.Error	Correct Std.Error
α	2.59465	.1283	.7151
β_k	.378094	.1070	.3228
β_l	1.13653	.1117	.4121
λ	.124	----	.2482
σ^2	.036922	----	.0179

Estimated Asymptotic Covariance Matrix

	α	β_k	β_l	λ	σ^2
α	.5114				
β_k	.2203	.1042			
β_l	.2612	.0951	.1698		
λ	.1747	.0730	.0953	.0617	
σ^2	.0104	.0044	.0059	.0038	.00032

The output elasticities for this function evaluated at the sample means, $\overline{K} = .175905$, $\overline{L} = .737988$, $\overline{Y} = 2.870777$, are $\partial\ln Y/\partial\ln K = b_k(K/Y)^\lambda = .2674$

$\partial\ln Y/\partial\ln L = b_l(L/Y)^\lambda = .9017$.

These are quite similar to the estimates given above. The sum of the two output elasticities for the states given in the example in the text are given below for the model estimated with and without transforming the dependent variable. Note that the first of these makes the model look much more similar to the Cobb Douglas model for which this sum is constant.

State	Full Box-Cox Model	lnQ on left hand side
Florida	1.2840	1.6598
Louisiana	1.2019	1.4239
California	1.1574	1.1176
Maryland	1.1657	1.0261
Ohio	1.1899	.9080
Michigan	1.1604	.8506

Once again, we are interested in testing the hypothesis that $\lambda = 0$. The Wald test statistic is $W = (.123 / .2482)^2 = .2455$. We would now not reject the hypothesis that $\lambda = 0$. This is a surprising outcome. The likelihood ratio statistic is based on both models. The sum of squared residuals for the restricted model is given above. The sum of the logs of the outputs is 19.29336, so the restricted log-likelihood is $\ln L^0 = (0\text{-}1)(19.29336) - (25/2)[1 + \ln(2\pi) + \ln(.781403/25)] = -11.44757$. The likelihood ratio statistic is $-2[-11.13758 - (-11.44757)] = .61998$. Once again, the statistic is small. Finally, to compute the Lagrange multiplier statistic, we now use the method described in Example 10.12. The result is $LM = 1.5621$. All of these suggest that the log-linear model is not a significant restriction on the Box-Cox model. This rather peculiar outcome would appear to arise because of the rather substantial reduction in the log-likelihood function which occurs when the dependent variable is transformed along with the right hand side. This is not a contradiction because the model with only the right hand side transformed is not a parametric restriction on the model with both sides transformed. Some further evidence is given in the next exercise.

6. Verify the following differential equation which applies to the Box-Cox transformation

$$d^i x^{(\lambda)}/d\lambda^i = (1/\lambda)[x^\lambda(\ln x)^i - i d^{i-1}x^{(\lambda)}/d\lambda^{i-1}] \tag{9-33}$$

Show that the limiting sequence for $\lambda = 0$ is

$$d^i x^{(\lambda)}/d\lambda^i|_{\lambda=0} = (\ln x)^i/(i+1). \tag{9-34}$$

(These results can be used to great advantage in deriving the actual second derivatives of the log likelihood function for the Box-Cox model. **Hint:** See Example 10.11.)

The proof can be done by mathematical induction. For convenience, denote the ith derivative by f_i. The first derivative appears in Equation (9-34). Just by plugging in i=1, it is clear that f_1 satisfies the relationship. Now, use the chain rule to differentiate f_1,

$$f_2 = (-1/\lambda^2)[x^\lambda(\ln x) - x^{(\lambda)}] + (1/\lambda)[(\ln x)x^\lambda(\ln x) - f_1]$$

Collect terms to yield $f_2 = (-1/\lambda)f_1 + (1/\lambda)[x^\lambda(\ln x)^2 - f_1] = (1/\lambda)[x^\lambda(\ln x)^2 - 2f_1]$.

So, the relationship holds for i = 0, 1, and 2. We now assume that it holds for $i = K$-1, and show that if so, it also holds for $i = K$. This will complete the proof. Thus, assume

$$f_{K\text{-}1} = (1/\lambda)[x^{\lambda}(\ln x)^{K\text{-}1} - (K\text{-}1)f_{K\text{-}2}]$$

Differentiate this to give $f_K = (-1/\lambda)f_{K\text{-}1} + (1/\lambda)[(\ln x)x^{\lambda}(\ln x)^{K\text{-}1} - (K\text{-}1)f_{K\text{-}1}]$.

Collect terms to give $f_K = (1/\lambda)[x^{\lambda}(\ln x)^K - Kf_{K\text{-}1}]$, which completes the proof for the general case.

Now, we take the limiting value

$$\lim_{\lambda\to 0} f_i = \lim_{\lambda\to 0} [x^{\lambda}(\ln x)^i - if_{i\text{-}1}]/\lambda.$$

Use L'Hospital's rule once again.

$$\lim_{\lambda\to 0} f_i = \lim_{\lambda\to 0} d\{[x^{\lambda}(\ln x)^i - if_{i\text{-}1}]/d\lambda\}/\lim_{\lambda\to 0} d\lambda/d\lambda.$$

Then, $\lim_{\lambda\to 0} f_i = \lim_{\lambda\to 0} \{[x^{\lambda}(\ln x)^{i+1} - if_i]\}$

Just collect terms, $(i+1)\lim_{\lambda\to 0} f_i = \lim_{\lambda\to 0} [x^{\lambda}(\ln x)^{i+1}]$

or $\lim_{\lambda\to 0} f_{\mathrm{i}} = \lim_{\lambda\to 0} [x^{\lambda}(\ln x)^{i+1}]/(i+1) = (\ln x)^{i+1}/(i+1)$.

Chapter 10

Nonspherical Disturbances - The Generalized Regression Model

1. What is the covariance matrix, $\text{Cov}[\hat{\beta}, \hat{\beta} - \mathbf{b}]$, of the GLS estimator $\hat{\beta} = (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{y}$ and the difference between it and the OLS estimator, $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$? The result plays a pivotal role in the development of specification tests in Hausman (1978).

Write the two estimators as $\hat{\beta} = \beta + (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}\varepsilon$ and $\mathbf{b} = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon$. Then,

$(\hat{\beta} - \mathbf{b}) = [(\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}^{-1} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\varepsilon$ has $E[\hat{\beta} - \mathbf{b}] = \mathbf{0}$ since both estimators are unbiased. Therefore,

$$\text{Cov}[\hat{\beta}, \hat{\beta} - \mathbf{b}] = E[(\hat{\beta} - \beta)(\hat{\beta} - \mathbf{b})'].$$

Then,

$$\begin{aligned}
&E\{(\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}\varepsilon\varepsilon'[(\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}^{-1} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']'\} \\
&= (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}(\sigma^2\mathbf{\Omega})[\mathbf{\Omega}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] \\
&= \sigma^2(\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{\Omega}\mathbf{\Omega}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1} - (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{\Omega}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})(\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1} - (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1} = \mathbf{0}
\end{aligned}$$

once the inverse matrices are multiplied.

2.This and the next two exercises are based on the test statistic usually used to test a set of J linear restrictions in the generalized regression model:

$$F[J, n-K] = \frac{(\mathbf{R}\hat{\beta} - \mathbf{q})'[\mathbf{R}(\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\beta} - \mathbf{q})/J}{(\mathbf{y} - \mathbf{X}\hat{\beta})'\mathbf{\Omega}^{-1}(\mathbf{y} - \mathbf{X}\hat{\beta})/(n-K)},$$

where $\hat{\beta}$ is the GLS estimator. Show that if $\mathbf{\Omega}$ is known, if the disturbances are normally distributed and if the null hypothesis, $\mathbf{R}\beta = \mathbf{q}$, is true, then this statistic is exactly distributed as F with J and $n - K$ degrees of freedom. What assumptions about the regressors are needed to reach this conclusion? Need they be nonstochastic?

First, $(\mathbf{R}\hat{\beta} - \mathbf{q}) = \mathbf{R}[\beta + (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}\varepsilon)] - \mathbf{q} = \mathbf{R}(\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}\varepsilon$ if $\mathbf{R}\beta - \mathbf{q} = \mathbf{0}$.

Now, use the inverse square root matrix of $\mathbf{\Omega}$, $\mathbf{P} = \mathbf{\Omega}^{-1/2}$ to obtain the transformed data,

$$\mathbf{X}^* = \mathbf{P}\mathbf{X} = \mathbf{\Omega}^{-1/2}\mathbf{X},\quad \mathbf{y}^* = \mathbf{P}\mathbf{y} = \mathbf{\Omega}^{-1/2}\mathbf{y},\ \text{and}\ \ \varepsilon^* = \mathbf{P}\varepsilon = \mathbf{\Omega}^{-1/2}\varepsilon.$$

Then, $E[\varepsilon^*\varepsilon^{*\prime}] = E[\mathbf{\Omega}^{-1/2}\varepsilon\varepsilon'\mathbf{\Omega}^{-2}] = \mathbf{\Omega}^{-1/2}(\sigma^2\mathbf{\Omega})\mathbf{\Omega}^{-1/2} = \sigma^2\mathbf{I}$,

and, $\hat{\beta} = (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{y} = (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{y}^*$

= the OLS estimator in the regression of $\mathbf{y}^*$ on $\mathbf{X}^*$.

Then, $\mathbf{R}\hat{\beta} - \mathbf{q} = \mathbf{R}(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\varepsilon^*$

and the numerator is $\varepsilon^{*\prime}\mathbf{X}^*(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{R}']^{-1}\mathbf{R}(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\varepsilon^* / J$. By multiplying it out, we find that the matrix of the quadratic form above is idempotent. Therefore, this is an idempotent quadratic form in a normally distributed random vector. Thus, its distribution is that of σ^2 times a chi-squared variable with degrees of freedom equal to the rank of the matrix. To find the rank of the matrix of the quadratic form, we can find its trace. That is

$$\begin{aligned}
&\text{tr}\{\mathbf{X}^*(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{R}']^{-1}\mathbf{R}(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^*\} \\
&= \text{tr}\{(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{R}']^{-1}\mathbf{R}(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{X}^*\} \\
&= \text{tr}\{(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{R}']^{-1}\mathbf{R}\} \\
&= \text{tr}\{[\mathbf{R}(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{R}'][\mathbf{R}(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{R}']^{-1}\} = \text{tr}\{\mathbf{I}_J\} = J,
\end{aligned}$$

which might have been expected. Before proceeding, we should note, we could have deduced this outcome from the form of the matrix. The matrix of the quadratic form is of the form $\mathbf{Q} = \mathbf{X}^*\mathbf{A}\mathbf{B}\mathbf{A}'\mathbf{X}^{*\prime}$ where $\mathbf{B}$ is the nonsingular matrix in the square brackets and $\mathbf{A} = (\mathbf{X}^{*\prime}\mathbf{X}^{*})^{-1}\mathbf{R}'$, which is a $K\times J$ matrix which cannot have rank higher than J. Therefore, the entire product cannot have rank higher than J. Continuing, we now find that the numerator (apart from the scale factor, σ^2) is the ratio of a chi-squared[J] variable to its degrees of freedom.

We now turn to the denominator. By multiplying it out, we find that the denominator is $(\mathbf{y}^* - \mathbf{X}^*\hat{\beta})'(\mathbf{y}^* - \mathbf{X}^*\hat{\beta})/(n - K)$. This is exactly the sum of squared residuals in the least squares regression of $\mathbf{y}^*$ on $\mathbf{X}^*$. Since $\mathbf{y}^* = \mathbf{X}^*\beta + \varepsilon^*$ and $\hat{\beta} = (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{y}^*$ the denominator is $\varepsilon^{*\prime}\mathbf{M}^*\varepsilon^*/(n - K)$, the familiar form of the sum of squares. Once again, this is an idempotent quadratic form in a normal vector (and, again, apart from the scale factor, σ^2, which now cancels). The rank of the $\mathbf{M}$ matrix is $n - K$, as always, so the denominator is also a chi-squared variable divided by its degrees of freedom.

It remains only to show that the two chi-squared variables are independent. We know they are if the two matrices are orthogonal. They are since $\mathbf{M}^*\mathbf{X}^* = \mathbf{0}$. This completes the proof, since all of the requirements for the F distribution have been shown.

3. Now suppose that the disturbances are not normally distributed, although $\mathbf{\Omega}$ is still known. Show that the limiting distribution of previous statistic is $(1/J)$ times a chi-squared variable with J degrees of freedom. (Hint: The denominator converges to σ^2.) Conclude that in the generalized regression model, the limiting distribution of the Wald statistic

$$W = (\mathbf{R}\hat{\beta} - \mathbf{q})'\{\mathbf{R}(\text{Est. Var}[\hat{\beta}])\mathbf{R}'\}^{-1}(\mathbf{R}\hat{\beta} - \mathbf{q})$$

is chi-squared with J degrees of freedom, regardless of the distribution of the disturbances, as long as the data are otherwise well behaved. Note that in a finite sample, the true distribution may be approximated with an $F[J,n-K]$ distribution. It is a bit ambiguous, however, to interpret this fact as implying that the statistic is asymptotically distributed as F with J and $n-K$ degrees of freedom, because the limiting distribution used to obtain our result is the chi-squared, not the F. In this instance, the $F[J, n-K]$ is a random variable that tends asymptotically to the chi-squared variate.

First, we know that the denominator of the F statistic converges to σ^2. Therefore, the limiting distribution of the F statistic is the same as the limiting distribution of the statistic which results when the denominator is replaced by σ^2. It is useful to write this modified statistic as

$$W^* = (1/\sigma^2)(\mathbf{R}\hat{\beta} - \mathbf{q})'[\mathbf{R}(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\beta} - \mathbf{q})/J.$$

Now, incorporate the results from the previous problem to write this as

$$W^* = \varepsilon^{*\prime}\mathbf{X}^*(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{R}'[\mathbf{R}\sigma^2(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{R}']^{-1}\mathbf{R}(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\varepsilon/J$$

Let $\qquad \varepsilon^0 = \mathbf{R}(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\varepsilon^*$.

Note that this is a $J\times 1$ vector. By multiplying it out, we find that $E[\varepsilon^0\varepsilon^{0\prime}] = \text{Var}[\varepsilon^0] = \mathbf{R}\{\sigma^2(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\}\mathbf{R}'$. Therefore, the modified statistic can be written as $W^* = \varepsilon^{0\prime}\text{Var}[\varepsilon^0]^{-1}\varepsilon^0/J$. This is the 'full rank quadratic form' discussed in Appendix B. For convenience, let $\mathbf{C} = \text{Var}[\varepsilon^0]$, $\mathbf{T} = \mathbf{C}^{-1/2}$, and $\mathbf{v} = \mathbf{T}\varepsilon^0$. Then, $W^* = \mathbf{v}'\mathbf{v}$. By construction, $\mathbf{v} = \text{Var}[\varepsilon^0]^{-1/2}\varepsilon^0$, so $E[\mathbf{v}] = \mathbf{0}$ and $\text{Var}[\mathbf{v}] = \mathbf{I}$. The limiting distribution of $\mathbf{v}'\mathbf{v}$ is chi-squared J if the limiting distribution of $\mathbf{v}$ is standard normal. All of the conditions for the central limit theorem apply to $\mathbf{v}$, so we do have the result we need. This implies that as long as the data are well behaved, the numerator of the F statistic will converge to the ratio of a chi-squared variable to its degrees of freedom.

4. Finally, suppose that $\mathbf{\Omega}$ must be estimated, but that assumptions (10-27) and (10-31) are met by the estimator. What changes are required in the development of the previous problem?

The development is unchanged. As long as the limiting behavior of $(1/n)\hat{\mathbf{X}}'\hat{\mathbf{X}} = (1/n)\mathbf{X}'\hat{\mathbf{\Omega}}^{-1}\mathbf{X}$ is the same as that of $(1/n)\mathbf{X}^{*\prime}\mathbf{X}^*$, the limiting distribution of the test statistic will be the same as if the true $\mathbf{\Omega}$ were used instead of the estimate $\hat{\mathbf{\Omega}}$.

5. In the generalized regression model, if the K columns of $\mathbf{X}$ are characteristic vectors of $\boldsymbol{\Omega}$, then ordinary least squares and generalized least squares are identical. (The result is actually a bit broader; $\mathbf{X}$ may be any linear combination of exactly K characteristic vectors. This result is **Kruskal's Theorem**.)

a. Prove the result directly using matrix algebra.

b. Prove that if $\mathbf{X}$ contains a constant term and if the remaining columns are in deviation form (so that the column sum is zero), then the model of Exercise 8 below is one of these cases. (The seemingly unrelated regressions model with identical regressor matrices, discussed in Chapter 14, is another.)

First, in order to simplify the algebra somewhat without losing any generality, we will scale the columns of $\mathbf{X}$ so that for each $\mathbf{x}_k$, $\mathbf{x}_k'\mathbf{x}_k = 1$. We do this by beginning with our original data matrix, say, $\mathbf{X}^0$ and obtaining $\mathbf{X}$ as $\mathbf{X} = \mathbf{X}^0\mathbf{D}^{-1/2}$, where $\mathbf{D}$ is a diagonal matrix with diagonal elements $\mathbf{D}_{kk} = \mathbf{x}_k^{0\prime}\mathbf{x}_k^0$. By multiplying it out, we find that the GLS slopes based on $\mathbf{X}$ instead of $\mathbf{X}^0$ are

$$\hat{\boldsymbol{\beta}} = [(\mathbf{X}^0\mathbf{D}^{-1/2})'\boldsymbol{\Omega}^{-1}(\mathbf{X}^0\mathbf{D}^{-1/2})]^{-1}[(\mathbf{X}^0\mathbf{D}^{-1/2})'\boldsymbol{\Omega}^{-1}\mathbf{y}] = \mathbf{D}^{1/2}[\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}](\mathbf{D}')^{1/2}(\mathbf{D}')^{-1/2}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y} = \mathbf{D}^{1/2}\,\hat{\boldsymbol{\beta}}^{\,0}$$

with variance $\text{Var}[\hat{\boldsymbol{\beta}}] = \mathbf{D}^{1/2}\sigma^2[\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}]^{-1}(\mathbf{D}')^{1/2} = \mathbf{D}^{1/2}\text{Var}[\hat{\boldsymbol{\beta}}^{\,0}](\mathbf{D}')^{1/2}$. Likewise, the OLS estimator based on $\mathbf{X}$ instead of $\mathbf{X}^0$ is $\mathbf{b} = \mathbf{D}^{1/2}\mathbf{b}^0$ and has variance $\text{Var}[\mathbf{b}] = \mathbf{D}^{1/2}\text{Var}[\mathbf{b}^0](\mathbf{D}')^{1/2}$. Since the scaling affects both estimators identically, we may ignore it and simply assume that $\mathbf{X}'\mathbf{X} = \mathbf{I}$.

If each column of $\mathbf{X}$ is a characteristic vector of $\boldsymbol{\Omega}$, then, for the kth column, $\mathbf{x}_k$, $\boldsymbol{\Omega}\mathbf{x}_k = \lambda_k\mathbf{x}_k$. Further, $\mathbf{x}_k'\boldsymbol{\Omega}\mathbf{x}_k = \lambda_k$ and $\mathbf{x}_k'\boldsymbol{\Omega}\mathbf{x}_j = 0$ for any two different columns of $\mathbf{X}$. (We neglect the scaling of $\mathbf{X}$, so that $\mathbf{X}'\mathbf{X} = \mathbf{I}$, which we would usually assume for a set of characteristic vectors. The implicit scaling of $\mathbf{X}$ is absorbed in the characteristic roots.) Recall that the characteristic vectors of $\boldsymbol{\Omega}^{-1}$ are the same as those of $\boldsymbol{\Omega}$ while the characteristic roots are the reciprocals. Therefore, $\mathbf{X}'\boldsymbol{\Omega}\mathbf{X} = \boldsymbol{\Lambda}_K$, the diagonal matrix of the K characteristic roots which correspond to the columns of $\mathbf{X}$. In addition, $\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X} = \boldsymbol{\Lambda}_K^{-1}$, so $(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1} = \boldsymbol{\Lambda}_K$, and $\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y} = \boldsymbol{\Lambda}_K^{-1}\mathbf{X}'\mathbf{y}$. Therefore, the GLS estimator is simply $\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$ with variance $\text{Var}[\hat{\boldsymbol{\beta}}] = \sigma^2\boldsymbol{\Lambda}_K$. The OLS estimator is $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{y}$. Its variance is $\text{Var}[\mathbf{b}] = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2\boldsymbol{\Lambda}_K$, which means that OLS and GLS are identical in this case.

6. In the generalized regression model, suppose that $\boldsymbol{\Omega}$ is known.

a. What is the covariance matrix of the OLS and GLS estimators of $\boldsymbol{\beta}$?

b. What is the covariance matrix of the OLS residual vector $\mathbf{e} = \text{y - Xb}$?

c. What is the covariance matrix of the GLS residual vector $\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$?

d. What is the covariance matrix of the OLS and GLS residual vectors?

Write $\mathbf{b} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}$ and $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}$. The covariance matrix is

$$\text{E}[(\mathbf{b} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'] = E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{\Omega}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\sigma^2\boldsymbol{\Omega})\boldsymbol{\Omega}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1} = \sigma^2(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}.$$

For part (b), $\mathbf{e} = \mathbf{M}\boldsymbol{\varepsilon}$ as always, so $E[\mathbf{ee}'] = \sigma^2\mathbf{M}\boldsymbol{\Omega}\mathbf{M}$. No further simplification is possible for the general case.

For part (c),

$$\begin{aligned}
\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} &= \mathbf{y} - \mathbf{X}[\boldsymbol{\beta} + (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-1}\boldsymbol{\varepsilon}] \\
&= \mathbf{X}\boldsymbol{\beta} + \varepsilon - \mathbf{X}[\boldsymbol{\beta} + (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-1}\boldsymbol{\varepsilon}] \\
&= [\mathbf{I} - \mathbf{X}(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-1}]\varepsilon.
\end{aligned}$$

Thus,

$$\begin{aligned}
E[\hat{\boldsymbol{\varepsilon}}\,\hat{\boldsymbol{\varepsilon}}'] &= [\mathbf{I} - \mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}]E[\varepsilon\varepsilon'][\mathbf{I} - \mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}]' \\
&= [\mathbf{I} - \mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}](\sigma^2\boldsymbol{\Omega})[\mathbf{I} - \mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}]' \\
&= [\sigma^2\boldsymbol{\Omega} - \sigma^2\mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'][\mathbf{I} - \mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}]' \\
&= [\sigma^2\boldsymbol{\Omega} - \sigma^2\mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'][\mathbf{I} - \boldsymbol{\Omega}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'] \\
&= \sigma^2\boldsymbol{\Omega} - \sigma^2\mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}' - \sigma^2\mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}' + \sigma^2\mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1})\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}' \\
&= \sigma^2[\boldsymbol{\Omega} - \mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}']
\end{aligned}$$

The GLS residual vector appears in the preceding part. As always, the OLS residual vector is $\mathbf{e} = \mathbf{M}\boldsymbol{\varepsilon} = [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\boldsymbol{\varepsilon}$. The covariance matrix is

$$
\begin{aligned}
E[\mathbf{e}\hat{\boldsymbol{\varepsilon}}'] &= E[(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'(\mathbf{I} - \mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1})'] \\
&= (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')(\sigma^2\boldsymbol{\Omega})(\mathbf{I} - \Omega^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}') \\
&= \sigma^2\boldsymbol{\Omega} - \sigma^2\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega} - \sigma^2\boldsymbol{\Omega}\boldsymbol{\Omega}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}' + \sigma^2\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}\boldsymbol{\Omega}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}' \\
&= \sigma^2\boldsymbol{\Omega} - \sigma^2\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\
&= \sigma^2\mathbf{M}\boldsymbol{\Omega}.
\end{aligned}
$$

7. Suppose that y has the pdf $f(y \mid \mathbf{x}) = (1/\mathbf{x}'\boldsymbol{\beta})e^{-y/(\boldsymbol{\beta}'\mathbf{x})}$, $y > 0$.

Then $E[y \mid \mathbf{x}] = \boldsymbol{\beta}'\mathbf{x}$ and $\text{Var}[y \mid \mathbf{x}] = (\boldsymbol{\beta}'\mathbf{x})^2$. For this model, prove that GLS and MLE are the same, even though this distribution, like the one in Exercise 2, involves the same parameters in the conditional mean function and the disturbance variance.

The GLS estimator is $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'^{-1}\mathbf{y} = [\Sigma_i\mathbf{x}_i\mathbf{x}_i'/(\boldsymbol{\beta}'\mathbf{x}_i)^2]^{-1}[\Sigma_i\mathbf{x}_iy_i/(\boldsymbol{\beta}'\mathbf{x}_i)^2]$. The log-likelihood for this model is $\ln L = -\Sigma_i\ln(\boldsymbol{\beta}'\mathbf{x}_i) - \Sigma_iy_i/(\boldsymbol{\beta}'\mathbf{x}_i)$.

The likelihood equations are

$$\partial\ln L/\partial\boldsymbol{\beta} = -\Sigma_i(1/\boldsymbol{\beta}'\mathbf{x}_i)\mathbf{x}_i + \Sigma_i[y_i/(\boldsymbol{\beta}'\mathbf{x}_i)^2]\mathbf{x}_i = \mathbf{0}$$

or $\quad \Sigma_i(\mathbf{x}_iy_i/(\boldsymbol{\beta}'\mathbf{x}_i)^2) = \Sigma_i\mathbf{x}_i/(\boldsymbol{\beta}'\mathbf{x}_i)$.

Now, write $\quad \Sigma_i\mathbf{x}_i/(\boldsymbol{\beta}'\mathbf{x}_i) = \Sigma_i\mathbf{x}_i\mathbf{x}_i'\boldsymbol{\beta}/(\boldsymbol{\beta}'\mathbf{x}_i)^2$,

so the likelihood equations are equivalent to $\Sigma_i(\mathbf{x}_iy_i/(\boldsymbol{\beta}'\mathbf{x}_i).^2) = \Sigma_i\mathbf{x}_i\mathbf{x}_i'\boldsymbol{\beta}/(\boldsymbol{\beta}'\mathbf{x}_i).^2$, or $\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y} = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})\boldsymbol{\beta}$. These are the normal equations for the GLS estimator, so the two estimators are the same. We should note, the solution is only implicit, since $\boldsymbol{\Omega}$ is a function of $\boldsymbol{\beta}$. For another more common application, see the discussion of the FIML estimator for simultaneous equations models in Chapter 15.

8. Suppose that the regression model is $y = \mu + \varepsilon$, where ε has a zero mean, constant variance, and equal correlation ρ across observations. Then $\text{Cov}[\varepsilon_i,\varepsilon_j] = \sigma^2\rho$ if $i \neq j$. Prove that the least squares estimator of μ is inconsistent. Find the characteristic roots of $\boldsymbol{\Omega}$ and show that Condition 2. after Theorem 10.2 is violated.

The covariance matrix is

$$\sigma^2\boldsymbol{\Omega} = \sigma^2\begin{bmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \rho & \rho & 1 & \cdots & \rho \\ & & & \vdots & \\ \rho & \rho & \rho & \cdots & 1 \end{bmatrix}.$$

The matrix $\mathbf{X}$ is a column of 1s, so the least squares estimator of μ is $\bar{y}$. Inserting this $\boldsymbol{\Omega}$ into (10-5), we obtain $\text{Var}[\bar{y}] = \dfrac{\sigma^2}{n}(1 - \rho + n\rho)$. The limit of this expression is $\rho\sigma^2$, not zero. Although ordinary least squares is unbiased, it is not consistent. For this model, $\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}/n = 1 + \rho(n - 1)$, which does not converge. Using Theorem 10.2 instead, $\mathbf{X}$ is a column of 1s, so $\mathbf{X}'\mathbf{X} = n$, a scalar, which satisfies condition 1. To find the characteristic roots, multiply out the equation $\boldsymbol{\Omega}\mathbf{x} = \lambda\mathbf{x} = (1-\rho)\mathbf{I}\mathbf{x} + \rho\mathbf{i}\mathbf{i}'\mathbf{x} = \lambda\mathbf{x}$. Since $\mathbf{i}'\mathbf{x} = \Sigma_i x_i$, consider any vector $\mathbf{x}$ whose elements sum to zero. If so, then it's obvious that $\lambda = \rho$. There are n-1 such roots. Finally, suppose that $\mathbf{x} = \mathbf{i}$. Plugging this into the equation produces $\lambda = 1 - \rho + n\rho$. The characteristic roots of $\boldsymbol{\Omega}$ are $(1 - \rho)$ with multiplicity $n - 1$ and $(1 - \rho + n\rho)$, which violates condition 2.

Chapter 11

Heteroscedasticity

1. Suppose the regression model is $y_i = \mu + \varepsilon_i$, where $E[\varepsilon_i|x_i] = 0$, but $\text{Var}[\varepsilon_i|x_i] = \sigma^2 x_i^2$, $x_i > 0$.
(a) Given a sample of observations on y_i and x_i, what is the most efficient estimator of ε? What is its variance?
(b) What is the ordinary least squares estimator of μ and what is the variance of the ordinary least squares estimator?
(c) Prove that the estimator in (a) is at least as efficient as the estimator in (b).

This is a heteroscedastic regression model in which the matrix **X** is a column of ones. The efficient estimator is the GLS estimator, $\hat{\beta} = (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{y} = [\Sigma_i 1 y_i/x_i^2] / [\Sigma_i 1^2/\boldsymbol{x}_i^2] = [\Sigma_i(y_i/x_i^2)] / [\Sigma_i(1/x_i^2)]$. As always, the variance of the estimator is $\text{Var}[\hat{\beta}] = \sigma^2(\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1} = \sigma^2/[\Sigma_i(1/x_i^2)]$. The ordinary least squares estimator is $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \bar{y}$. The variance of $\bar{y}$ is $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{\Omega}\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1} = (\sigma^2/n^2)\Sigma_i x_i^2$. To show that the variance of the OLS estimator is greater than or equal to that of the GLS estimator, we must show that $(\sigma^2/n^2)\Sigma_i x_i^2 \geq \sigma^2/\Sigma_i(1/x_i^2)$ or $(1/n^2)(\Sigma_i x_i^2)(\Sigma_i(1/x_i^2)) \geq 1$ or $\Sigma_i\Sigma_j(x_i^2/x_j^2) \geq n^2$. The double sum contains n terms equal to one. There remain $n(n-1)/2$ pairs of the form $(x_i^2/x_j^2 + x_j^2/x_i^2)$. If it can be shown that each of these sums is greater than or equal to 2, the result is proved. Just let $z_i = x_i^2$. Then, we require $z_i/z_j + z_j/z_i - 2 \geq 0$. But, this is equivalent to $(z_i^2 + z_j^2 - 2z_iz_j) / z_iz_j \geq 0$ or $(z_i - z_j)^2/z_iz_j \geq 0$, which is certainly true if z_i and z_j are positive. They are since z_i equals x_i^2. This completes the proof.

2. For the model in the previous exercise, what is the probability limit of $s^2 = (1/(n-1))\Sigma_i(y_i - \bar{y})^2$? Note that this is the least squares estimate of the residual variance. It is also n times the conventional estimator of the variance of the OLS estimator, $\text{Est.Var}[\bar{y}] = s^2(\mathbf{X}'\mathbf{X})^{-1} = s^2/n$. How does this compare to the true value you found in part (b) of Exercise 1? Does the conventional estimator produce the correct estimate of the true asymptotic variance of the least squares estimator?

Consider, first, $\bar{y}$. We saw earlier that $\text{Var}[\bar{y}] = (\sigma^2/n^2)\Sigma_i x_i^2 = (\sigma^2/n)(1/n)\Sigma_i x_i^2$. The expected value is $E[\bar{y}] = E[(1/n)\Sigma_i y_i] = \alpha$. If the mean square of x converges to something finite, then $\bar{y}$ is consistent for α. That is, if $\text{plim}(1/n)\Sigma_i x_i^2 = \bar{q}$ where $\bar{q}$ is some finite number, then, plim $\bar{y} = \alpha$. As such, it follows that s^2 and $s_*^2 = (1/(n-1))\Sigma_i(y_i - \alpha)^2$ have the same probability limit. We consider, therefore, plim $s_*^2 = \text{plim}(1/(n-1))\Sigma_i\varepsilon_i^2$. The expected value of s_*^2 is $E[(1/(n-1))\,\Sigma_i\varepsilon_i^2] = \sigma^2(1/\Sigma_i x_i^2)$. Once again, nothing more can be said without some assumption about x_i. Thus, we assume again that the average square of x_i converges to a finite, positive constant, $\bar{q}$. Of course, the result is unchanged by division by $(n-1)$ instead of n, so $\lim_{n\to\infty} E[s_*^2] = \sigma^2\bar{q}$. The variance of s_*^2 is $\text{Var}[s_*^2] = \Sigma_i\text{Var}[\varepsilon_i^2]/(n-1)^2$. To characterize this, we will require the variances of the squared disturbances, which involves their fourth moments. But, if we assume that every fourth moment is finite, then the preceding is $(n/(n-1)^2)$ times the average of these fourth moments. If every fourth moment is finite, then the term is dominated by the leading $(n/(n-1)^2)$ which converges to zero. It follows that plim $s_*^2 = \sigma^2\bar{q}$. Therefore, the conventional estimator estimates $\text{Asy.Var}[\bar{y}] = \sigma^2\bar{q}/n$.

The appropriate variance of the least squares estimator is $\text{Var}[\bar{y}] = (\sigma^2/n^2)\Sigma_i x_i^2$, which is, of course, precisely what we have been analyzing above. It follows that the conventional estimator of the variance of the OLS estimator in this model is an appropriate estimator of the true variance of the least squares estimator. This follows from the fact that the regressor in the model, **i**, is unrelated to the source of heteroscedasticity, as discussed in the text.

3. Two samples of 50 observations each produce the following moment matrices: (In each case, **X** is a constant and one variable.)

	Sample 1	Sample 2
$\mathbf{X'X}$	$\begin{bmatrix} 50 & 300 \\ 300 & 1200 \end{bmatrix}$	$\begin{bmatrix} 50 & 300 \\ 300 & 1200 \end{bmatrix}$
$\mathbf{y'X}$	[300 2000]	[300 2200]
$\mathbf{y'y}$	2100	2800

(a) Compute the least squares regression coefficients and the residual variances, s^2, for each data set. Compute the R^2 for each regression.
(b) Compute the OLS estimate of the coefficient vector assuming that the coefficients and disturbance variance are the same in the two regressions. Also compute the estimate of the asymptotic covariance matrix of the estimator.
(c) Test the hypothesis that the variances in the two regressions are the same without assuming that the coefficients are the same in the two regressions.
(d) Compute the two step feasible GLS estimator of the coefficients in the regression assuming that the constant and slope are the same in both regressions. Compute the estimate of the covariance matrix and compare it to the result of (b)

The sample moments are obtained using, for example, $S_{xx} = \mathbf{x'x} - n\bar{x}^2$ and so on. For the two samples, we obtain

	$\bar{y}$	$\bar{x}$	S_{xx}	S_{yy}	S_{xy}
Sample 1	6	6	300	300	200
Sample 2	6	6	300	1000	400

The parameter estimates are computed directly using the results of Chapter 6.

	Intercept	Slope	R^2	s^2
Sample 1	2	2/3	4/9	(1500/9)/48 = 3.472
Sample 2	-2	4/3	16/30	(4200/9)/48 = 9.722

The pooled moments based on 100 observations are $\mathbf{X'X} = \begin{bmatrix} 100 & 600 \\ 600 & 4200 \end{bmatrix}$, $\mathbf{X'y} = \begin{bmatrix} 600 \\ 4200 \end{bmatrix}$, $\mathbf{y'y} = 4900$. The coefficient vector based on these data is $[a,b] = [0,1]$. This might have been predicted since the two $\mathbf{X'X}$ matrices are identical. OLS which ignores the heteroscedasticity would simply average the estimates. The sum of squared residuals would be $\mathbf{e'e} = \mathbf{y'y} - \mathbf{b'X'y} = 4900 - 4200 = 700$, so the estimate of σ^2 is $s^2 = 700/98 = 7.142$. Note that the earlier values obtained were 3.472 and 9.722, so the pooled estimate is between the two, once again, as might be expected. The asymptotic covariance matrix of these estimates is $s^2(\mathbf{X'X})^{-1}$ $= 7.142\begin{bmatrix} .07 & -.01 \\ -.01 & .167 \end{bmatrix}$.

To test the equality of the variances, we can use the Goldfeld and Quandt test. Under the null hypothesis of equal variances, the ratio $F = [\mathbf{e}_1'\mathbf{e}_1/(n_1 - 2)]/[\mathbf{e}_2'\mathbf{e}_2/(n_2 - 2)]$ (or vice versa for the subscripts) is the ratio of two independent chi-squared variables each divided by their respective degrees of freedom. Although it might seem so from the discussion in the text (and the literature) there is nothing in the test which requires that the coefficient vectors be assumed equal across groups. Since for our data, the second sample has the larger residual variance, we refer $F[48,48] = s_2^2/s_1^2 = 9.722 / 3.472 = 2.8$ to the F table. The critical value for 95% significance is 1.61, so the hypothesis of equal variances is rejected.

The two step estimator is $\hat{\boldsymbol\beta} = [(1/s_1^2)\mathbf{X}_1'\mathbf{X}_1 + (1/s_2^2)\mathbf{X}_2'\mathbf{X}_2]^{-1}[(1/s_1^2)\mathbf{X}_1'\mathbf{y}_1 + (1/s_2^2)\mathbf{X}_2'\mathbf{y}_2]$. The $\mathbf{X'X}$ matrices are the same in this problem, so this simplifies to

$\hat{\boldsymbol\beta} = [(1/s_1^2 + 1/s_2^2)\mathbf{X'X}]^{-1}[(1/s_1^2)\mathbf{X}_1'\mathbf{y}_1 + (1/s_2^2)\mathbf{X}_2'\mathbf{y}_2]$. . The estimator is, therefore

$$\left[\left(\frac{1}{3.472}+\frac{1}{9.722}\right)\begin{pmatrix} 50 & 300 \\ 300 & 2100 \end{pmatrix}\right]^{-1}\left[\frac{1}{3.472}\begin{pmatrix} 300 \\ 2000 \end{pmatrix}+\frac{1}{9.722}\begin{pmatrix} 300 \\ 2200 \end{pmatrix}\right]=\begin{pmatrix} .9469 \\ .8422 \end{pmatrix}.$$

4. Using the data in the previous exercise, use the Oberhofer-Kmenta method to compute the maximum likelihood estimate of the common coefficient vector.

The estimator must be based on maximum likelihood estimators of the two disturbance variances, so they must be recomputed first. Our initial estimators of them are s_1^2 = (1500/9)/50 = 3.3333 and s_2^2 = (4200/9)/50 = 9.3333. Beginning from this point, we iterate between the estimator of the coefficient vector described above and the two variance estimators $s_j^2 = (1/50)[(\mathbf{y}'\mathbf{y})_j - 2\hat{\boldsymbol{\beta}}'(\mathbf{X}'\mathbf{y})_j + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}]$. The iterations are

	s_1^2	s_2^2	$\hat{\alpha}$	$\hat{\beta}$	
(0)	3.333333	9.33333	.947366	.824106	
(1)	3.518005	10.78117	1.01588	.830686	
(3)	3.494747	10.84926	1.02545	.829092	
(4)	3.491626	10.85889	1.02676	.828873	
(5)	3.491199	10.86021	1.02694	.828843	
(6)	3.491141	10.86039	1.02697	.828839	
(7)	3.491134	10.86042	1.02697	.828839	(Converged)

5. This exercise is based on the following data set:

50 observations on *Y*:

-1.42	2.75	2.10	-5.08	1.49	1.00	.16	-1.11	1.66
-.26	-4.87	5.94	2.21	-6.87	.90	1.61	2.11	-3.82
-.62	7.01	26.14	7.39	.79	1.93	1.97	-23.17	-2.52
-1.26	-.15	3.41	-5.45	1.31	1.52	2.04	3.00	6.31
5.51	-15.22	-1.47	-1.48	6.66	1.78	2.62	-5.16	-4.71
-.35	-.48	1.24	.69	1.91				

50 observations on X_1:

-1.65	1.48	.77	.67	.68	.23	-.40	-1.13	.15
-.63	.34	.35	.79	.77	-1.04	.28	.58	-.41
-1.78	1.25	.22	1.25	-.12	.66	1.06	-.66	-1.18
-.80	-1.32	.16	1.06	-.60	.79	.86	2.04	-.51
.02	.33	-1.99	.70	-.17	.33	.48	1.90	-.18
-.18	-1.62	.39	.17	1.02				

50 observations on X_2:

-.67	.70	.32	2.88	-.19	-1.28	-2.72	-.70	-1.55
-.74	-1.87	1.56	.37	-2.07	1.20	.26	-1.34	-2.10
.61	2.32	4.38	2.16	1.51	.30	-.17	7.82	-1.15
1.77	2.92	-1.94	2.09	1.50	-.46	.19	-.39	1.54
1.87	-3.45	-.88	-1.53	1.42	-2.70	1.77	-1.89	-1.85
2.01	1.26	-2.02	1.91	-2.23				

(a) Compute the ordinary least squares regression of *Y* on a constant, X_1, and X_2. Be sure to compute the conventional estimator of the asymptotic covariance matrix of the OLS estimator also.
(b) Compute the White estimator of the appropriate asymptotic covariance matrix for the OLS estimates. (See (12-9).)
(c) Test for the presence of heteroscedasticity using White's general test. Do your results suggest the nature of the heteroscedasticity?
(d) Use the Breusch and Pagan Lagrange multiplier test to test for heteroscedasticity.
(e) Sort the data keying on X_1 and use the Goldfeld-Quandt test to test for heteroscedasticity. Repeat using X_2. What do you find?
(f) Use one of Glesjer's tests to test for heteroscedasticity.

The ordinary least squares regression of Y on a constant, X_1, and X_2 produces the following results:

Sum of squared residuals	1911.9275		
R^2	.03790		
Standard error of regression	6.3780		
Variable	**Coefficient**	**Standard Error**	**t-ratio**
One	.190394	.9144	.208
X_1	1.13113	.9826	1.151
X_2	.376825	.4399	.857

Covariance Matrix			**White's Corrected Matrix**		
.836212			.524589		
-.115451	.96551		.076578	.282366	
-.047133	.051081	.193532	.399218	-.091608	1.14447

To apply White's test, we first obtain the residuals from the regression of Y on a constant, X_1, and X_2. Then, we regress the squares of these residuals on a constant, X_1, X_2, X_1^2, X_2^2, and X_1X_2. The R^2 in this regression is .78296, so the chi-squared statistic is $50\times0.78296 = 39.148$. The critical value from the table of chi-squared with 6 degrees of freedom is 12.5916, so we would conclude that there is evidence of heteroscedasticity.

Referring back to the ordinary least squares regression, we now compute the mean squared residual, 1911.9275/50 = 38.23855. Then, we compute $v_i = (1/38.23855)e_i^2$ for each observation. In the regression of $\mathbf{v}$ on a constant, X_1, and X_2, the regression sum of squares is 145.551, so the chi-squared statistic is 145.551/2 = 72.775. We reach the same conclusion as in the previous paragraph. In this case, the degrees of freedom for the test are only two, so the conclusion is somewhat stronger.

To carry out the Goldfeld-Quandt test, we order the data first based on X_1 then on X_2. The regressions are computed using the first and last 17 observations, so the F statistic in each case is $F[14,14] = \mathbf{e}_1'\mathbf{e}_1 / \mathbf{e}_2'\mathbf{e}_2$ where $\mathbf{e}_1'\mathbf{e}_1$ is the larger of the two sums of squares and $\mathbf{e}_2'\mathbf{e}_2$ is the smaller. For our data set, we find

	Sorted on X_1	Sorted on X_2
$\mathbf{e}'\mathbf{e}$ for obs. 1-17	385.974	246.026
$\mathbf{e}'\mathbf{e}$ for obs. 34-50	273.404	1165.683
$F[14,14]$	1.412	4.738

The second is considerably larger than the critical value of 2.48. We would conclude, therefore, that there is evidence of heteroscedasticity and it is related to X_2 but not X_1. In view of this finding, it is instructive to go back to the White and Breusch and Pagan tests considered earlier. Consider the Lagrange multiplier test, first. In the regression of the squared residuals from the original regression divided by $\mathbf{e}'\mathbf{e}/n$ first on 1 and X_1 then on 1 and X_2, the regression sums of squares are .01805 and 105.906, respectively. Therefore, the chi-squared statistics are .009025 and 52.953, respectively. The implication is, once again, that there is substantial heteroscedasticity, of the form $\sigma_i^2 = \sigma^2 h(1+\gamma X_{i2})$. The White test involves regressing the squared residuals first on 1, X_1, and X_1^2 then on 1, X_2, and X_2^2. The R^2s in these two regressions are .02216 and .61380, respectively. The test statistics are thus 1.108 and 30.69. The conclusion is the same.

Finally, we compute Glesjer's test statistics for the three models discussed in Section 14.3.5. We regress e^2, $|e|$, and $\log|e|$ on 1, X_1, and X_2. We use the White estimator for the covariance matrix of the parameter estimates in these regressions as there is ample evidence now that the disturbances are heteroscedastic related to X_2. To compute the Wald statistic, we need the two slope coefficients, which we denote $\mathbf{q}$, and the 2×2 submatrix of the 3×3 covariance matrix of the coefficients, which we denote $\mathbf{V}_q$. The statistic is $W = \mathbf{q}'\mathbf{V}_q^{-1}\mathbf{q}$. For the three regressions, the values are 4.13, 6.51, and 6.60, respectively. The critical value from the chi-squared distribution with 2 degrees of freedom is 5.99, so the second and third are statistically significant while the first is not.

The disturbance variance underlying these data is, in fact, $\sigma_i^2 = \sigma^2(1 + \gamma X_{i2}^2)$ so the Goldfeld-Quandt and Glejser tests have given the right diagnosis. For the Glejser test, the finding that the linear model is inappropriate makes sense since X_2 takes negative values.

6. Using the data of Exercise 5, reestimate the parameters, using a two step feasible GLS estimator. Try (12-19), (12-20), and (12-21). Which one appears to be most appropriate?

The ordinary least squares estimates are given above. The estimates of the disturbance variances are based on the residuals from this regression. For the three models, the disturbance variances are estimated as follows:

(1) $\sigma_i^2 = \sigma^2(\boldsymbol{\alpha}'\mathbf{z}_i)$: Regress e^2 on 1, X_1, X_2. Estimates of σ_i^2 are the fitted values in this regression. This produces numerous negative values and is clearly inappropriate.

(2) $\sigma_i^2 = \sigma^2(\boldsymbol{\alpha}'\mathbf{z}_i)^2$: Regress $|e_i|$ on 1, X_1, and X_2. The estimates of σ_i^2 are the squares of the predicted values from this regression.

(3) $\sigma_i^2 = \sigma^2\exp(\boldsymbol{\alpha}'\mathbf{z}_i)$: Regress $\log(e_i^2)$ on 1, X_1, and X_2. The estimates of σ_i^2 are the exponents of the fitted values in this regression.

Weighted least squares regressions based on the second and third sets of weights ($1/\sigma_i^2$) produces the following estimates (standard errors are shown in parentheses). In each case, the weights are the reciprocals of the estimated standard deviations as described above.

	Unweighted	(2)	(3)
One	.190394 (.9144)	1.48129 (1.817)	.166626 (.7198)
X_1	1.13113 (.9826)	1.44651 (.9716)	.776487 (.6388)
X_2	.376825 (.4399)	.894613 (.7451)	.847177 (.3633)

There is little in the way of guidance as to which model is the better one. The OLS estimates are suggestive since they are consistent under all specifications. The second set of estimates resemble the OLS estimates slightly more than the third. As we discussed above, model (2) is, in fact, the right one. Unfortunately, one would be hard pressed to reach that as a firm conclusion based on just these results. Of course, the results of the tests in the previous exercise are much more convincing. □

7. For the model in Exercise 1, suppose , is normally distributed with mean zero and variance $\sigma^2(1 + (\gamma x)^2)$. Show that σ^2 and γ^2 can be consistently estimated by a regression of the least squares residuals on a constant and x^2. Is this estimator efficient?

The residuals from the least squares regression are $e_i = y_i - \bar{y} = \alpha + \varepsilon_i - (\alpha + \bar{\varepsilon}) = \varepsilon_i - \bar{\varepsilon}$. The expected value of the squared residual is

$$E[e_i^2] = E[\varepsilon_i^2] + E[\bar{\varepsilon}^2] - 2E[\varepsilon_i\bar{\varepsilon}] = \sigma_i^2 + (1/n^2)E[(\Sigma_i\varepsilon_i)^2] - (2/n)E[\varepsilon_i(\Sigma_j\varepsilon_j)]$$
$$= \sigma_i^2 + (1/n^2)\Sigma_i E[\varepsilon_i^2] - (2/n)E[\varepsilon_i^2]$$

since the disturbances are uncorrelated. We can write this as

$$E[e_i^2] = \sigma^2 + \sigma^2\gamma^2 x_i^2 + (1/n)\{[(1/n)\Sigma_i\sigma_i^2] - [2\sigma_i^2]\}.$$

And, of course, $e_i^2 = E[e_i^2] + (e_i^2 - E[e_i^2]) = E[e_i^2] + v_i$, where v_i is uncorrelated with $E[e_i^2]$ by construction. Now, if we regress e_i^2 on a constant and x_i^2, the estimates of σ^2 and $(\sigma^2\gamma^2)$ will be biased in a finite sample because of the left out variable, namely the term multiplied by $(1/n)$ in the expression for $E[e_i^2]$. But, if the two terms inside the curled brackets above converge to finite quantities as $n\rightarrow\infty$, then the entire term will vanish, and the omitted variable problem will vanish with it. Surely the second does since it is the variance of ε_i^2, assuming that x_i^2 is finite. To make the first converge, we will require that $(1/n)\Sigma_i[\sigma^2 + \sigma^2\gamma^2 x_i^2] = \sigma^2 + \sigma^2\gamma^2(1/n)\Sigma_i x_i^2$ converge to a finite quantity, or that the mean square of the xs converge to a finite quantity. This is a minimal requirement for a heteroscedastic regression, and would surely be met. As such, if e_i^2 is regressed on a constant and x_i^2, we obtain consistent estimators of σ^2 and $\sigma^2\gamma^2$. The estimator of σ is the square root of the ratio of the slope to the constant.

The estimator is not efficient. The expected fourth moment of a normally distributed variable is 3 times the square of the variance. Therefore, in the regression above, the variance of v_i must be a function of $[\sigma^2(1 + \gamma^2 x_i^2)]^2$. Since the regression is heteroscedastic in a way which is not dependent on the sample size, OLS will not be efficient, but it will be consistent.

8. Derive the log-likelihood function, first order conditions for maximization, and information matrix for the model $y_i = \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i$, $\varepsilon_i \sim N[0,\sigma^2(\boldsymbol{\gamma}'\mathbf{z}_i)^2]$.

First, $\log L = -(n/2)\log(2\pi) - 2\Sigma_i \log\sigma_i^2 - 2\Sigma_i(\varepsilon_i^2/\sigma_i^2)$

$= -(n/2)\ln(2\pi) - (n/2)\log\sigma^2 - 2\Sigma_i\log(\boldsymbol{\gamma}'\mathbf{z}_i)^2 - [1/(2\sigma^2)]\Sigma_i[(y_i - \boldsymbol{\beta}'\mathbf{x}_i)^2/(\boldsymbol{\gamma}'\mathbf{z}_i)^2]$

$\partial\log L/\partial\boldsymbol{\beta} = -(1/(2\sigma^2))\Sigma_i(1/(\boldsymbol{\gamma}'\mathbf{z}_i)^2)(2)(y_i - \boldsymbol{\beta}'\mathbf{x}_i)(-\mathbf{x}_i) = \Sigma_i\{\varepsilon_i/[\sigma^2(\boldsymbol{\gamma}'\mathbf{z}_i)^2]\}\mathbf{x}_i$

$\partial\log L/\partial\boldsymbol{\gamma} = -2\Sigma_i(1/(\boldsymbol{\gamma}'\mathbf{z}_i)^2)(2\boldsymbol{\gamma}'\mathbf{z}_i)\mathbf{z}_i - (1/(2\sigma^2))\Sigma_i(y_i - \boldsymbol{\beta}'\mathbf{x}_i)^2[-2/(\boldsymbol{\gamma}'\mathbf{z}_i)^3]\mathbf{z}_i$

$= -\Sigma_i[1/(\boldsymbol{\gamma}'\mathbf{z}_i)]\mathbf{z}_i + (1/\sigma^2)\Sigma_i[\varepsilon_i^2/(\boldsymbol{\gamma}'\mathbf{z}_i)^2][1/(\boldsymbol{\gamma}'\mathbf{z}_i)]\mathbf{z}_i$

$= \Sigma_i\{\varepsilon_i^2/[\sigma^2(\boldsymbol{\gamma}'\mathbf{z}_i)^2] - 1\}[1/(\boldsymbol{\gamma}'\mathbf{z}_i)]\mathbf{z}_i$

It is useful to note in passing that we can write this as $= 2\Sigma_i\{\varepsilon_i^2/[\sigma^2(\boldsymbol{\gamma}'\mathbf{z}_i)^2] - 1\}[1/(\boldsymbol{\gamma}'\mathbf{z}_i)^2]2(\boldsymbol{\gamma}'\mathbf{z}_i)\mathbf{z}_i$ which is what appears in (12-28). It is obvious that these first derivatives have expectation zero, the first since $E[\varepsilon_i] = 0$ and the second because the first term in large brackets has expectation zero. Finally,

$\partial\log L/\partial\sigma^2 = -n/(2\sigma^2) + (1/(2\sigma^4))\Sigma_i[\varepsilon_i^2/(\boldsymbol{\gamma}'\mathbf{z}_i)^2)]$.

Since $E[\varepsilon_i^2] = \sigma^2(\boldsymbol{\gamma}'\mathbf{z}_i)^2$, this also has expected value zero. The second derivatives are

$\partial^2\log L/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}' = \Sigma_i\mathbf{x}_i(\partial\varepsilon_i/\partial\boldsymbol{\beta}')/[\sigma^2(\boldsymbol{\gamma}'\mathbf{z}_i)^2] = -\Sigma_i[1/\sigma^2(\boldsymbol{\gamma}'\mathbf{z}_i)^2]\mathbf{x}_i\mathbf{x}_i'$

$\partial^2\log L/\partial\boldsymbol{\beta}\partial\sigma^2 = \Sigma_i\{-\varepsilon_i/[(\sigma^2)^2(\boldsymbol{\gamma}'\mathbf{z}_i)^2]\}\mathbf{x}_i$

$\partial^2\log L/\partial\boldsymbol{\beta}\partial\boldsymbol{\gamma}' = \Sigma E_i\{-2\varepsilon_i/[\sigma^2(\boldsymbol{\gamma}'\mathbf{z}_i)^3]\}\mathbf{x}_i\mathbf{z}_i'$

$\partial^2\log L/\partial(\sigma^2)^2 = n/(2\sigma^4) - (1/\sigma^4)\Sigma_i\{\varepsilon_i^2/[\sigma^2(\boldsymbol{\gamma}'\mathbf{z}_i)^2]\}$

$\partial^2\log L/\partial F^2\partial\boldsymbol{\gamma}' = [1/(2\sigma^4)]\Sigma_i\{\varepsilon_i^2[-2/(\boldsymbol{\gamma}'\mathbf{z}_i)^3)]\}\mathbf{z}_i = [-1/(2\sigma^2)]\Sigma_i\{\varepsilon_i^2[-2/(\boldsymbol{\gamma}'\mathbf{z}_i)^2)]\}\{1/(\boldsymbol{\gamma}'\mathbf{z}_i)^2\}\{2\boldsymbol{\gamma}'\mathbf{z}_i\}\mathbf{z}_i$

$\partial^2\log L/\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}' = \Sigma_i\{\varepsilon_i^2/[\sigma^2(\boldsymbol{\gamma}'\mathbf{z}_i)^2] - 1\}[-1/(\boldsymbol{\gamma}'\mathbf{z}_i)^2]\mathbf{z}_i\mathbf{z}_i' + \Sigma_i\{-2\varepsilon_i^2/[\sigma^2(\boldsymbol{\gamma}'\mathbf{z}_i)^3]\}[1/(\boldsymbol{\gamma}'\mathbf{z}_i)]\mathbf{z}_i\mathbf{z}_i'$

In the notation of (12-26) - (12-28), $f_i = (\boldsymbol{\gamma}'\mathbf{z}_i)^2$ and $\mathbf{g}_i = 2(\boldsymbol{\gamma}'\mathbf{z}_i)\mathbf{z}_i$. Therefore, $\boldsymbol{\Omega} = \text{diag}[(\boldsymbol{\gamma}'\mathbf{z}_i)^2]$ and $\mathbf{G}$ = the matrix whose rows are $\mathbf{g}_i'$. The negatives of the expected second derivatives are simple to obtain since $E[\varepsilon_i] = 0$ and $E[\varepsilon_i^2] = \sigma^2(\boldsymbol{\gamma}'\mathbf{z}_i)^2$. We will write them in the form of (14-29). Thus,

$-E[\partial^2\log L/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'] = (1/\sigma^2)\Sigma_i[1/(\boldsymbol{\gamma}'\mathbf{z}_i)^2]\mathbf{x}_i\mathbf{x}_i'$

$-E[\partial^2\log L/\partial\boldsymbol{\beta}\partial\sigma^2] = \mathbf{0}$

$-E[\partial^2\log L/\partial\boldsymbol{\beta}\partial\boldsymbol{\gamma}'] = \mathbf{0}$

$-E[\partial^2\log L/\partial(\sigma^2)^2] = n/(2\sigma^4)$

$-E[\partial^2\log L/\partial\sigma^2\partial\boldsymbol{\gamma}'] = [1/2\sigma^2]\Sigma_i\{1/(\boldsymbol{\gamma}'\mathbf{z}_i)^2\}2(\boldsymbol{\gamma}'\mathbf{z}_i)\mathbf{z}_i = [1/\sigma^2]\Sigma_i\{1/(\boldsymbol{\gamma}'\mathbf{z}_i)\}\mathbf{z}_i$

$-E[\partial^2\log L/\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}'] = 2\Sigma_i\{1/(\boldsymbol{\gamma}'\mathbf{z}_i)^4\}[2(\boldsymbol{\gamma}'\mathbf{z}_i)\mathbf{z}_i][2(\boldsymbol{\gamma}'\mathbf{z}_i)\mathbf{z}_i]' = 2\Sigma_i[1/(\boldsymbol{\gamma}'\mathbf{z}_i)^2]\mathbf{z}_i\mathbf{z}_i'$. □

9. See Exercise 7 of Chapter 10.

10. In the discussion of Harvey's model in Section 11.7, it is noted that the initial estimator of γ_1, the constant term in the regression of $\log e_i^2$ on a constant and $\mathbf{z}_i$ is inconsistent by the amount 1.2704. Harvey points out that if the purpose of this initial regression is only to obtain starting values for the interations, then the correction is not necessary. Explain why this would be the case.

The constant is just an estimate of σ^2 in $\sigma^2\boldsymbol{\Omega}$. Let $\mathbf{W}$ equal $\sigma^2\boldsymbol{\Omega}$. Then, the GLS estimator is $\hat{\boldsymbol{\beta}} = [\mathbf{X}'\mathbf{W}^{-1}\mathbf{X}]^{-1}[\mathbf{X}'\mathbf{W}^{-1}\mathbf{y}]$. The scale factor is immaterial. The estimator will be the same whether $\mathbf{W}$ is scaled or not, since the scale factor will fall out of the result.

11. (This exercise requires appropriate computer software. The computations required can be done with *RATS*, *EViews*, *Stata*, *TSP, LIMDEP*, and a variety of other software using only preprogrammed procedures.) Quarterly data on the consumer price index for 1950.1 to 2000.4 are given in Appendix Table F5.1. Use these data to fit the model proposed by Engle and Kraft (1983). The model is

$$\pi_t = \beta_0 + \beta_1\pi_{t-1} + \beta_2\pi_{t-2} + \beta_3\pi_{t-3} + \beta_4\pi_{t-4} + \varepsilon_t$$

where $\pi_t = 100\ln[p_t/p_{t-1}]$ and p_t is the price index.

a. Fit the model by ordinary least squares, then use the tests suggested in the text to see if ARCH effects appear to be present.

b. The authors fit an ARCH(8) model with declining weights,

$$\sigma_t^2 = \alpha_0 + \Sigma_{i=1}^8\left(\frac{9-i}{36}\right)\varepsilon_{t-i}^2$$

Fit this model. If the software does not allow constraints on the coefficients, you can still do this with a two step least squares procedure, using the least squares residuals from the first step. What do you find?

c. Bollerslev (1986) recomputed this model as a GARCH(1,1). Use the GARCH(1,1) form and refit your model.

a. We used LIMDEP with the macroeconomics data in table F5.1. The rate of inflation was computed with all observations, then observations 6 to 204 were used to remove the missing data due to lags. Least squares results were obtained first. The residuals were then computed and squared. Using observations 15-204, we then computed a regression of the squared residual on a constant and 8 lagged values. The chi-squared statistic with 8 degrees of freedom is 28.24. The critical value from the table for 95% significance and 8 degrees of freedom is 15.51, so at this level of significance, the hypothesis of no GARCH effects is rejected.

```
crea;pt=100*log(cpi_u/cpi_u[-1])$
crea;pt1=pt[-1];pt2=pt[-2];pt3=pt[-3];pt4=pt[-4]$
samp;6-204$
regr;lhs=pt;rhs=one,pt1,pt2,pt3,pt4;res=et$$
crea;vt=et*et$
crea;vt1=vt[-1];vt2=vt[-2];vt3=vt[-3];vt4=vt[-4];vt5=vt[-5];vt6=vt[-6];vt7=vt[-
7];vt8=vt[-8]$
samp;15-204$
regr;lhs=vt;rhs=one,vt1,vt2,vt3,vt4,vt5,vt6,vt7,vt8$
calc;list;lm=n*rsqrd$
+-----------------------------------------------------------------------+
| Ordinary    least squares regression    Weighting variable = none     |
| Dep. var. = PT       Mean=   .9589185961    , S.D.=   .8318268241     |
| Model size: Observations =     199, Parameters =   5, Deg.Fr.=    194 |
| Residuals:  Sum of squares= 61.97028507    , Std.Dev.=         .56519 |
| Fit:        R-squared=  .547673, Adjusted R-squared =          .53835 |
| Model test: F[  4,    194] =   58.72,    Prob value =          .00000 |
| Diagnostic: Log-L =   -166.2871, Restricted(b=0) Log-L =    -245.2254 |
|             LogAmemiyaPrCrt.=   -1.116, Akaike Info. Crt.=      1.721 |
| Autocorrel: Durbin-Watson Statistic =   1.80740,   Rho =       .09630 |
+-----------------------------------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |t-ratio |P[|T|>t] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant     .1296044455   .67521735E-01    1.919   .0564
 PT1          .2856136998   .69863942E-01    4.088   .0001     .97399582
 PT2          .1237760914   .70647061E-01    1.752   .0813     .98184918
 PT3          .2516837602   .70327318E-01    3.579   .0004     .99074774
 PT4          .1824670634   .69251374E-01    2.635   .0091     .98781131
 LM      =  .28240022492847690D+02
```

For the second step, we need an estimate of α_0, which is the unconditional variance if there are no ARCH effects. We computed this based on the ARCH specification by a regression of $e_t^2 - (8/36)e_{t-1}^2 - \ldots - (1/36)e_{t-8}^2$ on just a constant term. This produces a negative estimate of α_0, but this is not the variance, so we retain the result. We note, the problem that this reflects is probably the specific, doubtless unduly restrictive, ARCH structure assumed.

```
samp;6-204$
crea;vt=et*et$
crea;ht=vt-8/36*vt[-1]-7/36*vt[-2]-6/36*vt[-3]-5/36*vt[-4]-4/36*vt[-5]-3/36*vt[-
6]-2/36*vt[-7]-1/36*vt[-8]$
samp;15-204$
calc;list;a0=xbr(ht)$
samp;6-204$
crea;qt=a0+8/36*vt[-1]+7/36*vt[-2]+6/36*vt[-3]+5/36*vt[-4]+4/36*vt[-5]+3/36*vt[-
6]+2/36*vt[-7]+1/36*vt[-8]$
samp;15-204$
plot;rhs=qt$
crea;wt=1/qt$
regr;lhs=pt;rhs=one,pt1,pt2,pt3,pt4;wts=wt$
regr;lhs=pt;rhs=one,pt1,pt2,pt3,pt4;model=garch(1,1)$
```

Once we have an estimate of α_0 in hand, we then computed the set of variances according to the ARCH(8) model, using the lagged squared residuals. Finally, we used these variance estimators to compute a weighted least squares regression accounting for the heteroscedasticity. This regression is based on observations 15-204, again because of the lagged values. Finally, using the same sample, a GARCH(1,1) model is fit by maximum likelihood.

```
+-----------------------------------------------------------------------+
| Ordinary    least squares regression    Weighting variable = WT       |
| Dep. var. = PT       Mean=   .8006997687    , S.D.=   .6327877239     |
| Model size: Observations =     190, Parameters =   5, Deg.Fr.=   185 |
| Residuals:  Sum of squares= 38.67492770    , Std.Dev.=         .45722 |
| Fit:        R-squared=  .488964, Adjusted R-squared =          .47791 |
| Model test: F[  4,    185] =   44.25,    Prob value =          .00000 |
| Diagnostic: Log-L =   -147.7324, Restricted(b=0) Log-L =    -211.5074 |
|             LogAmemiyaPrCrt.=   -1.539, Akaike Info. Crt.=      1.608 |
| Autocorrel: Durbin-Watson Statistic =   1.90310,   Rho =       .04845 |
+-----------------------------------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |t-ratio |P[|T|>t] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant     .1468553158   .60127085E-01    2.442   .0155
 PT1      .9760051110E-01  .88469908E-01    1.103   .2714     .77755556
 PT2          .3328520370   .86772549E-01    3.836   .0002     .76745308
 PT3          .1428889148   .85420554E-01    1.673   .0961     .76271761
 PT4          .2878686524   .84090832E-01    3.423   .0008     .74173558
```

The 8 period ARCH model produces quite a substantial change in the estimates. Once again, this probably results from the restrictive assumption about the lag weights in the ARCH model. The GARCH model follows.

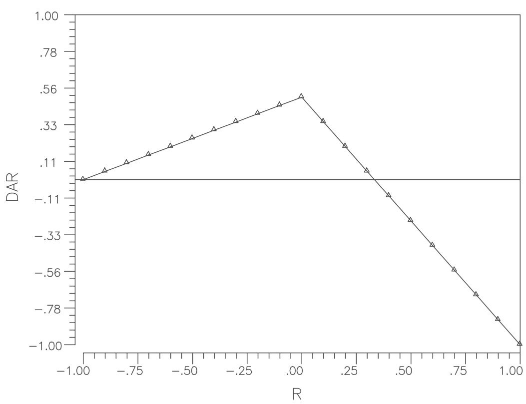
```
+---------------------------------------------+
| GARCH MODEL                                 |
| Maximum Likelihood Estimates                |
| Model estimated: Jul 31, 2002 at 01:19:14PM.|
| Dependent variable                  PT      |
| Weighting variable                None      |
| Number of observations             190      |
| Iterations completed                22      |
| Log likelihood function      -135.5043      |
| Restricted log likelihood    -147.6465      |
| Chi squared                   24.28447      |
| Degrees of freedom                   2      |
| Prob[ChiSqd > value] =        .5328953E-05 |
| GARCH Model, P = 1, Q = 1                   |
| Wald statistic for GARCH =      521.483     |
+---------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
          Regression parameters
 Constant     .1308478127   .61887183E-01    2.114   .0345
 PT1          .1749239917   .70912277E-01    2.467   .0136     .98810078
 PT2          .2532191617   .73228319E-01    3.458   .0005     .98160455
 PT3          .1552879436   .68274176E-01    2.274   .0229     .97782066
 PT4          .2751467919   .63910272E-01    4.305   .0000     .97277700
          Unconditional Variance
 Alpha(0)  .1005125676E-01  .11653271E-01     .863   .3884
          Lagged Variance Terms
 Delta(1)     .8556879884   .89322732E-01    9.580   .0000
          Lagged Squared Disturbance Terms
 Alpha(1)     .1077364862   .60761132E-01    1.773   .0762
          Equilibrium variance, a0/[1-D(1)-A(1)]
 EquilVar     .2748082674       2.0559946     .134   .8937
```

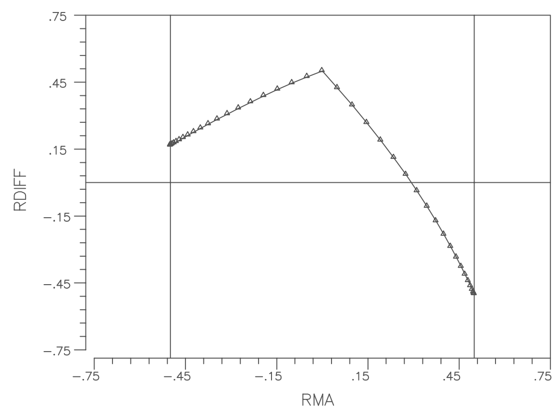
Chapter 12

Autocorrelation

1. Does first differencing reduce autocorrelation? Consider the models $y_t = \boldsymbol{\beta}'\mathbf{x}_t + \varepsilon_t$, where $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$ and $\varepsilon_t = u_t - \lambda u_{t-1}$. Compare the autocorrelation of ε_t in the original model to that of v_t in $y_t - y_{t-1} = \boldsymbol{\beta}'(\mathbf{x}_t - \mathbf{x}_{t-1}) + v_t$ where $v_t = \varepsilon_t - \varepsilon_{t-1}$.

For the first order autoregressive model, the autocorrelation is ρ. Consider the first difference, $v_t = \varepsilon_t - \varepsilon_{t-1}$ which has $\text{Var}[v_t] = 2\text{Var}[\varepsilon_t] - 2\text{Cov}[(\varepsilon_t,\varepsilon_{t-1})] = 2\sigma_u^2[1/(1-\rho^2) - \rho/(1-\rho^2)] = 2\sigma_u^2/(1+\rho)$ and $\text{Cov}[v_t,v_{t-1}] = 2\text{Cov}[\varepsilon_t,\varepsilon_{t-1}] - \text{Var}[\varepsilon_t] - \text{Cov}[\varepsilon_t,\varepsilon_{t-1}] = \sigma_u^2[1/(1-\rho^2)][2\rho - 1 - \rho^2] = \sigma_u^2[(\rho-1)/(1+\rho)]$. Therefore, the autocorrelation of the differenced process is $\text{Cov}[v_t,v_{t-1}] / \text{Var}[v_t] = (\rho - 1)/2$. As the figure below on the left shows, first differencing reduces the absolute value of the autocorrelation coefficient when ρ is greater than 1/3. For economic data, this is likely to be fairly common.





For the moving average process, the first order autocorrelation is $\text{Cov}[(\varepsilon_t,\varepsilon_{t-1})]/\text{Var}[\varepsilon_t] = -\lambda/(1+\lambda^2)$. To obtain the autocorrelation of the first difference, write $\varepsilon_t - \varepsilon_{t-1} = u_t - (1+\lambda)u_{t-1} + \lambda u_{t-2}$ and $\varepsilon_{t-1} - \varepsilon_{t-2} = u_{t-1} - (1+\lambda)u_{t-2} + \lambda u_{t-3}$. The variance of the difference is $\text{Var}[\varepsilon_t - \varepsilon_{t-1}] = \sigma_u^2[(1+\lambda)^2 + (1+\lambda^2)]$. The covariance can be found by taking the expected product of terms with equal subscripts. Thus, $\text{Cov}[\varepsilon_t - \varepsilon_{t-1},\varepsilon_{t-1} - \varepsilon_{t-2}] = -\sigma_u^2(1+\lambda)^2$. The autocorrelation is $\text{Cov}[\varepsilon_t - \varepsilon_{t-1},\varepsilon_{t-1} - \varepsilon_{t-2}]/\text{Var}[\varepsilon_t - \varepsilon_{t-1}] = -(1+\lambda)^2/[(1+\lambda)^2 + (1+\lambda^2)]$. A plot of the relationship between the differenced and undifferenced series is shown in the right panel above. The horizontal axis plots the autocorrelation of the original series. The values plotted are the absolute values of the difference between the autocorrelation of the differenced series and the original series. The results are similar to those for the AR(1) model. For most of the range of the autocorrelation of the original series, differencing increases autocorrelation. But, for most of the range of values that are economically meaningful, differencing reduces autocorrelation.

2. Derive the disturbance covariance matrix for the model $y_t = \boldsymbol{\beta}'\mathbf{x}_t + \varepsilon_t$, $\varepsilon_t = \rho\varepsilon_{t-1} + u_t - \lambda u_{t-1}$. What parameter is estimated by the regression of the ordinary least squares residuals on their lagged values?

Solve the disturbance process in its moving average form. Write the process as $\varepsilon_t - \rho\varepsilon_{t-1} = u_t - \lambda u_{t-1}$ or, using the lag operator, $\varepsilon_t(1-\rho L) = u_t - \lambda u_{t-1}$ or $\varepsilon_t = u_t/(1-\rho L) - \lambda u_{t-1}/(1-\rho L)$. After multiplying these out, we obtain

$$\begin{aligned}
\varepsilon_t &= u_t + \rho u_{t-1} + \rho^2 u_{t-2} + \rho^3 u_{t-3} + \ldots - \lambda u_{t-1} - \rho\lambda u_{t-2} - \rho^2\lambda u_{t-3} - \ldots \\
&= u_t + (\rho-\lambda)u_{t-1} + \rho(\rho-\lambda)u_{t-2} + \rho^2(\rho-\lambda)u_{t-3} + \ldots
\end{aligned}$$

Therefore,

$$\begin{aligned}
\text{Var}[\varepsilon_t] &= \sigma_u^2(1 + (\rho-\lambda)^2)(1 + \rho^2 + \rho^4 + \ldots) = \sigma_u^2(1 + (\rho-\lambda)^2/(1-\rho^2)) \\
&= \sigma_u^2(1 + \lambda^2 - 2\rho\lambda)/(1-\rho^2) \\
\text{Cov}[\varepsilon_t,\varepsilon_{t-1}] &= \rho\text{Var}[\varepsilon_{t-1}] + \text{Cov}[\varepsilon_{t-1},u_t] - \lambda\text{Cov}[\varepsilon_{t-1},u_{t-1}].
\end{aligned}$$

To evaluate this expression, write

$$\varepsilon_{t-1} = u_{t-1} + (\rho-\lambda)u_{t-2} + \rho(\rho-\lambda)u_{t-3} + \rho^2(\rho-\lambda)u_{t-4} + \ldots$$

Therefore, the middle term is zero and the third is simply $\lambda\sigma_u^2$. Thus,

$$\text{Cov}[\varepsilon_t,\varepsilon_{t-1}] = \sigma_u^2\{[\rho(1 + \lambda^2 - 2\rho\lambda)]/(1 - \rho^2) - \lambda]\} = \sigma_u^2[(\rho - \lambda)(1 - \lambda\rho)/(1 - \rho^2)]$$

For lags greater than 1, $\text{Cov}[\varepsilon_t,\varepsilon_{t-j}] = \rho\text{Cov}[\varepsilon_{t-1},\varepsilon_{t-j}] + \text{Cov}[\varepsilon_{t-j},u_t] - \lambda\text{Cov}[\varepsilon_{t-j},u_{t-1}]$.
Since ε_{t-j} involves only us up to its current period, ε_{t-j} is uncorrelated with u_t and u_{t-1} if j is greater than 1. Therefore, after the first lag, the autocovariances behave in the familiar fashion, $\text{Cov}[\varepsilon_t,\varepsilon_{t-j}] = \rho\text{Cov}[\varepsilon_t,\varepsilon_{t-j+1}]$
The autocorrelation coefficient of the residuals estimates $\text{Cov}[\varepsilon_t,\varepsilon_{t-1}]/\text{Var}[\varepsilon_t] = (\rho - \lambda)(1 - \rho\lambda)/(1 + \lambda^2 - 2\rho\lambda)$.

3. The following regression is obtained by ordinary least squares using 21 observations: (Estimated asymptotic standard deviations are shown in parentheses). $y_t = 1.3(0.3) + .97(0.18)y_{t-1} + 2.31(1.04)x_t$, DW = 1.21. Test for the presence of autocorrelation in the disturbances.

Since the regression contains a lagged dependent variable, we cannot use the Durbin-Watson statistic directly. The h statistic in (15-34) would be $h = (1 - 1.21/2)[21 / (1 - 21(.18^2)]^2 = 3.201$.
The 95% critical value from the standard normal distribution for this one-tailed test would be 1.645. Therefore, we would reject the hypothesis of no autocorrelation.

4. It is commonly asserted that the Durbin-Watson statistic is only appropriate for testing for first order autoregressive disturbances. What combination of the coefficients of the model is estimated by the Durbin-Watson statistic in each of the following cases: AR(1), AR(2), MA(1)? In each case, assume that the regression model does not contain a lagged dependent variable. Comment on the impact on your results of relaxing this assumption.

In each case, plim $d = 2 - 2\rho_1$ where $\rho_1 = \text{Corr}[\varepsilon_t,\varepsilon_{t-1}]$. The first order autocorrelations are as follows: AR(1): ρ (see (15-9)) and AR(2): $\theta_1/(1 - \theta_2)$. For the AR(2), a proof is as follows: First, $\varepsilon_t = \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + u_t$. Denote $\text{Var}[\varepsilon_t]$ as c_0 and $\text{Cov}[\varepsilon_t,\varepsilon_{t-1}]$ as c_1. Then, it follows immediately that $c_1 = \theta_1 c_0 + \theta_2 c_1$ since u_t is independent of ε_{t-1}. Therefore $\rho_1 = c_1/c_0 = \theta_1/(1 - \theta_2)$. For the MA(1): $-\lambda / (1 + \lambda^2)$ (See (15-43)). To prove this, write $\varepsilon_t = u_t - \lambda u_{t-1}$. Then, since the us are independent, the result follows just by multiplying out $\rho_1 = \text{Cov}[\varepsilon_t,\varepsilon_{t-1}]/\text{Var}[\varepsilon_t] = -\lambda\text{Var}[u_{t-1}]/\{\text{Var}[u_t] + \lambda^2\text{Var}[u_{t-1}]\} = -\lambda/(1 + \lambda^2)$.

5. The data used to fit the expectations augmented Phillips curve in Example 12.3 are given in Table F5.1. Using these data, reestimate the model given in the example. Carry out a formal test for first order autocorrelation using the LM statistic. Then, reestimate the model using an AR(1) model for the disturbance process. Since the sample is large, the Prais-Winsten and Cochrane-Orcutt estimators should give essentially the same answer. Do they? After fitting the model, obtain the transformed residuals and examine them for first order autocorrelation. Does the AR(1) model appear to have adequately "fixed" the problem?

```
--> date;1950.1$
--> peri;1950.1-2000.4$
--> crea;dp=infl-infl[-1]$
--> crea;dy=loggdp-loggdp[-1]$
--> peri;1950.3-2000.4$
--> regr;lhs=dp;rhs=one,unemp$;ar1;res=u$
```

```
+-----------------------------------------------------------------------+
| Ordinary    least squares regression    Weighting variable = none     |
| Dep. var. = DP       Mean=  -.1926996283E-01, S.D.=    2.818214558     |
| Model size: Observations =     202, Parameters =   2, Deg.Fr.=    200 |
| Residuals:  Sum of squares= 1592.321197    , Std.Dev.=        2.82163 |
| Fit:        R-squared=  .002561, Adjusted R-squared =        -.00243 |
| Model test: F[  1,    200] =      .51,    Prob value =         .47449 |
| Diagnostic: Log-L =   -495.1583, Restricted(b=0) Log-L =    -495.4173 |
|             LogAmemiyaPrCrt.=    2.084, Akaike Info. Crt.=      4.922 |
| Autocorrel: Durbin-Watson Statistic =   2.82755,   Rho =      -.41378 |
+-----------------------------------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |t-ratio |P[|T|>t] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant     .4918922148       .74047944      .664   .5073
 UNEMP    -.9013159906E-01      .12578616     -.717   .4745     5.6712871
```

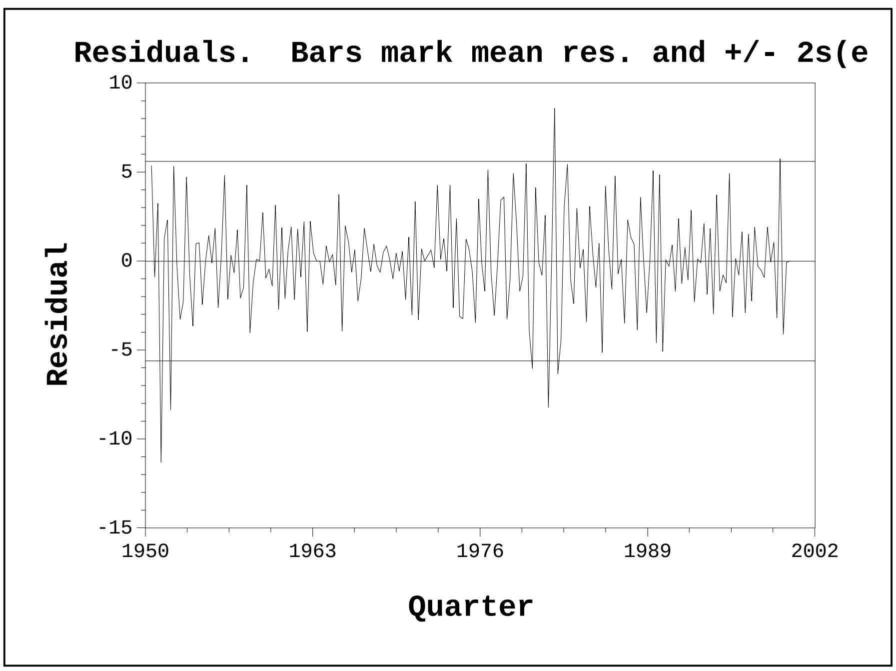
```
--> peri;1951.2-2000.4$
--> regr;lhs=u;rhs=one,u[-1],u[-2]$

+-----------------------------------------------------------------------+
| Ordinary    least squares regression    Weighting variable = none     |
| Dep. var. = U        Mean=  -.3890391012E-01, S.D.=   2.799476915     |
| Model size: Observations =     199, Parameters =   3, Deg.Fr.=    196 |
| Residuals:  Sum of squares= 1079.052269    , Std.Dev.=        2.34635 |
| Fit:        R-squared=  .304618, Adjusted R-squared =          .29752 |
| Model test: F[  2,    196] =   42.93,    Prob value =          .00000 |
| Diagnostic: Log-L =   -450.5769, Restricted(b=0) Log-L =    -486.7246 |
|             LogAmemiyaPrCrt.=    1.721, Akaike Info. Crt.=      4.559 |
| Autocorrel: Durbin-Watson Statistic =   1.99273,   Rho =       .00363 |
+-----------------------------------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |t-ratio |P[|T|>t] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant -.5048615289E-01      .16633422     -.304   .7618
 U[-1]       -.5946344724   .65920584E-01   -9.020   .0000 -.10234931E-01
 U[-2]       -.3824653303   .65904378E-01   -5.803   .0000 -.14370453E-01
 (Note: E+nn or E-nn means multiply by 10 to + or -nn power.)

--> calc;list;lm=n*rsqrd$
    LM      =  .60618960968412850D+02
+---------------------------------------------+
| AR(1) Model:      e(t) = rho * e(t-1) + u(t) |
| Initial value of rho       =        -.41378 |
| Maximum iterations         =            100 |
| Method = Prais - Winsten                    |
| Iter=  1, SS=   1299.275, Log-L=-474.710175 |
| Final value of Rho    =            -.413779 |
| Iter=  1, SS=   1299.275, Log-L=-474.710175 |
| Durbin-Watson:   e(t) =            2.827557 |
| Std. Deviation:  e(t) =            2.799716 |
| Std. Deviation:  u(t) =            2.548799 |
| Durbin-Watson:   u(t) =            2.340706 |
| Autocorrelation: u(t) =            -.170353 |
| N[0,1] used for significance levels         |
+---------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant      .4704274598       .47671946      .987   .3237
 UNEMP    -.8709854633E-01  .80962277E-01   -1.076   .2820     5.6712871
 RHO         -.4137785986   .64213081E-01   -6.444   .0000
```

Regression results are almost unchanged. Autocorrelation of transformed residuals is -.17, less than -.41 in original model.

6. Data for fitting an improved Phillips curve model can be obtained from many sources, including the Bureau of Economic Analysis's (BEA) own website, Economagic.com, and so on. Obtain the necessary data and expand the model of example 12.3. Does adding additional explanatory variables to the model reduce the extreme pattern of the OLS residuals that appears in Figure 12.3?

We added a dummy variable for the period after the 1973 oil shock. The new variable did not seem to improve the model much, and the pattern of the residuals was unchanged.

```
--> crea;newecon=dmy(1974.1,2000.4)$
--> regr;lhs=dp;rhs=one,unemp,newecon;plot$
+-----------------------------------------------------------------------+
| Ordinary    least squares regression    Weighting variable = none     |
| Dep. var. = DP       Mean=  -.1926996283E-01, S.D.=   2.818214558     |
| Model size: Observations =     202, Parameters =   3, Deg.Fr.=    199 |
| Residuals:  Sum of squares= 1586.260338    , Std.Dev.=        2.82332 |
| Fit:        R-squared=  .006357, Adjusted R-squared =         -.00363 |
| Model test: F[  2,    199] =     .64,    Prob value =          .53017 |
| Diagnostic: Log-L =   -494.7731, Restricted(b=0) Log-L =    -495.4173 |
|             LogAmemiyaPrCrt.=    2.091, Akaike Info. Crt.=      4.928 |
| Autocorrel: Durbin-Watson Statistic =   2.83473,   Rho =      -.41737 |
+-----------------------------------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |t-ratio |P[|T|>t] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant      .5507626279       .74399306      .740   .4600
 UNEMP    -.9835166981E-01       .12621412     -.779   .4368     5.6712871
 NEWECON     -2.474910396       2.8382661     -.872   .3843  .49504950E-02
```



Chapter 13

Models for Panel Data

1. The following is a panel of data on investment (y) and profit (x) for n=3 firms over T=10 periods.

	i=1 y	i=1 x	i=2 y	i=2 x	i=3 y	i=3 x
t= 1	13.32	12.85	20.30	22.93	8.85	8.65
t= 2	26.30	25.69	17.47	17.96	19.60	16.55
t= 3	2.62	5.48	9.31	9.16	3.87	1.47
t= 4	14.94	13.79	18.01	18.73	24.19	24.91
t= 5	15.80	15.41	7.63	11.31	3.99	5.01
t= 6	12.20	12.59	19.84	21.15	5.73	8.34
t= 7	14.93	16.64	13.76	16.13	26.68	22.70
t= 8	29.82	26.45	10.00	11.61	11.49	8.36
t= 9	20.32	19.64	19.51	19.55	18.49	15.44
t=10	4.77	5.43	18.32	17.06	20.84	17.87

(a) Pool the data and compute the least squares regression coefficients of the model $y_{it} = \alpha + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it}$.
(b) Estimate the fixed effects model of (13-2), then test the hypothesis that the constant term is the same for all three firms.
(c) Estimate the random effects model of (13-18), then carry out the Lagrange multiplier test of the hypothesis that the classical model without the common effect applies.
(d) Carry out Hausman's specification test for the random versus the fixed model.

The pooled least squares estimator is

$$\hat{y} = \underset{(.95595)}{-.747476} + \underset{(.058656)}{1.058959}x, \quad \mathbf{e}'\mathbf{e} = 120.6687$$

The fixed effects regression can be computed just by including the three dummy variables since the sample sizes are quite small. The results are

$$\hat{y} = -1.4684i_1 - 2.8362i_2 + .12166i_3 + \underset{(.050719)}{1.102192}x \qquad \mathbf{e}'\mathbf{e} = 79.183.$$

The F statistic for testing the hypothesis that the constant terms are all the same is

$F[26,2] = [(120.6687 - 79.183)/2]/[79.183/26] = 6.811.$

The critical value from the F table is 19.458, so the hypothesis is not rejected.

In order to estimate the random effects model, we need some additional parameter estimates. The group means are

	$\bar{y}$	$\bar{x}$
Group 1	15.502	14.962
Group 2	15.415	16.559
Group 3	14.373	12.930

In the group means regression using these three observations, we obtain

$$\bar{y}_{i.} = 10.665 + .29909\bar{x}_{i.} \text{ with } \mathbf{e}_{**}'\mathbf{e}_{**} = .19747.$$

There is only one degree of freedom, so this is the candidate for estimation of $\sigma_\varepsilon^2/T + \sigma_u^2$. In the least squares dummy variable (fixed effects) regression, we have an estimate of σ_ε^2 of $79.183/26 = 3.045$. Therefore, our estimate of σ_u^2 is $\hat{\sigma}_u^2 = .19747/1 - 3.045/10 = -.6703$. Obviously, this won't do. Before abandoning the random effects model, we consider an alternative consistent estimator of the constant and slope, the pooled ordinary least squares estimator. Using the group means above, we find

$$\Sigma_{i=1}^{3}[\bar{y}_{i.} - (-.747476) - 1.058959\bar{x}_{i.}]^2 = 3.9273.$$

One ought to proceed with some caution at this point, but it is difficult to place much faith in the group means regression with but a single degree of freedom, so this is probably a preferable estimator in any event. (The true model underlying these data -- using a random number generator -- has a slope, β of 1.000 and a true constant of zero. Of course, this would not be known to the analyst in a real world situation.) Continuing, we now use $\hat{\sigma}_u^2 = 3.9273 - 3.045/10 = 3.6227$ as the estimator. (The true value of $\rho = \sigma_u^2/(\sigma_u^2+\sigma_\varepsilon^2)$ is .5.) This leads to $\theta = 1 - [3.0455^{1/2}/(10(3.6227) + 3.045)^{1/2}] = .721524$. Finally, the FGLS estimator computed according to (16-48) is $\hat{y} = -1.3415(.786) + 1.0987\ (.028998)x$.

For the LM test, we return to the pooled ordinary least squares regression. The necessary quantities are $\mathbf{e}'\mathbf{e} = 120.6687$, $\Sigma_t e_{1t} = -.55314$, $\Sigma_t e_{2t} = -13.72824$, $\Sigma_t e_{3t} = 14.28138$. Therefore,

$$LM = \{[3(10)]/[2(9)]\}\{[(-.55314)^2 + (13.72824)^2 + (14.28138)^2]/120.687 - 1\}^2 = 8.4683$$

The statistic has one degree of freedom. The critical value from the chi-squared distribution is 3.84, so the hypothesis of no random effect is rejected. Finally, for the Hausman test, we compare the FGLS and least squares dummy variable estimators. The statistic is $\chi^2 = [(1.0987 - 1.058959)^2]/[(.058656)^2 - (.05060)^2] = 1.794373$. This is relatively small and argues (once again) in favor of the random effects model.

2. Suppose that the model of (13-2) is formulated with an overall constant term and n-1 dummy variables (dropping, say, the last one). Investigate the effect that this has on the set of dummy variable coefficients and on the least squares estimates of the slopes.

There is no effect on the coefficients of the other variables. For the dummy variable coefficients, with the full set of n dummy variables, each coefficient is

$\bar{y}_i^*$ = mean residual for the ith group in the regression of y on the xs omitting the dummy variables.

(We use the partitioned regression results of Chapter 6.) If an overall constant term and n-1 dummy variables (say the last n-1) are used, instead, the coefficient on the ith dummy variable is simply $\bar{y}_i^* - \bar{y}_1^*$ while the constant term is still $\bar{y}_1^*$ For a full proof of these results, see the solution to Exercise 5 of Chapter 8 earlier in this book.

3. Use the data in Section 13.9.7 (these are the Grunfeld data) to fit the random and fixed effects models. There are five firms and 20 years of data for each. Use the F, LM, and/or Hausman statistics to determine which model, the fixed or random effects model, is preferable for these data.

The regression model is $I_{it} = \beta_{1i} + \beta_2 F_{it} + \beta_3 C_{it} + \beta_{it}$. We first fit the model by pooled OLS, ignoring the specific firm effect.

```
+--------------------------------------------------------------------+
| OLS Without Group Dummy Variables                                  |
| Ordinary    least squares regression                               |
| Dependent variable is I         Mean =  248.95700, S.D. =    267.8654 |
| Model size: Observations =     100, Parameters =   3, Deg.Fr. =    97 |
| Residuals:  Sum of squares=   0.157088E+07 Std.Dev. =       127.25831 |
| Fit:        R-squared = 0.77886, Adjusted R-squared =         0.77430 |
| Model test: F[  2,     97] =  170.81,    Prob value =         0.00000 |
| Diagnostic: Log-L =   -624.9928, Restricted(ß=0) Log-L =    -700.4398 |
| Panel Data Analysis of I          [ONE way]                        |
|           Unconditional ANOVA (No regressors)                      |
| Source      Variation        Deg. Free.     Mean Square            |
| Between      0.487817E+07           4.       0.121954E+07          |
| Residual     0.222527E+07          95.         23423.9             |
| Total        0.710344E+07          99.         71751.9             |
+--------------------------------------------------------------------+
  Variable  Coefficient   Standard Error  t-ratio  P[*T*$t]   Mean of X
  ---------------------------------------------------------------------
  F           0.10509        0.11378E-01    9.236   0.00000   1922.
  C           0.30537        0.43508E-01    7.019   0.00000   311.1
  Constant   -48.030          21.480       -2.236   0.02764
```

The least squares regression with firm specific effects is

```
+----------------------------------------------------------------------+
| Least Squares with Group Dummy Variables                             |
| Ordinary    least squares regression    Weighting variable = ONE     |
| Dependent variable is I         Mean =  248.95700, S.D. =    267.8654 |
| Model size: Observations =     100, Parameters =   7, Deg.Fr. =    93 |
| Residuals:  Sum of squares=    444288.     Std.Dev. =        69.11797 |
| Fit:        R-squared = 0.93745, Adjusted R-squared =         0.93342 |
| Model test: F[  6,     93] =  232.32,    Prob value =         0.00000 |
| Diagnostic: Log-L =   -561.8468, Restricted(ß=0) Log-L =    -700.4398 |
+----------------------------------------------------------------------+
  Variable  Coefficient   Standard Error  t-ratio  P[*T*$t]   Mean of X
  ---------------------------------------------------------------------
  F          0.10598         0.15891E-01    6.669   0.00000   1922.
  C          0.34666         0.24161E-01   14.348   0.00000   311.1
```

To estimate the variance components for the random effects model, we also computed the group means regression. The sum of squared residuals from the LSDV estimator is 444,288. The sum of squares from the group means regression is 22382.1. The estimate of σ_ε^2 is 444,288/93 = 4777.29. The estimate of σ_u^2 is 22,382.1/2 - (1/20)4777.29 = 10,952.2. The model is then reestimated by FGLS using these estimates:

```
            +--------------------------------------------------+
            | Random Effects Model: v(i,t) = e(i,t) + u(i)     |
            | Estimates:  Var[e]              =  0.477729E+04  |
            |             Var[u]              =  0.109522E+05  |
            |             Corr[v(i,t),v(i,s)] =  0.696284      |
            | Lagrange Multiplier Test vs. Model (3) =  453.82 |
            | ( 1 df, prob value = 0.000000)                   |
            | Fixed vs. Random Effects (Hausman)     =    3.14 |
            | ( 2 df, prob value = 0.208081)                   |
            +--------------------------------------------------+

  Variable  Coefficient   Standard Error  z=b/s.e. P[*Z*$z]   Mean of X
  ---------------------------------------------------------------------
  F          0.10489         0.14711E-01    7.130   0.00000   1922.
  C          0.34602         0.24112E-01   14.350   0.00000   311.1
  Constant   -60.291          54.167       -1.113   0.26568
```

The *F* and LM statistics are not useful for comparing the fixed and random effects models. The Hausman statistic can be used. The value appears above. Since the Hausman statistic is small (only 3.14 with two degrees of freedom), we conclude that the GLS estimator is consistent. The statistic would be large if the two estimates were significantly different. Since they are not, we conclude that the evidence favors the random effects model.

4. Derive the log-likelihood function for the model in (13-18) assuming that „it and u_i are normally distributed. [**Hints:** Write the log-likelihood function as $\ln L = \Sigma_{i=1}^{n} \ln L_i$ where $\ln L_i$ is the log-likelihood function for the *T* observations in group *i*. These *T* observations are joint normally distributed with covariance matrix given in (14-20).] The log-likelihood is the sum of the logs of the joint normal densities of the *n* sets of *T* observations,

$$\varepsilon_{it} + u_i = y_{it} - \alpha - \boldsymbol{\beta}'\mathbf{x}_{it}.$$

This will involve the inverse and determinant of $\boldsymbol{\Omega}$. Use (2-66) to prove that

$$\boldsymbol{\Omega}^{-1} = (1/\sigma_\varepsilon^2)\{\mathbf{I} - [\sigma_u^2/(\sigma_\varepsilon^2 + T\sigma_u^2)]\mathbf{ii}'\}$$

To find the determinant, use the product of the characteristic roots. Note first that

$$|\sigma\mathbf{I} + \sigma\mathbf{ii}'| = (\sigma_\varepsilon^2)|\mathbf{I} + (\sigma_u^2/\sigma_\varepsilon^2)\mathbf{ii}'|.$$

The roots are determined by $[\mathbf{I} + (\sigma_u^2/\sigma_\varepsilon^2)\mathbf{ii}']\mathbf{c} = \lambda\mathbf{c}$ or $(\sigma_u^2/\sigma_\varepsilon^2)\mathbf{ii}'\mathbf{c} = (\lambda\text{-}1)\mathbf{c}$. Any vector whose elements sum to zero is a solution. There are *T*-1 such independent vectors, so *T*-1 characteristic roots are $(\lambda\text{-}1)=0$ or $\lambda=1$. Premultiply the expression by $\mathbf{i}'$ to obtain the remaining characteristic root. (Remember to add 1 to the result.) Now, collect terms to obtain the log-likelihood. The *i*th group of *T* observations,

$\mathbf{w}_i = \mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} = \begin{bmatrix} \varepsilon_{i1}+u_i \\ \varepsilon_{i2}+u_i \\ \cdots \\ \varepsilon_{iT}+u_i \end{bmatrix} = \begin{bmatrix} y_{i1}-\boldsymbol{\beta}'\mathbf{x}_{i1} \\ y_{i2}-\boldsymbol{\beta}'\mathbf{x}_{i2} \\ \\ y_{iT}-\boldsymbol{\beta}'\mathbf{x}_{iT} \end{bmatrix}$, is normally distributed with mean vector $\mathbf{0}$ and the covariance matrix given in (14-20). We have included the constant term in $\mathbf{X}_i$. The joint density of these T observations is, therefore, $L_i = f(\mathbf{w}_i) = (2\pi)^{-T/2}|\boldsymbol{\Omega}|^{-1/2}exp[(-1/2)\mathbf{w}_i'\boldsymbol{\Omega}^{-1}\mathbf{w}_i]$. The log of the joint density is

$$\ln L_i = -(T/2)\ln(2\pi) - (1/2)\ln|\boldsymbol{\Omega}| - (1/2)\mathbf{w}_i'\boldsymbol{\Omega}^{-1}\mathbf{w}_i$$

and, finally, for the full sample, $\ln L = \Sigma_i \ln L_i$. Consider the log-determinant first. As suggested above, we write $\boldsymbol{\Omega} = \sigma_\varepsilon^2[\mathbf{I} + (\sigma_u^2/\sigma_\varepsilon^2)\mathbf{ii}']$. Then, $|\boldsymbol{\Omega}| = (\sigma_\varepsilon^2)^T|\mathbf{I} + (\sigma_u^2/\sigma_\varepsilon^2)\mathbf{ii}'|$ or $\ln|\boldsymbol{\Omega}| = T\ln\sigma_\varepsilon^2 + \ln|\mathbf{I} + (\sigma_u^2/\sigma_\varepsilon^2)\mathbf{ii}'|$. The determinant of a matrix equals the product of its characteristic roots, so the log determinant equals the sum of the logs of the roots. The characteristic roots of the matrix above remain to be determined. As shown in the exercise, T-1 of the T roots equal 1. Therefore, the logs of these roots are zero, so the log-determinant equals the log of the remaining root. It remains only to find the other characteristic root. Premultiply the result

$(\sigma_u^2/\sigma_\varepsilon^2)\mathbf{ii}'\mathbf{c} = (\lambda-1)\mathbf{c}$ by $\mathbf{i}'$ to obtain $(\sigma_u^2/\sigma_\varepsilon^2)\mathbf{i}'\mathbf{ii}'\mathbf{c} = (\lambda-1)\mathbf{i}'\mathbf{c}$.

Now, $\mathbf{i}'\mathbf{i} = T$. Divide both sides of the equation by $\mathbf{i}'\mathbf{c}$ -- remember, we now seek the characteristic root which corresponds to the characteristic vector whose elements do not sum to zero -- and obtain

$$T(\sigma_u^2/\sigma_\varepsilon^2) = \lambda - 1 \text{ or } \lambda = 1 + T(\sigma_u^2/\sigma_\varepsilon^2).$$

Therefore, $\ln|\boldsymbol{\Omega}| = T\ln\sigma_\varepsilon^2 + \ln[1 + T(\sigma_u^2/\sigma_\varepsilon^2)]$.

By writing $1 + T(\sigma_u^2/\sigma_\varepsilon^2) = (1/\sigma_\varepsilon^2)[\sigma_\varepsilon^2 + T\sigma_u^2]$

we obtain $\ln|\boldsymbol{\Omega}| = (T-1)\ln\sigma_\varepsilon^2 + \ln[\sigma_\varepsilon^2 + T\sigma_u^2]$

We now turn to the exponential term. The inverse matrix is given in the exercise, so it remains only to multiply it out. Thus, $\mathbf{w}_i'\boldsymbol{\Omega}^{-1}\mathbf{w}_i = \mathbf{w}_i'\mathbf{w}_i/\sigma_\varepsilon^2 - (\mathbf{w}_i'\mathbf{i})^2/[\sigma_\varepsilon^2 + T\sigma_u^2]$

Since $\mathbf{w}_i = \mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}$ $\quad \mathbf{w}_i'\boldsymbol{\Omega}^{-1}\mathbf{w}_i = (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})/\sigma_\varepsilon^2 - [\mathbf{i}'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})]^2/[\sigma_\varepsilon^2 + T\sigma_u^2]$.

The first term is the usual sum of squared deviations. The numerator in the second can be written as

$$[\mathbf{i}'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})]^2 = [T(\bar{y}_{i.} - \boldsymbol{\beta}'\bar{\mathbf{x}}_{i.})]^2.$$

Collecting terms, $\ln L_i = -(T/2)\ln(2\pi) - [(T-1)/2]\ln\sigma_\varepsilon^2 - \frac{1}{2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})/\sigma_\varepsilon^2 - \frac{1}{2}[T(\bar{y}_{i.} - \boldsymbol{\beta}'\bar{\mathbf{x}}_{i.})]^2/[\sigma_\varepsilon^2 + T\sigma_u^2]$. Finally, the log-likelihood for the full sample is the sum of these expressions over the i=1 to n groups.

5. *Unbalanced design for random effects.* Suppose that the random effects model of Section 13.4 is to be estimated with a panel in which the groups have different numbers of observations. Let T_i be the number of observations in group i.

(a) Show that the pooled least squares estimator in (13-11) is unbiased and consistent in spite of this complication.

(b) Show that the estimator in (13-29) based on the pooled least squares estimator of $\boldsymbol{\beta}$ (or, for that matter, *any* consistent estimator of $\boldsymbol{\beta}$) is a consistent estimator of σ_ε^2.

The model in (13-11) is a generalized regression model. As we saw in Chapter 10, OLS is consistent in the GR model. The unequal group sizes here does not have any effect on the result. The residual using any estimators of α and $\boldsymbol{\beta}$ is $e_{it} = y_{ti} - \hat{\alpha} - \hat{\boldsymbol{\beta}}'\mathbf{x}_{it}$ and $\bar{e}_{i.} = \bar{y}_{i.} - \hat{\alpha} - \hat{\boldsymbol{\beta}}'\bar{\mathbf{x}}_{i.}$. Thus the estimator in (13-29) is $[1/(nT-n-K)]\Sigma_i\Sigma_t(e_{it} - \bar{e}_{i.})^2 = [1/(nT-n-K)]\Sigma_i\Sigma_t[(y_{it} - \bar{y}_{i.}) - \hat{\boldsymbol{\beta}}'(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i.})]^2$. The probability limit is the same as the probability limit of the statistic which results when $\hat{\boldsymbol{\beta}}$ is replaced with its probability limit. If $\hat{\boldsymbol{\beta}}$ is a consistent estimator of $\boldsymbol{\beta}$, then the estimator converges to plim$[1/(nT-n-K)]\Sigma_i\Sigma_t[(y_{it} - \bar{y}_{i.}) - \boldsymbol{\beta}'(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i.})]^2$. But, $(y_{it} - \bar{y}_{i.}) - \boldsymbol{\beta}'(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i.}) = \varepsilon_{it} - \bar{\varepsilon}_{i.}$ So, our estimator has the same limiting behavior as $\hat{\sigma}_\varepsilon^2 = [1/(nT-n-K)]\Sigma_i\Sigma_t(\varepsilon_{it} - \bar{\varepsilon}_{i.})^2$. Write this as $\hat{\sigma}_\varepsilon^2 = (1/n)\Sigma_{i=1}^n[\Sigma_t(\varepsilon_{it} - \bar{e}_{i.})^2]/[(T-1) - K/n]$. The expected value of sum of squared deviations in the brackets is $(T-1)\sigma_\varepsilon^2$. Each term in the outer sum has the same expectation, so the exact expectation is $1/n$ times n times this expectation, or

$$E[\hat{\sigma}_\varepsilon^2] = [(T\text{-}1)\,\sigma_\varepsilon^2]/[(T\text{-}1) - K/n]$$

This obviously converges to σ_ε^2 as $n\to\infty$. The exact variance of the estimator depends upon what we assume about the fourth moment of ε_{it}. If we assume only that the fourth moment of ε_{it} is finite, then the variance of each term in the inner sum is of the form

$$[T/(T - 1 - K/n)][\phi_1/T + \phi_2/T^2 + \phi_3/T^3] = \phi.$$

If ϕ is finite, then the variance of the entire expression is ϕ/n which converges to 0. This completes the proof. To summarize the argument, we have shown that the limiting behavior of the statistic in (13-27) based on any consistent estimator of $\boldsymbol{\beta}$ is the same as that of a statistic which converges in mean square to σ_ε^2 if the fourth moment of ε is finite.

6. What are the probability limits of $(1/n)$LM, where LM is defined in (13-31) under the null hypotheses that $\sigma_u^2 = 0$ and under the alternative that $\sigma_u^2 \neq 0$?

To find plim$(1/n)$LM = plim $[T/(2(T\text{-}1))]\{[\Sigma_i(\Sigma_t e_{it})^2]/[\Sigma_i\Sigma_t e_{it}^2] - 1\}^2$ we can concentrate on the sums inside the curled brackets. First, $\Sigma_i(\Sigma_t e_{it})^2 = nT^2\{(1/n)\Sigma_i[(1/T)\Sigma_t e_{it}]^2\}$ and $\Sigma_i\Sigma_t e_{it}^2 = nT(1/(nT))\Sigma_i\Sigma_t e_{it}^2$. The ratio equals $[\Sigma_i(\Sigma_t e_{it})^2]/[\Sigma_i\Sigma_t e_{it}^2] = T\{(1/n)\Sigma_i[(1/T)\Sigma_t e_{it}]^2\}/\{(1/(nT))\Sigma_i\Sigma_t e_{it}^2\}$. Using the argument used in Exercise 8 to establish consistency of the variance estimator, the limiting behavior of this statistic is the same as that which is computed using the true disturbances since the OLS coefficient estimator is consistent. Using the true disturbances, the numerator may be written $(1/n)\Sigma_i[(1/T)\Sigma_t\varepsilon_{it}]^2 = (1/n)\Sigma_i\bar{\varepsilon}_{i.}^2$. Since $E[\bar{\varepsilon}_{i.}] = 0$, plim$(1/n)\Sigma_i\bar{\varepsilon}_{i.}^2 = \text{Var}[\bar{\varepsilon}_{i.}] = \sigma_\varepsilon^2 T + \sigma_u^2$The denominator is simply the usual variance estimator, so plim$(1/(nT))\Sigma_i\Sigma_t\varepsilon_{it}^2 = \text{Var}[\varepsilon_{it}] = \sigma_\varepsilon^2 + \sigma_u^2$Therefore, inserting these results in the expression for LM, we find that plim $(1/n)$LM $= [T/(2(T\text{-}1))]\{[T(\sigma_\varepsilon^2 T + \sigma_u^2)]/[\sigma_\varepsilon^2 + \sigma_u^2] - 1\}^2$. Under the null hypothesis that $\sigma_u^2 = 0$, this equals 0. By expanding the inner term then collecting terms, we find that under the alternative hypothesis that σ_u^2 is not equal to 0, plim $(1/n)$LM $= [T(T\text{-}1)/2][\sigma_u^2/(\sigma_\varepsilon^2+\sigma_u^2)]^2$. Within group i, $\text{Corr}^2[\varepsilon_{it},\varepsilon_{is}] = \rho^2 = \sigma_u^2/(\sigma_u^2+\sigma_\varepsilon^2)$ so plim $(1/n)$LM $= [T(T\text{-}1)/2](\rho^2)^2$. It is worth noting what is obtained if we do not divide the LM statistic by n at the outset. Under the null hypothesis, the limiting distribution of LM is chi-squared with one degree of freedom. This is a random variable with mean 1 and variance 2, so the statistic, itself, does not converge to a constant; it converges to a random variable. Under the alternative, the LM statistic has mean and variance of order n (as we see above) and hence, explodes. It is this latter attribute which makes the test a consistent one. As the sample size increases, the power of the LM test must go to 1.

7. *A two way fixed effects model:* Suppose the fixed effects model is modified to include a time specific dummy variable as well as an individual specific variable. Then, $y_{it} = \alpha_i + \gamma_t + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it}$. At every observation, the individual- and time-specific dummy variables sum to one, so there are some redundant coefficients. The discussion in Section 13.3.3 shows one way to remove the redundancy. Another useful way to do this is to include an overall constant and to drop one of the time specific *and* one of the time-dummy variables. The model is, thus, $y_{it} = \delta + (\alpha_i - \alpha_1) + (\gamma_t - \gamma_1) + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it}$. (Note that the respective time or individual specific variable is zero when t or i equals one.) Ordinary least squares estimates of $\boldsymbol{\beta}$ can be obtained by regression of $y_{it} - \bar{y}_{i.} - \bar{y}_{.t} + \bar{\bar{y}}$ on $\mathbf{x}_{it} - \bar{\mathbf{x}}_{i.} - \bar{\mathbf{x}}_{.t} + \bar{\bar{\mathbf{x}}}$ Then, $(\alpha_i\text{-}\alpha_1)$ and $(\gamma_t\text{-}\gamma_1)$ are estimated using the expressions in (13-17) while $d = \bar{\bar{y}} - \mathbf{b}'\bar{\bar{\mathbf{x}}}$.

Using the following data, estimate the full set of coefficients for the least squares dummy variable model:

	t=1	t=2	t=3	t=4	t=5	t=6	t=7	t=8	t=9	t=10
					i=1					
y	21.7	10.9	33.5	22.0	17.6	16.1	19.0	18.1	14.9	23.2
x_1	26.4	17.3	23.8	17.6	26.2	21.1	17.5	22.9	22.9	14.9
x_2	5.79	2.60	8.36	5.50	5.26	1.03	3.11	4.87	3.79	7.24
					i=2					
y	21.8	21.0	33.8	18.0	12.2	30.0	21.7	24.9	21.9	23.6
x_1	19.6	22.8	27.8	14.0	11.4	16.0	28.8	16.8	11.8	18.6
x_2	3.36	1.59	6.19	3.75	1.59	9.87	1.31	5.42	6.32	5.35
					i=3					
y	25.2	41.9	31.3	27.8	13.2	27.9	33.3	20.5	16.7	20.7
x_1	13.4	29.7	21.6	25.1	14.1	24.1	10.5	22.1	17.0	20.5
x_2	9.57	9.62	6.61	7.24	1.64	5.99	9.00	1.75	1.74	1.82

i=4

y	15.3	25.9	21.9	15.5	16.7	26.1	34.8	22.6	29.0	37.1
x_1	14.2	18.0	29.9	14.1	18.4	20.1	27.6	27.4	28.5	28.6
x_2	4.09	9.56	2.18	5.43	6.33	8.27	9.16	5.24	7.92	9.63

Test the hypotheses that (1) the "period" effects are all zero, (2) the "group" effects are all zero, and (3) both period and group effects are zero. Use an F test in each case.

The ordinary least squares regression results are

R^2 = .92803, $\mathbf{e}'\mathbf{e}$ = 146.761, 40 observations

Variable	**Coefficient**	**Standard Error**
X_1	.446845	.07887
X_2	1.83915	.1534
Constant	3.60568	2.555
Period 1	-3.57906	1.723
Period 2	-1.49784	1.716
Period 3	2.00677	1.760
Period 4	-3.03206	1.731
Period 5	-5.58937	1.768
Period 6	-1.49474	1.714
Period 7	1.52021	1.714
Period 8	-2.25414	1.737
Period 9	-3.29360	1.722
Group 1	-.339998	1.135
Group 2	4.39271	1.183
Group 3	5.00207	1.125

Estimated covariance matrix for the slopes:

	β_1	β_2
β_1	.0062209	
β_2	.00030947	.023523

For testing the hypotheses that the sets of dummy variable coefficients are zero, we will require the sums of squared residuals from the restrictions. These are

Regression	**Sum of squares**
All variables included	146.761
Period variables omitted	318.503
Group variables omitted	369.356
Period and group variables omitted	585.622

The F statistics are therefore,

(1) $F[9,25]$ = [(318.503 - 146.761)/9]/[146.761/25] = 3.251

(2) $F[3,25]$ = [(369.356 - 146.761)/3]/[146.761/25] = 12.639

(3) $F[12,25]$ = [(585.622 - 146.761)/12]/[146.761/25] = 6.23

The critical values for the three distributions are 2.283, 2.992, and 2.165, respectively. All sample statistics are larger than the table value, so all of the hypotheses are rejected. □

8. *Two way random effects model:* We modify the random effects model by the addition of a time specific disturbance. Thus, $y_{it} = \alpha + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it} + u_i + v_t$, where $E[\varepsilon_{it}] = E[u_i] = E[v_t] = 0$, $E[\varepsilon_{it}u_t] = E[\varepsilon_{it}v_s] = E[u_i v_t] = 0$, for all i,j,t,s, $\text{Var}[\varepsilon_{it}] = \sigma_\varepsilon^2 \text{Cov}[\varepsilon_{it},\varepsilon_{js}] = 0$ for all t,j,s, $\text{Var}[u_i] = \sigma_u^2 \text{Cov}[u_i,u_j] = 0$ for all i,j $\text{Var}[v_t] = \sigma_v^2$, $\text{Cov}[v_t,v_s] = 0$ for all t,s. Write out the full covariance matrix for a data set with n=2 and T=2.

The covariance matrix would be

$$
\begin{array}{ccccc}
 & i=1,t=1 & i=1,t=2 & i=2,t=1 & i=2,t=2 \\
i=1,t=1 & \sigma_\varepsilon^2+\sigma_u^2+\sigma_v^2 & \sigma_u^2 & \sigma_v^2 & 0 \\
i=1,t=2 & \sigma_u^2 & \sigma_\varepsilon^2+\sigma_u^2+\sigma_v^2 & 0 & \sigma_v^2 \\
i=2,t=1 & \sigma_v^2 & 0 & \sigma_\varepsilon^2+\sigma_u^2+\sigma_v^2 & \sigma_u^2 \\
i=2,t=2 & 0 & \sigma_v^2 & \sigma_u^2 & \sigma_\varepsilon^2+\sigma_u^2+\sigma_v^2
\end{array}
$$

9. The model $\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix} = \begin{bmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{bmatrix}\beta + \begin{bmatrix}\varepsilon_1\\ \varepsilon_2\end{bmatrix}$ satisfies the groupwise heteroscedastic regression model of Section 11.7.2. All variables have zero means. The following sample second moment matrix is obtained from a sample of 20 observations:

	y_1	y_2	x_1	x_2
y_1	20	6	4	3
y_2	6	10	3	6
x_1	4	3	5	2
x_2	3	6	2	10

(a) Compute the two separate OLS estimates of β, their sampling variances, the estimates of σ_1^2 and σ_2^2, and the R^2s in the two regressions.
(b) Carry out the Lagrange Multiplier test of the hypothesis that $\sigma_1^2 = \sigma_2^2$.
(c) Compute the two step FGLS estimate of β and an estimate of its sampling variance. Test the hypothesis that β equals one.
(d) Carry out the Wald test of equal disturbance variances.
(e) Compute the maximum likelihood estimates of β, σ_1^2, and σ_2^2 by iterating the FGLS estimates to convergence.
(f) Carry out a likelihood ratio test of equal disturbance variances.
(g) Compute the two step FGLS estimate of β assuming that the model in (14-7) applies. [That is, allow for cross sectional correlation.] Compare your results to those of part (c).

The two separate regressions are as follows:

	Sample 1	**Sample 2**
$b = \mathbf{x}'\mathbf{y}/\mathbf{x}'\mathbf{x}$	4/5 = .8	6/10 = .6
$\mathbf{e}'\mathbf{e} = \mathbf{y}'\mathbf{y} - b\mathbf{x}'\mathbf{y}$	20 - 4(4/5) = 84/5	10 - 6(6/10) = 64/10
$R^2 = 1 - \mathbf{e}'\mathbf{e}/\mathbf{y}'\mathbf{y}$	1 - (84/5)/20 = .16	1 - (64/10)/10 = .36
$s^2 = \mathbf{e}'\mathbf{e}/(n\text{-}1)$	(84/5)/19 = .88421	(64/10)/19 = .33684
Est.Var$[b] = s^2/\mathbf{x}'\mathbf{x}$	.88421/5 = .17684	.33684/10 = .033684

To carry out a Lagrange multiplier test of the hypothesis of equal variances, we require the separate and common variance estimators based on the restricted slope estimator. This, in turn, is the pooled least squares estimator. For the combined sample, we obtain

$$b = [\mathbf{x}_1'\mathbf{y}_1 + \mathbf{x}_2'\mathbf{y}_2]/[\mathbf{x}_1'\mathbf{x}_1 + \mathbf{x}_2'\mathbf{x}_2] = (4 + 6) / (5 + 10) = 2/3.$$

Then, the variance estimators are based on this estimate. For the hypothesized common variance,

$$\mathbf{e}'\mathbf{e} = (\mathbf{y}_1'\mathbf{y}_1 + \mathbf{y}_2'\mathbf{y}_2) - b(\mathbf{x}_1'\mathbf{y}_1 + \mathbf{x}_2'\mathbf{y}_2) = (20 + 10) - (2/3)(4 + 6) = 70/3,$$

so the estimate of the common variance is $\mathbf{e}'\mathbf{e}/40 = (70/3)/40 = .58333$. Note that the divisor is 40, not 39, because we are comptuting maximum likelihood estimators. The individual estimators are

$$\mathbf{e}_1'\mathbf{e}_1/20 = (\mathbf{y}_1'\mathbf{y}_1 - 2b(\mathbf{x}_1'\mathbf{y}_1) + b^2(\mathbf{x}_1'\mathbf{x}_1))/20 = (20 - 2(2/3)4 + (2/3)^2 5)/20 = .84444$$

and $$\mathbf{e}_2'\mathbf{e}_2/20 = (\mathbf{y}_2'\mathbf{y}_2 - 2b(\mathbf{x}_2'\mathbf{y}_2) + b^2(\mathbf{x}_2'\mathbf{x}_2))/20 = (10 - 2(2/3)6 + (2/3)^2 10)/20 = .32222.$$

The LM statistic is given in Example 16.3,

$$LM = (T/2)[(s_1^2/s^2 - 1)^2 + (s_2^2/s^2 - 1)^2] = 10[(.84444/.58333 - 1)^2 + (.32222/.58333 - 1)^2] = 4.007.$$

This has one degree of freedom for the single restriction. The critical value from the chi-squared table is 3.84, so we would reject the hypothesis.

In order to compute a two step GLS estimate, we can use either the original variance estimates based on the separate least squares estimates or those obtained above in doing the LM test. Since both pairs are consistent, both FGLS estimators will have all of the desirable asymptotic properties. For our estimator, we used $\hat{\sigma}_1^2 = \mathbf{e}_j'\mathbf{e}_j/T$ from the original regressions. Thus, $\hat{\sigma}_1^2 = .84$ and $\hat{\sigma}_2^2 = .32$. The GLS estimator is

$$\hat{\boldsymbol{\beta}} = [(1/\hat{\sigma}_1^2)\mathbf{x}_1'\mathbf{y}_1 + (1/\hat{\sigma}_2^2)\mathbf{x}_2'\mathbf{y}_2]/[(1/\hat{\sigma}_1^2)\mathbf{x}_1'\mathbf{x}_1 + (1/\hat{\sigma}_2^2)\mathbf{x}_2'\mathbf{x}_2] = [4/.84 + 6/.32]/[5/.84 + 10/.32] = .632.$$

The estimated sampling variance is $1/[(1/\hat{\sigma}_1^2)\mathbf{x}_1'\mathbf{x}_1 + (1/\hat{\sigma}_2^2)\mathbf{x}_2'\mathbf{x}_2] = .02688$. This implies an asymptotic standard error of $(.02688)^2 = .16395$. To test the hypothesis that β = 1, we would refer z = (.632 - 1) / .16395 = -2.245 to a standard normal table. This is reasonably large, and at the usual significance levels, would lead to rejection of the hypothesis.

The Wald test is based on the unrestricted variance estimates. Using b = .632, the variance estimators are $\hat{\sigma}_1^2 = [\mathbf{y}_1'\mathbf{y}_1 - 2b(\mathbf{x}_1'\mathbf{y}_1) + b^2(\mathbf{x}_1'\mathbf{x}_1)]/20 = .847056$

and $\hat{\sigma}_2^2 = [\mathbf{y}_2'\mathbf{y}_2 - 2b(\mathbf{x}_2'\mathbf{y}_2) + b^2(\mathbf{x}_2'\mathbf{x}_2)]/20 = .320512$

while the pooled estimator would be $\hat{\sigma}^2 = [\mathbf{y}'\mathbf{y} - 2b(\mathbf{x}'\mathbf{y}) + b^2(\mathbf{x}'\mathbf{x})]/40 = .583784$. The statistic is given at the end of Example 16.3,

$$W = (T/2)[(\hat{\sigma}/\hat{\sigma}_1^2 - 1)^2 + (\hat{\sigma}/\hat{\sigma}_2^2 - 1)^2]$$
$$= 10[(.583784/.847056 - 1)^2 + (.583784/.320512 - 1)^2] = 7.713.$$

We reach the same conclusion as before.

To compute the maximum likelihood estimators, we begin our iterations from the two separate ordinary least squares estimates of b which produce estimates $\hat{\sigma}_1^2 = .84$ and $\hat{\sigma}_2^2 = .32$. The iterations are

Iteration	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$	$\hat{\beta}$
0	.840000	.320000	.632000
1	.847056	.320512	.631819
2	.847071	.320506	.631818
3	.847071	.320506	converged

Now, to compute the likelihood ratio statistic for a likelihood ratio test of the hypothesis of equal variances, we refer χ^2 = 40ln.58333 - 20ln.847071 - 20ln.320506 to the chi-squared table. (Under the null hypothesis, the pooled least squares estimator is maximum likelihood.) Thus, χ^2 = 4.5164, which is roughly equal to the LM statistic and leads once again to rejection of the null hypothesis.

Finally, we allow for cross sectional correlation of the disturbances. Our initial estimate of b is the pooled least squares estimator, 2/3. The estimates of the two variances are .84444 and .32222 as before while the cross sectional covariance estimate is

$$\mathbf{e}_1'\mathbf{e}_2/20 = [\mathbf{y}_1'\mathbf{y}_2 - b(\mathbf{x}_1'\mathbf{y}_2 + \mathbf{x}_2'\mathbf{y}_1) + b^2(\mathbf{x}_1'\mathbf{x}_2)]/20 = .14444.$$

Before proceeding, we note, the estimated squared correlation of the two disturbances is

$$r = .14444 / [(.84444)(.32222)]^{1/2} = .277,$$

which is not particularly large. The LM test statistic given in (16-14) is 1.533, which is well under the critical value of 3.84. Thus, we would not reject the hypothesis of zero cross section correlation. Nonetheless, we proceed. The estimator is shown in (16-6). The two step FGLS and iterated maximum likelihood estimates appear below.

Iteration	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$	$\hat{\sigma}_{12}$	$\hat{\beta}$
0	.84444	.32222	.14444	.5791338
1	.8521955	.3202177	.1597994	.5731058
2	.8528702	.3203616	.1609133	.5727069
3	.8529155	.3203725	.1609873	.5726805
4	.8529185	.3203732	.1609921	.5726788
5	.8529187	.3203732	.1609925	converged

Because the correlation is relatively low, the effect on the previous estimate is relatively minor. □

10. Suppose that in the model of Section 15.2.1, $\mathbf{X}_i$ is the same for all i. What is the generalized least squares estimator of β? How would you compute the estimator if it were necessary to estimate σ_i^2?

If all of the regressor matrices are the same, the estimator in (15-6) reduces to

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \Sigma_{i=1}^{n} \{(1/\sigma_i^2)/[\Sigma_{j=1}^{n} (1/\sigma_j^2)]\}\mathbf{X}'\mathbf{y}_i = \Sigma_{i=1}^{n} w_i\mathbf{b}_i$$

a weighted average of the ordinary least squares estimators, $\mathbf{b}_i = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_i$ with weights $w_i = (1/\sigma_i^2)/[\Sigma_{j=1}^{n}(1/\sigma_j^2)]$. If it were necessary to estimate the weights, a simple two step estimator could be based on individual variance estimators. Either of $s_i^2 = \mathbf{e}_i'\mathbf{e}_i/T$ based on separate least squares regressions (with different estimators of β) or based on residuals computed from a common pooled ordinary least squares slope estimator could be used.

11. Repeat Exercise for the model of Section 13.9.1.

The estimator is shown in (15-11). If all of the **X** matrices are the same, the estimator reduces to a weighted average of the OLS estimators again. Using (15-11) directly with a common **X**,

$$\hat{\boldsymbol\beta} = [\Sigma_i\Sigma_j\sigma^{ij}\mathbf{X}'\mathbf{X}]^{-1}[\Sigma_i\Sigma_j\sigma^{ij}\mathbf{X}'\mathbf{y}_j] = [1/\Sigma_i\Sigma_j\sigma^{ij}][\Sigma_i\Sigma_j\sigma^{ij}[\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{y}_j = [1/\Sigma_j(\Sigma_i\sigma^{ij})][\Sigma_j(\Sigma_i\sigma^{ij})\mathbf{b}_j]$$

The disturbance variances and covariances can be estimated as suggested in the previous exercise. ~

12. The following table presents a hypothetical panel of data:

	i=1		i=2		i=3	
	y	**x**	**y**	**x**	**y**	**x**
t= 1	30.27	24.31	38.71	28.35	37.03	21.16
t= 2	35.59	28.47	29.74	27.38	43.82	26.76
t= 3	17.90	23.74	11.29	12.74	37.12	22.21
t= 4	44.90	25.44	26.17	21.08	24.34	19.02
t= 5	37.58	20.80	5.85	14.02	26.15	18.64
t= 6	23.15	10.55	29.01	20.43	26.01	18.97
t= 7	30.53	18.40	30.38	28.13	29.64	21.35
t= 8	39.90	25.40	36.03	21.78	30.25	21.34
t= 9	20.44	13.57	37.90	25.65	25.41	15.86
t=10	36.85	25.60	33.90	11.66	26.04	13.28

(a) Estimate the groupwise heteroscedastic model of Section 11.7.2. Include an estimate of the asymptotic variance of the slope estimator. Use a two step procedure, basing the FGLS estimator at the second step on residuals from the pooled least squares regression.
(b) Carry out the Wald, Lagrange multiplier, and likelihood ratio tests of the hypothesis that the variances are all equal. For the likelihood ratio test, use the FGLS estimates.
(c) Carry out a Lagrange multiplier test of the hypothesis that the disturbances are uncorrelated across individuals.

The various least squares estimators of the parameters are

	Sample 1	**Sample 2**	**Sample 3**	**Pooled**
a	11.6644	5.42213	1.41116	8.06392
	(9.658)	(10.46)	(7.328)	
b	.926881	1.06410	1.46885	1.05413
	(.4328)	(.4756)	(.3590)	
e′e	452.206	673.409	125.281	
	(464.288)	(732.560)	(171.240)	(1368.088)

(Values of **e′e** in parentheses above are based on the pooled slope estimator.) The FGLS estimator and its estimated asymptotic covariance matrix are

$$\mathbf{b} = \begin{pmatrix} 7.17889 \\ 1.13792 \end{pmatrix}, \quad \text{Est.Asy.Var}[\mathbf{b}] = \begin{bmatrix} 22.8049 & -1.0629 \\ -1.0629 & 0.05197 \end{bmatrix}$$

Note that the FGLS estimator of the slope is closer to the 1.46885 of sample 3 (the highest of the three OLS estimates). This is to be expected since the third group has the smallest residual variance. The LM test statistic is based on the pooled regression,

$$LM = (10/2)\{[(464.288/10)/(1368.088/30) - 1]^2 + ...\} = 3.7901$$

To compute the Wald statistic, we require the unrestricted regression. The parameter estimates are given above. The sums of squares are 465.708, 785.399, and 145.055 for i = 1, 2, and 3, respectively. For the common estimate of σ^2, we use the total sum of squared GLS residuals, 1396.162. Then,

$$W = (10/2)\{[(1396.162/30)/(465.708/10) - 1]^2 + ...\} = 25.21.$$

The Wald statistic is far larger than the *LM* statistic. Since there are two restrictions, at significance levels of 95% or 99% with critical values of 5.99 or 9.21, the two tests lead to different conclusions. The likelihood ratio statistic based on the FGLS estimates is $\chi^2 = 30\ln(1396.162/30) - 10\ln(465.708/10) ... = 6.42$ which is between the previous two and between the 95% and 99% critical values.

The correlation matrix for the residuals from the pooled OLS regression is

$$\mathbf{R} = \begin{matrix} 1.000 & -.0704 & -.7619 \\ -.0704 & 1.000 & -.0825 \\ -.7619 & -.0825 & 1.000 \end{matrix}$$

so the LM statistic is $LM = 10[(-.0704)^2 + (-.7619)^2 + (-.0825)^2] = 5.9225$. The 95% critical value from the chi-squared distribution with 3 degrees of freedom is 7.82, so we would not reject the hypothesis of uncorrelated disturbances.

Chapter 14

Systems of Regression Equations

1. A sample of 100 observations produces the following sample data:

$$\bar{y}_1 = 1,\ \bar{y}_2 = 2,\ \mathbf{y}'_1\mathbf{y}_1 = 150 \qquad \mathbf{y}'_2\mathbf{y}_2 = 550, \qquad \mathbf{y}'_1\mathbf{y}_2 = 260$$

The underlying bivariate regression model is $y_1 = \mu + \varepsilon_1$, $y_2 = \mu + \varepsilon_2$.
(a) Compute the ordinary least squares estimate of μ and estimate the sampling variance of this estimator.
(b) Compute the FGLS estimate of μ and the sampling variance of your estimator.

The model can be written as $\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix} = \begin{bmatrix}\mathbf{i}\\ \mathbf{i}\end{bmatrix}\mu + \begin{bmatrix}\boldsymbol{\varepsilon}_1\\ \boldsymbol{\varepsilon}_2\end{bmatrix}$. Therefore, the OLS estimator is

$$m = (\mathbf{i}'\mathbf{i} + \mathbf{i}'\mathbf{i})^{-1}(\mathbf{i}'\mathbf{y}_1 + \mathbf{i}'\mathbf{y}_2) = (n\bar{y}_1 + n\bar{y}_2)/(n+n) = (\bar{y}_1 + \bar{y}_2)/2 = 1.5.$$

The sampling variance would be $\text{Var}[m] = (1/2)^2\{\text{Var}[\bar{y}_1] + \text{Var}[\bar{y}_2] + 2\text{Cov}[(\bar{y}_{1\ 1}, \bar{y}_2)]\}$.

We would estimate the parts with
$\text{Est.Var}[\bar{y}_1] = s_{11}/n = ((150 - 100(1)^2)/99)/100 = .0051$
$\text{Est.Var}[\bar{y}_2] = s_{22}/n = ((550 - 100(2)^2)/99)/100 = .0152$
$\text{Est.Cov}[\bar{y}_1, \bar{y}_2] = s_{12}/n = ((260 - 100(1)(2))/99)/100 = .0061$

Combining terms, $\text{Est.Var}[m] = .0079$.

The GLS estimator would be

$$[(\sigma^{11} + \sigma^{12})\mathbf{i}'\mathbf{y}_1 + (\sigma^{22} + \sigma^{12})\mathbf{i}'\mathbf{y}_2]/[(\sigma^{11} + \sigma^{12})\mathbf{i}'\mathbf{i} + (\sigma^{22} + \sigma^{12})\mathbf{i}'\mathbf{i}] = w\bar{y}_1 + (1-w)\bar{y}_2$$

where $w = (\sigma^{11} + \sigma^{12})/(\sigma^{11} + \sigma^{22} + 2\sigma^{12})$. Denoting $\boldsymbol{\Sigma} = \begin{bmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{12} & \sigma_{22}\end{bmatrix}$, $\boldsymbol{\Sigma}^{-1} = \dfrac{1}{\sigma_{11}\sigma_{22} - \sigma_{12}^2}\begin{bmatrix}\sigma_{22} & -\sigma_{12}\\ -\sigma_{12} & \sigma_{11}\end{bmatrix}$.

The weight simplifies a bit as the determinant appears in both the denominator and the numerator. Thus, $w = (\sigma_{22} - \sigma_{12})/(\sigma_{11} + \sigma_{22} - 2\sigma_{12})$. For our sample data, the two step estimator would be based on the variances computed above and $s_{11} = .5051$, $s_{22} = 1.5152$, $s_{12} = .6061$. Then, $w = 1.1250$. The FGLS estimate is $1.125(1) + (1 - 1.125)(2) = .875$. The sampling variance of this estimator is $w^2\text{Var}[\bar{y}_1] + (1-w)^2\text{Var}[\bar{y}_2] + 2w(1-w)\text{Cov}[\bar{y}_1, \bar{y}_2] = .0050$ as compared to .0079 for the OLS estimator.

2. Consider estimation of the following two equation model $\begin{array}{l} y_1 = \beta_1 + \varepsilon_1 \\ y_2 = x\beta_2 + \varepsilon_2 \end{array}$. A sample of 50 observations produces the following moment matrix:

	1	y_1	y_2	x
1	50			
y_1	150	500		
y_2	50	40	90	
x	100	60	50	100

(a) Write out the explicit formula for the GLS estimator of $[\beta_1,\beta_2]$. What is the asymptotic covariance matrix of the estimator?
(b) Derive the OLS estimator and its sampling variance in this model.
(c) Obtain the OLS estimates of β_1 and β_2 and estimate the sampling covariance matrix of the two estimators. Use n instead of $(n-1)$ as the divisor to compute the estimates of the disturbance variances.
(d) Compute the FGLS estimates of β_1 and β_2 and the estimated sampling covariance matrix.
(e) Test the hypothesis that $\beta_2 = 1$.

The model is $\mathbf{y} = \begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix} = \mathbf{X}\beta + \varepsilon = \begin{bmatrix}\mathbf{i} & \mathbf{0}\\ \mathbf{0} & \mathbf{x}\end{bmatrix}\begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix} + \begin{bmatrix}\varepsilon_1\\ \varepsilon_2\end{bmatrix}$, $\sigma^2\mathbf{\Omega} = \begin{bmatrix}\sigma_{11}\mathbf{I} & \sigma_{12}\mathbf{I}\\ \sigma_{12}\mathbf{I} & \sigma_{22}\mathbf{I}\end{bmatrix}$.

The generalized least squares estimator is

$$\hat{\beta} = [\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X}]^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{y} = \begin{bmatrix}\sigma^{11}\mathbf{i}'\mathbf{i} & \sigma^{12}\mathbf{i}'\mathbf{x}\\ \sigma^{12}\mathbf{i}'\mathbf{x} & \sigma^{22}\mathbf{x}'\mathbf{x}\end{bmatrix}^{-1}\begin{pmatrix}\sigma^{11}\mathbf{i}'\mathbf{y}_1 + \sigma^{12}\mathbf{i}'\mathbf{y}_2\\ \sigma^{12}\mathbf{x}'\mathbf{y}_1 + \sigma^{22}\mathbf{x}'\mathbf{y}_2\end{pmatrix}$$

$$= \left[n\begin{pmatrix}\sigma^{11} & \sigma^{12}\bar{x}\\ \sigma^{12}\bar{x} & \sigma^{22}s_{xx}\end{pmatrix}\right]^{-1}\left[n\begin{pmatrix}\sigma^{11}\bar{y}_1 + \sigma^{12}\bar{y}_2\\ \sigma^{12}s_{x1} + \sigma^{22}s_{x2}\end{pmatrix}\right]$$

where $s_{xx} = \mathbf{x}'\mathbf{x}/n$, $s_{x1} = \mathbf{x}'\mathbf{y}_1/n$, $s_{x2} = \mathbf{x}'\mathbf{y}_2/n$

and σ^{ij} = the ijth element of the 2×2 Σ^{-1}.

To obtain the explicit form, note, first, that all terms σ^{ij} are of the form $\sigma_{ji}/(\sigma_{11}\sigma_{22} - \sigma^2_{12})$ But, the denominator in these ratios will be cancelled as it appears in both the inverse matrix and in the vector. Therefore, in terms of the original parameters, (after cancelling n), we obtain

$$\hat{\beta} = \begin{bmatrix}\sigma_{22} & -\sigma_{12}\bar{x}\\ -\sigma_{12}\bar{x} & \sigma_{11}s_{xx}\end{bmatrix}^{-1}\begin{bmatrix}\sigma_{22}\bar{y}_1 - \sigma_{12}\bar{y}_2\\ -\sigma^{12}s_{x1} + \sigma_{11}s_{x2}\end{bmatrix} = \frac{1}{\sigma_{11}\sigma_{22}s_{xx} - (\sigma_{12}\bar{x})^2}\begin{bmatrix}\sigma_{11}s_{xx} & \sigma_{12}\bar{x}\\ \sigma_{12}\bar{x} & \sigma_{22}\end{bmatrix}\begin{pmatrix}\sigma_{22}\bar{y}_1 - \sigma_{12}\bar{y}_2\\ -\sigma_{12}s_{x1} + \sigma_{11}s_{x2}\end{pmatrix}.$$

The two elements are $\hat{\beta}_1 = [\sigma_{11}s_{xx}(\sigma_{22}\bar{y}_1 - \sigma_{12}\bar{y}_2) - \sigma_{12}\bar{x}(\sigma_{12}s_{x1} - \sigma_{11}s_{x2})]/[\sigma_{11}\sigma_{22}s_{xx} - (\sigma_{12}\bar{x})^2]$

$\hat{\beta}_2 = [\sigma_{12}\bar{x}(\sigma_{22}\bar{y}_1 - \sigma_{12}\bar{y}_2) - \sigma_{22}(\sigma_{12}s_{x1} - \sigma_{11}s_{x2})]/[\sigma_{11}\sigma_{22}s_{xx} - (\sigma_{12}\bar{x})^2]$

The asymptotic covariance matrix is

$$[\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X}]^{-1} = \left[n\begin{pmatrix}\sigma^{11} & \sigma^{12}\bar{x}\\ \sigma^{12}\bar{x} & \sigma^{22}s_{xx}\end{pmatrix}\right]^{-1} = \left[\frac{n}{\sigma_{11}\sigma_{22} - \sigma_{12}^2}\begin{pmatrix}\sigma_{22} & -\sigma_{12}\bar{x}\\ -\sigma_{12}\bar{x} & \sigma_{11}s_{xx}\end{pmatrix}\right]^{-1}$$

The OLS estimator is $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \begin{pmatrix}\bar{y}_1\\ \mathbf{x}'\mathbf{y}/\mathbf{x}'\mathbf{x}\end{pmatrix}$. The sampling variance is

$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix}n & 0\\ 0 & ns_{xx}\end{bmatrix}^{-1}\begin{bmatrix}\sigma_{11}n & \sigma_{12}n\bar{x}\\ \sigma_{12}n\bar{x} & \sigma_{22}ns_{xx}\end{bmatrix}\begin{bmatrix}n & 0\\ 0 & ns_{xx}\end{bmatrix}^{-1}$. The ns are carried outside the product and reduce to $(1/n)$. This leaves Var[**b**] $= \begin{bmatrix}\sigma_{11}/n & \sigma_{12}\bar{x}/(ns_{xx})\\ \sigma_{12}\bar{x}/(ns_{xx}) & \sigma_{22}/(ns_{xx})^2\end{bmatrix}$.

Using the results above, the OLS coefficients are $b_1 = \bar{y}_1 = 150/50 = 3$ and $b_2 = \mathbf{x}'\mathbf{y}_2/\mathbf{x}'\mathbf{x} = 50/100 = 1/2$. The estimators of the disturbance (co-)variances are

$s_{11} = \Sigma_i(y_{i1} - \bar{y}_1)^2/n = (500 - 50(3)2)/50 = 1$

$s_{22} = \Sigma_i(y_{i2} - b_2x_i)^2/n = (90 - (1/2)50)/50 = 1.3$

$s_{12} = \Sigma_i(y_{i1} - \bar{y}_1)(y_{i2} - b_2x_i)^2/n = [\mathbf{y}_1'\mathbf{y}_2 - n\bar{y}_1\ \bar{y}_2 - b_2\mathbf{x}'\mathbf{y}_1 + nb_2\bar{y}_1\ \bar{x}]/n$

$= (40 - 50(3)(1) - (1/2)60 + 50(1/2)(3)(2)/50 = .2$

Therefore, we estimate the asymptotic covariance matrix of the OLS estimates as

$$\text{Est.Var}[\mathbf{b}] = \begin{bmatrix}1/50 & .2(2)[50(90)]\\ .2(2)[50/90] & 1.3/90\end{bmatrix} = \begin{bmatrix}.02 & .0000888\\ .0000888 & .01444\end{bmatrix}.$$

To compute the FGLS estimates, we use our results from part a. The necessary statistics for the computation are $s_{11} = 1$, $s_{22} = 1.3$, $s_{11} = .2, s_{xx} = 100/50 = 2$, $\bar{x} = 100/50 = 2$,

$\bar{y}_1 = 150/50 = 3$, $\bar{y}_2 = 50/50 = 1$

$s_{x1} = 60/50 = 1.2$, $s_{x2} = 50/50 = 1$

Then, $\hat{\beta}_1 = \{1(2)[1.3(3) - .2(1)] - .2(2)[.2(1.2) - 1(1)]\}/\{1(1.3) - [.2(2)]^2\} = 3.157$

$$\hat{\beta}_2 = \{2(2)[1.3(3) - .2(1)] - 1.3[.2(1.2) - 1(1)]\}/\{1(1.3) - [.2(2)]^2\} = 1.011$$

The estimate of the asymptotic covariance matrix is

$$(1/50)[1(1.3) - (.2)^2]/\{1(1.3)2 - [.2(2)]^2\}\begin{bmatrix} 1(2) & .2(2) \\ .2(2) & 1.3 \end{bmatrix} = \begin{bmatrix} .020656 & .004131 \\ .004131 & .007945 \end{bmatrix}.$$

Notice that the estimated variance of the FGLS estimator of the parameter of the first equation is larger. The result for the *true* GLS estimator based on known values of the disturbance variances and covariance does not guarantee that the *estimated* variances will be smaller in a finite sample. However, the estimated variance of the second parameter is considerably smaller than that for the OLS estimate.

Finally, to test the hypothesis that $\beta_2 = 1$ we use the z-statistic (asymptotically distributed as standard normal), $z = (1.011 - 1) / (.007945)^2 = .123$. The hypothesis cannot be rejected.

3. The model $\begin{matrix} y_1 = \beta_1 x_1 + \varepsilon_1 \\ y_2 = \beta_2 x_2 + \varepsilon_2 \end{matrix}$ satisfies all of the assumptions of the classical multivariate regression model. All variables have zero means. The following sample second moment matrix is obtained from a sample of 20 observations:

	y_1	y_2	x_1	x_2
y_1	20	6	4	3
y_2	6	10	3	6
x_1	4	3	5	2
x_2	3	6	2	10

(Note: These are the data from Exercise 1 in Chapter 16.)

(a) Compute the FGLS estimates of β_1 and β_2.
(b) Test the hypothesis that $\beta_1 = \beta_2$.
(c) Compute the maximum likelihood estimates of the model parameters.
(d) Use the likelihood ratio test to test the hypothesis in part (b).

The ordinary least squares estimates of the parameters are

$$b_1 = \mathbf{x}_1'\mathbf{y}_1/\mathbf{x}_1'\mathbf{x}_1 = 4/5 = .8 \quad \text{and} \quad b_2 = \mathbf{x}_2'\mathbf{y}_2/\mathbf{x}_2'\mathbf{x}_2 = 6/10 = .6$$

Then, the variances and covariance of the disturbances are

$$s_{11} = (\mathbf{y}_1'\mathbf{y}_1 - b_1\mathbf{x}_1'\mathbf{y}_1)/n = (20 - .8(4))/20 = .84$$
$$s_{22} = (\mathbf{y}_2'\mathbf{y}_2 - b_2\mathbf{x}_2'\mathbf{y}_2)/n = (10 - .6(6))/20 = .32$$
$$s_{12} = (\mathbf{y}_1'\mathbf{y}_2 - b_2\mathbf{x}_2'\mathbf{y}_1 - b_1\mathbf{x}_1'\mathbf{y}_2 + b_1b_2\mathbf{x}_1'\mathbf{x}_2)/n = (6 - .6(3) - .8(3) + .8(.6)(2))/20 = .246$$

We will require $\mathbf{S}^{-1} = \begin{bmatrix} .84 & .246 \\ .246 & .32 \end{bmatrix}^{-1} = \begin{bmatrix} s^{11} & ^{12} \\ s^{12} & s^{11} \end{bmatrix}$. Then, the FGLS estimator is

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{bmatrix} s^{11}\mathbf{x}_1'\mathbf{x}_1 & s^{12}\mathbf{x}_1'\mathbf{x}_2 \\ s^{12}\mathbf{x}_1'\mathbf{x}_2 & s^{22}\mathbf{x}_2'\mathbf{x}_2 \end{bmatrix}^{-1} \begin{bmatrix} s^{11}\mathbf{x}_1'\mathbf{y}_1 + s^{12}\mathbf{x}_1'\mathbf{y}_2 \\ s^{12}\mathbf{x}_2'\mathbf{y}_1 + s^{22}\mathbf{x}_2'\mathbf{y}_2 \end{bmatrix}.$$

Inserting the values given in the problem produces the FGLS estimates, $\hat{\beta}_1 = .505335$, $\hat{\beta}_2 = .541741$ with estimated asymptotic covariance matrix equal to the inverse matrix shown above, $\text{Est.Var}\begin{bmatrix}\hat{\boldsymbol{\beta}}\end{bmatrix} = \begin{bmatrix} .132565 & .0077645 \\ .0077645 & .0252505 \end{bmatrix}$. To test the hypothesis, we use the t statistic, $t = (.505335 - .541741)/[.132565 + .0252505 - 2(.0077645)]^2 = -.0965$ which is quite small. We would not reject the hypothesis.

To compute the maximum likelihood estimates, we would begin with the OLS estimates of σ_{11}, σ_{22}, and σ_{12}. Then, we iterate between the following calculations

(1) Compute the 2×2 matrix, $\mathbf{S}^{-1}$

(2) Compute the 2×2 matrix $[\mathbf{X}'(\mathbf{S}^{-1}\otimes\mathbf{I})\mathbf{X}] = \begin{bmatrix} s^{11}\mathbf{x}_1'\mathbf{x}_1 & s^{12}\mathbf{x}_1'\mathbf{x}_2 \\ s^{12}\mathbf{x}_1'\mathbf{x}_2 & s^{22}\mathbf{x}_2'\mathbf{x}_2 \end{bmatrix}$

$$[\mathbf{X}'(\mathbf{S}^{-1}\otimes\mathbf{I})\mathbf{y}] = \begin{bmatrix} s^{11}\mathbf{x}_1'\mathbf{y}_1 + s^{12}\mathbf{x}_1'\mathbf{y}_2 \\ s^{12}\mathbf{x}_2'\mathbf{y}_1 + s^{22}\mathbf{x}_2'\mathbf{y}_2 \end{bmatrix}$$

(3) Compute the coefficient vector $\hat{\boldsymbol\beta} = [\mathbf{X}'(\mathbf{S}^{-1}\otimes\mathbf{I})\mathbf{X}]^{-1}[\mathbf{X}'(\mathbf{S}^{-1}\otimes\mathbf{I})\mathbf{y}]$

Compare this estimate to the previous one. If they are similar enough, exit the iterations.

(4) Recompute $\mathbf{S}$ using $s_{ij} = \mathbf{y}_i'\mathbf{y}_j - \hat{\beta}_i\mathbf{x}_i'\mathbf{y}_j - \hat{\beta}_j\mathbf{x}_j'\mathbf{y}_i + \hat{\beta}_i\hat{\beta}_j\mathbf{x}_i'\mathbf{x}_j$, i,j = 1,2.

(5) Go back to step (1) and continue.

Our iterations produce the two slope estimates

1: .505335 .541741
2: .601889 .564998
3: .614884 .566875
4: .616559 .567186
5: .616775 .567227
6: .616803 .567232
7: .616807 .567232 converged.

At convergence, we find the estimate of the asymptotic covariance matrix of the estimates as

$[\mathbf{X}'(\mathbf{S}^{-1}\otimes\mathbf{I})\mathbf{X}]^{-1} = \begin{bmatrix} .155355 & .00576887 \\ .00576887 & .029348 \end{bmatrix}$ and $\mathbf{S} = \begin{bmatrix} .8483899 & .1573814 \\ .1573814 & .3205369 \end{bmatrix}$.

To use the likelihood ratio method to test the hypothesis, we will require the restricted maximum likelihood estimate. Under the hypothesis,the model is the one in Section 15.2.2. The restricted estimate is given in (15-12) and the equations which follow. To obtain them, we make a small modification in our algorithm above. We replace step (3) with

(3') $\hat{\beta} = [s^{11}\mathbf{x}_1'\mathbf{y}_1 + s^{22}\mathbf{x}_2'\mathbf{y}_2 + s^{12}(\mathbf{x}_1'\mathbf{y}_2 + \mathbf{x}_2'\mathbf{y}_1)]/[s^{11}\mathbf{x}_1'\mathbf{x}_1 + s^{22}\mathbf{x}_2'\mathbf{x}_2 + 2s^{12}\mathbf{x}_1'\mathbf{x}_2]$.

Step 4 is then computed using this common estimate for both $\hat{\beta}_1$ and $\hat{\beta}_2$. The iterations produce

1: .5372671
2: .5703837
3: .5725274
4: .5726687
5: .5726780
6: .5726786 converged.

At this estimate, the estimate of $\boldsymbol\Sigma$ is $\begin{bmatrix} .8529188 & .1609926 \\ .1609926 & .3203732 \end{bmatrix}$. The likelihood ratio statistic is given in (15-56).

Using our unconstrained and constrained estimates, we find $|\mathbf{W}_u|$ = .2471714 and $|\mathbf{W}_r|$ = .2473338. The statistic is λ = 20(ln.2473338 - ln.2471714) = .0131. This is far below the critical value of 3.84, so once again, we do not reject the hypothesis.

4. Prove that in the model $\begin{array}{l}\mathbf{y}_1 = \mathbf{X}_1\boldsymbol\beta_1 + \boldsymbol\varepsilon_1 \\ \mathbf{y}_2 = \mathbf{X}_2\boldsymbol\beta_2 + \boldsymbol\varepsilon_2\end{array}$, generalized least squares is equivalent to equation by equation ordinary least squares if $\mathbf{X}_1 = \mathbf{X}_2$. Does your result hold if it is also known that $\boldsymbol\beta_1 = \boldsymbol\beta_2$?

The GLS estimator is

$$\hat{\boldsymbol\beta} = \begin{bmatrix} \sigma^{11}\mathbf{X'X} & \sigma^{12}\mathbf{X'X} \\ \sigma^{12}\mathbf{X'X} & \sigma^{22}\mathbf{X'X} \end{bmatrix}^{-1} \begin{bmatrix} \sigma^{11}\mathbf{X'y}_1 + \sigma^{12}\mathbf{X'y}_2 \\ \sigma^{12}\mathbf{X'y}_1 + \sigma^{22}\mathbf{X'y}_2 \end{bmatrix}$$

The matrix to be inverted equals $[\boldsymbol\Sigma^{-1}\otimes\mathbf{X}'\mathbf{X}]^{-1}$. But, $[\boldsymbol\Sigma^{-1}\otimes\mathbf{X}'\mathbf{X}]^{-1} = \boldsymbol\Sigma\otimes(\mathbf{X}'\mathbf{X})^{-1}$. (See (2-76).) Therefore,

$$\hat{\boldsymbol\beta} = \begin{bmatrix} \sigma_{11}(\mathbf{X'X})^{-1} & \sigma_{12}(\mathbf{X'X})^{-1} \\ \sigma_{12}(\mathbf{X'X})^{-1} & \sigma_{22}(\mathbf{X'X})^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \sigma^{11}\mathbf{X'y}_1 + \sigma^{12}\mathbf{X'y}_2 \\ \sigma^{12}\mathbf{X'y}_1 + \sigma^{22}\mathbf{X'y}_2 \end{bmatrix}$$

We now make the replacements $\mathbf{X}'\mathbf{y}_1 = (\mathbf{X}'\mathbf{X})\mathbf{b}_1$ and $\mathbf{X}'\mathbf{y}_2 = (\mathbf{X}'\mathbf{X})\mathbf{b}_2$. After multiplying out the product, we find that

$$\hat{\boldsymbol\beta} = \begin{bmatrix} \sigma_{11}\sigma^{11}\mathbf{b}_1 + \sigma_{11}\sigma^{12}\mathbf{b}_2 + \sigma_{12}\sigma^{12}\mathbf{b}_1 + \sigma_{12}\sigma^{22}\mathbf{b}_2 \\ \sigma_{12}\sigma^{11}\mathbf{b}_1 + \sigma_{12}\sigma^{12}\mathbf{b}_2 + \sigma_{22}\sigma^{12}\mathbf{b}_1 + \sigma_{22}\sigma^{22}\mathbf{b}_2 \end{bmatrix} = \begin{bmatrix} (\sigma_{11}\sigma^{11} + \sigma_{12}\sigma^{12})\mathbf{b}_1 + (\sigma_{11}\sigma^{12} + \sigma_{12}\sigma^{22})\mathbf{b}_2 \\ (\sigma_{12}\sigma^{11} + \sigma_{22}\sigma^{12})\mathbf{b}_1 + (\sigma_{12}\sigma^{12} + \sigma_{22}\sigma^{22})\mathbf{b}_2 \end{bmatrix}$$

The four scalar terms in the matrix product are the corresponding elements of $\Sigma\Sigma^{-1} = \mathbf{I}$. Therefore, $\hat{\beta} = \begin{pmatrix}\mathbf{b}_1\\ \mathbf{b}_2\end{pmatrix}$.

5. Consider the two equation system $\begin{matrix} y_1 \\ y_2 \end{matrix} = \begin{matrix} \beta_1 x_1 + \varepsilon_1 \\ \beta_2 x_2 + \beta_3 x_3 + \varepsilon_2 \end{matrix}$. Assume the disturbance variances and covariance are known. Now, suppose that the analyst of this model applies GLS, but erroneously omits x_3 from the second equation. What effect does this specification error have on the consistency of the estimator of β_1?

The algebraic result is a little tedious, but straightforward. The GLS estimator which is computed is

$$\begin{pmatrix}\hat{\beta}_1\\ \hat{\beta}_2\end{pmatrix} = \begin{bmatrix}\sigma^{11}\mathbf{x}_1'\mathbf{x}_1 & \sigma^{12}\mathbf{x}_1'\mathbf{x}_2\\ \sigma^{12}\mathbf{x}_2'\mathbf{x}_1 & \sigma^{22}\mathbf{x}_2'\mathbf{x}_2\end{bmatrix}^{-1}\begin{bmatrix}\sigma^{11}\mathbf{x}_1'\mathbf{y}_1 + \sigma^{12}\mathbf{x}_1'\mathbf{y}_2\\ \sigma^{12}\mathbf{x}_2'\mathbf{y}_1 + \sigma^{22}\mathbf{x}_2'\mathbf{y}_2\end{bmatrix}.$$

It helps at this point to make some simplifying substitutions. The elements in the inverse matrix, σ^{ij}, are all equal to elements of the original matrix divided by the determinant. But, the determinant appears in the leading matrix, which is inverted and in the trailing vector (which is not). Therefore, the determinant will cancel out. Making the substitutions, $\begin{pmatrix}\hat{\beta}_1\\ \hat{\beta}_2\end{pmatrix} = \begin{bmatrix}\sigma_{22}\mathbf{x}_1'\mathbf{x}_1 & -\sigma_{12}\mathbf{x}_1'\mathbf{x}_2\\ -\sigma_{12}\mathbf{x}_2'\mathbf{x}_1 & \sigma_{11}\mathbf{x}_2'\mathbf{x}_2\end{bmatrix}^{-1}\begin{bmatrix}\sigma_{22}\mathbf{x}_1'\mathbf{y}_1 - \sigma_{12}\mathbf{x}_1'\mathbf{y}_2\\ -\sigma_{12}\mathbf{x}_2'\mathbf{y}_1 + \sigma_{22}\mathbf{x}_2'\mathbf{y}_2\end{bmatrix}$. Now, we are concerned with probability limits. We divide every element of the matrix to be inverted by n, then because of the inversion, divide the vector on the right by n as well. Suppose, for simplicity, that $\lim_{n\to\infty}\mathbf{x}_i'\mathbf{x}_j/n = q_{ij}$, i,j = 1,2,3. Then, $plim\begin{pmatrix}\hat{\beta}_1\\ \hat{\beta}_2\end{pmatrix} = \begin{bmatrix}\sigma_{22}q_{11} & -\sigma_{12}q_{12}\\ -\sigma_{12}q_{12} & \sigma_{11}q_{22}\end{bmatrix}^{-1} plim\begin{bmatrix}\sigma_{22}\mathbf{x}_1'\mathbf{y}_1/n - \sigma_{12}\mathbf{x}_1'\mathbf{y}_2/n\\ -\sigma_{12}\mathbf{x}_2'\mathbf{y}_1/n + \sigma_{11}\mathbf{x}_2'\mathbf{y}_2/n\end{bmatrix}$

Then, we will use $plim\ (1/n)\mathbf{x}_1'\mathbf{y}_1 = \beta_1 q_{11} + plim\ (1/n)\mathbf{x}_1 N\varepsilon_1 = \beta_1 q_{11}$

$plim\ (1/n)\mathbf{x}_1'\mathbf{y}_2 = \beta_2 q_{12} + \beta_3 q_{13}$

$plim\ (1/n)\mathbf{x}_2'\mathbf{y}_1 = \beta_1 q_{12}$

$plim\ (1/n)\mathbf{x}_2'\mathbf{y}_2 = \beta_2 q_{22} + \beta_3 q_{23}$.

Therefore, after multiplying out all the terms,

$$plim\begin{pmatrix}\hat{\beta}_1\\ \hat{\beta}_2\end{pmatrix} = \begin{bmatrix}\sigma_{22}q_{11} & -\sigma_{12}q_{12}\\ -\sigma_{12}q_{12} & \sigma_{11}q_{22}\end{bmatrix}^{-1}\begin{bmatrix}\beta_1\sigma_{22}q_{11} - \beta_2\sigma_{12}q_{12} - \beta_3\sigma_{12}q_{13}\\ -\beta_1\sigma_{12}q_{12} + \beta_2\sigma_{11}q_{22} + \beta_3\sigma_{11}q_{23}\end{bmatrix}.$$

The inverse matrix is $\dfrac{1}{\sigma_{11}\sigma_{22}q_{11}q_{22} - (\sigma_{12}q_{12})^2}\begin{bmatrix}\sigma_{11}q_{22} & \sigma_{12}q_{12}\\ \sigma_{12}q_{12} & \sigma_{22}q_{22}\end{bmatrix}$, so with $\Delta = (\sigma_{11}F_{22}q_{11}q_{22} - (F_{12}q_{12})^2)$

$$\text{plim}\begin{pmatrix}\hat{\beta}_1\\ \hat{\beta}_2\end{pmatrix} = \left[\frac{1}{\Delta}\begin{pmatrix}\sigma_{11}q_{22} & \sigma_{12}q_{12}\\ \sigma_{12}q_{12} & \sigma_{22}q_{11}\end{pmatrix}\right]^{-1}\begin{bmatrix}\beta_1\sigma_{22}q_{11} - \beta_2\sigma_{12}q_{12} - \beta_3\sigma_{12}q_{13}\\ -\beta_1\sigma_{12}q_{12} + \beta_2\sigma_{11}q_{22} + \beta_3\sigma_{11}q_{23}\end{bmatrix}.$$

Taking the first coefficient separately and collecting terms,

$\text{plim}\,\hat{\beta}_1 = \beta_1[\sigma_{11}\sigma_{22}q_{11}q_{22} - (\sigma_{12}q_{12})^2]/\Delta + \beta_2[\sigma_{11}q_{22}\sigma_{12}q_{12} + \sigma_{12}q_{12}\sigma_{11}q_{22}]/\Delta + \beta_3[\sigma_{11}q_{22}\sigma_{12}q_{13} + \sigma_{12}q_{12}\sigma_{11}q_{23}]/\Delta$

The first term in brackets equals Δ while the second equals 0. That leaves

$\text{plim}\ \hat{\beta}_1 = \beta_1 - \beta_3[\sigma_{11}\sigma_{12}(q_{22}q_{13} - q_{12}q_{23})]/\Delta$ which is not equal to β_1. There are two special cases worthy of note, though. The right hand side does equal β_1 if either (1) $\sigma_{12} = 0$; the regressions are actually unrelated, or (2) $q_{12} = q_{13} = 0$; the regressors in the two equations are uncorrelated. The second of these is similar to our finding for omitted variables in the classical regression model.

6. Consider the system $\begin{matrix} y_1 = \alpha_1 + \beta x + \varepsilon_1 \\ y_2 = \alpha_2 + \quad\ \varepsilon_2 \end{matrix}$ .The disturbances are freely correlated. Prove that GLS applied to the system leads to the OLS estimates of α_1 and α_2 but to a mixture of the least squares slopes in the regressions of y_1 and y_2 on x as the estimator of β. What is the mixture? To simplify the algebra, assume (with no loss of generality) that $\bar{x} = 0$.

The model is $\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{i} & \mathbf{x} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{i} \end{bmatrix} \begin{pmatrix} \alpha_1 \\ \beta \\ \alpha_2 \end{pmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \end{bmatrix}$. The GLS estimator of the full coefficient vector, θ,

is $\hat{\theta} = \begin{bmatrix} \sigma^{11}\begin{pmatrix} n & n\bar{x} \\ n\bar{x} & \mathbf{x}'\mathbf{x} \end{pmatrix} & \sigma^{12}\begin{pmatrix} n \\ n\bar{x} \end{pmatrix} \\ \sigma^{12}\begin{pmatrix} n & n\bar{x} \end{pmatrix} & \sigma^{22}n \end{bmatrix}^{-1} \begin{bmatrix} \sigma^{11}\begin{pmatrix} n\bar{y}_1 \\ \mathbf{x}'\mathbf{y}_1 \end{pmatrix} + \sigma^{12}\begin{pmatrix} n\bar{y}_2 \\ \mathbf{x}'\mathbf{y}_2 \end{pmatrix} \\ \sigma^{12}n\bar{y}_1 + \sigma^{22}n\bar{y}_2 \end{bmatrix}$. Let q_{xx} equal $\mathbf{x}'\mathbf{x}/n$, $q_{x1} = \mathbf{x}'\mathbf{y}_1/n$ and, q_{x2} $= \mathbf{x}'\mathbf{y}_2/n$. The ns in the inverse and in the vector cancel. Also, as suggested, we assume that $\bar{x} = 0$. As in the previous exercise, we replace elements of the inverse with elements from the original matrix and cancel the determinant which multiplies the matrix (after inversion) and divides the vector. Thus, $\hat{\theta} = \begin{bmatrix} \sigma_{11} & 0 & -\sigma_{12} \\ 0 & \sigma_{22}q_{xx} & 0 \\ -\sigma_{12} & 0 & \sigma_{11} \end{bmatrix}^{-1} \begin{bmatrix} \sigma_{22}\bar{y}_1 - \sigma_{12}\bar{y}_2 \\ \sigma_{11}q_{x1} - \sigma_{12}q_{x2} \\ -\sigma_{12}\bar{y}_1 + \sigma_{11}\bar{y}_2 \end{bmatrix}$. The inverse of the matrix is straightforward. Proceeding

directly, we obtain $\hat{\theta} = \dfrac{1}{\sigma_{22}q_{xx}(\sigma_{11}\sigma_{22} - \sigma_{12}^2)} \begin{bmatrix} \sigma_{11}\sigma_{22}q_{xx} & 0 & \sigma_{12}\sigma_{22}q_{xx} \\ 0 & \sigma_{11}\sigma_{22} - \sigma_{12}^2 & 0 \\ \sigma_{12}\sigma_{22}q_{xx} & 0 & \sigma_{22}q_{xx} \end{bmatrix}^{-1} \begin{bmatrix} \sigma_{22}\bar{y}_1 - \sigma_{12}\bar{y}_2 \\ \sigma_{11}q_{x1} - \sigma_{12}q_{x2} \\ -\sigma_{12}\bar{y}_1 + \sigma_{11}\bar{y}_2 \end{bmatrix}$.

It remains only to multiply the matrices and collect terms. The result is

$$\hat{\alpha}_1 = \bar{y}_1,\ \hat{\alpha}_2 = \bar{y}_2,\ \hat{\beta} = [(q_{x1}/q_{xx}) - (\sigma_{12}\sigma_{22})(q_{x2}/q_{xx})] = b_1 - \gamma b_2.$$

7. For the model
$$\begin{aligned} y_1 &= \alpha_1 + \beta x + \varepsilon_1 \\ y_2 &= \alpha_2 \qquad\ + \varepsilon_2 \\ y_3 &= \alpha_3 \qquad\ + \varepsilon_3 \end{aligned}$$
assume that $y_{i2} + y_{i3} = 1$ at every observation. Prove that the sample covariance matrix of the least squares residuals from the three equations will be singular, thereby precluding computation of the FGLS estimator. How could you proceed in this case?

Once again, nothing is lost by assuming that $\bar{x} = 0$. Now, the OLS estimators are

$$a_1 = \bar{y}_1,\ a_2 = \bar{y}_2,\ a_3 = \bar{y}_3,\ b = \mathbf{x}'\mathbf{y}_1/\mathbf{x}'\mathbf{x}.$$

The vector of residuals is
$$\begin{aligned} e_{i1} &= y_{i1} - \bar{y}_1 - bx_i \\ e_{i2} &= y_{i2} - \bar{y}_2 \\ e_{i3} &= y_{i3} - \bar{y}_3 \end{aligned}$$

Now, if $y_{i2} + y_{i3} = 1$ at every observation, then $(1/n)\Sigma_i(y_{i2} + y_{i3}) = \bar{y}_2 + \bar{y}_3 = 1$ as well. Therefore, by just adding the two equations, we see that $e_{i2} + e_{i3} = 0$ for every observation. Let $\mathbf{e}_i$ be the 3×1 vector of residuals. Then, $\mathbf{e}_i'\mathbf{c} = 0$, where $\mathbf{c} = [0,1,1]'$. The sample covariance matrix of the residuals is

$\mathbf{S} = [(1/n)\Sigma_i\,\mathbf{e}_i\mathbf{e}_i']$. Then, $\mathbf{Sc} = [(1/n)\Sigma_i\,\mathbf{e}_i\mathbf{e}_i']\mathbf{c} = [(1/n)\Sigma_i\,\mathbf{e}_i\mathbf{e}_i'\mathbf{c}] = [(1/n)\Sigma_i\,\mathbf{e}_i\times 0] = \mathbf{0}$, which means, by definition, that $\mathbf{S}$ is singular.

We can proceed simply by dropping the third equation. The adding up condition implies that $\alpha_3 = 1 - \alpha_2$. So, we can treat the first two equations as a seemingly unrelated regression model and estimate a_3 using the estimate of α_2.

8. Continuing the analysis of Section 14.3.2, we find that a translog cost function for one output and three factor inputs which does not impose constant returns to scale is

$$\ln C = \alpha + \beta_1\ln p_1 + \beta_2\ln p_2 + \beta_3\ln p_3 + \delta_{11}(\ln^2 p_1)/2 + \delta_{12}\ln p_1\ln p_2 + \delta_{13}\ln p_1\ln p_3 + \delta_{22}(\ln^2 p_2)/2 + \delta_{23}\ln p_2\ln p_3 + \delta_{33}(\ln^2 p_3)/2 + \gamma_{y1}\ln Y\ln p_1 + \gamma_{y2}\ln Y\ln p_2 + \gamma_{y3}\ln Y\ln p_3 + \beta_y\ln Y + \beta_{yy}(\ln^2 Y)/2 + \varepsilon_c$$

The factor share equations are

$$S_1 = \beta_1 + \delta_{11}\ln p_1 + \delta_{12}\ln p_2 + \delta_{13}\ln p_3 + \gamma_{y1}\ln Y + \varepsilon_1$$
$$S_2 = \beta_2 + \delta_{12}\ln p_1 + \delta_{22}\ln p_2 + \delta_{23}\ln p_3 + \gamma_{y2}\ln Y + \varepsilon_2$$
$$S_3 = \beta_3 + \delta_{13}\ln p_1 + \delta_{23}\ln p_2 + \delta_{33}\ln p_3 + \gamma_{y3}\ln Y + \varepsilon_3.$$

[See Christensen and Greene (1976) for analysis of this model.]

(a) The three factor shares must add identically to 1. What restrictions does this place on the model parameters?

(b) Show that the adding up condition in (14-39) can be imposed directly on the model by specifying the translog model in (C/p_3), (p_1/p_3), and (p_2/p_3) and dropping the third share equation. (See Example17.10.) Notice that this reduces the number of free parameters in the model to 10.

(c) Continuing part (b), the model as specified, with the symmetry and equality restrictions has 15 parameters. By imposing the constraints, you reduce this to 10 in the estimating equations? How would you obtain estimates of the parameters not estimated directly?

The remaining parts of this exercise will require specialized software. The *TSP*, *LIMDEP*, *Shazam*, *Gauss*, and *E-Views* programs are five that could be used. All estimation is to be done using the data in Section 14.3.1.

(d) Estimate each of the three equations you obtained in part (b) by ordinary least squares. Do the estimates appear to satisfy the cross equation equality and symmetry equations implied by the theory?

(e) Using the data in Section 14.3.1, estimate the full system of three equations (cost and the two independent shares) imposing the symmetry and cross equation equality constraints.

(f) Using your parameter estimates, compute the estimates of the elasticities in (15-7) at the means of the variables.

(g) Use a likelihood ratio to test the joint hypothesis that $\gamma_{yi} = 1$, i=1,2,3. [**Hint**: just drop the relevant variables from the model.]

By adding the share equations vertically, we find the restrictions

$$\beta_1 + \beta_2 + \beta_3 = 1$$
$$\delta_{11} + \delta_{12} + \delta_{13} = 0$$
$$\delta_{12} + \delta_{22} + \delta_{23} = 0$$
$$\delta_{13} + \delta_{23} + \delta_{33} = 0$$
$$\gamma_{y1} + \gamma_{y2} + \gamma_{y3} = 0.$$

Note that the adding up condition also implies $\varepsilon_1 + \varepsilon_2 + \varepsilon_3 = 0$.

We will eliminate the third share equation. The restrictions imply

$$\beta_3 = 1 - \beta_1 - \beta_2$$
$$\delta_{13} = -\delta_{11} - \delta_{12}$$
$$\delta_{23} = -\delta_{12} - \delta_{22}$$
$$\delta_{33} = -\delta_{13} - \delta_{23} = \delta_{11} + \delta_{22} + 2\delta_{12}$$
$$\gamma_{y3} = -\gamma_{y1} - \gamma_{y2}.$$

By inserting these in the three share equations, we find

$$\begin{aligned}
S_1 &= \beta_1 + \delta_{11}\ln p_1 + \delta_{12}\ln p_2 - \delta_{11}\ln p_3 - \delta_{12}\ln p_3 + \gamma_{y1}\ln Y + \varepsilon_1 \\
&= \beta_1 + \delta_{11}\ln(p_1/p_3) + \delta_{12}\ln(p_2/p_3) + \gamma_{y1}\ln Y + \varepsilon_1 \\
S_2 &= \beta_2 + \delta_{12}\ln p_1 + \delta_{22}\ln p_2 - \delta_{12}\ln p_3 - \delta_{22}\ln p_3 + \gamma_{y2}\ln Y + \varepsilon_2 \\
&= \beta_2 + \delta_{12}\ln(p_1/p_3) + \delta_{22}\ln(p_2/p_3) + \gamma_{y2}\ln Y + \varepsilon_2 \\
S_3 &= 1 - \beta_1 - \beta_2 - \delta_{11}\ln p_1 - \delta_{12}\ln p_1 - \delta_{12}\ln p_2 - \delta_{22}\ln p_2 + \delta_{11}\ln p_3 + \delta_{12}\ln p_3 + \delta_{12}\ln p_3 \\
&\qquad + \delta_{22}\ln p_3 - \gamma_{y1}\ln p_3 - \gamma_{y2}\ln p_3 - \varepsilon_1 - \varepsilon_2 \\
&= 1 - S_1 - S_2
\end{aligned}$$

For the cost function, making the substitutions for β_3, δ_{13}, δ_{23}, δ_{33}, and γ_{y3} produces

$$\begin{aligned}\ln C &= \alpha + \beta_1(\ln p_1 - \ln p_3) + \beta_2(\ln p_2 - \ln p_3)\\ &\quad + \delta_{11}((\ln^2 p_1)/2 - \ln p_1\ln p_3 + (\ln^2 p_3)/2)\\ &\quad + \delta_{22}((\ln^2 p_2)/2 - \ln p_2\ln p_3 + (\ln^2 p_3)/2) + \delta_{12}(\ln p_1\ln p_2 - \ln p_1\ln p_3 - \ln p_2\ln p_3 + (\ln^2 p_3))\\ &\quad + \gamma_{y1}\ln Y(\ln p_1 - \ln p_3) + \gamma_{y2}\ln Y(\ln p_2 - \ln p_3) + \beta_y\ln Y + \beta_{yy}(\ln^2 Y)/2 + \varepsilon_c\\ &= \alpha + \beta_1\ln(p_1/p_3) + \beta_2\ln(p_2/p_3)\\ &\quad + \delta_{11}(\ln^2(p_1/p_3))/2 + \delta_{22}(\ln^2(p_2/p_3))/2 + \delta_{12}\ln(p_1/p_3)\ln(p_2/p_3)\\ &\quad + \gamma_{y1}\ln Y\ln(p_1/p_3) + \gamma_{y2}\ln Y\ln(p_2/p_3) + \beta_y\ln Y + \beta_{yy}(\ln^2 Y)/2 + \varepsilon_c\end{aligned}$$

The system of three equations (cost and two shares) can be estimated as discussed in the text. Invariance is achieved by using a maximum likelihood estimator. The five parameters eliminated by the restrictions can be estimated after the others are obtained just by using the restrictions. The restrictions are linear, so the standard errors are also striaghtforward to obtain.

The least squares estimates are shown below. Estimated standard errors appear in parentheses.

Variable	**Cost Function**	**Capital Share**	**Labor Share**
One	51.32 (45.91)	-.0174 (.4697)	.2172 (.2408)
$\ln(p_k/p_f)$	-21.74 (20.14)	.2380 (.1045)	.0033 (.0534)
$\ln(p_l/p_f)$	32.39 (21.81)	.0065 (.1059)	.0168 (.0542)
$\ln^2(p_k/p_f)/2$	4.596 (4.604)	-.0007 (.0098)	-.0117 (.0050)
$\ln^2(p_l/p_f)/2$	8.216 (5.159)		
$\ln(p_k/p_f)\ln(p_l/p_f)$	-6.238 (4.684)		
$\ln Y$	1.674 (.9297)		
$\ln^2 Y/2$	,006997 (.0313)		
$\ln Y\ln(p_k/p_f)$	-.3223 (.2652)		
$\ln Y\ln(p_l/p_f)$	.08631 (.1981)		

The estimates do not even come close to satisfying the cross equation restrictions. The parameters in the cost function are extremely large, owing primarily to rather severe multicollinearity among the price terms.

The results of estimation of the system by direct maximum likelihood are shown. The convergence criterion is the value of Belsley (discussed near the end of Section 5.5). The value α shown below is $\mathbf{g}'\mathbf{H}^{-1}\mathbf{g}$ where $\mathbf{g}$ is the gradient and $\mathbf{H}$ is the Hessian of the log-likelihood.

```
Iteration  0, F=46.76391, ln*S*= -7.514268, α= 2.054399
Iteration  1, F=136.7448, ln*S*= -16.51236, α= .5796486
Iteration  2, F=146.9803, ln*S*= -17.53591, α= .02179947
Iteration  3, F=147.2268, ln*S*= -17.56055, α= .0004222
```

Residual covariance matrix

	Cost	Capital	Labor
Cost	.0145572		
Capital	.000304768	.00303853	
Labor	-.000317554	-.000887258	.000798128

Coefficient	Estimate	Std. Error
α	-6.41878	.6637
β_k	-.0546555	.2422
β_l	.250976	.2138
δ_{kk}	.245259	.06904
δ_{ll}	.0245770	.04788
δ_{kl}	-.00403448	.04779
β_y	.572452	.1340
β_{yy}	.0456587	.01908
γ_{yk}	-.00124236	.008409
γ_{yl}	-.0116921	.004442
β_f	.8036795	
δ_{kf}	-.2412245	
δ_{lf}	-.0205425	
δ_{ff}	.261767	
γ_{yf}	.0129345	

The means of the variables are: $\overline{Y} = 3531.8$, $\overline{p}_k = 169.35$, $\overline{p}_l = 2.039$, $\overline{p}_f = 26.41$. The three factor shares computed at these means are $S_k = .4182$, $S_l = .0865$, $S_f = .4953$. (The sample means are .411, .0954, and .4936.) The matrix of elasticities computed according to (15-72) is

$$\Sigma = \begin{array}{ccc|c} k & l & f & \\ .01115 & & & k \\ .8885 & -7.2756 & & l \\ -.1646 & .5206 & .04819 & f \end{array}$$

(Two of the three diagonals have the `wrong' sign. This may be due to the very small sample size. The cross elasticities however do conform to what one might expect, the primary one being the evident substitution between capital and fuel.

To test the hypothesis that $\gamma_{yi} = 0$, we reestimate the model without the interaction terms between ln$Y$ and the prices in the cost function and without ln$Y$ in the factor share equations. The iterations for this restricted model are shown below.

Iter.= 0, F=46.76391, log|**S**|= -7.514268, α= 1.912223
Iter.= 1, F=123.7521, log|**S**|= -15.21308, α= .5888180
Iter.= 2, F=136.3410, log|**S**|= -16.47198, α= .2771995
Iter.= 3, F=141.3491, log|**S**|= -16.97279, α= .08024513
Iter.= 4, F=142.5591, log|**S**|= -17.09379, α= .01636212
Converged achieved

Since we are interested only in the test statistic, we have not listed the parameter estimates. The test statistic given in (17-26) is $\lambda = T(\ln|\mathbf{S}_r| - \ln|\mathbf{S}_u|) = 20(-17.09379 - (-17.56055)) = 9.3352$. There are two restrictions since only two of the three parameters are free. The critical value from the chi-squared table is 5.99, so we would reject the hypothesis.

Chapter 15

Simultaneous Equations Models

1. Consider the following two equation model:

$$y_1 = \gamma_1 y_2 + \beta_{11}x_1 + \beta_{21}x_2 + \beta_{31}x_3 + \varepsilon_1$$
$$y_2 = \gamma_2 y_1 + \beta_{12}x_1 + \beta_{22}x_2 + \beta_{32}x_3 + \varepsilon_2.$$

(a) Verify that as stated, neither equation is identified.

(b) Establish whether or not the following restrictions are sufficient to identify (or partially identify) the model:

(1) $\beta_{21} = \beta_{32} = 0$,
(2) $\beta_{12} = \beta_{22} = 0$,
(3) $\gamma_1 = 0$,
(4) $\gamma_1 = \gamma_2$ and $\beta_{32} = 0$,
(5) $\sigma_{12} = 0$ and $\beta_{31} = 0$,
(6) $\gamma_1 = 0$ and $\sigma_{12} = 0$,
(7) $\beta_{21} + \beta_{22} = 1$,
(8) $\sigma_{12} = 0$, $\beta_{21} = \beta_{22} = \beta_{31} = \beta_{32} = 0$,
(9) $\sigma_{12} = 0$, $\beta_{11} = \beta_{21} = \beta_{22} = \beta_{31} = \beta_{32} = 0$.

Since nothing is excluded from either equation and there are no other restrictions, neither equation passes the order condition for identification.

(1) We use (15-12) and the equations which follow it. For the first equation, $[\mathbf{A}_3',\mathbf{A}_5'] = \beta_{22}$, a scalar which has rank M-1 = 1 unless $\beta_{22} = 0$. For the second, $[\mathbf{A}_3',\mathbf{A}_5'] = \beta_{31}$. Thus, both equations are identified.

(2) This restriction does not restrict the first equation, so it remains unidentified. The second equation is now identified, as $[\mathbf{A}_3',\mathbf{A}_5'] = [\beta_{11},\beta_{21}]$ has rank 1 if either of the two ceofficients are nonzero.

(3) If γ_1 equals 0, the model becomes partially recursive. The first equation becomes a regression which can be estimated by ordinary least squares. However, the second equation continues to fail the order condition. To see the problem, consider that even with the restriction, any linear combination of the two equations has the same variables as the original second eqation.

(4) We know from above that if $\beta_{32} = 0$, the second equation is identifiable. If it is, then γ_2 is identified. We may treat it as known. As such, γ_1 is known. By regressing $\mathbf{y}_1 - \gamma_1\mathbf{y}_2$ on the **x**s, we would obtain estimates of the remaining parameters, so these restrictions identify the model. It is instructive to analyze this from the standpoint of false structures as done in the text. A false structure which incorporates the known restrictions would be $\begin{bmatrix} 1 & -\gamma \\ -\lambda & 1 \\ \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \\ \beta_{31} & 0 \end{bmatrix} \times \begin{bmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{bmatrix}$. If the false structure is to obey the restrictions, then $f_{11} - \gamma f_{21} = 1$, $f_{22} - \gamma f_{12} = 1$, $f_{21} - \gamma f_{11} = f_{12} - \gamma f_{22}$, $\beta_{31}f_{12} = 0$. It follows then that $f_{12} = 0$ so $f_{11} = 1$. Then, $f_{21} - \gamma f_{11} = -\gamma$ or $f_{21} = (f_{11} - 1)\gamma$ so that $f_{11} - \gamma^2(f_{11} - 1) = 1$. This can only hold for all values of γ if $f_{11} = 1$ and, then, $f_{21} = 0$. Therefore, $\mathbf{F} = \mathbf{I}$ which establishes identification.

(5) If $\beta_{31} = 0$, the first equation is identified by the usual rank and order conditions. Consider, then, the off-diagonal element of $\mathbf{\Sigma} = \mathbf{\Gamma}'\mathbf{\Omega}\mathbf{\Gamma}$. $\mathbf{\Omega}$ is identified since it is the reduced form covariance matrix. The off-diagonal element is $\sigma_{12} = \omega_{11} + \omega_{22} - (\gamma_1 + \gamma_2)\omega_{12} = 0$. Since γ_1 is zero, $\gamma_2 = \omega_{12}/(\omega_{11} + \omega_{22})$. With γ_2 known, the remaining parameters are estimable by least squares regression of $(\mathbf{y}_2 - \gamma_2\mathbf{y}_1)$ on the **x**s. Therefore, the restrictions identify the model.

(6) Since this is only a single restriction, it will not likely identify the entire model. Consider again the false structure. The restrictions implied by the theory are $f_{11} - \gamma_2 f_{21} = 1$, $f_{22} - \gamma_1 f_{12} = 1$, $\beta_{21}f_{11} + \beta_{22}f_{21} =$

$\beta_{21}f_{12} + \beta_{22}f_{22}$. The three restrictions on four unknown elements of **F** do not serve to pin down any of them. This restriction does not even partially identify the model.

(7) The last four restrictions remove x_2 and x_3 from the model. The remaining model is not identified by the usual rank and order conditions. From part (5), we see that the first restriction implies $\sigma_{12} = \omega_{11} + \omega_{22} - (\gamma_1 + \gamma_2)\omega_{12} = 0$. But, with neither γ_1 nor γ_2 specified, this does not identify either parameter.

(8) The first equation is identified by the conventional rank and order conditions. The second equation fails the order condition. But, the restriction $\sigma_{12} = 0$ provides the necessary additional information needed to identify the model. For simplicity, write the model with the restrictions imposed as

$$y_1 = \gamma_1 y_2 + \varepsilon_1 \text{ and } y_2 = \gamma_2 y_1 + \beta x + \varepsilon_2.$$

The reduced form is $y_1 = \pi_1 x + v_1$ and $y_2 = \pi_2 x + v_2$

where $\pi_1 = \gamma_1\beta/\Delta$ and $\pi_2 = \beta/\Delta$ with $\Delta = (1 - \gamma_1\gamma_2)$, and $v_1 = (\varepsilon_1 + \gamma_1\varepsilon_2)/\Delta$ and $v_2 = (\varepsilon_2 + \gamma_2\varepsilon_1)/\Delta$. The reduced form variances and covariances are $\omega_{11} = (\gamma_1^2\sigma_{22} + \sigma_{11})/\Delta^2$, $\omega_{22} = (\gamma_2^2\sigma_{11} + \sigma_{22})/\Delta^2$, $\omega_{12} = (\gamma_1\sigma_{22} + \gamma_2\sigma_{11})/\Delta^2$. All reduced form parameters are estimable directly by using least squares, so the reduced form is identified in all cases. Now, $\gamma_1 = \pi_1/\pi_2$. σ_{11} is the residual variance in the euqation $(y_1 - \gamma_1 y_2) = \varepsilon_1$, so σ_{11} must be estimable (identified) if γ_1 is. Now, with a bit of manipulation, we find that $\gamma_1\omega_{12} - \omega_{11} = -\sigma_{11}/\Delta$. Therefore, with σ_{11} and γ_1 "known" (identified), the only remaining unknown is γ_2, which is therefore identified. With γ_1 and γ_2 in hand, β may be deduced from π_2. With γ_2 and β in hand, σ_{22} is the residual variance in the equation $(y_2 - \beta x - \gamma_2 y_1) = \varepsilon_2$, which is directly estimable, therefore, identified.

2. Verify the rank and order conditions for identification of the second and third behavioral equation in Klein's Model I. [**Hint:** See Example 15.6.]

Following the method in Example 15.6, for identification of the investment equation, we require that the matrix $\begin{bmatrix} (1) & (2) & (3) & (4) & (5) & (6) & (7) & (8) & (9) \\ -1 & \alpha_3 & 0 & 0 & \alpha_3 & 0 & 0 & 0 & 0 \\ 0 & -1 & \gamma_1 & 0 & 0 & 0 & 0 & \gamma_3 & \gamma_2 \\ 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$ have rank 5. Columns (1), (4), (6), (7), and (8) each have one element in a different row, so they are linearly independent. Therefore, the matrix has rank five. For the third equation, the required matrix is $\begin{bmatrix} (1) & (2) & (3) & (4) & (5) & (6) & (7) & (8) & (9) & (10) \\ -1 & 0 & \alpha_1 & 0 & \alpha_3 & 0 & 0 & 0 & \alpha_2 & 0 \\ 0 & -1 & \beta_1 & 0 & 0 & 0 & 0 & 0 & \beta_2 & \beta_3 \\ 1 & 1 & 0 & 0 & 0 & 01 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$.

Columns (4), (6), (7), (9), and (10) are linearly independent.

3. Check the identifiability of the parameters of the following model:

$$[y_1\ y_2\ y_3\ y_4]\begin{bmatrix} 1 & \gamma_{12} & 0 & 0 \\ \gamma_{21} & 1 & \gamma_{23} & \gamma_{24} \\ 0 & \gamma_{32} & 1 & \gamma_{34} \\ \gamma_{41} & \gamma_{42} & 0 & 1 \end{bmatrix} + [x_1\ x_2\ x_3\ x_4\ x_5]\begin{bmatrix} 0 & \beta_{12} & \beta_{13} & \beta_{14} \\ \beta_{21} & 1 & 0 & \beta_{24} \\ \beta_{31} & \beta_{32} & \beta_{33} & 0 \\ 0 & 0 & \beta_{43} & \beta_{44} \\ 0 & \beta_{52} & 0 & 0 \end{bmatrix} = [\varepsilon_1,\varepsilon_2,\varepsilon_3,\varepsilon_4].$$

We find $[\mathbf{A}_3',\mathbf{A}_5']'$ for each equation.

(1) (2) (3) (4)

$$\begin{bmatrix} \gamma_{32} & 1 & \gamma_{34} \\ \beta_{12} & \beta_{13} & \beta_{14} \\ 0 & \beta_{43} & \beta_{4} \\ \beta_{32} & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & \beta_{43} & \beta_{44} \end{bmatrix}, \begin{bmatrix} 1 & \gamma_{12} & 0 \\ \gamma_{41} & \gamma_{42} & 1 \\ \beta_{21} & 1 & 0 \\ 0 & \beta_{52} & 00 \end{bmatrix}, \begin{bmatrix} 1 & \gamma_{12} & 0 \\ \beta_{31} & \beta_{32} & \beta_{33} \\ 0 & \beta_{52} & 0 \end{bmatrix}$$

Identification requires that the rank of each matrix be M-1 = 3. The second is obviously not identified. In (1), none of the three columns can be written as a linear combination of the other two, so it has rank 3. (Although the second and last columns have nonzero elements in the same positions, for the matrix to have short rank, we would require that the third column be a multiple of the second, since the first cannot appear in the linear combination which is to replicate the second column.) By the same logic, (3) and (4) are identified.

4. Obtain the reduced form for the model in Exercise 1 under each of the assumptions made in parts (a) and (b1), (b6), and (b9).

(1). The model is $y_1 = \gamma_1 y_2 + \beta_{11}x_1 + \beta_{21}x_2 + \beta_{31}x_3 + \varepsilon_1$

$y_2 = \gamma_2 y_1 + \beta_{12}x_1 + \beta_{22}x_2 + \beta_{32}x_3 + \varepsilon_2$.

Therefore, $\mathbf{\Gamma} = \begin{bmatrix} 1 & -\gamma_2 \\ -\gamma_1 & 1 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} -\beta_{11} & -\beta_{12} \\ 0 & -\beta_{22} \\ -\beta_{31} & 0 \end{bmatrix}$ and $\mathbf{\Sigma}$ is unrestricted. The reduced form is

$$\mathbf{\Pi} = \frac{1}{1-\gamma_1\gamma_2}\begin{bmatrix} \beta_{11}+\gamma_1\beta_{21} & \gamma_2\beta_{11}+\beta_{12} \\ \gamma_1\beta_{22} & \beta_{22} \\ \beta_{31} & \gamma_2\beta_{31} \end{bmatrix}$$ and

$$\mathbf{\Omega} = (\mathbf{\Gamma}^{-1})'\mathbf{\Sigma}(\mathbf{\Gamma}^{-1}) = \frac{1}{(1-\gamma_1\gamma_2)^2}\begin{bmatrix} \sigma_{11}+\gamma_1^2\sigma_{22} + 2\gamma_1\sigma_{12} & \gamma_2\sigma_{11}+\gamma_1\sigma_{22} + (\gamma_1+\gamma_2)\sigma_{12} \\ \gamma_2\sigma_{11}+\gamma_1\sigma_{22} + (\gamma_1+\gamma_2)\sigma_{12} & \gamma_2^2\sigma_{11}+\sigma_{22} + 2\gamma_1\sigma_{12} \end{bmatrix}$$

(6) The model is $y_1 = \beta_{11}x_1 + \beta_{21}x_2 + \beta_{31}x_3 + \varepsilon_1$

$y_2 = \gamma_2 y_1 + \beta_{12}x_1 + \beta_{22}x_2 + \beta_{32}x_3 + \varepsilon_2$

The first equation is already a reduced form. Substituting it into the second provides the second reduced form.

The coefficient matrix is $\mathbf{P} = \begin{bmatrix} \beta_{11} & \beta_{12}+\gamma_2\beta_{11} \\ \beta_{21} & \beta_{22}+\gamma_2\beta_{21} \\ \beta_{31} & \beta_{32}+\gamma_2\beta_{31} \end{bmatrix}$, $\mathbf{\Gamma}^{-1} = \begin{bmatrix} 1 & \gamma_2 \\ 0 & 1 \end{bmatrix}$ so $\mathbf{\Omega} = (\mathbf{\Gamma}^{-1})'\mathbf{\Sigma}(\mathbf{\Gamma}^{-1}) = \begin{bmatrix} \sigma_{11} & \gamma_2\sigma_{11} \\ \gamma_2\sigma_{11} & \gamma_2^2\sigma_{11}+\sigma_{22} \end{bmatrix}$

(9) The model is

$y_1 = \gamma_1 y_2 + \varepsilon_1$

$y_2 = \gamma_2 y_1 + \beta_{12}x_1 + \varepsilon_2$

Then, $\mathbf{\Pi} = -\mathbf{B}\Gamma^{-1} = [\beta_{12}\gamma_1/(1-\gamma_1\gamma_2) \quad \beta_{12}/(1-\gamma_1\gamma_2)]$ and $\mathbf{\Omega} = \begin{bmatrix} \sigma_{11}+\gamma_1^2\sigma_{22} & \gamma_2\sigma_{11}+\gamma_1\sigma_{22} \\ \gamma_2\sigma_{11}+\gamma_1\sigma_{22} & \gamma_2^2\sigma_{11}+\sigma_{22} \end{bmatrix}$.

5. The following model is specified:

$$y_1 = \gamma_1 y_2 + \beta_{11}x_1 + \varepsilon_1$$
$$y_2 = \gamma_2 y_1 + \beta_{22}x_2 + \beta_{32}x_3 + \varepsilon_2$$

All variables are measured as deviations from their means. The sample of 25 observations produces the following matrix of sums of squares and cross products:

	y_1	y_2	x_1	x_2	x_3
y_1	20	6	4	3	5
y_2	6	10	3	6	7
x_1	4	3	5	2	3
x_2	3	6	2	10	8
x_3	5	7	3	8	15

(a) Estimate the two equations by ordinary least squares.
(b) Estimate the parameters of the two equations by two stage least squares. Also estimate the asymptotic covariance matrix of the two stage least squares estimates.
(c) Obtain the LIML estimates of the parameters of the first equation.
(d) Estimate the two equations by three stage least squares.
(e) Estimate the reduced form coefficient matrix by ordinary least squares and indirectly by using your structural estimates from part b.

The relevant submatrices are $\mathbf{X}'\mathbf{X} = \begin{bmatrix} 5 & 2 & 3 \\ 2 & 10 & 8 \\ 3 & 8 & 15 \end{bmatrix}$, $\mathbf{X}'\mathbf{y}_1 = \begin{bmatrix} 4 \\ 3 \\ 5 \end{bmatrix}$, $\mathbf{X}'\mathbf{y}_2 = \begin{bmatrix} 3 \\ 6 \\ 7 \end{bmatrix}$, $\mathbf{y_1}'\mathbf{y_1} = 20$, $\mathbf{y}_2'\mathbf{y}_2 = 10$, $\mathbf{y}_1'\mathbf{y}_2 = 6$, $\mathbf{X}'\mathbf{Z}_1 = \begin{bmatrix} 3 & 5 \\ 6 & 2 \\ 7 & 3 \end{bmatrix}$, $\mathbf{X}'\mathbf{Z}_2 = \begin{bmatrix} 4 & 2 & 3 \\ 3 & 10 & 8 \\ 5 & 8 & 15 \end{bmatrix}$ $\mathbf{Z}_1'\mathbf{Z}_1 = \begin{bmatrix} 10 & 3 \\ 3 & 5 \end{bmatrix}$, $\mathbf{Z}_2'\mathbf{Z}_2 = \begin{bmatrix} 10 & 3 & 5 \\ 3 & 10 & 8 \\ 5 & 8 & 15 \end{bmatrix}$,

$\mathbf{Z}_1'\mathbf{Z}_2 = \begin{bmatrix} 6 & 6 & 7 \\ 4 & 2 & 3 \end{bmatrix}$, $\mathbf{Z}_1'\mathbf{y}_1 = \begin{bmatrix} 6 \\ 4 \end{bmatrix}$, $\mathbf{Z}_1'\mathbf{y}_2 = \begin{bmatrix} 10 \\ 3 \end{bmatrix}$, $\mathbf{Z}_2'\mathbf{y}_1 = \begin{bmatrix} 20 \\ 3 \\ 5 \end{bmatrix}$, $\mathbf{Z}_2'\mathbf{y}_2 = \begin{bmatrix} 6 \\ 6 \\ 7 \end{bmatrix}$.

The two OLS coefficient vectors are

$$\mathbf{d}_1 = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_1 = [.439024,.536585]'$$
$$\mathbf{d}_2 = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_2 = [.193016,.384127,.19746]'.$$

The two stage least squares estimators are

$$\hat{\delta}_1 = [\mathbf{Z}_1'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}_1]^{-1}[\mathbf{Z}_1'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_1] = [.368816,.578711]'.$$

$$\hat{\delta}_2 = [\mathbf{Z}_2'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}_2]^{-1}[\mathbf{Z}_2'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_2] = [.484375,.367188,.109375]'.$$

$$\hat{\sigma}_{11} = (\mathbf{y}_1'\mathbf{y}_1 - 2\mathbf{y}_1'\mathbf{Z}\hat{\delta}_1 + \hat{\delta}_1'\mathbf{Z}_1'\mathbf{Z}_1\hat{\delta}_1) / 25 = .610397, \quad \hat{\sigma}_{22} = .268384.$$

The estimated asymptotic covariance matrices are

$$\text{Est.Var}[\hat{\delta}_1] = \hat{\sigma}_{11}[\mathbf{Z}_1'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}_1]^{-1} = \begin{bmatrix} .215858 & .129035 \\ .129036 & .1995 \end{bmatrix}$$

$$\text{Est.Var}[\text{Est.Var}[\hat{\delta}_2]] = \begin{bmatrix} .132423 & -.007699 & -.040035 \\ -.007688 & .047259 & -.022538 \\ -.040035 & -.022638 & .043311 \end{bmatrix}.$$

The three stage least squares estimate is

$$\begin{bmatrix} \hat{\sigma}^{11}[\mathbf{Z}_1'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}_1] & \hat{\sigma}^{12}[\mathbf{Z}_1'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}_2] \\ \hat{\sigma}^{12}[\mathbf{Z}_2'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}_1] & \hat{\sigma}^{22}[\mathbf{Z}_2'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}_2] \end{bmatrix}^{-1} \begin{bmatrix} \hat{\sigma}^{11}[\mathbf{Z}_1'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_1]+ \\ \hat{\sigma}^{12}[\mathbf{Z}_1'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_2] \\ \\ \hat{\sigma}^{12}[\mathbf{Z}_2'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_1]+ \\ \hat{\sigma}^{22}[\mathbf{Z}_2'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}_2] \end{bmatrix}$$

$$= [.368817,.578708,.4706,.306363,.168294]'.$$

The estimated standard errors are the square roots of the diagonal elements of the inverse matrix, [.4637,.4466,.3626,.1716,.1628], compared to the 2SLS values, [.4637,.4466,.3639,.2174,.2081].

To compute the limited information maximum likelihood estimator, we require the matrix of sums of squares and cross products of residuals of the regressions of $\mathbf{y}_1$ and $\mathbf{y}_2$ on $\mathbf{x}_1$ and on $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$. These are

$$\mathbf{W}^0 = \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{x}_1(\mathbf{x}_1'\mathbf{x}_1)^{-1}\mathbf{x}_1'\mathbf{Y} = \begin{bmatrix} 16.5 & 3.60 \\ 3.60 & 8.20 \end{bmatrix}, \mathbf{W}^1 = \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \begin{bmatrix} 16.2872 & 2.55312 \\ 2.55312 & 5.3617 \end{bmatrix}.$$

The two characteristic roots of $(\mathbf{W}^1)^{-1}\mathbf{W}^0$ are 1.53157 and 1.00837. We carry the smaller one into the *k*-class computation [see, for example, Theil (1971) or Judge, et al (1985)];

$$\hat{\delta}_{1k} = \begin{bmatrix} 10-1.00837(5.3617) & 3 \\ 3 & 5 \end{bmatrix}^{-1} \begin{bmatrix} 6-1.00837(2.55312) \\ 4 \end{bmatrix} = \begin{bmatrix} .367116 \\ .57973 \end{bmatrix}$$

Finally, the two estimates of the reduced form are

$$\text{(OLS)} \quad \mathbf{P} = \begin{bmatrix} .680851 & .329787 \\ .010638 & .37243 \\ .191489 & .202128 \end{bmatrix}$$

$$\text{and} \quad \text{(2SLS)} \quad \hat{\Pi} = \begin{bmatrix} -.578711 & 0 \\ 0 & -.367188 \\ 0 & -.109375 \end{bmatrix} \begin{bmatrix} 1 & -.484375 \\ -.368816 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} .704581 & .341281 \\ .104880 & .447051 \\ .049113 & .133164 \end{bmatrix}.$$

6. For the model

$$y_1 = \gamma_1 y_2 + \beta_{11}x_1 + \beta_{21}x_2 + \varepsilon_1$$
$$y_2 = \gamma_2 y_1 + \beta_{32}x_3 + \beta_{42}x_4 + \varepsilon_2$$

show that there are two restrictions on the reduced form coefficients. Describe a procedure for estimating the model while incorporating the restrictions.

$$\text{The structure is } [y_1\, y_2] \begin{bmatrix} 1 & -\gamma_2 \\ -\gamma_1 & 1 \end{bmatrix} + [x_1\ x_2\ x_3\ x_4] \begin{bmatrix} \beta_{11} & 0 \\ \beta_{21} & 0 \\ 0 & \beta_{32} \\ 0 & \beta_{42} \end{bmatrix} = [\varepsilon_1\ \ \varepsilon_1].$$

or $\mathbf{y}'\,\boldsymbol{\Gamma} + \mathbf{x}'\mathbf{B} = \varepsilon'$. The reduced form coefficient matrix is

$$\boldsymbol{\Pi} = -\mathbf{B}\boldsymbol{\Gamma}^{-1} = \frac{1}{1-\gamma_1\gamma_2} \begin{bmatrix} \beta_{11} & \gamma_2\beta_{11} \\ \beta_{21} & \gamma_2\beta_{21} \\ \gamma_1\beta_{32} & \beta_{32} \\ \gamma_1\beta_{42} & \beta_{42} \end{bmatrix} = \begin{bmatrix} \pi_{11} & \pi_{21} \\ \pi_{21} & \pi_{22} \\ \pi_{31} & \pi_{32} \\ \pi_{41} & \pi_{42} \end{bmatrix}$$

The two restrictions are $\pi_{12}/\pi_{11} = \pi_{22}/\pi_{21}$ and $\pi_{31}/\pi_{32} = \pi_{41}/\pi_{42}$. If we write the reduced form as

$$y_1 = \pi_{11}x_1 + \pi_{21}x_2 + \pi_{31}x_3 + \pi_{41}x_4 + v_1$$
$$y_2 = \pi_{12}x_1 + \pi_{22}x_2 + \pi_{32}x_3 + \pi_{42}x_4 + v_2.$$

We could treat the system as a nonlinear seemingly unrelated regressions model. One possible way to handle the restrictions is to eliminate two parameters directly by making the substitutions

$$\pi_{12} = \pi_{11}\pi_{22}/\pi_{21} \quad \text{and} \quad \pi_{31} = \pi_{32}\pi_{41}/\pi_{42}.$$

The pair of equations would be

$$y_1 = \pi_{11}x_1 + \pi_{21}x_2 + (\pi_{32}\pi_{41}/\pi_{42})x_3 + \pi_{41}x_4 + v_1$$
$$y_2 = (\pi_{11}\pi_{22}/\pi_{21})x_1 + \pi_{22}x_2 + \pi_{32}x_3 + \pi_{42}x_4 + v_2.$$

This nonlinear system could now be estimated by nonlinear GLS. The function to be minimized would be

$$\Sigma_{i=1}^{n}\ v_{i1}^2\sigma^{11} + v_{i2}^2\sigma^{22} + 2v_{i1}v_{i2}\sigma^{12} = n\text{tr}(\Sigma^{-1}\mathbf{W}).$$

Needless to say, this would be quite involved.

7. An updated version of Klein's Model I was estimated. Using the two stage least squares estimates, the relevant submatrix of Δ (see (15-42)) is $\Delta_1 = \begin{bmatrix} -.1899 & -.9471 & -.8991 \\ 0 & 1.0287 & 0 \\ -.0656 & -.0791 & .0952 \end{bmatrix}$. Is the model stable?

We would require that all three characteristic roots have modulus less than one. An intuitive guess that the diagonal element greater than one would preclude this would be correct. The roots are the solutions to $det\begin{bmatrix} -.1899-\lambda & -.9471 & -.8991 \\ 0 & 1.0287-\lambda & 0 \\ -.0656 & -.0791 & .0952-\lambda \end{bmatrix} = 0$. Expanding this produces $-(.1899 + \lambda)(1.0287 - \lambda)(.0952 - \lambda) - .0565(1.0287 - \lambda).8991 = 0$. There is no need to go any further. It is obvious that $\lambda = 1.0287$ is a solution, so there is at least one characteristic root larger than 1. The system is unstable.

8. Prove plim $\mathbf{Y}_j'\varepsilon/T = \boldsymbol{\omega}_j - \boldsymbol{\Omega}_{jj}\gamma_j$.

Consistent with the partitioning $\mathbf{y}' = [y_j\ \mathbf{Y}_j'\ \mathbf{Y}_i^{*\prime}]$, partition $\boldsymbol{\Omega}$ into

$$\boldsymbol{\Omega} = \begin{matrix} \omega_{jj} & \omega_j' & \omega^{*\prime}_j \\ \omega_j & \Omega_{jj} & \Omega_j' \\ \omega^*_j & \Omega^*_j & \Omega_j^* \end{matrix}$$

and, as in the equation preceding (15-8), partition the jth column of $\boldsymbol{\Gamma}$ as $\boldsymbol{\Gamma}_j = \begin{bmatrix} 1 \\ -\gamma \\ 0 \end{bmatrix}$. Since the full set of reduced form disturbances is $\mathbf{V} = \mathbf{E}\boldsymbol{\Gamma}^{-1}$, it follows that $\mathbf{E} = \mathbf{V}\boldsymbol{\Gamma}$. In particular, the jth column of $\mathbf{E}$ is $\boldsymbol{\varepsilon}_j = \mathbf{V}\boldsymbol{\Gamma}_j$. In the reduced form, now referring to (15-8), $\mathbf{Y}_j = \mathbf{X}\boldsymbol{\Pi}_j + \mathbf{V}_j$, where $\boldsymbol{\Pi}_j$ is the M_j columns of $\boldsymbol{\Pi}$ corresponding to the included endogenous variables and $\mathbf{V}_j$ is the $T\times M_j$ matrix of their reduced form disturbances. Since $\mathbf{X}$ is uncorrelated with all columns of $\mathbf{E}$, we have

$$\text{plim } \mathbf{Y}_j'\varepsilon_j/T = \text{plim } \mathbf{V}_j'\boldsymbol{\Gamma}_j/T = [\boldsymbol{\omega}_j\ \ \boldsymbol{\Omega}_{jj}\ \ \boldsymbol{\Omega}_j^*]\begin{bmatrix} 1 \\ -\gamma \\ 0 \end{bmatrix} = \boldsymbol{\omega}_j - \boldsymbol{\Omega}_{jj}\gamma_j \text{ as required.}$$ □

9. Prove that an underidentified equation cannot be estimated by two stage least squares.

If the equation fails the order condition, then the number of excluded exogenous variables is less than the number of included endogenous. The matrix of instrumental variables to be used for two stage least squares is of the form $\hat{\mathbf{Z}} = [\mathbf{XA},\mathbf{X}_j]$, where $\mathbf{XA}$ is M_j linear combination of all K columns in $\mathbf{X}$ and $\mathbf{X}_j$ is K_j columns of $\mathbf{X}$. In total, $K = K_j^* + K_j$. If the equation fails the order condition, then $K_j^* < M_j$, so $\hat{\mathbf{Z}}$ is $M_j + K_j$ columns which are linear combinations of $K = K_j^* + K_j < M_j + K_j$. Therefore, $\hat{\mathbf{Z}}$ cannot have full column rank. In order to compute the two stage least squares estimator, we require $(\hat{\mathbf{Z}}'\hat{\mathbf{Z}})^{-1}$, which cannot be computed.

Chapter 16

Estimation Frameworks in Econometrics

1. Compare the fully parametric and semiparametric approaches to estimation of a discrete choice model such as the multinomial logit model discussed in Chapter 21. What are the benefits and costs of the semiparametric approach?

A fully parametric model/estimator provides consistent, efficient, and comparatively precise results. The semiparametric model/estimator, by comparison, is relatively less precise in general terms. But, the payoff to this imprecision is that the semiparametric formulation is more likely to be robust to failures of the assumptions of the parametric model. Consider, for example, the binary probit model of Chapter 21, which makes a strong assumption of normality and homoscedasticity. If the assumptions are correct, the probit estimator is the most efficient use of the data. However, if the normality assumption or the homoscedasticity assumption are incorrect, then the probit estimator becomes inconsistent in an unknown fashion. Lewbel's semiparametric estimator for the binary choice model, in contrast, is not very precise in comparison to the probit model. But, it will remain consistent if the normality assumption is violated, and it is even robust to certain kinds of heteroscedasticity.

2. Asymptotics take on a different meaning in the Bayesian estimation context, since parameter estimators do not "converge" to a population quantity. Nonetheless, in a Bayesian estimation setting, as the sample size increases, the likelihood function will dominate the posterior density. What does this imply about the Bayesian "estimator" when this occurs.

The Bayesian estimator must "converge" to the maximum likelihood estimator as the sample size grows. The posterior mean will generally be a mixture of the prior and the maximizer of the likelihood function. We do note, however, that the likelihood will only dominate an informative prior asymptotically - the Bayesian estimator in this case will ultimately be a mixture of a prior with a finite precision and a likelihood based estimator whose variance converges to zero (thus, whose precision grows infinitely). Thus, the domination will not be complete in a finite sample.

3. Referring to the situation in question 2, one might think that an informative prior would outweigh the effect of the increasing sample size. With respect to the Bayesian analysis of the linear regression, analyze the way in which the likelihood and an informative prior will compete for dominance in the posterior mean.

The Bayesian estimator with an informative prior in (16-10) is

$$E[\beta|\text{data},\sigma^2] = \mathbf{F}\beta_0 + (\mathbf{I}-\mathbf{F})\mathbf{b}$$

where β_0 is the prior mean, **b** is the least squares estimator and $\mathbf{F} = [\boldsymbol{\Sigma}_0^{-1} + [\sigma^{-2}(\mathbf{X}'\mathbf{X})^{-1}]^{-1}]^{-1}\boldsymbol{\Sigma}_0^{-1}$ where $\boldsymbol{\Sigma}_0$ is the prior variance. Now, with well behaved data, **F** must ultimately converge to a zero matrix because the OLS estimator's variance is shrinking, so it's inverse is increasing inside the large square brackets.

The following exercises require specific software. The relevant techniques are available in several packages that might be in use, such as SAS, Stata, or LIMDEP. The exercises are suggested as departure points for explorations using a few of the many estimation techniques listed in this chapter.

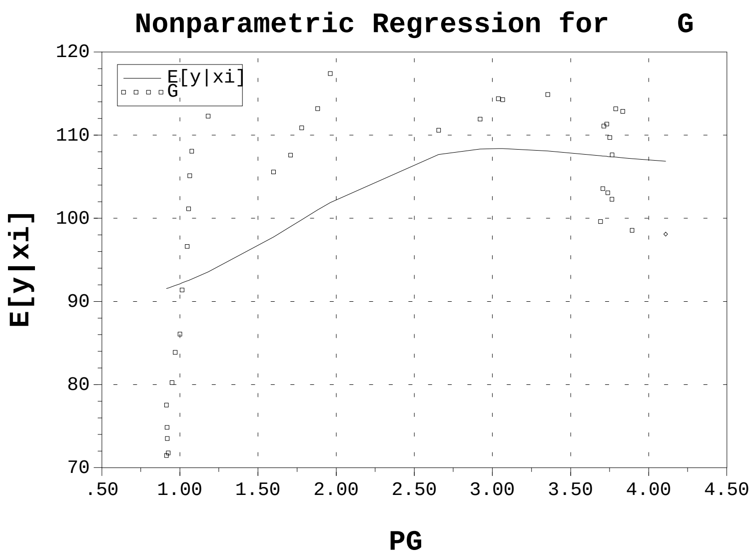
4. Using the gasoline market data in Appendix Table F2.2, use the partially linear regression method in Section 16.3.3 to fit an equation of the form

$$\ln(G/Pop) = \beta_1\ln(Income) + \beta_2\ln P_{new\ cars} + \beta_3\ln P_{used\ cars} + g(\ln P_{gasoline}) + \varepsilon$$

```
crea;gp=lg;ip=ly;ncp=lpnc;upp=lpuc;pgp=lpg$
sort;lhs=pgp;rhs=gp,ip,ncp,upp$
crea;dgp=.809*gp - .5*gp[-1] - .309*gp[-2]$
crea;dip=.809*ip - .5*ip[-1] - .309*ip[-2]$
crea;dnc=.809*ncp -.5*ncp[-1]-.309*ncp[-2]$
crea;duc=.809*upp -.5*upp[-1]-.309*upp[-2]$
samp;3-36$
regr;lhs=dgp;rhs=dip,dnc,duc;res=e$
+-----------------------------------------------------------------------+
| Ordinary    least squares regression    Weighting variable = none     |
| Dep. var. = DGP      Mean=   .9708646870E-02, S.D.=   .4738748109E-01 |
| Model size: Observations =      34, Parameters =   3, Deg.Fr.=     31 |
| Residuals:  Sum of squares= .1485994289E-01, Std.Dev.=         .02189 |
| Fit:        R-squared=  .799472, Adjusted R-squared =          .78653 |
| Model test: F[  2,     31] =   61.80,    Prob value =          .00000 |
| Diagnostic: Log-L =     83.2587, Restricted(b=0) Log-L =      55.9431 |
|             LogAmemiyaPrCrt.=   -7.559, Akaike Info. Crt.=     -4.721 |
| Model does not contain ONE. R-squared and F can be negative!          |
| Autocorrel: Durbin-Watson Statistic =   1.34659,   Rho =       .32671 |
+-----------------------------------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |t-ratio |P[|T|>t] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 DIP          .9629902959       .11631885    8.279   .0000  .14504254E-01
 DNC         -.1010972781   .87755182E-01   -1.152   .2581  .20153536E-01
 DUC      -.3197058148E-01  .51875022E-01    -.616   .5422  .35656776E-01
--> matr;varpl={1+1/(2*2)}*varb$
--> matr;stat(b,varpl)$
+-----------------------------------------------+
|Number of observations in current sample =   34 |
|Number of parameters computed here       =    3 |
|Number of degrees of freedom             =   31 |
+-----------------------------------------------+
+---------+--------------+----------------+--------+---------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] |
+---------+--------------+----------------+--------+---------+
 B_1          .9629902959       .13004843    7.405   .0000
 B_2         -.1010972781   .98113277E-01   -1.030   .3028
 B_3      -.3197058148E-01  .57998037E-01    -.551   .5815
```

5. To continue the analysis in question 4, consider a nonparametric regression of G/Pop on the price. Using the nonparametric estimation method in Section 16.4.2, fit the nonparametric estimator using a range of bandwidth values to explore the effect of bandwidth

```
+---------------------------------------+
| Nonparametric Regression for G        |
| Observations       =           36     |
| Points plotted     =           36     |
| Bandwidth          =       .468092    |
| Statistics for abscissa values----    |
| Mean               =      2.316611    |
| Standard Deviation =      1.251735    |
| Minimum            =       .914000    |
| Maximum            =      4.109000    |
| ----------------------------------    |
| Kernel Function    =      Logistic    |
| Cross val. M.S.E.  =    121.084982    |
| Results matrix     =        KERNEL    |
+---------------------------------------+
```



6. (You might find it useful to read the early sections of Chapter 21 for this exercise.) The extramarital affairs data analyzed in Section 22.3.7 can be reinterpreted in the context of a binary choice model. The dependent variable in the analysis is a count of events. Using these data, first recode the dependent variable 0 for none and 1 for more than zero. Now, first using the binary probit estimator, fit a binary choice model using the same independent variables as in the example discussed in Section 22.3.7. Then using a semiparametric or nonparametric estimator, estimate the same binary choice model. A model for binary choice can be fit for at least two purposes, for estimation of interesting coefficients or for prediction of the dependent variable. Use your estimated models for these two purposes and compare the two models.
A. Using the probit model and the Klein and Spady semiparametric models, the two sets of coefficient estimates are somewhat similar.

```
+---------------------------------------------+
| Binomial Probit Model                       |
| Maximum Likelihood Estimates                |
| Model estimated: Jul 31, 2002 at 05:16:40PM.|
| Dependent variable                   P      |
| Weighting variable                None      |
| Number of observations             601      |
| Iterations completed                 5      |
| Log likelihood function      -307.2955      |
| Restricted log likelihood    -337.6885      |
| Chi squared                   60.78608      |
| Degrees of freedom                   5      |
| Prob[ChiSqd > value] =        .0000000      |
| Hosmer-Lemeshow chi-squared =  5.74742      |
| P-value=  .67550 with deg.fr. =      8      |
+---------------------------------------------+
+---------+-------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+-------------+----------------+--------+---------+----------+
         Index function for probability
 Z2      -.2202376072E-01  .10177371E-01   -2.164   .0305     32.487521
 Z3       .5990084920E-01  .17086004E-01    3.506   .0005     8.1776955
 Z5        -.1836462412    .51493239E-01   -3.566   .0004     3.1164725
 Z7       .3751312008E-01  .32844576E-01    1.142   .2534     4.1946755
 Z8        -.2729824396    .52473295E-01   -5.202   .0000     3.9317804
 Constant     .9766647244      .36104809    2.705   .0068
+---------------------------------------------+
| Seimparametric Binary Choice Model          |
| Maximum Likelihood Estimates                |
| Model estimated: Jul 31, 2002 at 11:01:24PM.|
| Dependent variable                   P      |
| Weighting variable                None      |
| Number of observations             601      |
| Iterations completed                13      |
| Log likelihood function      -334.7367      |
| Restricted log likelihood    -337.6885      |
| Chi squared                   5.903551      |
| Degrees of freedom                   4      |
| Prob[ChiSqd > value] =        .2064679      |
| Hosmer-Lemeshow chi-squared = 118.69649     |
| P-value=  .00000 with deg.fr. =      8      |
| Logistic kernel fn. Bandwidth =  .34423     |
+---------------------------------------------+
+---------+-------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+-------------+----------------+--------+---------+----------+
         Characteristics in numerator of Prob[Y = 1]
 Z2      -.3284308221E-01  .52254249E-01    -.629   .5297     32.487521
 Z3         .1089817386    .86483083E-01    1.260   .2076     8.1776955
 Z5        -.2384951835        .23320058   -1.023   .3064     3.1164725
 Z7        -.1026067037        .17130225    -.599   .5492     4.1946755
 Z8        -.1892263132        .21598982    -.876   .3810     3.9317804
 Constant     .0000000000 ........(Fixed Parameter)........
```

The probit model produces a set of marginal effects, as discussed in the text. These cannot be computed for the Klein and Spady estimator.

```
+-------------------------------------------+
| Partial derivatives of E[y] = F[*]   with |
| respect to the vector of characteristics. |
| They are computed at the means of the Xs. |
| Observations used for means are All Obs.  |
+-------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
          Index function for probability
 Z2       -.6695300413E-02  .30909282E-02   -2.166   .0303     32.487521
 Z3        .1821006800E-01  .51704684E-02    3.522   .0004     8.1776955
 Z5       -.5582910069E-01  .15568275E-01   -3.586   .0003     3.1164725
 Z7        .1140411992E-01  .99845393E-02    1.142   .2534     4.1946755
 Z8       -.8298761795E-01  .15933104E-01   -5.209   .0000     3.9317804
 Constant     .2969094977       .11108860    2.673   .0075
```

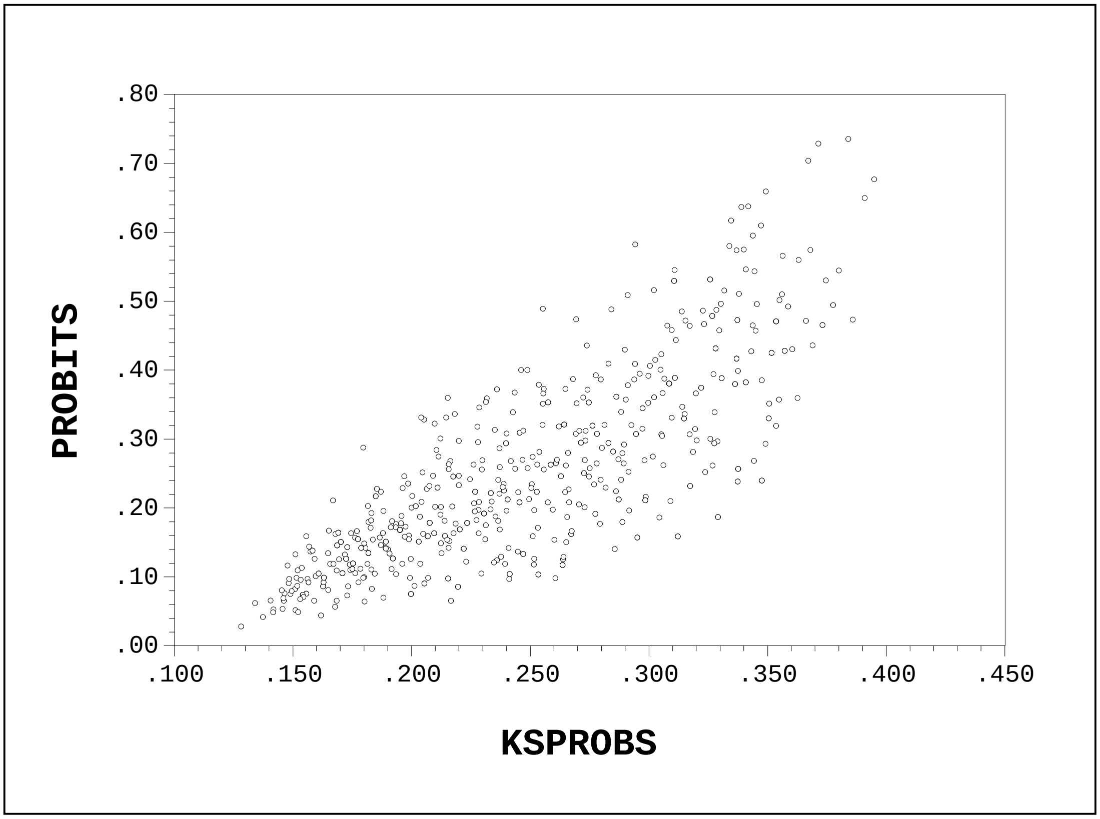
These are the various fit measures for the probit model

```
+----------------------------------------+
| Fit Measures for Binomial Choice Model |
| Probit   model for variable P          |
+----------------------------------------+
| Proportions P0= .750416   P1= .249584  |
| N =     601 N0=     451   N1=     150  |
| LogL =  -307.29545 LogL0 =  -337.6885  |
| Estrella = 1-(L/L0)^(-2L0/n) = .10056  |
+----------------------------------------+
|     Efron |  McFadden  |  Ben./Lerman  |
|    .10905 |    .09000  |       .66451  |
|    Cramer | Veall/Zim. |     Rsqrd_ML  |
|    .10486 |    .17359  |       .09619  |
+----------------------------------------+
| Information  Akaike I.C. Schwarz I.C.  |
| Criteria        1.04258    652.98248   |
+----------------------------------------+
Frequencies of actual & predicted outcomes
Predicted outcome has maximum probability.
Threshold value for predicting Y=1 = .5000
            Predicted
------  ----------  +  -----
Actual       0    1  |  Total
------  ----------  +  -----
  0        437   14  |    451
  1        130   20  |    150
------  ----------  +  -----
Total      567   34  |    601
```

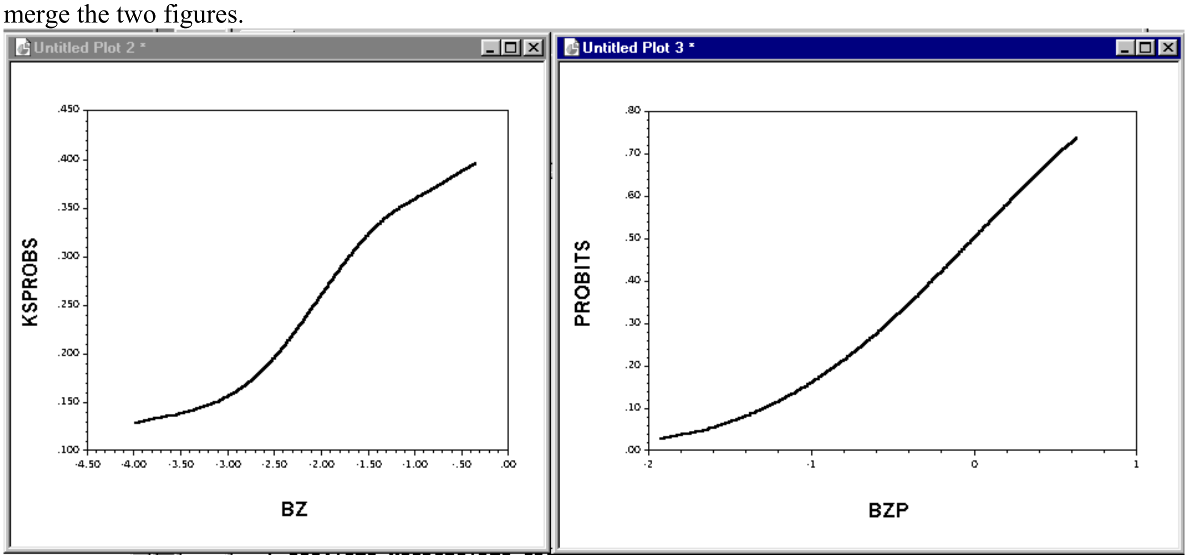
These are the fit measures for the probabilities computed for the Klein and Spady model. The probit model fits better by all measures computed.

```
+----------------------------------------+
| Fit Measures for Binomial Choice Model |
| Observed = P        Fitted = KSPROBS   |
+----------------------------------------+
| Proportions P0= .750416   P1= .249584  |
| N =     601 N0=     451   N1=     150  |
| LogL =  -320.37513 LogL0 =  -337.6885  |
| Estrella = 1-(L/L0)^(-2L0/n) = .05743  |
+----------------------------------------+
|     Efron |  McFadden  |  Ben./Lerman  |
|    .05686 |    .05127  |       .64117  |
|    Cramer | Veall/Zim. |     Rsqrd_ML  |
|    .03897 |    .10295  |       .05599  |
+----------------------------------------+
```

The first figure below plots the probit probabilities against the Klein and Spady probabilities. The models are obviously similar, though there is substantial difference in the fitted values.



Finally, these two figures plot the predicted probabilities from the two models against the respective index functions, b'x. Note that the two plots are based on different coefficient vectors, so it is not possible to merge the two figures.



Chapter 17

Maximum Likelihood Estimation

1. Assume the distribution of x is $f(x) = 1/\theta, 0 \le x \le \theta$. In random sampling from this distribution, prove that the sample maximum is a consistent estimator of θ. Note: you can prove that the maximum is the maximum likelihood estimator of θ. But, the usual properties do not apply here. Why not? (**Hint:** Attempt to verify that the expected first derivative of the log-likelihood with respect to θ is zero.)

Using the result of the previous problem, the density of the maximum is

$$n[z/\theta]^{n-1}(1/\theta), \quad 0 < z < \theta.$$

Therefore, the expected value is $E[z] = \int_0^\theta z^n dz = [\theta^{n+1}/(n+1)][n/\theta^n] = n\theta/(n+1)$. The variance is found likewise, $E[z^2] = \int_0^\theta z^2 n(z/n)^{n-1}(1/\theta)dz = n\theta^2/(n+2)$ so $\text{Var}[z] = E[z^2] - (E[z])^2 = n\theta^2/[(n+1)^2(n+2)]$. Using mean squared convergence we see that $\lim_{n\to\infty} E[z] = \theta$ and $\lim_{n\to\infty} \text{Var}[z] = 0$, so that plim $z = \theta$.

2. In random sampling from the exponential distribution, $f(x) = \frac{1}{\theta}e^{\frac{-x}{\theta}}$, $x > 0$, $\theta > 0$, find the maximum likelihood estimator of θ and obtain the asymptotic distribution of this estimator.

The log-likelihood is $\ln L = -n\ln\theta - (1/\theta)\sum_{i=1}^n x_i$. The maximum likelihood estimator is obtained as the solution to $\partial\ln L/\partial\theta = -n/\theta + (1/\theta^2)\sum_{i=1}^n x_i = 0$, or $\hat{\theta}_{ML} = (1/n)\sum_{i=1}^n x_i = \bar{x}$. The asymptotic variance of the MLE is $\{-E[\partial^2\ln L/\partial\theta^2]\}^{-1} = \{-E[n/\theta^2 - (2/\theta^3)\sum_{i=1}^n x_i]\}^{-1}$. To find the expected value of this random variable, we need $E[x_i] = \theta$. Therefore, the asymptotic variance is θ^2/n. The asymptotic distribution is normal with mean θ and this variance.

3. Suppose the joint distribution of the two random variables x and y is

$$f(x,y) = \theta e^{-(\beta+\theta)y}(\beta y)^x / x! \quad \beta,\theta\ \ 0,\ y \$ 0, x = 0,1,2,...$$

(a) Find the maximum likelihood estimators of β and θ and their asymptotic joint distribution.
(b) Find the maximum likelihood estimator of θ/(β+θ) and its asymptotic distribution.
(c) Prove that $f(x)$ is of the form $f(x) = \gamma(1-\gamma)^x$, $x = 0,1,2,...$
Then, find the maximum likelihood estimator of γ and its asymptotic distribution.
(d) Prove that f(y*x) is of the form $\lambda e^{-\lambda y}(\lambda y)^x/x!$ Prove that $f(y|x)$ integrates to 1. Find the maximum likelihood estimator of λ and its asymptotic distribution. (**Hint**: In the conditional distribution, just carry the xs along as constants.)
(e) Prove that $f(y) = \theta e^{-\theta y}$ then find the maximum likelihood estimator of θ and its asymptotic variance.
(f) Prove that $f(x|y) = e^{-\beta y}(\beta y)^x/x!$. Based on this distribution, what is the maximum likelihood estimator of β?

The log-likelihood is $\ln L = n\ln\theta - (\beta+\theta)\sum_{i=1}^n y_i + \ln\beta\sum_{i=1}^n x_i + \sum_{i=1}^n x_i \log y_i - \sum_{i=1}^n \log(x_i!)$

The first and second derivatives are

$$\partial\ln L/\partial\theta = n/\theta - \sum_{i=1}^n y_i$$
$$\partial\ln L/\partial\beta = -\sum_{i=1}^n y_i + \sum_{i=1}^n x_i/\beta$$
$$\partial^2\ln L/\partial\theta^2 = -n/\theta^2$$
$$\partial^2\ln L/\partial\beta^2 = -\sum_{i=1}^n x_i/\beta^2$$
$$\partial^2\ln L/\partial\beta\partial\theta = 0.$$

Therefore, the maximum likelihood estimators are $\hat{\theta} = 1/\bar{y}$ and $\hat{\beta} = \bar{x}/\bar{y}$ and the asymptotic covariance matrix is the inverse of $E\begin{bmatrix} n/\theta^2 & 0 \\ 0 & \sum_{i=1}^n x_i/\beta^2 \end{bmatrix}$. In order to complete the derivation, we will require the expected value of $\sum_{i=1}^n x_i = nE[x_i]$. In order to obtain $E[x_i]$, it is necessary to obtain the marginal distribution of x_i, which is f(x) = $\int_0^\infty \theta e^{-(\beta+\theta)y}(\beta y)^x / x!\,dy = \beta^x(\theta/x!)\int_0^\infty e^{-(\beta+\theta)y} y^x dy$. This is $\beta^x(\theta/x!)$ times a gamma integral. This is $f(x) = \beta^x(\theta/x!)[\Gamma(x+1)]/(\beta+\theta)^{x+1}$. But, $\Gamma(x+1) = x!$, so the expression reduces to

$$f(x) = [\theta/(\beta+\theta)][\beta/(\beta+\theta)]^x.$$

Thus, x has a geometric distribution with parameter $\pi = \theta/(\beta+\theta)$. (This is the distribution of the number of tries until the first success of independent trials each with success probability $1-\pi$. Finally, we require the expected value of x_i, which is $E[x] = [\theta/(\beta+\theta)]\sum_{x=0}^\infty x[\beta/(\beta+\theta)]^x = \beta/\theta$. Then, the required asymptotic covariance matrix is $\begin{bmatrix} n/\theta^2 & 0 \\ 0 & n(\beta/\theta)/\beta^2 \end{bmatrix}^{-1} = \begin{bmatrix} \theta^2/n & 0 \\ 0 & \beta\theta/n \end{bmatrix}$.

The maximum likelihood estimator of $\theta/(\beta+\theta)$ is is

$$\widehat{\theta/(\beta+\theta)} = (1/\bar{y})/[\bar{x}/\bar{y} + 1/\bar{y}] = 1/(1+\bar{x}).$$

Its asymptotic variance is obtained using the variance of a nonlinear function

$$V = [\beta/(\beta+\theta)]^2(\theta^2/n) + [-\theta/(\beta+\theta)]^2(\beta\theta/n) = \beta\theta^2/[n(\beta+\theta)^3].$$

The asymptotic variance could also be obtained as $[-1/(1+E[x])^2]^2$Asy.Var$[\bar{x}]$.)

For part (c), we just note that $\gamma = \theta/(\beta+\theta)$. For a sample of observations on x, the log-likelihood would be

$$\ln L = n\ln\gamma + \ln(1-\gamma)\sum_{i=1}^n x_i$$

$$\partial\ln L/d\gamma = n/\gamma - \sum_{i=1}^n x_i/(1-\gamma).$$

A solution is obtained by first noting that at the solution, $(1-\gamma)/\gamma = \bar{x} = 1/\gamma - 1$. The solution for γ is, thus, $\hat{\gamma} = 1/(1+\bar{x})$.Of course, this is what we found in part b., which makes sense.

For part (d) $f(y|x) = \dfrac{f(x,y)}{f(x)} = \dfrac{\theta e^{-(\beta+\theta)y}(\beta y)^x(\beta+\theta)^x(\beta+\theta)}{x!\ \ \theta\ \ \beta x}$. Cancelling terms and gathering the remaining like terms leaves $f(y|x) = (\beta+\theta)[(\beta+\theta)y]^x e^{-(\beta+\theta)y}/x!$ so the density has the required form with $\lambda = (\beta+\theta)$. The integral is $\left\{[\lambda^{x+1}]/x!\right\}\int_0^\infty e^{-\lambda y}y^x dy$. This integral is a Gamma integral which equals $\Gamma(x+1)/\lambda^{x+1}$, which is the reciprocal of the leading scalar, so the product is 1. The log-likelihood function is

$$\ln L = n\ln\lambda - \lambda\sum_{i=1}^n y_i + \ln\lambda\sum_{i=1}^n x_i - \sum_{i=1}^n \ln x_i!$$

$$\partial\ln L/\partial\lambda = (\sum_{i=1}^n x_i + n)/\lambda - \sum_{i=1}^n y_i.$$

$$\partial^2\ln L/\partial\lambda^2 = -(\sum_{i=1}^n x_i + n)/\lambda^2.$$

Therefore, the maximum likelihood estimator of λ is $(1+\bar{x})/\bar{y}$ and the asymptotic variance, conditional on the xs is Asy.Var.$\left[\hat{\lambda}\right] = (\lambda^2/n)/(1+\bar{x})$

Part (e.) We can obtain $f(y)$ by summing over x in the joint density. First, we write the joint density as $f(x,y) = \theta e^{-\theta y}e^{-\beta y}(\beta y)^x/x!$. The sum is, therefore, $f(y) = \theta e^{-\theta y}\sum_{x=0}^\infty e^{-\beta y}(\beta y)^x/x!$. The sum is that of the probabilities for a Poisson distribution, so it equals 1. This produces the required result. The maximum likelihood estimator of θ and its asymptotic variance are derived from

$$\ln L = n\ln\theta - \theta\sum_{i=1}^{n} y_i$$

$$\partial\ln L/\partial\theta = n/\theta - \sum_{i=1}^{n} y_i$$

$$\partial^2\ln L/\partial\theta^2 = -n/\theta^2.$$

Therefore, the maximum likelihood estimator is $1/\bar{y}$ and its asymptotic variance is θ^2/n. Since we found $f(y)$ by factoring $f(x,y)$ into $f(y)f(x|y)$ (apparently, given our result), the answer follows immediately. Just divide the expression used in part e. by $f(y)$. This is a Poisson distribution with parameter βy. The log-likelihood function and its first derivative are

$$\ln L = -\beta\sum_{i=1}^{n} y_i + \ln\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} x_i \ln y_i - \sum_{i=1}^{n} \ln x_i\,!$$

$$\partial\ln L/\partial\beta = -\sum_{i=1}^{n} y_i + \sum_{i=1}^{n} x_i/\beta,$$

from which it follows that $\hat{\beta} = \bar{x}/\bar{y}$.

4. Suppose x has the Weibull distribution, $f(x) = \alpha\beta x^{\beta-1}\exp(-\alpha x^{\beta})$, $x, \alpha, \beta > 0$.

(a) Obtain the log-likelihood function for a random sample of n observations.

(b) Obtain the likelihood equations for maximum likelihood estimation of α and β. Note that the first provides an explicit solution for α in terms of the data and β. But, after inserting this in the second, we obtain only an implicit solution for β. How would you obtain the maximum likelihood estimators?

(c) Obtain the second derivatives matrix of the log-likelihood with respect to α and β. The exact expectations of the elements involving β involve the derivatives of the Gamma function and are quite messy analytically. Of course, your exact result provides an empirical estimator. How would you estimate the asymptotic covariance matrix for your estimators in part (b)?

(d) Prove that $\alpha\beta\text{Cov}[\ln x, x^{\beta}] = 1$. (**Hint:** Use the fact that the expected first derivatives of the log-likelihood function are zero.)

The log-likelihood and its two first derivatives are

$$\log L = n\log\alpha + n\log\beta + (\beta-1)\sum_{i=1}^{n}\log x_i - \alpha\sum_{i=1}^{n} x_i^{\beta}$$

$$\partial\log L/\partial\alpha = n/\alpha - \sum_{i=1}^{n} x_i^{\beta}$$

$$\partial\log L/\partial\beta = n/\beta + \sum_{i=1}^{n}\log x_i - \alpha\sum_{i=1}^{n}(\log x_i)x_i^{\beta}$$

Since the first likelihood equation implies that at the maximum, $\hat{\alpha} = n/\sum_{i=1}^{n} x_i^{\beta}$, one approach would be to scan over the range of β and compute the implied value of α. Two practical complications are the allowable range of β and the starting values to use for the search.

The second derivatives are

$$\partial^2\ln L/\partial\alpha^2 = -n/\alpha^2$$

$$\partial^2\ln L/\partial\beta^2 = -n/\beta^2 - \alpha\sum_{i=1}^{n}(\log x_i)^2 x_i^{\beta}$$

$$\partial^2\ln L/\partial\alpha\partial\beta = -\sum_{i=1}^{n}(\log x_i)x_i^{\beta}\ .$$

If we had estimates in hand, the simplest way to estimate the expected values of the Hessian would be to evaluate the expressions above at the maximum likelihood estimates, then compute the negative inverse. First, since the expected value of $\partial\ln L/\partial\alpha$ is zero, it follows that $E[x_i^{\beta}] = 1/\alpha$. Now,

$$E[\partial\ln L/\partial\beta] = n/\beta + E[\sum_{i=1}^{n}\log x_i] - \alpha E[\sum_{i=1}^{n}(\log x_i)x_i^{\beta}] = 0$$

as well. Divide by n, and use the fact that every term in a sum has the same expectation to obtain

$$1/\beta + E[\ln x_i] - E[(\ln x_i)x_i^{\beta}]/E[x_i^{\beta}] = 0.$$

Now, multiply through by $E[x_i^{\beta}]$ to obtain $E[x_i^{\beta}] = E[(\ln x_i)x_i^{\beta}] - E[\ln x_i]E[x_i^{\beta}]$

or $\qquad 1/(\alpha\beta) = \text{Cov}[\ln x_i, x_i^{\beta}]$.

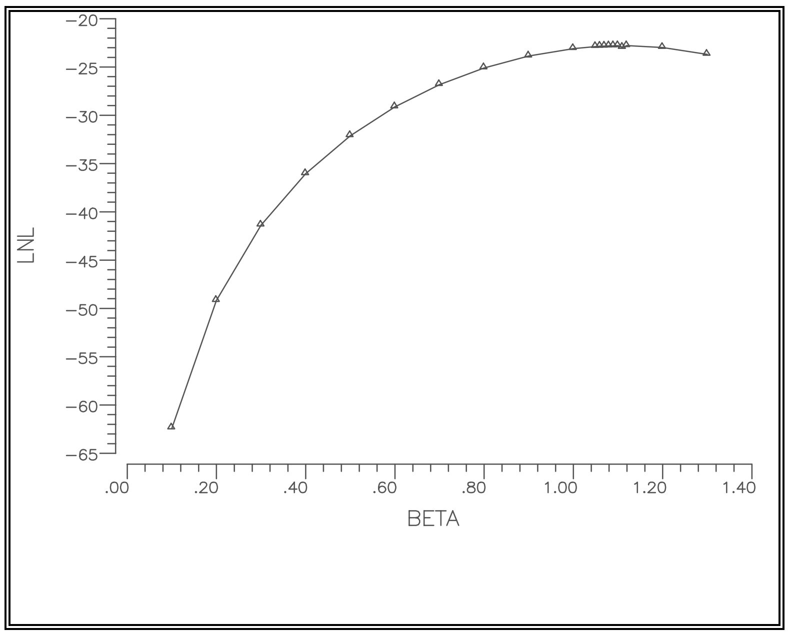
5. The following data were generated by the Weibull distribution of Exercise 17:

1.3043	.49254	1.2742	1.4019	.32556	.29965	.26423
1.0878	1.9461	.47615	3.6454	.15344	1.2357	.96381
.33453	1.1227	2.0296	1.2797	.96080	2.0070	

(a) Obtain the maximum likelihood estimates of α and β and estimate the asymptotic covariance matrix for the estimates.
(b) Carry out a Wald test of the hypothesis that $\beta = 1$.
(c) Obtain the maximum likelihood estimate of α under the hypothesis that $\beta = 1$.
(d) Using the results of a. and c. carry out a likelihood ratio test of the hypothesis that $\beta = 1$.
(e) Carry out a Lagrange multiplier test of the hypothesis that $\beta = 1$.

As suggested in the previous problem, we can concentrate the log-likelihood over α. From $\partial \log L/\partial\alpha = 0$, we find that at the maximum, $\alpha = 1/[(1/n)\sum_{i=1}^{n} x_i^{\beta}]$. Thus, we scan over different values of β to seek the value which maximizes $\log L$ as given above, where we substitute this expression for each occurrence of α. Values of β and the log-likelihood for a range of values of β are listed and shown in the figure below.

β	$\log L$
0.1	-62.386
0.2	-49.175
0.3	-41.381
0.4	-36.051
0.5	-32.122
0.6	-29.127
0.7	-26.829
0.8	-25.098
0.9	-23.866
1.0	-23.101
1.05	-22.891
1.06	-22.863
1.07	-22.841
1.08	-22.823
1.09	-22.809
1.10	-22.800
1.11	-22.796
1.12	-22.797
1.2	-22.984
1.3	-23.693



The maximum occurs at $\beta = 1.11$. The implied value of α is 1.179. The negative of the second derivatives matrix at these values and its inverse are

$$\mathbf{I}\left(\hat{\alpha},\hat{\beta}\right) = \begin{bmatrix} 25.55 & 9.6506 \\ 9.6506 & 27.7552 \end{bmatrix} \text{ and } \mathbf{I}^{-1}\left(\hat{\alpha},\hat{\beta}\right) = \begin{bmatrix} .04506 & -.2673 \\ -.2673 & .04148 \end{bmatrix}.$$

The Wald statistic for the hypothesis that $\beta = 1$ is $W = (1.11 - 1)^2/.041477 = .276$. The critical value for a test of size .05 is 3.84, so we would not reject the hypothesis.

If $\beta = 1$, then $\hat{\alpha} = n/\sum_{i=1}^{n} x_i = 0.88496$. The distribution specializes to the geometric distribution if $\beta = 1$, so the restricted log-likelihood would be

$$\log L_r = n\log\alpha - \alpha\sum_{i=1}^{n} x_i = n(\log\alpha - 1) \text{ at the MLE.}$$

$\log L_r$ at $\alpha = .88496$ is -22.44435. The likelihood ratio statistic is $-2\log\lambda = 2(23.10068 - 22.44435) = 1.3126$. Once again, this is a small value. To obtain the Lagrange multiplier statistic, we would compute

$$\begin{bmatrix} \partial \log L/\partial\alpha & \partial \log L/\partial\beta \end{bmatrix} \begin{bmatrix} -\partial^2 \log L/\partial\alpha^2 & -\partial^2 \log L/\partial\alpha\partial\beta \\ -\partial^2 \log L/\partial\alpha\partial\beta & -\partial^2 \log L/\partial\beta^2 \end{bmatrix}^{-1} \begin{bmatrix} \partial \log L/\partial\alpha \\ \partial \log L/\partial\beta \end{bmatrix}$$

at the restricted estimates of α = .88496 and β = 1. Making the substitutions from above, at these values, we would have

$$\partial \log L/\partial\alpha \;=\; 0$$

$$\partial \log L/\partial\beta \;=\; n + \sum_{i=1}^{n} \log x_i - \frac{1}{\bar{x}}\sum_{i=1}^{n} x_i \log x_i \;=\; 9.400342$$

$$\partial^2 \log L/\partial\alpha^2 \;=\; -n\bar{x}^{-2} \;=\; -25.54955$$

$$\partial^2 \log L/\partial\beta^2 \;=\; -n - \frac{1}{\bar{x}}\sum_{i=1}^{n} x_i (\log x_i)^2 \;=\; -30.79486$$

$$\partial^2 \log L/\partial\alpha\partial\beta \;=\; -\sum_{i=1}^{n} x_i \log x_i \;=\; -8.265.$$

The lower right element in the inverse matrix is .041477. The LM statistic is, therefore, $(9.40032)^2.041477$ = 2.9095. This is also well under the critical value for the chi-squared distribution, so the hypothesis is not rejected on the basis of any of the three tests.

6. (**Limited Information Maximum Likelihood Estimation**). Consider a bivariate distribution for x and y that is a function of two parameters, α and β. The joint density is $f(x,y|\alpha,\beta)$. We consider maximum likelihood estimation of the two parameters. The full information maximum likelihood estimator is the now familiar maximum likelihood estimator of the two parameters. Now, suppose that we can factor the joint distribution as done in Exercise 3, but in this case, we have $f(x,y|\alpha,\beta) = f(y|x,\alpha,\beta)f(x|\alpha)$. That is, the conditional density for y is a function of both parameters, but the marginal distribution for x involves only α.

a. Write down the general form for the log likelihood function using the joint density.

b. Since the joint density equals the product of the conditional times the marginal, the log likelihood function can be written equivalently in terms of the factored density. Write this down, in general terms.

c. The parameter α can be estimated by itself using only the data on x and the log likelihood formed using the marginal density for x. It can also be estimated with β by using the full log likelihood function and data on both y and x. Show this.

d. Show that the first estimator in part c has a larger asymptotic variance than the second one. This is the difference between a limited information maximum likelihood estimator and a full information maximum likelihood estimator.

e. Show that if $\partial^2 \ln f(y|x,\alpha,\beta)/\partial\alpha\partial\beta = 0$, then the result in d. is no longer true.

a. The full log likelihood is $\log L = \Sigma \log f_{yx}(y,x|\alpha,\beta)$.

b. By factoring the density, we obtain the equivalent $\log L = \Sigma[\log f_{y|x}(y|x,\alpha,\beta) + \log f_x(x|\alpha)]$

c. We can solve the first order conditions in each case. From the marginal distribution for x,

$$\Sigma\, \partial \log f_x(x|\alpha)/\partial\alpha \;=\; 0$$

provides a solution for α. From the joint distribution, factored into the conditional plus the marginal, we have

$$\Sigma[\,\partial \log f_{y|x}(y|x,\alpha,\beta)/\partial\alpha + \partial \log f_x(x|\alpha)/\partial\alpha \;=\; 0$$

$$\Sigma[\,\partial \log f_{y|x}(y|x,\alpha,\beta)/\partial\beta \;=\; 0$$

d. The asymptotic variance obtained from the first estimator would be the negative inverse of the expected second derivative, Asy.Var[a] = $\{[-E[\Sigma^2\partial \log f_x(x|\alpha)/\partial\alpha^2]\}^{-1}$. Denote this $A_{\alpha\alpha}^{-1}$. Now, consider the second estimator for α and β jointly. The negative of the expected Hessian is shown below. Note that the $A_{\alpha\alpha}$ from the marginal distribution appears there, as the marginal distribution appears in the factored joint distribution.

$$-E\left[\frac{\partial^2 \log L}{\partial\begin{pmatrix}\alpha\\ \beta\end{pmatrix}\begin{pmatrix}\alpha\\ \beta\end{pmatrix}'}\right]=\begin{bmatrix}B_{\alpha\alpha} & B_{\alpha\beta}\\ B_{\beta\alpha} & B_{\beta\beta}\end{bmatrix}+\begin{bmatrix}A_{\alpha\alpha} & 0\\ 0 & 0\end{bmatrix}=\begin{bmatrix}A_{\alpha\alpha}+B_{\alpha\alpha} & B_{\alpha\beta}\\ B_{\beta\alpha} & B_{\beta\beta}\end{bmatrix}$$

The asymptotic covariance matrix for the joint estimator is the inverse of this matrix. To compare this to the asymptotic variance for the marginal estimator of α, we need the upper left element of this matrix. Using the formula for the partitioned inverse, we find that this upper left element in the inverse is

$$[(A_{\alpha\alpha}+B_{\alpha\alpha}) - (B_{\alpha\beta}B_{\beta\beta}^{-1}B_{\beta\alpha})]^{-1} = [A_{\alpha\alpha} + (B_{\alpha\alpha} - B_{\alpha\beta}B_{\beta\beta}^{-1}B_{\beta\alpha})]^{-1}$$

which is smaller than $A_{\alpha\alpha}$ as long as the second term is positive.

e. (Unfortunately, this is an error in the text.) In the preceding expression, $B_{\alpha\beta}$ is the cross derivative. Even if it is zero, the asymptotic variance from the joint estimator is still smaller, being $[A_{\alpha\alpha} + B_{\alpha\alpha}]^{-1}$. This makes sense. If α appears in the conditional distribution, then there is additional information in the factored joint likelihood that is not in the marginal distribution, and this produces the smaller asymptotic variance.

7. Show that the likelihood inequality in Theorem 17.3 holds for the Poisson distribution used in Section 17.3 by showing that $E[(1/n)\ln L(\theta \mid \mathbf{y})]$ is uniquely maximized at $\theta = \theta_0$. Hint: First show that the expectation is $-\theta + \theta_0\ln\theta - E_0[\ln y_i!]$.

The log likelihood for the Poisson model is

$$\text{LogL} = -n\lambda + \log\lambda\Sigma_i\, y_i - \Sigma_i \log y_i!$$

The expected value of 1/n times this function with respect to the true distribution is

$$E[(1/n)\text{logL}] = -\lambda + \log\lambda\, E_0[\,\bar{y}\,] - E_0\,(1/n)\Sigma_i \log y_i!$$

The first expectation is λ_0. The second expectation can be left implicit since it will not affect the solution for λ - it is a function of the true λ_0. Maximizing this function with respect to λ produces the necessary condition

$$\partial E_0\,(1/n)\text{logL}]/\partial\lambda = -1 + \lambda_0/\lambda = 0$$

which has solution $\lambda = \lambda_0$ which was to be shown.

8. Show that the likelihood inequality in Theorem 17.3 holds for the normal distribution.

The log likelihood for a sample from the normal distribution is

$$\text{LogL} = -(n/2)\log2\pi - (n/2)\log\sigma^2 - 1/(2\sigma^2)\,\Sigma_i\,(y_i - \mu)^2.$$

$$E_0\,[(1/n)\text{logL}] = -(1/2)\log2\pi - (1/2)\log\sigma^2 - 1/(2\sigma^2)\,E_0[(1/n)\,\Sigma_i\,(y_i - \mu)^2].$$

The expectation term equals $E_0[(y_i - \mu)^2] = E_0[(y_i - \mu_0)^2] + (\mu_0 - \mu)^2 = \sigma_0^2 + (\mu_0 - \mu)^2$. Collecting terms,

$$E_0\,[(1/n)\text{logL}] = -(1/2)\log2\pi - (1/2)\log\sigma^2 - 1/(2\sigma^2)[\,\sigma_0^2 + (\mu_0 - \mu)^2]$$

To see where this is maximized, note first that the term $(\mu_0 - \mu)^2$ enters negatively as a quadratic, so the maximizing value of μ is obviously μ_0. Since this term is then zero, we can ignore it, and look for the σ^2 that maximizes $-(1/2)\log2\pi - (1/2)\log\sigma^2 - \sigma_0^2/(2\sigma^2)$. The $-1/2$ is irrelevant as is the leading constant, so we wish to minimize (after changing sign) $\log\sigma^2 + \sigma_0^2/\sigma^2$ with respect to σ^2. Equating the first derivative to zero produces $1/\sigma^2 = \sigma_0^2/(\sigma^2)^2$ or $\sigma^2 = \sigma_0^2$, which gives us the result.

9. For random sampling from the classical regression model in (17-3), reparameterize the likelihood function in terms of $\eta = 1/\sigma$ and $\boldsymbol{\delta} = (1/\sigma)\boldsymbol{\beta}$. Find the maximum likelihood estimators of η and $\boldsymbol{\delta}$ and obtain the asymptotic covariance matrix of the estimators of these parameters.

The log likelihood for the classical normal regression model is

$$\text{LogL} = \Sigma_i -(1/2)[\log 2\pi + \log\sigma^2 + (1/\sigma^2)(y_i - x_i'\beta)^2]$$

If we reparameterize this in terms of $\eta = 1/\sigma$ and $\delta = \beta/\sigma$, then after a bit of manipulation,

$$\text{LogL} = \Sigma_i -(1/2)[\log 2\pi - \log\eta^2 + (\eta y_i - x_i'\delta)^2]$$

The first order conditions for maximizing this with respect to η and δ are

$$\partial\text{logL}/\partial\eta = n/\eta - \Sigma_i y_i (\eta y_i - x_i'\delta) = 0$$

$$\partial\text{logL}/\partial\delta = \Sigma_i x_i (\eta y_i - x_i'\delta) = 0$$

Solve the second equation for δ, which produces $\delta = \eta (X'X)^{-1}X'y = \eta$ b. Insert this implicit solution into the first equation to produce $n/\eta = \Sigma_i y_i (\eta y_i - \eta x_i'b)$. By taking η outside the summation and multiplying the entire expression by η, we obtain $n = \eta^2 \Sigma_i y_i (y_i - x_i'b)$ or $\eta^2 = n/[\Sigma_i y_i (y_i - x_i'b)]$. This is an analytic solution for η that is only in terms of the data – b is a sample statistic. Inserting the square root of this result into the solution for δ produces the second result we need. By pursuing this a bit further, you canshow that the solution for η^2 is just n/e′e from the original least squares regression, and the solution for δ is just b times this solution for η. The second derivatives matrix is

$$\partial^2\text{logL}/\partial\eta^2 = -n/\eta^2 - \Sigma_i y_i^2$$

$$\partial^2\text{logL}/\partial\delta\,\partial\delta' = -\Sigma_i x_i x_i'$$

$$\partial^2\text{logL}/\partial\delta\,\partial\eta = \Sigma_i x_i y_i.$$

We'll obtain the expectations conditioned on X. $E[y_i|x_i]$ is $x_i'\beta$ from the original model, which equals $x_i'\delta/\eta$. $E[y_i^2|x_i] = 1/\eta^2 (\delta'x_i)^2 + 1/\eta^2$. (The cross term has expectation zero.) Summing over observations and collecting terms, we have, conditioned on X,

$$E[\partial^2\text{logL}/\partial\eta^2|X] = -2n/\eta^2 - (1/\eta^2)\delta'X'X\delta$$

$$E[\partial^2\text{logL}/\partial\delta\,\partial\delta'|X] = -X'X$$

$$E[\partial^2\text{logL}/\partial\delta\,\partial\eta|X] = (1/\eta)X'X\delta$$

The negative inverse of the matrix of expected second derivatives is

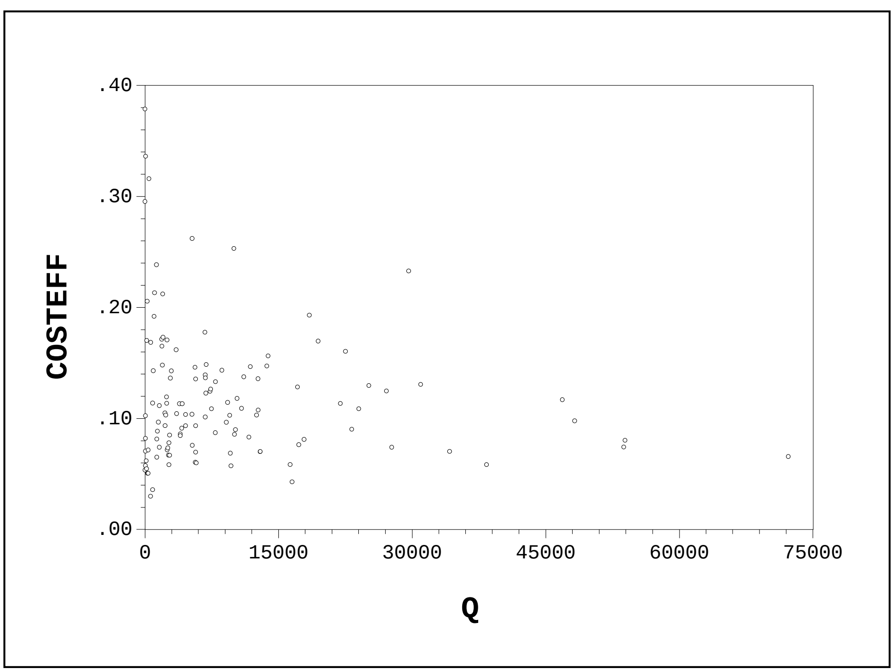
$$Asy.Var[\mathbf{d}, h] = \begin{bmatrix} \mathbf{X'X} & -(1/\eta)\mathbf{X'X}\boldsymbol{\delta} \\ -(1/\eta)\boldsymbol{\delta}'\mathbf{X'X} & (1/\eta^2)[2n + \boldsymbol{\delta}\mathbf{X'X}\boldsymbol{\delta}] \end{bmatrix}^{-1}$$

(The off diagonal term does not vanish here as it does in the original parameterization.)

10. Section 14.3.1 presents estimates of a Cobb-Douglas cost function using Nerlove's 1955 data on the U.S. electric power industry. Christensen and Greene's 1976 update of this study used 1970 data for this industry. The Christensen and Greene data are given in Table F5.2. These data have provided a standard test data set for estimating different forms of production and cost functions, including the stochastic frontier model examined in Example 17.5. It has been suggested that one explanation for the apparent finding of economies of scale in these data is that the smaller firms were inefficient for other reasons. The stochastic frontier might allow one to disentangle these effects. Use these data to fit a frontier cost function which includes a quadratic term in log output in addition to the linear term and the factor prices. Then examine the estimated Jondrow et al. residuals to see if they do indeed vary negatively with output, as suggested. (This will require either some programming on your part or specialized software. The stochastic frontier model is provided as an option in TSP and LIMDEP. Or, the likelihood function can be programmed fairly easily for RATS or GAUSS. Note, for a cost frontier as opposed to a production frontier, it is necessary to reverse the sign on the argument in the Φ function.)

We used LIMDEP to fit the cost frontier. The dependent variable is log(Cost/Pfuel). The regressors are a constant, log(Pcapital/Pfuel), log(Plabor/Pfuel), logQ and $\log^2 Q$. The Jondrow measure was then computed and plotted against output. There does not appear to be any relationship, though the weak relationship such as it is, is indeed, negative.

```
+----------------------------------------------+
| Limited Dependent Variable Model - FRONTIER |
| Dependent variable                  LCF     |
| Number of observations              123     |
| Log likelihood function        66.86502     |
| Variances: Sigma-squared(v)=       .01185   |
|            Sigma-squared(u)=       .02233   |
|            Sigma(v)        =       .10884   |
|            Sigma(u)        =       .14944   |
| Sigma = Sqr[(s^2(u)+s^2(v)]=       .18488   |
| Stochastic Cost Frontier, e=v+u.            |
+----------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
          Primary Index Equation for Model
 Constant    -7.494211759       .30737742  -24.381   .0000
 LPK       .5531289074E-01  .70211904E-01     .788   .4308     .88666047
 LPL          .2605889758   .67708437E-01    3.849   .0001     5.5808828
 LQ           .4109789313   .29495035E-01   13.934   .0000     8.1794715
 LQ2       .6058235980E-01  .43732083E-02   13.853   .0000     35.112527
          Variance parameters for compound error
 Lambda       1.373117163       .33353523    4.117   .0000
 Sigma        .1848750589   .28257115E-01    6.543   .0000
```



11. Consider, sampling from a multivariate normal distribution with mean vector $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_M)$ and covariance matrix $\sigma^2\mathbf{I}$. The log likelihood function is

$$\ln L = \frac{-nM}{2}\ln(2\pi) - \frac{nM}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(\mathbf{y}_i - \boldsymbol{\mu})'(\mathbf{y}_i - \boldsymbol{\mu}).$$

Show that the maximum likelihood estimates of the parameters are

$$\hat{\sigma}^2_{ML} = \frac{\sum_{i=1}^{n}\sum_{m=1}^{M}(y_{im} - \overline{y}_{n,m})^2}{nM} = \frac{1}{M}\sum_{m=1}^{M}\frac{1}{n}\sum_{i=1}^{n}(y_{im} - \overline{y}_{n,m})^2 = \frac{1}{M}\sum_{m=1}^{M}\hat{\sigma}^2_m.$$

Derive the second derivatives matrix and show that the asymptotic covariance matrix for the maximum likelihood estimators is

$$\left\{-E\left[\frac{\partial^2 \ln L}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}'}\right]\right\}^{-1} = \begin{bmatrix}\sigma^2\mathbf{I}/n & \mathbf{0}\\ \mathbf{0} & 2\sigma^4/(nM)\end{bmatrix}.$$

Suppose that we wished to test the hypothesis that the means of the M distributions were all equal. First, we might have a particular value μ^0 in mind. Show that the Wald statistic would be

$$\boldsymbol{W} = (\overline{\mathbf{x}} - \mu^0\mathbf{i})'\left(\frac{\hat{\sigma}^2}{n}\mathbf{I}\right)^{-1}(\overline{\mathbf{x}} - \mu^0\mathbf{i}), = \left(\frac{n}{s^2}\right)(\overline{\mathbf{x}} - \mu^0\mathbf{i})'(\overline{\mathbf{x}} - \mu^0\mathbf{i}),$$

where $\overline{\mathbf{x}}$ is the vector of sample means.

The first derivatives of the log likelihood function are $\partial\log L/\partial\mu = -(1/2\sigma^2)\,\Sigma_i\, -2(\mathbf{y}_i - \boldsymbol{\mu})$. Equating this to zero produces the vector of means for the estimator of $\boldsymbol{\mu}$. The first derivative with respect to σ^2 is

$\partial\log L/\partial\sigma^2 = -nM/(2\sigma^2) + 1/(2\sigma^4)\Sigma_i\,(\mathbf{y}_i - \boldsymbol{\mu})'(\mathbf{y}_i - \boldsymbol{\mu})$. Each term in the sum is $\Sigma_m\,(y_{im} - \mu_m)^2$. We already deduced that the estimators of μ_m are the sample means. Inserting these in the solution for σ^2 and solving the likelihood equation produces the solution given in the problem. The second derivatives of the log likelihood are

$\partial^2\log L/\partial\boldsymbol{\mu}\partial\boldsymbol{\mu}' = (1/\sigma^2)\Sigma_i\,-\mathbf{I}$

$\partial^2\log L/\partial\boldsymbol{\mu}\partial\sigma^2 = (1/2\sigma^4)\,\Sigma_i\,-2(\mathbf{y}_i - \boldsymbol{\mu})$

$\partial^2\log L/\partial\sigma^2\partial\sigma^2 = nM/(2\sigma^4) - 1/\sigma^6\,\Sigma_i\,(\mathbf{y}_i - \boldsymbol{\mu})'(\mathbf{y}_i - \boldsymbol{\mu})$

The expected value of the first term is $(-n/\sigma^2)\mathbf{I}$. The second term has expectation zero. Each term in the summation in the third term has expectation $M\sigma^2$, so the summation has expected value $nM\sigma^2$. Adding gives the expectation for the third term of $-nM/(2\sigma^4)$. Assembling these in a block diagonal matrix, then taking the negative inverse produces the result given earlier.

For the Wald test, the restriction is

H_0: $\boldsymbol{\mu} - \mu^0\mathbf{i} = \mathbf{0}$.

The unrestricted estimator of μ is $\overline{\mathbf{x}}$. The variance of $\overline{\mathbf{x}}$ is given above, so the Wald statistic is simply $(\overline{\mathbf{x}} - \mu^0\mathbf{i})'\,\mathrm{Var}[(\overline{\mathbf{x}} - \mu^0\mathbf{i})]^{-1}(\overline{\mathbf{x}} - \mu^0\mathbf{i})$. Inserting the covariance matrix given above produces the suggested statistic.

Chapter 18

The Generalized Method of Moments

1. For the normal distribution $\mu_{2k} = \sigma^{2k}(2k)!/(k!2^k)$ and $\mu_{2k+1} = 0$, $k = 0, 1, \ldots$. Use this result to analyze the two estimators

$$\sqrt{b_1} = \frac{m_3}{m_2^{3/2}} \text{ and } b_2 = \frac{m_4}{m_2^2}.$$

where $m_k = \frac{1}{n}\Sigma_{i=1}^n (x_i - \bar{x})^k$. The following result will be useful:

$$Asy.Cov[\sqrt{n}m_j, \sqrt{n}m_k] = \mu_{j+k} - \mu_j\mu_k + jk\mu_2\mu_{j-1}\mu_{k-1} - j\mu_{j-1}\mu_{k+1} - k\mu_{k-1}\mu_{j+1}.$$

Use the delta method to obtain the asymptotic variances and covariance of these two functions assuming the data are drawn from a normal distribution with mean μ and variance σ^2. (Hint: Under the assumptions, the sample mean is a consistent estimator of μ, so for purposes of deriving asymptotic results, the difference between $\bar{x}$ and μ may be ignored. As such, no generality is lost by assuming the mean is zero, and proceeding from there. Obtain **V**, the 3×3 covariance matrix for the three moments, then use the delta method to show that the covariance matrix for the two estimators is

$$\mathbf{JVJ'} = \begin{bmatrix} 6 & 0 \\ 0 & 24 \end{bmatrix}$$

where **J** is the 2×3 matrix of derivatives.

The elements of **J** are

$$\frac{\partial\sqrt{b_1}}{\partial m_2} = m_3(-3/2)m_2^{-5/2} \quad \frac{\partial\sqrt{b_1}}{\partial m_3} = m_2^{-3/2} \quad \frac{\partial\sqrt{b_1}}{\partial m_4} = 0$$

$$\frac{\partial b_2}{\partial m_2} = m_4(-2)m_2^{-3} \quad \frac{\partial b_2}{\partial m_3} = 0 \quad \frac{\partial b_2}{\partial m_4} = m_2^{-2}$$

Using the formula given for the moments, we obtain, $\mu_2 = \sigma^2$, $\mu_3 = 0$, $\mu_4 = 3\sigma_4$. Insert these in the derivatives above to obtain

$$\mathbf{J} = \begin{bmatrix} 0 & \sigma^{-3} & 0 \\ -6\sigma^{-2} & 0 & \sigma^{-4} \end{bmatrix}.$$

Since the rows of J are orthogonal, we know that the off diagonal term in **JVJ′** will be zero, which simplifies things a bit. Taking the parts directly, we can see that the asymptotic variance of $\sqrt{b_1}$ will be σ^{-6} Asy.Var[m_3], which will be

Asy.Var[$\sqrt{b_1}$] = $\sigma^{-6}(\mu_6 - \mu_3^2 + 9\mu_2^3 - 3\mu_2\mu_4 - 3\mu_2\mu_4)$.

The parts needed, using the general result given earlier, are $\mu_6 = 15\sigma^6$, $\mu_3 = 0$, $\mu_2 = \sigma^2$, $\mu_4 = 3\sigma^4$. Inserting these in the parentheses and multiplying it out and collecting terms produces the upper left element of JVJ′ equal to 6, which is the desired result. The lower right element will be

Asy.Var[b_2] = $36\sigma^{-4}$ Asy.Var[m_2] + σ^{-8}Asy.Var[m_4] - $2(6)\sigma^{-6}$Asy.Cov[m_2,m_4].

The needed parts are

Asy.Var[m_2] = $2\sigma^4$

Asy.Var[m_4] = $\mu_8 - \mu_4^2 = 105\sigma^8 - (3\sigma^4)^2$

Asy.Cov[m_2,m_4] = $\mu_6 - \mu_2\mu_4 = 15\sigma^6 - \sigma^2(3\sigma^4)$.

Inserting these parts in the expansion, multiplying it out and collecting terms produces the lower right element equal to 24, as expected.

2. Using the results in Example 18.7, estimate the asymptotic covariance matrix of the method of moments estimators of P and λ based on m_1' and m_2'. [Note: You will need to use the data in Example C.1 to estimate **V**.]

The necessary data are given in Examples 18.5 and 18.7. The two moments are m_1' =31.278 and m_2'.=1453.96. Based on the theoretical results $m_1' = P/\lambda$ and $m_2' = P(P+1)/\lambda^2$, the solutions are $P = \mu_1'^2/(\mu_2' - \mu_1'^2)$ and $\lambda = \mu_1'/(\mu_2' - \mu_1'^2)$. Using the sample moments produces estimates P = 2.05682 and λ = 0.065759. The matrix of derivatives is

$$\mathbf{G} = \begin{bmatrix} \partial\mu_1'/\partial P & \partial\mu_1'/\partial\lambda \\ \partial\mu_2'/\partial P & \partial\mu_2'/\partial\lambda \end{bmatrix} = \begin{bmatrix} 1/\lambda & -P/\lambda^2 \\ (2P+1)/\lambda^2 & -2P(P+1)/\lambda^3 \end{bmatrix} = \begin{bmatrix} 15.207 & -475.648 \\ 1{,}182.551 & -44{,}221.20 \end{bmatrix}.$$

The covariance matrix for the moments is given in Example 18.7;

$$\frac{1}{20}\Phi = \begin{bmatrix} 25.0339 & 2313.4163 \\ 2313.4163 & 228047.8 \end{bmatrix}$$

The estimated asymptotic covariance matrix for the two estimators is

$(1/20)[\mathbf{G}'\mathbf{\Phi}^{-1}\mathbf{G}]^{-1} = \begin{bmatrix} .003752 & .0084037 \\ .0084037 & .266831 \end{bmatrix}$ so the two standard errors for the estimators of λ and P are 0.01937 and 0.51656, respectively.

3. **Exponential Families of Distributions**) For each of the following distributions, determine whether it is an exponential family by examining the log likelihood function. Then, identify the sufficient statistics.

a. Normal distribution with mean μ and variance σ^2.

b. The Weibull distribution in Exercise 4 in Chapter 17.

c. The mixture distribution in Exercise 3 in Chapter 17.

a. The log likelihood for sampling from the normal distribution is

$$\log L = (-1/2)[n\log 2\pi + n\log\sigma^2 + (1/\sigma^2)\Sigma_i (x_i - \mu)^2]$$

write the summation in the last term as $\Sigma x_i^2 + n\mu^2 - 2\mu\Sigma_i x_i$. Thus, it is clear that the log likelihood is of the form for an exponential family, and the sufficient statistics are the sum and sum of squares of the observations.

b. The log of the density for the Weibull distribution is

$$\log f(x) = \log\alpha + \log\beta + (\beta-1)\log x_i - \alpha\Sigma_i x_i^{\beta}.$$

The log likelihood is found by summing these functions. The third term does not factor in the fashion needed to produce an exponential family. There are no sufficient statistics for this distribution.

c. The log of the density for the mixture distribution is

$$\log f(x,y) = \log\theta - (\beta+\theta)y_i + x_i\log\beta + x_i\log y_i - \log(x!)$$

This is an exponential family; the sufficient statistics are $\Sigma_i y_i$ and $\Sigma_i x_i$..

4. In the classical regression model with heteroscedasticity, which is more efficient, ordinary least squares or GMM? Obtain the two estimators and their respective asymptotic covariance matrices, then prove your assertion.

The question is (deliberately) misleading. We showed in Chapter 11 that in the classical regression model with heteroscedasticity, the OLS estimator is the GMM estimator. The asymptotic covariance matrix of the OLS estimator is given in Section 11.2. The estimator of the asymptotic covariance matrices are $s^2(\mathbf{X}'\mathbf{X})^{-1}$ for OLS and the White estimator for GMM.

5. Consider the probit model analyzed in Section 17.8. The model states that for given vector of independent variables,

$$\text{Prob}[y_i = 1 \mid \mathbf{x}_i] = \Phi[\mathbf{x}_i'\beta],\ \text{Prob}[y_i = 0 \mid \mathbf{x}_i] = 1 - \text{Prob}[y_i = 1 \mid \mathbf{x}_i].$$

We have considered maximum likelihood estimation of the parameters of this model at several points. Consider, instead, a GMM estimator based on the result that

$$E[y_i \mid \mathbf{x}_i] = \Phi(\mathbf{x}_i'\boldsymbol{\beta})$$

This suggests that we might base estimation on the orthogonality conditions

$$E\left[\left(y_i - \Phi(\mathbf{x}_i'\boldsymbol{\beta})\right)\mathbf{x}_i\right] = \mathbf{0}$$

Construct a GMM estimator based on these results. Note that this is not the nonlinear least squares estimator. Explain – what would the orthogonality conditions be for nonlinear least squares estimation of this model?

The GMM estimator would be chosen to minimize the criterion

$q = n\,\mathbf{m}'\mathbf{W}\mathbf{m}$

where $\mathbf{W}$ is the weighting matrix and $\mathbf{m}$ is the empirical moment,

$\mathbf{m} = (1/n)\Sigma_i\,(y_i - \Phi(\mathbf{x}_i'\boldsymbol{\beta}))\mathbf{x}_i$

For the first pass, we'll use $\mathbf{W} = \mathbf{I}$ and just minimize the sumof squares. This provides an initial set of estimates that can be used to compute the optimal weighting matrix. With this first round estimate, we compute

$\mathbf{W} = [(1/n^2)\,\Sigma_i\,(y_i - \Phi(\mathbf{x}_i'\boldsymbol{\beta}))^2\,\mathbf{x}_i\,\mathbf{x}_i']^{-1}$

then return to the optimization problem to find the optimal estimator. The asymptotic covariance matrix is computed from the first order conditions for the optimization. The matrix of derivatives is

$\mathbf{G} = \partial\mathbf{m}/\partial\boldsymbol{\beta}' = (1/n)\Sigma_i\,-\phi(\mathbf{x}_i'\boldsymbol{\beta})\mathbf{x}_i\mathbf{x}_i'$

The estimator of the asymptotic covariance matrix will be

$\mathbf{V} = (1/n)[\mathbf{G}'\mathbf{W}\mathbf{G}]^{-1}$

6. Consider GMM estimation of a regression model as shown at the beginning of Example 18.8. Let $\mathbf{W}_1$ be the optimal weighting matrix based on the moment equations. Let $\mathbf{W}_2$ be some other positive definite matrix. Compare the asymptotic covariance matrices of the two proposed estimators. Show conclusively that the asymptotic covariance matrix of the estimator based on $\mathbf{W}_1$ is not larger than that based on $\mathbf{W}_2$.

This is the comparison between (18-12) and (18-11). The proof can be done by comparing the inverses of the two covariance matrices.

7. Suppose in a sample of 500 observations from a normal distribution with mean μ and standard deviation σ, you are told that 35% of the observations are less than 2.1 and 55% of the observations are less than 3.6. Estimate μ and σ.

If 35% of the observations are less than 2.1, we would infer that

$$\Phi[(2.1 - \mu)/\sigma] = .35, \text{ or } (2.1 - \mu)/\sigma = -.385 \Rightarrow 2.1 - \mu = -.385\sigma.$$

Likewise, $$\Phi[(3.6 - \mu)/\sigma] = .55, \text{ or } (3.6 - \mu)/\sigma = .126 \Rightarrow 3.6 - \mu = .126\sigma.$$

The joint solution is $\hat{\mu} = 3.2301$ and $\hat{\sigma} = 2.9354$. It might not seem obvious, but we can also derive asymptotic standard errors for these estimates by constructing them as method of moments estimators. Observe, first, that the two estimates are based on moment estimators of the probabilities. Let x_i denote one of the 500 observations drawn from the normal distribution. Then, the two proportions are obtained as follows: Let $z_i(2.1) = \mathbf{1}[x_i < 2.1]$ and $z_i(3.6) = \mathbf{1}[x_i < 3.6]$ be indicator functions. Then, the proportion of 35% has been obtained as $\bar{z}(2.1)$ and .55 is $\bar{z}(3.6)$. So, the two proportions are simply the means of functions of the sample observations. Each z_i is a draw from a Bernoulli distribution with success probability $\pi(2.1) = \Phi((2.1-\mu)/\sigma)$ for $z_i(2.1)$ and $\pi(3.6) = \Phi((3.6-\mu)/\sigma)$ for $z_i(3.6)$. Therefore, $E[\bar{z}(2.1)] = \pi(2.1)$, and $E[\bar{z}(3.6)] = \pi(3.6)$. The variances in each case are $\text{Var}[\bar{z}(.)] = 1/n[\pi(.)(1-\pi(.))]$. The covariance of the two sample means is a bit trickier, but we can deduce it from the results of random sampling. $\text{Cov}[\bar{z}(2.1), \bar{z}(3.6)]]$ $= 1/n\ \text{Cov}[z_i(2.1), z_i(3.6)]$, and, since in random sampling sample moments will converge to their population counterparts, $\text{Cov}[z_i(2.1), z_i(3.6)] = \text{plim}\ [\{(1/n)\sum_{i=1}^{n} z_i(2.1)z_i(3.6)\} - \pi(2.1)\pi(3.6)]$. But, $z_i(2.1)z_i(3.6)$ must equal $[z_i(2.1)]^2$ which, in turn, equals $z_i(2.1)$. It follows, then, that $\text{Cov}[z_i(2.1), z_i(3.6)] = \pi(2.1)[1 - \pi(3.6)]$. Therefore, the asymptotic covariance matrix for the two sample proportions is $Asy.Var[p(2.1), p(3.6)] = \mathbf{\Sigma} = \frac{1}{n}\begin{bmatrix} \pi(2.1)(1-\pi(2.1)) & \pi(2.1)(1-\pi(3.6)) \\ \pi(2.1)(1-\pi(3.6)) & \pi(3.6)(1-\pi(3.6)) \end{bmatrix}$. If we insert our sample estimates, we obtain $Est.Asy.Var[p(2.1), p(3.6)] = \mathbf{S} = \begin{bmatrix} 0.000455 & 0.000315 \\ 0.000315 & 0.000495 \end{bmatrix}$. Now, ultimately, our estimates of μ and σ are found as functions of $p(2.1)$ and $p(3.6)$, using the method of moments. The moment equations are

$$m_{2.1} = \left[\frac{1}{n}\sum_{i=1}^{n} z_i(2.1)\right] - \Phi\left[\frac{2.1-\mu}{\sigma}\right] = 0,$$

$$m_{3.6} = \left[\frac{1}{n}\sum_{i=1}^{n} z_i(3.6)\right] - \Phi\left[\frac{3.6-\mu}{\sigma}\right] = 0.$$

Now, let $\mathbf{\Gamma} = \begin{bmatrix} \partial m_{2.1}/\partial\mu & \partial m_{2.1}/\partial\sigma \\ \partial m_{3.6}/\partial\mu & \partial m_{3.61}/\partial\sigma \end{bmatrix}$ and let $\mathbf{G}$ be the sample estimate of Γ. Then, the estimator of the asymptotic covariance matrix of $(\hat{\mu}, \hat{\sigma})$ is $[\mathbf{GS}^{-1}\mathbf{G}']^{-1}$. The remaining detail is the derivatives, which are just $\partial m_{2.1}/\partial\mu = (1/\sigma)\phi((2.1-\mu)/\sigma)$ and $\partial m_{2.1}/\partial\sigma = (2.1-\mu)/\sigma[\partial m_{2.1}/\partial\sigma]$ and likewise for $m_{3.6}$. Inserting our sample estimates produces $\mathbf{G} = \begin{bmatrix} 0.37046 & -0.14259 \\ 0.39579 & 0.04987 \end{bmatrix}$. Finally, multiplying the matrices and computing the necessary inverses produces $[\mathbf{GS}^{-1}\mathbf{G}']^{-1} = \begin{bmatrix} 0.10178 & -0.12492 \\ -0.12492 & 0.16973 \end{bmatrix}$. The asymptotic distribution would be normal, as usual. Based on these results, a 95% confidence interval for μ would be $3.2301 \pm 1.96(.10178)^2 = 2.6048$ to 3.8554.

Chapter 19

Models with Lagged Variables

1. Obtain the mean lag and the long and short run multipliers for the following distributed lag models.

(a) $y_t = .55[.02x_t + .15x_{t-1} + .43x_{t-2} + .23x_{t-3} + .17x_{t-4}] + \varepsilon_t$.

(b) The model in Exercise 5.

(6) The model in Exercise 8. (Do for either x or z.)

For the first, the mean lag is $.55(.02)(0) + .55(.15)(1) + ... + .55(.17)(4) = 1.31$ periods. The impact multiplier is $.55(.02) = .011$ while the long run multiplier is the sum of the coefficients, .55.

For the second, the coefficient on x_t is .6, so this is the impact multiplier. The mean lag is found by applying (18-9) to $B(L) = [.6 + 2L]/[1 - .6L + .5L^2] = A(L)/D(L)$. Then, $B'(1)/B(1) = \{[D(1)A'(1) - A(1)D'(1)]/[D(1)]^2\} / [A(1)/D(1)] = A'(1)/A(1) - D'(1)/D(1) = (2/2.6) / (.4/.9) = 1.731$ periods. The long run multiplier is $B(1) = 2.6/.9 = 2.888$ periods.

For the third, since we are interested only in the coefficients on x_t, write the model as $y_t = \alpha + \beta x_t[1 + \gamma L + \gamma^2 L^2 + ...] + \delta z_t^* + u_t$. The lag coefficients on x_t are simply β times powers of γ.

2. Explain how to estimate the parameters of the following model:

$$y_t = \alpha + \beta x_t + \gamma y_{t-1} + \delta y_{t-2} + \varepsilon_t, \quad \varepsilon_t = \rho\varepsilon_{t-1} + u_t.$$

Is there any problem with ordinary least squares? Using the method you have described, fit the model above to the data in Table F5.1. Report your results.

Because the model has both lagged dependent variables and autocorrelated disturbances, ordinary least squares will be inconsistent. Consistent estimates could be obtained by the method of instrumental variables. We can use x_{t-1} and x_{t-2} as the instruments for y_{t-1} and y_{t-2}. Efficient estimates can be obtained by a two step procedure. We write the model as $y_t - \rho y_{t-1} = \alpha(1-\rho) + \beta(x_t - \rho x_{t-1}) + \gamma(y_{t-1} - \rho y_{t-2}) + \delta(y_{t-2} - \rho y_{t-3}) + u_t$. With a consistent estimator of ρ, we could use FGLS. The residuals from the *IV* estimator can be used to estimate ρ. Then OLS using the transformed data is asymptotically equivalent to GLS. The method of Hatanaka discussed in the text is another possibility.

Using the real consumption and real disposable income data in Table F5.1, we obtain the following results: Estimated standard errors are shown in parentheses. (The estimated autocorrelation based on the IV estimates is .9172.) All three sets of estimates are based on the last 201 observations, 1950.4 to 2000.4

	OLS	IV	2 Step FGLS
$\hat{\alpha}$	-1.4946	-64.5073	-4.6614
	(3.8291)	(46.1075)	(3.2041)
$\hat{\beta}$	.007569	.7003	.3477
	(.001662)	(.4910)	(.0432)
$\hat{\gamma}$	1.1977	.5726	.2332
	(.006921)	(.9043)	(.05933)
$\hat{\delta}$	-0.1988	-.3324	.4072
	(.07109)	(.4962)	(.05500)

3. Show how to estimate a polynomial distributed lag model with lags of 6 periods and a third order polynomial using restricted least squares.

Using (18-22), we would regress y_t on a constant, $x_t, x_{t-1}, ..., x_{t-6}$. Constrained least squares using

$$\mathbf{R} = \begin{matrix} 1 & -5 & 10 & -10 & 5 & -1 & 0 & 0 \\ 0 & 1 & -5 & 10 & -10 & 5 & -1 & 0 \\ 0 & 0 & 1 & -5 & 10 & -10 & 5 & -1 \end{matrix}, \quad \mathbf{q} = \begin{matrix} 0 \\ 0 \\ 0 \end{matrix}$$

would produce the PDL estimates. □

4. Expand the rational lag model $y_t = \{[.6 + 2L]/[1 - .6L + .5L^2]\}\, x_t + \varepsilon_t$.
What are the coefficients on x_t, x_{t-1}, x_{t-2}, x_{t-3}, and x_{t-4}?

The ratio of polynomials will equal $B(L) = [.6 + 2L]/[1 - .6L + .5L^2]$. This will expand to $B(L) = \beta_0 + \beta_1 L + \beta_2 L^2 + ...$ Multiply both sides of the equation by $(1 - .6L + .5L^2)$ to obtain $(\beta_0 + \beta_1 L + \beta_2 L^2 + ...)(1 - .6L + .5L^2) = .6 + 2L$. Since the two sides must be equal, it follows that $\beta_0 = .6$ (the only term not involving L) $-.6\beta_0 + \beta_1 = 2$ (the only term involving only L. Therefore, $\beta_1 = 2.36$. All remaining terms, involving L^2, L^3, ... must equal zero. Therefore, $\beta_j - .6\beta_{j-1} + .5\beta_{j-2} = 0$ for all $j > 1$, or $\beta_j = .6\beta_{j-1} - .5\beta_{j-2}$. This provides a recursion for all remaining coefficients. For the specified coefficients, $\beta_2 = .6(2.36) - .5(.3) = 1.266$. $\beta_3 = .6(1.266) - .5(2.36) = -.4204$, $\beta_4 = .6(-.4204) - .5(1.266) = -.88524$ and so on.

5. Suppose the model of Exercise 4 were respecified as

$$y_t = \alpha + \{[\beta + \gamma L]/[1 + \delta_1 L + \delta_2 L^2]\}\, x_t + \varepsilon_t.$$

Describe a method of estimating the parameters. Is ordinary least squares consistent?

By multiplying through by the denominator of the lag function, we obtain an autoregressive form

$$\begin{aligned} y_t &= \alpha(1+\delta_1+\delta_2) + \beta x_t + \gamma x_{t-1} - \delta_1 y_{t-1} - \delta_2 y_{t-2} + \varepsilon_t + \delta_1\varepsilon_{t-1} + \delta_2\varepsilon_{t-2} \\ &= \alpha(1+\delta_1+\delta_2) + \beta x_t + \gamma x_{t-1} - \delta_1 y_{t-1} - \delta_2 y_{t-2} + v_t \end{aligned}$$

The model cannot be estimated consistently by ordinary least squares because there is autocorrelation in the presence of a lagged dependent variable. There are two approaches possible. Nonlinear least squares could be applied to the moving average (distributed lag) form. This would be fairly complicated, though a method of doing so is described by Maddala and Rao (1973). A much simpler approach would be to estimate the model in the autoregressive form using an instrumental variables estimator. The lagged variables x_{t-2} and x_{t-3} can be used for the lagged dependent variables.

6. Describe how to estimate the parameters of the model $y_t = \alpha + \beta x_t/(1 - \gamma L) + \delta z_t(1 - \phi L) + \varepsilon_t$ where ε_t is a serially uncorrelated, homoscedastic classical disturbance.

The model can be estimated as an autoregressive or distributed lag equation. Consider, first, the autoregressive form. Multiply through by $(1 - \gamma L)(1 - \phi L)$ to obtain

$$y_t = \alpha(1-\gamma)(1-\phi) + \beta x_t - (\beta\phi)x_{t-1} + \delta z_t - (\delta\gamma)z_{t-1} + (\gamma + \phi)y_{t-1} - (\gamma\phi)y_{t-2} + \varepsilon_t -(\gamma+\phi)\varepsilon_{t-1} + (\gamma\phi)\varepsilon_{t-2}.$$

Clearly, the model cannot be estimated by ordinary least squares, since there is an autocorrelated disturbance and a lagged dependent variable. The parameters can be estimated consistently, but inefficiently by linear instrumental variables. The inefficiency arises from the fact that the parameters are overidentified. The linear estimator estimates seven functions of the five underlying parameters. One possibility is a GMM estimator. Let $v_t = \varepsilon_t -(\gamma+\phi)\varepsilon_{t-1} + (\gamma\phi)\varepsilon_{t-2}$. Then, a GMM estimator can be defined in terms of, say, a set of moment equations of the form $E[v_t w_t] = 0$, where w_t is current and lagged values of x and z. A minimum distance estimator could then be used for estimation.

The distributed lag approach might be taken, instead. Each of the two regressors produces a recursions $x_t^* = x_t + \gamma x_{t-1}^*$ and $z_t^* = z_t + \gamma z_{t-1}^*$. Thus, values of the moving average regressors can be built up recursively. Note that the model is linear in 1, x_t^*, and z_t^*. Therefore, an approach is to search a grid of values of (γ,ϕ) to minimize the sum of squares.

7. We are interested in the long run multiplier in the model

$$y_t = \alpha + \sum_{j=0}^{6} \beta_j x_{t-j} + \varepsilon_t$$

Assume that x_t is an autoregressive series, $x_t = rx_{t-1} + v_t$ where $|r| < 1$.

a. What is the long run multiplier in this model?

b. How would you estimate the long run multiplier in this model?

c. Suppose you that the preceding is the true model but you linearly regress y_t only on a constant and the first 5 lags of x_t. How does this affect your estimate of the long run multiplier?

d. Same as c. for 4 lags instead of 5.

e. Using the macroeconomic data in Appendix F5.1, let y_t be the log of real investment and x_t be the log of real output. Carry out the computations suggested and report your findings. Specifically, how does the omission of a lagged value affect estimates of the short run and long run multipliers in the unrestricted lag model.

The long run multiplier is $\beta_0 + \beta_1 + ... + \beta_6$. The model is a classical regression, so it can be estimated by ordinary least squares. The estimator of the long run multiplier would be the sum of the least squares coefficients. If the sixth lag is omitted, then the standard omitted variable result applies, and all the coefficients are biased. The orthogonality result needed to remove the bias explicitly fails here, since x_t is an AR(1) process. All the lags are correlated. Since the form of the relationship is, in fact, known, we can derive the omitted variable formula. In particular, by construction, x_t will have mean zero. By implication, y_t will also, so we lose nothing by assuming that the constant term is zero. To save some cumbersome algebra, we'll also assume with no loss of generality that the unconditional variance of x_t is 1. Let $X_1 = [x_t, x_{t-1}, ..., x_{t-5}]$ and $X_2 = x_{t-6}$. Then, for the regression of y on X_1, we have by the omitted variable formula,

$$E\begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{bmatrix} | X_1 = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix} + \begin{bmatrix} 1 & r & r^2 & r^3 & r^4 & r^5 \\ r & 1 & r & r^2 & r^3 & r^4 \\ r^2 & r & 1 & r & r^2 & r^3 \\ r^3 & r^2 & r & 1 & r & r^2 \\ r^4 & r^3 & r^2 & r & 1 & r \\ r^5 & r^4 & r^3 & r^2 & r & 1 \end{bmatrix}^{-1} \begin{bmatrix} r^6 \\ r^5 \\ r^4 \\ r^3 \\ r^2 \\ r \end{bmatrix} \beta_6$$

We can derive a formal solution to the bias in this estimator. Note that the column that is to the right of the inverse matrix is r times the last column matrix. Therefore, the matrix product is r times the last column of an identity matrix. This gives us the complete result,

$$E\begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{bmatrix} | X_1 = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ r \end{bmatrix} \beta_6.$$

Therefore, the first 5 coefficients are unbiased, and the last one is an estimator of $\beta_5 + r\beta_6$. Adding these up, we see that when the last lag is omitted from the model, the estimator of the long run multiplier is biased downware by $(1-r)\beta_6$. For part d, we will use a similar construction. But, now there are five variables in X_1 and x_{t-5} and x_{t-6} in X_2. The same kind of computation will show that the first four coefficients are unbiased while the fifth now estimates $\beta_4 + r\beta_5 + r^2\beta_6$. The long run multiplier is estimated with downward bias equal to $(1-r)\beta_5 + (1-r^2)\beta_6$.

```
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |t-ratio |P[|T|>t] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 XT          .9726595701      1.9258818      .505   .6141     8.3384522
 XT1         .7709686332      3.1555811      .244   .8072     8.3301663
 XT2         .5450409860      3.1761465      .172   .8639     8.3218191
 XT3        -.6061007409      3.1903388     -.190   .8495     8.3134324
 XT4        -.2272352746      3.1729930     -.072   .9430     8.3050260
 XT5        -1.916555094      3.1414210     -.610   .5425     8.2964570
 XT6         1.218771893      1.8814874      .648   .5179     8.2878393
Matrix LRM      has  1 rows and  1 columns.
               1
        +--------------
       1|      .7575
 XT          1.101551478      1.9126777      .576   .5653     8.3384522
 XT1         .6941982792      3.1485851      .220   .8257     8.3301663
 XT2         .5287939572      3.1712435      .167   .8677     8.3218191
 XT3        -.7300170198      3.1797815     -.230   .8187     8.3134324
 XT4        -.5552651191      3.1275848     -.178   .8593     8.3050260
 XT5        -.2826674399      1.8697065     -.151   .8800     8.2964570
Matrix LRM      has  1 rows and  1 columns.
               1
        +--------------
       1|      .7566
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |t-ratio |P[|T|>t] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 XT          1.077633667      1.9012923      .567   .5715     8.3384522
 XT1         .7070443138      3.1394606      .225   .8221     8.3301663
 XT2         .5633400685      3.1549830      .179   .8585     8.3218191
 XT3        -.6608149939      3.1386871     -.211   .8335     8.3134324
 XT4        -.9304013056      1.8990464     -.490   .6247     8.3050260
Matrix LRM      has  1 rows and  1 columns.
               1
        +--------------
       1|      .7568
--> calc;list;cor(xt,xt1)$
    Result  =  .99978740920470700D+00
```

The results of the three suggested regressions are shown above. We used observations 7 - 204 of the logged real investment and real GDP data in deviations from the means for all regressions. Note that although there are some large changes in the estimated individual parameters, the long run multiplier is almost identical in all cases. Looking at the analytical results we can see why this would be the case. The correlation between current and lagged log gdp is r = 0.9998. Therefore, the biases that we found, $(1-r)\beta_6$ and $(1-r)\beta_5 + (1-r^2)\beta_6$ are trivial.

Chapter 20

Time Series Models

1. Find the autocorrelations and partial autocorrelations for the MA(2) process
$\varepsilon_t = v_t - \theta_1 v_{t-1} - \theta_2 v_{t-2}$.

The autocorrelations are simple to obtain just by multiplying out v_t^2, $v_t v_{t-1}$ and so on. The autocovariances are $1+\theta_1^2 + \theta_2^2$, $-\theta_2(1 - \theta_2)$, $-\theta_2$, 0, 0, 0... which provides the autocorrelations by division by the first of these. The partial autocorrelations are messy, and can be obtained by the Yule Walker equations. Alternatively (and much more simply), we can make use of the observation on page 615 that the partial autocorrelations for the MA(2) process mirror tha autocorrelations for an AR(2). Thus, the results on page 615 for the AR(2) can be used directly.

2. Carry out the ADF test for a unit root in the bond yield data of Example 20.1.

The regression results are shown below. We fit the regression using a constant, a time trend, the lagged dependent variable and three lagged first differences. The coefficient on “R1” is used for the test.

```
+-----------------------------------------------------------------------+
| Ordinary    least squares regression    Weighting variable = none     |
| Dep. var. = R        Mean=   8.212678571    , S.D.=   .7762719558     |
| Model size: Observations =       56, Parameters =   6, Deg.Fr.=     50 |
| Residuals:  Sum of squares= .9651001703    , Std.Dev.=         .13893 |
| Fit:        R-squared=  .970881, Adjusted R-squared =          .96797 |
| Model test: F[  5,     50] =  333.41,    Prob value =          .00000 |
| Diagnostic: Log-L =     34.2439, Restricted(b=0) Log-L =     -64.7739 |
|             LogAmemiyaPrCrt.=   -3.846, Akaike Info. Crt.=     -1.009 |
| Autocorrel: Durbin-Watson Statistic =   1.91589,   Rho =       .04205 |
+-----------------------------------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |t-ratio |P[|T|>t] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant     .2565690959       .47172815      .544   .5889
 T         .4401352136E-03  .25092142E-02      .175   .8615     32.500000
 R1           .9653227410   .48183346E-01   20.034   .0000     8.2305357
 DR1          .5600009441       .14342088    3.905   .0003 -.12321429E-01
 DR2        -.1739775168       .14781417   -1.177   .2448 -.20535714E-01
 DR3     -.7792177815E-03      .11072916    -.007   .9944 -.11607143E-01
 (Note: E+nn or E-nn means multiply by 10 to + or -nn power.)

--> wald;fn1=b_r1-1$
+-----------------------------------------------+
| WALD procedure. Estimates and standard errors |
| for nonlinear functions and joint test of     |
| nonlinear restrictions.                       |
| Wald Statistic            =       .51796     |
| Prob. from Chi-squared[ 1] =       .47171     |
+-----------------------------------------------+
+---------+--------------+----------------+--------+---------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] |
+---------+--------------+----------------+--------+---------+
 Fncn(1)  -.3467725900E-01  .48183346E-01    -.720   .4717
```

Unit root hypothesis is definitely not rejected.

3. Using the macroeconomic data in Appendix Table F5.1, estimate by least squares the parameters of the model

$$c_t = \beta_0 + \beta_1 y_t + \beta_2 c_{t-1} + \beta_3 c_{t-2} + \varepsilon_t,$$

where c_t is the log of real consumption and y_t is the log of real disposable income.

a. Use the Breusch and Pagan test to examine the residuals for autocorrelation.

b. Is the estimated equation stable? What is the characteristic equation for the autoregressive part of this model? What are the roots of the characteristic equation, using your estimated parameters?

c. What is your implied estimate of the short run (impact) multiplier for change in y_t on c_t? Compute the estimated long run multiplier.

```
--> samp;1-204$
--> crea;c=log(realcons);y=log(realdpi)$
--> crea;c1=c[-1];c2=c[-2]$
--> samp;3-204$
--> regr;lhs=c;rhs=one,y,c1,c2$
+----------------------------------------------------------------------+
| Ordinary    least squares regression    Weighting variable = none     |
| Dep. var. = C        Mean=   7.889033683    , S.D.=   .5102401315     |
| Model size: Observations =     202, Parameters =   4, Deg.Fr.=    198 |
| Residuals:  Sum of squares= .1519097328E-01, Std.Dev.=         .00876 |
| Fit:        R-squared=  .999710, Adjusted R-squared =          .99971 |
| Model test: F[  3,    198] =********,    Prob value =          .00000 |
| Diagnostic: Log-L =    672.4019, Restricted(b=0) Log-L =    -150.2038 |
|             LogAmemiyaPrCrt.=   -9.456, Akaike Info. Crt.=     -6.618 |
| Autocorrel: Durbin-Watson Statistic =   1.89384,   Rho =       .05308 |
+----------------------------------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |t-ratio |P[|T|>t] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant  .8165780259E-03  .10779352E-01     .076   .9397
 Y         .7869591065E-01  .29020268E-01    2.712   .0073     7.9998985
 C1          .9680839747    .72732869E-01   13.310   .0000     7.8802520
 C2       -.4701660339E-01  .70076193E-01    -.671   .5030     7.8714299
--> crea;e1=e[-1];e2=e[-3];e3=e[-3]$
--> crea;e1=e[-1];e2=e[-2];e3=e[-3]$
--> regr;lhs=e;rhs=one,e1,e2,e3$
+----------------------------------------------------------------------+
| Ordinary    least squares regression    Weighting variable = none     |
| Dep. var. = E        Mean=  -.6947138134E-15, S.D.=   .8693502258E-02 |
| Model size: Observations =     202, Parameters =   4, Deg.Fr.=    198 |
| Residuals:  Sum of squares= .1339943625E-01, Std.Dev.=         .00823 |
| Fit:        R-squared=  .117934, Adjusted R-squared =          .10457 |
| Model test: F[  3,    198] =     8.82,    Prob value =          .00002 |
| Diagnostic: Log-L =    685.0763, Restricted(b=0) Log-L =     672.4019 |
|             LogAmemiyaPrCrt.=   -9.581, Akaike Info. Crt.=     -6.743 |
| Autocorrel: Durbin-Watson Statistic =   1.85371,   Rho =       .07314 |
+----------------------------------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |t-ratio |P[|T|>t] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant  .2437121418E-04  .57884755E-03     .042   .9665
 E1       -.2553462753E-01  .70917392E-01    -.360   .7192 -.21497022E-04
 E2          .3385045374    .66904365E-01    5.060   .0000 -.56959898E-04
 E3        .6894158132E-01  .71101163E-01     .970   .3334 -.81793147E-04
--> calc;list;chisq=n*rsqrd$
    CHISQ    =   .23822731697405480D+02
--> matrix ; g=[.968083974,-.04701660339];root(g)$
```

The two roots of the matrix of the characteristic equation are 1.0688 and 19.8378. Since the smallest is larger than one, the equation is stable.

The short run multiplier is $\beta = .07869$. Long run is $\beta/(1-\gamma_1 - \gamma_2) = 12.669$. (Not very plausible.)

4. Verify the result in (20-10).

The model is $y_t = \mu + \gamma_1 y_{t-1} + \gamma_2 y_{t-2} + \varepsilon_t - \theta\varepsilon_{t-1}$. Write the MA(1) disturbance as $(1 - \theta L)\varepsilon_t$ where L is the lag operator. Now, divide both sides of the equation by $(1 - \theta L)$. This produces

$$y_t/(1 - \theta L) = \mu/(1 - \theta L) + \gamma_1 y_{t-1}/(1 - \theta L) + \gamma_2 y_{t-2}/(1 - \theta L) + \varepsilon_t.$$

Recall $1/(1 - \theta L) = 1 + \theta L + \theta^2 L^2 + ...$ Multiply out all the terms and assemble the sums to obtain

$$\sum_{i=0}^{\infty} \theta^i y_{t-i} = \frac{\mu}{1-\theta} + \gamma_1 \sum_{i=1}^{\infty} \theta^i y_{t-i} + \gamma_2 \sum_{i=2}^{\infty} \theta^i y_{t-i} + \varepsilon_t$$

By expanding the sums and collecting the term in the respective lags of y_t, we find the coefficients for the first several lags are $(\gamma_1 - \theta)$ for lag 1, $(\gamma_1\theta + \gamma_2 - \theta^2)$ for lag 2, $(\gamma_1\theta^2 + \gamma_2\theta - \theta^3)$ for lag 3, and so on. This is the pattern suggested in the text. The constant term is obvious, as given.

5. Show the Yule-Walker equations for an ARMA(1,1) process.

These are given on page 616 of the text.

6. Carry out an ADF test for a unit root in the rate of inflation using the subset of the data in Table F5.1 since 1974I. (This is the first quarter after the oil shock of 1973.)

To carry out the test, the rate of inflation is regressed on a constant, a time trend, the previous year's value of the rate of inflation, and three lags of the first difference. The test statistic for the ADF is (0.7290534455-1)/.011230759 = -2.373. The critical value in the lower part of Table 20.4 with about 100 observations is -3.45. Since our value is large than this, it follows that the hypothesis of a unit root cannot be rejected.

```
--> samp;1-204$
--> crea;ddp1=infl[-1]-infl[-2]$
--> crea;ddp2=ddp1[-1]$
--> crea;ddp3=ddp1[-2]$
--> crea;dp=infl[-1]$
--> samp;97-204$
--> crea;t=trn(1,1)$
--> regr;lhs=infl;rhs=one,t,dp,ddp1,ddp2,ddp3$
+-----------------------------------------------------------------------+
| Ordinary    least squares regression    Weighting variable = none     |
| Dep. var. = INFL     Mean=   4.907672727    , S.D.=   3.617392978     |
| Model size: Observations =     108, Parameters =   6, Deg.Fr.=    102 |
| Residuals:  Sum of squares= 608.5020156    , Std.Dev.=        2.44248 |
| Fit:        R-squared=  .565403, Adjusted R-squared =          .54410 |
| Model test: F[  5,    102] =   26.54,    Prob value =          .00000 |
+-----------------------------------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |t-ratio |P[|T|>t] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant     2.226039717       1.1342702    1.963   .0524
 T        -.1836785769E-01  .11230759E-01   -1.635   .1050     54.500000
 DP           .7290534455       .11419140    6.384   .0000     4.9830886
 DDP1        -.4744389916       .12707255   -3.734   .0003 -.58569323E-01
 DDP2        -.4273030624       .11563482   -3.695   .0004 -.46827528E-01
 DDP3        -.2248432703   .98954483E-01   -2.272   .0252 -.86558444E-02
--> wald;fn1=b_dp-1$
+---------+--------------+----------------+--------+---------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] |
+---------+--------------+----------------+--------+---------+
 Fncn(1)     -.2709465545       .11419140   -2.373   .0177
```

7. Estimate the parameters of the model in Example 15.1 using two stage least squares. Obtain the residuals from the two equations. Do these residuals appear to be white noise series? Based on your findings, what do you conclude about the specification of the model?

The two stage least squares regressions are shown below. Box-Jenkins analyses of the residuals shows fairly small, but significant autocorrelation of both sets of residuals. Thus, the specification may not be quite complete. There may be missing lags or other variables.

```
--> samp;1-204$
--> crea;ct=realcons;yt=realgdp;gt=realgovt;rt=tbilrate$
--> crea;ct1=ct[-1];yt1=yt[-1]$
--> samp;2-204$
--> samp;1-204$
--> crea;ct=realcons;yt=realgdp;gt=realgovt;rt=tbilrate;it=realinvs$
--> crea;ct1=ct[-1];yt1=yt[-1]$
--> crea;dy=yt-yt1$
--> samp;2-204$
--> name;x=one,rt,ct1,yt1,gt$
--> 2sls;lhs=ct;rhs=one,yt,ct1;inst=x;res=ec$
--> 2sls;lhs=it;rhs=one,rt,dy;inst=x;res=ei$
--> iden;rhs=ec;pds=10$
--> iden;rhs=ei;pds=10$
```

```
+-----------------------------------------------------------------------+
| Two stage   least squares regression    Weighting variable = none     |
| Dep. var. = CT       Mean=   3008.995074    , S.D.=   1456.900152     |
| Model size: Observations =     203, Parameters =   3, Deg.Fr.=    200 |
| Residuals:  Sum of squares= 96595.67529    , Std.Dev.=       21.97677 |
| Fit:        R-squared=  .999771, Adjusted R-squared =          .99977 |
|             (Note:  Not using OLS.  R-squared is not bounded in [0,1] |
| Model test: F[  2,    200] =********,    Prob value =          .00000 |
| Diagnostic: Log-L =   -913.8005, Restricted(b=0) Log-L =   -1766.2087 |
|             LogAmemiyaPrCrt.=    6.195, Akaike Info. Crt.=      9.033 |
| Autocorrel: Durbin-Watson Statistic =   1.61078,   Rho =       .19461 |
+-----------------------------------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant     6.666079115       8.6211817     .773   .4394
 YT       -.2932041745E-01  .35260653E-01    -.832   .4057     4577.1882
 CT1          1.051478712   .51482187E-01   20.424   .0000     2982.9744
 (Note: E+nn or E-nn means multiply by 10 to + or -nn power.)
```

```
+-----------------------------------------------------------------------+
| Two stage   least squares regression    Weighting variable = none     |
| Dep. var. = IT       Mean=   654.5295567    , S.D.=   391.3705005     |
| Model size: Observations =     203, Parameters =   3, Deg.Fr.=    200 |
| Residuals:  Sum of squares= 54658669.31    , Std.Dev.=      522.77466 |
| Fit:        R-squared= -.793071, Adjusted R-squared =         -.81100 |
|             (Note:  Not using OLS.  R-squared is not bounded in [0,1] |
| Diagnostic: Log-L =  -1557.1409, Restricted(b=0) Log-L =   -1499.3832 |
|             LogAmemiyaPrCrt.=   12.533, Akaike Info. Crt.=     15.371 |
| Autocorrel: Durbin-Watson Statistic =   1.49055,   Rho =       .25473 |
+-----------------------------------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant    -141.8297176       103.57113   -1.369   .1709
 RT           52.04340559       12.971223    4.012   .0001     5.2499007
 DY           13.80361384       1.7499250    7.888   .0000     37.898522
```

```
Time series identification for EC
Box-Pierce Statistic =    40.8498      Box-Ljung Statistic  =    41.7842
Degrees of freedom   =        10       Degrees of freedom   =        10
Significance level   =      .0000      Significance level   =      .0000
* => |coefficient| > 2/sqrt(N) or > 95% significant.
PACF is computed using Yule-Walker equations.
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
Lag |  Autocorrelation Function     |Box/Prc|    Partial Autocorrelations   X
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
  1 | .194*|           |**          |  7.65*| .194*|            |**         X
  2 | .264*|           |***         | 21.82*| .236*|            |***        X
  3 | .273*|           |***         | 36.93*| .207*|            |**         X
  4 | .067 |           |*           | 37.85*|-.063 |          * |           X
  5 | .054 |           |*           | 38.44*|-.068 |          * |           X
  6 | .073 |           |*           | 39.52*| .018 |            |*          X
  7 | .009 |           |*           | 39.53*| .003 |            |*          X
  8 |-.078 |          *|            | 40.78*|-.109 |          * |           X
  9 | .019 |           |*           | 40.85*| .023 |            |*          X
 10 | .002 |           |*           | 40.85*| .050 |            |*          X
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
Time series identification for EI
Box-Pierce Statistic =    27.4753      Box-Ljung Statistic  =    28.3566
Degrees of freedom   =        10       Degrees of freedom   =        10
Significance level   =      .0022      Significance level   =      .0016
* => |coefficient| > 2/sqrt(N) or > 95% significant.
PACF is computed using Yule-Walker equations.
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
Lag |  Autocorrelation Function     |Box/Prc|    Partial Autocorrelations   X
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
  1 | .244*|           |***         | 12.13*| .244*|            |***        X
  2 | .143*|           |**          | 16.27*| .096 |            |*          X
  3 | .037 |           |*           | 16.55*|-.019 |          * |           X
  4 |-.001 |          *|            | 16.55*|-.017 |          * |           X
  5 |-.066 |          *|            | 17.42*|-.078 |          * |           X
  6 | .003 |           |*           | 17.43*| .043 |            |*          X
  7 |-.042 |          *|            | 17.79*|-.033 |          * |           X
  8 |-.107 |          *|            | 20.10*|-.107 |          * |           X
  9 | .108 |           |*           | 22.46*| .194*|            |**         X
 10 | .157*|           |**          | 27.48*| .142*|            |**         X
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
```

Chapter 21
Models for Discrete Choice

1. A binomial probability model is to be based on the following index function model:

$$y^* = \alpha + \beta d + \varepsilon$$
$$y = 1 \text{ if } y^* > 0,$$
$$y = 0 \text{ otherwise.}$$

The only regressor, d, is a dummy variable. The data consist of 100 observations that have the following:

$$\begin{array}{c|cc} d \backslash y & 0 & 1 \\ 0 & 24 & 28 \\ 1 & 32 & 26 \end{array}$$

. Obtain the maximum likelihood estimators of α and β and estimate the asymptotic standard errors of your estimates. Test the hypothesis that β equals zero by using a Wald test (asymptotic t test) and a likelihood ratio test. Use the probit model and then repeat, using the logit model. Do your results change? [**Hint:** Formulate the log-likelihood in terms of α and $\delta = \alpha+\beta$.]

The log-likelihood is

$\ln L = \Sigma_{0,0}\ln\text{Prob}[y=0,d=0] + \Sigma_{0,1}\ln\text{Prob}[y=0,d=1] + \Sigma_{1,0}\ln\text{Prob}[y=1,d=0] + \Sigma_{1,1}\ln\text{Prob}[y=1,d=1]$

where $\Sigma_{i,j}$ indicates the sum over observations for which $y = i$ and $d = j$. Since there are no other regressors, this reduces to $\ln L = 24\ln(1 - F(\alpha)) + 32\ln(1 - F(\delta)) + 28\ln F(\alpha) + 16\ln F(\delta)$. Although it is straightforward to maximize the log-likelihood directly in terms of α and δ, an alternative, convenient approach is to estimate $F(\alpha)$ and $F(\delta)$. These functions can then be inverted to estimate the original parameters. The invariance of maximum likelihood estimators to transformation will justify this approach. One virtue of this approach is that the same procedure is used for both probit and logit models. Let $A = F(\alpha)$ and $D = F(\delta)$. Then, the log likelihood is simply $\ln L = 24\ln(1 - A) + 32\ln(1 - D) + 28\ln A + 16\ln D$. The necessary conditions are

$$\partial\ln L/\partial A = -24/(1 - A) + 28/A = 0$$
$$\partial\ln L/\partial D = -32/(1 - D) + 16/D = 0.$$

Simple manipulations produce the two solutions $A = 28/(24+28) = .539$ and $D = 16/(32+16) = .333$. Then, these functions can be inverted to produce the MLEs of α and β. Thus, $\hat{\alpha} = F^{-1}(A)$ and $\hat{\beta} = F^{-1}(D) - \hat{\alpha}$. The two inverse functions are $\Phi^{-1}(A)$ for the probit model, which must be approximated, and $\ln[F/(1-F)]$ for the logit model. The estimates are,

	Probit	Logit
α	.098	.156
δ	-.431	-.694
β	-.529	-.850

(Notice the proportionality relationship, .156/.098 = 1.592 and -.848/-.529 = 1.607.)

We will compute the asymptotic covariance matrix for $\hat{\alpha}$ and $\hat{\beta}$ directly using (19-24) for the probit model and (19-22) for the logit model. We will require $h_i = \partial^2\ln L/\partial(\alpha+\beta d)^2$ for the four cells. For the computation, we will require $\phi(c)/\Phi(c)$ and $-\phi(c)/[1-\Phi(c)]$. These are listed in the table below.

y	d	$\alpha+\beta d$	Φ	ϕ	λ_1 ϕ/Φ	λ_0 $-\phi/(1-\Phi)$	$\lambda_0\lambda_1$
0	0	.098	.539	.397	.737	-.861	-.636
1	0	.098	.539	.397	.737	-.861	-.636
0	1	-.431	.333	.364	1.093	-.546	-.597
1	1	-.431	.333	.364	1.093	-.546	-.597

The estimated asymptotic covariance matrix is the inverse of the estimate of $-E[\mathbf{H}]$.

$$-\hat{\mathbf{H}} = 24(.636)\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix} + 28(.636)\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix} + 32(.597)\begin{bmatrix}1 & 1\\1 & 1\end{bmatrix} + 16(.597)\begin{bmatrix}1 & 1\\1 & 1\end{bmatrix}.$$ Then,

$$\left[-\hat{\mathbf{H}}\right]^{-1} = \begin{bmatrix} 61.728 & 28.656 \\ 28.656 & 28.656 \end{bmatrix}^{-1} = \begin{bmatrix} .03024 & -.03024 \\ -.03024 & .06513 \end{bmatrix}.$$ The asymptotic standard errors are the square roots of the diagonal elements, which are .1739 and .2552, respectively. To test the hypothesis that $\beta = 0$, we would refer $z = -.529 / .2552 = -2.073$ to the standard normal table. This is larger than the 1.96 critical value, so we would reject the hypothesis. To compute the likelihood ratio statistic, we will require the two log-likelihoods. The restricted log-likelihood (for both the probit and logit models) is given in (19-28). This would be $\ln L_0 = 100[.44\ln.44 + .56\ln.56] = -68.593$. Let the predicted values above be denoted

P_{00} = Prob[y=0,d=0] = .461 (i.e., 1 - .539)
P_{10} = Prob[y=1,d=0] = .539
P_{01} = Prob[y=0,d=1] = .667
P_{11} = Prob[y=0,d=1] = .333

and let n_{ij} equal the number of observations in each cell Then, the unrestricted log-likelihood is $\ln L = 24\ln.461 + 28\ln.539 + 32\ln.667 + 16\ln.333 = -66.442$. The likelihood ratio statistic would be $\lambda = -2(-66.6442 - (-68.593)) = 4.302$. The critical value from the chi-squared distribution with one degree of freedom is 3.84, so once again, the test statistic is slightly larger than the table value.

We now compute the Hessian for the logit model. The predicted probabilities are

Prob[y = 0 , d = 0] = P_{00} = $1/(1 + e^{.156})$ = .462
Prob[y = 1 , d = 0] = P_{10} = $1 - P_{00}$ = .538
Prob[y = 0 , d = 1] = P_{01} = $1/(1 + e^{-.431})$ = .667
Prob[y = 1 , d = 1] = P_{11} = $1 - P_{01}$ = .333.

Notice that in spite of the quite different coefficients, these are identical to the results for the probit model. Remember that we originally estimated the probabilities, not the parameters, and these were independent of the distribution. Then, the Hessian is computed in the same manner as for the probit model using $h_{ij} = F_{ij}(1-F_{ij})$ instead of $\lambda_0\lambda_1$ in each cell. The asymptotic covariance matrix is the inverse of

$$(28+24)(.462)(.538)\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + (32+16)(.667)(.333)\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$ The standard errors are .2782 and .4137. For testing the hypothesis that β equals zero, the t-statistic is $z = -.850/.4137 = -2.055$, which is almost the same as that for the probit model. The unrestricted log-likelihood is $\ln L = 24\ln.4285 + ... + 16\ln.3635 = -66.442$ (again). The chi-squared statistic is 4.302, as before.

2. Suppose that a linear probability model is to be fit to a set of observations on a dependent variable, y, which takes values zero and one, and a single regressor, x, which varies continuously across observations. Obtain the exact expressions for the least squares slope in the regression in terms of the mean(s) and variance of x and interpret the result.

Using the usual regression statistics, we would have $a = \bar{y} - b\bar{x}$, $b = \Sigma_i (x_i - \bar{x})(y_i - \bar{y}) / \Sigma_i (x_i - \bar{x})^2$.

For data in which y is a binary variable, we can decompose the numerator somewhat further. First, divide both numerator and denominator by the sample size. Second, since only one variable need be in deviation form, drop the deviation in x. That leaves $b = \left[\Sigma_i x_i (y_i - \bar{y}) / n\right] / \left[\Sigma_i (x_i - \bar{x})^2 / n\right]$. The denominator is the sample variance of x. Since y_i is only 0s and 1s, $\bar{y}$ is the proportion of 1s in the sample, P. Thus, the numerator is $(1/n)\Sigma_i x_i y_i - (1/n)\Sigma_i x_i \bar{y} = (1/n)\Sigma_1 x_i - P\bar{x} = (n_1/n)\bar{x}_1 - P[P\bar{x} + (1-P)\bar{x}_0] = P(1 - P)(\bar{x}_1 - \bar{x}_0)$.

Therefore, the regression is essentially measuring how much the mean of x varies across the two groups of observations. The constant term does not simplify into any intuitively useful form.

3. Given the following data set:

Y	1	0	0	1	1	0	0	1	1	1
X	9	2	5	4	6	7	3	5	2	6

Estimate a probit model, and test the hypothesis that X is not influential in determining the probability that Y equals one.

The model was estimated using Newton's method as described in the text. The estimated coefficients and their standard are shown below:

$$\hat{y} = -.51274 + .15964X$$
$$(1.042) \quad (.202)$$

Log-likelihood = -6.403 Restricted log-likelihood = -6.9315.

The t-ratio for testing the hypothesis is .15964/.202 = .79. The chi-squared for the likelihood ratio test is 1.057. Neither is large enough to lead to rejection of the hypothesis.

4. Construct the Lagrange multiplier statistic for testing the hypothesis that all of the slopes (but not the constant term) equal zero in the binomial logit model. Prove that the Lagrange multiplier statistic is nR^2 in the regression of $(y_i - P)$ on the xs, where P is the sample proportion of ones.

The derivatives of the log-likelihood are given in (19-19). If all coefficients except the constant term are zero, then the first order condition for maximizing the log-likelihood would be $\partial \ln L/\partial\beta = \Sigma_i(y_i - \lambda)(1) = 0$ since with no regressors, λ_i will not vary with i. This leads to the constrained maximum $\hat{\lambda} = \Sigma_i y_i/n = P$, which might be expected. Thus, we estimate the constant term such that $P = \dfrac{e^{\hat{\alpha}}}{1+e^{\hat{\alpha}}}$, or $\hat{\alpha} = \text{logit}(P)$.

The LM statistic based on the BHHH estimator of the covariance matrix of the first derivatives would be

$$\text{LM} = [\Sigma_i\mathbf{g}_i]'[\Sigma_i\mathbf{g}_i\mathbf{g}_i']^{-1}[\Sigma_i\mathbf{g}_i] \quad \text{where} \quad \mathbf{g}_i = \Sigma_i(y_i - P)\mathbf{x}_i.$$

In full, the statistic is $\quad \text{LM} = [\Sigma_i(y_i - P)\mathbf{x}_i]'[\Sigma_i(y_i - P)^2\mathbf{x}_i\mathbf{x}_i']^{-1}[\Sigma_i(y_i - P)\mathbf{x}_i]$.

The actual (and expected) Hessian can be used instead by replacing $(y_i - P)^2$ with $P(1 - P)$ in the inverse matrix. The statistic could then be written

$$\text{LM} = [\mathbf{X}'(\mathbf{y} - P\mathbf{i})]'[(\mathbf{X}'\mathbf{X})^{-1}][\mathbf{X}'(\mathbf{y} - P\mathbf{i})]/P(1 - P) = \mathbf{e}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}/P(1 - P)$$

In the preceding, $\mathbf{e}'\mathbf{e} = \Sigma_i(y_i - P)^2 = nP(1 - P)$. Therefore, $\text{LM} = n[\mathbf{e}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}/\mathbf{e}'\mathbf{e}]$, which establishes the result. To compute the statistic, we regress $(y_i - P)$ on the $\mathbf{x}$s, then carry nR^2 into the chi-squared table.

5. We are interested in the ordered probit model. Our data consist of 250 observations, of which the responses are

Y	0	1	2	3	4
n	50	40	45	80	35

Using the data above, obtain maximum likelihood estimates of the unknown parameters of the model. [**Hint:** Consider the probabilities as the unknown parameters.]

Since there is no regressor, we may write the log-likelihood as

$$\begin{aligned}\ln L = \; & 50\ln\Phi(-\alpha) + 40\ln[\Phi(\mu_1-\alpha) - \Phi(-\alpha)] + 45\ln[\Phi(\mu_2-\alpha) - \Phi(\mu_1-\alpha)] + \\ & 80\ln[\Phi(\mu_3-\alpha) - \Phi(\mu_2-\alpha)] + 35\ln[1 - \Phi(\mu_3-\alpha)].\end{aligned}$$

There are four unknown parameters to estimate and four free probabilities. Suppose, then, we treat $\Phi(-\alpha)$, $\Phi(\mu_1-\alpha)$, $\Phi(\mu_2-\alpha)$, and $\Phi(\mu_3-\alpha)$ as the unknown parameters, π_0, π_1, π_2, and π_3, respectively. If we can find estimators of these, we can solve for the underlying parameters. We may write the log-likelihood as

$$\ln L = 50\ln\pi 0 + 40\ln(\pi 1 - \pi 0) + 45\ln(\pi_2 - \pi_1) + 80\ln(\pi_3 - \pi_2) + 35\ln(1 - \pi_3).$$

The necessary conditions are

$$\begin{aligned}\partial\ln L/\partial\pi_0 &= 50/\pi_0 - 40/(\pi_1-\pi_0) &= 0\\ \partial\ln L/\partial\pi_1 &= 40/(\pi_1 - \pi_0) - 45/(\pi_2 - \pi_1) &= 0\\ \partial\ln L/\partial\pi_2 &= 45/(\pi_2 - \pi_1) - 80/(\pi_3 - \pi_2) &= 0\\ \partial\ln L/\partial\pi_3 &= 80/(\pi_3 - \pi_2) - 35/(1 - \pi_3) &= 0.\end{aligned}$$

By a simple rearrangement, these can be recast as a set of linear equations. Thus,

$$\begin{aligned}
90\pi_0 - 50\pi_1 \quad &= 0\\
45\pi_0 - 85\pi_1 + 40\pi_2 \quad &= 0\\
80\pi_1 - 125\pi_2 + 45\pi_3 \quad &= 0\\
- 35\pi_2 + 115\pi_3 \quad &= 80
\end{aligned}$$

or

$$\begin{bmatrix} 90 & -50 & 0 & 00 \\ 45 & -85 & 40 & 0 \\ 0 & 80 & -125 & 45 \\ 0 & 0 & -35 & 115 \end{bmatrix} \begin{bmatrix} \pi_0 \\ \pi_1 \\ \pi_2 \\ \pi_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 80 \end{bmatrix}$$

The solution (as might be expected) is

$$\begin{aligned}
\pi_0 &= .2 \quad (50/250)\\
\pi_1 &= .36 \quad ((50+40)/250)\\
\pi_2 &= .54 \quad ((50+40+45)/250)\\
\pi_3 &= .86 \quad ((50+40+45+80)/250).
\end{aligned}$$

Now, we can solve for the underlying parameters.

$$\begin{aligned}
-\alpha &= \Phi^{-1}(.2) = -.841, \text{ so } \alpha = .841.\\
\mu_1-\alpha &= \Phi^{-1}(.36) = -.358, \text{ so } \mu_1 = .483\\
\mu_2-\alpha &= \Phi^{-1}(.54) = .101, \text{ so } \mu_2 = .942\\
\mu_3-\alpha &= \Phi^{-1}(.86) = 1.081, \text{ so } \mu_3 = 1.922.
\end{aligned}$$

6. The following hypothetical data give the participation rates in a particular type of recycling program and the number of trucks purchased for collection by ten towns in a small mid-Atlantic state:

Town	**Trucks**	**Participation**
1	160	11%
2	250	74%
3	170	8%
4	365	87%
5	210	62%
6	206	83%
7	203	48%
8	305	84%
9	270	71%
10	340	79%

The town of Eleven is contemplating initiating a recycling program but wishes to achieve a 95% rate of participation. Using a probit model for your analysis,

(a) How many trucks would the town expect to have to purchase in order to achieve their goal? [**Hint:** See Section 19.4.3.] Note that you will use $n_i = 1$.

(b) If trucks cost $20,000 each, is a goal of 90% reachable within a budget of $6.5 million? (That is, should they *expect* to reach the goal.)

(c) According to your model, what is the marginal value of the 301st truck in terms of the increase in the percentage participation?

To estimate the coefficients, we will use a two step FGLS procedure. Ordinary least squares estimates based on Section 19.4.3 are consistent, but inefficient. The OLS regression produces

$$\Phi^{-1}(P_i) = \hat{z}_i = \underset{(.7404)}{-2.18098} + \underset{(.002883)}{.0098898}T.$$

The predicted values from this regression can then be used to compute the weights in (21-39). The weighted least squares regression produces

$$\hat{z}_i = \underset{(.8103)}{-2.3116} + \underset{(.003322)}{.010646}T$$

In order to achieve a predicted proportion of 95%, we will require $z_i = 1.645$. The T required to achieve this is

$$T = (1.645 + 2.3116) / .010646 = 372.$$

The z_i which corresponds to 90% is 1.282. Doing the same calculation as above, this requires T = 338 trucks. At \$20,000 per truck, this requires \$6.751 million, so the budget is inadequate.

The marginal effect is $\partial\Phi_i/\partial T = .010646\phi(z_i)$. At $T = 300$, $z_i = .8822$, so $\phi(z_i) = .2703$ and the marginal effect is .00288.

7. A data set consists of $n = n_1 + n_2 + n_3$ observations on y and x. For the first n_1 observations, $y = 1$ and $x = 1$. For the next n_2 observations, $y = 0$ and $x = 1$. For the last n_3 observations, $y = 0$ and $x = 0$. Prove that neither (19-19) nor (19-21) has a solution.

This is similar to Exercise 1. It is simplest to prove it in that framework. Since the model has only a dummy variable, we can use the same log likelihood as in Exercise 1. But, in this exercise, there are no observations in the cell (y=1,x=0). The resulting log likelihood is, therefore,

$$\ln L = \Sigma_{0,0}\ln\text{Prob}[y=0,x=0] + \Sigma_{0,1}\ln\text{Prob}[y=0,x=1] + \Sigma_{1,1}\ln\text{Prob}[y=1,x=1]$$

or

$$\ln L = n_3\ln\text{Prob}[y=0,x=0] + n_2\ln\text{Prob}[y=0,x=1] + n_1\ln\text{Prob}[y=1,x=1].$$

Now, let $\delta = \alpha + \beta$. The log likelihood function is $\ln L = n_3\ln(1 - F(\alpha)) + n_2\ln(1 - F(\delta)) + n_1\ln F(\delta)$. For estimation, let $A = F(\alpha)$ and $D = F(\delta)$. We can estimate A and D, then $\alpha = F^{-1}(A)$ and $\beta = F^{-1}(D) - \alpha$. The first order condition for estimation of A is $\partial\ln L/\partial A = -n_3/(1 - A) = 0$, which obviously has no solution. If A cannot be estimated then α cannot either, nor, in turn, can β. This applies to both probit and logit models.

8. Data on t = strike duration and x = unanticipated industrial production for a number of strikes in each of 9 years are given in Table 22.7. Use the Poisson regressin model discussed in Section 21.8 to determine whether x is a significant determinant of the *number of strikes* in a given year.

Maximum likelihood estimates of the Poisson regression parameters are given below.

$$\ln\hat{\lambda} = \underset{(.1299)}{1.90854} + \underset{(2.51307)}{5.16577}x$$

The log-likelihood function at the maximum likelihood estimates is -28.993171. For the model with only a constant term, the value is -31.19884. The t statistic for testing the hypothesis that β equals zero is 5.16577/2.51307 = 2.056. This is a bit larger than the critical value of 1.96, though our use of the asymptotic distribution for a sample of 10 observations might be a bit optimistic. The chi squared value for the likelihood ratio test is 4.411, which is larger than the 95% critical value of 3.84, so the hypothesis that β equals zero is rejected on the basis of these two tests.

9. Asymptotics. Explore whether averaging individual marginal effects gives the same answer as computing the marginal effect at the mean.

In general, the conditional mean function in the discrete choice models is of the form $E[y|\mathbf{x}] = g(\mathbf{x},\beta)$ where g is a smooth and continuous function of both x and the parameters, β. Thus, the marginal effect is $\partial g(\mathbf{x},\beta)/\partial x = \mathbf{h}(\mathbf{x},\beta)$ which we have to assume is a set of also smooth and continuous functions in $\mathbf{x}$ and β. The question then is whether evaluating $\mathbf{h}(\mathbf{x},\beta)$ at the mean of the xs gives the same answer as averaging the sample values of $\mathbf{h}(\mathbf{x},\beta)$ each evaluated at the individual data points. The answer certainly is no in a finite sample since $\mathbf{h}(\mathbf{x},\beta)$ is nonlinear. Do they converge to the same thing? Suppose we assume that the data are well behaved so that the sample mean of the $\mathbf{x}$s, $\bar{\mathbf{x}}$ converges to a true mean vector, μ. Then, the question is whether $\mathbf{h}(\bar{\mathbf{x}},\beta)$ converges to the same thing as $(1/n)\Sigma\mathbf{h}(\mathbf{x}_i,\beta)$. Since $\mathbf{h}(\mathbf{x}_i,\beta)$ is continuous and smooth, $\mathbf{h}(\bar{\mathbf{x}},\beta)$ converges to $\mathbf{h}(\mu,\beta)$ by the Slutsky theorem. Write each term in the average marginal effect as a linear Taylor series, and use the mean value theorem, so $\mathbf{h}(\mathbf{x}_i,\beta) = \mathbf{h}(\mu,\beta) + \mathbf{H}(\mu^*,\beta)(\mathbf{x}_i-\mu)$ where μ^* is a point somewhere between $\mathbf{x}$ and μ and $\mathbf{H}$ is the second derivatives matrix. Presumably, $\mathbf{H}(\mu^*,\beta)$ is a matrix of constants. When we average $\mathbf{h}(\mathbf{x}_i,\beta)$, the first term is constant and the second term converges to zero by our assumption of well behaved data. So, at least in large samples, the answer is yes.

10. Prove (21-28). We'll do this more generally for any model $F(\alpha)$. Since the 'model' contains only a constant, the log likelihood is $\log L = \Sigma_0 \log[1-F(\alpha)] + \Sigma_1 \log F(\alpha) = n_0 \log[1-F(\alpha)]+n_1 \log F(\alpha)$. The likelihood equation is $\partial \log L/\partial\alpha = \Sigma_0[-f(\alpha)/[1-F(\alpha)] + \Sigma_1 f(\alpha)/F(\alpha) = 0$ where $f(\alpha)$ is the density (derivative of $F(\alpha)$ so that at the solution, $n_0 f(\alpha)/[1-F(\alpha)] = n_1 f(\alpha)/F(\alpha)$. Divide both sides of this equation by $f(\alpha)$ and solve it for $F(\alpha) = n_1/(n_0+n_1)$, as might be expected. You can then insert this solution for $F(\alpha)$ back into the log likelihood, and (21-28) follows immediately.

11. In the panel data models estimated in Example 21.5.1, neither the logit nor the probit model provides a framework for applying a Hausman test to determine whether fixed or random effects is preferred. Explain. (Hint: Unlike our application in the linear model, the incidental parameters problem persists here.) Look at the two cases. Neither case has an estimator which is consistent in both cases. In both cases, the unconditional fixed effects effects estimator is inconsistent, so the rest of the analysis falls apart. This is the incidental parameters problem at work. Note that the fixed effects estimator is inconsistent because in both models, the estimator of the constant terms is a function of 1/T. Certainly in both cases, if the fixed effects model is appropriate, then the random effects estimator is inconsistent, whereas if the random effects model is appropriate, the maximum likelihood random effects estimator is both consistent and efficient. Thus, in this instance, the random effects satisfies the requirements of the test. In fact, there does exist a consistent estimator for the logit model with fixed effects - see the text. However, this estimator must be based on a restricted sample observations with the sum of the ys equal to zero or T muust be discarded, so the mechanics of the Hausman test are problematic. This does not fall into the template of computations for the Hausman test.

Chapter 22

Limited Dependent Variable and Duration Models

1. The following 20 observations are drawn from a censored normal distribution:
3.8396, 7.2040, .00000, .00000, 4.4132, 8.0230, 5.7971, 7.0828, .00000, .80260, 13.0670, 4.3211, .00000, 8.6801, 5.4571, .00000, 8.1021, .00000, 1.2526, 5.6015. The applicable model is

$$y_i^* = \mu + \varepsilon_i$$
$$y_i = y_i^* \text{ if } \mu + \varepsilon_i > 0, \ 0 \text{ otherwise.}$$
$$\varepsilon_i \sim N[0,\sigma^2].$$

All exercises in this section are based on the preceding.

The OLS estimator of μ in the context of this tobit model is simply the sample mean. Compute the mean of all 20 observations. Would you expect this estimator to over- or underestimate μ? If we consider only the nonzero observations, the truncated regression model applies. The sample mean of the nonlimit observations is the least squares estimator in this context. Compute it, then comment on whether this should be an overestimate or an underestimate of the true mean.

The sample mean of all 20 observations is 4.18222. For the 14 nonzero observations, the mean is (20/14)4.18222 = 5.9746. Both of these should overestimate μ. In the first case, all negative values have been transformed to zeroes. Therefore, if we had had the original data, our estimator would include the negative values as well as the positive ones. Since we have only the zeroes, instead, our estimator includes, for every negative y^* a number which is larger than the true y^*. This will inflate the estimate. Likewise, for the truncated mean, whereas a complete sample might include some negative values, the observed one will not. Once again, this will serve to inflate the estimator of the mean.

2. We now consider the tobit model that applies to the full data set.
(a) Formulate the log-likelihood for this very simple tobit model.
(b) Reformulate the log-likelihood in terms of $\theta = 1/\sigma$ and $\gamma = \mu/\sigma$. Then, derive the necessary conditions for maximizing the log-likelihood with respect to θ and γ.
(c) Discuss how you would obtain the values of θ and γ to solve the problem in part (b).
(d) Obtain the maximum likelihood estimators of μ and σ.

The log-likelihood for the Tobit model is given in (22-13). With only a constant term, this is

$$\ln L = (-n_1/2)[\ln(2\pi) + \ln\sigma^2] - (1/(2\sigma^2))\Sigma_1(y_i - \mu)^2 + \Sigma_0\ln\Phi(-\mu/\sigma)$$

In terms of γ and θ, this is
$$\ln L = (-n_1/2)[\ln(2\pi) - \ln\theta^2] - (1/2)\Sigma_1(\theta y_i - \gamma)^2 + \Sigma_0\ln\Phi(-\gamma)$$
$$= (-n_1/2)\ln(2\pi) + n_1\ln\theta - (1/2)\Sigma_1(\theta y_i - \gamma)^2 + \Sigma_0\ln\Phi(-\gamma).$$

The necessary conditions for maximizing this with respect to γ and θ are

$$\partial\ln L/\partial\gamma = \Sigma_1(\theta y_i - \gamma) - \Sigma_0\phi(-\gamma)/\Phi(-\gamma) = \theta\Sigma_1 y_i - n_1\gamma - n_0[\phi(-\gamma)/\Phi(\gamma)] = 0$$
$$\partial\ln L/\partial\theta = n_1/\theta - \Sigma_1 y_i(\theta y_i - \gamma) = n_1/\theta - \theta\Sigma_1 y_i^2 + \gamma\Sigma_1 y_i = 0.$$

There are a few different ways one might solve these two equations. A grid search over the values of γ and θ is a possibility. A direct maximum likelihood estimator for the tobit model is the simpler choice if one is available. The model with only a constant term is otherwise the same as the usual model. Using the data above, the tobit maximum likelihood estimates are $\hat{\mu} = 3.2731$, $\hat{\sigma} = 5.0303$.

3. Using only the nonlimit observations, repeat the Exercise 2 in the context of the truncated regression model. Estimate μ and σ by using the method of moments estimator outlined in Example 20.4. Compare your results to those in the previous problems.

The log-likelihood for the truncated regression is given in (20-9). With only a constant term,

$$\ln L = (-n/2)[\ln(2\pi) + \ln\sigma^2] - (1/(2\sigma^2))\Sigma_1(y_i - \mu)^2 - \Sigma_i \ln\Phi(\mu/\sigma)$$

Once again transforming to γ and σ, this is

$$\ln L = -(n/2)\ln(2\pi) + n\ln\theta - (1/2)\Sigma_i(\theta y_i - \gamma)^2 - n\ln\Phi(\gamma).$$

The necessary conditions for maximizing this are

$$\partial\ln L/\partial\gamma = \Sigma_i(\theta y_i - \gamma) - n\phi(\gamma)/\Phi(\gamma) = 0$$

$$\partial\ln L/\partial\theta = n/\theta - \Sigma_i y_i(\theta y_i - \gamma)$$

The first of the two equations can be $\bar{y} = \gamma/\theta + \lambda/\theta$, where $\lambda = \phi(\gamma)/\Phi(\gamma)$. Now, reverting back to μ and σ, this is

$\bar{y} = \mu + \sigma\lambda$ which is (20-5). The second equation can be manipulated to produce $\Sigma y_i^2/n - \mu\bar{y} = \sigma^2$. Once again, trial and error could be used to find a solution. As before, estimating the model as a truncated regression with only a constant term will also produce a solution. The solution by this method is $\hat{\mu} = 3.3439$, $\hat{\sigma} = 5.6368$.

With the data of the first problem, we would have the following: Estimated Prob[$y^* > 0$] = 14/20 = .7. This is an estimate of $\Phi(\mu/\sigma)$, so we would have $\mu/\sigma = \Phi^{-1}(.7) = .525$ or $\mu = .525\sigma$. Now, we can use the relationship

$E[y|y > 0] = \mu + \sigma\phi(\mu/\sigma)/\Phi(\mu/\sigma) = \mu + \sigma\lambda$. Since μ/σ is now known, we have $\lambda = \phi(.525) / \Phi(.525) = .496$ so a second equation is $5.9746 = \mu + .496\sigma$. The joint solution is $\hat{\mu} = 3.0697$, $\hat{\sigma} = 5.8470$. The three solutions are surprisingly close.

4. Continuing to use the data in Exercise 1, consider, once again, only the nonzero observations. Suppose that the sampling mechanism is as follows: y^* and another normally distributed random variable, z, have population correlation 0.7. The two variables, y^* and z are sampled jointly. When z is greater than zero, y is reported. When z is less than zero, both z and y^* are discarded. Exactly 35 draws were required in order to obtain the preceding sample. Estimate μ and σ. [**Hint:** Use Theorem 20.4.]

Using Theorem 21.4, we have $1 - \Phi(\alpha_z) = 14/35 = .4$, $\alpha_z = \Phi^{-1}(.6) = .253$, $\lambda(\alpha_z) = .9659$, $\delta(\alpha_z) = .6886$. The two moment equations are based on the mean and variance of y in the observed data, 5.9746 and 9.869, respectively. The equations would be $5.9746 = \mu + \sigma(.7)(.9659)$ and $9.869 = \sigma^2(1 - .7^2(.6886))$. The joint solution is $\hat{\mu} = 3.3651$, $\hat{\sigma} = 3.8594$.

5. Derive the marginal effects for the tobit model with heteroscedasticity that is described in Section 22.3.4.a.

The conditional mean function is $E[y|x] = \Phi(\beta'x_i/\sigma_i)\beta'x + \sigma_i\Phi(\beta'x_i/\sigma_i)$ using the equation before (22-12). Suppose that $\sigma_i = \sigma\exp(\alpha'x_i)$ for the same vector x_i. (We'll relax that assumption shortly.) Now, differentiate this expression with respect to x. We differentiate the two parts, first with respect to $\beta'x$ then with respect to σ_i.

$$\frac{\partial E[y_i|\mathbf{x}_i]}{\partial \mathbf{x}_i} = \Phi\left(\frac{\beta'\mathbf{x}_i}{\sigma_i}\right)\beta + \left(\beta'\mathbf{x}_i\right)\phi\left(\frac{\beta'\mathbf{x}_i}{\sigma_i}\right)\frac{1}{\sigma_i}\beta + \sigma_i\left[-\left(\frac{\beta'\mathbf{x}_i}{\sigma_i}\right)\phi\left(\frac{\beta'\mathbf{x}_i}{\sigma_i}\right)\right]\frac{1}{\sigma_i}\beta$$

$$+\left(\beta'\mathbf{x}_i\right)\phi\left(\frac{\beta'\mathbf{x}_i}{\sigma_i}\right)\left(\frac{-1}{\sigma_i}\right)\left(\frac{\beta'\mathbf{x}_i}{\sigma_i}\right)\sigma_i\alpha + \phi\left(\frac{\beta'\mathbf{x}_i}{\sigma_i}\right)\sigma_i\alpha + \sigma_i\left[-\left(\frac{\beta'\mathbf{x}_i}{\sigma_i}\right)\phi\left(\frac{\beta'\mathbf{x}_i}{\sigma_i}\right)\right]\left(\frac{-1}{\sigma_i}\right)\left(\frac{\beta'\mathbf{x}_i}{\sigma_i}\right)\sigma_i\alpha$$

After collecting the terms, we obtain $\partial E[y_i|x_i]/\partial x_i = \Phi(a_i)\beta + \sigma_i\phi(a_i)\alpha$ where $a_i = \beta'x_i/\sigma_i$. Thus, the marginal effect has two parts. one for β and one for α. Now, if a variable appears in σ_i but not in x_i, then only the second term appears while if a variable appears only in x_i and not in σ_i, then only the first term appears in the marginal effect.

6. Prove that the Hessian for the tobit model in (22-14) is negative definite after Olsen's transformation is applied to the parameters.

The transformed log likelihood function is

$$\log L = \Sigma_{y>0}(-1/2)[\log 2\pi - \log\theta^2 + (\theta y - x'\gamma)^2] + \Sigma_{y=0}\log[1-\Phi(x'\gamma)]$$

It will be convenient to define $a_i = x_i'\gamma$. Note also that $1 - \Phi(a_i) = \Phi(-a_i)$. The first derivatives and Hessian in the transformed parameters are

$$\frac{\partial \log L}{\partial \theta} = \sum\nolimits_{y_i>0} (1/\theta) - y_i\left(\theta y_i - a_i\right)$$

$$\frac{\partial \log L}{\partial \gamma} = \sum\nolimits_{y_i>0} \mathbf{x}_i\left(\theta y_i - a_i\right) + \sum\nolimits_{y_i=0} \left[\phi(-a_i)/\Phi(-a_i)\right](-\mathbf{x}_i)$$

$$\frac{\partial^2 \log L}{\partial \theta^2} = \sum\nolimits_{y_i>0} -1/\theta^2 - y_i^2$$

$$\frac{\partial^2 \log L}{\partial \gamma \partial \gamma'} = \sum\nolimits_{y_i>0} -\mathbf{x}_i\mathbf{x}_i' + \sum\nolimits_{y_i=0} -\left[\phi(-a_i)/\Phi(-a_i)\right]\{-a_i + \left[\phi(-a_i)/\Phi(-a_i)\right]\}\mathbf{x}_i\mathbf{x}_i'$$

$$\frac{\partial^2 \log L}{\partial \gamma \partial \theta} = \sum\nolimits_{y_i>0} -\mathbf{x}_i y_i$$

The second derivatives can be collected in a matrix format:

$$\frac{\partial \log L}{\partial\begin{pmatrix}\gamma\\ \theta\end{pmatrix}\partial\begin{pmatrix}\gamma\\ \theta\end{pmatrix}'} = \sum\nolimits_{y>0}\left[-\begin{pmatrix}\mathbf{x}_i\\ -y_i\end{pmatrix}\begin{pmatrix}\mathbf{x}_i\\ -y_i\end{pmatrix}' - \begin{pmatrix}0\\ \theta\end{pmatrix}\begin{pmatrix}0\\ \theta\end{pmatrix}'\right] + \sum\nolimits_{y=0}\delta_i\begin{pmatrix}\mathbf{x}_i\\ 0\end{pmatrix}\begin{pmatrix}\mathbf{x}_i\\ 0\end{pmatrix}'$$

where δ_i is the last scalar term in $\partial^2\log L/\partial\delta\partial\gamma'$. By Theorem 22.2 (see (22-4)), we know that δ_i is negative. Thus, all three parts of the matrix are negative semidefinite. Assuming the data are not linearly dependent and there are more than K observations, the Hessian will have full rank and be negative definite.

Appendix A

Matrix Algebra

1. For the matrices $\mathbf{A} = \begin{bmatrix} 1 & 3 & 3 \\ 2 & 4 & 1 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 2 & 4 \\ 1 & 5 \\ 6 & 2 \end{bmatrix}$ compute **AB**, **A′B′**, and **BA**.

$$\mathbf{AB} = \begin{bmatrix} 23 & 25 \\ 14 & 30 \end{bmatrix},\ \mathbf{BA} = \begin{bmatrix} 10 & 22 & 10 \\ 11 & 23 & 8 \\ 10 & 26 & 20 \end{bmatrix},\ \mathbf{A}'\mathbf{B}' = (\mathbf{BA})' = \begin{bmatrix} 10 & 11 & 10 \\ 22 & 23 & 26 \\ 10 & 8 & 20 \end{bmatrix}.$$

2. Prove that $tr(\mathbf{AB}) = tr(\mathbf{BA})$ where **A** and **B** are any two matrices that are conformable for both multiplications. They need not be square.

The ith diagonal element of **AB** is $\sum_j a_{ij}b_{ji}$. Summing over i produces $tr(\mathbf{AB}) = \sum_i \sum_i a_{ij}b_{ji}$. The jth diagonal element of **BA** is $\sum_j b_{ji}a_{ij}$. Summing over i produces $tr(\mathbf{BA}) = \sum_i \sum_j b_{ji}a_{ij}$.

3. Prove that $tr(\mathbf{A}'\mathbf{A}) = \sum_i \sum_j a_{ij}^2$.

The jth diagonal element of **A′A** is the inner product of the jth column of **A**, or $\sum_i a_{ij}^2$. Summing over j produces $tr(\mathbf{A}'\mathbf{A}) = \sum_j \sum_i a_{ij}^2 = \sum_i \sum_j a_{ij}^2$.

4. Expand the matrix product $\mathbf{X} = \{[\mathbf{AB} + (\mathbf{CD})'][(\mathbf{EF})^{-1} + \mathbf{GH}]\}'$. Assume that all matrices are square and **E** and **F** are nonsingular.

In parts, $(\mathbf{CD})' = \mathbf{D}'\mathbf{C}'$ and $(\mathbf{EF})^{-1} = \mathbf{F}^{-1}\mathbf{E}^{-1}$. Then, the product is

$$\begin{aligned}\{[\mathbf{AB} + (\mathbf{CD})'][(\mathbf{EF})^{-1} + \mathbf{GH}]\}' &= (\mathbf{ABF}^{-1}\mathbf{E}^{-1} + \mathbf{ABGH} + \mathbf{D}'\mathbf{C}'\mathbf{F}^{-1}\mathbf{E}^{-1} + \mathbf{D}'\mathbf{C}'\mathbf{GH})' \\ &= (\mathbf{E}^{-1})'(\mathbf{F}^{-1})'\mathbf{B}'\mathbf{A}' + \mathbf{H}'\mathbf{G}'\mathbf{B}'\mathbf{A}' + (\mathbf{E}^{-1})'(\mathbf{F}^{-1})'\mathbf{CD} + \mathbf{H}'\mathbf{G}'\mathbf{CD}.\end{aligned}$$

5. Prove for that for $K\times 1$ column vectors, $\mathbf{x}_i$ $i = 1,\ldots,n$, and some nonzero vector, **a**,

$$\sum_{i=1}^n (\mathbf{x}_i - \mathbf{a})(\mathbf{x}_i - \mathbf{a})' = \mathbf{X}'\mathbf{M}^0\mathbf{X} + n(\bar{\mathbf{x}} - \mathbf{a})(\bar{\mathbf{x}} - \mathbf{a})'.$$

Write $\mathbf{x}_i - \mathbf{a}$ as $[(\mathbf{x}_i - \bar{\mathbf{x}}) + (\bar{\mathbf{x}} - \mathbf{a})]$. Then, the sum is

$$\begin{aligned}\sum_{i=1}^n [(\mathbf{x}_i - \bar{\mathbf{x}}) + (\bar{\mathbf{x}} - \mathbf{a})]\,[(\mathbf{x}_i - \bar{\mathbf{x}}) + (\bar{\mathbf{x}} - \mathbf{a})]' = \\ \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})' + \sum_{i=1}^n (\bar{\mathbf{x}} - \mathbf{a})(\bar{\mathbf{x}} - \mathbf{a})' \\ + \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}} - \mathbf{a})' + \sum_{i=1}^n (\bar{\mathbf{x}} - \mathbf{a})(\mathbf{x}_i - \bar{\mathbf{x}})'\end{aligned}$$

Since $(\bar{\mathbf{x}} - \mathbf{a})$ is a vector of constants, it may be moved out of the summations. Thus, the fourth term is $(\bar{\mathbf{x}} - \mathbf{a})\sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})' = \mathbf{0}$. The third term is likewise. The first term is $\mathbf{X}'\mathbf{M}^0\mathbf{X}$ by the definition while the second is $n(\bar{\mathbf{x}} - \mathbf{a})(\bar{\mathbf{x}} - \mathbf{a})'$.

6. Let **A** be any square matrix whose columns are [$\mathbf{a}_1,\mathbf{a}_2,...,\mathbf{a}_M$] and let **B** be any rearrangement of the columns of the $M\times M$ identity matrix. What operation is performed by the multiplication **AB**? What about **BA**?

B is called a permutation matrix. Each column of **B**, say, $\mathbf{b}_i$, is a column of an identity matrix. The jth column of the matrix product **AB** is **A** $\mathbf{b}_i$ which is the jth column of **A**. Therefore, post multiplication of **A** by **B** simply rearranges (permutes) the columns of **A** (hence the name). Each row of the product **BA** is one of the rows of **A**, so the product **BA** is a rearrangement of the rows of **A**. Of course, **A** need not be square for us to permute its rows or columns. If not, the applicable permutation matrix will be of different orders for the rows and columns.

7. Consider the 3×3 case of the matrix **B** in Exercise 6. For example, $\mathbf{B} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ Compute $\mathbf{B}^2$ and $\mathbf{B}^3$. Repeat for a 4×4 matrix. Can you generalize your finding?

$$\mathbf{B}^2 = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \mathbf{B}^3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Since each power of **B** is a rearrangement of **I**, some power of **B** will equal **I**. If n is this power, we also find, therefore, that $\mathbf{B}^{n-1} = \mathbf{B}^{-1}$. This will hold generally.

8. Calculate |**A**|, tr(**A**) and $\mathbf{A}^{-1}$ for $\mathbf{A} = \begin{bmatrix} 1 & 4 & 7 \\ 3 & 2 & 5 \\ 5 & 2 & 8 \end{bmatrix}$.

$$|\mathbf{A}| = 1(2)(8)+4(5)(5)+3(2)(7)-5(2)(7)-1(5)(2)-3(4)(8) = -18,$$
$$tr(\mathbf{A}) = 1 + 2 + 8 = 11$$
$$\mathbf{A}^{-1} = \frac{-1}{18}\begin{bmatrix} \det\begin{pmatrix} 2 & 5 \\ 2 & 8 \end{pmatrix} & -\det\begin{pmatrix} 4 & 7 \\ 2 & 8 \end{pmatrix} & \det\begin{pmatrix} 4 & 7 \\ 2 & 5 \end{pmatrix} \\ -\det\begin{pmatrix} 3 & 5 \\ 5 & 8 \end{pmatrix} & \det\begin{pmatrix} 1 & 7 \\ 5 & 8 \end{pmatrix} & -\det\begin{pmatrix} 1 & 7 \\ 3 & 5 \end{pmatrix} \\ \det\begin{pmatrix} 3 & 2 \\ 5 & 2 \end{pmatrix} & -\det\begin{pmatrix} 1 & 4 \\ 5 & 2 \end{pmatrix} & \det\begin{pmatrix} 1 & 4 \\ 3 & 2 \end{pmatrix} \end{bmatrix} = \begin{bmatrix} -6/18 & 18/18 & -6/18 \\ -1/18 & 27/18 & -16/18 \\ 4/18 & -18/18 & 10/18 \end{bmatrix}.$$

9. Obtain the Cholesky decomposition of the matrix $\mathbf{A} = \begin{bmatrix} 25 & 7 \\ 7 & 13 \end{bmatrix}$.

Recall that the Cholesky decomposition of a matrix, **A**, is the matrix product **LU** = **A** where **L** is a lower triangular matrix and **U** = **L′**. Write the decomposition as $\begin{bmatrix} 25 & 7 \\ 7 & 13 \end{bmatrix} = \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & \lambda_{22} \end{bmatrix} \begin{bmatrix} \lambda_{11} & \lambda_{21} \\ 0 & \lambda_{22} \end{bmatrix}$. By direct multiplication, 25 = λ_{11}^2 so $\lambda_{11} = 5$. Then, $\lambda_{11}\lambda_{21} = 7$, so $\lambda_{21} = 7/5 = 1.4$. Finally, $\lambda_{21}^2 + \lambda_{22}^2 = 13$, so $\lambda_{22} = 3.322$.

10. A symmetric positive definite matrix, **A**, can also be written as **A** = **UL**, where **U** is an upper triangular matrix and **L** = **U′**. This is not the Cholesky decomposition, however. Obtain this decomposition of the matrix in Exercise 9.

Using the same logic as in the previous problem, $\begin{bmatrix} 25 & 7 \\ 7 & 13 \end{bmatrix}. = \begin{bmatrix} \mu_{11} & \mu_{12} \\ 0 & \mu_{22} \end{bmatrix}.\begin{bmatrix} \mu_{11} & 0 \\ \mu_{12} & \mu_{22} \end{bmatrix}.$ Working from the bottom up, $\mu_{22} = \sqrt{13}$ = 3.606. Then, 7 = $\mu_{12}\mu_{22}$ so μ_{12} = $7/\sqrt{13}$ = 1.941. Finally, 25 = $\mu_{11}^2 + \mu_{12}^2$ so μ_{11}^2 = 25 - 49/13 = 21.23, or μ_{11} = 4.61.

11. What operation is performed by postmultiplying a matrix by a diagonal matrix? What about premultiplication?

The columns are multiplied by the corresponding diagonal element. Premultiplication multiplies the rows by the corresponding diagonal element.

12. Are the following quadratic forms positive for all values of **x**?

(a) $y = x_1^2 - 28x_1x_2 + (11x_2^2)$,

(b) $y = 5x_1^2 + x_2^2 + 7x_3^2 + 4x_1x_2 + 6x_1x_3 + 8x_2x_3$?

The first may be written $\begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} 1 & -14 \\ -14 & 11 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$. The determinant of the matrix is 121 - 196 = -75, so it is not positive definite. Thus, the first quadratic form need not be positive. The second uses the matrix $\begin{bmatrix} 5 & 2 & 3 \\ 2 & 1 & 4 \\ 3 & 4 & 7 \end{bmatrix}$. There are several ways to check the definiteness of a matrix. One way is to check the signs of the principal minors, which must be positive. The first two are 5 and 5(1)-2(2)=1, but the third, the determinant, is -34. Therefore, the matrix is not positive definite. Its three characteristic roots are 11.1, 2.9, and -1. It follows, therefore, that there are values of x_1, x_2, and x_3 for which the quadratic form is negative.

13. Prove that $tr(\mathbf{A}\otimes\mathbf{B}) = tr(\mathbf{A})tr(\mathbf{B})$.

The jth diagonal block of the product is $a_{jj}\mathbf{B}$. Its ith diagonal element is $a_{jj}b_{ii}$. If we sum in the jth block, we obtain $\sum_i a_{jj}b_{ii} = a_{jj}\sum_i b_{ii}$. Summing down the diagonal blocks gives the trace, $\sum_j a_{jj}\sum_i b_{ii}$ = $tr(\mathbf{A})tr(\mathbf{B})$.

14. A matrix, **A**, is *nilpotent* if $\lim_{k\to\infty} \mathbf{A}^k = \mathbf{0}$. Prove that a necessary and sufficient condition for a symmetric matrix to be nilpotent is that all of its characteristic roots be less than one in absolute value. (For an application, see Section 17.3.3.)

Use the spectral decomposition to write **A** as $\mathbf{C\Lambda C'}$ where Λ is the diagonal matrix of characteristic roots. Then, the Kth power of **A** is $\mathbf{C\Lambda}^K\mathbf{C'}$. Sufficiency is obvious. Also, since if some λ is greater than one, Λ^K must explode, the condition is necessary as well.

15. Compute the characteristic roots of $\mathbf{A} = \begin{bmatrix} 2 & 4 & 3 \\ 4 & 8 & 6 \\ 3 & 6 & 5 \end{bmatrix}$.

The roots are determined by $|\mathbf{A} - \lambda\,\mathbf{I}| = 0$. For the matrix above, this is

$$|\mathbf{A} - \lambda\mathbf{I}| = (2-\lambda)(8-\lambda)(5-\lambda) + 72 + 72 - 9(8-\lambda) - 36(2-\lambda) - 16(5-\lambda)$$
$$= -\lambda^3 + 15\lambda^2 - 5\lambda = -\lambda(\lambda^2 - 15\lambda + 5) = 0.$$

One solution is obviously zero. (This might have been apparent. The second column of the matrix is twice the first, so it has rank no more than two, and therefore no more than two nonzero roots.) The other two roots are $\left(15 \pm \sqrt{205}\right)/2$ = .341 and 4.659.

16. Suppose $\mathbf{A} = \mathbf{A}(z)$ where z is a scalar. What is $\partial\mathbf{x}'\mathbf{A}\mathbf{x}/\partial z$? Now, suppose each element of $\mathbf{x}$ is also a function of z. Once again, what is $\partial\mathbf{x}'\mathbf{A}\mathbf{x}/\partial z$?

The quadratic form is $\sum_i \sum_j x_i x_j a_{ij}$, so

$\partial\mathbf{x}'\mathbf{A}(z)\mathbf{x}/\partial z = \sum_i \sum_j x_i x_j (\partial a_{ij} / \partial z) = \mathbf{x}'(\partial\mathbf{A}(z)/\partial z)\mathbf{x}$ where $\partial\mathbf{A}(z)/\partial z$ is a matrix of partial derivatives.

Now, if each element of $\mathbf{x}$ is also a function of z, then,

$$\partial\mathbf{x}'\mathbf{A}\mathbf{x}/\partial z = \sum_i \sum_j x_i x_j (\partial a_{ij} / \partial z) + \sum_i \sum_j (\partial x_i / \partial z) x_j a_{ij} + \sum_i \sum_j x_i (\partial x_j / \partial z) a_{ij}$$
$$= \mathbf{x}'(\partial\mathbf{A}(z)/\partial z)\mathbf{x} + (\partial\mathbf{x}(z)/\partial z)'\mathbf{A}(z)\mathbf{x}(z) + \mathbf{x}(z)'\mathbf{A}(z)(\partial\mathbf{x}(z)/\partial z)$$

If $\mathbf{A}$ is symmetric, this simplifies a bit to $\mathbf{x}'(\partial\mathbf{A}(z)/\partial z)\mathbf{x} + 2(\partial\mathbf{x}(z)/\partial z)'\mathbf{A}(z)\mathbf{x}(z)$.

17. Show that the solutions to the determinantal equations $|\mathbf{B} - \lambda\mathbf{A}| = 0$ and $|\mathbf{A}^{-1}\mathbf{B} - \lambda\mathbf{I}| = 0$ are the same. How do the solutions to this equation relate to those of the equation $|\mathbf{B}^{-1}\mathbf{A} - \mu\mathbf{I}| = 0$? (For an application of the first of these equations, see Section 16.5.2d.)

Since $\mathbf{A}$ is assumed to be nonsingular, we may write

$$\mathbf{B} - \lambda\mathbf{A} = \mathbf{A}(\mathbf{A}^{-1}\mathbf{B} - \lambda\mathbf{I}). \text{ Then, } |\mathbf{B} - \lambda\mathbf{A}| = |\mathbf{A}|\times|\mathbf{A}^{-1}\mathbf{B} - \lambda\mathbf{I}|.$$

The determinant of $\mathbf{A}$ is nonzero if $\mathbf{A}$ is nonsingular, so the solutions to the two determinantal equations must be the same. $\mathbf{B}^{-1}\mathbf{A}$ is the inverse of $\mathbf{A}^{-1}\mathbf{B}$, so its characteristic roots must be the reciprocals of those of $\mathbf{A}^{-1}\mathbf{B}$. There might seem to be a problem here since these two matrices need not be symmetric, so the roots could be complex. But, for the application noted, both $\mathbf{A}$ and $\mathbf{B}$ are symmetric and positive definite. As such, it can be shown (see Section 16.5.2d) that the solution is the same as that of a third determinantal equation involving a symmetric matrix.

18. Using the matrix $\mathbf{A}$ in Exercise 9, find the vector $\mathbf{x}$ that minimizes $y = \mathbf{x}'\mathbf{A}\mathbf{x} + 2x_1 + 3x_2 - 10$. What is the value of y at the minimum? Now, minimize y subject to the constraint $x_1 + x_2 = 1$. Compare the two solutions.

The solution which minimizes $y = \mathbf{x}'\mathbf{A}\mathbf{x} + \mathbf{b}'\mathbf{x} + d$ will satisfy $\partial y\partial\mathbf{x} = 2\mathbf{A}\mathbf{x} + \mathbf{b} = \mathbf{0}$. For this problem, $\mathbf{A} = \begin{bmatrix} 25 & 7 \\ 7 & 13 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$, and $\mathbf{A}^{-1} = \begin{bmatrix} 13/276 & -7/276 \\ -7/276 & 25/276 \end{bmatrix}$, so the solution is x_1 =-5/552 = -.0090597 and x_2 = -61/552 = -.110507.

The constrained maximization problem may be set up as a Lagrangean,

$L^* = \mathbf{x}'\mathbf{A}\mathbf{x} + \mathbf{b}'\mathbf{x} + d + \lambda(\mathbf{c}'\mathbf{x} - 1)$ where $\mathbf{c} = [1,1]'$. The necessary conditions for the solution are

$$\partial L^*/\partial\mathbf{x} = 2\mathbf{A}\mathbf{x} + \mathbf{b} + \lambda\mathbf{c} = \mathbf{0}$$
$$\partial L^*/\partial\lambda = \mathbf{c}'\mathbf{x} - 1 = 0,$$

or,

$$\begin{bmatrix} 2\mathbf{A} & \mathbf{c} \\ \mathbf{c}' & 0 \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \lambda \end{bmatrix} = \begin{bmatrix} -\mathbf{b} \\ 1 \end{bmatrix}.$$

Inserting **A, b,** and **c** produces the solution $\begin{bmatrix}50 & 14 & 1\\ 14 & 26 & 1\\ 1 & 1 & 0\end{bmatrix}\begin{bmatrix}x_1\\ x_2\\ \lambda\end{bmatrix} = \begin{bmatrix}-2\\ -3\\ 1\end{bmatrix}$. The solution to the three equations is obtained by premultiplying the vector on the right by the inverse of the matrix on the left. The solutions are 0.27083, 0.72917, and, -25.75. The function value at the constrained solution is 4.240, which is larger than the unconstrained value of -10.00787.

19. What is the Jacobian for the following transformations? (A note for aspiring technical writers, about a common error in the literature. A Jacobian is a determinant. The term "Jacobian determinant" has superfluous redundancy.) $y_1 = x_1/x_2$,
$\ln y_2 = \ln x_1 - \ln x_2 + \ln x_3$,
and $y_3 = x_1x_2x_3$.

Let capital letters denote logarithms. Then, the three transformations can be written as

$$Y_1 = X_1 - X_2$$
$$Y_2 = X_1 - X_2 + X_3$$
$$Y_3 = X_1 + X_2 + X_3.$$

This linear transformation is $\mathbf{Y} = \begin{bmatrix}1 & -1 & 0\\ 1 & -1 & 1\\ 1 & 1 & 1\end{bmatrix}\mathbf{X} = \mathbf{JX}$. The inverse transformation is

$\mathbf{X} = \begin{bmatrix}1 & -1/2 & 1/2\\ 0 & -1/2 & 1/2\\ 1 & 1 & 0\end{bmatrix}\mathbf{Y} = \mathbf{J}^{-1}\mathbf{Y}$. In terms of the original variables, then, $x_1 = y_1(y_2/y_3)^{1/2}$, $x_2 = (y_3/y_2)^{1/2}$, and

$x_3 = y_1y_2$. The matrix of partial derivatives can be obtained directly, but an algebraic shortcut will prove useful for obtaining the Jacobian. Note first that $\partial x_i/\partial y_j = (x_i/y_j)(\partial \log x_i/\partial \log y_j)$. Therefore, the elements of the partial derivatives of the inverse transformations are obtained by multiplying the ith row by x_i, where we will substitute the expression for x_i in terms of the ys, then multiplying the jth column by $(1/y_j)$. Thus, the result of Exercise 11 will be useful here. The matrix of partial derivatives will be

$$\begin{bmatrix}\partial x_1/\partial y_1 & \partial x_1/\partial y_2 & \partial x_1/\partial y_3\\ \partial x_2/\partial y_1 & \partial x_2/\partial y_2 & \partial x_2/\partial y_3\\ \partial x_3/\partial y_1 & \partial x_3/\partial y_2 & \partial x_3/\partial y_3\end{bmatrix} = \begin{bmatrix}x_1 & 0 & 0\\ 0 & x_2 & 0\\ 0 & 0 & x_3\end{bmatrix}\begin{bmatrix}1 & -1/2 & 1/2\\ 0 & -1/2 & 1/2\\ 1 & 1 & 0\end{bmatrix}\begin{bmatrix}1/y_1 & 0 & 0\\ 0 & 1/y_2 & 0\\ 0 & 0 & 1/y_3\end{bmatrix}.$$

The determinant of the product matrix is the product of the three determinants. The determinant of the center matrix is -1/2. The determinants of the diagonal matrices are the products of the diagonal elements. Therefore, the Jacobian is $J = \text{abs}(|\partial\mathbf{x}/\partial\mathbf{y}'|) = ½(x_1x_2x_3)/(y_1y_2y_3) = 2(y_1/y_2)$ (after making the substitutions for x_i).

20. Prove that exchanging two columns of a square matrix reverses the sign of its determinant. (**Hint:** use a permutation matrix. See Exercise 6.)

Exchanging the first two columns of a matrix is equivalent to postmultiplying it by a permutation matrix $\mathbf{B} = [\mathbf{e}_2,\mathbf{e}_1,\mathbf{e}_3,\mathbf{e}_4,...]$ where $\mathbf{e}_i$ is the ith column of an identity matrix. Thus, the determinant of the matrix is $|\mathbf{AB}| = |\mathbf{A}|\,|\mathbf{B}|$. The question turns on the determinant of **B**. Assume that **A** and **B** have n columns. To obtain the determinant of **B**, merely expand it along the first row. The only nonzero term in the determinant is $(-1)|\mathbf{I}_{n-1}| = -1$, where $\mathbf{I}_{n-1}$ is the $(n-1)\times(n-1)$ identity matrix. This completes the proof.

21. Suppose $\mathbf{x}=\mathbf{x}(z)$ where z is a scalar. What is $\partial[(\mathbf{x}'\mathbf{Ax})/(\mathbf{x}'\mathbf{Bx})]/z$?

The required derivatives are given in Exercise 16. Let $\mathbf{g} = \partial\mathbf{x}/\partial z$ and let the numerator and denominator be a and b, respectively. Then,

$$\partial(a/b)/\partial z = [b(\partial a/\partial z) - a(\partial b/\partial z)]/b^2$$
$$= [\mathbf{x}'\mathbf{Bx}(2\mathbf{x}'\mathbf{Ag}) - \mathbf{x}'\mathbf{Ax}(2\mathbf{x}'\mathbf{Bg})] / (\mathbf{x}'\mathbf{Bx})^2 = 2[\mathbf{x}'\mathbf{Ax}/\mathbf{x}'\mathbf{Bx}][\mathbf{x}'\mathbf{Ag}/\mathbf{x}'\mathbf{Ax} - \mathbf{x}'\mathbf{Bg}/\mathbf{x}'\mathbf{Bx}].$$

22. Suppose $\mathbf{y}$ is an $n\times 1$ vector and $\mathbf{X}$ is an $n\times K$ matrix. The projection of $\mathbf{y}$ into the column space of $\mathbf{X}$ is defined in the text after equation (2-55), $\hat{\mathbf{y}} = \mathbf{Xb}$. Now, consider the projection of $\mathbf{y}^* = c\mathbf{y}$ into the column space of $\mathbf{X}^* = \mathbf{XP}$ where c is a scalar and $\mathbf{P}$ is a nonsingular $K\times K$ matrix. Find the projection of $\mathbf{y}^*$ into the column space of $\mathbf{X}^*$. Prove that the cosine of the angle between $\mathbf{y}^*$ and its projection into the column space of $\mathbf{X}^*$ is the same as that between $\mathbf{y}$ and its projection into the column space of $\mathbf{X}$. How do you interpret this result?

The projection of $\mathbf{y}^*$ into the column space of $\mathbf{X}^*$ is $\mathbf{X}^*\mathbf{b}^*$ where $\mathbf{b}^*$ is the solution to the set of equations $\mathbf{X}^{*\prime}\mathbf{y}^* = \mathbf{X}^{*\prime}\mathbf{X}^*\mathbf{b}^*$ or $\mathbf{P}'\mathbf{X}'(c\mathbf{y}) = \mathbf{P}'\mathbf{X}'\mathbf{XPb}^*$. Since $\mathbf{P}$ is nonsingular, $\mathbf{P}'$ has an inverse. Premultiplying the equation by $(\mathbf{P}')^{-1}$, we have $c\mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{X}(\mathbf{Pb}^*)$ or $\mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{X}[(1/c)\mathbf{Pb}^*]$. Therefore, in terms of the original $\mathbf{y}$ and $\mathbf{X}$, we see that $\mathbf{b} = (1/c)\mathbf{Pb}^*$ which implies $\mathbf{b}^* = c\mathbf{P}^{-1}\mathbf{b}$. The projection is $\mathbf{X}^*\mathbf{b}^* = (\mathbf{XP})(c\mathbf{P}^{-1}\mathbf{b}) = c\mathbf{Xb}$. We conclude, therefore, that the projection of $\mathbf{y}^*$ into the column space of $\mathbf{X}^*$ is a multiple c of the projection of $\mathbf{y}$ into the space of $\mathbf{X}$. This makes some sense, since, if $\mathbf{P}$ is a nonsingular matrix, the column space of $\mathbf{X}^*$ is exactly the same as the same as that of $\mathbf{X}$. The cosine of the angle between $\mathbf{y}^*$ and its projection is that between $c\mathbf{y}$ and $c\mathbf{Xb}$. Of course, this is the same as that between $\mathbf{y}$ and $\mathbf{Xb}$ since the length of the two vectors is unrelated to the cosine of the angle between them. Thus,
$\cos\theta = (c\mathbf{y})'(c\mathbf{X}\mathrm{b}))/(\|c\mathbf{y}\|\times\|c\mathbf{Xb}\|) = (\mathbf{y}'\mathbf{Xb}))/(\|\mathbf{y}\|\times\|\mathbf{Xb}\|)$.

23. For the matrix $\mathbf{X}' = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 4 & -2 & 3 & -5 \end{bmatrix}$, compute $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ and $\mathbf{M} = (\mathbf{I} - \mathbf{P})$. Verify that $\mathbf{MP} = \mathbf{0}$.

Let $\mathbf{Q} = \begin{bmatrix} 1 & 3 \\ 2 & 8 \end{bmatrix}$ (**Hint:** Show that $\mathbf{M}$ and $\mathbf{P}$ are idempotent.)

(a) Compute the $\mathbf{P}$ and $\mathbf{M}$ based on $\mathbf{XQ}$ instead of $\mathbf{X}$.
(b) What are the characteristic roots of $\mathbf{M}$ and $\mathbf{P}$?

First, $\mathbf{X}'\mathbf{X} = \begin{bmatrix} 4 & 0 \\ 0 & 54 \end{bmatrix}$, $(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 1/4 & 0 \\ 0 & 1/54 \end{bmatrix}$,

$$\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \begin{bmatrix} 1 & 4 \\ 1 & -2 \\ 1 & 3 \\ 1 & -5 \end{bmatrix}\begin{bmatrix} 1/4 & 0 \\ 0 & 1/54 \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 4 & -2 & 3 & -5 \end{bmatrix} = \frac{1}{108}\begin{bmatrix} 59 & 11 & 51 & -13 \\ 11 & 35 & 15 & 47 \\ 51 & 15 & 45 & -3 \\ -13 & 47 & -3 & 77 \end{bmatrix} = \mathbf{P}$$

$$\mathbf{M} = \mathbf{I} - \mathbf{P} = \frac{1}{108}\begin{bmatrix} 49 & -11 & -51 & 13 \\ -11 & 73 & -15 & -47 \\ -51 & -15 & 63 & 3 \\ 13 & -47 & 3 & 31 \end{bmatrix}$$

(a) There is no need to recompute the matrices $\mathbf{M}$ and $\mathbf{P}$ for $\mathbf{XQ}$, they are the same. Proof: The counterpart to $\mathbf{P}$ is $(\mathbf{XQ})[(\mathbf{XQ})'(\mathbf{XQ})]^{-1}(\mathbf{XQ})' = \mathbf{XQ}[\mathbf{Q}'\mathbf{X}'\mathbf{XQ}]^{-1}\mathbf{Q}'\mathbf{X}' =$
$\mathbf{XQQ}^{-1}(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{Q}')^{-1}\mathbf{Q}'\mathbf{X}' = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. The $\mathbf{M}$ matrix would be the same as well. This is an application of the result found in the previous exercise. The $\mathbf{P}$ matrix is the projection matrix, and, as we found, the projection into the space of $\mathbf{X}$ is the same as the projection into the space of $\mathbf{XQ}$.

(b) Since $\mathbf{M}$ and $\mathbf{P}$ are idempotent, their characteristic roots must all be either 0 or 1. The trace of the matrix equals the sum of the roots, which tells how many are 1 and 0. For the matrices above, the traces of both $\mathbf{M}$ and $\mathbf{P}$ are 2, so each has 2 unit roots and 2 zero roots.

24. Suppose that $\mathbf{A}$ is an $n\times n$ matrix of the form $\mathbf{A} = (1\text{-}\rho\mathbf{I}) + \rho\mathbf{ii}'$, where $\mathbf{i}$ is a column of 1s and $0 < \rho < 1$. Write out the format of $\mathbf{A}$ explicitly for $n = 4$. Find all of the characteristic roots and vectors of $\mathbf{A}$. (**Hint:** There are only two distinct characteristic roots, which occur with multiplicity 1 and n-1. Every $\mathbf{c}$ of a certain type is a characteristic vector of $\mathbf{A}$.) For an application which uses a matrix of this type, see Section 14.5 on the random effects model.

For $n = 4$, $\mathbf{A} = \begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix}$. There are several ways to analyze this matrix. Here is a simple shortcut. The characteristic roots and vectors satisfy $[(1\text{-}\rho)\mathbf{I} + \rho\mathbf{ii}']\mathbf{c} = \lambda\mathbf{c}$. Multiply this out to obtain $(1\text{-}\rho)\mathbf{c} + \rho\mathbf{ii}'\mathbf{c} = \lambda\mathbf{c}$ or $\rho\mathbf{ii}'\mathbf{c} = [\lambda\text{-} (1\text{-}\rho)]\mathbf{c}$. Let $\mu = \lambda - (1\text{-}\rho)$, so $\rho\mathbf{ii}'\mathbf{c}=\mu\mathbf{c}$. We need only find the characteristic roots of $\rho\mathbf{ii}'$,μ. The characteristic roots of the original matrix are just $\lambda = \mu + (1\text{-}\rho)$. Now, $\rho\mathbf{ii}'$ is a matrix with rank one, since every column is identical. Therefore, n-1 of the μs are zero. Thus, the original matrix has n-1 roots equal to $0 + (1\text{-}\rho) = (1 \text{-}\rho)$. We can find the remaining root by noting that the sum of the roots of $\rho\mathbf{ii}'$ equals the trace of $\rho\mathbf{ii}'$. Since $\rho\mathbf{ii}'$ has only one nonzero root, that root is the trace, which is $n\rho$. Thus, the remaining root of the original matrix is $(1 - \rho + n\rho)$. The characteristic vectors satisfy the equation $\rho\mathbf{ii}'\mathbf{c} = \mu\mathbf{c}$. For the nonzero root, we have $\rho\mathbf{ii}'\mathbf{c} = n\rho\mathbf{c}$. Divide by $n\rho$ to obtain $\mathbf{i}(1/n)\mathbf{i}'\mathbf{c} = \mathbf{c}$. This equation states that for each element in the vector, $c_i = (1/n)\sum_i c_i$. This implies that every element in the characteristic vector corresponding to the root $(1\text{-}\rho+n\rho)$ is the same, or $\mathbf{c}$ is a multiple of a column of ones. In particular, so that it will have unit length, the vector is $(1/\sqrt{n})\,\mathbf{i}$. For the remaining zero roots, the characteristic vectors must satisfy $\rho\mathbf{i}(\mathbf{i}'\mathbf{c}) = 0\mathbf{c} = \mathbf{0}$. If the characteristic vector is not to be a column of zeroes, the only way to make this an equality is to require $\mathbf{i}'\mathbf{c}$ to be zero. Therefore, for the remaining n-1 characteristic vectors, we may use *any* set of orthogonal vectors whose elements sum to zero and whose inner products are one. There are an infinite number of such vectors. For example, let $\mathbf{D}$ be any arbitrary set of n-1 vectors containing n elements. Transform all columns of $\mathbf{D}$ into deviations from their own column means. Thus, we let $\mathbf{F} = \mathbf{M}^0\mathbf{D}$ where $\mathbf{M}^0$ is defined in Section 2.3.6. Now, let $\mathbf{C} = \mathbf{F}(\mathbf{F}'\mathbf{F})^{-2}$. $\mathbf{C}$ is a linear combination of the columns of $\mathbf{F}$, so its columns sum to zero. By multiplying it out and using the results of Section 2.7.10, you will find that $\mathbf{C}'\mathbf{C} = \mathbf{I}$, so the columns are orthogonal and have unit length.

25. Find the inverse of the matrix in Exercise 24. [Hint: Use (2-66).]

Using the hint, the inverse is

$$[(1-\rho)\mathbf{I}]^{-1} \;-\; \frac{[(1-\rho)\mathbf{I}]^{-1}[\rho\mathbf{ii}'][(1-\rho)\mathbf{I}]^{-1}}{1+\left(\sqrt{\rho}\mathbf{i}\right)'[(1-\rho)\mathbf{I}]^{-1}\left(\sqrt{\rho}\mathbf{i}\right)} \;=\; \frac{1}{1-\rho}\{\mathbf{I} \;-\; [\rho/(1-\rho+n\rho)]\mathbf{ii}'\}$$

26. Prove that every matrix in the sequence of matrices $\mathbf{H}_{i+1} = \mathbf{H}_i + \mathbf{d}_i\mathbf{d}_i'$, where $\mathbf{H}_0 = \mathbf{I}$, is positive definite. For an application, see Section 5.5. For an extension, prove that every matrix in the sequence of matrices defined in (5-22) is positive definite if $\mathbf{H}_0 = \mathbf{I}$.

By repeated substitution, we find $\mathbf{H}_{i+1} = \mathbf{I} + \sum_{j=1}^{i}\mathbf{d}_j\mathbf{d}_j'$. A quadratic form in $\mathbf{H}_{i+1}$ is, therefore

$$\mathbf{x}'\mathbf{H}_{i+1}\mathbf{x} \;=\; \mathbf{x}'\mathbf{x} \;+\; \sum_{j=1}^{i}(\mathbf{x}'\mathbf{d}_j)(\mathbf{d}_j'\mathbf{x}) \;=\mathbf{x}'\mathbf{x} \;+\; \sum_{j=1}^{i}(\mathbf{x}'\mathbf{d}_j)^2$$

This is obviously positive for all $\mathbf{x}$. A simple way to establish this for the matrix in (5-22) is to note that in spite of its complexity, it is of the form $\mathbf{H}_{i+1} = \mathbf{H}_i + \mathbf{d}_i\mathbf{d}_i' + \mathbf{f}_i\mathbf{f}_i'$. If this starts with a positive definite matrix, such as $\mathbf{I}$, then the identical argument establishes its positive definiteness.

27. What is the inverse matrix of $\mathbf{P} = \begin{bmatrix} \cos(x) & \sin(x) \\ -\sin(x) & \cos(x) \end{bmatrix}$? What are the characteristic roots of $\mathbf{P}$?

The determinant of $\mathbf{P}$ is $\cos^2(x) + \sin^2(x) = 1$, so the inverse just reverses the signs of the two off diagonal elements. The two roots are the solutions to $|\mathbf{P}\text{-}\lambda\mathbf{I}| = 0$, which is $\cos^2(x) + \sin^2(x) - 2\lambda\cos(x) + \lambda^2 = 0$. This simplifies because $\cos^2(x) + \sin^2(x) = 1$. Using the quadratic formula, then, $\lambda = \cos(x) \pm (\cos^2(x) - 1)^{1/2}$. But, $\cos^2(x) - 1 = -\sin^2(x)$. Therefore, the imaginary solutions to the resulting quadratic are $\lambda_1,\lambda_2 = \cos(x) \pm i\sin(x)$.

28. Derive the off diagonal block of $\mathbf{A}^{-1}$ in Section 2.6.4.

For the simple 2×2 case, $\mathbf{F}_2$ is derived explicitly in the text, as $\mathbf{F}_2 = (\mathbf{x}'\mathbf{M}^0\mathbf{x})^{-1} = 1/\sum_i (x_i - \bar{x})^2$. Using (2-74), the off diagonal element is just $\mathbf{F}_2(\sum_i x_i)/n = \bar{x}/\sum_i (x_i - \bar{x})^2$. To extend this to a matrix containing a constant and K-1 variables, use the result at the end of the section. The off diagonal vector in $\mathbf{A}^{-1}$ when there is a constant and K-1 other variables is $-\mathbf{F}_2\mathbf{A}_{21}(\mathbf{A}_{11})^{-1} = [\mathbf{X}'\mathbf{M}^0\mathbf{X}]^{-1}\bar{\mathbf{x}}$. In all cases, $\mathbf{A}_{11}$ is just n, so $(\mathbf{A}_{11})^{-1}$ is $1/n$.

29. (This requires a computer.) For the $\mathbf{X}'\mathbf{X}$ matrix at the end of Section 6.4.1,

(a) Compute the characteristic roots of $\mathbf{X}'\mathbf{X}$.

(b) Compute the condition number of $\mathbf{X}'\mathbf{X}$. (Do not forget to scale the columns of the matrix so that the diagonal elements are 1.)

The matrix is
$$\begin{bmatrix} 15.000 & 120.00 & 19.310 & 111.79 & 99.770 \\ 120.00 & 1240.0 & 164.30 & 1035.9 & 875.60 \\ 19.310 & 164.30 & 25.218 & 148.98 & 131.22 \\ 111.79 & 1035.9 & 148.98 & 943.86 & 799.02 \\ 99.770 & 875.60 & 131.22 & 799.02 & 716.67 \end{bmatrix}$$

Its characteristic roots are 2486, 72.96, 19.55, 2.027, and .007354. To compute the condition number, we first extract $\mathbf{D}$ = diag(15,1240,25.218,943.86,716.67). To scale the matrix, we compute $\mathbf{V} = \mathbf{D}^{-2}\mathbf{X}'\mathbf{X}\mathbf{D}^{-2}$.

The resulting matrix is
$$\begin{bmatrix} 1 & .8798823 & .992845 & .939515 & .962265 \\ .879883 & 1 & .929119 & .957532 & .928828 \\ .992845 & .929119 & 1 & .965648 & .976079 \\ .939515 & .957532 & .965648 & 1 & .971503 \\ .962265 & .928828 & .976079 & .971503 & 1 \end{bmatrix}.$$

The characteristic roots of this matrix are 4.801, .1389, .03716, .02183, and .0003527. The square root of the largest divided by the smallest is 116.675. These data are highly collinear by this measure

Appendix B

Probability and Distribution Theory

1. How many different 5 card poker hands can be dealt from a deck of 52 cards?

There are $\binom{52}{5}$ = (52×51×51...×1)/[(5×4×3×2×1)(47×46×...×1)] = 2,598,960 possible hands.

2. Compute the probability of being dealt 4 of a kind in a poker hand.

There are 48(13) possible hands containing 4 of a kind and any of the remaining 48 cards. Thus, given the answer to the previous problem, the probability of being dealt one of these hands is 48(13)/2598960 =.00024, or less than one chance in 4000. □

3. Suppose a lottery ticket costs $1 per play. The game is played by drawing 6 numbers without replacement from the numbers 1 to 48. If you guess all six numbers, you win the prize. Now, suppose that N = the number of tickets sold and P = the size of the prize. N and P are related by

$$N = 5 + 1.2P$$
$$P = 1 + .4N$$

N and P are in millions. What is the expected value of a ticket in this game? (Don't forget that you might have to share the prize with other winners.)

The size of the prize and number of tickets sold are jointly determined. The solutions to the two equations are N = 11.92 million tickets and P = $5.77 million. The number of possible combinations of 48 numbers without replacement is $\binom{48}{6}$ = (48×47×46...×1)/[(6×5×4×3×2×1)(42×41×...×1)] = 12,271,512 so the probability of making the right choice is 1/12271512 = .000000081. The expected number of winners is the expected value of a binomial random variable with N trials and this success probability, which is N times the probability, or 11.92/12.27 = .97, or roughly 1. Thus, one would not expect to have to share the prize. Now, the expected value of a ticket is Prob[win](5.77 million - 1) + Prob[lose](-1) . -53 cents.

4. If x has a normal distribution with mean 1 and standard deviation 3, what are

(a) Prob[$|x| > 2$].

(b) Prob[$x > -1 \mid x < 1.5$].

Using the normal table,

$$\begin{aligned}
\text{(a) Prob}[|x| > 2] &= 1 - \text{Prob}[|x| \leq 2] \\
&= 1 - \text{Prob}[-2 \leq x \leq 2] \\
&= 1 - \text{Prob}[(-2-1)/3 \leq z \leq (2-1)/3] \\
&= 1 - [\text{F}(1/3) - \text{F}(-1)] = 1 - .6306 + .1587 = .5281 \\
\text{(b) Prob}[x > -1 \mid x < 1.5] &= \text{Prob}[-1 < x < 1.5] \,/\, \text{Prob}[x < 1.5] \\
\text{Prob}[-1 < x < 1.5] &= \text{Prob}[(-1-1)/3 < z < (1.5-1)/3)] \\
&= \text{Prob}[z < 1/6] - \text{Prob}[z < -2/3] \\
&= .5662 - .2525 = .3137.
\end{aligned}$$

The conditional probability is .3137/.5662 = .5540.

5. Approximately what is the probability that a random variable with chi-squared distribution with 264 degrees of freedom is less than 297?

We use the approximation in (3-37), $z = [2(297)]^2 - [2(264) - 1]^2 = 1.4155$, so the probability is approximately .9215. To six digits, the approximation is .921539 while the correct value is .921559.

6. Chebychev Inequality For the following two probability distributions, find the lower limit of the probability of the indicated event using the Chebychev inequality (3-18) and the exact probability using the appropriate table:

(a) $x \sim$ Normal$[0,3^2]$, and $-4 < x < 4$.

(b) $x \sim$ chi-squared, 8 degrees of freedom, $0 < x < 16$.

The inequality given in (3-18) states that Prob$[|x - \mu| \leq k\sigma] \geq 1 - 1/k^2$. Note that the result is not informative if k is less than or equal to 1.

(a) The range is 4/3 standard deviations, so the lower limit is $1 - (3/4)^2$ or 7/16 = .4375. From the standard normal table, the actual probability is 1 - 2Prob[z < -4/3] = .8175.

(b) The mean of the distribution is 8 and the standard deviation is 4. The range is, therefore, $\mu \pm 2\sigma$. The lower limit according to the inequality is $1 - (1/2)^2 = .75$. The actual probability is the cumulative chi-squared(8) at 16, which is a bit larger than .95. (The actual value is .9576.)

7. Given the following joint probability distribution,

		X		
		0	1	2
	0	.05	.1	.03
Y	1	.21	.11	.19
	2	.08	.15	.08

(a) Compute the following probabilities: Prob$[Y < 2]$, Prob$[Y < 2, X > 0]$, Prob$[Y = 1, X \geq 1]$.

(b) Find the marginal distributions of X and Y.

(c) Calculate $E[X]$, $E[Y]$, Var$[X]$, Var$[Y]$, Cov$[X,Y]$, and $E[X^2Y^3]$.

(d) Calculate Cov$[Y,X^2]$.

(e) What are the conditional distributions of Y given $X = 2$ and of X given $Y > 0$?

(f) Find $E[Y|X]$ and Var$[Y|X]$. Obtain the two parts of the variance decomposition
Var$[Y]$ = E_x[Var$[Y|X]$] + Var$_x$[$E[Y|X]$].

We first obtain the marginal probabilities. For the joint distribution, these will be

```
X:  P(0) = .34, P(1) = .36, P(2) = .30
Y:  P(0) = .18, P(1) = .51, P(2)  = .31
```

Then,

(a) Prob$[Y < 2]$ = .18 + .51 = .69.
Prob$[Y < 2, X > 0]$ = .1 + .03 + .11 + .19 = .43.
Prob[Y = 1, X $ 1] = .11 + .19 = .30.

(b) They are shown above.

(c) $E[X]$ = 0(.34) + 1(.36) + 2(.30) = .96
$E[Y]$ = 0(.18) + 1(.51) + 2(.31) = 1.13
$E[X^2]$ = $0^2(.34) + 1^2(.36) + 2^2(.30)$ = 1.56
$E[Y^2]$ = $0^2(.18) + 1^2(.51) + 2^2(.31)$ = 1.75
Var$[X]$ = $1.56 - .96^2$ = .6384
Var$[Y]$ = $1.75 - 1.13^2$ = .4731
$E[XY]$ = 1(1)(.11)+1(2)(.15)+2(1)(.19)+2(2)(.08) = 1.11
Cov$[X,Y]$ = 1.11 - .96(1.13) = .0252
$E[X^2Y^3]$ = .11 + 8(.15) + 4(.19) + 32(.08) = 4.63.

(d) E$[YX^2]$ = $1(1^2).11+1(2^2).19+2(1^2).15+2(2^2).08$ = 1.81
Cov$[Y,X^2]$ = 1.81 - 1.13(1.56) = .0472.

(e) Prob[Y = 0 * X = 2] = .03/.3 = .1
Prob[Y = 1 * X = 2] = .19/.3 = .633
Prob[Y = 1 * X = 2] = .08/.3 = .267
Prob[X = 0 * Y > 0] = (.21 + .08)/(.51 + .31) = .3537
Prob[X = 1 * Y > 0] = (.11 + .15)/(.51 + .31) = .3171
Prob[X = 2 * Y > 0] = (.19 + .08)/(.51 + .31) = .3292.

(f) $E[Y * X=0]$ = $0(.05/.34)+1(.21/.34)+2(.08/.34) = 1.088$
$E[Y^2 * X=0]$ = $1^2(.21/.34)+2^2(.08/.34) = 1.559$
$\text{Var}[Y* X=0]$ = $1.559 - 1.088^2 = .3751$
$E[Y*X=1]$ = $0(.1/.36)+1(.11/.36)+2(.15/.36) = 1.139$
$E[Y^2*X=1]$ = $1^2(.11/.36)+2^2(.15/.36) = 1.972$
$\text{Var}[Y*X=1]$ = $1.972 - 1.139^2 = .6749$
$E[Y*X=2]$ = $0(.03/.30)+1(.19/.30)+2(.08/.30) = 1.167$
$E[Y^2*X=2]$ = $1^2(.19/.30)+2^2(.08/.30) = 1.700$
$\text{Var}[Y*X=2]$ = $1.700 - 1.167^2 = .6749 = .3381$
$E[\text{Var}[Y*X]]$ = $.34(.3751)+.36(.6749)+.30(.3381) = .4719$
$\text{Var}[E[Y*X]]$ = $.34(1.088^2)+.36(1.139^2)+.30(1.167^2) - 1.13^2 = 1.2781 - 1.2769 = .0012$
$E[\text{Var}[Y*X]] + \text{Var}[E[Y*X]] = .4719 + .0012 = .4731 = \text{Var}[Y]$.

8. **Minimum mean squared error predictor**. For the joint distribution in Exercise 7, compute $E[y - E[y|x]]^2$. Now, find the a and b which minimize the function $E[y - a - bx]^2$. Given the solutions, verify that $E[y - E[y|x]]^2 \leq E[y - a - bx]^2$. The result is fundamental in least squares theory. Verify that the a and b which you found satisfy (3-68) and (3-69).

		$(x=0)$	$(x=1)$	$(x=2)$
$E[y - E[y\|x]]^2 =$	$(y=0)$	$.05(0 - 1.088)^2$	$+ .10(0 - 1.139)^2$	$+ .03(0 - 1.167)^2$
	$(y=1)$	$+ .21(1 - 1.088)^2$	$+ .11(1 - 1.139)^2$	$+ .19(1 - 1.167)^2$
	$(y=2)$	$+ .08(2 - 1.088)^2$	$+ .15(2 - 1.139)^2$	$+ .08(2 - 1.167)^2$

$= .4719 = E[\text{Var}[y|x]]$.

The necessary conditions for minimizing the function with respect to a and b are

$$\partial E[y - a - bx]^2/\partial a = 2E\{[y - a - bx](-1)\} = 0$$
$$\partial E[y - a - bx]^2/\partial b = 2E\{[y - a - bx](-x)\} = 0.$$

First dividing by -2, then taking expectations produces

$$E[y] - a - bE[x] = 0$$
$$E[xy] - aE[x] - bE[x^2] = 0.$$

Solve the first for $a = E[y] - bE[x]$ and substitute this in the second to obtain

$$E[xy] - E[x](E[y] - bE[x]) - bE[x^2] = 0$$

or $\quad (E[xy] - E[x]E[y]) = b(E[x^2] - (E[x])^2)$

or $\quad b = \text{Cov}[x,y] / \text{Var}[x] = -.0708 / .4731 = -.150$

and $\quad a = E[y] - bE[x] = 1.13 - (-.1497)(.96) = 1.274.$

The linear function compared to the conditional mean produces

	$x=0$	$x=1$	$x=2$
$E[y\|x]$	1.088	1.139	1.167
$a + bx$	1.274	1.124	.974

Now, repeating the calculation above using $a + bx$ instead of $E[y|x]$ produces

		$(x=0)$	$(x=1)$	$(x=2)$
$E[y - a - bx]^2 =$	$(y=0)$	$.05(0 - 1.274)^2$	$+ .10(0 - 1.124)^2$	$+ .03(0 - .974)^2$
	$(y=1)$	$+ .21(1 - 1.274)^2$	$+ .11(1 - 1.124)^2$	$+ .19(1 - .974)^2$
	$(y=2)$	$+ .08(2 - 1.274)^2$	$+ .15(2 - 1.124)^2$	$+ .08(2 - .974)^2$

$= .4950 > .4719.$

9. Suppose x has an exponential distribution, $f(x) = \theta e^{-\theta x}$, $x \geq 0$. (For an application, see Examples 3.5, 3.8, and 3.10.) Find the mean, variance, skewness, and kurtosis of x. (**Hints:** The latter two are defined in Section 3.3. The Gamma integral in Section 5.4.2b will be useful for finding the raw moments.)

In order to find the central moments, we will use the raw moments, $E[x^r] = \int_0^\infty \theta x^r e^{-\theta x} dx$. These can be obtained by using the gamma integral. Making the appropriate substitutions, we have

$$E[x^r] = [\theta\Gamma(r+1)]/\theta^{r+1} = r!/\theta^r.$$

The first four moments are: $E[x] = 1/\theta$, $E[x^2] = 2/\theta^2$, $E[x^3] = 6/\theta^3$, and $E[x^4] = 24/\theta^4$. The mean is, thus, $1/\theta$ and the variance is $2/\theta^2 - (1/\theta)^2 = 1/\theta^2$. For the skewness and kurtosis coefficients, we have

$$E[x - 1/\theta]^3 = E[x^3] - 3E[x^2]/\theta + 3E[x]/\theta^2 - 1/\theta^3 = 2/\theta^3.$$

The normalized skewness coefficient is 2. The kurtosis coefficient is

$$E[x - 1/\theta]^4 = E[x^4] - 4E[x^3]/\theta + 6E[x^2]/\theta^2 - 4E[x]/\theta^3 + 1/\theta^4 = 9/\theta^4.$$

The degree of excess is 6.

10. For the random variable in Exercise 9, what is the probability distribution of the random variable $y = e^{-x}$? What is $E[y]$? Prove that the distribution of this y is a special case of the beta distribution in (3-40).

If $y = e^{-x}$, then $x = -\ln y$, so the Jacobian is $|dx/dy| = 1/y$. The distribution of y is, therefore,

$$f(y) = \theta e^{-\theta(-\ln y)}(1/y) = (\theta y^{\theta})/y = \theta y^{\theta-1} \text{ for } 0 < y < 1.$$

This is in the form of (3-40) with y instead of x, $c = 1$, $\beta = 1$, and $\alpha = \theta$.

11. If the probability density of y is $\alpha y^2(1-y)^3$ for y between 0 and 1, what is α? What is the probability that y is between .25 and .75?

This is a beta distribution of the form in (3-40) with $\alpha = 3$ and $\beta = 4$. Therefore, the constant is $\Gamma(3+4)/(\Gamma(3)\Gamma(4)) = 60$. The probability is

$$\int_{.25}^{.75} 60y^2(1-y)^3dy = 60\int_{.25}^{.75} (y^2 - 3y^3 + 3y^4 - y^5)dy = 60(y^3/3 - 3y^4/4 + 3y^5/5 - y^6/6)\Big|_{.25}^{.75} = .79296.$$

12. Suppose x has the following discrete probability distribution:

X	1	2	3	4
Prob[$X = x$]	.1	.2	.4	.3.

Find the exact mean and variance of X. Now, suppose $Y = 1/X$. Find the exact mean and variance of Y. Find the mean and variance of the linear and quadratic approximations to $Y = f(X)$. Are the mean and variance of the quadratic approximation closer to the true mean than those of the linear approximation?

We will require a number of moments of x, which we derive first:

$$E[x] = .1(1) + .2(2) + .4(3) + .3(4) = 2.9 = \mu$$
$$E[x^2] = .1(1) + .2(4) + .4(9) + .3(16) = 9.3$$
$$\text{Var}[x] = 9.3 - 2.9^2 = .89 = \sigma^2.$$

For later use, we also obtain

$$E[x - \mu]^3 = .1(1 - 2.9)^3 + ... = -.432$$
$$E[x - \mu]^4 = .1(1 - 2.9)^4 + ... = 1.8737.$$

The approximation is $y = 1/x$. The exact mean and variance are

$$E[y] = .1(1) + .2(1/2) + .4(1/3) + .3(1/4) = .40833$$
$$\text{Var}[y] = .1(12) + .2(1/4) + .4(1/9) + .3(1/16) - .40833^2 = .04645.$$

The linear Taylor series approximation around μ is $y \approx 1/\mu + (-1/\mu^2)(x - \mu)$. The mean of the linear approximation is $1/\mu = .3448$ while its variance is $(1/\mu^4)\text{Var}[x-\mu] = \sigma^2/\mu^4 = .01258$. The quadratic approximation is

$$y \approx 1/\mu + (-1/\mu^2)(x - \mu) + (1/2)(2/\mu^3)(x - \mu)^2$$
$$= 1/\mu - (1/\mu^2)(x - \mu) + (1/\mu^3)(x - \mu)^2.$$

The mean of this approximation is $E[y] \approx 1/\mu + \sigma^2/\mu^3 = .3813$ while the variance is approximated by the variance of the right hand side,

$$(1/\mu^4)\text{Var}[x - \mu] + (1/\mu^6)\text{Var}[x - \mu]^2 - (2/\mu^5)\text{Cov}[(x-\mu),(x-\mu)^2]$$
$$= (1/\mu^4)\sigma^2 + (1/\mu^6)(E[x - \mu]^4 - \sigma^4] - (2/\mu^5)E[x - \mu]^3$$
$$= .01498.$$

Neither approximation provides a close estimate of the variance. Note that in both cases, it would be possible simply to evaluate the approximations at the four values of x and compute the means and variances directly. The virtue of the approach above is that it can be applied when there are many values of x, and is necessary when the distribution of x is continuous.

13. **Interpolation in the chi-squared table.** In order to find a percentage point in the chi-squared table which is between two values, we interpolate linearly between the *reciprocals* of the degrees of freedom. The chi-squared distribution is defined for noninteger values of the degrees of freedom parameter [see (3-39)], but your table does not contain critical values for noninteger values. Using linear interpolation, find the 99% critical value for a chi-squared variable with degrees of freedom parameter 11.3. (For an application of this calculation, see Section 8.5.1. and Example 8.6.)

The 99% critical values for 11 and 12 degrees of freedom are 24.725 and 26.217. To interpolate linearly between these values for the value corresponding to 11.3 degrees of freedom, we use

$$c = 26.217 + \frac{(111.3 - 1/12)}{(1/11 - 1/12)}(24.725 - 26.217) = 25.2009.$$

14. Suppose x has a standard normal distribution. What is the pdf of the following random variable?

$y = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}, 0 < y < \frac{1}{\sqrt{2\pi}}$. [**Hints:** You know the distribution of $z = x^2$ from (3-30). The density of this z is given in (3-39). Solve the problem in terms of $y = g(z)$.]

We know that $z = x^2$ is distributed as chi-squared with 1 degree of freedom. We seek the density of $y = ke^{-z/2}$ where $k = (2\pi)^{-2}$. The inverse transformation is $z = 2\ln k - 2\ln y$, so the Jacobian is $|-2/y| = 2/y$. The density of z is that of Gamma with parameters 1/2 and 1/2. [See (3-39) and the succeeding discussion.] Thus,

$$f(z) = \frac{(1/2)^{1/2}}{\Gamma(1/2)}e^{-z/2}z^{-1/2}, z > 0.$$

Note, $\Gamma(1/2) = \sqrt{\pi}$. Making the substitution for z and multiplying by the Jacobian produces

$$f(y) = \frac{(1/2)^{1/2}}{\Gamma(1/2)}\frac{2}{y}e^{(-1/2)(2\ln k - 2\ln y)}(2\ln k - 2\ln y)^{-1/2}$$

The exponential term reduces to y/k. The scale factor is equal to $2k/y$. Therefore, the density is simply

$f(y) = 2(2\ln k - 2\ln y)^{-1/2} = \sqrt{2}\,(\ln k - \ln y)^{-1/2} = \{2/[\ln(1/(y(2\pi)^{1/2}))]\}, 0 < y < (2\pi)^{-1/2}$.

15. **The fundamental probability transformation**. Suppose that the continuous random variable x has cumulative distribution $F(x)$. What is the probability distribution of the random variable $y = F(x)$? (**Observation:** This result forms the basis of the simulation of draws from many continuous distributions.)

The inverse transformation is $x(y) = F^{-1}(y)$, so the Jacobian is $dx/dy = F^{-1\prime}(y) = 1/f(x(y))$ where $f(.)$ is the density of x. The density of y is $f(y) = f[F^{-1}(y)] \times 1/f(x(y)) = 1, 0 \leq y \leq 1$. Thus, y has a continuous uniform distribution. Note, then, for purposes of obtaining a random sample from the distribution, we can sample $y_1,...,y_n$ from the distribution of y, the continuous uniform, then obtain $x_1 = x_1(y_1), \ldots x_n = x_n(y_n)$.

16. **Random number generators**. Suppose x is distributed uniformly between 0 and 1, so $f(x) = 1, 0 \leq x \leq 1$. Let θ be some positive constant. What is the pdf of $y = -(1/\theta)\ln x$. (**Hint**: See Section 3.5.) Does this suggest a means of simulating draws from this distribution if one has a random number generator which will produce draws from the uniform distribution? To continue, suggest a means of simulating draws from a logistic distribution, $f(x) = e^{-x}/(1+e^{-x})^2$.

The inverse transformation is $x = e^{-\theta y}$ so the Jacobian is $dx/dy = \theta e^{-\theta y}$. Since $f(x) = 1$, this Jacobian is also the density of y. One can simulate draws y from any exponential distribution with parameter θ by drawing observations x from the uniform distribution and computing $y = -(1/\theta)\ln x$. Likewise, for the logistic distribution, the CDF is $F(x) = 1/(1 + e^{-x})$. Thus, draws y from the uniform distribution may be taken as draws on $F(x)$. Then, we may obtain x as $x = \ln[F(x)/(1 - F(x)] = \ln[y/(1 - y)]$.

17. Suppose that x_1 and x_2 are distributed as independent standard normal. What is the joint distribution of $y_1 = 2 + 3x_1 + 2x_2$ and $y_2 = 4 + 5x_1$? Suppose you were able to obtain two samples of observations from independent standard normal distributions. How would you obtain a sample from the bivariate normal distribution with means 1 and 2 variances 4 and 9 and covariance 3?

We may write the pair of transformations as

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix} + \begin{bmatrix} 3 & 2 \\ 5 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{b} + \mathbf{Ax}.$$

The problem also states that $\mathbf{x} \sim N[\mathbf{0},\mathbf{I}]$. From (3-103), therefore, we have $\mathbf{y} \sim N[\mathbf{b} + \mathbf{A0}, \mathbf{AIAN}]$ where

$$E[\mathbf{y}] = \mathbf{b} + \mathbf{A0} = \mathbf{b} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}, \text{Var}[\mathbf{y}] = \mathbf{AA}' = \begin{bmatrix} 13 & 15 \\ 15 & 25 \end{bmatrix}.$$

For the second part of the problem, using our result above, we would require the $\mathbf{A}$ and $\mathbf{b}$ such that $\mathbf{b} + \mathbf{A0} = (1,2)'$ and $\mathbf{AA}' = \begin{bmatrix} 4 & 3 \\ 3 & 9 \end{bmatrix}$. The vector is obviously $\mathbf{b} = (1,2)'$. In order to find the elements of $\mathbf{A}$, there are a few ways to proceed. The Cholesky factorization used in Exercise 9 is probably the simplest. Let $y_1 = 1 + 2x_1$. Thus, y_1 has mean 1 and variance 4 as required. Now, let $y_2 = 2 + w_1x_1 + w_2x_2$. The covariance between y_1 and y_2 is $2w_1$, since x_1 and x_2 are uncorrelated. Thus, $2w_1 = 3$, or $w_1 = 1.5$. Now, $\text{Var}[y_2] = w_1^2 + w_2^2 = 9$, so $w_2^2 = 9 - 1.5^2 = 6.75$. The transformation matrix is, therefore, $\mathbf{A} = \begin{bmatrix} 2 & 0 \\ 1.5 & 2.598 \end{bmatrix}$. This is the Cholesky factorization of the desired $\mathbf{AA}'$ above. It is worth noting, this provides a simple method of finding the requisite $\mathbf{A}$ matrix for any number of variables. Finally, an alternative method would be to use the characteristic roots and vectors of $\mathbf{AA}'$. The inverse square root defined in Section 2.7.12 would also provide a method of transforming $\mathbf{x}$ to obtain the desired covariance matrix.

18. The density of the standard normal distribution, denoted $\phi(x)$, is given in (3-28). The function based on the ith derivative of the density given by $H_i = [(-1)^i d^i\phi(x)/dx^i]/\phi(x)$, $i = 0,1,2,...$ is called a *Hermite polynomial*. By definition, $H_0 = 1$.

(a) Find the next three Hermite polynomials.

(b) A useful device in this context is the differential equation

$$d^r\phi(x)/dx^r + xd^{r-1}\phi(x)/dx^{r-1} + (r-1)d^{r-2}\phi(x)/dx^{r-2} = 0.$$

Use this result and the results of part a. to find H_4 and H_5.

The crucial result to be used in the derivations is $d\phi(x)/dx = -x\phi(x)$. Therefore,

$$d^2\phi(x)/dx^2 = (x^2 - 1)\phi(x)$$

and $d^3\phi(x)/dx^3 = (3x - x^3)\phi(x)$.

The polynomials are $H_1 = x$, $H_2 = x^2 - 1$, and $H_3 = x^3 - 3x$.

For part (b), we solve for $d^r\phi(x)/dx^r = -xd^{r-1}\phi(x)/dx^{r-1} - (r-1)d^{r-2}\phi(x)/dx^{r-2}$

Therefore, $d^4\phi(x)/dx^4 = -x(3x - x^3)\phi(x) - 3(x^2 - 1)\phi(x) = (x^4 - 6x^2 + 3)\phi(x)$

and $d^5\phi(x)/dx^5 = (-x^5 + 10x^3 - 15x)\phi(x)$.

Thus, $H_4 = x^4 - 6x^2 + 3$ and $H_5 = x^5 - 10x^3 + 15x$.

19. Continuation: *orthogonal polynomials*: The Hermite polynomials are orthogonal if x has a standard normal distribution. That is, $E[H_iH_j] = 0$ if $i \neq j$. Prove this for the H_1, H_2, and H_3 which you obtained above.

$$E[H_1(x)H_2(x)] = E[x(x^2 - 1)] = E[x^3 - x] = 0$$

since the normal distribution is symmetric. Then,

$$E[H_1(x)H_3(x)] = E[x(x^3 - 3x)] = E[x^4 - 3x^2] = 0.$$

The fourth moment of the standard normal distribution is 3 times the variance. Finally,

$$E[H_2(x)H_3(x)] = E[(x^2 - 1)(x^3 - 3x)] = E[x^5 - 4x^3 + 3x] = 0$$

because all odd order moments of the normal distribution are zero. (The general result for extending the preceding is that in a product of Hermite polynomials, if the sum of the subscripts is odd, the product will be a sum of odd powers of x, and if even, a sum of even powers. This provides a method of determining the higher moments of the normal distribution if they are needed. (For example, $E[H_1H_3] = 0$ implies that $E[x^4] = 3E[x^2]$.)

20. If x and y have means μ_x and μ_y and variances σ_x^2 and σ_y^2 and covariance σ_{xy}, what is the approximation of the covariance matrix of the two random variables $f_1 = x/y$ and $f_2 = xy$?

The elements of $\mathbf{J\Sigma J'}$ are $(1,1) = \dfrac{\sigma_x^2}{\mu_y^2} + \dfrac{\sigma_y^2\mu_2^x}{\mu_y^4} - \dfrac{2\sigma_{xy}\mu_x}{\mu_y^3}$

$(1,2) = \sigma_x^2 - \sigma_y^2\ \mu_x^2 /\ \mu_y^4$

$(2,2) = \sigma_x^2\ \mu_y^4\ + \sigma_y^2\ \mu_x^2\ + 2\sigma_{xy}\mu_x\mu_y$.

21. ***Factorial Moments.*** For finding the moments of a distribution such as the Poisson, a useful device is the factorial moment. (The Poisson distribution is given in Example 3.1.) The density is

$$f(x) = e^{-\lambda}\lambda^x / x!,\ x = 0,1,2,...$$

To find the mean, we can use

$$\begin{aligned} E[x] &= \sum\nolimits_{x=0}^{\infty} xf(x) = \sum\nolimits_{x=0}^{\infty} xe^{-\lambda}\lambda^x / x! \\ &= \sum\nolimits_{x=1}^{\infty} e^{-\lambda}\lambda^{x-1} / (x-1)! \\ &= \lambda\sum\nolimits_{y=0}^{\infty} e^{-\lambda}\lambda^{y} / y! \\ &= \lambda, \end{aligned}$$

since the probabilities sum to 1. To find the variance, we will extend this method by finding $E[x(x\text{-}1)]$, and likewise for other moments. Use this method to find the variance and third central moment of the Poisson distribution. (Note that this device is used to transform the factorial in the denominator in the probability.)

Using the same technique,

$$\begin{aligned} E[x(x\text{-}1)] &= \sum\nolimits_{x=0}^{\infty} x(x-1)f(x) = \sum\nolimits_{x=0}^{\infty} x(x-1)e^{-\lambda}\lambda^x / x! \\ &= \sum\nolimits_{x=2}^{\infty} e^{-\lambda}\lambda^{x-2} / (x-2)! \\ &= \lambda^2\sum\nolimits_{y=0}^{\infty} e^{-\lambda}\lambda^{y} / y! \\ &= \lambda^2 \\ &= \mathrm{E}[x^2]\ -\ \mathrm{E}[x] \end{aligned}$$

So, $E[x^2] = \lambda^2 + \lambda$.

Since $E[x] = \lambda$, it follows that $\mathrm{Var}[x] = (\lambda^2 + \lambda) - \lambda^2 = \lambda$. Following the same pattern, the preceding produces

$$\begin{aligned} E[x(x\text{-}1)(x\text{-}2)] &= E[x^3] - 3E[x^2] + 2E[x]. \\ &= \lambda^3. \end{aligned}$$

Therefore, $E[x^3] = \lambda^3 + 3(\lambda + \lambda^2) - 2\lambda = \lambda^3 + 3\lambda^2 + \lambda$.

Then, $E[x - E[x]]^3 = E[x^3] - 3\lambda E[x^2] + 3\lambda^2 E[x] - \lambda^3 = \lambda$.

22. If x has a normal distribution with mean μ and standard deviation σ, what is the probability distribution of $y = e^x$?

If $y = e^x$, then $x = \ln y$ and the Jacobian is $dx/dy = 1/y$. Making the substitution,

$$f(y) = \frac{1}{\sigma y\sqrt{2\pi}} e^{-\frac{1}{2}\left[(\ln y - \mu)/\sigma\right]^2}$$

This is the density of the lognormal distribution.

23. If y has a lognormal distribution, what is the probability distribution of y^2?

Let $z = y^2$. Then, $y = \sqrt{z}$ and $dy/dz = 1/(2\sqrt{z})$. Inserting these in the density above, we find

$$f(z) \quad = \frac{1}{\sigma\sqrt{2\pi}}\frac{1}{\sqrt{z}}\frac{1}{2\sqrt{z}}e^{-\frac{1}{2}\left[\left(\frac{1}{2}\ln z-\mu\right)/\sigma\right]^2}, z > 0$$

$$= \frac{1}{(2\sigma)z\sqrt{2\pi}}e^{-\frac{1}{2}[(\ln z-2\mu)/(2\sigma)]^2}, z > 0.$$

Thus, z has a lognormal distribution with parameters 2μ and 2σ. The general result is that if y has a lognormal distribution with parameters μ and σ, y^r has a lognormal distribution with parameters $r\mu$ and $r\sigma$.

24. Suppose y, x_1, and x_2 have a joint normal distribution with parameters $\mu' = [1, 2, 4]$

and covariance matrix $\Sigma = \begin{bmatrix} 2 & 3 & 1 \\ 3 & 5 & 2 \\ 1 & 2 & 6 \end{bmatrix}$

(a) Compute the intercept and slope in the function $E[y|x_1]$, $\text{Var}[y|x_1]$, and the coefficient of determination in this regression. (**Hint**: See Section 3.10.1.)

(b) Compute the intercept and slopes in the conditional mean function, $E[y|x_1,x_2]$. What is $E[y|x_1=2.5,x_2=3.3]$? What is $\text{Var}[y|x_1=2.5,x_2=3.3]$?

First, for normally distributed variables, we have from (3-102),

$$E[y|\mathbf{x}] = \mu_y + \text{Cov}[y,\mathbf{x}]\{\text{Var}[\mathbf{x}]\}^{-1}(\mathbf{x} - \mu_x)$$

and
$$\text{Var}[y|\mathbf{x}] = \text{Var}[y] - \text{Cov}[y,\mathbf{x}]\{\text{Var}[\mathbf{x}]\}^{-1}\text{Cov}[\mathbf{x},y]$$

and
$$COD = \text{Var}[E[y|\mathbf{x}]] / \text{Var}[y]$$
$$= \text{Cov}[y,\mathbf{x}]\{\text{Var}[\mathbf{x}]\}^{-1}\text{Cov}[\mathbf{x},y] / \text{Var}[y].$$

We may just insert the figures above to obtain the results.

$$E[y|x_1] = 1 + (3/5)(x_1 - 2) = -.2 + .6x_1,$$
$$\text{Var}[y|x_1] = 2 - 3(1/5)3 = 1/5 = .2$$
$$COD = .6^2(5) / 2 = .9$$
$$\text{E}[y|x_1,x_2] = 1 + \begin{bmatrix} 3 & 1 \end{bmatrix}\begin{bmatrix} 5 & 2 \\ 2 & 6 \end{bmatrix}^{-1}\begin{bmatrix} 3 \\ 1 \end{bmatrix}$$
$$= -.4615 + .6154x_1 - .03846x_2,$$
$$\text{Var}[y|x_1,x_2] = 2 - (.6154,-.03846)(3,1)' = .1923.$$
$$E[y|x_1=2.5,x_2=3.3] = 1.3017.$$

The conditional variance is not a function of x_1 or x_2. ~

25. What is the density of $y = 1/x$ if x has a chi-squared distribution?

The density of a chi-squared variable is a gamma variable with parameters 1/2 and $n/2$ where n is the degrees of freedom of the chi-squared variable. Thus,

$$f(x) = \frac{(1/2)^{n/2}}{\Gamma(n/2)}e^{-\frac{1}{2}x}x^{\frac{n}{2}-1}, x > 0.$$

If $y = 1/x$ then $x = 1/y$ and $|dx/dy| = 1/y^2$. Therefore, after multiplying by the Jacobian,

$$f(y) = \frac{(1/2)^{n/2}}{\Gamma(n/2)}e^{-\frac{1}{2y}}\left(\frac{1}{y}\right)^{\frac{n}{2}+1}, y > 0.$$

26. What is the density and what are the mean and variance of $y = 1/x$ if x has the gamma distribution described in Section 3.4.5.

The density of x is $f(x) = \dfrac{\lambda^P}{\Gamma(P)} e^{-\lambda x} x^{P-1}, x > 0$. If $y = 1/x$, then $x = 1/y$, and the Jacobian is $|dx/dy| = 1/y^2$. Using the change of variable formula, as usual, the density of y is $f(y) = \dfrac{\lambda^P}{\Gamma(P)} \dfrac{1}{y^2} e^{-\lambda/y} \left(\dfrac{1}{y}\right)^{P-1}, y > 0$. The mean is $E(y) = \int_0^\infty y \dfrac{\lambda^P}{\Gamma(P)} \dfrac{1}{y^2} e^{-\lambda/y} \left(\dfrac{1}{y}\right)^{P-1} dy$. This is a gamma integral (see Section 5.2.4b). Combine terms to obtain $E(y) = \int_0^\infty \dfrac{\lambda^P}{\Gamma(P)} e^{-\lambda/y} \left(\dfrac{1}{y}\right)^{P} dy$. Now, in order to use the results for the gamma integral, we will have to make a change of variable. Let $z = 1/y$, so $|dy/dz| = 1/z^2$. Making the change of variable, we find $E(y) = \int_0^\infty \dfrac{\lambda^P}{\Gamma(P)} e^{-\lambda z} z^P \left(\dfrac{1}{z^2}\right) dz = \int_0^\infty \dfrac{\lambda^P}{\Gamma(P)} e^{-\lambda z} z^{P-2} dz$. Now, we can use the gamma integral directly, to find $E(y) = \dfrac{\lambda^P}{\Gamma(P)} \times \dfrac{\Gamma(P-1)}{\lambda^{P-1}} = \dfrac{\lambda}{P-1}$. Note that for this to exist, P must be greater than one. We can use the same approach to find the variance. We start by finding $E[y^2]$. First, $E(y^2) = \int_0^\infty y^2 \dfrac{\lambda^P}{\Gamma(P)} \dfrac{1}{y^2} e^{-\lambda/y} \left(\dfrac{1}{y}\right)^{P-1} dy = \int_0^\infty \dfrac{\lambda^P}{\Gamma(P)} e^{-\lambda/y} \left(\dfrac{1}{y}\right)^{P-1} dy$. Once again, this is a gamma integral, which we can evaluate by first making the change of variable to $z = 1/y$. The integral is $E(y^2) = \int_0^\infty \dfrac{\lambda^P}{\Gamma(P)} e^{-\lambda z} z^{P-1} \left(\dfrac{1}{z^2}\right) dz = \int_0^\infty \dfrac{\lambda^P}{\Gamma(P)} e^{-\lambda z} z^{P-3} dz$. This is $\dfrac{\lambda^P}{\Gamma(P)} \times \dfrac{\Gamma(P-2)}{\lambda^{P-2}} = \dfrac{\lambda^2}{(P-1)(P-2)}$.

Now, $\text{Var}[y] = E[y^2] - E^2[y] = \dfrac{\lambda^3}{(P-1)^2(P-2)}, P > 2$.

27. Suppose x_1 and x_2 have the bivariate normal distribution described in Section 3.8. Consider an extension of Example 3.4, where the bivariate normal distribution is obtained by transforming two independent standard normal variables. Obtain the distribution of $z = \exp(y_1)\exp(y_2)$ where y_1 and y_2 have a bivariate normal distribution and are correlated. Solve this problem in two ways. First, use the transformation approach described in Section 3.6.4. Second, note that $z = \exp(y_1+y_2) = \exp(w)$, so you can first find the distribution of w, then use the results of Section 3.5 (and, in fact, Section 3.4.4 as well).

The (extremely) hard way to proceed is to define the joint transformations $z_1 = \exp(y_1)\exp(y_2)$ and $z_2 = \exp(y_2)$. The Jacobian is $1/(z_1z_2)$. The joint distribution is the Jacobian times the bivariate normal distribution, evaluated at $y_1 = \log z_1 - \log z_2$ and $y_2 = \log z_2$, from which it is now necessary to integrate out z_2. Obviously, this is going to be tedious, but the hint gives a much simpler way to proceed. The variable $w = y_1+y_2$ has a normal distribution with mean $\mu = \mu_1+\mu_2$ and variance $\sigma^2 = (\sigma_1^2 + \sigma_2^2 + 2\sigma_{12})$. We already have a simple result for $\exp(w)$ in Exercise 22; this has a lognormal distribution.

28. **Probability Generating Function**. For a discrete random variable, x, the function

$$E[t^x] = \sum_{x=0}^{\infty} t^x \text{Prob}[X = x]$$

is called the **probability generating function** because in the function, the coefficient on t^i is Prob[X=i]. Suppose that x is the number of the repetitions of an experiment with probability π of success upon which the first success occurs. The density of x is the *geometric distribution*,

$$\text{Prob}[X=x] = (1 - \pi)^{x-1}\pi.$$

What is the probability generating function?

$$\begin{aligned} E[t^x] &= \sum_{x=0}^{\infty} t^x (1-\pi)^{x-1}\pi \\ &= \frac{\pi}{(1-\pi)} \sum_{x=0}^{\infty} [t(1-\pi)]^x \\ &= \frac{\pi}{(1-\pi)} \frac{1}{1-t(1-\pi)}. \end{aligned}$$

29. **Moment Generating Function**. For the random variable X, with probability density function $f(x)$, if the function $M(t) = E[e^{tx}]$ exists, it is the moment generating function. Assuming the function exists, it can be shown that $d^rM(t)/dt^r|_{t=0} = E[x^r]$. Find the moment generating functions for

(a) The Exponential distribution of Exercise 9.
(b) The Poisson distribution of Exercise 21.

For the continuous variable in (a), For $f(x) = \theta\exp(-\theta x)$, $M(t) = \int_0^{\infty} e^{tx}\theta e^{-\theta\, x}dx = \int_0^{\infty} \theta e^{-(\theta\, - t)x}dx$.

This is θ times a Gamma integral (see Section 5.4.2b) with p=1, c=1, and $a = (\theta\text{-}t)$. Therefore, $M(t) = \theta/(\theta - t)$.

For the Poisson distribution,

$$\begin{aligned} M(t) &= \sum_{x=0}^{\infty} e^{tx} e^{-\lambda}\lambda^x / x! = \sum_{x=0}^{\infty} e^{-\lambda}(\lambda e^t)^x / x! \\ &= \sum_{x=0}^{\infty} e^{-\lambda} e^{\lambda e^t} e^{-\lambda e^t} (\lambda e^t)^x / x! \\ &= e^{-\lambda+\lambda e^t} \sum_{x=0}^{\infty} e^{-\lambda e^t} (\lambda e^t)^x / x! \end{aligned}$$

The sum is the sum of probabilities for a Poisson distribution with parameter λe^t, which equals 1, so the term before the summation sign is the moment generating function, $M(t) = \exp[\lambda(e^t - 1)]$.

28. **Moment generating function for a sum of variables.** When it exists, the moment generating function has a one to one correspondence with the distribution. Thus, for example, if we begin with some random variable and find that a transformation of it has a particular MGF, we may infer that the function of the random variable has the distribution associated with that MGF. A useful application is the following:
If x and y are independent, the MGF of $x + y$ is $M_x(t)M_y(t)$.

(a) Use this result to prove that the sum of Poisson random variables has a Poisson distribution.
(b) Use the result to prove that the sum of chi-squared variables has a chi-squared distribution. [Note, you must first find the MGF for a chi-squared variate. The density is given in (3-39).]
(c) The MGF for the standard normal distribution is $M_z = \exp(-t^2/2)$. Find the MGF for the N[μ,σ^2] distribution, then find the distribution of a sum of normally distributed variables.

(a) From the previous problem, $M_x(t) = \exp[\lambda(e^t - 1)]$. Suppose y is distributed as Poisson with parameter μ. Then, $M_y(t)=\exp[\mu(e^t-1)]$. The product of these two moment generating functions is $M_x(t)M_y(t)= \exp[\lambda(e^t - 1)]\exp[\mu(e^t - 1)] = \exp[(\lambda+\mu)(e^t - 1)]$, which is the moment generating function of the Poisson distribution with parameter $\lambda+\mu$. Therefore, on the basis of the theorem given in the problem, it follows that x+y has a Poisson distribution with parameter $\lambda+\mu$.

(b) The density of the Chi-squared distribution with n degrees of freedom is [from (3-39)]

$$f(x) = \frac{(1/2)^{n/2}}{\Gamma(n/2)} e^{-\frac{1}{2}x} x^{\frac{n}{2}-1}, x > 0.$$

Let the constant term be k for the present. The moment generating function is

$$M(t) = k\int_0^\infty e^{tx}e^{-x/2}x^{(n/2)-1}dx$$
$$= k\int_0^\infty e^{-x(1/2-t)}x^{(n/2)-1}dx\ .$$

This is a gamma integral which reduces to $M(t) = k(1/2 - t)^{-n/2}\Gamma(n/2)$. Now, reinserting the constant k and simplifying produces the moment generating function $M(t) = (1 - 2t)^{-n/2}$. Suppose that x_i is distributed as chi-squared with n_i degrees of freedom. The moment generating function of $\Sigma_i x_i$ is

$$\Pi_i M_i(t) = (1-2t)^{-\sum_i n_i/2}$$

which is the MGF of a chi-squared variable with $n = \Sigma_i n_i$ degrees of freedom.

(c) We let $y = \sigma z + \mu$. Then, $M_y(t) = E[\exp(ty)] = E\left[e^{t(\sigma z+\mu)}\right] = e^{t\mu}E\left[e^{\sigma tz}\right] = e^{t\mu}E\left[e^{(\sigma t)z}\right]$

$$= e^{\mu t}e^{-(\sigma t)^2/2} = \exp\left[\mu t - (\sigma^2 t^2)/2\right]$$

Using the same approach as in part b., it follows that the moment generating function for a sum of random variables with means μ_i and standard deviations σ_i is

$$M_{\sum_i x_i} = \exp\left[\sum_i \mu_i - \frac{1}{2}\left(\sum_i \sigma_i^2\right)t^2\right].$$

Appendix C

Estimation and Inference

1. The following sample is drawn from a normal distribution with mean μ and standard deviation σ:

$$x = 1.3, 2.1, .4, 1.3, .5, .2, 1.8, 2.5, 1.9, 3.2.$$

Compute the mean, median, variance, and standard deviation of the sample.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = 1.52,$$

$$s^2 = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n-1} = .9418,$$

$$s = .97$$

$$median = 1.55, \text{ midway between 1.3 and 1.8.}$$

2. Using the data in the previous exercise, test the following hypotheses:

(a) $\mu > 2$.

(b) $\mu < .7$.

(c) $\sigma^2 = .5$.

(d) Using a likelihood ratio test, test the following hypothesis $\mu = 1.8$, $\sigma^2 = .8$.

(a) We would reject the hypothesis if 1.52 is too small relative to the hypothesized value of 2. Since the data are sampled from a normal distribution, we may use a *t* test to test the hypothesis. The *t* ratio is

$$t[9] = (1.52 - 2) / [.97/\sqrt{10}\,] = -1.472.$$

The 95% critical value from the *t* distribution for a one tailed test is -1.833. Therefore, we would not reject the hypothesis at a significance level of 95%.

(b) We would reject the hypothesis if 1.52 is excessively large relative to the hypothesized mean of .7. The *t* ratio is $t[9] = (1.52 - .7) / [.97/\sqrt{10}\,] = 2.673$. Using the same critical value as in the previous problem, we would reject this hypothesis.

(c) The statistic $(n\text{-}1)s^2/\sigma^2$ is distributed as χ^2 with 9 degrees of freedom. This is 9(.94)/.5 = 16.920. The 95% critical values from the chi-squared table for a two tailed test are 2.70 and 19.02. Thus we would not reject the hypothesis.

(d) The log-likelihood for a sample from a normal distribution is

$$\ln L = -(n/2)\ln(2\pi) - (n/2)\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

The sample values are $\hat{\mu} = \bar{x} = 1.52$, $\hat{\sigma}^2 = \dfrac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n} = .8476$.

The maximized log-likelihood for the sample is -13.363. A useful shortcut for computing the log-likelihood at the hypothesized values is $\sum_{i=1}^{n}(x_i - \mu)^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 + n\left(\bar{x} - \mu\right)^2$. For the hypothesized value of μ = 1.8, this is $\sum_{i=1}^{n}(x_i - 1.8)^2 = 9.26$. The log-likelihood is -5(ln(2π) - 5(ln.8) - (1/1.6)9.26 = -13.861. The likelihood ratio statistic is $-2(\ln L_r - \ln L_u) = .996$. The critical value for a chi-squared with 2 degrees of freedom is 5.99, so we would not reject the hypothesis.

3. Suppose that the following sample is drawn from a normal distribution with mean μ and standard deviation σ: y = 3.1, -.1, .3, 1.4, 2.9, .3, 2.2, 1.5, 4.2, .4. Test the hypothesis that the mean of the distribution which produced these data is the same as that which produced the data in Exercise 1. Test the hypothesis assuming that the variances are the same. Test the hypothesis that the variances are the same using an F test and using a likelihood ratio test. (Do not assume that the means are the same.)

If the variances are the same,

$$\overline{x}_1 \sim N[\mu_1, \sigma_1^2 / n_1] \text{ and } \overline{x}_2 \sim N[\mu_2, \sigma_2^2 / n_2],$$

$$\overline{x_1} - \overline{x_2} \sim N[\mu_1 - \mu_2, \sigma^2\{(1/n_1) + (1/n_2)\}],$$

$$(n_1\text{-}1)s_1^2/\sigma^2 \sim \chi^2[n_1\text{-}1] \text{ and } (n_2\text{-}1)s_2^2/\sigma^2 \sim \chi^2[n_2\text{-}1]$$

$$(n_1\text{-}1)s_1^2/\sigma^2 + (n_2\text{-}1)s_2^2/\sigma^2 \sim \chi^2[n_1 + n_2 - 2]$$

Thus, the statistic
$$t = \frac{\left\{\left(\overline{x_1} - \overline{x_2}\right) - \left(\mu_1 - \mu_2\right)\right\} / \sqrt{\sigma^2\left[(1/n_1) + (1/n_2)\right]}}{\sqrt{\left\{(n_1 - 1)s_1^2 / \sigma^2 + (n_2 - 1)s_2^2 / \sigma^2\right\} / (n_1 + n_2 - 2)}}$$

is the ratio of a standard normal variable to the square root of a chi-squared variable divided by its degrees of freedom which is distributed as t with $n_1 + n_2 - 2$ degrees of freedom. Under the hypothesis that the means are equal, the statistic is
$$t = \frac{\left(\overline{x_1} - \overline{x_2}\right) / \sqrt{(1/n_1) + (1/n_2)}}{\sqrt{\left\{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2\right\} / (n_1 + n_2 - 2)}}$$

The sample statistics are $n_1 = 10$, $\overline{x}_1 = 1.52$, $s_1^2 = .9418$

$n_2 = 10$, $\overline{x}_2 = 1.62$, $s_2^2 = 2.0907$

so $t[18]$ = .1816. This is quite small, so we would not reject the hypothesis of equal means.

For random sampling from two normal distributions, under the hypothesis of equal variances, the statistic $F[n_1\text{-}1, n_2\text{-}1] = \dfrac{\left[(n_1 - 1)s_1^2 / \sigma^2\right] / (n_1 - 1)}{\left[(n_2 - 1)s_2^2 / \sigma^2\right] / (n_2 - 1)}$ is the ratio of two independent chi-squared variables, each divided by its degrees of freedom. This has the F distribution with n_1-1 and n_2-1 degrees of freedom. If $n_1 = n_2$, the statistic reduces to $F[n_1\text{-}1, n_2\text{-}1] = s_1^2 / s_2^2$. For our purposes, it is more convenient to put the larger variance in the denominator. Thus, for our sample data, $F[9,9]$ = 2.0907 / .9418 = 2.2199. The 95% critical value from the F table is 3.18. Thus, we would not reject the hypothesis of equal variances.

The likelihood ratio test is based on the test statistic $\lambda = -2(\ln L_r - \ln L_u)$. The log-likelihood for the joint sample of 20 observations is the sum of the two separate log-likelihoods if the samples are assumed to be independent. A useful shortcut for computing the log-likelihood arises when the maximum likelihood estimates are inserted: At the maximum likelihood estimates, $\ln L = (-n/2)[1 + \ln(2\pi) + \ln\hat{\sigma}^2]$. So, the log-likelihood for the sample is $\ln L_2 = (-5/2)[1 + \ln(2\pi) + \ln((9/10)2.0907)] = -17.35007$. (Remember, we don't make the degrees of freedom correction for the variance estimator.) The log-likelihood function for the sample of 20 observations is just the sum of the two log-likelihoods if the samples are completely independent. The unrestricted log-likelihood function is, thus, -13.363+(-17.35001) = -30.713077. To compute the restricted log-likelihood function, we need the pooled estimator which does not assume that the means are identical. This would be

$$\hat{\sigma}^2 = [(n_1\text{-}1)s_1^2 + (n_2\text{-}1)s_2^2]/[n_1 + n_2]$$
$$= [9(.9418) + 9(2.0907)]/20 = 1.36463.$$

So, the restricted log-likelihood is $\ln L_r = (-20/2)[1 + \ln(2\pi) + \ln(1.36463)] = -31.4876$. Minus twice the difference is λ = -2[-31.4876 - (-30.713077)] = 1.541. This is distributed as chi-squared with one degree of freedom. The critical value is 3.84, so we would not reject the hypothesis.

4. A common method of simulating random draws from the standard normal distribution is to compute the sum of 12 draws from the uniform [0,1] distribution and subtract 6. Can you justify this procedure?

The uniform distribution has mean 2 and variance 1/12. Therefore, the statistic $12(\bar{x} - 1/2) = \sum_{i=1}^{12} x_i - 6$ is equivalent to $z = \sqrt{n}\,(\bar{x} - \mu) / \sigma$. As $n\to\infty$, this converges to a standard normal variable. Experience suggests that a sample of 12 is large enough to approximate this result. However, more recently developed random number generators usually use different procedures based on the truncation error which occurs in representing real numbers in a digital computer.

5. Using the data in Exercise 1, form confidence intervals for the mean and standard deviation.

Since the underlying distribution is normal, we may use the t distribution. Using (4-57), we obtain a 95% confidence interval for the mean of $1.52 - 2.262[.97/\sqrt{10}\,] \leq \mu \leq 1.52 + 2.262[.97/\sqrt{10}\,]$ or $.826 \leq \mu \leq 2.214$. Using the procedure in Example 4.30, we obtain a 95% confidence for σ^2 of $9(.941)/19.02 \leq \sigma^2 \leq 9(.941)/2.70$ or $.445 \leq \sigma^2 \leq 3.137$. Taking square roots gives the confidence interval for σ, $.667 \leq \sigma \leq 1.771$.

6. Based on a sample of 65 observations from a normal distribution, you obtain a *median* of 34 and a standard deviation of 13.3. Form a confidence interval for the mean. (**Hint**: Use the asymptotic distribution. See Example 4.15.) Compare your confidence interval to the one you would have obtained had the estimate of 34 been the sample mean instead of the sample median.

The asymptotic variance of the median is $\pi\sigma^2/(2n)$. Using the asymptotic normal distribution instead of the t distribution, the confidence interval is $34 - 1.96(13.3^2\pi/130)^2 \leq \mu \leq 34 + 1.96(13.3^2\pi/130)^2$ or $29.95 \leq \mu \leq 38.052$. Had the estimator been the mean instead of the median, the appropriate asymptotic variance would be σ^2/n, instead, which we would estimate with $13.3^2/65 = 2.72$ compared to 4.274 for the median. The confidence interval would have been (30.77,37.24), which is somewhat narrower.

7. The random variable x has a continuous distribution $f(x)$ and cumulative distribution function $F(x)$. What is the probability distribution of the sample maximum? (**Hint:** In a random sample of n observations, $x_1, x_2, ..., x_n$, if z is the maximum, then every observation in the sample is less than or equal to z. Use the cdf.)

If z is the maximum, then every sample observation is less than or equal to z. The probability of this is $\text{Prob}[x_1 \# z, x_2 \# z, ..., x_n \# z] = F(z)F(z)...F(z) = [F(z)]^n$. The density is the derivative, $n[F(z)]^{n-1}f(z)$.

8. Assume the distribution of x is $f(x) = 1/\theta, 0 \leq x \leq \theta$. In random sampling from this distribution, prove that the sample maximum is a consistent estimator of θ. Note: you can prove that the maximum is the maximum likelihood estimator of θ. But, the usual properties do not apply here. Why not? (**Hint:** Attempt to verify that the expected first derivative of the log-likelihood with respect to θ is zero.)

Using the result of the previous problem, the density of the maximum is

$$n[z/\theta]^{n-1}(1/\theta),\ 0 < z < \theta.$$

Therefore, the expected value is $E[z] = \int_0^\theta z^n dz = [\theta^{n+1}/(n+1)][n/\theta^n] = n\theta/(n+1)$. The variance is found likewise, $E[z^2] = \int_0^\theta z^2 n(z/n)^{n-1}(1/\theta)dz = n\theta^2/(n+2)$ so $\text{Var}[z] = E[z^2] - (E[z])^2 = n\theta^2/[(n+1)^2(n+2)]$. Using mean squared convergence we see that $\lim_{n\to\infty} E[z] = \theta$ and $\lim_{n\to\infty} \text{Var}[z] = 0$, so that $\text{plim}\, z = \theta$.

9. In random sampling from the exponential distribution, $f(x) = \frac{1}{\theta}e^{\frac{-x}{\theta}}$, $x > 0$, $\theta > 0$, find the maximum likelihood estimator of θ and obtain the asymptotic distribution of this estimator.

The log-likelihood is $\ln L = -n\ln\theta - (1/\theta)\sum_{i=1}^n x_i$. The maximum likelihood estimator is obtained as the solution to $\partial\ln L/\partial\theta = -n/\theta + (1/\theta^2)\sum_{i=1}^n x_i = 0$, or $\hat{\theta}_{ML} = (1/n)\sum_{i=1}^n x_i = \bar{x}$. The asymptotic variance of the MLE is $\{-E[\partial^2\ln L/\partial\theta^2]\}^{-1} = \{-E[n/\theta^2 - (2/\theta^3)\sum_{i=1}^n x_i\,]\}^{-1}$. To find the expected value of this random variable, we need $E[x_i] = \theta$. Therefore, the asymptotic variance is θ^2/n. The asymptotic distribution is normal with mean θ and this variance.

10. Suppose in a sample of 500 observations from a normal distribution with mean μ and standard deviation σ, you are told that 35% of the observations are less than 2.1 and 55% of the observations are less than 3.6. Estimate μ and σ.

If 35% of the observations are less than 2.1, we would infer that

$$\Phi[(2.1 - \mu)/\sigma] = .35, \text{ or } (2.1 - \mu)/\sigma = -.385 \Rightarrow 2.1 - \mu = -.385\sigma.$$

Likewise,

$$\Phi[(3.6 - \mu)/\sigma] = .55, \text{ or } (3.6 - \mu)/\sigma = .126 \Rightarrow 3.6 - \mu = .126\sigma.$$

The joint solution is $\hat{\mu} = 3.2301$ and $\hat{\sigma} = 2.9354$. It might not seem obvious, but we can also derive asymptotic standard errors for these estimates by constructing them as method of moments estimators. Observe, first, that the two estimates are based on moment estimators of the probabilities. Let x_i denote one of the 500 observations drawn from the normal distribution. Then, the two proportions are obtained as follows: Let $z_i(2.1) = \mathbf{1}[x_i < 2.1]$ and $z_i(3.6) = \mathbf{1}[x_i < 3.6]$ be indicator functions. Then, the proportion of 35% has been obtained as $\bar{z}(2.1)$ and .55 is $\bar{z}(3.6)$. So, the two proportions are simply the means of functions of the sample observations. Each z_i is a draw from a Bernoulli distribution with success probability $\pi(2.1) = \Phi((2.1-\mu)/\sigma)$ for $z_i(2.1)$ and $\pi(3.6) = \Phi((3.6-\mu)/\sigma)$ for $z_i(3.6)$. Therefore, $E[\bar{z}(2.1)] = \pi(2.1)$, and $E[\bar{z}(3.6)] = \pi(3.6)$. The variances in each case are $\text{Var}[\bar{z}(.)] = 1/n[\pi(.)(1-\pi(.))]$. The covariance of the two sample means is a bit trickier, but we can deduce it from the results of random sampling. $\text{Cov}[\bar{z}(2.1), \bar{z}(3.6)]]$ $= 1/n\ \text{Cov}[z_i(2.1),z_i(3.6)]$, and, since in random sampling sample moments will converge to their population counterparts, $\text{Cov}[z_i(2.1),z_i(3.6)] = \text{plim}\ [\{(1/n)\sum_{i=1}^{n} z_i(2.1)z_i(3.6)\} - \pi(2.1)\pi(3.6)]$. But, $z_i(2.1)z_i(3.6)$ must equal $[z_i(2.1)]^2$ which, in turn, equals $z_i(2.1)$. It follows, then, that $\text{Cov}[z_i(2.1),z_i(3.6)] = \pi(2.1)[1 - \pi(3.6)]$. Therefore, the asymptotic covariance matrix for the two sample proportions is $Asy.Var[p(2.1), p(3.6)] = \mathbf{\Sigma} = \frac{1}{n}\begin{bmatrix} \pi(2.1)(1-\pi(2.1)) & \pi(2.1)(1-\pi(3.6)) \\ \pi(2.1)(1-\pi(3.6)) & \pi(3.6)(1-\pi(3.6)) \end{bmatrix}$. If we insert our sample estimates, we obtain $Est.Asy.Var[p(2.1), p(3.6)] = \mathbf{S} = \begin{bmatrix} 0.000455 & 0.000315 \\ 0.000315 & 0.000495 \end{bmatrix}$. Now, ultimately, our estimates of μ and σ are found as functions of $p(2.1)$ and $p(3.6)$, using the method of moments. The moment equations are

$$m_{2.1} = \left[\frac{1}{n}\sum_{i=1}^{n} z_i(2.1)\right] - \Phi\left[\frac{2.1-\mu}{\sigma}\right] = 0,$$

$$m_{3.6} = \left[\frac{1}{n}\sum_{i=1}^{n} z_i(3.6)\right] - \Phi\left[\frac{3.6-\mu}{\sigma}\right] = 0.$$

Now, let $\mathbf{\Gamma} = \begin{bmatrix} \partial m_{2.1}/\partial\mu & \partial m_{2.1}/\partial\sigma \\ \partial m_{3.6}/\partial\mu & \partial m_{3.61}/\partial\sigma \end{bmatrix}$ and let $\mathbf{G}$ be the sample estimate of Γ. Then, the estimator of the asymptotic covariance matrix of $(\hat{\mu}, \hat{\sigma})$ is $[\mathbf{GS}^{-1}\mathbf{G}']^{-1}$. The remaining detail is the derivatives, which are just $\partial m_{2.1}/\partial\mu = (1/\sigma)\phi((2.1-\mu)/\sigma)$ and $\partial m_{2.1}/\partial\sigma = (2.1-\mu)/\sigma[\text{M}m_{2.1}/\text{M}\sigma]$ and likewise for $m_{3.6}$. Inserting our sample estimates produces $\mathbf{G} = \begin{bmatrix} 0.37046 & -0.14259 \\ 0.39579 & 0.04987 \end{bmatrix}$. Finally, multiplying the matrices and computing the necessary inverses produces $[\mathbf{GS^{-1}G'}]^{-1} = \begin{bmatrix} 0.10178 & -0.12492 \\ -0.12492 & 0.16973 \end{bmatrix}$. The asymptotic distribution would be normal, as usual. Based on these results, a 95% confidence interval for μ would be $3.2301 \pm 1.96(.10178)^2 =$ 2.6048 to 3.8554.

11. For random sampling from a normal distribution with nonzero mean μ and standard deviation σ, find the asymptotic joint distribution of the maximum likelihood estimators of σ/μ and μ^2/σ^2.

The maximum likelihood estimators, $\hat{\mu} = (1/n)\sum_{i=1}^{n} x_i$ and $\hat{\sigma}^2 = (1/n)\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$ were given in (4-49). By the invariance principle, we know that the maximum likelihood estimators of μ/σ and μ^2/σ^2 are $\hat{\mu}/\hat{\sigma}$ and $\hat{\mu}/\hat{\sigma}^2$ and the maximum likelihood estimate of σ is $\sqrt{\hat{\sigma}}$. To obtain the asymptotic joint distribution of the two functions of $\hat{\mu}$ and $\hat{\sigma}$, we first require the asymptotic joint distribution of $\hat{\mu}$ and $\hat{\sigma}^2$. This is normal with mean vector (μ,σ^2) and covariance matrix equal to the inverse of the information matrix. This is the inverse of

$$-E\begin{bmatrix} \partial^2 \log L/\partial\mu^2 & \partial^2 \log L/\partial\mu\partial\sigma^2 \\ \partial^2 \log L/\partial\sigma^2\partial\mu & \partial^2 \log L/\partial(\sigma^2)^2 \end{bmatrix} = \begin{bmatrix} -n/\sigma^2 & -(1/\sigma^3)\sum_{i=1}^{n}(x_i-\mu) \\ -(1/\sigma^3)\sum_{i=1}^{n}(x_i-\mu) & n/(2\sigma^4)-(1/\sigma^6)\sum_{i=1}^{n}\left(x_i-\mu\right)^2 \end{bmatrix}$$

The off diagonal term has expected value 0. Each term in the sum in the lower right has expected value σ^2, so, after collecting terms, taking the negative, and inverting, we obtain the asymptotic covariance matrix,

$\mathbf{V} = \begin{bmatrix} \sigma^2/n & 0 \\ 0 & 2\sigma^4/n \end{bmatrix}$. To obtain the asymptotic joint distribution of the two nonlinear functions, we use the multivariate version of Theorem 4.4. Thus, we require $\mathbf{H} = \mathbf{JVJ}'$ where

$\mathbf{J} = \begin{bmatrix} \partial(\mu/\sigma)/\partial\mu & \partial(\mu/\sigma)/\partial\sigma^2 \\ \partial(\mu^2/\sigma^2)/\partial\mu & \partial(\mu^2/\sigma^2)/\partial\sigma^2 \end{bmatrix} = \begin{bmatrix} 1/\sigma & -\mu/(2\sigma^3) \\ 2\mu/\sigma^2 & -\mu/\sigma^4 \end{bmatrix}$. The product is

$$\mathbf{H} = \frac{1}{n}\begin{bmatrix} 1+\mu^2/(2\sigma^2) & 2\mu/\sigma+(\mu/\sigma)^3 \\ 2\mu/\sigma+(\mu/\sigma)^3 & 4\mu^2/\sigma^2+2\mu^4/\sigma^4 \end{bmatrix}.$$

12. The random variable x has the following distribution: $f(x) = e^{-\lambda}\lambda^x / x!$, $x = 0,1,2,...$ The following random sample is drawn: 1,1,4,2,0,0,3,2,3,5,1,2,1,0,0. Carry out a Wald test of the hypothesis that λ= 2.

For random sampling from the Poisson distribution, the maximum likelihood estimator of λ is $\bar{x}$ = 25/15. (See Example 4.18.) The second derivative of the log-likelihood is $-\sum_{i=1}^{n} x_i/\lambda^2$, so the the asymptotic variance is λ/n. The Wald statistic would be

$$W = \frac{\left(\bar{x}-2\right)^2}{\hat{\lambda}/n} = [(25/15 - 2)^2]/[(25/15)/15] = 1.0.$$

The 95% critical value from the chi-squared distribution with one degree of freedom is 3.84, so the hypothesis would not be rejected. Alternatively, one might estimate the variance of with s^2/n = 2.38/15 = 0.159. Then, the Wald statistic would be $(1.6 - 2)^2/.159$ = 1.01. The conclusion is the same.

13. Based on random sampling of 16 observations from the exponential distribution of Exercise 9, we wish to test the hypothesis that θ =1. We will reject the hypothesis if $\bar{x}$ is greater than 1.2 or less than .8. We are interested in the power of this test.

(a) Using the asymptotic distribution of $\bar{x}$ graph the asymptotic approximation to the true power function.

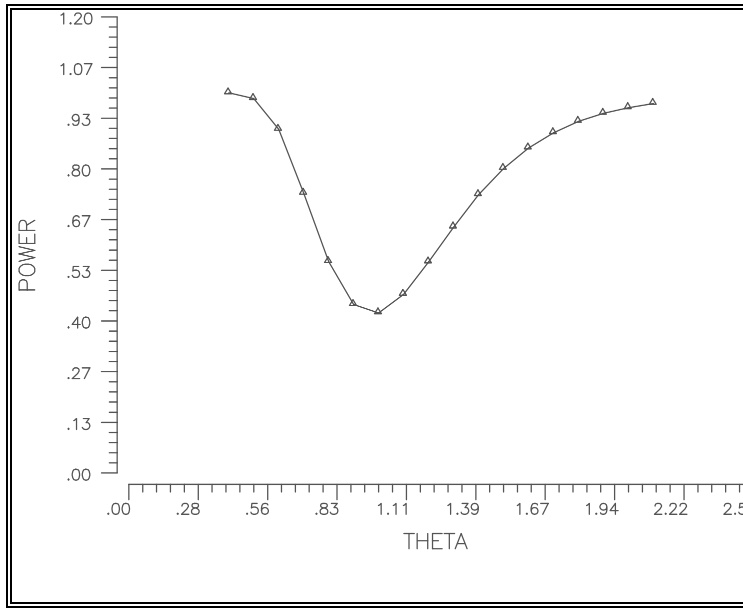
(b) Using the result discussed in Example 4.17, describe how to obtain the true power function for this test.

The asymptotic distribution of $\bar{x}$ is normal with mean θ and variance θ^2/n. Therefore, the power function based on the asymptotic distribution is the probability that a normally distributed variable with mean equal to θ and variance equal to θ^2/n will be greater than 1.2 or less than .8. That is,

$$\text{Power} = \Phi[(.8 - \theta)/(\theta/4)] + 1 - \Phi[(1.2 - \theta)/(\theta/4)].$$

Some values of this power function and a sketch are given below:

θ	Approx. Power	True Power
.4	1.000	1.000
.5	.992	.985
.6	.908	.904
.7	.718	.736
.8	.522	.556
.9	.420	.443
1.0	.423	.421
1.1	.496	.470
1.2	.591	.555
1.3	.685	.647
1.4	.759	.732
1.5	.819	.801
1.6	.864	.855
1.7	.897	.895
1.8	.922	.925
1.9	.940	.946
2.0	.954	.961
2.1	.963	.972



Note that the power function does not have the symmetric shape of Figure 4.7 because both the variance and the mean are changing as θ changes. Moreover, the power is not the lowest at the value of $\theta = 1$, but at about $\theta = .9$. That means (assuming that the normal distribution is appropriate) that the test is slightly biased. The size of the test is its power at the hypothesized value, or .423, and there are points at which the power is less than the size.

According to the example cited, the true distribution of $\bar{x}$ is that of $\theta/(2n)$ times a chi-squared variable with $2n$ degrees of freedom. Therefore, we could find the true power by finding the probability that a chi-squared variable with $2n$ degrees of freedom is less than $.8(2n/\theta)$ or greater than $1.2(2n/\theta)$. Thus,

$$\text{True power} = F(25.6/\theta) + 1 - F(38.4/\theta)$$

where $F(.)$ is the CDF of the chi-squared distribution with 32 degrees of freedom. Values for the correct power function are shown above. Given that the sample is only 16 observations, the closeness of the asymptotic approximation is quite impressive.

14. For the normal distribution, $\mu_{2k} = \sigma^{2k}(2k)!/(k!2^k)$ and $\mu_{2k+1} = 0$, $k = 0,1,...$ Use this result to show that in Example 4.27, $\theta_1 = 0$ and $\theta_2 = 3$, and $\mathbf{JVJ'} = \begin{bmatrix} 6 & 0 \\ 0 & 24 \end{bmatrix}$.

For θ_1 and θ_2, just plug in the result above using k = 2, 3, and 4. The example involves 3 moments, m_2, m_3, and m_4. The asymptotic covariance matrix for these three moments can be based on the formulas given in Example 4.26. In particular, we note, first, that for the normal distribution, Asy.Cov$[m_2,m_3]$ and Asy.Cov$[m_3,m_4]$ will be zero since they involve only odd moments, which are all zero. The necessary even moments are $\mu_2 = \sigma^2$, $\mu_4 = 3\sigma^4$. $\mu_6 = 15\sigma^6$, $\mu_8 = 105\sigma^8$. The three variances will be

$$n[\text{Asy.Var}(m_2)] = \mu_4 - \mu_2^2 = 3\sigma^4 - (\sigma^2)^2 = 2\sigma^4$$
$$n[\text{Asy.Var}(m_3)] = \mu_6 - \mu_3^2 - 6\mu_4\mu_2 + 9\mu_2^3 = 6\sigma^6$$
$$n[\text{Asy.Var}(m_4)] = \mu_8 - \mu_4^2 - 8\mu_5\mu_3 + 16\mu_2\mu_3^2 = 96\sigma^8$$

and $$n[\text{Asy.Cov}(m_2,m_4)] = \mu_6 - \mu_2\mu_4 - 4\mu_3^2 = 12\sigma^6.$$

The elements of **J** are given in Example 4.27. For the normal distribution, this matrix would be $\mathbf{J} = \begin{bmatrix} 0 & 1/\sigma^3 & 0 \\ -6/\sigma^2 & 0 & 1/\sigma^4 \end{bmatrix}$. Multiplying out **JVJN** produces the result given above.

15. Testing for normality. One method that has been suggested for testing whether the distribution underlying a sample is normal is to refer the statistic $L = n\{\text{skewness}^2/6 + (\text{kurtosis-3})^2/24\}$ to the chi-squared distribution with 2 degrees of freedom. Using the data in Exercise 1, carry out the test.

The skewness coefficient is .14192 and the kurtosis is 1.8447. (These are the third and fourth moments divided by the third and fourth power of the sample standard deviation.) Inserting these in the expression above produces L = 10{.14192^2/6 + (1.8447 - 3)^2/24} = .59. The critical value from the chi-squared distribution with 2 degrees of freedom (95%) is 5.99. Thus, the hypothesis of normality cannot be rejected.

16. Suppose the joint distribution of the two random variables x and y is

$$f(x,y) = \theta e^{-(\beta+\theta)y}(\beta y)^x / x! \quad \beta,\theta \;\; 0,\; y \; \$ \; 0,\; x = 0,1,2,...$$

(a) Find the maximum likelihood estimators of β and θ and their asymptotic joint distribution.
(b) Find the maximum likelihood estimator of θ/(β+θ) and its asymptotic distribution.
(c) Prove that $f(x)$ is of the form $f(x) = \gamma(1-\gamma)^x$, x = 0,1,2,...
Then, find the maximum likelihood estimator of γ and its asymptotic distribution.
(d) Prove that f(y*x) is of the form $\lambda e^{-\lambda y}(\lambda y)^x/x!$ Prove that $f(y|x)$ integrates to 1. Find the maximum likelihood estimator of λ and its asymptotic distribution. (**Hint**: In the conditional distribution, just carry the xs along as constants.)
(e) Prove that $f(y) = \theta e^{-\theta y}$ then find the maximum likelihood estimator of θ and its asymptotic variance.
(f) Prove that $f(x|y) = e^{-\beta y}(\beta y)^x/x!$. Based on this distribution, what is the maximum likelihood estimator of β?

The log-likelihood is $\ln L = n\ln\theta - (\beta+\theta)\sum_{i=1}^n y_i + \ln\beta\sum_{i=1}^n x_i + \sum_{i=1}^n x_i \log y_i - \sum_{i=1}^n \log(x_i!)$

The first and second derivatives are

$$\partial\ln L/\partial\theta = n/\theta - \sum_{i=1}^n y_i$$
$$\partial\ln L/\partial\beta = -\sum_{i=1}^n y_i + \sum_{i=1}^n x_i/\beta$$
$$\partial^2\ln L/\partial\theta^2 = -n/\theta^2$$
$$\partial^2\ln L/\partial\beta^2 = -\sum_{i=1}^n x_i/\beta^2$$
$$\partial^2\ln L/\partial\beta\partial\theta = 0.$$

Therefore, the maximum likelihood estimators are $\hat{\theta} = 1/\bar{y}$ and $\hat{\beta} = \bar{x}/\bar{y}$ and the asymptotic covariance matrix is the inverse of $E\begin{bmatrix} n/\theta^2 & 0 \\ 0 & \sum_{i=1}^n x_i/\beta^2 \end{bmatrix}$. In order to complete the derivation, we will require the expected value of $\sum_{i=1}^n x_i = nE[x_i]$. In order to obtain $E[x_i]$, it is necessary to obtain the marginal distribution of x_i, which is f(x) = $\int_0^\infty \theta e^{-(\beta+\theta)y}(\beta y)^x / x!\,dy = \beta^x(\theta/x!)\int_0^\infty e^{-(\beta+\theta)y}y^x dy$. This is $\beta^x(\theta/x!)$ times a gamma integral. This is $f(x) = \beta^x(\theta/x!)[\Gamma(x+1)]/(\beta+\theta)^{x+1}$. But, $\Gamma(x+1) = x!$, so the expression reduces to

$$f(x) = [\theta/(\beta+\theta)][\beta/(\beta+\theta)]^x.$$

Thus, x has a geometric distribution with parameter π = θ/(β+θ). (This is the distribution of the number of tries until the first success of independent trials each with success probability 1-π. Finally, we require the expected value of x_i, which is $E[x] = [\theta/(\beta+\theta)]\sum_{x=0}^\infty x[\beta/(\beta+\theta)]^x = \beta/\theta$. Then, the required asymptotic covariance matrix is $\begin{bmatrix} n/\theta^2 & 0 \\ 0 & n(\beta/\theta)/\beta^2 \end{bmatrix}^{-1} = \begin{bmatrix} \theta^2/n & 0 \\ 0 & \beta\theta/n \end{bmatrix}$.

The maximum likelihood estimator of θ/(β+θ) is

$$\widehat{\theta/(\beta+\theta)} = (1/\bar{y})/[\bar{x}/\bar{y} + 1/\bar{y}] = 1/(1+\bar{x}).$$

Its asymptotic variance is obtained using the variance of a nonlinear function

$$V = [\beta/(\beta+\theta)]^2(\theta^2/n) + [-\theta/(\beta+\theta)]^2(\beta\theta/n) = \beta\theta^2/[n(\beta+\theta)^3].$$

The asymptotic variance could also be obtained as $[-1/(1 + E[x])^2]^2$Asy.Var[$\bar{x}$].)

For part (c), we just note that $\gamma = \theta/(\beta+\theta)$. For a sample of observations on x, the log-likelihood would be

$$\ln L = n\ln\gamma + \ln(1-\gamma)\sum_{i=1}^{n} x_i$$

$$\partial \ln L/d\gamma = n/\gamma - \sum_{i=1}^{n} x_i /(1-\gamma).$$

A solution is obtained by first noting that at the solution, $(1-\gamma)/\gamma = \bar{x} = 1/\gamma - 1$. The solution for γ is, thus, $\hat{\gamma} = 1/(1+\bar{x})$.Of course, this is what we found in part b., which makes sense.

For part (d) $f(y|x) = \dfrac{f(x,y)}{f(x)} = \dfrac{\theta e^{-(\beta+\theta)y}(\beta y)^x(\beta+\theta)^x(\beta+\theta)}{x! \quad \theta \quad \beta x}$. Cancelling terms and gathering the remaining like terms leaves $f(y|x) = (\beta+\theta)[(\beta+\theta)y]^x e^{-(\beta+\theta)y}/x!$ so the density has the required form with $\lambda = (\beta+\theta)$. The integral is $\left\{[\lambda^{x+1}]/x!\right\}\int_0^\infty e^{-\lambda y}y^x dy$. This integral is a Gamma integral which equals $\Gamma(x+1)/\lambda^{x+1}$, which is the reciprocal of the leading scalar, so the product is 1. The log-likelihood function is

$$\ln L = n\ln\lambda - \lambda\sum_{i=1}^{n} y_i + \ln\lambda\sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\ln x_i!$$

$$\partial\ln L/\partial\lambda = (\sum_{i=1}^{n} x_i + n)/\lambda - \sum_{i=1}^{n} y_i .$$

$$\partial^2\ln L/\partial\lambda^2 = -(\sum_{i=1}^{n} x_i + n)/\lambda^2.$$

Therefore, the maximum likelihood estimator of λ is $(1+\bar{x})/\bar{y}$ and the asymptotic variance, conditional on the xs is Asy.Var.$\left[\hat{\lambda}\right] = (\lambda^2/n)/(1+\bar{x})$

Part (e.) We can obtain $f(y)$ by summing over x in the joint density. First, we write the joint density as $f(x,y) = \theta e^{-\theta y}e^{-\beta y}(\beta y)^x/x!$. The sum is, therefore, $f(y) = \theta e^{-\theta y}\sum_{x=0}^{\infty} e^{-\beta y}(\beta y)^x/x!$. The sum is that of the probabilities for a Poisson distribution, so it equals 1. This produces the required result. The maximum likelihood estimator of θ and its asymptotic variance are derived from

$$\ln L = n\ln\theta - \theta\sum_{i=1}^{n} y_i$$

$$\partial\ln L/\partial\theta = n/\theta - \sum_{i=1}^{n} y_i$$

$$\partial^2\ln L/\partial\theta^2 = -n/\theta^2.$$

Therefore, the maximum likelihood estimator is $1/\bar{y}$ and its asymptotic variance is θ^2/n. Since we found $f(y)$ by factoring $f(x,y)$ into $f(y)f(x|y)$ (apparently, given our result), the answer follows immediately. Just divide the expression used in part e. by $f(y)$. This is a Poisson distribution with parameter βy. The log-likelihood function and its first derivative are

$$\ln L = -\beta\sum_{i=1}^{n} y_i + \ln\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} x_i \ln y_i - \sum_{i=1}^{n}\ln x_i!$$

$$\partial\ln L/\partial\beta = -\sum_{i=1}^{n} y_i + \sum_{i=1}^{n} x_i /\beta,$$

from which it follows that $\hat{\beta} = \bar{x}/\bar{y}$.

17. Suppose x has the Weibull distribution, $f(x) = \alpha\beta x^{\beta-1}\exp(-\alpha x^\beta)$, $x, \alpha, \beta > 0$.
(a) Obtain the log-likelihood function for a random sample of n observations.
(b) Obtain the likelihood equations for maximum likelihood estimation of α and β. Note that the first provides an explicit solution for α in terms of the data and β. But, after inserting this in the second, we obtain only an implicit solution for β. How would you obtain the maximum likelihood estimators?
(c) Obtain the second derivatives matrix of the log-likelihood with respect to α and β. The exact expectations of the elements involving β involve the derivatives of the Gamma function and are quite messy analytically. Of course, your exact result provides an empirical estimator. How would you estimate the asymptotic covariance matrix for your estimators in part (b)?

(d) Prove that $\alpha\beta\text{Cov}[\ln x, x^{\beta}] = 1$. (**Hint:** Use the fact that the expected first derivatives of the log-likelihood function are zero.)

The log-likelihood and its two first derivatives are

$$\log L = n\log\alpha + n\log\beta + (\beta-1)\sum_{i=1}^{n}\log x_i - \alpha\sum_{i=1}^{n} x_i^{\beta}$$

$$\partial\log L/\partial\alpha = n/\alpha - \sum_{i=1}^{n} x_i^{\beta}$$

$$\partial\log L/\partial\beta = n/\beta + \sum_{i=1}^{n}\log x_i - \alpha\sum_{i=1}^{n}(\log x_i)x_i^{\beta}$$

Since the first likelihood equation implies that at the maximum, $\hat{\alpha} = n/\sum_{i=1}^{n} x_i^{\beta}$, one approach would be to scan over the range of β and compute the implied value of α. Two practical complications are the allowable range of β and the starting values to use for the search.

The second derivatives are

$$\partial^2\ln L/\partial\alpha^2 = -n/\alpha^2$$

$$\partial^2\ln L/\partial\beta^2 = -n/\beta^2 - \alpha\sum_{i=1}^{n}(\log x_i)^2 x_i^{\beta}$$

$$\partial^2\ln L/\partial\alpha\partial\beta = -\sum_{i=1}^{n}(\log x_i)x_i^{\beta}\ .$$

If we had estimates in hand, the simplest way to estimate the expected values of the Hessian would be to evaluate the expressions above at the maximum likelihood estimates, then compute the negative inverse. First, since the expected value of $\partial\ln L/\partial\alpha$ is zero, it follows that $E[x_i^{\beta}] = 1/\alpha$. Now,

$$E[\partial\ln L/\partial\beta] = n/\beta + E[\sum_{i=1}^{n}\log x_i] - \alpha E[\sum_{i=1}^{n}(\log x_i)x_i^{\beta}] = 0$$

as well. Divide by n, and use the fact that every term in a sum has the same expectation to obtain

$$1/\beta + E[\ln x_i] - E[(\ln x_i)x_i^{\beta}]/E[x_i^{\beta}] = 0.$$

Now, multiply through by $E[x_i^{\beta}]$ to obtain $E[x_i^{\beta}] = E[(\ln x_i)x_i^{\beta}] - E[\ln x_i]E[x_i^{\beta}]$

or $\quad 1/(\alpha\beta) = \text{Cov}[\ln x_i, x_i^{\beta}]$.

18. The following data were generated by the Weibull distribution of Exercise 17:

1.3043	.49254	1.2742	1.4019	.32556	.29965	.26423
1.0878	1.9461	.47615	3.6454	.15344	1.2357	.96381
.33453	1.1227	2.0296	1.2797	.96080	2.0070	

(a) Obtain the maximum likelihood estimates of α and β and estimate the asymptotic covariance matrix for the estimates.

(b) Carry out a Wald test of the hypothesis that $\beta = 1$.

(c) Obtain the maximum likelihood estimate of α under the hypothesis that $\beta = 1$.

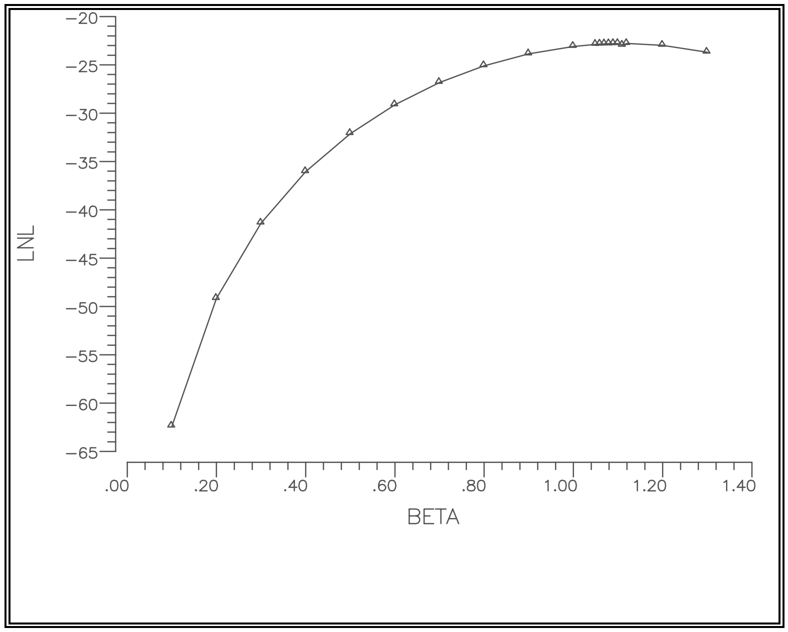
(d) Using the results of a. and c. carry out a likelihood ratio test of the hypothesis that $\beta = 1$.

(e) Carry out a Lagrange multiplier test of the hypothesis that $\beta = 1$.

As suggested in the previous problem, we can concentrate the log-likelihood over α. From $\partial\log L/\partial\alpha = 0$, we find that at the maximum, $\alpha = 1/[(1/n)\sum_{i=1}^{n} x_i^{\beta}]$. Thus, we scan over different values of β to seek the value which maximizes $\log L$ as given above, where we substitute this expression for each occurrence of α.

Values of β and the log-likelihood for a range of values of β are listed and shown in the figure below.

β	log$L$
0.1	-62.386
0.2	-49.175
0.3	-41.381
0.4	-36.051
0.5	-32.122
0.6	-29.127
0.7	-26.829
0.8	-25.098
0.9	-23.866
1.0	-23.101
1.05	-22.891
1.06	-22.863
1.07	-22.841
1.08	-22.823
1.09	-22.809
1.10	-22.800
1.11	-22.796
1.12	-22.797
1.2	-22.984
1.3	-23.693



The maximum occurs at β = 1.11. The implied value of α is 1.179. The negative of the second derivatives matrix at these values and its inverse are

$$\mathbf{I}\left(\hat{\alpha},\hat{\beta}\right) = \begin{bmatrix} 25.55 & 9.6506 \\ 9.6506 & 27.7552 \end{bmatrix} \text{ and } \mathbf{I}^{-1}\left(\hat{\alpha},\hat{\beta}\right) = \begin{bmatrix} .04506 & -.2673 \\ -.2673 & .04148 \end{bmatrix}.$$

The Wald statistic for the hypothesis that β = 1 is $W = (1.11 - 1)^2/.041477 = .276$. The critical value for a test of size .05 is 3.84, so we would not reject the hypothesis.

If β = 1, then $\hat{\alpha} = n / \sum_{i=1}^{n} x_i = 0.88496$. The distribution specializes to the geometric distribution if β = 1, so the restricted log-likelihood would be

$$\log L_r = n\log\alpha - \alpha \sum_{i=1}^{n} x_i = n(\log\alpha - 1) \text{ at the MLE.}$$

$\log L_r$ at α = .88496 is -22.44435. The likelihood ratio statistic is -2logλ = 2(23.10068 - 22.44435) = 1.3126. Once again, this is a small value. To obtain the Lagrange multiplier statistic, we would compute

$$\left[\partial \log L / \partial\alpha \quad \partial \log L / \partial\beta\right] \begin{bmatrix} -\partial^2 \log L / \partial\alpha^2 & -\partial^2 \log L / \partial\alpha\partial\beta \\ -\partial^2 \log L / \partial\alpha\partial\beta & -\partial^2 \log L / \partial\beta^2 \end{bmatrix}^{-1} \begin{bmatrix} \partial \log L / \partial\alpha \\ \partial \log L / \partial\beta \end{bmatrix}$$

at the restricted estimates of α = .88496 and β = 1. Making the substitutions from above, at these values, we would have

$$\partial\log L/\partial\alpha = 0$$

$$\partial\log L/\partial\beta = n + \sum_{i=1}^{n} \log x_i - \frac{1}{\bar{x}} \sum_{i=1}^{n} x_i \log x_i = 9.400342$$

$$\partial^2\log L/\partial\alpha^2 = -n\bar{x}^2 = -25.54955$$

$$\partial^2\log L/\partial\beta^2 = -n - \frac{1}{\bar{x}} \sum_{i=1}^{n} x_i (\log x_i)^2 = -30.79486$$

$$\partial^2\log L/\partial\alpha\partial\beta = -\sum_{i=1}^{n} x_i \log x_i = -8.265.$$

The lower right element in the inverse matrix is .041477. The LM statistic is, therefore, $(9.40032)^2.041477 = 2.9095$. This is also well under the critical value for the chi-squared distribution, so the hypothesis is not rejected on the basis of any of the three tests.

19. We consider forming a confidence interval for the variance of a normal distribution. As shown in Example 4.29, the interval is formed by finding c_{lower} and c_{upper} such that Prob[$c_{lower} < \chi^2[n-1] < c_{upper}$] = 1 - α. The endpoints of the confidence interval are then $(n-1)s^2/c_{upper}$ and $(n-1)s^2/c_{lower}$. How do we find the narrowest interval? Consider simply minimizing the width of the interval, c_{upper} - c_{lower} subject to the constraint that the probability contained in the interval is (1-α). Prove that for symmetric and asymmetric distributions alike, the narrowest interval will be such that the density is the same at the two endpoints.

The general problem is to minimize Upper - Lower subject to the constraint F(Upper) - F(Lower) = 1 - α, where F(.) is the appropriate chi-squared distribution. We can set this up as a Lagrangean problem,

$$\min_{L,U} L_* = U - L + \lambda\{(F(U) - F(L)) - (1 - \alpha)\}$$

The necessary conditions are

$$\partial L_*/\partial U = 1 + \lambda f(U) = 0$$
$$\partial L_*/\partial L = -1 - \lambda f(L) = 0$$
$$\partial L_*/\partial \lambda = (F(U) - F(L)) - (1 - \alpha) = 0$$

It is obvious from the first two that at the minimum, $f(U)$ must equal $f(L)$.

20. Using the results in Example 4.26, and Section 4.7.2, estimate the asymptotic covariance matrix of the method of moments estimators of P and λ based on m_{-1}' and m_2' . (**Note:** You will need to use the data in Table 4.1 to estimate **V**.)

Using the income data in Table 4.1, (1/n) times the covariance matrix of $1/x_i$ and x_i^2 is

$\mathbf{V} = \begin{bmatrix} .000068456 & -2.811 \\ -2.811 & 228050. \end{bmatrix}$. The moment equations used to estimate P and λ are $E[m_{-1}' - \lambda/(P-1)] = 0$ and $E[m_2' - P(P+1)/\lambda] = 0$. The matrix of derivatives with respect to P and λ is $\mathbf{G} = \begin{bmatrix} \lambda/(P-1)^2 & -\lambda/(P-1) \\ -(2P+1)/\lambda^2 & 2P(P+1)/\lambda^3 \end{bmatrix}$. The estimated asymptotic covariance matrix is

$[\mathbf{G}\mathbf{V}^{-1}\mathbf{G}']^{-1} = \begin{bmatrix} .17532 & .0073617 \\ .0073617 & .00041871 \end{bmatrix}$.

Appendix D

Large Sample Distribution Theory

There are no exercises for Appendix D.

Appendix E

Computation and Optimization

1. Show how to maximize the function

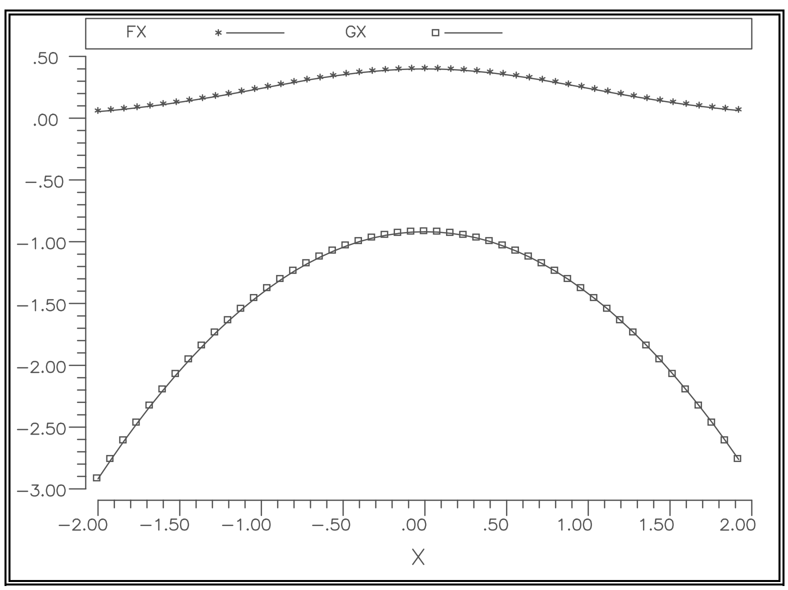
$$f(\beta) = \frac{1}{\sqrt{2\pi}} e^{-(\beta-c)^2/2}$$

with respect to β for a constant, c, using Newton's method. Show that maximizing log$f$(β) leads to the same solution. Plot f(β) and log$f$(β).

The necessary condition for maximizing f(β) is

$$df(\beta)/d\beta = \frac{1}{\sqrt{2\pi}} e^{-(\beta-c)^2/2} [-(\beta - c)] = 0 = -(\beta - c)f(\beta).$$

The exponential function can never be zero, so the only solution to the necessary condition is β = c. The second derivative is $d^2f(\beta)/d\beta^2 = -(\beta-c)df(\beta)/d\beta - f(\beta) = [(\beta-c)^2 - 1]f(\beta)$. At the stationary value $b = c$, the second derivative is negative, so this is a maximum. Consider instead the function $g(\beta) = \log f(\beta) = -(1/2)\ln(2\pi) - (1/2)(\beta - c)^2$. The leading constant is obviously irrelevant to the solution, and the quadratic is a negative number everywhere except the point β = c. Therefore, it is obvious that this function has the same maximizing value as f(β). Formally, $dg(\beta)/d\beta = -(\beta - c) = 0$ at β = c, and $d^2g(\beta)/d\beta^2 = -1$, so this is indeed the maximum. A sketch of the two functions appears below.



Note that the transformed function is concave everywhere while the original function has inflection points.

2. Prove that Newton's method for minimizing the sum of squared residuals in the linear regression model will converge to the minimum in one iteration.

The function to be maximized is $f(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$. The required derivatives are $\partial f(\boldsymbol{\beta})/\partial\boldsymbol{\beta} = -\mathbf{X}'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ and $\partial^2 f(\boldsymbol{\beta})/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}\partial = \mathbf{X}'\mathbf{X}$. Now, consider beginning a Newton iteration at an arbitrary point, $\boldsymbol{\beta}^0$. The iteration is defined in (12-17),
$\boldsymbol{\beta}^1 = \boldsymbol{\beta}^0 - (\mathbf{X}'\mathbf{X})^{-1}\{-\mathbf{X}'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^0)\} = \boldsymbol{\beta}^0 + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta}^0 = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{b}$.
Therefore, regardless of the starting value chosen, the next value will be the least squares coefficient vector.

3. For the Poisson regression model, $\text{Prob}[Y_i = y_i|\mathbf{x}_i] = \dfrac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}$ where $\lambda_i = e^{\boldsymbol{\beta}'\mathbf{x}_i}$. The log-likelihood function is $\ln L = \sum_{i=1}^n \log\text{Prob}[Y_i = y_i|\mathbf{x}_i]$.

(a) Insert the expression for λ_i to obtain the log-likelihood function in terms of the observed data.
(b) Derive the first order conditions for maximizing this function with respect to $\boldsymbol{\beta}$.
(c) Derive the second derivatives matrix of this criterion function with respect to $\boldsymbol{\beta}$. Is this matrix negative definite?
(d) Define the computations for using Newton's method to obtain estimates of the unknown parameters.
(e) Write out the full set of steps in an algorithm for obtaining the estimates of the parameters of this model. Include in your algorithm a test for convergence of the estimates based on Belsley's suggested criterion.
(f) How would you obtain starting values for your iterations?
(g) The following data are generated by the Poisson regression model with $\log\lambda = \alpha + \beta x$.

y	6	7	4	10	10	6	4	7	2	3	6	5	3	3	4
x	1.5	1.8	1.8	2.0	1.3	1.6	1.2	1.9	1.8	1.0	1.4	.5	.8	1.1	.7

Use your results from parts (a) - (f) to compute the maximum likelihood estimates of α and β. Also obtain estimates of the asymptotic covariance matrix of your estimates.

The log-likelihood is

$$\log L = \sum_{i=1}^n [-\lambda_i + y_i\ln\lambda_i - \ln y_i!] = -\sum_{i=1}^n e^{\boldsymbol{\beta}'\mathbf{x}_i} + \sum_{i=1}^n y_i(\boldsymbol{\beta}'\mathbf{x}_i) - \sum_{i=1}^n \log y_i!$$
$$= -\sum_{i=1}^n e^{\boldsymbol{\beta}'\mathbf{x}_i} + \boldsymbol{\beta}'\sum_{i=1}^n \mathbf{x}_i y_i - \sum_{i=1}^n \log y_i!$$

The necessary condition is $M\ln L/M\boldsymbol{\beta} = -\sum_{i=1}^n \mathbf{x}_i e^{\boldsymbol{\beta}'\mathbf{x}_i} + \sum_{i=1}^n \mathbf{x}_i y_i = \mathbf{0}$ or $\mathbf{X}N\mathbf{y} = \sum_{i=1}^n \mathbf{x}_i\lambda_i$. It is useful to note, since $E[y_i*\mathbf{x}_i] = \lambda_i = e^{\boldsymbol{\beta}N\mathbf{x}i}$, the first order condition is equivalent to $\sum_{i=1}^n \mathbf{x}_i y_i = \sum_{i=1}^n \mathbf{x}_i E[y_i*\mathbf{x}_i]$ or $\mathbf{X}N\mathbf{y} = \mathbf{X}NE[\mathbf{y}]$, which makes sense. We may write the first order condition as $M\ln L/M\boldsymbol{\beta} = \sum_{i=1}^n \mathbf{x}_i(y_i - \lambda_i) = \mathbf{0}$
which is quite similar to the counterpart for the classical regression if we view $(y_i - \lambda_i) = (y_i - E[y_i*\mathbf{x}_i])$ as a residual. The second derivatives matrix is $M\ln L/M\boldsymbol{\beta}M\boldsymbol{\beta}N = -\sum_{i=1}^n \left(e^{\boldsymbol{\beta}'\mathbf{x}_i}\right)\mathbf{x}_i\mathbf{x}_i' = -\sum_{i=1}^n \lambda_i\mathbf{x}_i\mathbf{x}_i'$. This is a negative definite matrix. To prove this, note, first, that λ_i must always be positive. Then, let $\boldsymbol{\Omega}$ be a diagonal matrix whose ith diagonal element is $\sqrt{\lambda_i}$ and let $\mathbf{Z} = \boldsymbol{\Omega}\mathbf{X}$. Then, $M\ln L/M\boldsymbol{\beta}M\boldsymbol{\beta}N = -\mathbf{Z}N\mathbf{Z}$ which is clearly negative definite. This implies that the log-likelihood function is globally concave and finding its maximum using NewtonNs method will be straightforward and reliable.

The iteration for NewtonNs method is defined in (5-17). We may apply it directly in this problem. The computations involved in using Newton's method to maximize $\ln L$ will be as follows:

(1) Obtain starting values for the parameters. Because the log-likelihood function is globally concave, it will usually not matter what values are used. Most applications simply use zero. One suggestion which does appear in the literature is $\boldsymbol{\beta}^0 = \left[\sum_{i=1}^n q_i\mathbf{x}_i\mathbf{x}_i'\right]^{-1}\left[\sum_{i=1}^n q_i\mathbf{x}_i y_i\right]$ where $q_i = \log(\max(1, y_i))$.

(2) The iteration is computed as $\hat{\boldsymbol{\beta}}_{t+1} = \hat{\boldsymbol{\beta}}_t + \left[\sum_{i=1}^n \hat{\lambda}_i\mathbf{x}_i\mathbf{x}_i'\right]^{-1}\left[\sum_{i=1}^n \mathbf{x}_i(y_i - \hat{\lambda}_i)\right]$.

(3) Each time we compute $\hat{\beta}_{t+1}$, we should check for convergence. Some possibilities are

(a) Gradient: Are the elements of $\text{M}\ln L/\text{M}\beta$ small?

(b) Change: Is $\hat{\beta}_{t+1} - \hat{\beta}_t$ small?

(c) Function rate of change: Check the size of

$$\delta_t = \left[\sum_{i=1}^{n} \mathbf{x}_i(y_i - \hat{\lambda}_i)\right]' \left[\sum_{i=1}^{n} \hat{\lambda}_i \mathbf{x}_i \mathbf{x}_i'\right]^{-1} \left[\sum_{i=1}^{n} \mathbf{x}_i(y_i - \hat{\lambda}_i)\right]$$

before computing $\hat{\beta}_{t+1}$. This measure describes what will happen to the function at the next value of β. This is Belsley's criterion.

(4) When convergence has been achieved, the asymptotic covariance matrix for the estimates is estimated with the inverse matrix used in the iterations.

Using the data given in the problem, the results of the above computations are

Iter.	α	β	$\ln L$	$\text{M}\ln L/\text{M}\alpha$	$\text{M}\ln L/\text{M}\beta$	Change
0	0	0	-102.387	65.	95.1	296.261
1	1.37105	2.17816	-1442.38	-1636.25	-2788.5	1526.36
2	.619874	2.05865	-461.989	-581.966	-996.711	516.92
3	.210347	1.77914	-141.022	-195.953	-399.751	197.652
4	.351893	1.26291	-51.2989	-57.9294	-102.847	30.616
5	.824956	.698768	-33.5530	-12.8702	-23.1932	2.75855
6	1.05288	.453352	-32.0824	-1.28785	-2.29289	.032399
7	1.07777	.425239	-32.0660	-.016067	-.028454	.0000051
8	1.07808	.424890	-32.0660	0	0	0

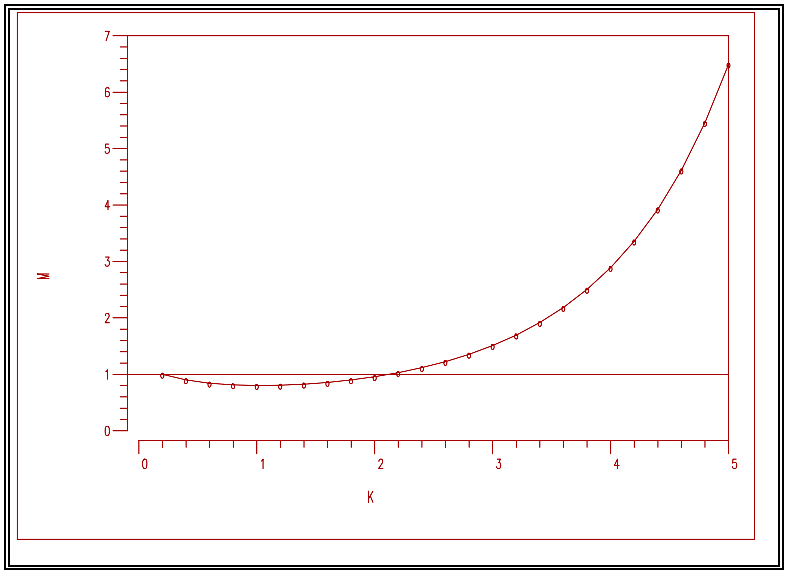
At the final values, the negative inverse of the second derivatives matrix is

$$\left[\sum_{i=1}^{n} \hat{\lambda}_i \mathbf{x}_i \mathbf{x}_i'\right]^{-1} = \begin{bmatrix} .151044 & -.095961 \\ -.095961 & .0664665 \end{bmatrix}.$$

4. Use Monte Carlo Integration to plot the function $g(r) = E[x^r * x>0]$ for the standard normal distribution. The expected value from the truncated normal distribution is

$$E[x^r|x>0] = \int_0^\infty x^r f(x|x>0)dx = \frac{\int_0^\infty x^r \phi(x)dx}{\int_0^\infty \phi(x)dx} = \frac{2}{\sqrt{\pi}}\int_0^\infty x^r e^{-\frac{x^2}{2}}dx.$$

To evaluate this expectation, we first sampled 1,000 observations from the truncated standard normal distribution using (5-1). For the standard normal distribution, $\mu = 0$, $\sigma = 1$, $P_L = \Phi((0 - 0)/1) = 2$, and $P_U = \Phi((+4 - 0)/1) = 1$. Therefore, the draws are obtained by transforming draws from U(0,1) (denoted F_i) to $x_i = \Phi[2(1 + F_i)]$. Since $0 < F_i < 1$, the argument in brackets must be greater than 2, so $x_i > 0$, which is to be expected. Using the same 1,000 draws each time (so as to obtain smoothness in the figure), we then plot the values of $\bar{x}_r = \frac{1}{1000}\sum_{i=1}^{1000} x_i^r$, r = 0, .2, .4,.6, ..., 5.0. As an additional experiment, we generated a second sample of 1,000 by drawing observations from the standard normal distribution and discarding them and redrawing if they were not positive. The means and standard deviations of the two samples were (0.8097,0.6170) for the first and (0.8059,0.6170) for the second. Drawing the second sample takes approximately twice as long as the second. Why?



5. For the model in Example 5.10, derive the LM statistic for the test of the hypothesis that μ=0.

The derivatives of the log-likelihood with μ = 0 imposed are $g_\mu = n\bar{x}/\sigma^2$ and $g_{\sigma^2} = \dfrac{-n}{2\sigma^2} + \dfrac{\sum_{i=1}^n x_i^2}{2\sigma^4}$. The estimator for σ^2 will be obtained by equating the second of these to 0, which will give (of course), $v = \mathbf{x}'\mathbf{x}/n$. The terms in the Hessian are $H_{\mu\mu} = -n/\sigma^2$, $H_{\mu\sigma^2} = -n\bar{x}/\sigma^4$, and $H_{\sigma^2\sigma^2} = n/(2\sigma^4)-\mathbf{x}'\mathbf{x}/\sigma^6$. At the MLE, $g_{\sigma^2} = 0$, exactly. The off diagonal term in the expected Hessian is also zero. Therefore, the LM statistic is $LM = \begin{bmatrix} n\bar{x}/v & 0 \end{bmatrix} \begin{bmatrix} \dfrac{n}{v} & 0 \\ 0 & \dfrac{n}{2v^2} \end{bmatrix}^{-1} \begin{bmatrix} n\bar{x}/v \\ 0 \end{bmatrix} = \left[\dfrac{\bar{x}}{v/\sqrt{n}}\right]^2.$

This resembles the square of the standard t-ratio for testing the hypothesis that μ = 0. It would be exactly that save for the absence of a degrees of freedom correction in v. However, since we have not estimated μ with $\bar{x}$ in fact, LM *is* exactly the square of a standard normal variate divided by a chi-squared variate over its degrees of freedom. Thus, in this model, LM is exactly an F statistic with 1 degree of freedom in the numerator and n degrees of freedom in the denominator.

6. In Example 5.10, what is the concentrated over μ log likelihood function?

It is obvious that whatever solution is obtained for σ^2, the MLE for μ will be $\bar{x}$, so the concentrated log-likelihood function is $\log L_c = \dfrac{-n}{2}\left(\log 2\pi + \log\sigma^2\right) - \dfrac{1}{2\sigma^2}\sum_{i=1}^n \left(x_i - \bar{x}\right)^2$

7. In Example 5.13, suppose that $E[y_i]$ = μ, for a nonzero mean.
(a) Extend the model to include this new parameter. What are the new log likelihood, likelihood equation, Hessian, and expected Hessian?
(b) How are the iterations carried out to estimate the full set of parameters?
(c) Show how the *LIMDEP* program should be modified to include estimation of μ.
(d) Using the same data set, estimate the full set of parameters.

If y_i has a nonzero mean, μ, then the log-likelihood is

$$\ln L(\gamma,\mu|\mathbf{Z}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}\log\sigma_i^2 - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{(y_i-\mu)^2}{\sigma_i^2}\right)$$

$$= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}\mathbf{z}_i'\gamma - \frac{1}{2}\sum_{i=1}^{n}(y_i-\mu)^2\exp(-\mathbf{z}_i'\gamma).$$

The likelihood equations are

$$\frac{\partial \ln L}{\partial \gamma} = \frac{1}{2}\sum_{i=1}^{n}\mathbf{z}_i\left(\frac{(y_i-\mu)^2}{\sigma_i^2}-1\right) = -\frac{1}{2}\sum_{i=1}^{n}\mathbf{z}_i + \frac{1}{2}\sum_{i=1}^{n}(y_i - \mu)^2\mathbf{z}_i\exp(-\mathbf{z}_i'\gamma)$$

$$= \mathbf{g}_\gamma(\gamma,\mu) = \mathbf{0}$$

and

$$\frac{\partial \ln L}{\partial \mu} = \sum_{i=1}^{n}(y_i-\mu)\exp(-\mathbf{z}_i'\gamma) = g_\mu(\gamma,\mu) = 0.$$

The Hessian is

$$\frac{\partial^2 \ln L}{\partial\gamma\partial\gamma'} = -\frac{1}{2}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{z}_i'\left(\frac{(y_i-\mu)^2}{\sigma_i^2}\right) = -\frac{1}{2}\sum_{i=1}^{n}(y_i - \mu)^2\mathbf{z}_i\mathbf{z}_i'\exp(-\mathbf{z}_i'\gamma) = \mathbf{H}_{\gamma\gamma}.$$

$$\frac{\partial^2 \ln L}{\partial\gamma\partial\mu} = -\sum\nolimits_{i=1}^{n}\mathbf{z}_i(y_i-\mu)\exp(-\mathbf{z}_i'\gamma) = \mathbf{H}_{\gamma\mu}$$

$$\frac{\partial^2 \ln L}{\partial\mu\partial\mu} = -\sum\nolimits_{i=1}^{n}\exp(-\mathbf{z}_i'\gamma) = \mathbf{H}_{\mu\mu}$$

The expectations in the Hessian are found as follows: Since $E[y_i] = \mu$, $E[\mathbf{H}_{\gamma\mu}] = \mathbf{0}$. There are no stochastic terms in $\mathbf{H}_{\mu\mu}$, so $E[\mathbf{H}_{\mu\mu}] = \mathbf{H}_{\mu\mu} = -\sum\nolimits_{i=1}^{n}\frac{1}{\sigma_i^2}$. Finally, $E[(y_i - \mu)^2] = \sigma_i^2$, so $E[\mathbf{H}_{\gamma\gamma}] = -1/2(\mathbf{Z}'\mathbf{Z})$.

There is more than one way to estimate the parameters. As in Example 5.13, the method of scoring (using the expected Hessian) will be straightforward in principle - though in our example, it does not work well in practice, so we use Newton's method instead. The iteration, in which we use index 't' to indicate the estimate at iteration t, will be

$$\begin{bmatrix}\mu\\ \gamma\end{bmatrix}_{(t+1)} = \begin{bmatrix}\mu\\ \gamma\end{bmatrix}_{(t)} - E[\mathbf{H}(t)]^{-1}\,\mathbf{g}(t).$$

If we insert the expected Hessians and first derivatives in this iteration, we obtain

$$\begin{bmatrix}\mu\\ \gamma\end{bmatrix}_{(t+1)} = \begin{bmatrix}\mu\\ \gamma\end{bmatrix}_{(t)} + \begin{bmatrix}\sum_{i=1}^{n}\frac{1}{\sigma_i^2(t)} & 0\\ 0 & \frac{1}{2}\mathbf{Z}'\mathbf{Z}\end{bmatrix}^{-1}\begin{bmatrix}\sum_{i=1}^{n}\frac{y_i-\mu(t)}{\sigma_i^2(t)}\\ \frac{1}{2}\sum_{i=1}^{n}\mathbf{z}_i\left(\frac{(y_i-\mu(t))^2}{\sigma_i^2(t)}-1\right)\end{bmatrix}.$$

The zero off diagonal elements in the expected Hessian make this convenient, as the iteration may be broken into two parts. We take the iteration for μ first. With current estimates $\mu(t)$ and $\gamma(t)$, the method of scoring produces this iteration: $\mu(t+1) = \mu(t) + \dfrac{\sum_{i=1}^{n}\frac{y_i-\mu(t)}{\sigma_i^2(t)}}{\sum_{i=1}^{n}\frac{1}{\sigma_i^2(t)}}$. As will be explored in Chapters 12 and 13, this is generalized least squares. Let $\mathbf{i}$ denote an $n\times 1$ vector of ones, let $e_i(t) = y_i - \mu(t)$ denote the 'residual' at iteration t and let $\mathbf{e}(t)$ denote the $n\times 1$ vector of residuals. Let $\mathbf{\Omega}(t)$ denote a diagonal matrix which has σ_i^2 on its diagonal (and zeros elsewhere). Then, the iteration for μ is
$\mu(t+1) = \mu(t) + [\mathbf{i}'\mathbf{\Omega}(t)^{-1}\mathbf{i}]^{-1}[\mathbf{i}'\mathbf{\Omega}(t)^{-1}\mathbf{e}(t)]$. This shows how to compute $\mu(t+1)$. The iteration for $\gamma(t+1)$ is exactly as was shown in Example 5.13, save for the single change that in the computation, y_i^2 is changed to $(y_i - \mu(t))^2$. Otherwise, the computation is identical. Thus, we would have
$\gamma(t+1) = \gamma(t) + (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{v}(\gamma(t),\mu(t))$, where $v_i(\gamma(t),\mu(t))$ is the term in parentheses in the iteration shown above. This shows how to compute $\gamma(t+1)$.

```
/*=================================================================
Program Code for Estimation of Harvey's Model
The data set for this model is 100 observations from Greene (1992)
Variables are: Y  = Average monthly credit card expenditure
               Q1 = Age in years+ 12ths of a year
               Q2 = Income, divided by 10,000
               Q3 = OwnRent; individual owns (1) or rents (0) home
               Q4 = Self employed (1=yes, 0=no)
Read     ; Nobs = 200 ; Nvar = 6 ; Names = y,q1,q2,q3,q4
         ; file=d:\DataSets\A5-1.dat$
Namelist ; Z = One,q1,q2,q3,q4 $
================================================================
Step 1 is to get the starting values and set some values for the
iterations- iter=iteration counter, delta=value for convergence.
*/
Create   ; y0 = y - Xbr(y) ; ui = log(y0^2) $
Matrix   ; gamma0 = <Z'Z> * Z'ui ; EH = 2*<Z'Z> $
Calc     ; c0 = gamma0(1)+1.2704     ? Correction to start value
         ; s20 = y0'y0/n ; delta = 1 ; iter=0 $
Create   ; vi0 = y0^2 / s20 - 1 $  (Used in LM statistic)
? Correct first element in gamma, then set starting vector.
Matrix   ; Gamma0(1) = c0  ; Gamma = Gamma0 $ Start value for gamma
Calc     ; mu0 = Xbr(y); mu = mu0$            Start value for mu
Procedure ----------[This does the iterations]--------------------
Create   ; vari = exp(Z'Gamma)  ;  ei = y-mu  ; varinv=1/vari
         ; hi   = ei^2 / vari
         ; gigamma = .5*(hi - 1); gimu = ei/vari
         ; logli = -.5*(log(2*pi) + log(vari) + hi) $
Matrix   ; ggamma = Z'gigamma ; gmu= 1'gimu
         ; H = 2*<Z'[hi]Z> ; gupdate = H*ggamma
? scoring, update = EH*ggamma
         ; Gamma = Gamma + gupdate $
Calc     ; muupdate = Sum(gimu)/Sum(varinv) ; mu = mu + muupdate $
Matrix   ; update = [gupdate/muupdate] ; g = [ggamma/gmu] $
Calc     ; list ; Iter = Iter+1 ; LogLU = Sum(logli);delta=g'update$
EndProcedure
Execute  ; While  delta > .00001 $ ------------------------------
Matrix   ; Stat (Gamma,H) $
Calc     ; list ; mu ; vmu = 1/Sum(varinv)   ; tmu = mu/Sqr(Vmu) $
Calc     ; list ; Sigmasq = Exp(Gamma(1))    ; K = Col(Z)
                ; SE = Sigmasq * Sqr(H(1,1)) ; TRSE = Sigmasq/SE
                ; LogLR = -n/2*(1 + log(2*pi)+ log(s20))
                ; LRTest = -2*(LogLR - LogLU) $
Matrix          ; Alpha = Gamma(2:K) ; VAlpha = Part(H,2,K,2,K)
         ; list ; WaldTest = Alpha ' <VAlpha> Alpha
                ; LMTest = .5* vi0'Z * <Z'Z> * Z'vi0
                ; EH ; H ; VB = BHHH(Z,gi) ; <VB> $
```

In the Example in the text, μ was constrained to equal $\bar{y}$. In the program, μ is allowed to be a free parameter. The comparison of the two sets of results appears below.

Iteration	(Constrained model, $\mu = \bar{y}$) log likelihood	δ	(Unconstrained model) log-l;ikelihood	δ
1	-698.3888	19.7022	-692.2987	22.8406
2	-692.2986	4.5494	-683.2320	6.9005
3	-689.7029	0.406881	-680.7028	2.7494
4	-689.4980	0.01148798	-679,7461	0.63453
5	-689.4741	0.0000125995	-679.4856	0.27023
6	-689.47407	0.000000000016	-679.4856	0.08124
			-679.4648	0.03079
			-679.4568	0.0101793
			-679.4542	0.00364255
			-679.4533	0.001240906
			-679.4530	0.00043431
			-679.4529	0.0001494193
			-679.4528	0.00005188501
			-679.4528	0.00001790973
			-679.4528	0.00000620193

Estimated Paramaters

Variable	**Estimate**	**Std Error**	**t-ratio**			
Age	0.013042	0.02310	0.565	-0.0134	0.0244	-0.550
Income	0.6432	0.120001	5.360	0.9953	0.1375	7.236
Ownrent	-0.2159	0.3073	-0.703	0.0774	0.3004	0.258
SelfEmployed	-0.4273	0.6677	-0.640	-1.3117	0.6719	-1.952
γ_1	8.465			7.867		
σ^2	4,745.92			2609.72		
μ	189.02	fixed		91.874	15.247	6.026

Tests of the joint hypothesis that all slope coefficients are zero:

LW	40.716	60.759
Wald:	39.024	69.515
LM	35.115	35.115 (same by construction).