Solutions and Applications Manual

Econometric Analysis

Sixth Edition

William H. Greene

New York University

Prentice Hall, Upper Saddle River, New Jersey 07458

Contents and Notation

This book presents solutions to the end of chapter exercises and applications in Econometric Analysis. There are no exercises in the text for Appendices A – E. For the instructor or student who is interested in exercises for this material, I have included a number of them, with solutions, in this book. The various computations in the solutions and exercises are done with the *NLOGIT* Version 4.0 computer package (Econometric Software, Inc., Plainview New York, www.nlogit.com). In order to control the length of this document, only the solutions and not the questions from the exercises and applications are shown here. In some cases, the numerical solutions for the in text examples shown here differ slightly from the values given in the text. This occurs because in general, the derivative computations in the text are done using the digits shown in the text, which are rounded to a few digits, while the results shown here are based on internal computations by the computer that use all digits.

In the solutions, we denote:

- scalar values with italic, lower case letters, as in a,
- column vectors with boldface lower case letters, as in $\mathbf{b}$,
- row vectors as transposed column vectors, as in $\mathbf{b}'$,
- matrices with boldface upper case letters, as in $\mathbf{M}$ or $\boldsymbol{\Sigma}$,
- single population parameters with Greek letters, as in θ,
- sample estimates of parameters with Roman letters, as in $\mathbf{b}$ as an estimate of β,
- sample estimates of population parameters with a caret, as in $\hat{\alpha}$ or $\hat{\beta}$,
- cross section observations with subscript i, as in y_i,
  time series observations with subscript t, as in z_t and
  panel data observations with x_{it} or $x_{i,t-1}$ when the comma is needed to remove ambiguity.
  Observations that are vectors are denoted likewise, for example, $\mathbf{x}_{it}$ to denote a column vector of observations.

These are consistent with the notation used in the text.

Chapter 1

Introduction

There are no exercises or applications in Chapter 1.

Chapter 2

The Classical Multiple Linear Regression Model

There are no exercises or applications in Chapter 2.

Chapter 3

Least Squares

Exercises

1. Let $\mathbf{X} = \begin{bmatrix} 1 & x_1 \\ ... & ... \\ 1 & x_n \end{bmatrix}$.

(a) The normal equations are given by (3-12), $\mathbf{X'e} = \mathbf{0}$ (we drop the minus sign), hence for each of the columns of **X**, $\mathbf{x_k}$, we know that $\mathbf{x_k}'$e = 0. This implies that $\Sigma_{i=1}^n e_i = 0$ and $\Sigma_{i=1}^n x_i e_i = 0$.

(b) Use $\Sigma_{i=1}^n e_i$ to conclude from the first normal equation that $a = \bar{y} - b\bar{x}$.

(c) We know that $\Sigma_{i=1}^n e_i = 0$ and $\Sigma_{i=1}^n x_i e_i = 0$. It follows then that $\Sigma_{i=1}^n (x_i - \bar{x})e_i = 0$ because $\Sigma_{i=1}^n \bar{x}e_i = \bar{x}\Sigma_{i=1}^n e_i = 0$. Substitute e_i to obtain

$\Sigma_{i=1}^n (x_i - \bar{x})(y_i - a - bx_i) = 0$ or $\Sigma_{i=1}^n (x_i - \bar{x})(y_i - \bar{y} - b(x_i - \bar{x})) = 0$

Then, $\Sigma_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = b\Sigma_{i=1}^n (x_i - \bar{x})(x_i - \bar{x}))$ so $b = \dfrac{\Sigma_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\Sigma_{i=1}^n (x_i - \bar{x})^2}$.

(d) The first derivative vector of e′e is -2**X′e**. (The normal equations.) The second derivative matrix is ∂^2(**e′e**)/∂**b**∂**b′** = 2**X′X**. We need to show that this matrix is positive definite. The diagonal elements are 2*n* and $2\Sigma_{i=1}^n x_i^2$ which are clearly both positive. The determinant is (2n)($2\Sigma_{i=1}^n x_i^2$)-($2\Sigma_{i=1}^n x_i$)^2
= $4n\Sigma_{i=1}^n x_i^2$ -4($n\bar{x}$)^2 = $4n[(\Sigma_{i=1}^n x_i^2) - n\bar{x}^2] = 4n[(\Sigma_{i=1}^n (x_i - \bar{x})^2]$. Note that a much simpler proof appears after (3-6).

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

(**y** - **Xc**)′(**y** - **Xc**) = [**y** - **X**(**b** + (**c** - **b**))] ′[**y** - **X**(**b** + (**c** - **b**))] = [(**y** - **Xb**) + **X**(**c** - **b**)] ′[(**y** - **Xb**) + **X**(**c** - **b**)]
= (**y** - **Xb**) ′(**y** - **Xb**) + (**c** - **b**) ′**X′X**(**c** - **b**) + 2(**c** - **b**) ′**X′**(**y** - **Xb**).

But, the third term is zero, as 2(**c** - **b**) ′**X′**(**y** - **Xb**) = 2(**c** - **b**)**X′e** = **0**. Therefore,

(**y** - **Xc**) ′(**y** - **Xc**) = **e′e** + (**c** - **b**) ′**X′X**(**c** - **b**)

or (**y** - **Xc**) ′(**y** - **Xc**) - **e′e** = (**c** - **b**) ′**X′X**(**c** - **b**).

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

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

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

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

4. In the regression of **y** on **i** and **X**, the coefficients on **X** are **b** = $(\mathbf{X'M^0X})^{-1}\mathbf{X'M}^0\mathbf{y}$. $\mathbf{M}^0$ = **I** - $\mathbf{i}(\mathbf{i'i})^{-1}\mathbf{i'}$ 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 $[(\mathbf{X'M}^{0\prime})(\mathbf{M}^0\mathbf{X})]^{-1}(\mathbf{X'M}^{0\prime})(\mathbf{M}^0\mathbf{y})$ 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.

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

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

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

6. The original X matrix has n rows. We add an additional row, x_s'. The new y vector likewise has an additional element. Thus, $\mathbf{X}_{n,s} = \begin{bmatrix}\mathbf{X}_n \\ \mathbf{x}_s'\end{bmatrix}$ and $\mathbf{y}_{n,s} = \begin{bmatrix}\mathbf{y}_n \\ y_s\end{bmatrix}$. The new coefficient vector is

$\mathbf{b}_{n,s} = (\mathbf{X}_{n,s}'\mathbf{X}_{n,s})^{-1}(\mathbf{X}_{n,s}'\mathbf{y}_{n,s})$. The matrix is $\mathbf{X}_{n,s}'\mathbf{X}_{n,s} = \mathbf{X}_n'\mathbf{X}_n + \mathbf{x}_s\mathbf{x}_s'$. To invert this, use (A -66);

$(\mathbf{X}_{n,s}'\mathbf{X}_{n,s})^{-1} = (\mathbf{X}_n'\mathbf{X}_n)^{-1} - \dfrac{1}{1+\mathbf{x}_s'(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{x}_s}(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{x}_s\mathbf{x}_s'(\mathbf{X}_n'\mathbf{X}_n)^{-1}$. The vector is

$(\mathbf{X}_{n,s}'\mathbf{y}_{n,s}) = (\mathbf{X}_n'\mathbf{y}_n) + \mathbf{x}_s y_s$. Multiply out the four terms to get

$(\mathbf{X}_{n,s}'\mathbf{X}_{n,s})^{-1}(\mathbf{X}_{n,s}'\mathbf{y}_{n,s}) =$

$$\mathbf{b}_n - \frac{1}{1+\mathbf{x}_s'(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{x}_s}(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{x}_s\mathbf{x}_s'\mathbf{b}_n + (\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{x}_s y_s - \frac{1}{1+\mathbf{x}_s'(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{x}_s}(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{x}_s\mathbf{x}_s'(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{x}_s y_s$$

=

$$\mathbf{b}_n + (\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{x}_s y_s - \frac{\mathbf{x}_s'(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{x}_s}{1+\mathbf{x}_s'(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{x}_s}(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{x}_s y_s - \frac{1}{1+\mathbf{x}_s'(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{x}_s}(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{x}_s\mathbf{x}_s'\mathbf{b}_n$$

$$\mathbf{b}_n + \left[1 - \frac{\mathbf{x}_s'(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{x}_s}{1+\mathbf{x}_s'(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{x}_s}\right](\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{x}_s y_s - \frac{1}{1+\mathbf{x}_s'(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{x}_s}(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{x}_s\mathbf{x}_s'\mathbf{b}_n$$

$$\mathbf{b}_n + \frac{1}{1+\mathbf{x}_s'(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{x}_s}(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{x}_s y_s - \frac{1}{1+\mathbf{x}_s'(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{x}_s}(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{x}_s\mathbf{x}_s'\mathbf{b}_n$$

$$\mathbf{b}_n + \frac{1}{1+\mathbf{x}_s'(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{x}_s}(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{x}_s(y_s - \mathbf{x}_s'\mathbf{b}_n)$$

7. Define the data matrix as follows: $\mathbf{X} = \begin{bmatrix}\mathbf{i} & \mathbf{x} & \mathbf{0} \\ 1 & 0 & 1\end{bmatrix} = \begin{bmatrix}\mathbf{X}_1, & \mathbf{0} \\ & 1\end{bmatrix} = \begin{bmatrix}\mathbf{X}_1 & \mathbf{X}_2\end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix}\mathbf{y}_o \\ y_m\end{bmatrix}$. (The subscripts on the parts of $\mathbf{y}$ refer to the "observed" and "missing" rows of $\mathbf{X}$. We will use Frish-Waugh to obtain the first two columns of the least squares coefficient vector. $\mathbf{b}_1 = (\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1)^{-1}(\mathbf{X}_1'\mathbf{M}_2\mathbf{y})$. Multiplying it out, we find that $\mathbf{M}_2$ = an identity matrix save for the last diagonal element that is equal to 0.

$\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1 = \mathbf{X}_1'\mathbf{X}_1 - \mathbf{X}_1'\begin{bmatrix}\mathbf{0} & \mathbf{0} \\ \mathbf{0}' & 1\end{bmatrix}\mathbf{X}_1$. This just drops the last observation. $\mathbf{X}_1'\mathbf{M}_2\mathbf{y}$ is computed likewise. Thus, the coeffients on the first two columns are the same as if y_0 had been linearly regressed on $\mathbf{X}_1$. The denomonator of R^2 is different for the two cases (drop the observation or keep it with zero fill and the dummy variable). For the first strategy, the mean of the n-1 observations should be different from the mean of the full n unless the last observation happens to equal the mean of the first n-1.

For the second strategy, replacing the missing value with the mean of the other n-1 observations, we can deduce the new slope vector logically. Using Frisch-Waugh, we can replace the column of x's with deviations from the means, which then turns the last observation to zero. Thus, once again, the coefficient on the x equals what it is using the earlier strategy. The constant term will be the same as well.

8. 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

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

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

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

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

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

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

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

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

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

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

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

From the regressions, we have

	Cov[C,Y]/Var[Y] = Cov[C,Y] = .8.
But,	Cov[C,Y] = Cov[N,Y] + Cov[D,Y].
Also,	Cov[C,N]/Var[N] = Cov[C,N] = .5,
but,	Cov[C,N] = Var[N] + Cov[N,D] = 1 + Cov[N,D], so Cov[N,D] = -.5,
so that	Var[C] = 2(1 + -.5) = 1.
And,	Cov[D,Y]/Var[Y] = Cov[D,Y] = .4.
Since	Cov[C,Y] = .8 = Cov[N,Y] + Cov[D,Y], Cov[N,Y] = .4.
Finally,	Cov[C,D] = Cov[N,D] + Var[D] = -.5 + 1 = .5.

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

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

12. The relevant submatrices to be used in the calculations are

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

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

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

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

13. 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. Further discussion about the data themselves appeared in subsequent idscussion. [See Goldberger (1973) and Leff (1973).]

14. A proof of Theorem 3.1 provides a general statement of the observation made after (3-8). The counterpart for a multiple regression to the normal equations preceding (3-7) is

$$\begin{aligned}
&b_1 n + b_2\Sigma_i x_{i2} + b_3\Sigma_i x_{i3} + \ldots + b_K\Sigma_i x_{iK} = \Sigma_i y_i \\
&b_1\Sigma_i x_{i2} + b_2\Sigma_i x_{i2}^2 + b_3\Sigma_i x_{i2}x_{i3} + \ldots + b_K\Sigma_i x_{i2}x_{iK} = \Sigma_i x_{i2}y_i \\
&\ldots \\
&b_1\Sigma_i x_{iK} + b_2\Sigma_i x_{iK}x_{i2} + b_3\Sigma_i x_{iK}x_{i3} + \ldots + b_K\Sigma_i x_{iK}^2 = \Sigma_i x_{iK}y_i.
\end{aligned}$$

As before, divide the first equation by n, and manipulate to obtain the solution for the constant term, $b_1 = \bar{y} - b_2\bar{x}_2 - \ldots - b_K\bar{x}_K$. Substitute this into the equations above, and rearrange once again to obtain the equations for the slopes,

$$\begin{aligned}
&b_2\Sigma_i(x_{i2}-\bar{x}_2)^2 + b_3\Sigma_i(x_{i2}-\bar{x}_2)(x_{i3}-\bar{x}_3) + \ldots + b_K\Sigma_i(x_{i2}-\bar{x}_2)(x_{iK}-\bar{x}_K) = \Sigma_i(x_{i2}-\bar{x}_2)(y_i-\bar{y}) \\
&b_2\Sigma_i(x_{i3}-\bar{x}_3)(x_{i2}-\bar{x}_2) + b_3\Sigma_i(x_{i3}-\bar{x}_3)^2 + \ldots + b_K\Sigma_i(x_{i3}-\bar{x}_3)(x_{iK}-\bar{x}_K) = \Sigma_i(x_{i3}-\bar{x}_3)(y_i-\bar{y}) \\
&\ldots \\
&b_2\Sigma_i(x_{iK}-\bar{x}_K)(x_{i2}-\bar{x}_2) + b_3\Sigma_i(x_{iK}-\bar{x}_K)(x_{i3}-\bar{x}_3) + \ldots + b_K\Sigma_i(x_{iK}-\bar{x}_K)^2 = \Sigma_i(x_{iK}-\bar{x}_K)(y_i-\bar{y}).
\end{aligned}$$

If the variables are uncorrelated, then all cross product terms of the form $\Sigma_i(x_{ij}-\bar{x}_j)(x_{ik}-\bar{x}_k)$ will equal zero. This leaves the solution,

$$\begin{aligned}
&b_2\Sigma_i(x_{i2}-\bar{x}_2)^2 = \Sigma_i(x_{i2}-\bar{x}_2)(y_i-\bar{y}) \\
&b_3\Sigma_i(x_{i3}-\bar{x}_3)^2 = \Sigma_i(x_{i3}-\bar{x}_3)(y_i-\bar{y}) \\
&\ldots \\
&b_K\Sigma_i(x_{iK}-\bar{x}_K)^2 = \Sigma_i(x_{iK}-\bar{x}_K)(y_i-\bar{y}),
\end{aligned}$$

which can be solved one equation at a time for

$$b_k = \left[\Sigma_i(x_{ik}-\bar{x}_k)(y_i-\bar{y})\right]\Big/\left[\Sigma_i(x_{ik}-\bar{x}_k)^2\right],\ k = 2,\ldots,K.$$

Each of these is the slope coefficient in the simple of y on the respective variable.

Application

```
?=======================================================================
? Chapter 3 Application 1
?=======================================================================
Read $
(Data appear in the text.)
Namelist ; X1 = one,educ,exp,ability$
Namelist ; X2 = mothered,fathered,sibs$
?=======================================================================
? a.
?=======================================================================
Regress  ; Lhs = wage ; Rhs = x1$
+----------------------------------------------------+
| Ordinary    least squares regression               |
| LHS=WAGE     Mean                 =   2.059333     |
|              Standard deviation   =   .2583869     |
| WTS=none     Number of observs.   =         15     |
| Model size   Parameters           =          4     |
|              Degrees of freedom   =         11     |
| Residuals    Sum of squares       =   .7633163     |
|              Standard error of e  =   .2634244     |
| Fit          R-squared            =   .1833511     |
|              Adjusted R-squared   =  -.3937136E-01 |
| Model test   F[  3,    11] (prob) =    .82 (.5080) |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|    1.66364000       .61855318     2.690   .0210
 EDUC    |     .01453897       .04902149      .297   .7723   12.8666667
 EXP     |     .07103002       .04803415     1.479   .1673   2.80000000
 ABILITY |     .02661537       .09911731      .269   .7933    .36600000
?=======================================================================
? b.
?=======================================================================
Regress  ; Lhs = wage ; Rhs = x1,x2$
+----------------------------------------------------+
| Ordinary    least squares regression               |
| LHS=WAGE     Mean                 =   2.059333     |
|              Standard deviation   =   .2583869     |
| WTS=none     Number of observs.   =         15     |
| Model size   Parameters           =          7     |
|              Degrees of freedom   =          8     |
| Residuals    Sum of squares       =   .4522662     |
|              Standard error of e  =   .2377673     |
| Fit          R-squared            =   .5161341     |
|              Adjusted R-squared   =   .1532347     |
| Model test   F[  6,     8] (prob) =   1.42 (.3140) |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|     .04899633       .94880761      .052   .9601
 EDUC    |     .02582213       .04468592      .578   .5793   12.8666667
 EXP     |     .10339125       .04734541     2.184   .0605   2.80000000
 ABILITY |     .03074355       .12120133      .254   .8062    .36600000
 MOTHERED|     .10163069       .07017502     1.448   .1856   12.0666667
 FATHERED|     .00164437       .04464910      .037   .9715   12.6666667
 SIBS    |     .05916922       .06901801      .857   .4162   2.20000000
?=======================================================================
? c.
?=======================================================================
```

```
Regress  ; Lhs = mothered ; Rhs = x1 ; Res = meds $
Regress  ; Lhs = fathered ; Rhs = x1 ; Res = feds $
Regress  ; Lhs = sibs     ; Rhs = x1 ; Res = sibss $
Namelist ; X2S = meds,feds,sibss $
Matrix   ; list ; Mean(X2S) $
Matrix Result   has  3 rows and  1 columns.
               1
        +--------------
       1| -.1184238D-14
       2|  .1657933D-14
       3| -.5921189D-16
```
The means are (essentially) zero. The sums must be zero, as these new variables are orthogonal to the columns of X1. The first column in X1 is a column of ones, so this means that these residuals must sum to zero.

```
?=======================================================================
? d.
?=======================================================================
Namelist ; X = X1,X2 $
Matrix   ; i = init(n,1,1) $
Matrix   ; M0 = iden(n) - 1/n*i*i' $
Matrix   ; b12 = <X'X>*X'wage$
Calc     ; list ; ym0y =(N-1)*var(wage) $
Matrix   ; list ; cod = 1/ym0y * b12'*X'*M0*X*b12 $
Matrix COD      has  1 rows and  1 columns.
               1
        +--------------
       1|     .51613
Matrix   ; e = wage - X*b12 $
Calc     ; list ; cod = 1 - 1/ym0y * e'e $
+------------------------------------+
 COD     =       .516134
```
The R squared is the same using either method of computation.

```
Calc     ; list ; RsqAd = 1 - (n-1)/(n-col(x))*(1-cod)$
+------------------------------------+
 RSQAD   =       .153235
? Now drop the constant
Namelist ; X0 = educ,exp,ability,X2 $
Matrix   ; i = init(n,1,1) $
Matrix   ; M0 = iden(n) - 1/n*i*i' $
Matrix   ; b120 = <X0'X0>*X0'wage$
Matrix   ; list ; cod = 1/ym0y * b120'*X0'*M0*X0*b120 $
Matrix COD      has  1 rows and  1 columns.
               1
        +--------------
       1|     .52953
Matrix   ; e0 = wage - X0*b120 $
Calc     ; list ; cod = 1 - 1/ym0y * e0'e0 $
+------------------------------------+
| Listed Calculator Results          |
+------------------------------------+
 COD     =       .515973
```
The R squared now changes depending on how it is computed. It also goes up, completely artificially.

```
?=======================================================================
? e.
?=======================================================================
```
The R squared for the full regression appears immediately below.

```
? f.
Regress ; Lhs = wage ; Rhs = X1,X2 $
+----------------------------------------------------+
| Ordinary    least squares regression               |
| WTS=none     Number of observs.   =        15      |
| Model size   Parameters           =         7      |
|              Degrees of freedom   =         8      |
| Fit          R-squared            =   .5161341     |
+----------------------------------------------------+
+--------+-------------+---------------+--------+--------+----------+
```

```
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|     .04899633       .94880761      .052   .9601
 EDUC    |     .02582213       .04468592      .578   .5793   12.8666667
 EXP     |     .10339125       .04734541     2.184   .0605   2.80000000
 ABILITY |     .03074355       .12120133      .254   .8062    .36600000
 MOTHERED|     .10163069       .07017502     1.448   .1856   12.0666667
 FATHERED|     .00164437       .04464910      .037   .9715   12.6666667
 SIBS    |     .05916922       .06901801      .857   .4162   2.20000000
Regress ; Lhs = wage ; Rhs = X1,X2S $
+-----------------------------------------------------+
| Ordinary    least squares regression                 |
| WTS=none     Number of observs.   =         15      |
| Model size   Parameters           =          7      |
|              Degrees of freedom   =          8      |
| Fit          R-squared            =   .5161341      |
|              Adjusted R-squared   =   .1532347      |
+-----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|    1.66364000       .55830716     2.980   .0176
 EDUC    |     .01453897       .04424689      .329   .7509   12.8666667
 EXP     |     .07103002       .04335571     1.638   .1400   2.80000000
 ABILITY |     .02661537       .08946345      .297   .7737    .36600000
 MEDS    |     .10163069       .07017502     1.448   .1856 -.118424D-14
 FEDS    |     .00164437       .04464910      .037   .9715  .165793D-14
 SIBSS   |     .05916922       .06901801      .857   .4162 -.592119D-16
```

In the first set of results, the first coefficient vector is
$\mathbf{b}_1 = (\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{M}_2\mathbf{y}$ and
$b_2 = (\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2)^{-1}\mathbf{X}_2'M_1\mathbf{y}$
In the second regression, the second set of regressors is M_1X_2, so
$\mathbf{b}_1 = (\mathbf{X}_1'\mathbf{M}_{12}\,\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{M}_{12}\mathbf{y}$ where $\mathbf{M}_{12} = \mathbf{I} - (\mathbf{M}_1\mathbf{X}_2)[(M_1\mathbf{X}_2)'(\mathbf{M}_1\mathbf{X}_2)]^{-1}(\mathbf{M}_1\mathbf{X}_2)'$
Thus, because the "M" matrix is different, the coefficient vector is different. The second set of coefficients in the second regression is
$\mathbf{b}_2 = [(\mathbf{M}_1\mathbf{X}_2)'\mathbf{M}_1(\mathbf{M}_1\mathbf{X}_2)]^{-1}\,(\mathbf{M}_1\mathbf{X}_2)\mathbf{M}_1\mathbf{y} = (\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{M}_1\mathbf{y}$ because $\mathbf{M}_1$ is idempotent.

Chapter 4

Statistical Properties of the Least Squares Estimator

Exercises

1. 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 $\text{Min}_c L_* = c_1^2 v_1 + (1 - c_1)^2 v_2$. The necessary condition is $\partial L_*/\partial c_1 = 2c_1 v_1 - 2(1 - c_1)v_2 = 0$ which implies $c_1 = v_2 / (v_1 + v_2)$. A more intuitively appealing form is obtained by dividing numerator and denominator by $v_1 v_2$ to obtain $c_1 = (1/v_1) / [1/v_1 + 1/v_2]$. Thus, the weight is proportional to the inverse of the variance. The estimator with the smaller variance gets the larger weight.

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

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

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

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

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

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

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

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

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

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

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

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 $\mathbf{x}'\mathbf{x}$ grows, in which case both estimators are consistent.

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

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

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

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

But, $\Sigma_{i=1}^{n}(x_i - \bar{x})^2 = \mathbf{x'x} + n\bar{x}^2$

so the ratio is $\text{Var}[b]/\text{Var}[b'] = [\mathbf{x'x} + n\bar{x}^2]/\mathbf{x'x} = 1 - n\bar{x}^2/\mathbf{x'x} = 1 - \{n\bar{x}^2/[S_{xx} + n\bar{x}^2]\} \leq 1$

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

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

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

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

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 $S_{xx}\,\Sigma_{i=1}^{n}(x_i - \bar{x})^2$. The solution is, therefore, $\begin{pmatrix}\lambda_1 \\ \lambda_2\end{pmatrix} = \dfrac{1}{nS_{xx}}\begin{bmatrix} \Sigma_i x_i^2 & -n\bar{x} \\ -n\bar{x} & 0 \end{bmatrix}\begin{bmatrix} -2\sigma^2 \\ 0 \end{bmatrix}$

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

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

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

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 $\Sigma_i d_i y_i = (1/n)\Sigma_i y_i - \bar{x}\,\Sigma_{i=1}^{n}(x_i - \bar{x})y_i/S_{xx} = \bar{y} - b\bar{x}$ which was to be shown.

6. Let $q = E[Q]$. Then, $q = \alpha + \beta P$, or $P = (-\alpha/\beta) + (1/\beta)q$.

Using a well known result, for a linear demand curve, marginal revenue is $MR = (-\alpha/\beta) + (2/\beta)q$. 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. a. The sample means are (1/100) times the elements in the first column of $\mathbf{X'X}$. The sample covariance matrix for the three regressors is obtained as $(1/99)[(\mathbf{X}'\mathbf{X})_{ij} - 100\,\bar{x}_i\bar{x}_j]$.

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

$$\begin{bmatrix} 1 & .07971 & .78043 \\ .07971 & 1 & .68167 \\ .78043 & .68167 & 1 \end{bmatrix}$$

b. The vector of slopes is $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ = [-.4022, 6.123, 5.910, -7.525]′.

c. 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]′

d. 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: $[(1/99(.496969))/ 4.292]^2 = .068$.

e. 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. We consider two regressions. In the first, $\mathbf{y}$ is regressed on K variables, $\mathbf{X}$. The variance of the least squares estimator, $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, Var[$\mathbf{b}$] $= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$. In the second, $\mathbf{y}$ is regressed on $\mathbf{X}$ and an additional variable, $\mathbf{z}$. Using results for the partitioned regression, the coefficients on $\mathbf{X}$ when $\mathbf{y}$ is regressed on $\mathbf{X}$ and $\mathbf{z}$ 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 Var[$\mathbf{b}$,c] $= s^2\begin{bmatrix} \mathbf{X'X} & \mathbf{X'z} \\ \mathbf{z'X} & \mathbf{z'X} \end{bmatrix}^{-1}$. But, we have already found this above. The submatrix is Var[$\mathbf{b}_{.z}$] = $s^2(\mathbf{X}'\mathbf{M}_z\mathbf{X})^{-1}$. We can show that the second matrix is larger than the first by showing that its inverse is smaller. (See (A-120).) 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 Var[$\mathbf{b}$]$^{-1}$ is larger than Var[$\mathbf{b}_{.z}$]$^{-1}$, which implies that Var[$\mathbf{b}$] is smaller.

Although the true variance of $\mathbf{b}$ 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 $\mathbf{z}$ 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 $\mathbf{z}$ is larger than one in absolute value, then s_z^2 will be smaller than s^2. Thus, in the comparison, Est.Var[$\mathbf{b}$] $= s^2(\mathbf{X}'\mathbf{X})^{-1}$ is based on a smaller matrix, but a larger scale factor than 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 $\mathbf{z}$ and $\mathbf{X}$ were uncorrelated. Then, $\mathbf{XNM}_z\mathbf{X}$ equals $\mathbf{XNX}$. Then, the estimated variance of $\mathbf{b}_{.z}$ would be less than that of $\mathbf{b}$ without $\mathbf{z}$ 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. The F ratio is computed as $[\mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b}/K]/[\mathbf{e}'\mathbf{e}/(n - K)]$. We substitute $\mathbf{e} = \mathbf{M}\boldsymbol{\varepsilon}$, and

$\mathbf{b} = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon$. Then, $F = [\varepsilon'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon/K]/[\varepsilon'\mathbf{M}\varepsilon/(n - K)] =$ $[\varepsilon'(\mathbf{I} - \mathbf{M})\varepsilon/K]/[\varepsilon'\mathbf{M}\varepsilon/(n - K)]$.

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

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

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

The trace of the inverse equals the sum of the characteristic roots of the inverse, which are the reciprocals of the characteristic roots of $\mathbf{X}'\mathbf{X}$.

11. 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 $\mathbf{b} = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon$. Then, $F = [\varepsilon'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon/K]/[\varepsilon'\mathbf{M}\varepsilon/(n - K)] =$ $[\varepsilon'(\mathbf{I} - \mathbf{M})\varepsilon/K]/[\varepsilon'\mathbf{M}\varepsilon/(n - K)]$. The 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.

12. We can write e_i as $e_i = y_i - \mathbf{b}'\mathbf{x}_i = (\beta'\mathbf{x}_i + \varepsilon_i) - \mathbf{b}'\mathbf{x}_i = \varepsilon_i + (\mathbf{b} - \beta)'\mathbf{x}_i$
We know that plim $\mathbf{b} = \beta$, and $\mathbf{x}_i$ is unchanged as n increases, so as $n\rightarrow\infty$, e_i is arbitrarily close to ε_i.

13. 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 $\mathrm{Var}[\bar{y}] = (1/n^2)\Sigma_i\Sigma_j \mathrm{Cov}[\varepsilon_i,\varepsilon_j] = \sigma^2/n$. Since the mean equals μ and the variance vanishes as $n\rightarrow\infty$, $\bar{y}$ is mean square 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 $\mathrm{Var}[\hat{\mu}] = \Sigma_i w_i^2\sigma^2 = \sigma^2\Sigma_i w_i^2$. The sum of squares of the weights is $\Sigma_i w_i^2 = \Sigma_i i^2/[\Sigma_i i]^2 = [n(n+1)(2n+1)/6]/[n(n+1)/2]^2 =$ $[2(n^2 + 3n/2 + 1/2)]/[1.5n(n^2 + 2n + 1)]$. As $n\rightarrow\infty$, the fraction will be dominated by the term $(1/n)$ and will tend to zero. This establishes the consistency of this estimator. The last expression also provides the asymptotic variance. The large sample variance can be found as $\mathrm{Asy.Var}[\hat{\mu}] = (1/n)\lim_{n\rightarrow\infty}\mathrm{Var}[\sqrt{n}(\hat{\mu} - \mu)]$. For the estimator above, we can use $\mathrm{Asy.Var}[\hat{\mu}] = (1/n)\lim_{n\rightarrow\infty} n\mathrm{Var}[\hat{\mu} - \mu] = (1/n)\lim_{n\rightarrow\infty}\sigma^2[2(n^2 +$

$3n/2 + 1/2)]/[1.5(n^2 + 2n + 1)] = 1.3333\sigma^2$. Notice that this is unambiguously larger than the variance of the sample mean, which is the ordinary least squares estimator.

14. To obtain the asymptotic distribution, write the result already in hand as $\mathbf{b} = (\beta + \mathbf{Q}^{-1}\gamma) + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon} - \mathbf{Q}^{-1}\boldsymbol{\varepsilon}$. 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 $\mathbf{b}$. Write the preceding in the form $\mathbf{b} - \theta = (\mathbf{X}'\mathbf{X}/n)^{-1}(\mathbf{X}'\varepsilon/n) - \mathbf{Q}^{-1}\gamma$. Since plim$(\mathbf{X}'\mathbf{X}/n) = \mathbf{Q}$, the large sample behavior of the right hand side is the same as that of 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 $\mathbf{Q}$ in our derivation. Then, we seek the asymptotic distribution of $\sqrt{n}\,(\mathbf{b} - \theta)$ which is the same as that of
$\sqrt{n}\,[\mathbf{Q}^{-1}\text{plim}(\mathbf{X}'\varepsilon/n) - \mathbf{Q}^{-1}\gamma] = \mathbf{Q}^{-1}\sqrt{n}\ (1/n)\Sigma_{i=1}^{n}\left(\mathbf{x}_i\varepsilon_i\ -\ \gamma\right)$. 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.

15. a. To solve this, we will use an extension of Exercise 6 in Chapter 3 (adding one row of data), and the necessary matrix result, (A-66b) in which B will be $\mathbf{X}_m$ and $\mathbf{C}$ will be $\mathbf{I}$. Bypassing the matrix algebra, which will be essentially identical to the earlier exercise, we have

$$\mathbf{b}_{c,m} = \mathbf{b}_c + [\mathbf{I} + \mathbf{X}_m(\mathbf{X}_c'\mathbf{X}_c)^{-1}\mathbf{X}_m]^{-1}(\mathbf{X}_c'\mathbf{X}_c)^{-1}\mathbf{X}_m'(\mathbf{y}_m - \mathbf{X}_m\mathbf{b}_c)$$

But, in this case, $\mathbf{y}_m$ is precisely $\mathbf{X}_m\mathbf{b}_c$, so the ending vector is zero. Thus, the coefficient vector is the same. b. The model applies to the first n_c observations, so $\mathbf{b}_c$ is the least squares estimator for those observations. Yes, it is unbiased.
c. The residuals at the second step are $\mathbf{e}_c$ and $(\mathbf{X}_m\mathbf{b}_c - \mathbf{X}_m\mathbf{b}_c) = (\mathbf{e}_c', \mathbf{0}')'$. Thus, the sum of squares is the same at both steps.
d. The numerator of s^2 is the same in both cases, however, for the second one, the degrees of freedom is larger. The first is unbiased, so the second one must be biased downward.

Applications

```
?=======================================================================
? Chapter 4 Application 1
?=======================================================================
Read $
Year   GasExp Pop    Gasp   Income PNC    PUC    PPT    PD     PN     PS
1953   7.4    159565 16.668 8883   47.2   26.7   16.8   37.7   29.7   19.4
...
2004   224.5  293951 123.901 27113 133.9  133.3  209.1  114.8  172.2  222.8

Sample ; 1 - 52 $
Create ; G = 1000000*gasexp/(gasp*pop)$
Create ; t = year - 1952 $
Namelist ; X = one,income, gasp,pnc,puc,ppt,pd,pn,ps,t$
?=======================================================================
? a.  Basic regression
?=======================================================================
Regress ; Lhs = g ; Rhs = X $
+----------------------------------------------------+
| Ordinary    least squares regression               |
| LHS=G        Mean                  =   4.935619     |
|              Standard deviation    =   1.059105     |
| WTS=none     Number of observs.    =         52     |
| Model size   Parameters            =         10     |
|              Degrees of freedom    =         42     |
| Residuals    Sum of squares        =   .4985489     |
|              Standard error of e   =   .1089505     |
| Fit          R-squared             =   .9912852     |
|              Adjusted R-squared    =   .9894177     |
| Model test   F[  9,    42] (prob) = 530.82 (.0000) |
+----------------------------------------------------+
```

```
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|    1.10587817       .56937860     1.942   .0588
 INCOME  |     .00021575     .517619D-04     4.168   .0001   16805.0577
 GASP    |    -.01108386       .00397812    -2.786   .0080   51.3429615
 PNC     |     .00057735       .01284414      .045   .9644   87.5673077
 PUC     |    -.00587463       .00487032    -1.206   .2345   77.8000000
 PPT     |     .00690726       .00483613     1.428   .1606   89.3903846
 PD      |     .00122888       .01188175      .103   .9181   78.2692308
 PN      |     .01269051       .01259799     1.007   .3195   83.5980769
 PS      |    -.02802781       .00799625    -3.505   .0011   89.7769231
 T       |     .07250369       .01418280     5.112   .0000   26.5000000
?======================================================================
? b.  Hypothesis that b(NC) = b(UC) $
?======================================================================
Calc ; list ; (b(4)-b(5))/sqr(varb(4,4)+varb(5,5)-2*varb(4,5)) $
+------------------------------------+
| Listed Calculator Results          |
+------------------------------------+
 Result  =       .494883
?======================================================================
? c.  Elasticities.  In each case, elasticity = b*xbar/ybar
?======================================================================
Calc ; g2004   = g(52)$
Calc ; i2004   = income(52)$
Calc ; pg2004  = gasp(52)$
Calc ; ppt2004 = ppt(52)$
Calc ; list ; ei = b(2)*i2004/g2004
            ; ep = b(3)*pg2004/g2004
            ; eppt = b(6)*ppt2004/g2004$
+------------------------------------+
| Listed Calculator Results          |
+------------------------------------+
 EI      =       .948988
 EP      =      -.222792
 EPPT    =       .234311
?======================================================================
? d. Log regression
?======================================================================
Create ; logg = log(g) ; logpg = log(gasp) ; logi = log(income)
       ; logpnc=log(pnc) ; logpuc = log(puc) ; logppt = log(ppt)
       ; logpd = log(pd) ; logpn = log(pn) ; logps = log(ps) $
Namelist ; LogX = one,logi,logpg,logpnc,logpuc,logppt,logpd,logpn,logps,t$
Regress ; lhs = logg ; rhs = logx $
+----------------------------------------------------+
| Ordinary    least squares regression               |
| LHS=LOGG     Mean                 =   1.570475     |
|              Standard deviation   =   .2388115     |
| WTS=none     Number of observs.   =         52     |
| Model size   Parameters           =         10     |
|              Degrees of freedom   =         42     |
| Residuals    Sum of squares       =   .3812817E-01 |
|              Standard error of e  =   .3012994E-01 |
| Fit          R-squared            =   .9868911     |
|              Adjusted R-squared   =   .9840821     |
| Model test   F[  9,    42] (prob) = 351.33 (.0000) |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|   -7.28719016      2.52056245    -2.891   .0061
 LOGI    |     .99299135       .25037574     3.966   .0003   9.67214751
 LOGPG   |     .06051812       .05401018     1.120   .2689   3.72930296
 LOGPNC  |    -.15471632       .26696298     -.580   .5653   4.38036654
 LOGPUC  |    -.48909058       .08519952    -5.741   .0000   4.10544881
 LOGPPT  |     .01926966       .13644891      .141   .8884   4.14194132
```

```
 LOGPD   |    1.73205775       .25988611     6.665   .0000   4.23906603
 LOGPN   |    -.72953933       .26506853    -2.752   .0087   4.23689080
 LOGPS   |    -.86798166       .35291106    -2.459   .0181   4.17535768
 T       |     .03797198       .00751371     5.054   .0000   26.5000000
?=======================================================================
? e.  Correlations of Price Variables
?=======================================================================
Namelist ; Prices = pnc,puc,ppt,pd,pn,ps$
Matrix   ; list ; xcor(prices) $
Correlation Matrix for Listed Variables

              PNC      PUC      PPT       PD       PN       PS
     PNC  1.00000   .99387   .98074   .99327   .98853   .97849
     PUC   .99387  1.00000   .98242   .98783   .98220   .97685
     PPT   .98074   .98242  1.00000   .95847   .98986   .99751
      PD   .99327   .98783   .95847  1.00000   .97734   .95633
      PN   .98853   .98220   .98986   .97734  1.00000   .99358
      PS   .97849   .97685   .99751   .95633   .99358  1.00000
?=======================================================================
? f.  Renormalizations of price variables
?=======================================================================
/*
In the linear case, the coefficients would be divided by the same
scale factor, so that x*b would be unchanged, where x is a variable
and b is the coefficient.  In the loglinear case, since log(k*x)=
log(k)+log(x), the renomalization would simply affect the constant
term.  The price coefficients woulde be unchanged.
*/
?=======================================================================
? g.  Oaxaca decomposition
?=======================================================================
Dates ; 1953 $
Period ; 1953-1973 $
Matrix ; xb0 = Mean(logx)$
Regress ; lhs = logg ; rhs = logx $
Matrix ; b0 = b ; v0 = varb $
Calc ; yb0 = ybar $
Period ; 1974-2004 $
Matrix ; xb1 = mean(logx) $
Regress ; lhs = logg ; rhs = logx $
Matrix ; b1 = b ; v1 = varb $
Calc ; yb1 = ybar $
? Now the decomposition
Calc ; list ; dybar = yb1 - yb0 $ Total
Calc ; list ; dy_dx = b1'xb1 - b1'xb0 $ Change due to change in x
Calc ; list ; dy_db = b1'xb0 - b0'xb0 $
Matrix ; vdb = v1+v0 ; vdb = xb0'[vdb]xb0 $
Calc ; sdb = sqr(vdb)
     ; list ; lower = dy_db - 1.96*sqr(vdb)
            ; upper = dy_db + 1.96*sqr(vdb) $
+------------------------------------+
| Listed Calculator Results          |
+------------------------------------+
 DYBAR   =       .395377
 DY_DX   =       .122745
 DY_DB   =       .272631
 LOWER   =       .184844
 UPPER   =       .360419
```

```
?=======================================================================
? Chapter 4 Application 2
?=======================================================================
Create ; lc = log(cost/pf) ; lpl=log(pl/pf) ; lpk=log(pk/pf)$
Create ; lq = log(q) ; lqq = .5*lq*lq $
Namelist ; x = one,lq,lqq,lpk,lpl $
? a.  Cost function
Regress; lhs = lc ; rhs = x ; printvc $
+----------------------------------------------------+
| Ordinary    least squares regression               |
| LHS=LC       Mean                 =  -.3195570     |
|              Standard deviation   =   1.542364     |
| WTS=none     Number of observs.   =        158     |
| Model size   Parameters           =          5     |
|              Degrees of freedom   =        153     |
| Residuals    Sum of squares       =   2.904896     |
|              Standard error of e  =   .1377906     |
| Fit          R-squared            =   .9922222     |
|              Adjusted R-squared   =   .9920189     |
| Model test   F[  4,   153] (prob) =4879.59 (.0000) |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|   -6.81816332       .25243920   -27.009   .0000
 LQ      |     .40274543       .03148312    12.792   .0000   8.26548908
 LQQ     |     .06089514       .00432530    14.079   .0000   35.7912728
 LPK     |     .16203385       .04040556     4.010   .0001    .85978893
 LPL     |     .15244470       .04659735     3.272   .0013   5.58162250
             1             2             3             4             5
        +---------------------------------------------------------------------
       1|     .06373      -.00238       .00031       .00399      -.01047
       2|    -.00238       .00099      -.00013       .00010      -.00020
       3|     .00031      -.00013    .1870819D-04 -.1493338D-04  .2453652D-04
       4|     .00399       .00010   -.1493338D-04     .00163      -.00102
       5|    -.01047      -.00020    .2453652D-04    -.00102       .00217
?=======================================================================
? b.  capital price coefficient
?=======================================================================
Wald ; fn1 = 1 - b_lpk - b_lpl $
+-----------------------------------------------+
| WALD procedure. Estimates and standard errors |
| for nonlinear functions and joint test of     |
| nonlinear restrictions.                       |
| Wald Statistic            =    266.36109      |
| Prob. from Chi-squared[ 1] =       .00000     |
+-----------------------------------------------+
+--------+--------------+----------------+--------+--------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]|
+--------+--------------+----------------+--------+--------+
 Fncn(1) |     .68552145       .04200352    16.321   .0000
?=======================================================================
? c.  efficient scale
?=======================================================================
Wald ; fn1 = exp((1-b_lq)/b_lqq) $
+-----------------------------------------------+
| WALD procedure. Estimates and standard errors |
| for nonlinear functions and joint test of     |
| nonlinear restrictions.                       |
| Wald Statistic            =     21.74979      |
| Prob. from Chi-squared[ 1] =       .00000     |
+-----------------------------------------------+
+--------+--------------+----------------+--------+--------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]|
+--------+--------------+----------------+--------+--------+
 Fncn(1) |    18177.1045      3897.59890     4.664   .0000
Calc ; qstar = waldfns(1) ; vqstar = varwald(1,1)
```

```
     ; list ; lower = qstar - 1.96*sqr(vqstar)
            ; upper = qstar + 1.96*sqr(vqstar) $
?========================================================================
? d.  Raw data
?========================================================================
+------------------------------------+
| Listed Calculator Results          |
+------------------------------------+
 LOWER   =  10537.810653
 UPPER   =  25816.398344
Create ; output = q $
Sort ; lhs = output $
/*
```

The estimated efficient scale is 18177. There are 25 firms in the sample that have output larger than this. As noted in the problem, many of the largest firms in the sample are aggregates of smaller ones, so it is difficult to draw a conclusion here. However, some of the largest firms (Southern, American Electric power) are singly counted, and are much larger than this scale. The important point is that much of the output in the sample is produced by firms that are smaller than this efficient scale. There are unexploited economies of scale in this industry.

```
*/
```

Chapter 5

Inference and Prediction

Exercises

1. The estimated covariance matrix for the least squares estimator is

$$s^2(\mathbf{X'X})^{-1} = \frac{20}{3900}\begin{bmatrix} 3900/29 & 0 & 0 \\ 0 & 80 & -10 \\ 0 & -10 & 80 \end{bmatrix} = \begin{bmatrix} .69 & 0 & 0 \\ 0 & .40 & -.051 \\ 0 & -.051 & .256 \end{bmatrix}$$ 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. 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 $R^2 = 1 - \mathbf{e'e}/\mathbf{y'M^0y}$, so $\mathbf{y'M^0y} = 520 / (52/60) = 600$. 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. 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 $(\mathbf{X'X})^{-1}\mathbf{R'} = \begin{bmatrix} -(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2(\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2)^{-1} \\ (\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2)^{-1} \end{bmatrix}$ 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 $\mathbf{X}_2$. 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 $\boldsymbol{\beta}_2 = \boldsymbol{\beta}_2^0$ then the preceding is changed by replacing $\mathbf{R}\boldsymbol{\beta} - \mathbf{q} = \mathbf{0}$ with $\mathbf{R}\boldsymbol{\beta} - \boldsymbol{\beta}_2^0 = \mathbf{0}$. Thus, $\mathbf{Rb} - \mathbf{q} = \mathbf{b}_2 - \boldsymbol{\beta}_2^0$. Then, the constrained estimator is

$$\mathbf{b}_* = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} - \begin{bmatrix} -(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2(\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2)^{-1} \\ (\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2)^{-1} \end{bmatrix}(\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2)(\mathbf{b}_2 - \boldsymbol{\beta}_2^0)$$

or

$$\mathbf{b}_* = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} + \begin{bmatrix} (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2(\mathbf{b}_2 - \boldsymbol{\beta}_2^0) \\ (\boldsymbol{\beta}_2^0 - \mathbf{b}_2) \end{bmatrix}$$

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

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

which was to be shown.

4. By factoring the result in (5-14), 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

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

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

$$= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\{[\sigma^2(\mathbf{X}'\mathbf{X})^{-1}]^{-1} - \mathbf{R}'[\mathbf{R}\sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}\mathbf{R}\}\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

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

5. 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 Var[$\mathbf{b}$] to obtain Var[$\mathbf{b}_*$] is positive definite. Consider, then, a quadratic form in Var[$\mathbf{b}_*$]

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

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 Var[$\mathbf{b}_*$] is less than a quadratic form in Var[$\mathbf{b}$] in the same vector.

6. The result follows immediately from the result which precedes (5-19). Since the sum of squared residuals must be at least as large, the coefficient of determination, $COD = 1 -$ sum of squares / $\Sigma_i (y_i - \bar{y})^2$, must be no larger.

7. 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 $\text{Var}[\lambda] = \mathbf{FR}\sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\mathbf{F} = \sigma^2\mathbf{F}$. 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 (5-19) and (5-20). Finally, J times the F statistic in (5-20) equals the expression given above.

8. We use (5-19) 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. First, use (5-19) to write $\mathbf{e}_*'\mathbf{e}_* = \mathbf{e}'\mathbf{e} + (\mathbf{Rb} - \mathbf{q})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{Rb} - \mathbf{q})$. Now, the result that $E[\mathbf{e}'\mathbf{e}] = (n - K)\sigma^2$ obtained in Chapter 6 must hold here, so $E[\mathbf{e}_*'\mathbf{e}_*] = (n - K)\sigma^2 + E[(\mathbf{Rb} - \mathbf{q})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{Rb} - \mathbf{q})]$.
Now, $\mathbf{b} = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon$, so $\mathbf{Rb} - \mathbf{q} = \mathbf{R}\beta - \mathbf{q} + \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon$. But, $\mathbf{R}\beta - \mathbf{q} = \mathbf{0}$, so under the hypothesis, $\mathbf{Rb} - \mathbf{q} = \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon$. Insert this in the result above to obtain
$E[\mathbf{e}_*'\mathbf{e}_*] = (n-K)\sigma^2 + E[\varepsilon'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon]$. The quantity in square brackets is a scalar, so it is equal to its trace. Permute $\varepsilon'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'$ in the trace to obtain

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

We may now carry the expectation inside the trace and use $E[\varepsilon\varepsilon'] = \sigma^2\mathbf{I}$ 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 $E[\mathbf{e}_*'\mathbf{e}_*] = (n - K)\sigma^2 + \sigma^2\text{tr}[\mathbf{I}_J] = (n - K + J)\sigma^2$.

10. 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 $\mathbf{Rb} - \mathbf{q} = \mathbf{0}$ 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 $\mathbf{y} - \mathbf{x}_1 = \beta_2(\mathbf{x}_2 - \mathbf{x}_1) + \alpha\mathbf{i} + \boldsymbol{\varepsilon}$.

Applications

```
?=======================================================================
? Application 5.1 Wage Equation
?=======================================================================
Read;File="F:\Text-Revision\edition6\Solutions-and-Applications\time_var.dat";
nvar=5;nobs=17919$
? This creates the group count variable.
Regress ; Lhs = one ; Rhs = one ; Str = ID ; Panel $
? This READ merges the smaller file into the larger one.
Read;File="F:\Text-Revision\edition6\Solutions-and-Applications\time_invar.dat";
names=ability,med,fed,bh,sibs? ; group=_groupti ;nvar=5;nobs=2178$
Names=id,educ,lwage,pexp,t;
namelist ; x1=one,educ,pexp,ability$
namelist ; x2=med,fed,bh,sibs$
?=======================================================================
? a.  Long regression
?=======================================================================
regress ; lhs= lwage ; rhs = x1,x2 $
+----------------------------------------------------+
| Ordinary    least squares regression               |
| LHS=LWAGE    Mean                 =   2.296821     |
|              Standard deviation   =   .5282364     |
| WTS=none     Number of observs.   =      17919     |
| Model size   Parameters           =          8     |
|              Degrees of freedom   =      17911     |
| Residuals    Sum of squares       =   4119.734     |
|              Standard error of e  =   .4795950     |
| Fit          R-squared            =   .1760081     |
|              Adjusted R-squared   =   .1756861     |
| Model test   F[  7, 17911] (prob) = 546.55 (.0000) |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|      .98965433       .03389449    29.198   .0000
 EDUC    |      .07118866       .00225722    31.538   .0000   12.6760422
 PEXP    |      .03951038       .00089858    43.970   .0000   8.36268765
 ABILITY |      .07736880       .00493359    15.682   .0000    .05237402
 MED     |     .709887D-04      .00169543      .042   .9666   11.4719013
 FED     |      .00531681       .00133795     3.974   .0001   11.7092472
 BH      |     -.05286954       .00999042    -5.292   .0000    .15385903
 SIBS    |      .00487138       .00179116     2.720   .0065   3.15620291
?=======================================================================
? b.  F test
?=======================================================================
Calc ; list ; fstat = Rsqrd/(kreg-1)/((1-rsqrd)/(n-kreg)) $
+--------------------------------+
 FSTAT   =     14.025040
Calc ; r1 = rsqrd ; df1=n-kreg$
Matrix ; b1 = b ; v1 = varb $
Matrix ; b1 =b1(5:8) ; v1=varb(5:8,5:8)$
Regress ; lhs = lwage ; rhs = x1 $
+----------------------------------------------------+
```

```
| Ordinary    least squares regression                 |
| LHS=LWAGE    Mean                 =   2.296821     |
|              Standard deviation   =   .5282364     |
| WTS=none     Number of observs.   =      17919     |
| Model size   Parameters           =          4     |
|              Degrees of freedom   =      17915     |
| Residuals    Sum of squares       =   4132.637     |
|              Standard error of e  =   .4802919     |
| Fit          R-squared            =   .1734272     |
|              Adjusted R-squared   =   .1732888     |
| Model test   F[  3, 17915] (prob) =1252.94 (.0000) |
+----------------------------------------------------+
+--------+-------------+---------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+-------------+---------------+--------+--------+----------+
 Constant|    1.02722913       .03004146    34.194   .0000
 EDUC    |     .07376210       .00221425    33.312   .0000   12.6760422
 PEXP    |     .03948955       .00089835    43.958   .0000   8.36268765
 ABILITY |     .08289072       .00459996    18.020   .0000    .05237402
?=======================================================================
? c.  F test for hypothesis that coefficients on X2 are zero
?=======================================================================
Calc ; list ; fstat = (r1-rsqrd)/(col(x2))/((1-r1)/(df1)) $
+-------------------------------------+
 FSTAT   =     14.025040
?=======================================================================
? c.  Wald test for hypothesis that coefficients on X2 are zero
?=======================================================================
Matrix ; List ; Wald = b1'<v1>b1 $
Matrix WALD     has  1 rows and  1 columns.
               1
        +--------------
       1|   56.10016
Note Wald = 4*F, as expected.

?=======================================================================
? Application 5.2 Translog Cost Function
?=======================================================================
? First prepare the data
?
Create ; lpk=log(pk);lpl=log(pl);lpf=log(pf)$
create ; lpk2=.5*lpk^2 ; lpl2=.5*lpl^2 ; lpf2=.5*lpf^2$
Create ; lpkf=lpk*lpf ; lplf=lpl*lpf ; lpkl=lpk*lpl $
Create ; lq = log(q) ; lq2 = .5*lq^2 $
Create ; lqk=lq*lpk ; lql=lq*lpl ; lqf=lq*lpf $
Create ; lc = log(cost) $
Create ; lcpf = log(cost/pf) $
Create ; lpkpf=log(pk/pf) ; lplpf=log(pl/pf) $
Create ; lpkpf2=.5*lpkpf^2 ; lplpf2=.5*lplpf^2 ; lplfpkf=lplpf*lpkpf $
Create ; lqlpkf=lq*lpkpf ; lqlplf=lq*lplf $
?=======================================================================
? a.  Beta is a,b,dk,dl,df,pkk,pll,pff,pkl,pkf,plf,c,tqk,tql,tqf
?=======================================================================
Restrictions are
            0,0,1,1,1,0,0,0,0,0,0,0,0,0,0      1
            0,0,0,0,0,1,0,0,1,1,0,0,0,0,0      0
    R =     0,0,0,0,0,0,1,0,1,0,1,0,0,0,0  q = 0
            0,0,0,0,0,0,0,1,0,1,1,0,0,0,0      0
            0,0,0,0,0,0,0,0,0,0,0,0,1,1,1      0
?=======================================================================
? b. Testing the theory
?=======================================================================
Namelist ; X1=one,lq,lpk,lpl,lpf,lpk2,lpl2,lpf2,lpkl,lpkf,lplf,lq2,lqk,lq...
Namelist ; X0=one,lq,lpkf,lplf,lpkpf2,lplpf2,lplfpkf,lq2,lqlpkf,lqlplf$
Regress ; lhs = lc ; rhs=x0 $

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

```
| Ordinary    least squares regression                |
| LHS=LC       Mean                 =   3.071619     |
|              Standard deviation   =   1.542734     |
| WTS=none     Number of observs.   =        158     |
| Model size   Parameters           =         10     |
|              Degrees of freedom   =        148     |
| Residuals    Sum of squares       =   2.634416     |
|              Standard error of e  =   .1334170     |
| Fit          R-squared            =   .9929498     |
|              Adjusted R-squared   =   .9925211     |
| Model test   F[  9,   148] (prob) =2316.03 (.0000) |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|   -1.13340208      1.04296294    -1.087   .2789
 LQ      |     .02244828       .12717485      .177   .8601   8.26548908
 LPKF    |    -.02309567       .14153592     -.163   .8706   14.4192992
 LPLF    |    -.01690697       .09185395     -.184   .8542   30.4387314
 LPKPF2  |    -.04730093       .21017152     -.225   .8222    .42211776
 LPLPF2  |    -.03419034       .06850142     -.499   .6184   15.6173009
 LPLFPKF |    -.00741233       .11649585     -.064   .9494   4.84868706
 LQ2     |     .05544306       .00446607    12.414   .0000   35.7912728
 LQLPKF  |     .03562155       .02862683     1.244   .2153   7.15696461
 LQLPLF  |     .01279036       .00375187     3.409   .0008   251.570118
Calc ; ee0 = sumsqdev $
Regress ; lhs = lcpf ; rhs = x1 $
+----------------------------------------------------+
| Ordinary    least squares regression                |
| LHS=LCPF     Mean                 =  -.3195570     |
|              Standard deviation   =   1.542364     |
| WTS=none     Number of observs.   =        158     |
| Model size   Parameters           =         15     |
|              Degrees of freedom   =        143     |
| Residuals    Sum of squares       =   2.464348     |
|              Standard error of e  =   .1312753     |
| Fit          R-squared            =   .9934018     |
|              Adjusted R-squared   =   .9927558     |
| Model test   F[ 14,   143] (prob) =1537.82 (.0000) |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|   -76.2592615      38.2800363    -1.992   .0483
 LQ      |   -1.08042535       .37554512    -2.877   .0046   8.26548908
 LPK     |    6.38079702      4.52920686     1.409   .1611   4.25096457
 LPL     |    14.7182926      7.08482345     2.077   .0395   8.97279814
 LPF     |   -1.89473291      2.84231282     -.667   .5061   3.39117564
 LPK2    |    -.32741427       .44070869     -.743   .4587   9.05539681
 LPL2    |   -1.53852735       .69240298    -2.222   .0279   40.2700121
 LPF2    |    -.07350556       .18203881     -.404   .6870   5.78602018
 LPKL    |    -.57205049       .37189026    -1.538   .1262   38.1346773
 LPKF    |    -.02402470       .24632928     -.098   .9224   14.4192992
 LPLF    |     .16228289       .27007181      .601   .5489   30.4387314
 LQ2     |     .05297849       .00471336    11.240   .0000   35.7912728
 LQK     |     .04014440       .02979137     1.348   .1799   35.1677247
 LQL     |     .13104059       .03828401     3.423   .0008   74.2063474
 LQF     |     .05865220       .02554928     2.296   .0232   28.0107601
Calc ; ee1 = sumsqdev $
Calc ; list ; Fstat = ((ee0 - ee1)/5)/(ee1/(158-15))$
+------------------------------------+
FSTAT    =      1.973714
--> Calc ; list ; ftb(.95,5,143)$
+------------------------------------+
Result  =      2.277490
The F statistic is small; the theory is not rejected.
```

```
?=====================================================================
? c. Testing homotheticity
?=====================================================================
+----------------------------------------------------+
| Ordinary    least squares regression                |
| LHS=LCPF     Mean                 =  -.3195570      |
|              Standard deviation   =   1.542364      |
| WTS=none     Number of observs.   =        158      |
| Model size   Parameters           =         10      |
|              Degrees of freedom   =        148      |
| Residuals    Sum of squares       =   2.634223      |
|              Standard error of e  =   .1334121      |
| Fit          R-squared            =   .9929469      |
|              Adjusted R-squared   =   .9925180      |
| Model test   F[  9,   148] (prob) =2315.08 (.0000) |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|   -2.78239562      1.04292476    -2.668   .0085
 LQ      |     .01362521       .12717020      .107   .9148   8.26548908
 LPKF    |    -.06044098       .14153074     -.427   .6700   14.4192992
 LPLF    |    -.07639000       .09185059     -.832   .4069   30.4387314
 LPKPF2  |    -.10507269       .21016383     -.500   .6178    .42211776
 LPLPF2  |    -.00146323       .06849891     -.021   .9830   15.6173009
 LPLFPKF |     .01806822       .11649158      .155   .8770   4.84868706
 LQ2     |     .05565578       .00446590    12.462   .0000   35.7912728
 LQLPKF  |     .03824257       .02862578     1.336   .1836   7.15696461
 LQLPLF  |     .01296202       .00375173     3.455   .0007   251.570118
Regress ; lhs = lcpf ; Rhs = x0 ; cls:b(9)=0,b(10)=0$
+----------------------------------------------------+
| Linearly restricted regression                      |
| Ordinary    least squares regression                |
| LHS=LCPF     Mean                 =  -.3195570      |
|              Standard deviation   =   1.542364      |
| WTS=none     Number of observs.   =        158      |
| Model size   Parameters           =          8      |
|              Degrees of freedom   =        150      |
| Residuals    Sum of squares       =   2.896172      |
|              Standard error of e  =   .1389526      |
| Fit          R-squared            =   .9922456      |
|              Adjusted R-squared   =   .9918837      |
| Model test   F[  7,   150] (prob) =2741.96 (.0000) |
| Restrictns.  F[  2,   148] (prob) =   7.36 (.0009) |
| Not using OLS or no constant. Rsqd & F may be < 0. |
| Note, with restrictions imposed,  Rsqd may be < 0. |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|   -6.20547247       .37175165   -16.693   .0000
 LQ      |     .40111764       .03208201    12.503   .0000   8.26548908
 LPKF    |    -.05918207       .14502101     -.408   .6838   14.4192992
 LPLF    |     .03234530       .08668866      .373   .7096   30.4387314
 LPKPF2  |    -.20340518       .21249945     -.957   .3400    .42211776
 LPLPF2  |    -.00516132       .06888408     -.075   .9404   15.6173009
 LPLFPKF |     .08684971       .10534811      .824   .4110   4.84868706
 LQ2     |     .06103878       .00440807    13.847   .0000   35.7912728
 LQLPKF  |   -.138778D-16    .517639D-09      .000  1.0000   7.15696461
 LQLPLF  |       .000000     .915064D-10      .000  1.0000   251.570118
Calc ; list ; ftb(.95,2,148)$
+------------------------------------+
Result  =      3.057197
```

The F statistic of 7.36 is larger than the critical value of 3.057. The hypothesis is rejected.

```
?=====================================================================
? d.  Testing generalized Cobb-Douglas against full translog.
?=====================================================================
Regress ; lhs = lcpf ; rhs = x0 ;cls:b(5)=0,b(6)=0,b(7)=0,b(9)=0,b(10)=0$
+----------------------------------------------------+
| Linearly restricted regression                     |
| Ordinary    least squares regression               |
| LHS=LCPF     Mean                 =  -.3195570     |
|              Standard deviation   =   1.542364     |
| WTS=none     Number of observs.   =        158     |
| Model size   Parameters           =          5     |
|              Degrees of freedom   =        153     |
| Residuals    Sum of squares       =   3.191949     |
|              Standard error of e  =   .1444383     |
| Fit          R-squared            =   .9914536     |
|              Adjusted R-squared   =   .9912302     |
| Model test   F[  4,   153] (prob) =4437.33 (.0000) |
| Restrictns.  F[  5,   148] (prob) =   6.27 (.0000) |
| Not using OLS or no constant. Rsqd & F may be < 0. |
| Note, with restrictions imposed,  Rsqd may be < 0. |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|   -5.07718678       .18072495   -28.093   .0000
 LQ      |     .41724916       .03285950    12.698   .0000   8.26548908
 LPKF    |     .00903097       .01466874      .616   .5391   14.4192992
 LPLF    |    -.03131901       .00770196    -4.066   .0001   30.4387314
 LPKPF2  |   -.582867D-15    .127559D-07      .000  1.0000    .42211776
 LPLPF2  |   -.328730D-15    .986857D-08      .000  1.0000   15.6173009
 LPLFPKF |    .461436D-15    .201473D-07      .000  1.0000   4.84868706
 LQ2     |     .05956626       .00452575    13.162   .0000   35.7912728
 LQLPKF  |   -.555112D-16    .538074D-09      .000  1.0000   7.15696461
 LQLPLF  |   -.693889D-17    .223074D-09      .000  1.0000   251.570118

Calc ; list ; ftb(.95,5,148)$
+------------------------------------+
| Listed Calculator Results          |
+------------------------------------+
 Result  =      2.275319
The F statistic of 6.27 is larger than the critical value of 2.275.  The
hypothesis is rejected.

?=====================================================================
? e.  Testing Cobb-Douglas against full translog.
?=====================================================================
Matrix ; b2=b(5:10) ; v2=varb(5:10,5:10) $
Matrix ; list ; Fcd = 1/6 * b2'<v2>b2 $
Matrix FCD      has  1 rows and  1 columns.
               1
        +--------------
       1|   28.87144
Calc ; list ; ftb(.95,6,148)$
+------------------------------------+
| Listed Calculator Results          |
+------------------------------------+
 Result  =      2.160352
The F statistic of 28.871 is larger than the critical value of 2.16.  The
hypothesis is rejected.

?=====================================================================
? f. Testing generalized Cobb-Douglas against homothetic translog.
?=====================================================================
Regress ; Lhs = lcpf ; rhs = one,lq,lpkf,lplf,lpkpf2,lplpf2,lplfpkf,lq2
        ; cls:b(5)=0,b(6)=0,b(7)=0$
+----------------------------------------------------+
| Linearly restricted regression                     |
```

```
| Ordinary    least squares regression                |
| LHS=LCPF     Mean                 =  -.3195570      |
|              Standard deviation   =   1.542364      |
| WTS=none     Number of observs.   =        158      |
| Model size   Parameters           =          5      |
|              Degrees of freedom   =        153      |
| Residuals    Sum of squares       =   3.191949      |
|              Standard error of e  =   .1444383      |
| Fit          R-squared            =   .9914536      |
|              Adjusted R-squared   =   .9912302      |
| Model test   F[  4,   153] (prob) =4437.33 (.0000) |
| Restrictns.  F[  3,   150] (prob) =   5.11 (.0022) |
| Not using OLS or no constant. Rsqd & F may be < 0. |
| Note, with restrictions imposed,  Rsqd may be < 0. |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|   -5.07718678       .18072495    -28.093   .0000
 LQ      |     .41724916       .03285950     12.698   .0000   8.26548908
 LPKF    |     .00903097       .01466874       .616   .5391   14.4192992
 LPLF    |    -.03131901       .00770196     -4.066   .0001   30.4387314
 LPKPF2  |   -.199840D-14    .243505D-07       .000  1.0000    .42211776
 LPLPF2  |   -.746798D-15    .608762D-08       .000  1.0000   15.6173009
 LPLFPKF |    .140166D-14    .121752D-07       .000  1.0000   4.84868706
 LQ2     |     .05956626       .00452575     13.162   .0000   35.7912728

Calc ; list ; ftb(.95,3,150) $
+-----------------------------------+
| Listed Calculator Results          |
+-----------------------------------+
 Result  =      2.664907
?
?=======================================================================
? g.  We have not rejected the theory, but we have rejected all the
? functional forms
? except the nonhomothetic translog.  Just like Christensen and Greene.
?=======================================================================


?=======================================================================
? Application 5.3 Nonlinear restrictions
?=======================================================================
sample;1-52$
name;x=one,logpg,logi,logpnc,logpuc,logppt,t,logpd,logpn,logps$
?=======================================================================
? a.  Simple hypothesis test
?=======================================================================
Regr;lhs=logg;rhs=x$
+----------------------------------------------------+
| Ordinary    least squares regression                |
| LHS=LOGG     Mean                 =   1.570475      |
|              Standard deviation   =   .2388115      |
| WTS=none     Number of observs.   =         52      |
| Model size   Parameters           =         10      |
|              Degrees of freedom   =         42      |
| Residuals    Sum of squares       =   .3812817E-01 |
|              Standard error of e  =   .3012994E-01 |
| Fit          R-squared            =   .9868911      |
|              Adjusted R-squared   =   .9840821      |
| Model test   F[  9,    42] (prob) = 351.33 (.0000) |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|   -7.28719016      2.52056245    -2.891   .0061
 LOGPG   |     .06051812       .05401018     1.120   .2689   3.72930296
```

```
 LOGI    |     .99299135       .25037574     3.966   .0003   9.67214751
 LOGPNC  |    -.15471632       .26696298     -.580   .5653   4.38036654
 LOGPUC  |    -.48909058       .08519952    -5.741   .0000   4.10544881
 LOGPPT  |     .01926966       .13644891      .141   .8884   4.14194132
 T       |     .03797198       .00751371     5.054   .0000   26.5000000
 LOGPD   |    1.73205775       .25988611     6.665   .0000   4.23906603
 LOGPN   |    -.72953933       .26506853    -2.752   .0087   4.23689080
 LOGPS   |    -.86798166       .35291106    -2.459   .0181   4.17535768
Calc;r1=rsqrd$
Regr;lhs=logg;rhs=one,logpg,logi,logpnc,logpuc,logppt,t$
+-----------------------------------------------------+
| Ordinary    least squares regression                |
| LHS=LOGG     Mean                 =   1.570475      |
|              Standard deviation   =   .2388115      |
| WTS=none     Number of observs.   =         52      |
| Model size   Parameters           =          7      |
|              Degrees of freedom   =         45      |
| Residuals    Sum of squares       =   .1014368      |
|              Standard error of e  =   .4747790E-01  |
| Fit          R-squared            =   .9651249      |
|              Adjusted R-squared   =   .9604749      |
| Model test   F[  6,    45] (prob) = 207.55 (.0000)  |
+-----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|   -13.1396625      2.09171186    -6.282   .0000
 LOGPG   |    -.05373342       .04251099    -1.264   .2127   3.72930296
 LOGI    |    1.64909204       .20265477     8.137   .0000   9.67214751
 LOGPNC  |    -.03199098       .20574296     -.155   .8771   4.38036654
 LOGPUC  |    -.07393002       .10548982     -.701   .4870   4.10544881
 LOGPPT  |    -.06153395       .12343734     -.499   .6206   4.14194132
 T       |    -.01287615       .00525340    -2.451   .0182   26.5000000

Calc;r0=rsqrd$
Calc;list;f=((r1-r0)/2)/((1-r1)/(n-10))$
+-----------------------------------+
| Listed Calculator Results         |
+-----------------------------------+
 F       =     34.868735
```

The critical value from the *F* table is 2.827, so we would reject the hypothesis.

```
?========================================================================
? b.  Nonlinear restriction
?========================================================================
```

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

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

Thus, $\mathbf{f} = [-.17399, .10091]'$. The covariance matrix to use for the tests is $\mathbf{G}s^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{G}'$

The statistic for the joint test is $\chi^2 = \mathbf{f}'[\mathbf{G}s^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{G}']^{-1}\mathbf{f} = .4772$. 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 **f**. Thus,

$$z_1 = -.6053$$

and $$z_2 = .2898$$

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

```
?=======================================================================
? c.  Computations for nonlinear restriction
?=======================================================================
sample;1-52$
name;x=one,logpg,logi,logpnc,logpuc,logppt,t,logpd,logpn,logps$
Regr;lhs=logg;rhs=x$
+-----------------------------------------------------+
| Ordinary    least squares regression                |
| LHS=LOGG     Mean                 =   1.570475      |
|              Standard deviation   =   .2388115      |
| WTS=none     Number of observs.   =         52      |
| Model size   Parameters           =          7      |
|              Degrees of freedom   =         45      |
| Residuals    Sum of squares       =   .1014368      |
|              Standard error of e  =   .4747790E-01 |
| Fit          R-squared            =   .9651249      |
|              Adjusted R-squared   =   .9604749      |
| Model test   F[  6,    45] (prob) = 207.55 (.0000) |
+-----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|   -13.1396625      2.09171186    -6.282   .0000
 LOGPG   |    -.05373342       .04251099    -1.264   .2127   3.72930296
 LOGI    |    1.64909204       .20265477     8.137   .0000   9.67214751
 LOGPNC  |    -.03199098       .20574296     -.155   .8771   4.38036654
 LOGPUC  |    -.07393002       .10548982     -.701   .4870   4.10544881
 LOGPPT  |    -.06153395       .12343734     -.499   .6206   4.14194132
 T       |    -.01287615       .00525340    -2.451   .0182   26.5000000
Calc;r1=rsqrd$
Regr;lhs=logg;rhs=one,logpg,logi,logpnc,logpuc,logppt,t$
+-----------------------------------------------------+
| Ordinary    least squares regression                |
| LHS=LOGG     Mean                 =   1.570475      |
|              Standard deviation   =   .2388115      |
| WTS=none     Number of observs.   =         52      |
| Model size   Parameters           =          7      |
|              Degrees of freedom   =         45      |
| Residuals    Sum of squares       =   .1014368      |
|              Standard error of e  =   .4747790E-01 |
| Fit          R-squared            =   .9651249      |
|              Adjusted R-squared   =   .9604749      |
| Model test   F[  6,    45] (prob) = 207.55 (.0000) |
+-----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|   -13.1396625      2.09171186    -6.282   .0000
 LOGPG   |    -.05373342       .04251099    -1.264   .2127   3.72930296
 LOGI    |    1.64909204       .20265477     8.137   .0000   9.67214751
 LOGPNC  |    -.03199098       .20574296     -.155   .8771   4.38036654
 LOGPUC  |    -.07393002       .10548982     -.701   .4870   4.10544881
 LOGPPT  |    -.06153395       .12343734     -.499   .6206   4.14194132
 T       |    -.01287615       .00525340    -2.451   .0182   26.5000000
Calc;r0=rsqrd$
Calc;list;fstat=((r1-r0)/2)/((1-r1)/(n-10))$
+------------------------------------+
 FSTAT   =     34.868735
Calc;list;ftb(.95,3,42)$
+------------------------------------+
 Result  =      2.827049

REGR;Lhs=logg;rhs=x$
Calc ; ds=b(10);dd=-b(8);gpt=-b(6);gnc=b(4)$
Matr;gm=[0,0,0,1,-1,0,0,0,0,0 / 0,0,0,ds,0,dd,0,gpt,0,gnc]$
Calc;f1=b(4)-b(6) ; f2=b(4)*b(10)-b(6)*b(8)$
Matrix;list;f=[f1/f2]$
```

```
Matrix F        has  2 rows and  1 columns.
               1
        +--------------
       1|    -.17399
       2|     .10091
Matrix;list;vf=gm*varb*gm'$
Matrix VF       has  2 rows and  2 columns.
               1             2
        +----------------------------
       1|     .08263      -.08059
       2|    -.08059       .12129
Matrix;list;Wald=f'<vf>f$
Matrix WALD     has  1 rows and  1 columns.
               1
        +--------------
       1|     .47716
Calc;list;z1=f(1)/sqr(vf(1,1))$
+------------------------------------+
 Z1      =      -.605278
Calc;list;z2=f(2)/sqr(vf(2,2))$
+------------------------------------+
 Z2      =       .289760
```

Chapter 6

Functional Form and Structural Change

Exercises

1. T he 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.

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

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

We are interested in the second (lower) of the two coefficients. The matrix product at the end is $\mathbf{e}'\mathbf{e}$ 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 $c_* = -c - \mathbf{e}'\mathbf{e}$ times the off diagonal element in the inverse matrix. The off diagonal element is

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

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

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

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

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

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,

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

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.

3. 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 plim[b'] = Cov[y,x]/Var[y] = $\beta\sigma_*^2/(\beta^2\sigma_*^2 + \sigma_\varepsilon^2)$. Then,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.

4. 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. $plim\begin{bmatrix} \mathbf{x}'\mathbf{x}/n & \mathbf{x}'\mathbf{d}/n \\ \mathbf{d}'\mathbf{x}/n & \mathbf{d}'\mathbf{d}/n \end{bmatrix}^{-1} \begin{pmatrix} \mathbf{x}'\mathbf{y}/n \\ \mathbf{d}'\mathbf{y}/n \end{pmatrix} = \begin{bmatrix} \sigma_*^2 + \sigma_u^2 & 0 \\ 0 & \pi \end{bmatrix}^{-1} \begin{pmatrix} \beta\sigma_*^2 \\ \gamma\pi \end{pmatrix} = \begin{pmatrix} \beta/\left(1 + \sigma_u^2/\sigma_*^2\right) \\ \gamma \end{pmatrix}$. 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($\mathbf{x}'\mathbf{d}/n$). 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[$\mathbf{y}'\mathbf{d}/n$] is now βCov[x^*,d] + γplim($\mathbf{d}'\mathbf{d}/n$) = $\beta\pi\mu^1 + \gamma\pi$ and plim[$\mathbf{y}'\mathbf{x}^*/n$] 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

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

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

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

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

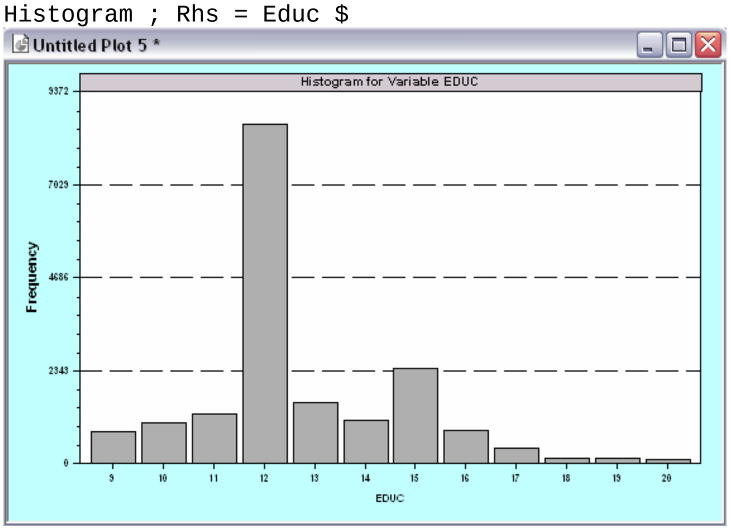
$$E[y|d=1] = \beta E[x^*|d=1] + \gamma = \beta\mu^1 + \gamma$$

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.) □

Applications

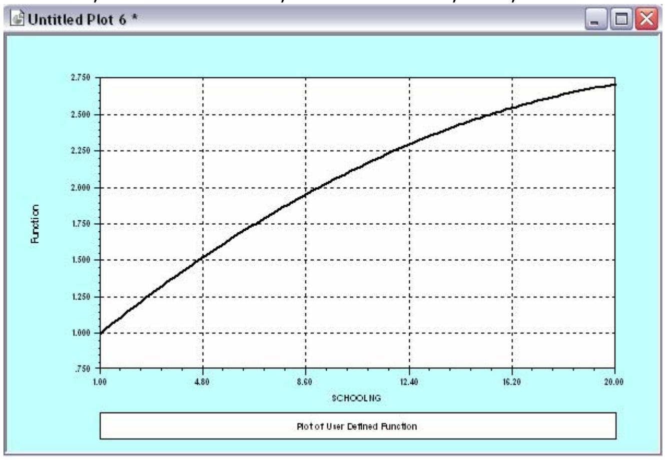
```
?=======================================================================
? Application 6.1
?=======================================================================
a. Wage equation
Namelist ; X = one,educ,ability,pexp,med,fed,bh,sibs$
Regress ; Lhs = lwage ; Rhs = x $
Calc ; xb = b(1)+b(2)*12+b(3)*xbr(ability)+b(4)*xbr(med)
           +b(5)*xbr(fed)+b(6)*0+b(7)*xbr(sibs) $
Calc ; list ; mv = exp(xb) * b(2) $
+----------------------------------------------------+
| Ordinary    least squares regression                |
| LHS=LWAGE    Mean                  =   2.296821     |
|              Standard deviation    =   .5282364     |
| WTS=none     Number of observs.    =      17919     |
| Model size   Parameters            =          7     |
|              Degrees of freedom    =      17912     |
| Residuals    Sum of squares        =   4126.175     |
|              Standard error of e   =   .4799564     |
| Fit          R-squared             =   .1747197     |
|              Adjusted R-squared    =   .1744433     |
| Model test   F[  6, 17912] (prob) = 632.02 (.0000) |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|     .96950956       .03370543    28.764   .0000
 EDUC    |     .07220350       .00225076    32.080   .0000   12.6760422
 ABILITY |     .07746781       .00493727    15.690   .0000    .05237402
 PEXP    |     .03950928       .00089926    43.936   .0000   8.36268765
 MED     |    -.00011702       .00169634     -.069   .9450   11.4719013
 FED     |     .00545695       .00133870     4.076   .0000   11.7092472
 SIBS    |     .00476557       .00179240     2.659   .0078   3.15620291

+------------------------------------+
| Listed Calculator Results          |
+------------------------------------+
 MV      =       .725176b. Step function

?=======================================================================
? b.
?=======================================================================
Histogram ; Rhs = Educ $
```



```
Create ; HS = Educ <= 12 $
Create ; Col = (Educ>12) * (educ <=16) $
Create ; Grad = Educ > 16 $
Regress ; Lhs=lwage ; Rhs = one,Col,Grad,ability,pexp,med,fed,bh,sibs $
+----------------------------------------------------+
| Ordinary    least squares regression               |
| LHS=LWAGE    Mean                 =   2.296821     |
|              Standard deviation   =   .5282364     |
| WTS=none     Number of observs.   =      17919     |
| Model size   Parameters           =          9     |
|              Degrees of freedom   =      17910     |
| Residuals    Sum of squares       =   4215.033     |
|              Standard error of e  =   .4851239     |
| Fit          R-squared            =   .1569472     |
|              Adjusted R-squared   =   .1565706     |
| Model test   F[  8, 17910] (prob) = 416.78 (.0000) |
+----------------------------------------------------+
+--------+-------------+---------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+-------------+---------------+--------+--------+----------+
 Constant|    1.81124933       .02069456    87.523   .0000
 COL     |     .17467913       .00872506    20.020   .0000    .32183716
 GRAD    |     .36244740       .02086328    17.373   .0000    .03493499
 ABILITY |     .10097636       .00486713    20.747   .0000    .05237402
 PEXP    |     .03814088       .00090643    42.078   .0000   8.36268765
 MED     |     .00081934       .00171488      .478   .6328   11.4719013
 FED     |     .00700641       .00135096     5.186   .0000   11.7092472
 BH      |    -.06962521       .01007870    -6.908   .0000    .15385903
 SIBS    |     .00371191       .00181156     2.049   .0405   3.15620291
```

```
c. Education squared
Create ; educsq = educ*educ $
Regress ; Lhs = lwage;rhs=one,educ,educsq,ability,pexp,med,fed,bh,sibs$
+----------------------------------------------------+
| Ordinary    least squares regression               |
| LHS=LWAGE    Mean                 =   2.296821     |
|              Standard deviation   =   .5282364     |
| WTS=none     Number of observs.   =      17919     |
| Model size   Parameters           =          9     |
|              Degrees of freedom   =      17910     |
| Residuals    Sum of squares       =   4114.269     |
|              Standard error of e  =   .4792902     |
| Fit          R-squared            =   .1771010     |
|              Adjusted R-squared   =   .1767334     |
| Model test   F[  8, 17910] (prob) = 481.81 (.0000) |
+----------------------------------------------------+
+--------+-------------+---------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+-------------+---------------+--------+--------+----------+
 Constant|     .42778242       .12008093     3.562   .0004
 EDUC    |     .15590624       .01751608     8.901   .0000   12.6760422
 EDUCSQ  |    -.00313261       .00064230    -4.877   .0000   164.377588
 ABILITY |     .07433494       .00496954    14.958   .0000    .05237402
 PEXP    |     .03962214       .00089830    44.108   .0000   8.36268765
 MED     |     .00030520       .00169504      .180   .8571   11.4719013
 FED     |     .00519423       .00133734     3.884   .0001   11.7092472
 BH      |    -.04957434       .01000691    -4.954   .0000    .15385903
 SIBS    |     .00499325       .00179020     2.789   .0053   3.15620291
Namelist ; x1 = one,educ,educsq,ability,pexp,med,fed,bh,sibs $
Matrix ; means = mean(x1)$
Matrix ; means(2)=0 $
Matrix ; means(3)=0$
Calc   ; a=means'b $
Calc   ; b2=b(2) ; b3=b(3) $
Sample ; 1 $
```

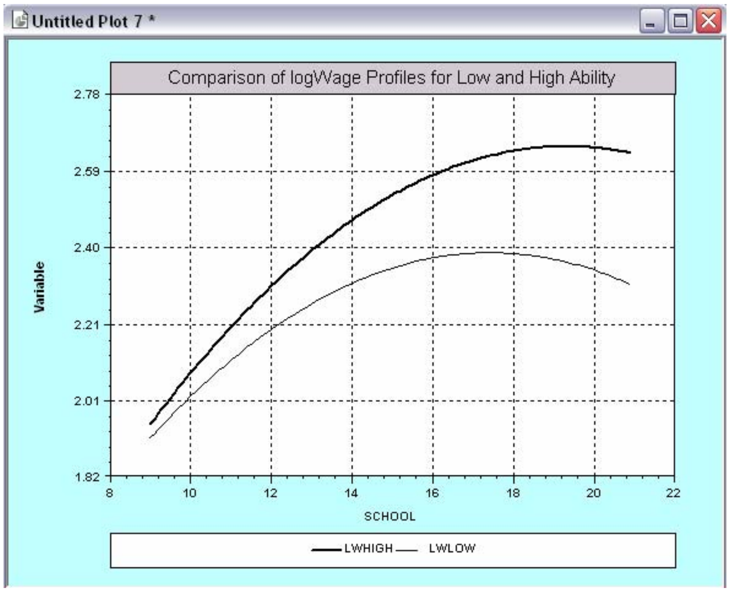
```
Fplot ; fcn = a + b2*schoolng + b3*schoolgn^2 ; pts=100
      ; start = 12 ; limits = 1,20 ; labels=schoolng ; plot(schoolng) $
```




```
d. Interaction.
Sample ; All $
Create ; EA = Educ*ability $
Regress ; Lhs = lwage;rhs=one,educ,ability,ea,pexp,med,fed,bh,sibs$
Calc ; abar =xbr(ability) $
Calc ; list ; me = b(2)+b(4)*abar $
Calc ; sdme = sqr(varb(2,2)+abar^2*varb(4,4) + 2*abar*varb(2,4))$
Calc ; list ; lower = me - 1.96*sdme ; upper = me + 1.96*sdme $
+----------------------------------------------------+
| Ordinary    least squares regression               |
| LHS=LWAGE    Mean                 =   2.296821     |
|              Standard deviation   =   .5282364     |
| WTS=none     Number of observs.   =      17919     |
| Model size   Parameters           =          9     |
|              Degrees of freedom   =      17910     |
| Residuals    Sum of squares       =   4119.377     |
|              Standard error of e  =   .4795877     |
| Fit          R-squared            =   .1760794     |
|              Adjusted R-squared   =   .1757113     |
| Model test   F[  8, 17910] (prob) = 478.44 (.0000) |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|    1.00190276       .03529335    28.388   .0000
 EDUC    |     .07006221       .00243183    28.811   .0000   12.6760422
 ABILITY |     .04693108       .02494471     1.881   .0599    .05237402
 EA      |     .00253975       .00204029     1.245   .2132   1.60372621
 PEXP    |     .03947437       .00089903    43.908   .0000   8.36268765
 MED     |    .542277D-04      .00169546      .032   .9745   11.4719013
 FED     |     .00534599       .00133813     3.995   .0001   11.7092472
 BH      |    -.05314420       .00999271    -5.318   .0000    .15385903
 SIBS    |     .00479076       .00179231     2.673   .0075   3.15620291

+------------------------------------+
| Listed Calculator Results          |
+------------------------------------+
 ME      =       .070195
 LOWER   =       .065503
 UPPER   =       .074888
```

e.

```
Regress ; Lhs = lwage;rhs=one,educ,educsq,ability,ea,pexp,med,fed,bh,sibs$
+----------------------------------------------------+
| Ordinary    least squares regression               |
| LHS=LWAGE    Mean                 =   2.296821     |
|              Standard deviation   =   .5282364     |
| WTS=none     Number of observs.   =      17919     |
| Model size   Parameters           =         10     |
|              Degrees of freedom   =      17909     |
| Residuals    Sum of squares       =   4106.031     |
|              Standard error of e  =   .4788235     |
| Fit          R-squared            =   .1787487     |
|              Adjusted R-squared   =   .1783360     |
| Model test   F[  9, 17909] (prob) = 433.11 (.0000) |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|    -.10514525       .14931731     -.704   .4813
 EDUC    |     .24088793       .02252126    10.696   .0000   12.6760422
 EDUCSQ  |    -.00654261       .00085754    -7.630   .0000   164.377588
 ABILITY |    -.12453442       .03354596    -3.712   .0002    .05237402
 EA      |     .01631824       .00272231     5.994   .0000   1.60372621
 PEXP    |     .03951247       .00089761    44.020   .0000   8.36268765
 MED     |     .00045246       .00169356      .267   .7893   11.4719013
 FED     |     .00524829       .00133606     3.928   .0001   11.7092472
 BH      |    -.04775208       .01000179    -4.774   .0000    .15385903
 SIBS    |     .00460796       .00178961     2.575   .0100   3.15620291
+-----------------------------------+
| Listed Calculator Results         |
+-----------------------------------+
 AVGLOW  =      -.798563
 AVGHIGH =       .717891
Create ; lowa = ability < xbr(ability) ; higha = 1 - lowa $
Calc   ; list ; avglow= lowa'ability / lowa'lowa ;
avghigh=higha'ability/higha'higha $
Calc   ; a = b(1) + b(6)*xbr(pexp)+b(7)*xbr(med)+
             b(8)*xbr(fed)+b(9)*xbr(bh)+b(10)*xbr(sibs)$
Calc   ; al=a+b(4)*avglow ; ah = a+b(4)*avghigh$
Samp;1-120$
Create ; school = trn(9,.1)$
Create ; lwlow  = al + b(2)*school+b(3)*school^2 + b(5)*avglow*school $
Create ; lwhigh = ah + b(2)*school+b(3)*school^2 + b(5)*avghigh*school $
Plot ; lhs = school ; rhs =lwhigh,lwlow ;fill ;grid
;Title=Comparison of logWage Profiles for Low and High Ability$
```



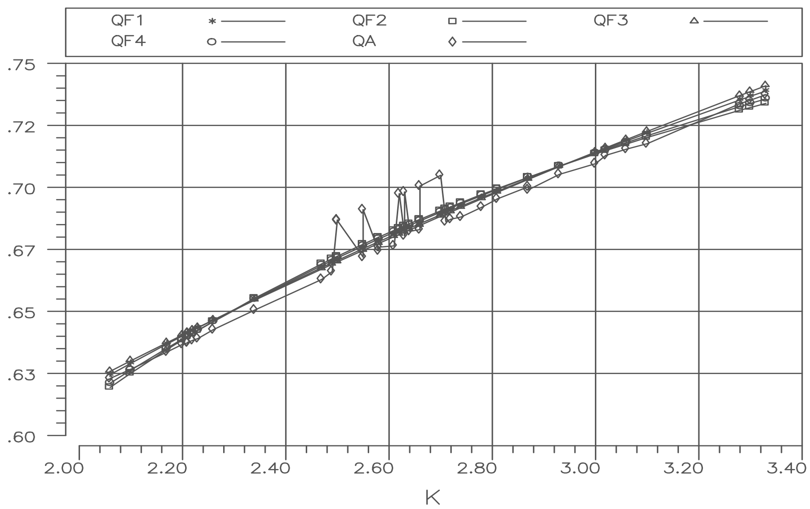

```
?======================================================================
? Application 6.2
?======================================================================
Sample ; All $
Namelist ; X = one,educ,ability,pexp,med,fed,sibs$
Regress ; For [bh=0] ; Lhs = lwage ; Rhs = x $
Calc ; ee0=sumsqdev $
Matrix ; b0=b ; v0=varb $
Regress ; For [bh=1] ; Lhs = lwage ; Rhs = x $
Calc ; ee1=sumsqdev $
Matrix ; b1=b ; v1=varb $
Regress ; Lhs = lwage ; Rhs = x $
Calc ; ee=sumsqdev $
Calc ; list ; chow = ((ee-ee0-ee1)/col(x))/ ((ee0+ee1)/(n-2*col(x))) $
+-------------------------------------+
| Listed Calculator Results           |
+-------------------------------------+
 CHOW    =      7.348379
Matrix ; db=b0-b1 ; vdb=v0+v1 $
Matrix ; list ; Wald = db'<vdb>db $
Matrix WALD     has  1 rows and  1 columns.
               1
        +--------------
       1|   50.57114
```

```
?=======================================================================
? Application 6.3
?=======================================================================
```

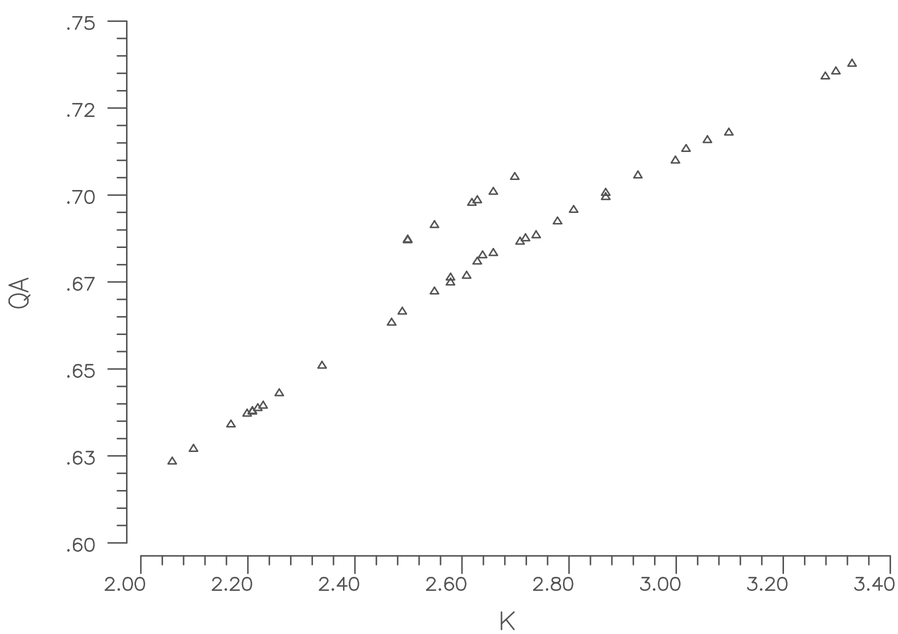
a. The least squares estimates of the four models are

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

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.



b. The scatter diagram is shown below. The last seven years of the data set show clearly the effect observed by Solow.



.

c. 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	**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)		

d. 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: $F = \quad = 120.43$
Model 3: $F = \quad = 1371.0$
Model 4: $F = \quad = 234.64$

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. □

```
?=====================================================================
? Application 6.4
?=====================================================================
```
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:

$$\text{type}: F[4,12] = \frac{(3925.2 - 660.9)/4}{660.9/12} = 14.82,$$

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

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.

Regression Coefficients				
	Full Model	***Time Effects***	***Type Effects***	***No Effects***
Constant	3.4	6.0	8.25	10.85
B	27.75	0	27.75	0
C	–7.0	0	–7.0	0
D	–4.5	0	–4.5	0
E	–3.25	0	–3.25	0
65–69	7.0	7.0	0	0
70–74	11.4	11.4	0	0
75–79	1.0	1.0	0	0
R^2	0.84823	0.0986	0.74963	0
$\mathbf{e}'\mathbf{e}$	660.9	3925.2	1090.2	4354.5

Chapter 7

Specification Analysis and Model Selection

Exercises

1. The result cited is E[$\mathbf{b}_1$] = β_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 $P_{1.2}$ 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

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

So, when we regress $\mathbf{y}$ on $\mathbf{X}_1$ alone and compute the predictions, we are computing an estimator of $\mathbf{X}_1(\beta_1 + \mathbf{P}_{1.2}\beta_2) = \mathbf{X}_1\beta_1 + E[\mathbf{X}_2|\mathbf{X}_1]\beta_2$. Both parts of the compound disturbance in this regression ε and $(X_2 - E[X_2|X_1])\beta_2$ have mean zero and are uncorrelated with $\mathbf{X}_1$ and $E[\mathbf{X}_2|\mathbf{X}_1]$, so the prediction error has mean zero. The implication is that the forecast is unbiased. Note that this is not true if $E[\mathbf{X}_2|\mathbf{X}_1]$ 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 $E[\mathbf{X}_2|\mathbf{X}_1]$ and $\mathbf{X}_1\mathbf{P}_{1.2}$.

2. 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; E[$\mathbf{b}_1$] = β_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 - Var[$\mathbf{b}_{1.2}$] $\geq$ 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 Var[$\mathbf{b}_1$] + **bias**×**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 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 log likelihood function at the maximum is

$$\begin{aligned}\ln L &= -n/2[1 + \ln 2\pi + \ln(\mathbf{e}'\mathbf{e}/n)]\\ &= -n/2\{1 + \ln 2\pi + \ln[nS_{yy}(1 - R^2)]\}\\ &= -n/2\{1 + \ln 2\pi + \ln(nS_{yy}) + \ln(1-R^2)\} \text{ where } S_{yy} = \Sigma_{i=1}^{n}(y_i - \bar{y})^2\end{aligned}$$

since $R^2 = 1 - e'e/S_{yy}$. The derivative of this expression is $\partial\ln L/\partial R^2 = (-n/2)\{1/(1-R^2)\}(-1)$ which is always positive. Therefore, the log likelihood increases when R^2 increases.

4. An inconvenient way to obtain the result is by repeated substitution of C_{t-1}, then C_{t-2} and so on. It is much easier and faster to introduce the lag operator used in Chapter 20. Thus, the alternative model is

$$C_t = \gamma_1 + \gamma_2 Y_t + \gamma_3 L C_t + \varepsilon_{1t} \text{ where } LC_t = C_{t-1}.$$

Then, $(1 - \gamma_3 L)C_t = \gamma_1 + \gamma_2 Y_t + \varepsilon_{1t}$.

Now, multiply both sides of the equation by $1/(1-\gamma_3 L) = 1 + \gamma_3 L + \gamma_3^2 L^2 + \ldots$ to obtain

$$\boldsymbol{C_t = \gamma_1/(1 - \gamma_3) + \gamma_2 Y_t + \gamma_2\gamma_3 Y_{t-1} + \Sigma_{s=2}^{\infty}\gamma_2\gamma_3^s Y_{t-s} + \Sigma_{s=0}^{\infty}\gamma_3^s\varepsilon_{t-s}}.$$

Application

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.

```
?=====================================
? Example 7.2 and Application 7.1
?=====================================
Dates  ; 1950.1 $
Period ; 1950.1 - 2000.4 $
Create ; Ct = Realcons ; Yt = RealDPI $
Create ; Ct1 = Ct[-1] ; Yt1 = Yt[-1] $
? Example 7.2
Period ; 1950.2 - 2000.4 $
Regress; Lhs = Ct ; Rhs = one,Yt,Yt1 ; Keep = CY $
Regress; Lhs = Ct ; Rhs = one,Yt,Ct1 ; Keep = CC $
Regress; Lhs = Ct ; Rhs = one,Yt,Yt1,CC $
+----------------------------------------------------+
| Ordinary    least squares regression               |
| Model was estimated May 12, 2007 at 08:56:19AM     |
| LHS=CT       Mean                 =   3008.995     |
|              Standard deviation   =   1456.900     |
| WTS=none     Number of observs.   =        203     |
| Model size   Parameters           =          4     |
|              Degrees of freedom   =        199     |
| Residuals    Sum of squares       =   73550.21     |
|              Standard error of e  =   19.22496     |
| Fit          R-squared            =   .9998285     |
|              Adjusted R-squared   =   .9998259     |
| Model test   F[  3,   199] (prob) =******* (.0000) |
| Diagnostic   Log likelihood       =  -886.1351     |
|              Restricted(b=0)      =  -1766.209     |
|              Chi-sq [  3]  (prob) =1760.15 (.0000) |
| Info criter. LogAmemiya Prd. Crt. =   5.931932     |
|              Akaike Info. Criter. =   5.931926     |
| Autocorrel   Durbin-Watson Stat.  =  2.0256102     |
|              Rho = cor[e,e(-1)]   =  -.0128051     |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|    -.60444607      3.43245774     -.176   .8604
 YT      |     .31456542       .04619552     6.809   .0000   3352.09360
 YT1     |    -.33004915       .04591940    -7.188   .0000   3325.25222
 CC      |    1.01450597       .01613899    62.861   .0000   3008.99507
Regress; Lhs = Ct ; Rhs = one,Yt,Ct1,CY $
+----------------------------------------------------+
| Ordinary    least squares regression               |
| Model was estimated May 12, 2007 at 08:56:19AM     |
| LHS=CT       Mean                 =   3008.995     |
|              Standard deviation   =   1456.900     |
| WTS=none     Number of observs.   =        203     |
| Model size   Parameters           =          4     |
|              Degrees of freedom   =        199     |
| Residuals    Sum of squares       =   73550.21     |
|              Standard error of e  =   19.22496     |
| Fit          R-squared            =   .9998285     |
|              Adjusted R-squared   =   .9998259     |
| Model test   F[  3,   199] (prob) =******* (.0000) |
| Diagnostic   Log likelihood       =  -886.1351     |
|              Restricted(b=0)      =  -1766.209     |
|              Chi-sq [  3]  (prob) =1760.15 (.0000) |
| Info criter. LogAmemiya Prd. Crt. =   5.931932     |
|              Akaike Info. Criter. =   5.931926     |
```

```
| Autocorrel   Durbin-Watson Stat.  =  2.0256102     |
|              Rho = cor[e,e(-1)]   =  -.0128051     |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|   -865.712368      120.569071    -7.180   .0000
 YT      |    9.82505250      1.36759557     7.184   .0000   3352.09360
 CT1     |    1.02780685       .01635059    62.861   .0000   2982.97438
 CY      |   -10.6765577      1.48541853    -7.188   .0000   3008.99507
? Application 7.1.  We use only the 1980 data, so we
? start in quarter 2 of 1980 even though data are
? available for the last quarter of 1979.
Period ; 1980.2 - 2000.4 $
Regress; Lhs = Ct ; Rhs = one,Yt,Yt1 ; Keep = CY $
Regress; Lhs = Ct ; Rhs = one,Yt,Ct1 ; Keep = CC $
Regress; Lhs = Ct ; Rhs = one,Yt,Yt1,CC $
+----------------------------------------------------+
| Ordinary    least squares regression               |
| Model was estimated May 12, 2007 at 08:58:19AM     |
| LHS=CT       Mean                 =   4503.230     |
|              Standard deviation   =   879.3593     |
| WTS=none     Number of observs.   =         83     |
| Model size   Parameters           =          4     |
|              Degrees of freedom   =         79     |
| Residuals    Sum of squares       =   43603.43     |
|              Standard error of e  =   23.49345     |
| Fit          R-squared            =   .9993123     |
|              Adjusted R-squared   =   .9992862     |
| Model test   F[  3,    79] (prob) =******* (.0000) |
| Diagnostic   Log likelihood       =  -377.7300     |
|              Restricted(b=0)      =  -679.9419     |
|              Chi-sq [  3]  (prob) = 604.42 (.0000) |
| Info criter. LogAmemiya Prd. Crt. =   6.360511     |
|              Akaike Info. Criter. =   6.360436     |
| Autocorrel   Durbin-Watson Stat.  =  1.8153241     |
|              Rho = cor[e,e(-1)]   =   .0923379     |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|    39.6958824      37.1402619     1.069   .2884
 YT      |     .20222923       .07364203     2.746   .0075   4987.32410
 YT1     |    -.25661196       .07221392    -3.553   .0006   4951.70482
 CC      |    1.04938412       .04670690    22.467   .0000   4503.23012
Regress; Lhs = Ct ; Rhs = one,Yt,Ct1,CY $
+----------------------------------------------------+
| Ordinary    least squares regression               |
| Model was estimated May 12, 2007 at 08:58:19AM     |
| LHS=CT       Mean                 =   4503.230     |
|              Standard deviation   =   879.3593     |
| WTS=none     Number of observs.   =         83     |
| Model size   Parameters           =          4     |
|              Degrees of freedom   =         79     |
| Residuals    Sum of squares       =   43603.43     |
|              Standard error of e  =   23.49345     |
| Fit          R-squared            =   .9993123     |
|              Adjusted R-squared   =   .9992862     |
| Model test   F[  3,    79] (prob) =******* (.0000) |
| Diagnostic   Log likelihood       =  -377.7300     |
|              Restricted(b=0)      =  -679.9419     |
|              Chi-sq [  3]  (prob) = 604.42 (.0000) |
| Info criter. LogAmemiya Prd. Crt. =   6.360511     |
|              Akaike Info. Criter. =   6.360436     |
| Autocorrel   Durbin-Watson Stat.  =  1.8153241     |
|              Rho = cor[e,e(-1)]   =   .0923379     |
+----------------------------------------------------+
```

```
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|   -856.107861      221.141722    -3.871   .0002
 YT      |    1.21490273       .32340906     3.757   .0003   4987.32410
 CT1     |     .98759074       .04395654    22.467   .0000   4465.65542
 CY      |   -1.13474451       .31933175    -3.553   .0006   4503.23012
?
? The results are essentially the same.  This suggests
? that neither model is right.
```

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.

Chapter 8

The Generalized Regression Model and Heteroscedasticity

Exercises

1. Write the two estimators as $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}$ and $\mathbf{b} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}$. Then, $(\hat{\boldsymbol{\beta}} - \mathbf{b}) = [(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\boldsymbol{\varepsilon}$ has $E[\hat{\boldsymbol{\beta}} - \mathbf{b}] = \mathbf{0}$ since both estimators are unbiased. Therefore,

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

Then,

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

once the inverse matrices are multiplied.

2 First, $(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q}) = \mathbf{R}[\boldsymbol{\beta} + (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon})] - \mathbf{q} = \mathbf{R}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}$ if $\mathbf{R}\boldsymbol{\beta} - \mathbf{q} = \mathbf{0}$.
Now, use the inverse square root matrix of $\boldsymbol{\Omega}$, $\mathbf{P} = \boldsymbol{\Omega}^{-1/2}$ to obtain the transformed data,

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

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

and,

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y} = (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{y}^* \\ &= \text{the OLS estimator in the regression of } \mathbf{y}^* \text{ on } \mathbf{X}^*. \end{aligned}$$

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

and the numerator is $\boldsymbol{\varepsilon}^{*\prime}\mathbf{X}^*(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{R}']^{-1}\mathbf{R}(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\boldsymbol{\varepsilon}^* / J$. 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 $\mathbf{Q} = \mathbf{X}^*\mathbf{ABA}'\mathbf{X}^{*\prime}$ 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 $(\mathbf{y}^* - \mathbf{X}^*\hat{\boldsymbol{\beta}})'(\mathbf{y}^* - \mathbf{X}^*\hat{\boldsymbol{\beta}})/(n - K)$. This is exactly the sum of squared residuals in the least squares regression of $\mathbf{y}^*$ on $\mathbf{X}^*$. Since $\mathbf{y}^* = \mathbf{X}^*\boldsymbol{\beta} + \boldsymbol{\varepsilon}^*$ and $\hat{\boldsymbol{\beta}} = (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{y}^*$ the denominator is $\boldsymbol{\varepsilon}^{*\prime}\mathbf{M}^*\boldsymbol{\varepsilon}^*/(n - K)$, 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. 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

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

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

Note that this is a $J\times 1$ vector. By multiplying it out, we find that $\mathrm{E}[\boldsymbol{\varepsilon}^0\boldsymbol{\varepsilon}^{0\prime}] = \mathrm{Var}[\boldsymbol{\varepsilon}^0] = \mathbf{R}\{\sigma^2(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\}\mathbf{R}'$. Therefore, the modified statistic can be written as $W^* = \boldsymbol{\varepsilon}^{0\prime}\mathrm{Var}[\boldsymbol{\varepsilon}^0]^{-1}\boldsymbol{\varepsilon}^0/J$. This is the 'full rank quadratic form' discussed in Appendix B. For convenience, let $\mathbf{C} = \mathrm{Var}[\boldsymbol{\varepsilon}^0]$, $\mathbf{T} = \mathbf{C}^{-1/2}$, and $\mathbf{v} = \mathbf{T}\boldsymbol{\varepsilon}^0$. Then, $W^* = \mathbf{v}'\mathbf{v}$. By construction, $\mathbf{v} = \mathrm{Var}[\boldsymbol{\varepsilon}^0]^{-1/2}\boldsymbol{\varepsilon}^0$, so $E[\mathbf{v}] = \mathbf{0}$ and $\mathrm{Var}[\mathbf{v}] = \mathbf{I}$. 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. The development is unchanged. As long as the limiting behavior of $(1/n)\hat{\mathbf{X}}'\hat{\mathbf{X}} = (1/n)\mathbf{X}'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{X}$ 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 $\boldsymbol{\Omega}$ were used instead of the estimate $\hat{\boldsymbol{\Omega}}$.

5. 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

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

with variance $\mathrm{Var}[\hat{\beta}] = \mathbf{D}^{1/2}\sigma^2[\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}]^{-1}(\mathbf{D}')^{1/2} = \mathbf{D}^{1/2}\mathrm{Var}[\hat{\beta}^0](\mathbf{D}')^{1/2}$. 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 $\mathrm{Var}[\mathbf{b}] = \mathbf{D}^{1/2}\mathrm{Var}[\mathbf{b}^0](\mathbf{D}')^{1/2}$. 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 $\hat{\beta} = \mathbf{X}'\mathbf{y}$ with variance $\mathrm{Var}[\hat{\beta}] = \sigma^2\boldsymbol{\Lambda}_K$. The OLS estimator is $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{y}$. Its variance is $\mathrm{Var}[\mathbf{b}] = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2\boldsymbol{\Lambda}_K$, which means that OLS and GLS are identical in this case.

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

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

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

Thus, $E[\hat{\boldsymbol{\varepsilon}}\hat{\boldsymbol{\varepsilon}}'] = [\mathbf{I} - \mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}]E[\boldsymbol{\varepsilon\varepsilon}'][\mathbf{I} - \mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}]'$

$$= [\mathbf{I} - \mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}](\sigma^2\boldsymbol{\Omega})[\mathbf{I} - \mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}]'$$
$$= [\sigma^2\boldsymbol{\Omega} - \sigma^2\mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'][\mathbf{I} - \mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}]'$$
$$= [\sigma^2\boldsymbol{\Omega} - \sigma^2\mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'][\mathbf{I} - \boldsymbol{\Omega}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}']$$
$$= \sigma^2\boldsymbol{\Omega} - \sigma^2\mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}' - \sigma^2\mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}' + \sigma^2\mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1})\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'$$
$$= \sigma^2[\boldsymbol{\Omega} - \mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}']$$

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

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

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

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

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

so the likelihood equations are equivalent to $\Sigma_i(\mathbf{x}_i y_i/(\boldsymbol{\beta}'\mathbf{x}_i).^2) = \Sigma_i\mathbf{x}_i\mathbf{x}_i'\boldsymbol{\beta}/(\boldsymbol{\beta}'\mathbf{x}_i).^2$, 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 13.

8. 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 8.2 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.

9. This is a heteroscedastic regression model in which the matrix $\mathbf{X}$ is a column of ones. The efficient estimator is the GLS estimator, $\hat{\beta} = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y} = [\Sigma_i 1y_i/x_i^2] / [\Sigma_i 1^2/x_i^2] = [\Sigma_i(y_i/x_i^2)] / [\Sigma_i(1/x_i^2)]$. As always, the variance of the estimator is $\text{Var}[\hat{\beta}] = \sigma^2(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1} = \sigma^2/[\Sigma_i(1/x_i^2)]$. The ordinary least squares estimator is $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \bar{y}$. The variance of $\bar{y}$ is $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\boldsymbol{\Omega}\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1} = (\sigma^2/n^2)\Sigma_i x_i^2$. 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.

10. Consider, first, $\bar{y}$. We saw earlier that 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 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 $s_*^2 = (1/(n\text{-}1))\Sigma_i(y_i - \alpha)^2$ have the same probability limit. We consider, therefore, plim s_*^2 = plim$(1/(n\text{-}1))\Sigma_i\varepsilon_i^2$. The expected value of s_*^2 is $E[(1/(n\text{-}1))\, \Sigma_i\varepsilon_i^2] = \sigma^2(1/\Sigma_i x_i^2)$. 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 Var[s_*^2] = Σ_iVar[ε_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 $(n/(n\text{-}1)^2)$ times the average of these fourth moments. If every fourth moment is finite, then the term is dominated by the leading $(n/(n\text{-}1)^2)$ which converges to zero. It follows that plim $s_*^2 = \sigma^2\bar{q}$. Therefore, the conventional estimator estimates Asy.Var[$\bar{y}$]= $\sigma^2\bar{q}/n$.

The appropriate variance of the least squares estimator is 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.

11. The sample moments are obtained using, for example, $S_{xx} = \mathbf{x}'\mathbf{x} - n\bar{x}^2$ 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 **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 method of Example 8.5 can be applied to this groupwise heteroscedastic model. 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 **X′X** matrices are the same in this problem, so this simplifies to $\hat{\beta} = [(1/s_1^2 + 1/s_2^2)\mathbf{X}'\mathbf{X}]^{-1}[(1/s_1^2)\mathbf{X}_1'\mathbf{y}_1 + (1/s_2^2)\mathbf{X}_2'\mathbf{y}_2]$. The estimator is, therefore $\left[\left(\frac{1}{3.472}+\frac{1}{9.722}\right)\begin{pmatrix}50 & 300\\ 300 & 2100\end{pmatrix}\right]^{-1}\left[\frac{1}{3.472}\begin{pmatrix}300\\2000\end{pmatrix}+\frac{1}{9.722}\begin{pmatrix}300\\2200\end{pmatrix}\right]=\begin{pmatrix}.9469\\.8422\end{pmatrix}$.

```
?=======================================================
? Application 8.1
?=======================================================
```

a. 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

b.

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

c. 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 50×0.78296 = 39.148. The critical value from the table of chi-squared with 5 degrees of freedom is 11.08, so we would conclude that there is evidence of heteroscedasticity.

d. Lagrange multiplier test.

```
Regress;Lhs=y;rhs=one,x1,x2 ; Res=e ; het $
create ; lmi=e*e/(sumsqdev/n) - 1 $
Name ; x=one,x1,x2 $
Calc ; list ; .5*xss(x,lmi)$
The result was reported with the regression,
| Br./Pagan LM Chi-sq [  2]  (prob) =  72.78 (.0000) |
```

e. Two step estimator

```
read;nobs=50;nvar=1;names=y;byva $
   -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
read;nobs=50;nvar=1;names=x1;byva $
   -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
read;nobs=50;nvar=1;names=x2;byva $
    -.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
Regress;Lhs=y;rhs=one,x1,x2 ; Res=e $
+----------------------------------------------------+
| Ordinary    least squares regression               |
| Model was estimated May 12, 2007 at 08:33:20PM     |
| LHS=Y        Mean                 =   .3938000     |
|              Standard deviation   =   6.368374     |
| WTS=none     Number of observs.   =         50     |
| Model size   Parameters           =          3     |
```

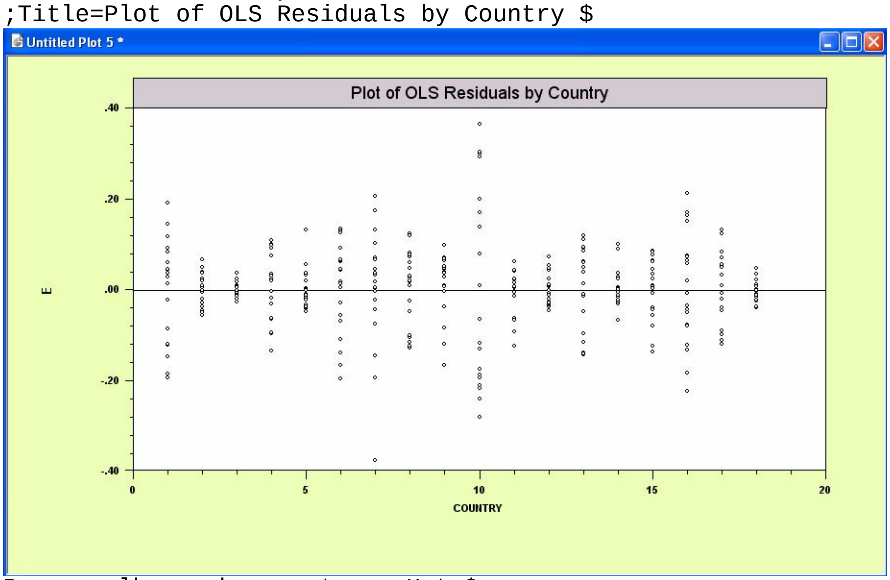
```
|                Degrees of freedom   =        47      |
| Residuals      Sum of squares       =   1911.928     |
|                Standard error of e  =   6.378033     |
| Fit            R-squared            =   .3790450E-01 |
|                Adjusted R-squared   =  -.3035736E-02 |
| Model test     F[  2,    47] (prob) =    .93 (.4033) |
| Diagnostic     Log likelihood       =  -162.0430     |
|                Restricted(b=0)      =  -163.0091     |
|                Chi-sq [  2]  (prob) =   1.93 (.3806) |
| Info criter.   LogAmemiya Prd. Crt. =   3.763988     |
|                Akaike Info. Criter. =   3.763844     |
| Autocorrel     Durbin-Watson Stat.  =  1.8560359     |
|                Rho = cor[e,e(-1)]   =   .0719820     |
+-----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|     .19039401       .91444640      .208   .8360
 X1      |    1.13113339       .98260352     1.151   .2555    .10820000
 X2      |     .37682493       .43992218      .857   .3960    .21500000
Create ; e2 = e*e $
Create ; loge2 = log(e2) $
Regress ; lhs = loge2 ; Rhs = one,x1,x2 ; keep=vi $
Create ; vi = 1/exp(vi) $
Regress ; Lhs = y ; rhs = one,x1,x2 ; wts = vi $
+-----------------------------------------------------+
| Ordinary    least squares regression                |
| Model was estimated May 12, 2007 at 08:33:20PM      |
| LHS=Y          Mean                 =  -.5316339     |
|                Standard deviation   =   4.535703     |
| WTS=VI         Number of observs.   =        50      |
| Model size     Parameters           =         3      |
|                Degrees of freedom   =        47      |
| Residuals      Sum of squares       =   890.9017     |
|                Standard error of e  =   4.353775     |
| Fit            R-squared            =   .1162193     |
|                Adjusted R-squared   =   .7861157E-01 |
| Model test     F[  2,    47] (prob) =   3.09 (.0548) |
| Diagnostic     Log likelihood       =  -150.0732     |
|                Restricted(b=0)      =  -153.1619     |
|                Chi-sq [  2]  (prob) =   6.18 (.0456) |
| Info criter.   LogAmemiya Prd. Crt. =   3.000355     |
|                Akaike Info. Criter. =   3.285051     |
| Autocorrel     Durbin-Watson Stat.  =  1.9978648     |
|                Rho = cor[e,e(-1)]   =   .0010676     |
+-----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|     .16662621       .71981411      .231   .8179
 X1      |     .77648745       .63883379     1.215   .2303   -.51884171
 X2      |     .84717700       .36328984     2.332   .0240   -.34867101
```

Applications

```
?=======================================================
? Application 8.2  Gasoline Consumption
?=======================================================
? Rename variable for convenience
Create ; y=lgaspcar $
? RHS of new regression
Namelist ; x = one,lincomep,lrpmg,lcarpcap $
? Base regression.  Is cars per capita significant?
Regress ; Lhs = y ; Rhs = x $
+-----------------------------------------------------+
| Ordinary    least squares regression                |
| LHS=Y        Mean                 =   4.296242      |
|              Standard deviation   =   .5489071      |
| WTS=none     Number of observs.   =        342      |
| Model size   Parameters           =          4      |
|              Degrees of freedom   =        338      |
| Residuals    Sum of squares       =   14.90436      |
|              Standard error of e  =   .2099898      |
| Fit          R-squared            =   .8549355      |
|              Adjusted R-squared   =   .8536479      |
| Model test   F[  3,   338] (prob) = 664.00 (.0000) |
+-----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|    2.39132562       .11693429    20.450   .0000
 LINCOMEP|     .88996166       .03580581    24.855   .0000  -6.13942544
 LRPMG   |    -.89179791       .03031474   -29.418   .0000   -.52310321
 LCARPCAP|    -.76337275       .01860830   -41.023   .0000  -9.04180473Calc ; r0
= rsqrd $
Namelist ; Cntry=c2,c3,c4,c5,c6,c7,c8,c9,c10,c11,c12,c13,c14,c15,c16,c17,c18$
Regress;lhs=y;rhs=x,cntry ; Res = e $
+-----------------------------------------------------+
| Ordinary    least squares regression                |
| LHS=Y        Mean                 =   4.296242      |
|              Standard deviation   =   .5489071      |
| WTS=none     Number of observs.   =        342      |
| Model size   Parameters           =         21      |
|              Degrees of freedom   =        321      |
| Residuals    Sum of squares       =   2.736491      |
|              Standard error of e  =   .9233035E-01 |
| Fit          R-squared            =   .9733657      |
|              Adjusted R-squared   =   .9717062      |
| Model test   F[ 20,   321] (prob) = 586.56 (.0000) |
+-----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|    2.28585577       .22832349    10.011   .0000
 LINCOMEP|     .66224966       .07338604     9.024   .0000  -6.13942544
 LRPMG   |    -.32170246       .04409925    -7.295   .0000   -.52310321
 LCARPCAP|    -.64048288       .02967885   -21.580   .0000  -9.04180473
 C2      |    -.12030455       .03414942    -3.523   .0005    .05555556
 C3      |     .75598453       .04074554    18.554   .0000    .05555556
 C4      |     .10360026       .03660467     2.830   .0049    .05555556
 C5      |    -.08108439       .03356343    -2.416   .0163    .05555556
 C6      |    -.13598740       .03187957    -4.266   .0000    .05555556
 C7      |     .05125389       .04152961     1.234   .2180    .05555556
 C8      |     .30646950       .03529373     8.683   .0000    .05555556
 C9      |    -.05330785       .03711258    -1.436   .1519    .05555556
 C10     |     .09007170       .03860659     2.333   .0203    .05555556
 C11     |    -.05106438       .03357607    -1.521   .1293    .05555556
 C12     |    -.06915517       .04040779    -1.711   .0880    .05555556
```

C13	-.60407878	.09122015	-6.622	.0000	.05555556
C14	.74048679	.18008419	4.112	.0000	.05555556
C15	.11664698	.03471246	3.360	.0009	.05555556
C16	.22413229	.04764432	4.704	.0000	.05555556
C17	.05959184	.03018816	1.974	.0492	.05555556
C18	.76939510	.04457642	17.260	.0000	.05555556

```
Calc ; r1 = rsqrd $
Calc ; list ; Fstat = ((r1 - r0)/17) / ((1-r1)/(n-4-17)) $
Calc ; list ; Fc    =ftb(.95,17,(n-4-17)) $
+------------------------------------+
| Listed Calculator Results          |
+------------------------------------+
 FSTAT    =     83.960798
 FC       =      1.654675
Plot ; lhs = country ; rhs = e ; Bars = 0
;Title=Plot of OLS Residuals by Country $
```




```
Regress;lhs=y;rhs=x,cntry ; Het $
+----------------------------------------------------+
| Ordinary    least squares regression               |
| LHS=Y        Mean                 =   4.296242     |
|              Standard deviation   =   .5489071     |
| WTS=none     Number of observs.   =        342     |
| Model size   Parameters           =         21     |
|              Degrees of freedom   =        321     |
| Residuals    Sum of squares       =   2.736491     |
|              Standard error of e  =   .9233035E-01 |
| Fit          R-squared            =   .9733657     |
|              Adjusted R-squared   =   .9717062     |
| Model test   F[ 20,    321] (prob) = 586.56 (.0000) |
| White heteroscedasticity robust covariance matrix  |
| Br./Pagan LM Chi-sq [ 20]  (prob) = 338.94 (.0000) |
+----------------------------------------------------+
```

Variable	Coefficient	Standard Error	t-ratio	P[|T|>t]	Mean of X
Constant	2.28585577	.22608070	10.111	.0000	
LINCOMEP	.66224966	.07277408	9.100	.0000	-6.13942544
LRPMG	-.32170246	.05381258	-5.978	.0000	-.52310321
LCARPCAP	-.64048288	.03876145	-16.524	.0000	-9.04180473
C2	-.12030455	.03160815	-3.806	.0002	.05555556
C3	.75598453	.03692877	20.471	.0000	.05555556
C4	.10360026	.03642008	2.845	.0047	.05555556
C5	-.08108439	.03252022	-2.493	.0132	.05555556
C6	-.13598740	.03504274	-3.881	.0001	.05555556
C7	.05125389	.05768530	.889	.3749	.05555556
C8	.30646950	.03516370	8.716	.0000	.05555556

```
 C9      |    -.05330785       .04078467    -1.307   .1921    .05555556
 C10     |     .09007170       .05606508     1.607   .1091    .05555556
 C11     |    -.05106438       .03228064    -1.582   .1147    .05555556
 C12     |    -.06915517       .03857838    -1.793   .0740    .05555556
 C13     |    -.60407878       .09798870    -6.165   .0000    .05555556
 C14     |     .74048679       .18836593     3.931   .0001    .05555556
 C15     |     .11664698       .03500336     3.332   .0010    .05555556
 C16     |     .22413229       .08147015     2.751   .0063    .05555556
 C17     |     .05959184       .03166823     1.882   .0608    .05555556
 C18     |     .76939510       .04121364    18.668   .0000    .05555556

Create ; e2 = e*e $
Regress ; Lhs = e2 ; Rhs = one,cntry $
Calc ; List ; White = n*rsqrd ;  ctb(.95,17) $
+------------------------------------+
| Listed Calculator Results          |
+------------------------------------+
 WHITE   =    131.209847
 Result  =     27.587112
Calc ; s2 = e'e/n $
Matrix ; s2g = {1/19} * cntry'e2
       ; s2g = 1/s2 * s2g
       ; g = s2g - 1
       ; List ; lmstat = {19/2}*g'g  $
Matrix LMSTAT   has  1 rows and  1 columns.
        +--------------
       1|  277.00947
Name   ; All = c1,cntry $
Matrix ; vg = 1/19*all'e2 $
Create ; wt = 1/vg(country) $
Regress ; Lhs = y ; rhs = x,cntry;wts=wt $
+-----------------------------------------------------+
| Ordinary    least squares regression                |
| LHS=Y        Mean                 =   4.460122      |
|              Standard deviation   =   .4535009      |
| WTS=WT       Number of observs.   =        342      |
| Model size   Parameters           =         21      |
|              Degrees of freedom   =        321      |
| Residuals    Sum of squares       =   .5901434      |
|              Standard error of e  =   .4287719E-01  |
| Fit          R-squared            =   .9915851      |
|              Adjusted R-squared   =   .9910608      |
| Model test   F[ 20,   321] (prob) =1891.29 (.0000)  |
+-----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|    2.43706653       .11308370    21.551   .0000
 LINCOMEP|     .57506962       .02926687    19.649   .0000  -5.84790214
 LRPMG   |    -.27967108       .03518536    -7.949   .0000   -.87736963
 LCARPCAP|    -.56540465       .01613491   -35.042   .0000  -8.34742189
 C2      |    -.12007208       .02789011    -4.305   .0000    .08866789
 C3      |     .76945446       .03011060    25.554   .0000    .34252221
 C4      |     .11000512       .03169158     3.471   .0006    .01995470
 C5      |    -.09845013       .02921659    -3.370   .0008    .05724878
 C6      |    -.13641007       .03387520    -4.027   .0001    .01079455
 C7      |     .13502296       .04413211     3.060   .0024    .00604952
 C8      |     .28669153       .03200056     8.959   .0000    .01577251
 C9      |    -.08901681       .03324265    -2.678   .0078    .01701683
 C10     |     .15281210       .05659004     2.700   .0073    .00228044
 C11     |    -.04087890       .02882321    -1.418   .1571    .03809105
 C12     |    -.05220341       .02952832    -1.768   .0780    .09438377
 C13     |    -.53400193       .06166458    -8.660   .0000    .01328985
 C14     |     .64117855       .10737812     5.971   .0000    .06594614
 C15     |     .12783552       .03189740     4.008   .0001    .02454617
 C16     |     .38638811       .05013313     7.707   .0000    .00712693
 C17     |     .04507072       .03121765     1.444   .1498    .01629698
```

```
 C18     |     .77812476       .03277077    23.744   .0000    .17152029


?========================================================
? Application 8.3  Iterative estimator
?========================================================
create ; logc = log(c) ; logq=log(q) ; logq2=logq^2 ; logp=log(pf) $
Name ; x = one,logq,logq2,logp $
Regress ; lhs = logc ; rhs = x ; Res = e $
Matrix ; b0=b $
Procedure$
Create ; e2 = e*e
; le = e2/(sumsqdev/n)-1 $ (MLE)
?le = log(e2) $            (Iterative two step)
Regress ; quiet ; lhs=le ; rhs=one,lf ; keep = s2i $
Create ; wi = 1/exp(s2i) $
Regress ; lhs = logc ; rhs = x ; wts=wi ; res=e $
Matrix ; db = b-b0 ; b0 = b $
Calc ; list ; db2 = db'db $
Endproc $
Exec ; n = 10 $
These are the two step estimators from Example 8.4
+-----------------------------------------------------+
| Ordinary    least squares regression                |
| LHS=LOGC     Mean                 =   12.92005      |
|              Standard deviation   =   1.192244      |
| WTS=WI       Number of observs.   =         90      |
| Model size   Parameters           =          4      |
|              Degrees of freedom   =         86      |
| Residuals    Sum of squares       =   1.212889      |
|              Standard error of e  =   .1187576      |
| Fit          R-squared            =   .9904126      |
|              Adjusted R-squared   =   .9900782      |
| Model test   F[  3,    86] (prob) =2961.37 (.0000) |
+-----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|    9.27731457       .20978736    44.222   .0000
 LOGQ    |     .91610564       .03299348    27.766   .0000  -1.56779393
 LOGQ2   |     .02164855       .01101812     1.965   .0527   3.87530677
 LOGP    |     .40174171       .01633292    24.597   .0000   12.4336185
These are the maximum likelihood estimates
+-----------------------------------------------------+
| Ordinary    least squares regression                |
| Residuals    Sum of squares       =   1.347926      |
|              Standard error of e  =   .1251941      |
| Fit          R-squared            =   .9892110      |
|              Adjusted R-squared   =   .9888346      |
| Model test   F[  3,    86] (prob) =2628.35 (.0000) |
+-----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|    9.24395222       .21962091    42.090   .0000
 LOGQ    |     .92163069       .03302261    27.909   .0000  -1.43646434
 LOGQ2   |     .02461767       .01143734     2.152   .0342   3.46800689
 LOGP    |     .40366011       .01701993    23.717   .0000   12.5455161
```

Chapter 9

Models for Panel Data

1. 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. 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. (a) The pooled OLS estimator will be $\mathbf{b} = \left[\Sigma_{i=1}^n \mathbf{X}_i'\mathbf{X}_i\right]^{-1}\left[\Sigma_{i=1}^n \mathbf{X}_i'\mathbf{y}_i\right]$ where X_i and y_i have T_i observations. It remains true that $\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i + u_i\mathbf{i}$, where Var[$\boldsymbol{\varepsilon}_i + u_i\mathbf{i}|X_i$] = Var[$\mathbf{w}_i|\mathbf{X}_i$] = $\sigma_\varepsilon^2\mathbf{I} + \sigma_u^2\mathbf{ii}'$ and, maintaining the assumptions, both ε_i and u_i are uncorrelated with X_i. Substituting the expression for y_i into that of b and collecting terms, we have

$$\mathbf{b} = \boldsymbol{\beta} + \left[\Sigma_{i=1}^n \mathbf{X}_i'\mathbf{X}_i\right]^{-1}\left[\Sigma_{i=1}^n \mathbf{X}_i'\mathbf{w}_i\right].$$

Unbiasedness follows immediately as long as E[$w_i|X_i$] equals zero, which it does by assumption. Consistency, as mentioned in Section 9.3.2, is covered in the discussion of Chapter 4. We would need for the matrix $\mathbf{Q}$ $= \left[\frac{1}{n}\Sigma_{i=1}^n \frac{1}{T_i}\mathbf{X}_i'\mathbf{X}_i\right]$ to converge to a matrix of constants, or not to degenerate to a matrix of zeros. The requirements for the large sample behavior of the vector in the second set of brackets is quite the same as in our earlier discussions of consistency. The vector $(1/n)\Sigma_{i=1}^n \mathbf{X}_i'\mathbf{w}_i = (1/n)\Sigma_{i=1}^n \mathbf{v}_i$ has mean zero. We would require the conditions of the Lindeberg-Feller version of the central theorem to apply, which could be expected.

(b) We seek to establish consistency, not unbiasedness. As such, we will ignore the degrees of freedom correction, -K, in (9-37). Use n(T-1) as the denominator. Thus, the question is whether

$$\text{plim}\frac{\Sigma_{i=1}^n\Sigma_{t=1}^T(e_{it} - \bar{e}_{i.})^2}{n(T-1)} = \sigma_\varepsilon^2$$

If so, then the estimator in (9-37) will be consistent. Using (9-33) and e_{it} - $\bar{e}_i = \bar{y}_i - \bar{\mathbf{x}}_i'\mathbf{b} - a_i$, it follows that $e_{it} - \bar{e}_i = \varepsilon_{it} - \bar{\varepsilon}_i - (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(\mathbf{b} - \boldsymbol{\beta})$. Summing the squares in (9-37), we find that the estimator in (9-37)

$$\frac{\Sigma_{i=1}^n\Sigma_{t=1}^T(e_{it} - \bar{e}_{i.})^2}{n(T-1)} = \frac{1}{n}\sum_{i=1}^n \hat{\sigma}^2(i) + (\mathbf{b}-\boldsymbol{\beta})'\left[\frac{1}{n}\sum_{i=1}^n\frac{1}{T}\sum_{t=1}^T(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'\right](\mathbf{b}-\boldsymbol{\beta})$$

$$- 2(\mathbf{b}-\boldsymbol{\beta})'\left[\frac{1}{n}\sum_{i=1}^n\frac{1}{T}\sum_{t=1}^T(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(\varepsilon_{it} - \bar{\varepsilon}_{i.})'\right]$$

The second term will converge to zero as the center matrix converges to a constant Q and the vectors converge to zero as b converges to β. (We use the Slutsky theorem.) The third term will converge to zero as both the leading vector converges to zero and the covariance vector between the regressors and the disturbances converges to zero. That leaves the first term, which is the average of the estimators in (9-34). The terms in the average are independent. Each has expected value exactly equal to σ_ε^2. So, if each estimator has finite variance, then the average will converge to its expectation. Appendix D discusses various different conditions underwhich a sample average will converge to its expectation. For example, finite fouth moment of ε_{it} would be sufficient here (though weaker conditions would also suffice). Note that this derivation follows through for any consistent estimator of β, not just for **b**.

4. 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$(1/n)\Sigma_i \bar{\varepsilon}_{i.}^2$ = Var[$\bar{\varepsilon}_{i.}$] = $\sigma_\varepsilon^2 T + \sigma_u^2$The denominator is simply the usual variance estimator, so plim$(1/(nT))\Sigma_i\Sigma_t\varepsilon_{it}^2$ = Var[ε_{it}] = $\sigma_\varepsilon^2 + \sigma_u^2$Therefore, inserting these results in the expression for LM, we find that plim $(1/n)$LM = $[T/(2(T-1))]\{[T(\sigma_\varepsilon^2 T + \sigma_u^2)]/[\sigma_\varepsilon^2 + \sigma_u^2] - 1\}^2$. 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 = $[T(T-1)/2][\sigma_u^2/(\sigma_\varepsilon^2+\sigma_u^2)]^2$. Within group i, Corr$^2[\varepsilon_{it},\varepsilon_{is}]$ = ρ^2 = $\sigma_u^2/(\sigma_u^2+\sigma_\varepsilon^2)$ so plim $(1/n)$LM = $[T(T-1)/2](\rho^2)^2$. 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. □

5. The ordinary least squares regression results are

R^2 = .92803, **e′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. □

6. The covariance matrix would be

	$i=1,t=1$	$i=1,t=2$	$i=2,t=1$	$i=2,t=2$
$i=1,t=1$	$\sigma_\varepsilon^2+\sigma_u^2+\sigma_v^2$	σ_u^2	σ_v^2	0
$i=1,t=2$	σ_u^2	$\sigma_\varepsilon^2+\sigma_u^2+\sigma_v^2$	0	σ_v^2
$i=2,t=1$	σ_v^2	0	$\sigma_\varepsilon^2+\sigma_u^2+\sigma_v^2$	σ_u^2
$i=2,t=2$	0	σ_v^2	σ_u^2	$\sigma_\varepsilon^2+\sigma_u^2+\sigma_v^2$

7. 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 $\beta = 1$, 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. □

8. If all of the regressor matrices are the same, the estimator in (8-35) reduces to

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

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. □

9. 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)	
$\mathbf{e}'\mathbf{e}$	452.206	673.409	125.281	
	(464.288)	(732.560)	(171.240)	(1368.088)

(Values of $\mathbf{e}'\mathbf{e}$ 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.

Applications

As usual, the applications below require econometric software. The computations can be done with any modern software package, so no specific program is recommended.

```
--> read $
Last observation read from data file was     200
End of data listing in edit window was reached
--> REGRESS ; Lhs = I ; Rhs = F,C,one $
+----------------------------------------------------+
| Ordinary    least squares regression              |
| LHS=I        Mean                    =   145.9582   |
|              Standard deviation      =   216.8753   |
| WTS=none     Number of observs.      =        200   |
| Model size   Parameters              =          3   |
|              Degrees of freedom      =        197   |
| Residuals    Sum of squares          =   1755850.   |
|              Standard error of e     =   94.40840   |
| Fit          R-squared               =   .8124080   |
|              Adjusted R-squared      =   .8105035   |
| Model test   F[  2,   197] (prob) = 426.58 (.0000) |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 F       |     .11556216       .00583571    19.803   .0000   1081.68110
 C       |     .23067849       .02547580     9.055   .0000   276.017150
 Constant|   -42.7143694      9.51167603    -4.491   .0000

--> CALC    ; R0=Rsqrd $
--> REGRESS ; Lhs = I ; Rhs = F,C,one ; Cluster = 20 $


+----------------------------------------------------+
| Ordinary    least squares regression              |
| LHS=I        Mean                    =   145.9582   |
|              Standard deviation      =   216.8753   |
| WTS=none     Number of observs.      =        200   |
| Model size   Parameters              =          3   |
|              Degrees of freedom      =        197   |
| Residuals    Sum of squares          =   1755850.   |
|              Standard error of e     =   94.40840   |
| Fit          R-squared               =   .8124080   |
|              Adjusted R-squared      =   .8105035   |
| Model test   F[  2,   197] (prob) = 426.58 (.0000) |
+----------------------------------------------------+
+---------------------------------------------------------------------+
| Covariance matrix for the model is adjusted for data clustering.    |
| Sample of    200 observations contained     10 clusters defined by  |
|     20 observations (fixed number) in each cluster.                 |
| Sample of    200 observations contained      1 strata defined by    |
|    200 observations (fixed number) in each stratum.                 |
+---------------------------------------------------------------------+
```

```
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 F       |     .11556216       .01589434     7.271   .0000   1081.68110
 C       |     .23067849       .08496711     2.715   .0072   276.017150
 Constant|   -42.7143694      20.4252029    -2.091   .0378
```

The standard errors increase substantially. This is at least suggestive that there is correlation across observations within the groups. A formal test would be based on one of the panel models below. When the random effects model is fit by maximum likelihood, for example, the log likelihood function is -1095.257. The log likelihood function for the pooled model is -1191.802. Thus, the correlation is highly significant. The Lagrange multiplier statistic reported below is 798.16, which is far larger than the critical value of 3.84. Once again, these results do suggest within groups correlation.

```
--> REGRESS ; Lhs = I ; Rhs = F,C,one  ; Panel ; Pds=20 ; Fixed $
+----------------------------------------------------+
| Least Squares with Group Dummy Variables           |
| Ordinary    least squares regression               |
| LHS=I        Mean                 =   145.9583     |
|              Standard deviation   =   216.8753     |
| WTS=none     Number of observs.   =        200     |
| Model size   Parameters           =         12     |
|              Degrees of freedom   =        188     |
| Residuals    Sum of squares       =   523478.1     |
|              Standard error of e  =   52.76797     |
| Fit          R-squared            =   .9440725     |
|              Adjusted R-squared   =   .9408002     |
| Model test   F[ 11,   188] (prob) = 288.50 (.0000) |
+----------------------------------------------------+
+----------------------------------------------------+
| Panel:Groups   Empty       0,   Valid data      10 |
|                Smallest   20,   Largest         20 |
|                Average group size             20.00 |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 F       |     .11012380       .01185669     9.288   .0000   1081.68110
 C       |     .31006534       .01735450    17.867   .0000   276.017150
+--------------------------------------------------------------------+
|             Test Statistics for the Classical Model                |
+--------------------------------------------------------------------+
|       Model            Log-Likelihood  Sum of Squares    R-squared |
|(1)  Constant term only    -1359.15096  .9359943929D+07    .0000000 |
|(2)  Group effects only    -1216.34872  .2244352274D+07    .7602173 |
|(3)  X - variables only    -1191.80236  .1755850484D+07    .8124080 |
|(4)  X and group effects   -1070.78103  .5234781474D+06    .9440725 |
+--------------------------------------------------------------------+
|                        Hypothesis Tests                            |
|         Likelihood Ratio Test           F Tests                    |
|         Chi-squared   d.f.  Prob.       F    num. denom.   P value |
|(2) vs (1)   285.604      9  .00000   66.932     9     190   .00000 |
|(3) vs (1)   334.697      2  .00000  426.576     2     197   .00000 |
|(4) vs (1)   576.740     11  .00000  288.500    11     188   .00000 |
|(4) vs (2)   291.135      2  .00000  309.014     2     188   .00000 |
|(4) vs (3)   242.043      9  .00000   49.177     9     188   .00000 |
+--------------------------------------------------------------------+
--> CALC    ; R1 = Rsqrd $
--> MATRIX  ; bf = b(1:2) ; vf = varb(1:2,1:2) $
--> CALC    ; List ; Fstat=((R1-R0)/9)/((1-R1)/(n-2-10))
    ; FC=Ftb(.95,9,(n-2-10)) $
+-----------------------------------+
| Listed Calculator Results         |
+-----------------------------------+
 FSTAT   =     49.176625
```

```
FC      =      1.929957
```

The F statistic of 49.18 is far larger than the critical value, so the hypothesis of equal constant terms is rejected.

```
--> REGRESS ; Lhs = I ; Rhs = F,C,one
    ; Panel ; Pds=20 ; Random $
+--------------------------------------------------+
| Random Effects Model: v(i,t) = e(i,t) + u(i)     |
| Estimates:  Var[e]              =   .278446D+04  |
|             Var[u]              =   .612849D+04  |
|             Corr[v(i,t),v(i,s)] =   .687594      |
| Lagrange Multiplier Test vs. Model (3) =  798.16 |
| ( 1 df, prob value =  .000000)                   |
| (High values of LM favor FEM/REM over CR model.) |
|             Sum of Squares          .184029D+07  |
|             R-squared               .803387D+00  |
+--------------------------------------------------+
+---------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+---------+--------------+----------------+--------+--------+----------+
 F       |     .10974919       .01031952    10.635   .0000   1081.68110
 C       |     .30780890       .01715154    17.946   .0000   276.017150
 Constant|   -57.7159079      27.1118671    -2.129   .0333
```

The LM statistic, as noted earlier, is very large, so the hypothesis of no effects is rejected.

```
--> MATRIX  ; br = b(1:2) ; vr = varb(1:2,1:2) $
--> MATRIX  ; db = bf-br ; vdb = vf-vr ; List ; Hausman=db'<vdb>db $
               1
        +--------------
       1|    2.45500
--> CALC    ; List ; Ctb(.95,2) $
+-------------------------------------+
| Listed Calculator Results           |
+-------------------------------------+
 Result  =      5.991465
```

The Hausman statistic is quite small, which suggests that the random effects approach is consistent with the data.

2.

```
create ; logc=log(cost/pfuel)
       ; logp1=log(pmtl/pfuel)
       ; logp2=log(peqpt/pfuel)
       ; logp3=log(plabor/pfuel)
       ; logp4=log(pprop/pfuel)
       ; logp5=log(kprice/pfuel)
       ; logq=log(output)
       ; logq2=.5*logq^2 $
Namelist ; cd = logp1,logp2,logp3,logp4,logp5 $
create
       ; p11=.5* logp1^2
       ; p22=.5* logp2^2
       ; p33=.5* logp3^2
       ; p44=.5* logp4^2
       ; p55=.5* logp5^2
       ; p12=logp1*logp2
       ; p13=logp1*logp3
       ; p14=logp1*logp4
       ; p15=logp1*logp5
       ; p23=logp2*logp3
       ; p24=logp2*logp4
       ; p25=logp2*logp5
       ; p34=logp3*logp4
       ; p35=logp3*logp5
       ; p45=logp4*logp5 $
Namelist ; tl = p11,p12,p13,p14,p15,p22,p23,p24,p25,p33,p34,p35,p44,p45,p55$
Namelist ; z = loadfctr,stage,points $
regress;lhs=logc;rhs=one,logq,logq2,cd,z $
+---------------------------------------------------+
| Ordinary    least squares regression              |
| LHS=LOGC     Mean                 =   .7723984     |
|              Standard deviation   =   1.074424     |
| WTS=none     Number of observs.   =        256     |
| Model size   Parameters           =         11     |
|              Degrees of freedom   =        245     |
| Residuals    Sum of squares       =   2.965806     |
|              Standard error of e  =   .1100242     |
| Fit          R-squared            =   .9899249     |
|              Adjusted R-squared   =   .9895136     |
| Model test   F[ 10,   245] (prob) =2407.23 (.0000) |
+---------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|    20.3856176      22.8643711      .892   .3735
 LOGQ    |     .95227889       .01832119    51.977   .0000  -1.11237037
 LOGQ2   |     .06568531       .01060839     6.192   .0000   1.45687077
 LOGP1   |    -.32662031      1.17956412     -.277   .7821    .37999226
 LOGP2   |    -.28619766       .56614750     -.506   .6136   -.25308254
 LOGP3   |     .16012937       .08634095     1.855   .0649    .66688211
 LOGP4   |    -.00519153       .07328859     -.071   .9436  -2.14504306
 LOGP5   |    1.43718160      1.78896723      .803   .4225  -12.6860637
 LOADFCTR|    -.94688632       .18441822    -5.134   .0000    .54786115
 STAGE   |    -.00021794     .402227D-04    -5.418   .0000   507.879666
 POINTS  |     .00199712       .00031682     6.304   .0000   72.9843750
?
? Turns out the translog model cannot be computed with the firm
? dummy variables.  I'll use the Cobb Douglas form.
?
regress;lhs=logc;rhs=  one,logq,logq2,cd ; panel ; pds=ti $
+---------------------------------------------------+
| OLS Without Group Dummy Variables                 |
| Ordinary    least squares regression              |
| LHS=LOGC     Mean                 =   .7723984     |
|              Standard deviation   =   1.074424     |
| WTS=none     Number of observs.   =        256     |
```

```
| Model size   Parameters           =          8     |
|              Degrees of freedom   =        248     |
| Residuals    Sum of squares       =   4.190133     |
|              Standard error of e  =   .1299834     |
| Fit          R-squared            =   .9857657     |
|              Adjusted R-squared   =   .9853639     |
| Model test   F[  7,   248] (prob) =2453.53 (.0000) |
+----------------------------------------------------+
+----------------------------------------------------+
| Panel Data Analysis of LOGC       [ONE way]        |
|           Unconditional ANOVA (No regressors)      |
| Source      Variation   Deg. Free.     Mean Square |
| Between       272.013          24.     11.3339     |
| Residual      22.3551         231.     .967752E-01 |
| Total         294.368         255.     1.15439     |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 LOGQ    |      .93708702       .01772733    52.861   .0000  -1.11237037
 LOGQ2   |      .07754607       .01211431     6.401   .0000   1.45687077
 LOGP1   |    -.94586281      1.38855410      -.681   .4964    .37999226
 LOGP2   |    -.79081045       .66530892    -1.189   .2357   -.25308254
 LOGP3   |      .01998606       .09963618      .201   .8412    .66688211
 LOGP4   |      .08893118       .08543313     1.041   .2989  -2.14504306
 LOGP5   |     2.63118115      2.10504302     1.250   .2125  -12.6860637
 Constant|    35.4178566      26.9017806      1.317   .1892
+----------------------------------------------------+
| Least Squares with Group Dummy Variables           |
| Ordinary    least squares regression               |
| LHS=LOGC     Mean                 =   .7723984     |
|              Standard deviation   =   1.074424     |
| WTS=none     Number of observs.   =        256     |
| Model size   Parameters           =         32     |
|              Degrees of freedom   =        224     |
| Residuals    Sum of squares       =   .9373686     |
|              Standard error of e  =   .6468911E-01 |
| Fit          R-squared            =   .9968157     |
|              Adjusted R-squared   =   .9963750     |
| Model test   F[ 31,   224] (prob) =2261.94 (.0000) |
+----------------------------------------------------+
+----------------------------------------------------+
| Panel:Groups   Empty       0,   Valid data      25 |
|                Smallest    2,   Largest         15 |
|                Average group size           10.24 |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 LOGQ    |      .66448665       .03580894    18.556   .0000  -1.11237037
 LOGQ2   |    -.00955723       .01280811     -.746   .4563   1.45687077
 LOGP1   |     1.84750938       .76113884     2.427   .0159    .37999226
 LOGP2   |      .73986763       .37612716     1.967   .0503   -.25308254
 LOGP3   |    -.05323942       .06396335     -.832   .4060    .66688211
 LOGP4   |      .22763995       .04625120     4.922   .0000  -2.14504306
 LOGP5   |    -1.83738098     1.16995945    -1.570   .1176  -12.6860637
+--------------------------------------------------------------------+
|             Test Statistics for the Classical Model               |
+--------------------------------------------------------------------+
|       Model            Log-Likelihood  Sum of Squares    R-squared |
|(1)  Constant term only     -381.12407  .2943684435D+03    .0000000 |
|(2)  Group effects only      -51.16832  .2235506489D+02    .9240575 |
|(3)  X - variables only      163.14470  .4190132631D+01    .9857657 |
|(4)  X and group effects     354.81332  .9373685874D+00    .9968157 |
+--------------------------------------------------------------------+
|                       Hypothesis Tests                             |
|         Likelihood Ratio Test           F Tests                    |
```

```
|          Chi-squared   d.f.  Prob.       F    num. denom.   P value |
|(2) vs (1)   659.911     24  .00000  117.116    24     231   .00000 |
|(3) vs (1)  1088.538      7  .00000 2453.527     7     248   .00000 |
|(4) vs (1)  1471.875     31  .00000 2261.945    31     224   .00000 |
|(4) vs (2)   811.963      7  .00000  731.160     7     224   .00000 |
|(4) vs (3)   383.337     24  .00000   32.388    24     224   .00000 |
+--------------------------------------------------------------------+
+--------------------------------------------------+
| Random Effects Model: v(i,t) = e(i,t) + u(i)     |
| Estimates:  Var[e]              =   .418468D-02  |
|             Var[u]              =   .127110D-01  |
|             Corr[v(i,t),v(i,s)] =   .752323      |
| Lagrange Multiplier Test vs. Model (3) =  479.37 |
| ( 1 df, prob value =  .000000)                   |
| (High values of LM favor FEM/REM over CR model.) |
| Baltagi-Li form of LM Statistic =         174.85 |
| Fixed vs. Random Effects (Hausman)     =   40.99 |
| ( 7 df, prob value =  .000001)                   |
| (High (low) values of H favor FEM (REM).)        |
|             Sum of Squares          .648771D+01  |
|             R-squared               .978056D+00  |
+--------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 LOGQ    |     .79769706       .02494671    31.976   .0000  -1.11237037
 LOGQ2   |     .02011534       .01130089     1.780   .0751   1.45687077
 LOGP1   |    1.11671466       .74579390     1.497   .1343    .37999226
 LOGP2   |     .27128619       .36294718      .747   .4548   -.25308254
 LOGP3   |    -.10761385       .06138583    -1.753   .0796    .66688211
 LOGP4   |     .18385724       .04550246     4.041   .0001  -2.14504306
 LOGP5   |    -.49374865      1.13625272     -.435   .6639  -12.6860637
 Constant|   -4.53328730      14.5229534     -.312   .7549
regress;lhs=logc;rhs=z,one,logq,logq2,cd ; panel ; pds=ti $
+----------------------------------------------------+
| OLS Without Group Dummy Variables                  |
| Ordinary    least squares regression               |
| LHS=LOGC     Mean                 =   .7723984     |
|              Standard deviation   =   1.074424     |
| WTS=none     Number of observs.   =        256     |
| Model size   Parameters           =         11     |
|              Degrees of freedom   =        245     |
| Residuals    Sum of squares       =   2.965806     |
|              Standard error of e  =   .1100242     |
| Fit          R-squared            =   .9899249     |
|              Adjusted R-squared   =   .9895136     |
| Model test   F[ 10,   245] (prob) =2407.23 (.0000) |
+----------------------------------------------------+
+----------------------------------------------------+
| Panel Data Analysis of LOGC       [ONE way]        |
|           Unconditional ANOVA (No regressors)      |
| Source      Variation   Deg. Free.     Mean Square |
| Between       272.013          24.     11.3339     |
| Residual      22.3551         231.     .967752E-01 |
| Total         294.368         255.     1.15439     |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 LOADFCTR|    -.94688632       .18441823    -5.134   .0000    .54786115
 STAGE   |    -.00021794     .402227D-04    -5.418   .0000   507.879666
 POINTS  |     .00199712       .00031682     6.304   .0000   72.9843750
 LOGQ    |     .95227889       .01832119    51.977   .0000  -1.11237037
 LOGQ2   |     .06568531       .01060839     6.192   .0000   1.45687077
 LOGP1   |    -.32662033      1.17956418     -.277   .7821    .37999226
 LOGP2   |    -.28619767       .56614753     -.506   .6136   -.25308254
 LOGP3   |     .16012937       .08634095     1.855   .0649    .66688211
```

```
 LOGP4   |    -.00519153       .07328859     -.071   .9436  -2.14504306
 LOGP5   |    1.43718164      1.78896732      .803   .4225  -12.6860637
 Constant|    20.3856181      22.8643723      .892   .3735
+----------------------------------------------------+
| Least Squares with Group Dummy Variables           |
| Ordinary    least squares regression               |
| LHS=LOGC     Mean                 =   .7723984     |
|              Standard deviation   =   1.074424     |
| WTS=none     Number of observs.   =        256     |
| Model size   Parameters           =         35     |
|              Degrees of freedom   =        221     |
| Residuals    Sum of squares       =   .7726037     |
|              Standard error of e  =   .5912651E-01 |
| Fit          R-squared            =   .9973754     |
|              Adjusted R-squared   =   .9969716     |
| Model test   F[ 34,   221] (prob) =2470.05 (.0000) |
+----------------------------------------------------+
+----------------------------------------------------+
| Panel:Groups   Empty       0,   Valid data      25 |
|                Smallest    2,   Largest         15 |
|                Average group size            10.24 |
+----------------------------------------------------+
+--------+-------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+-------------+----------------+--------+--------+----------+
 LOADFCTR|    -.89457348       .14242570    -6.281   .0000    .54786115
 STAGE   |    -.00022827      .894260D-04   -2.553   .0113   507.879666
 POINTS  |     .00010341       .00041551      .249   .8037   72.9843750
 LOGQ    |     .75278467       .03923479    19.187   .0000  -1.11237037
 LOGQ2   |    -.00324835       .01306645     -.249   .8039   1.45687077
 LOGP1   |    1.38217070       .72421015     1.909   .0575    .37999226
 LOGP2   |     .61609241       .35323609     1.744   .0824   -.25308254
 LOGP3   |     .00706546       .05918620      .119   .9051    .66688211
 LOGP4   |     .14433953       .04404683     3.277   .0012  -2.14504306
 LOGP5   |   -1.25331458      1.10477945    -1.134   .2577  -12.6860637
+--------------------------------------------------------------------+
|             Test Statistics for the Classical Model                |
+--------------------------------------------------------------------+
|       Model              Log-Likelihood  Sum of Squares    R-squared |
|(1)  Constant term only     -381.12407  .2943684435D+03    .0000000 |
|(2)  Group effects only      -51.16832  .2235506489D+02    .9240575 |
|(3)  X - variables only      207.37940  .2965806000D+01    .9899249 |
|(4)  X and group effects     379.55705  .7726036853D+00    .9973754 |
+--------------------------------------------------------------------+
|                          Hypothesis Tests                          |
|         Likelihood Ratio Test          F Tests                     |
|         Chi-squared   d.f.  Prob.       F    num. denom.   P value |
|(2) vs (1)   659.911     24  .00000  117.116    24     231   .00000 |
|(3) vs (1)  1177.007     10  .00000 2407.226    10     245   .00000 |
|(4) vs (1)  1521.362     34  .00000 2470.054    34     221   .00000 |
|(4) vs (2)   861.451     10  .00000  617.357    10     221   .00000 |
|(4) vs (3)   344.355     24  .00000   26.140    24     221   .00000 |
+--------------------------------------------------------------------+

+--------------------------------------------------+
| Random Effects Model: v(i,t) = e(i,t) + u(i)     |
| Estimates:  Var[e]              =   .349594D-02  |
|             Var[u]              =   .860939D-02  |
|             Corr[v(i,t),v(i,s)] =   .711206      |
| Lagrange Multiplier Test vs. Model (3) =  466.36 |
| ( 1 df, prob value =  .000000)                   |
| (High values of LM favor FEM/REM over CR model.) |
| Baltagi-Li form of LM Statistic =         170.10 |
| Fixed vs. Random Effects (Hausman)     =   44.65 |
| (10 df, prob value =  .000003)                   |
| (High (low) values of H favor FEM (REM).)        |
|             Sum of Squares          .451094D+01  |
```

```
|              R-squared               .984812D+00  |
+---------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 LOADFCTR|   -1.07921018       .13264921    -8.136   .0000    .54786115
 STAGE   |    -.00016415     .672354D-04    -2.441   .0146   507.879666
 POINTS  |     .00044792       .00035950     1.246   .2128   72.9843750
 LOGQ    |     .86611837       .02783747    31.113   .0000  -1.11237037
 LOGQ2   |     .02222380       .01102947     2.015   .0439   1.45687077
 LOGP1   |     .92719911       .70150544     1.322   .1863    .37999226
 LOGP2   |     .30782803       .33937387      .907   .3644   -.25308254
 LOGP3   |    -.02581955       .05671735     -.455   .6489    .66688211
 LOGP4   |     .09284095       .04277517     2.170   .0300  -2.14504306
 LOGP5   |    -.36595849      1.06514141     -.344   .7312  -12.6860637
 Constant|   -2.36774378      13.6315073     -.174   .8621
matrix ; List ; bz=b(1:3);vz=varb(1:3,1:3) ; wald = bz'<vz>bz $
Matrix WALD     has  1 rows and  1 columns.
               1
        +--------------
       1|   74.33957
```

Chapter 10

Systems of Regression Equations

1. 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

$$\begin{aligned}\text{Est.Var}[\bar{y}_1] &= s_{11}/n = ((150 - 100(1)^2)/99)/100 = .0051\\ \text{Est.Var}[\bar{y}_2] &= s_{22}/n = ((550 - 100(2)^2)/99)/100 = .0152\\ \text{Est.Cov}[\bar{y}_1, \bar{y}_2] &= s_{12}/n = ((260 - 100(1)(2))/99)/100 = .0061\end{aligned}$$

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

The generalized least squares estimator is

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

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

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 $\sigma_{ji}/(\sigma_{11}\sigma_{22} - \sigma_{12}^2)$ 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

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

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

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

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

$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Omega\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix}n & 0\\ 0 & ns_{xx}\end{bmatrix}^{-1}\begin{bmatrix}\sigma_{11}n & \sigma_{12}n\bar{x}\\ \sigma_{12}n\bar{x} & \sigma_{22}ns_{xx}\end{bmatrix}\begin{bmatrix}n & 0\\ 0 & ns_{xx}\end{bmatrix}^{-1}$. 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 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, $\text{Est.Var}\left[\hat{\boldsymbol{\beta}}\right] = \begin{bmatrix} .132565 & .0077645 \\ .0077645 & .0252505 \end{bmatrix}$. 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

$[\mathbf{X}\text{N}(\mathbf{S}^{-1}\otimes\mathbf{I})\mathbf{X}]^{-1} = \begin{bmatrix} .155355 & .00576887 \\ .00576887 & .029348 \end{bmatrix}$ 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 $\mathbf{\Sigma}$ 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. 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 $[\mathbf{\Sigma}^{-1} \otimes \mathbf{X'X}]^{-1}$. But, $[\mathbf{\Sigma}^{-1} \otimes \mathbf{X'X}]^{-1} = \mathbf{\Sigma} \otimes (\mathbf{X'X})^{-1}$. (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 $\mathbf{X'y}_1 = (\mathbf{X'X})\mathbf{b}_1$ and $\mathbf{X'y}_2 = (\mathbf{X'X})\mathbf{b}_2$. 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 $\mathbf{\Sigma\Sigma}^{-1} = \mathbf{I}$. Therefore, $\hat{\boldsymbol{\beta}} = \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{pmatrix}$.

5. 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 $plim\,(1/n)\mathbf{x}_1'\mathbf{y}_1 = \beta_1 q_{11} + plim\,(1/n)\mathbf{x}_1\text{N}\varepsilon_1 = \beta_1 q_{11}$

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

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

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

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 $\Delta = (\sigma_{11}\text{F}_{22}q_{11}q_{22} - (\text{F}_{12}q_{12})^2)$

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

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

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. 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

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

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. 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
$$e_{i1} = y_{i1} - \bar{y}_1 - bx_i$$
$$e_{i2} = y_{i2} - \bar{y}_2$$
$$e_{i3} = y_{i3} - \bar{y}_3$$

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.

Applications

1. 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

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

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

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

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 **g** is the gradient and **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
           βk       -.0546555     .2422
           βl        .250976      .2138
           δkk       .245259      .06904
           δll       .0245770     .04788
           δkl      -.00403448    .04779
           βy        .572452      .1340
           βyy       .0456587     .01908
           γyk      -.00124236    .008409
           γyl      -.0116921     .004442
           βf        .8036795
           δkf      -.2412245
           δlf      -.0205425
           δff       .261767
           γyf       .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.

```
?==========================================
? Application 10.2
?==========================================
? a.  Separate regressions and aggregation test.
?     This saves the residuals to be used later.
CALC    ; SS1=0 $
MATRIX  ; EOLS = Init(20,10,0) $
PROCEDURE $
Include ; new ; Firm = company $
REGRESS ; Lhs = I ; Rhs = F,C,one ; Res = e$
CALC    ; SS1=SS1 + Sumsqdev $
MATRIX  ; EOLS(*,company) = e $
ENDPROC $
EXECUTE ; Company=1,10 $
SAMPLE  ; 1-200 $
+----------------------------------------------------+
| Residuals    Sum of squares       =   143205.9     |
|              Standard error of e  =   91.78167     |
| Fit          R-squared            =   .9213540     |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 F       |     .11928083       .02583417     4.617   .0002   4333.84500
 C       |     .37144481       .03707282    10.019   .0000   648.435000
 Constant|   -149.782453      105.842125    -1.415   .1751
+----------------------------------------------------+
| Residuals    Sum of squares       =   158093.3     |
|              Standard error of e  =   96.43445     |
| Fit          R-squared            =   .4708624     |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 F       |     .17485602       .07419805     2.357   .0307   1971.82500
 C       |     .38964189       .14236688     2.737   .0140   294.855000
 Constant|   -49.1983219      148.075365     -.332   .7438
+----------------------------------------------------+
| Residuals    Sum of squares       =   13216.59     |
|              Standard error of e  =   27.88272     |
| Fit          R-squared            =   .7053067     |
|              Adjusted R-squared   =   .6706369     |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 F       |     .02655119       .01556610     1.706   .1063   1941.32500
 C       |     .15169387       .02570408     5.902   .0000   400.160000
 Constant|   -9.95630645      31.3742491     -.317   .7548
+----------------------------------------------------+
| Residuals    Sum of squares       =   2997.444     |
|              Standard error of e  =   13.27856     |
| Fit          R-squared            =   .9135784     |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 F       |     .07794782       .01997330     3.903   .0011   693.210000
 C       |     .31571819       .02881317    10.957   .0000   121.245000
 Constant|   -6.18996051      13.5064781     -.458   .6525
+----------------------------------------------------+
| Residuals    Sum of squares       =   1396.836     |
|              Standard error of e  =   9.064592     |
| Fit          R-squared            =   .6804076     |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
```

```
+--------+--------------+----------------+--------+--------+----------+
 F       |      .16237770       .05703645     2.847   .0111   231.470000
 C       |      .00310174       .02196531      .141   .8894   486.765000
 Constant|    22.7071160      6.87207605     3.304   .0042
+---------------------------------------------------+
| Residuals    Sum of squares       =   1110.533     |
|              Standard error of e  =   8.082418     |
| Fit          R-squared            =   .9521422     |
+---------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 F       |      .13145484       .03117234     4.217   .0006   419.865000
 C       |      .08537427       .10030597      .851   .4065   104.285000
 Constant|   -8.68554338      4.54516804    -1.911   .0730
+---------------------------------------------------+
| Residuals    Sum of squares       =   1507.403     |
|              Standard error of e  =   9.416516     |
| Fit          R-squared            =   .7635009     |
+---------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 F       |      .08752720       .06562593     1.334   .1999   149.790000
 C       |      .12378141       .01706483     7.254   .0000   314.945000
 Constant|   -4.49953436      11.2893942     -.399   .6952
+---------------------------------------------------+
| Residuals    Sum of squares       =   1773.234     |
|              Standard error of e  =   10.21312     |
| Fit          R-squared            =   .7444461     |
+---------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 F       |      .05289413       .01570650     3.368   .0037   670.910000
 C       |      .09240649       .05609897     1.647   .1179   85.6400000
 Constant|    -.50939018      8.01528894     -.064   .9501
+---------------------------------------------------+
| Residuals    Sum of squares       =   1407.360     |
|              Standard error of e  =   9.098674     |
| Fit          R-squared            =   .6655145     |
+---------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 F       |      .07538794       .03395227     2.220   .0403   333.650000
 C       |      .08210356       .02799168     2.933   .0093   297.900000
 Constant|   -7.72283708      9.35933952     -.825   .4207
+---------------------------------------------------+
| Residuals    Sum of squares       =   20.02673     |
|              Standard error of e  =   1.085377     |
| Fit          R-squared            =   .6431578     |
+---------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 F       |      .00457343       .02716079      .168   .8683   70.9210000
 C       |      .43736919       .07958891     5.495   .0000   5.94150000
 Constant|      .16151857      2.06556414      .078   .9386
```

```
+----------------------------------------------------+
| Ordinary    least squares regression               |
| LHS=I        Mean                 =   145.9582     |
|              Standard deviation   =   216.8753     |
| WTS=none     Number of observs.   =        200     |
| Model size   Parameters           =          3     |
|              Degrees of freedom   =        197     |
| Residuals    Sum of squares       =   1755850.     |
|              Standard error of e  =   94.40840     |
| Fit          R-squared            =   .8124080     |
|              Adjusted R-squared   =   .8105035     |
| Model test   F[  2,   197] (prob) = 426.58 (.0000) |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 F       |     .11556216       .00583571    19.803   .0000   1081.68110
 C       |     .23067849       .02547580     9.055   .0000   276.017150
 Constant|   -42.7143694      9.51167603    -4.491   .0000
? b.  Aggregation test
REGRESS ; LHS = I ; RHS = F,C,one $
CALC    ; SS0=Sumsqdev $
CALC    ; List ; Fstat = ((SS0 - SS1)/(9*3)) / (SS0/(n-10*3))
               ; FC    = Ftb(.95,27,170) $
+------------------------------------+
| Listed Calculator Results          |
+------------------------------------+
 FSTAT   =      5.131854
 FC      =      1.551534
? c.  SUR model
NAMELIST ; X1=F1,C1,one $
NAMELIST ; X2=F2,C2,one $
NAMELIST ; X3=F3,C3,one $
NAMELIST ; X4=F4,C4,one $
NAMELIST ; X5=F5,C5,one $
NAMELIST ; X6=F6,C6,one $
NAMELIST ; X7=F7,C7,one $
NAMELIST ; X8=F8,C8,one $
NAMELIST ; X9=F9,C9,one $
NAMELIST ; X10=F10,C10,one $
NAMELIST ; Y=I1,I2,I3,I4,I5,I6,I7,I8,I9,I10 $
SAMPLE ; 1 - 20 $
SURE ; Lhs = Y ; Eq1=X1;Eq2=X2;Eq3=X3;Eq4=X4;Eq5=X6;Eq6=X6
               ; Eq7=X7;Eq8=X8;Eq9=X9;Eq10=X10
     ; Maxit=0 ; OLS $
Criterion function for GLS is log-likelihood.
Iteration    0, GLS           =   -737.6463
Iteration    1, GLS           =   -730.1070
+----------------------------------------------------+
| Estimates for equation: I1                         |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 F1      |     .12472490       .01490044     8.371   .0000   4333.84500
 C1      |     .37951869       .02912686    13.030   .0000   648.435000
 Constant|   -178.611571      65.7890483    -2.715   .0066
+----------------------------------------------------+
| Estimates for equation: I2                         |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 F2      |     .16828512       .04057787     4.147   .0000   1971.82500
 C2      |     .33587688       .10299836     3.261   .0011   294.855000
 Constant|   -20.3887867      83.2537952     -.245   .8065
```

Estimates for equation: I3

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
F3	.03425481	.00925706	3.700	.0002	1941.32500
C3	.12538119	.02040101	6.146	.0000	400.160000
Constant	-14.3822597	20.6146424	-.698	.4854	

Estimates for equation: I4

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
F4	.06760969	.01597735	4.232	.0000	693.210000
C4	.30752805	.02536245	12.125	.0000	121.245000
Constant	1.96954637	11.0026359	.179	.8579	

Estimates for equation: I5

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
F6	.00635232	.02903793	.219	.8268	419.865000
C6	.12737505	.09456013	1.347	.1780	104.285000
Constant	45.8520779	4.86959707	9.416	.0000	

Estimates for equation: I6

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
F6	.12891587	.01798607	7.168	.0000	419.865000
C6	.06768693	.06029084	1.123	.2616	104.285000
Constant	-5.77499083	3.44886478	-1.674	.0940	

Estimates for equation: I7

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
F7	.09106397	.04535783	2.008	.0447	149.790000
C7	.12913287	.01446995	8.924	.0000	314.945000
Constant	-6.71472214	8.72476796	-.770	.4415	

Estimates for equation: I8

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
F8	.05179274	.00835658	6.198	.0000	670.910000
C8	.04729955	.03473521	1.362	.1733	85.6400000
Constant	4.09249729	5.09237714	.804	.4216	

Estimates for equation: I9

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
F9	.07275469	.02111017	3.446	.0006	333.650000
C9	.06640816	.02194422	3.026	.0025	297.900000
Constant	-2.16859331	7.30885683	-.297	.7667	

Estimates for equation: I10

```
+--------+-------------+---------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+-------------+---------------+--------+--------+----------+
 F10     |    -.01695668       .01550963    -1.093   .2743   70.9210000
 C10     |     .37466423       .05739586     6.528   .0000   5.94150000
 Constant|    2.06101718      1.16003699     1.777   .0756?

c.  Aggregation test according to (10-15)
MATRIX ; Z=Init(3,3,0) ; J=Iden(3); L=-1*J $
MATRIX ; R=[j,z,z,z,z,z,z,z,z,l /
            z,j,z,z,z,z,z,z,z,l /
            z,z,j,z,z,z,z,z,z,l /
            z,z,z,j,z,z,z,z,z,l /
            z,z,z,z,j,z,z,z,z,l /
            z,z,z,z,z,j,z,z,z,l /
            z,z,z,z,z,z,j,z,z,l /
            z,z,z,z,z,z,z,j,z,l /
            z,z,z,z,z,z,z,z,j,l ]
       ; d = R*b ; Vd = R*Varb*R'
       ; list ; AggF = 1/27 * d'<vd>d $
Matrix AGGF     has  1 rows and  1 columns.
               1
        +--------------
       1|   98.53777
CALC   ; List ; Ftb(.95,27,(200-10*3)) $
+----------------------------------+
| Listed Calculator Results          |
+----------------------------------+
 Result  =      1.551534
? d.  Using separate OLS regressions, compute LM statistic
? OLS residuals were saved in matrix EOLS earlier.
MATRIX ; VEOLS = 1/20*EOLS'EOLS
       ; VI = Diag(VEOLS) ; SDI = ISQR(VI)
       ; ROLS = SDI*VEOLS*SDI
       ; RR = ROLS' *ROLS $
CALC   ; List ; LMStat = (20/2)*(Trc(RR)-10)
              ; Ctb(.95, (9*10/2))$
+----------------------------------+
| Listed Calculator Results          |
+----------------------------------+
 LMSTAT  =     97.617948
 Result  =     61.656233

? Constrained Sur model with one coefficient vector.
? This is the unconstrained model in (10-19)-(10-21)
SAMPLE ; 1 - 200 $
REGRESS; Lhs = I ; Rhs = F,C,one $
+----------------------------------------------------+
| Ordinary    least squares regression                |
| LHS=I        Mean                 =   145.9582      |
|              Standard deviation   =   216.8753      |
| WTS=none     Number of observs.   =        200      |
| Model size   Parameters           =          3      |
|              Degrees of freedom   =        197      |
| Residuals    Sum of squares       =   1755850.      |
|              Standard error of e  =   94.40840      |
| Fit          R-squared            =   .8124080      |
|              Adjusted R-squared   =   .8105035      |
| Model test   F[  2,   197] (prob) = 426.58 (.0000)  |
+----------------------------------------------------+
+--------+-------------+---------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+-------------+---------------+--------+--------+----------+
 F       |     .11556216       .00583571    19.803   .0000   1081.68110
 C       |     .23067849       .02547580     9.055   .0000   276.017150
 Constant|   -42.7143694      9.51167603    -4.491   .0000
TSCS   ; Lhs = I ; Rhs = F,C,one ; Pds=20 ; Model=S2,R0 $
```

```
+----------------------------------------------------+
| Groupwise Regression Models                         |
| Estimator =                  2 Step GLS             |
| Groupwise Het. and Correlated       (S2)            |
| Nonautocorrelated disturbances      (R0)            |
| Test statistics against the correlation             |
| Deg.Fr. =   45 C*(.95) =  61.66 C*(.99) =  69.96    |
| Test statistics against the correlation             |
| Likelihood ratio    statistic      =    320.2052    |
| Log-likelihood function =           -853.084972     |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]|
+--------+--------------+----------------+--------+--------+
 F       |     .10806238       .00241169    44.808   .0000
 C       |     .15079551       .00386063    39.060   .0000
 Constant|   -20.1588844       .79950153   -25.214   .0000
CREATE ; WI = (SDI(firm,firm))^2 $
REGRESS; Lhs = I ; Rhs = F,C,one ; Wts = WI $
+-----------------------------------------------------+
| Ordinary    least squares regression                 |
| LHS=I        Mean                 =   6.993136       |
|              Standard deviation   =   18.01824       |
| WTS=WI       Number of observs.   =        200       |
| Model size   Parameters           =          3       |
|              Degrees of freedom   =        197       |
| Residuals    Sum of squares       =   11690.82       |
|              Standard error of e  =   7.703521       |
| Fit          R-squared            =   .8190465       |
|              Adjusted R-squared   =   .8172094       |
+-----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |t-ratio |P[|T|>t]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 F       |     .07847124       .00459121    17.092   .0000   96.8424912
 C       |     .09896094       .00761314    12.999   .0000   23.8374846
 Constant|   -2.96519441       .66964256    -4.428   .0000
```

Chapter 11

Nonlinear Regression Models

Exercises

1. 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 $\alpha^0 = \overline{y}$ and $\beta^0 = 0$ or $\beta^0 = 1$ and α^0 either $\mathbf{x'y}/\mathbf{x'x}$ or $\overline{y}/\overline{x}$. 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 Appendix E..

2. The proof can be done by mathematical induction. For convenience, denote the ith derivative by f_i. The first derivative appears in Equation (10-34). 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 $\lim_{\lambda\to 0} f_{\text{i}} = \lim_{\lambda\to 0} [x^\lambda(\ln x)^{i+1}]/(i+1) = (\ln x)^{i+1}/(i+1)$.

Applications

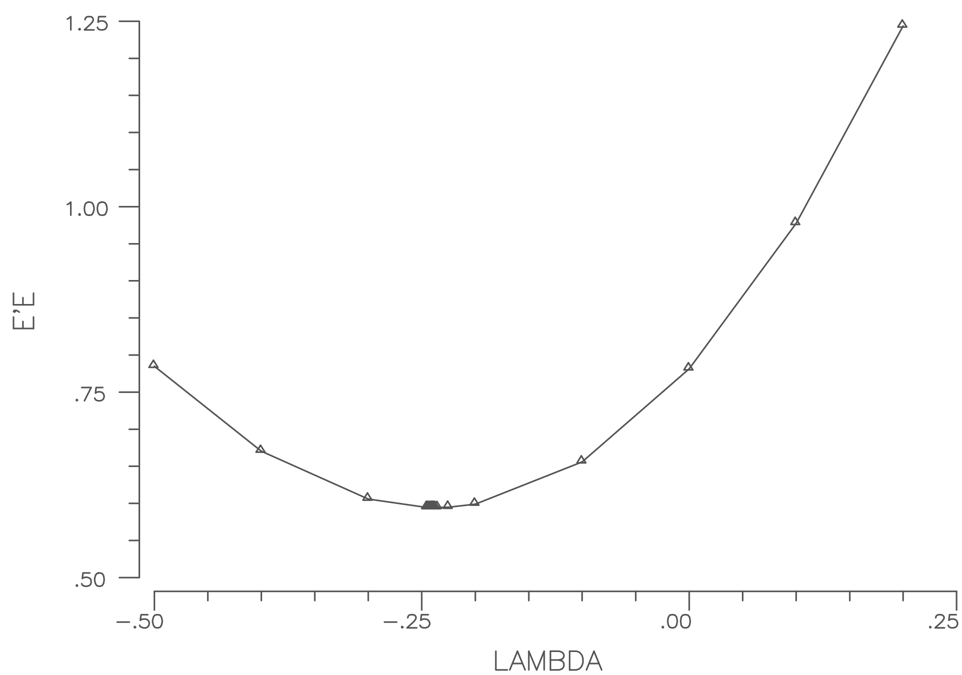
1. First, the two simple regressions produce

	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 λ = 1. Thus, the hypothesis that λ = 0 (the log-linear model) would not be rejected but the hypothesis that λ = 1 (the linear model) would be rejected using the Box-Cox model as a framework. □

2. 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 λ = -.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. □

3. 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 λ = 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. □

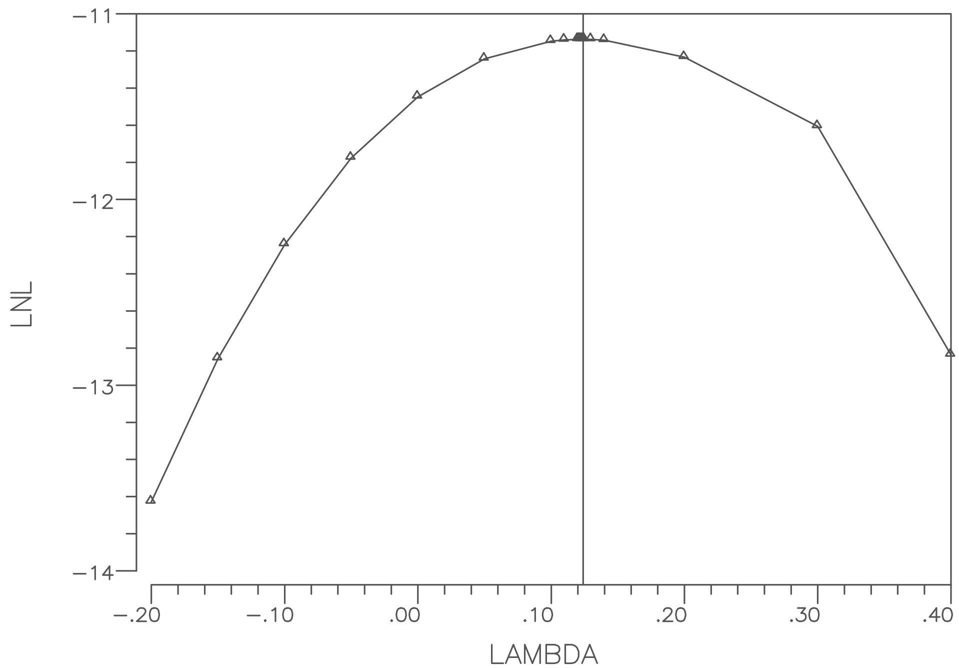
4. Instead of minimizing the sum of squared deviations, we now maximize the concentrated log-likelihood function, $\ln L = -(n/2)\ln(1+\ln(2\pi)) + (\lambda - 1)\Sigma_i \ln Y_i - (n/2)\ln(\varepsilon'\varepsilon/n)$.
The search for the maximum of $\ln L$ produced the results on the next page
The log-likelihood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

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 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 $\ln L^0 = (0-1)(19.29336) - (25/2)[1 + \ln(2\pi) + \ln(.781403/25)] = -11.44757$. 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 11.8. 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.

5. **--> nlsq ; lhs = y ; labels = b1,b2 ; fcn=b1*(1 - 1/sqr(1+2*b2*x)) ; start = 500,.0001 ;output=2$**

```
Begin NLSQ iterations. Linearized regression.
Iteration=  1; Sum of squares=  11603.0164    ; Gradient=  11602.9326
Iteration=  2; Sum of squares=  19821.5463    ; Gradient=  19821.4534
Iteration=  3; Sum of squares=  331169.005    ; Gradient=  331144.576
Iteration=  4; Sum of squares=  356630.271    ; Gradient=  356504.582
Iteration=  5; Sum of squares=  14997.8506    ; Gradient=  14938.8590
Iteration=  6; Sum of squares=  449.855530    ; Gradient=  442.701921
Iteration=  7; Sum of squares=  102026.884    ; Gradient=  102026.775
Iteration=  8; Sum of squares=  12887.7536    ; Gradient=  12886.6539
Iteration=  9; Sum of squares=  14263101.5    ; Gradient=  14263101.0
Iteration= 10; Sum of squares=  10203.1920    ; Gradient=  10202.6789
Iteration= 11; Sum of squares=  144.393444    ; Gradient=  144.338425
Iteration= 12; Sum of squares=  258.186688    ; Gradient=  258.145522
Iteration= 13; Sum of squares=  .154284512    ; Gradient=  .113316151
Iteration= 14; Sum of squares=  .409681292E-01; Gradient=  .129216769E-05
Iteration= 15; Sum of squares=  .409668370E-01; Gradient=  .439070450E-13
Iteration= 16; Sum of squares=  .409668370E-01; Gradient=  .211594637E-18
Iteration= 17; Sum of squares=  .409668370E-01; Gradient=  .107898463E-24
Convergence achieved
+-----------------------------------------------------+
| Nonlinear   least squares regression                |
| LHS=Y        Mean                 =   43.34071      |
|              Standard deviation   =   22.80652      |
| WTS=none     Number of observs.   =         14      |
| Model size   Parameters           =          2      |
|              Degrees of freedom   =         12      |
| Residuals    Sum of squares       =   .4096684E-01  |
|              Standard error of e  =   .5409439E-01  |
| Fit          R-squared            =   .9999939      |
| Not using OLS or no constant. Rsqd & F may be < 0.  |
+-----------------------------------------------------+
+--------+--------------+----------------+--------+--------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]|
+--------+--------------+----------------+--------+--------+
 B1      |    636.427250      4.31789336   147.393   .0000
 B2      |     .00020814     .164134D-05   126.809   .0000
```

--> nlsq ; lhs = y ; labels = b1,b2 ; fcn=b1*(1 - 1/sqr(1+2*b2*x)) ; start = 600,.0002 ;output=2$

```
Begin NLSQ iterations. Linearized regression.
Iteration=  1; Sum of squares=  262.456583    ; Gradient=  262.415454
Iteration=  2; Sum of squares=  .155984704    ; Gradient=  .115016579
Iteration=  3; Sum of squares=  .409675977E-01; Gradient=  .760690867E-06
Iteration=  4; Sum of squares=  .409668370E-01; Gradient=  .379981726E-13
Iteration=  5; Sum of squares=  .409668370E-01; Gradient=  .186919870E-18
Iteration=  6; Sum of squares=  .409668370E-01; Gradient=  .150578559E-23
Convergence achieved
+-----------------------------------------------------+
| Nonlinear   least squares regression                |
| LHS=Y        Mean                 =   43.34071      |
|              Standard deviation   =   22.80652      |
| Residuals    Sum of squares       =   .4096684E-01  |
|              Standard error of e  =   .5409439E-01  |
| Fit          R-squared            =   .9999939      |
|              Adjusted R-squared   =   .9999944      |
+-----------------------------------------------------+
+--------+--------------+----------------+--------+--------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]|
+--------+--------------+----------------+--------+--------+
 B1      |    636.427250      4.31789336   147.393   .0000
 B2      |     .00020814     .164134D-05   126.809   .0000
```

Chapter 12

Instrumental Variables Estimation

Exercises

1. There is no need for a separate proof different from the usual for OLS. Formally, however, it follows from the results at (12-4) that

$$\mathbf{b} \;=\; \beta + \left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1}\left(\frac{\mathbf{X}'\varepsilon}{n}\right)$$

Then,

$$\mathbf{b} - \text{plim}\,\mathbf{b} = \left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1}\left(\frac{\mathbf{X}'\varepsilon}{n}\right) - \mathbf{Q}_{\mathrm{XX}}^{-1}\gamma$$

and

$$\sqrt{n}\left(\mathbf{b} - \text{plim}\,\mathbf{b}\right) = \sqrt{n}\left[\left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1}\left(\frac{\mathbf{X}'\varepsilon}{n}\right) - \mathbf{Q}_{\mathrm{XX}}^{-1}\gamma\right]$$

The large sample distribution of this statistic will be the same as the large sample of the statistic with $X'X/n$ replaced with its probablity limit, which is $\mathbf{Q_{XX}}$. Thus,

$$\sqrt{n}\left(\mathbf{b} - \text{plim}\,\mathbf{b}\right) \rightarrow \mathbf{Q}_{\mathrm{XX}}^{-1}\sqrt{n}\left[\left(\frac{\mathbf{X}'\varepsilon}{n}\right) - \gamma\right]$$

To deduce the large sample behavior of this statistic, we can invoke the results from chapter 4. The only change here is the nonzero mean (probability limit) of the vector in brackets. [See (12-3).] Thus, the same proof applies. The consistency, asymptotic normality and asymptotic covariance matrix equal to Asy.Var[b] = $\sigma_\varepsilon^2\,(X'X)^{-1}$

2. A logical solution to this one is simple. For y and x*,

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

$\text{Cov}^2(y,x)\,/[\text{Var}(y)\text{Var}(x)] \;= \text{Cov}[\beta x^*+\varepsilon, x^*+u]\,/\,[\text{Var}(y)\text{Var}(x)]$

$\qquad\qquad = \{\text{Cov}[y,x^*] + \text{Cov}[y,u]\}^2\,/\,[\text{Var}(y)\text{Var}(x)]$.

The second term is zero, since $y=\beta x^*+\varepsilon$ which is uncorrelated with u. Thus,

$\text{Cov}^2(y,x)\,/[\text{Var}(y)\text{Var}(x)] \;= \text{Cov}[y,x^*]\,/\,[\text{Var}(y)\text{Var}(x)]$.

The numerator is the same. The denominator is larger, since [Var(y)Var(x)] = Var[y](Var[x*] + Var[u]), so the squared correlation must be smaller. If both variables are measured with errors, then we are comparing $\text{Cov}^2(y^*,x^*)/\{\text{Var}[y^*]\text{Var}[x^*]\}$ to $\text{Cov}^2(y,x)/\{\text{Var}[y]\text{Var}[x]\}$.

The numerator is the covariance of $(\beta x^* + \varepsilon + v)$ with $(x^* + u)$, so the numerator of the fraction is still $\beta^2(\sigma_*^2)^2$. The denominator is still obviously larger, so the same result holds when both variables are measured with error.

3. We work off (12-16), using repeatedly the result $\Sigma_{uu} = (\sigma_u j)(\sigma_u j)'$ where j has a 1 in the first position and 0 in the remaining K-1. From (12-16),

plim b = β - $[Q^* + \Sigma_{uu}]^{-1}\Sigma_{uu}\beta$. The vector is $\Sigma_{uu}\beta$ equals $[\sigma_u^2\beta_1,0,...,0]'$. The inverse matrix is

$$[Q^* + \Sigma_{uu}]^{-1} = \left[\left(\mathbf{Q}^*\right)^{-1} - \frac{1}{1+(\sigma_u\mathbf{j})'\left(\mathbf{Q}^*\right)^{-1}(\sigma_u\mathbf{j})}\left(\mathbf{Q}^*\right)^{-1}(\sigma_u\mathbf{j})(\sigma_u\mathbf{j})'\left(\mathbf{Q}^*\right)^{-1}\right]$$

This can be simplified since the quadratic form in the denominator just picks off the 1,1 diagonal element. Thus,

$[\mathbf{Q}^* + \mathbf{\Sigma}_{uu}]^{-1} = \left[(\mathbf{Q}^*)^{-1} - \frac{1}{1+\sigma_u^2 q^{*11}}(\mathbf{Q}^*)^{-1}(\sigma_u\mathbf{j})(\sigma_u\mathbf{j})'(\mathbf{Q}^*)^{-1}\right]$

Then

$[\mathbf{Q}^* + \mathbf{\Sigma}_{uu}]^{-1}\mathbf{\Sigma}_{uu}\boldsymbol{\beta} = \left[(\mathbf{Q}^*)^{-1} - \frac{1}{1+\sigma_u^2 q^{*11}}(\mathbf{Q}^*)^{-1}(\sigma_u\mathbf{j})(\sigma_u\mathbf{j})'(\mathbf{Q}^*)^{-1}\right](\sigma_u\mathbf{j})(\sigma_u\mathbf{j})'\beta$

$$= (\mathbf{Q}^*)^{-1}(\sigma_u\mathbf{j})(\sigma_u\mathbf{j})'\boldsymbol{\beta} - \frac{1}{1+\sigma_u^2 q^{*11}}(\mathbf{Q}^*)^{-1}(\sigma_u\mathbf{j})(\sigma_u\mathbf{j})'(\mathbf{Q}^*)^{-1}(\sigma_u\mathbf{j})(\sigma_u\mathbf{j})'\beta$$

$$= (\mathbf{Q}^*)^{-1}\mathbf{j}\,\sigma_u^2\beta_1 - \frac{\sigma_u^2 q^{*11}}{1+\sigma_u^2 q^{*11}}(\mathbf{Q}^*)^{-1}\mathbf{j}\,\sigma_u^2\beta_1$$

$$= (\mathbf{Q}^*)^{-1}\mathbf{j}\left[1 - \frac{\sigma_u^2 q^{*11}}{1+\sigma_u^2 q^{*11}}\right]\sigma_u^2\beta_1$$

$$= (\mathbf{Q}^*)^{-1}\mathbf{j}\left[\frac{1}{1+\sigma_u^2 q^{*11}}\right]\sigma_u^2\beta_1$$

$$= (\mathbf{Q}^*)^{-1}\mathbf{j}\left[\frac{\sigma_u^2\beta_1}{1+\sigma_u^2 q^{*11}}\right]$$

Finally, $(\mathbf{Q}^*)^{-1}\mathbf{j}$ equals the first column of $(\mathbf{Q}^*)^{-1}$ = [q^{*11}, q^{*21},...,q^{*k1}]. Therefore, the first element, given by (12-17a) is

$$\text{plim } b_1 = \beta_1 - \left[\frac{\sigma_u^2\beta_1}{1+\sigma_u^2 q^{*11}}\right]q^{*11} = \beta_1\left[1 - \frac{\sigma_u^2 q^{*11}}{1+\sigma_u^2 q^{*11}}\right]$$

For (12-17b),

$$\text{plim } b_2 = \beta_2 - \left[\frac{\sigma_u^2\beta_1}{1+\sigma_u^2 q^{*11}}\right]q^{*k1}$$

4. To obtain the result, note first:

plim $\mathbf{b} = \boldsymbol{\beta} + \mathbf{Q}_{XX}^{-1}\boldsymbol{\gamma}$

Asy.Var[$\mathbf{b}$] = $(\sigma^2/n)\mathbf{Q}_{XX}^{-1}$

Asy.Var[$\mathbf{b}_{2sls}$] = $(\sigma^2/n)\mathbf{Q}_{ZX}^{-1}\mathbf{Q}_{ZZ}\mathbf{Q}_{XZ}^{-1}$.

The mean squared error of the OLS estimator is the variance plus the squared bias,

$$M(b|\beta) = (\sigma^2/n)\mathbf{Q}_{XX}^{-1} + \mathbf{Q}_{XX}^{-1}\gamma\gamma'\mathbf{Q}_{XX}^{-1}$$

the mean squared error of the 2SLS estimator equals its variance. For OLS to be more precise then 2SLS, we would have to have

$$(\sigma^2/n)\mathbf{Q}_{XX}^{-1} + \mathbf{Q}_{XX}^{-1}\gamma\gamma'\mathbf{Q}_{XX}^{-1} << (\sigma^2/n)\mathbf{Q}_{ZX}^{-1}\mathbf{Q}_{ZZ}\mathbf{Q}_{XZ}^{-1}.$$

For convenience, let $\delta = \mathbf{Q}_{XX}^{-1}\gamma$ so $M(b|\beta) = (\sigma^2/n)\mathbf{Q}_{XX}^{-1} + \delta\delta'$. If the mean squared error matrix of the OLS estimator is smaller than that of the 2SLS estimator, then its inverse is larger. Use (A-66) to do the inversion. The result would be

$$[(\sigma^2/n)\mathbf{Q}_{XX}^{-1} + \delta\delta']^{-1} >> [(\sigma^2/n)\mathbf{Q}_{ZX}^{-1}\mathbf{Q}_{ZZ}\mathbf{Q}_{XZ}^{-1}]^{-1}$$

Now, use A-66

$$[(\sigma^2/n)\mathbf{Q}_{XX}^{-1} + \delta\delta']^{-1} = (n/\sigma^2)\,\mathbf{Q}_{XX} - \frac{1}{1+\delta'(n/\sigma^2)\mathbf{Q}_{XX}\delta}(n/\sigma^2)\,\mathbf{Q}_{XX}\delta\delta'(n/\sigma^2)\,\mathbf{Q}_{XX}$$

Reinsert $\delta = \mathbf{Q}_{XX}^{-1}\gamma$ and the right hand side above reduces to

$$(n/\sigma^2)\,\mathbf{Q}_{XX} - \frac{1}{1+(n/\sigma^2)\gamma'\mathbf{Q}_{XX}^{-1}\gamma}(n/\sigma^2)^2\,\gamma\gamma'$$

Therefore, if the mean squared error matrix of OLS is smaller, then

$$(n/\sigma^2)\,\mathbf{Q}_{XX} - \frac{1}{1+(n/\sigma^2)\gamma'\mathbf{Q}_{XX}^{-1}\gamma}(n/\sigma^2)^2\,\gamma\gamma' >> (n/\sigma^2)\mathbf{Q}_{XZ}\mathbf{Q}_{ZZ}^{-1}\mathbf{Q}_{ZX}$$

Collect the terms, and this implies

$$(n/\sigma^2)[\,\mathbf{Q}_{XX} - \mathbf{Q}_{XZ}\mathbf{Q}_{ZZ}^{-1}\mathbf{Q}_{ZX}] >> \frac{1}{1+(n/\sigma^2)\gamma'\mathbf{Q}_{XX}^{-1}\gamma}(n/\sigma^2)^2\,\gamma\gamma'$$

divide both sides by (n/σ^2),

$$\mathbf{Q}_{XX} - \mathbf{Q}_{XZ}\mathbf{Q}_{ZZ}^{-1}\mathbf{Q}_{ZX} >> \frac{(n/\sigma^2)}{1+(n/\sigma^2)\gamma'\mathbf{Q}_{XX}^{-1}\gamma}\gamma\gamma'$$

and divide numerator and denominator of the fraction by n/σ^2

$$\mathbf{Q}_{XX} - \mathbf{Q}_{XZ}\mathbf{Q}_{ZZ}^{-1}\mathbf{Q}_{ZX} >> \frac{1}{(\sigma^2/n)+\gamma'\mathbf{Q}_{XX}^{-1}\gamma}\gamma\gamma'$$

which is the desired result. Is it possible? It is possible, since

$$\begin{aligned}\mathbf{Q}_{XX} - \mathbf{Q}_{XZ}\mathbf{Q}_{ZZ}^{-1}\mathbf{Q}_{ZX} &= \text{plim}\,(1/n)[\mathbf{X}'\mathbf{X} - \mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}] \\ &= \text{plim}\,(1/n)\,\mathbf{X}'\mathbf{M}_Z\mathbf{X}\end{aligned}$$

which is a positive definite matrix. SInce γ varies independently of $\mathbf{Z}$ and $\mathbf{X}$, certainly there is some configuration of the data and parameters for which this is the case. The result is that it is, indeed, possible for OLS to be more precise, in the mean squared error sense, than 2SLS.

5. The matrices are X = [i,x] and Z = [i,z]. For the OLS estimators, we know from chapter 2 that

$$a = \bar{y} - b\bar{x} \text{ and } b = \text{Cov}[x,y]/\text{var}[x].$$

For the IV estimator, $(\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{y}$, we obtain the result in detail. Given the forms,

$$(\mathbf{Z}'\mathbf{X}) = \begin{bmatrix} n & \Sigma x_i \\ n_1 & \Sigma_{z=1}x_i \end{bmatrix} = \begin{bmatrix} n & n\bar{x} \\ n_1 & n_1\bar{x}_1 \end{bmatrix},\ (\mathbf{Z}'\mathbf{X})^{-1} = \frac{1}{nn_1(\bar{x}-\bar{x})}\begin{bmatrix} n_1\bar{x}_1 & -n\bar{x} \\ -n_1 & n \end{bmatrix},\ \mathbf{Z}'\mathbf{y} = \begin{bmatrix} n\bar{y} \\ n_1\bar{y}_1 \end{bmatrix}$$

where subscript 1 indicates the mean of the observations for which z equals 1, and n_1 is the number of observations. Multiplying the matrix times the vector and cancelling terms produces the solutions

$$a_{IV} = a_{IV} = \frac{\bar{x}_1\bar{y} - \bar{x}\,\bar{y}_1}{\bar{x}_1 - \bar{x}} \text{ and } b_{IV} = \frac{\bar{y}_1 - \bar{y}}{\bar{x}_1 - \bar{x}}$$

Application

a. The statement of the problem is actually a bit optimistic. GIven the way it is stated, it would imply that the exogenous variables in the "demand" equation would be, in principle, (Ed, Union, Fem) which are also in the supply equation, plus the remainder, (Exp, Exp^2, Occ, Ind, South, SMSA, Blk). The problem is that the model as stated would not be identified – the supply equation would, but the demand equation would not be. The way out would be to assume that at least one of (Ed, Union, Fem) does not appear in the demand equation. Since surely education would, that leaves one or both of Union and Fem. We will assume both of them are omitted. So, our equation is

$$\begin{aligned} \ln Wage_{it} = \; & \alpha_1 + \alpha_2 Ed_{it} + \alpha_3 Exp_{it} + \alpha_4 Exp_{it}^2 + \alpha_5 Occ_{it} + \\ & \alpha_6 Ind_{it} + \alpha_7 South_{it} + \alpha_8 SMSA_{it} + \alpha_9 Blk_{it} + \gamma\, Wks_{it} + u_{it}. \end{aligned}$$

```
NAMELIST ; X = one,Ed,Exp,Expsq,Occ,Ind,South,SMSA,Blk,Wks $
NAMELIST ; Z = one,Ed,Exp,expsq,Occ,Ind,south,SMSA,Blk,Union,Fem $
Regress  ; Lhs = lwage ; Rhs = X $
2SLS     ; Lhs = lwage ; Rhs = X ; Inst = Z $
REGRESS  ; Lhs = Wks ; Rhs = Z ; cls:b(10)=0,b(11)=0$
+----------------------------------------------------+
| Ordinary    least squares regression               |
| LHS=LWAGE    Mean                 =   6.676346     |
|              Standard deviation   =   .4615122     |
| WTS=none     Number of observs.   =       4165     |
| Model size   Parameters           =         10     |
|              Degrees of freedom   =       4155     |
| Residuals    Sum of squares       =   581.2717     |
|              Standard error of e  =   .3740280     |
| Fit          R-squared            =   .3446066     |
|              Adjusted R-squared   =   .3431870     |
| Model test   F[  9,  4155] (prob) = 242.74 (.0000) |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|    5.13171052       .07238152    70.898   .0000
 ED      |     .06112766       .00277226    22.050   .0000   12.8453782
 EXP     |     .04291665       .00229783    18.677   .0000   19.8537815
 EXPSQ   |    -.00070803     .506204D-04   -13.987   .0000   514.405042
 OCC     |    -.07814434       .01502100    -5.202   .0000    .51116447
 IND     |     .09066812       .01247863     7.266   .0000    .39543818
 SOUTH   |    -.07629062       .01318346    -5.787   .0000    .29027611
 SMSA    |     .13789225       .01278553    10.785   .0000    .65378151
 BLK     |    -.26269494       .02304380   -11.400   .0000    .07226891
 WKS     |     .00484184       .00113470     4.267   .0000   46.8115246
+----------------------------------------------------+
| Two stage   least squares regression               |
| LHS=LWAGE    Mean                 =   6.676346     |
|              Standard deviation   =   .4615122     |
| WTS=none     Number of observs.   =       4165     |
| Model size   Parameters           =         10     |
|              Degrees of freedom   =       4155     |
| Residuals    Sum of squares       =   602.3138     |
|              Standard error of e  =   .3807377     |
| Fit          R-squared            =   .3192467     |
|              Adjusted R-squared   =   .3177722     |
| Model test   F[  9,  4155] (prob) = 216.50 (.0000) |
+----------------------------------------------------+
| Instrumental Variables:
|ONE      ED       EXP      EXPSQ    OCC      IND      SOUTH    SMSA
|BLK      UNION    FEM
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|    4.46105888       .27680953    16.116   .0000
 ED      |     .06167266       .00283031    21.790   .0000   12.8453782
```

```
 EXP      |      .04207640       .00236282    17.808   .0000   19.8537815
 EXPSQ    |    -.00068241     .525268D-04   -12.992   .0000   514.405042
 OCC      |    -.07605669       .01531301    -4.967   .0000    .51116447
 IND      |     .08348143       .01302032     6.412   .0000    .39543818
 SOUTH    |    -.08242895       .01364036    -6.043   .0000    .29027611
 SMSA     |     .13244624       .01319402    10.038   .0000    .65378151
 BLK      |    -.25212290       .02383132   -10.579   .0000    .07226891
 WKS      |     .01922950       .00583960     3.293   .0010   46.8115246
```

This is the test of relevance of the instrumental variables. In the regression of WKS on the full set of exogenous variables, we test the hypothesis that the coefficients on the instruments, UNION and FEM are jointly zero. The results show that the hypothesis is rejected. We conclude that the instruments are relevant.

```
+-----------------------------------------------------+
| Linearly restricted regression                      |
| Ordinary    least squares regression                |
| LHS=WKS      Mean                 =   46.81152      |
|              Standard deviation   =   5.129098      |
| WTS=none     Number of observs.   =       4165      |
| Model size   Parameters           =          9      |
|              Degrees of freedom   =       4156      |
| Residuals    Sum of squares       =   108653.5      |
|              Standard error of e  =   5.113097      |
| Fit          R-squared            =   .8138966E-02   |
|              Adjusted R-squared   =   .6229705E-02   |
| Model test   F[  8,  4156] (prob) =   4.26 (.0000)  |
| Restrictns.  F[  2,  4154] (prob) =  84.57 (.0000)  |
| Not using OLS or no constant. Rsqd & F may be < 0.  |
| Note, with restrictions imposed,  Rsqd may be < 0.  |
+-----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|    46.6129896       .67547781    69.007   .0000
 ED       |    -.03787988       .03789322    -1.000   .3175   12.8453782
 EXP      |     .05840099       .03139904     1.860   .0629   19.8537815
 EXPSQ    |    -.00178055       .00069145    -2.575   .0100   514.405042
 OCC      |    -.14509978       .20533021     -.707   .4798    .51116447
 IND      |     .49950389       .17041135     2.931   .0034    .39543818
 SOUTH    |     .42663864       .18010107     2.369   .0178    .29027611
 SMSA     |     .37851979       .17468415     2.167   .0302    .65378151
 BLK      |    -.73479892       .31481083    -2.334   .0196    .07226891
 UNION    |    .444089D-15    .182255D-08      .000  1.0000    .36398559
 FEM      |       .000000    ......(Fixed Parameter).......
```

Chapter 13

Simultaneous Equations Models

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

(1) We use (13-12) 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 $\mathbf{x}$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 $\mathbf{x}$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 $\mathbf{F}$ 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. Following the method in Example 13.6, for identification of the investment equation, we require that the matrix
$$\begin{array}{ccccccccc} (1) & (2) & (3) & (4) & (5) & (6) & (7) & (8) & (9) \end{array}$$
$$\begin{bmatrix} -1 & \alpha_3 & 0 & 0 & \alpha_3 & 0 & 0 & 0 & 0 \\ 0 & -1 & \gamma_1 & 0 & 0 & 0 & 0 & \gamma_3 & \gamma_2 \\ 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$
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
$$\begin{array}{cccccccccc} (1) & (2) & (3) & (4) & (5) & (6) & (7) & (8) & (9) & (10) \end{array}$$
$$\begin{bmatrix} -1 & 0 & \alpha_1 & 0 & \alpha_3 & 0 & 0 & 0 & \alpha_2 & 0 \\ 0 & -1 & \beta_1 & 0 & 0 & 0 & 0 & 0 & \beta_2 & \beta_3 \\ 1 & 1 & 0 & 0 & 0 & 01 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$
Columns (4), (6), (7), (9), and (10) are linearly independent. □

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

$$\begin{array}{c} (1) \\ \begin{bmatrix} \gamma_{32} & 1 & \gamma_{34} \\ \beta_{12} & \beta_{13} & \beta_{14} \\ 0 & \beta_{43} & \beta_4 \\ \beta_{32} & 0 & 0 \end{bmatrix} \end{array}, \quad \begin{array}{c} (2) \\ \begin{bmatrix} 0 & \beta_{43} & \beta_{44} \end{bmatrix} \end{array}, \quad \begin{array}{c} (3) \\ \begin{bmatrix} 1 & \gamma_{12} & 0 \\ \gamma_{41} & \gamma_{42} & 1 \\ \beta_{21} & 1 & 0 \\ 0 & \beta_{52} & 00 \end{bmatrix} \end{array}, \quad \begin{array}{c} (4) \\ \begin{bmatrix} 1 & \gamma_{12} & 0 \\ \beta_{31} & \beta_{32} & \beta_{33} \\ 0 & \beta_{52} & 0 \end{bmatrix} \end{array}$$

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

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

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

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

5. The 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

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

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

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

The estimated asymptotic covariance matrices are

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

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

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.

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

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

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

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

Needless to say, this would be quite involved. □

7. 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 $\mathbf{Y}_j'\boldsymbol{\varepsilon}/T = \boldsymbol{\omega}_\mathbf{j} - \boldsymbol{\Omega}_{jj}\boldsymbol{\gamma}_\mathrm{j}$.

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

$$\boldsymbol{\Omega} = \begin{matrix} \omega_{jj} & \boldsymbol{\omega}_j' & \boldsymbol{\omega}^*_j{}' \\ \boldsymbol{\omega}_j & \boldsymbol{\Omega}_{jj} & \boldsymbol{\Omega}_j' \\ \boldsymbol{\omega}^*_j & \boldsymbol{\Omega}^*_j & \boldsymbol{\Omega}_j^* \end{matrix}$$

and, as in the equation preceding (13-8), 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

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

Application

```
?=========================================================
? Application 13.1 - Simultaneous Equations
?=========================================================
? Read the data
? For convenience, rename the variables so they correspond
? to the example in the text.
sample ; 1 - 204 $
create ; ct=realcons$
create ; it=realinvs$
create ; gt=realgovt$
create ; rt=tbilrate $
? Impose (artifically) the adding up condition on total demand.
create ; yt=ct+it+gt $
create ; ct1=ct[-1] $
create ; yt1 = yt[-1] $
create ; dyt = yt - yt1 $
sample ; 2-204 $
names  ; xt = one,gt,rt,ct1,yt1$
? Estimate equations by 2sls and save coefficients with
? the names used in the example.
2sls   ; lhs = ct ; rhs=one,yt,ct1 ; inst = xt $
+----------------------------------------------------+
| Two stage   least squares regression               |
| LHS=CT       Mean                 =   3008.995     |
|              Standard deviation   =   1456.900     |
| WTS=none     Number of observs.   =        203     |
| Model size   Parameters           =          3     |
|              Degrees of freedom   =        200     |
| Residuals    Sum of squares       =   75713.32     |
|              Standard error of e  =   19.45679     |
| Fit          R-squared            =   .9998208     |
|              Adjusted R-squared   =   .9998190     |
| Model test   F[  2,   200] (prob) =******* (.0000) |
+----------------------------------------------------+
| Instrumental Variables:
|ONE      GT       RT       CT1      YT1
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|   -13.8657181      5.31536302    -2.609   .0091
 YT      |     .05843862       .01790473     3.264   .0011   4663.67389
 CT1     |     .92200662       .02657199    34.698   .0000   2982.97438
calc   ; a0=b(1) ; a1=b(2) ; a2=b(3) $
2sls   ; lhs = it ; rhs=one,rt,dyt ; inst = xt $
+----------------------------------------------------+
| Two stage   least squares regression               |
| LHS=IT       Mean                 =   654.5296     |
|              Standard deviation   =   391.3705     |
| WTS=none     Number of observs.   =        203     |
| Model size   Parameters           =          3     |
|              Degrees of freedom   =        200     |
| Residuals    Sum of squares       =   .7744227E+08 |
|              Standard error of e  =   622.2631     |
| Fit          R-squared            =  -1.540485     |
|              Adjusted R-squared   =  -1.565889     |
+----------------------------------------------------+
| Instrumental Variables:
|ONE      GT       RT       CT1      YT1
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|   -300.699429      125.980850    -2.387   .0170
 RT      |    56.5192542      15.4643912     3.655   .0003   5.24965517
 DYT     |    16.5359646      2.02509785     8.166   .0000   39.8236453
calc   ; b0=b(1) ; b1=b(2) ; b2=b(3) $
```

```
?
? Create the coefficients of the reduced form.  We only need the parts
? for the dynamics.  These are in the second half of the example.
calc   ; a=1-a1-b2 $
?
? Construct the matrix that governs the dynamics of the system.  Note that
? the I equation is static.  It is a function of y(t-1) and c(t-1) but not
? of I(t-1).  This is the DELTA(1) submatrix in (13-42).  The dominant
? root is the largest rood of DELTA(1).
calc   ; list ; C11=(1-b2)/a ; C12=-a1*b2/a ; C21=a2/a ; C22=-b2/a $
matrix ; C = [c11,c12 / c21,c22] $
+----------------------------------+
| Listed Calculator Results        |
+----------------------------------+
 C11     =       .996253
 C12     =       .061967
 C21     =      -.059124
 C22     =      1.060378
Matrix ; list ; roots = cxrt(c)$
Calc   ; list ; domroot = sqr(roots(1,1)^2 + roots(1,2)^2)$
--> Matrix ; list ; roots = cxrt(c)$

Matrix ROOTS    has  2 rows and  2 columns.
               1             2
        +---------------------------
       1|    1.02832      -.05134
       2|    1.02832       .05134
--> Calc   ; list ; domroot = sqr(roots(1,1)^2 + roots(1,2)^2)$
+----------------------------------+
| Listed Calculator Results        |
+----------------------------------+
 DOMROOT =      1.029596

? The largest root is larger than on in absolute value.  The system is unstable.

3sls   ; lhs = ct,it ; eq1=one,yt,ct1 ; eq2=one,rt,dyt ; inst=xt ; maxit=0 $
+-------------------------------------------------+
| Estimates for equation: CT                      |
| InstVar/GLS least squares regression            |
| LHS=CT       Mean                 =   3008.995    |
| Residuals    Sum of squares       =   73370.06    |
|              Standard error of e  =   19.15334    |
+-------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|   -17.4780776      4.55837624    -3.834   .0001
 YT      |     .07312129       .01415744     5.165   .0000   4663.67389
 CT1     |     .90026227       .02103720    42.794   .0000   2982.97438
+-------------------------------------------------+
| Estimates for equation: IT                      |
| InstVar/GLS least squares regression            |
| LHS=IT       Mean                 =   654.5296    |
| Residuals    Sum of squares       =   .9735005E+08 |
|              Standard error of e  =   697.6749    |
+-------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|   -236.744328      122.661644    -1.930   .0536
 RT      |    30.5417941      12.9861014     2.352   .0187   5.24965517
 DYT     |    18.3544221      1.93633720     9.479   .0000   39.8236453
```

Chapter 14

Estimation Frameworks in Econometrics

Exercise

1. 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.

Applications

1. 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

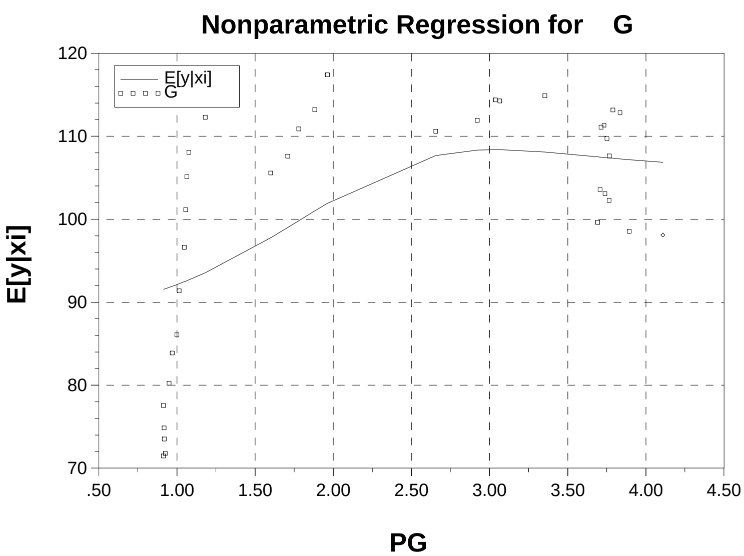
$$\ln(G/Pop) = \beta_1 \ln(Income) + \beta_2 \ln P_{new\ cars} + \beta_3 \ln P_{used\ cars} + g(\ln P_{gasoline}) + \varepsilon$$

```
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
```

2.

```
+--------------------------------------+
| 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   |
+--------------------------------------+
```



3. 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
```

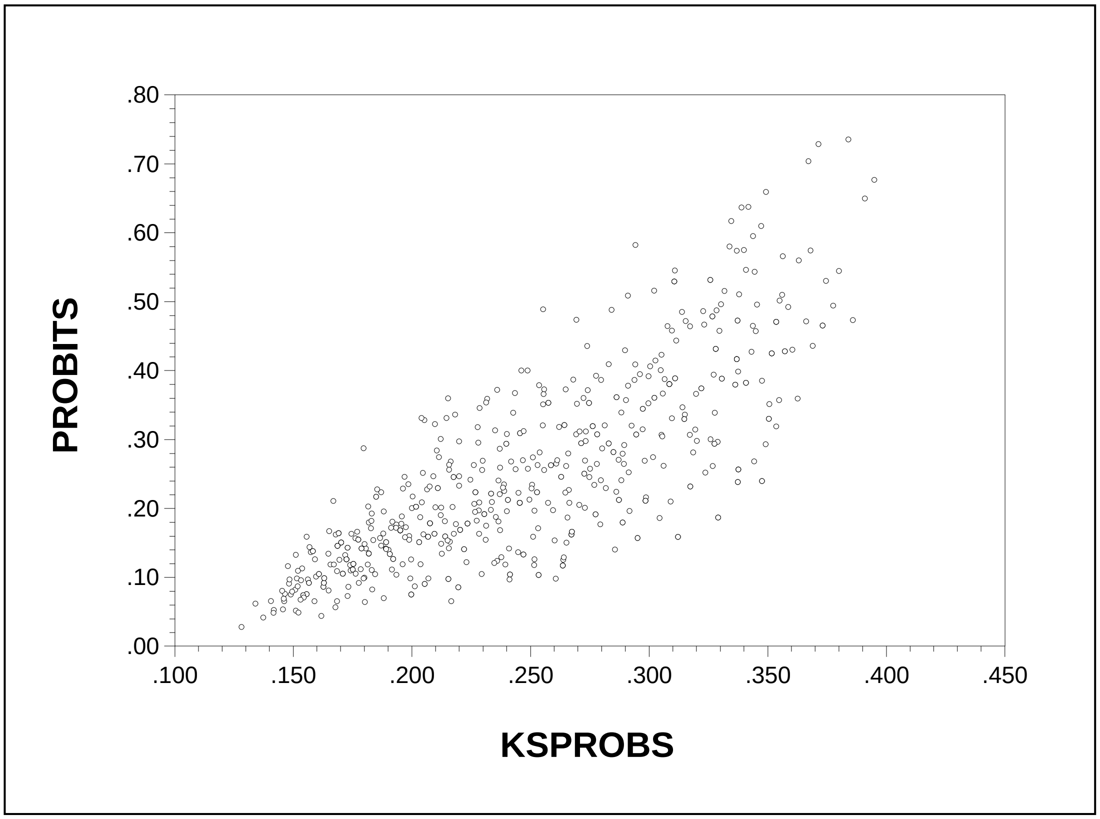
These are the various fit measures for the probit model

```
+----------------------------------------+
| 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
```

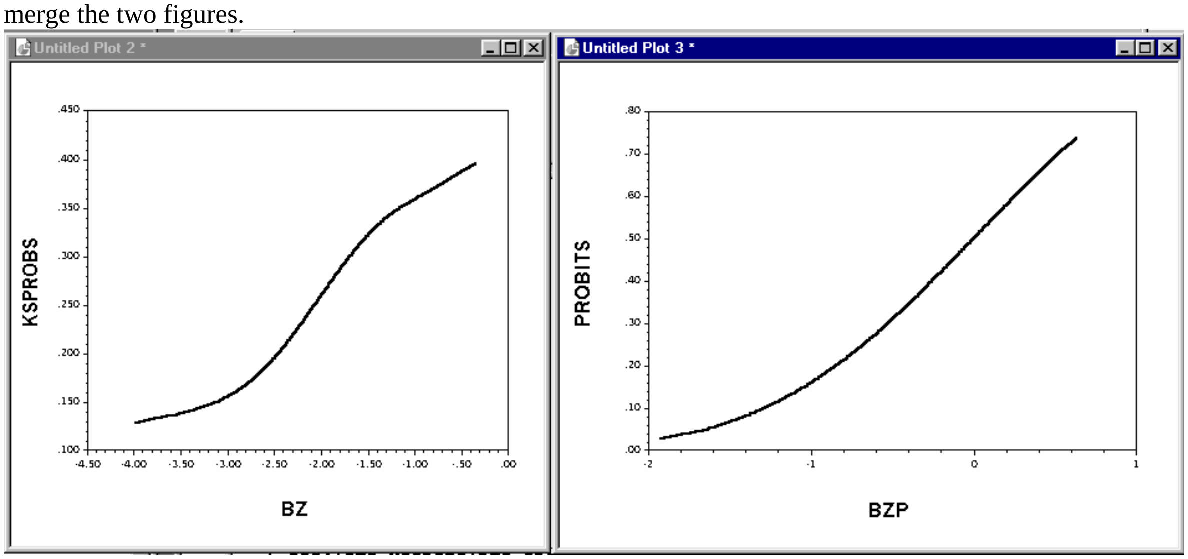
These are the fit measures for the probabilities computed for the Klein and Spady model. The probit model fits better by all measures computed.

```
+----------------------------------------+
| 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 15

Minimum Distance Estimation and The Generalized Method of Moments

Exercises

1. 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. The necessary data are given in Examples 15.5. 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

$$\mathbf{G} = \begin{bmatrix} \partial\mu_1'/\partial P & \partial\mu_1'/\partial\lambda \\ \partial\mu_2'/\partial P & \partial\mu_2'/\partial\lambda \end{bmatrix} = \begin{bmatrix} 1/\lambda & -P/\lambda^2 \\ (2P+1)/\lambda^2 & -2P(P+1)/\lambda^3 \end{bmatrix} = \begin{bmatrix} 15.207 & -475.648 \\ 1{,}182.54 & -44{,}220.08 \end{bmatrix}.$$

The covariance matrix for the moments is given in Example 18.7;

$$\Phi = \begin{bmatrix} 24.7051 & 2307.126 \\ 2307.126 & 229{,}609.5 \end{bmatrix}$$

3. 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. The question is (deliberately) misleading. We showed in Chapter 8 and in this chapter 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 8.2. The estimator of the asymptotic covariance matrices are $s^2(\mathbf{X}'\mathbf{X})^{-1}$ for OLS and the White estimator for GMM.

5. 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. This is the comparison between (15-12) and (15-11). The proof can be done by comparing the inverses of the two covariance matrices. Thus, if the claim is correct, the matrix in (15-11) is larger than that in (15-12), or its inverse is smaller. We can ignore the (1/n) as well. We require, then, that

$$\bar{\mathbf{G}}'\boldsymbol{\Phi}^{-1}\bar{\mathbf{G}} > \bar{\mathbf{G}}'\mathbf{W}\bar{\mathbf{G}}[\bar{\mathbf{G}}'\mathbf{W}\boldsymbol{\Phi}\mathbf{W}\bar{\mathbf{G}}]^{-1}\bar{\mathbf{G}}'\mathbf{W}\bar{\mathbf{G}}$$

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 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. Cov[$\bar{z}$(2.1), $\bar{z}$(3.6)]]
= $1/n$ Cov[$z_i(2.1),z_i(3.6)$], and, since in random sampling sample moments will converge to their population counterparts, Cov[$z_i(2.1),z_i(3.6)$] = 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
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 **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 **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 16

Maximum Likelihood Estimation

Exercises

1. 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 plim $z = \theta$.

2. 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. The log-likelihood is $\ln L = n\ln\theta - (\beta+\theta)\sum_{i=1}^n y_i + \ln\beta\sum_{i=1}^n x_i + \sum_{i=1}^n x_i \ln y_i - \sum_{i=1}^n \ln(x_i!)$

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 $\hat{\theta}_{ML} = 1/\bar{y}$ 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 $[-1/(1+E[x])^2]^2\text{Asy.Var}[\bar{x}]$.)

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 $\text{Asy.Var.}\left[\hat{\lambda}\right] = (\lambda^2/n)/(1+\bar{x})$

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. 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

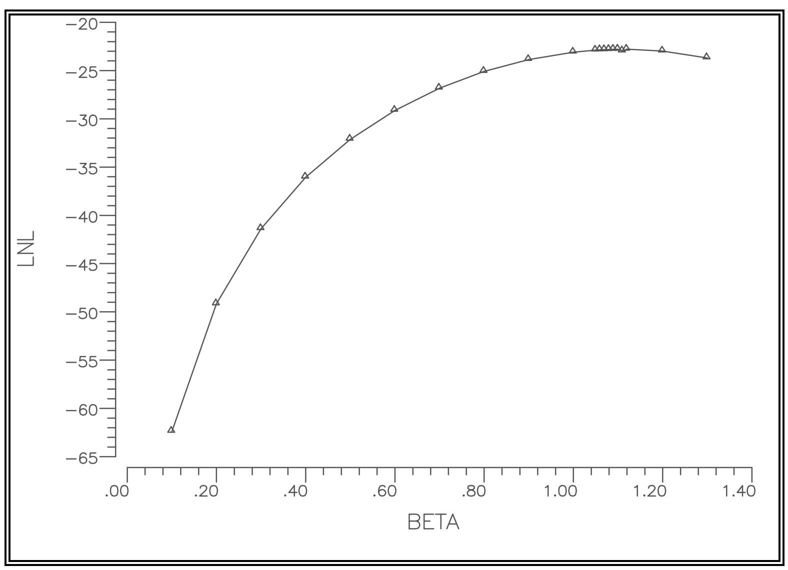
$$1/\beta + E[\ln x_i] - E[(\ln x_i)x_i^{\beta}]/E[x_i^{\beta}] = 0.$$

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 $\qquad 1/(\alpha\beta) \;=\; \mathrm{Cov}[\ln x_i, x_i^{\beta}]$. ~

5. 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 $\mathbf{I}\left(\hat{\alpha},\hat{\beta}\right) = \begin{bmatrix} 25.55 & 9.6506 \\ 9.6506 & 27.7552 \end{bmatrix}$ 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 $\beta = 1$, 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 $\alpha = .88496$ and $\beta = 1$. 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. 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, $\text{Asy.Var}[a] = \{[-E[\Sigma^2\partial \log f_x(x|\alpha)/\partial\alpha^2]\}^{-1}$. 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.

$$-E\frac{\partial^2 \ln L}{\partial\begin{pmatrix}\alpha\\ \beta\end{pmatrix}\begin{pmatrix}\alpha\\ \beta\end{pmatrix}'} = \begin{bmatrix} B_{\alpha\alpha} & B_{\alpha\beta} \\ B_{\beta\alpha} & B_{\beta\beta} \end{bmatrix} + \begin{bmatrix} A_{\alpha\alpha} & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} A_{\alpha\alpha} + B_{\alpha\alpha} & B_{\alpha\beta} \\ B_{\beta\alpha} & B_{\beta\beta} \end{bmatrix}$$

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. 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

$$E[(1/n)\log L] = -\lambda + \log\lambda\, E_0[\,\bar{y}\,] - E_0\,(1/n)\Sigma_i \log y_i!$$

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

$$\partial E_0\,(1/n)\log L]/\partial\lambda = -1 + \lambda_0/\lambda = 0$$

which has solution $\lambda = \lambda_0$ which was to be shown.

8. The log likelihood for a sample from the normal distribution is

$$\text{LogL} = -(n/2)\log 2\pi - (n/2)\log\sigma^2 - 1/(2\sigma^2)\, \Sigma_i\,(y_i - \mu)^2.$$

$$E_0\,[(1/n)\log L] = -(1/2)\log 2\pi - (1/2)\log\sigma^2 - 1/(2\sigma^2)\, E_0[(1/n)\, \Sigma_i\,(y_i - \mu)^2].$$

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,

$$E_0\,[(1/n)\log L] = -(1/2)\log 2\pi - (1/2)\log\sigma^2 - 1/(2\sigma^2)[\,\sigma_0^2 + (\mu_0 - \mu)^2]$$

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 $-(1/2)\log 2\pi - (1/2)\log\sigma^2 - \sigma_0^2/(2\sigma^2)$. 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. 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

$$\partial\log L/\partial\eta = n/\eta - \Sigma_i\, y_i\,(\eta y_i - x_i'\delta) = 0$$

$$\partial\log L/\partial\delta = \Sigma_i\, x_i\,(\eta y_i - x_i'\delta) = 0$$

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

$\partial^2 \log L/\partial\eta^2 = -n/\eta^2 - \Sigma_i y_i^2$

$\partial^2 \log L/\partial\delta\,\partial\delta' = -\Sigma_i x_i x_i'$

$\partial^2 \log L/\partial\delta\,\partial\eta = \Sigma_i x_i y_i$.

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,

$E[\partial^2 \log L/\partial\eta^2|X] = -2n/\eta^2 - (1/\eta^2)\delta' X' X\delta$

$E[\partial^2 \log L/\partial\delta\,\partial\delta'|X] = -X'X$

$E[\partial^2 \log L/\partial\delta\,\partial\eta|X] = (1/\eta)X'X\delta$

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**.** 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 $\bar{\mathbf{x}}$. The variance of $\bar{\mathbf{x}}$ is given above, so the Wald statistic is simply $(\bar{\mathbf{x}} - \mu^0\mathbf{i})'\,\mathrm{Var}[(\bar{\mathbf{x}} - \mu^0\mathbf{i})]^{-1}(\bar{\mathbf{x}} - \mu^0\mathbf{i})$. Inserting the covariance matrix given above produces the suggested statistic.

11. The asymptotic variance of the MLE is, in fact, equal to the Cramer-Rao Lower Bound for the variance of a consistent, asymptotically normally distributed estimator, so this completes the argument.

In example 4.9, we proposed a regression with a gamma distributed disturbance,

$$y_i = \alpha + \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i$$

where,

$$f(\varepsilon_i) = [\lambda^P/\Gamma(P)]\, \varepsilon_i^{P-1} \exp(-\lambda\varepsilon_i),\ \varepsilon_i \geq 0,\ \lambda > 0,\ P > 2.$$

(The fact that ε_i is nonnegative will shift the constant term, as shown in Example 4.9. The need for the restriction on P will emerge shortly.) It will be convenient to assume the regressors are measured in deviations from their means, so $\Sigma_i\mathbf{x}_i = \mathbf{0}$. The OLS estimator of $\boldsymbol{\beta}$ remains unbiased and consistent in this model, with variance

$$\text{Var}[\mathbf{b}|\mathbf{X}] = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

where $\sigma^2 = \text{Var}[\varepsilon_i|\mathbf{X}] = P/\lambda^2$. [You can show this by using gamma integrals to verify that $E[\varepsilon_i|\mathbf{X}] = P/\lambda$ and $E[\varepsilon_i^2|\mathbf{X}] = P(P+1)/\lambda^2$. See B-39 and (E-1) in Section E2.3. A useful device for obtaining the variance is $\Gamma(P) = (P-1)\Gamma(P-1)$.] We will now show that in this model, there is a more efficient consistent estimator of $\boldsymbol{\beta}$. (As we saw in Example 4.9, the constant term in this regression will be biased because $E[\varepsilon_i|\mathbf{X}] = P/\lambda$; a estimates $\alpha+P/\lambda$. In what follows, we will focus on the slope estimators.

The log likelihood function is

$$\text{Ln } L = \sum_{i=1}^{n} P\ln\lambda - \ln\Gamma(P) + (P-1)\ln\varepsilon_i - \lambda\varepsilon_i$$

The likelihood equations are

$$\begin{aligned}
\partial \ln L/\partial\alpha &= \Sigma_i\, [-(P-1)/\varepsilon_i + \lambda] = 0,\\
\partial \ln L/\partial\boldsymbol{\beta} &= \Sigma_i\, [-(P-1)/\varepsilon_i + \lambda]\mathbf{x}_i = \mathbf{0},\\
\partial \ln L/\partial\lambda &= \Sigma_i\, [P/\lambda - \varepsilon_i] = 0,\\
\partial \ln L/\partial P &= \Sigma_i\, [\ln\lambda - \psi(P) - \varepsilon_i] = 0.
\end{aligned}$$

The function $\psi(P) = d\ln\Gamma(P)/dP$ is defined in Section E2.3.) To show that these expressions have expectation zero, we use the gamma integral once again to show that $E[1/\varepsilon_i] = \lambda/(P-1)$. We used the result $E[\ln\varepsilon_i] = \psi(P)-\lambda$ in Example 15.5. So show that $E[\partial\ln L/\partial\boldsymbol{\beta}] = \mathbf{0}$, we only require $E[1/\varepsilon_i] = \lambda/(P-1)$ because $\mathbf{x}_i$ and ε_i are independent. The second derivatives and their expectations are found as follows: Using the gamma integral once again, we find $E[1/\varepsilon_i^2] = \lambda^2/[(P-1)(P-2)]$. And, recall that $\Sigma_i\mathbf{x}_i = \mathbf{0}$. Thus, conditioned on $\mathbf{X}$, we have

$$\begin{aligned}
-E[\partial^2\ln L/\partial\alpha^2] &= E[\Sigma_i\,(P-1)(1/\varepsilon_i^2)] &&= n\lambda^2/(P-2),\\
-E[\partial^2\ln L/\partial\alpha\partial\boldsymbol{\beta}] &= E[\Sigma_i\,(P-1)(1/\varepsilon_i^2)\mathbf{x}_i] &&= \mathbf{0},\\
-E[\partial^2\ln L/\partial\alpha\partial\lambda] &= E[\Sigma_i\,(-1)] &&= -n,\\
-E[\partial^2\ln L/\partial\alpha\partial P] &= E[\Sigma_i\,(1/\varepsilon_i)] &&= n\lambda/(P-1),\\
-E[\partial^2\ln L/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'] &= E[\Sigma_i\,(P-1)(1/\varepsilon_i^2)\mathbf{x}_i\mathbf{x}_i'] &&= \Sigma_i\,[\lambda^2/(P-2)]\mathbf{x}_i\mathbf{x}_i' = [\lambda^2/(P-2)](\mathbf{X}'\mathbf{X}),\\
-E[\partial^2\ln L/\partial\lambda\partial\boldsymbol{\beta}] &= E[\Sigma_i\,(-1)\mathbf{x}_i] &&= \mathbf{0},\\
-E[\partial^2\ln L/\partial P\partial\boldsymbol{\beta}] &= E[\Sigma_i\,(1/\varepsilon_i)\mathbf{x}_i] &&= \mathbf{0},\\
-E[\partial^2\ln L/\partial\lambda^2] &= E[\Sigma_i\,(P/\lambda^2)] &&= nP/\lambda^2,\\
-E[\partial^2\ln L/\partial\lambda\partial P] &= E[\Sigma_i\,(1/\lambda)] &&= n/\lambda,\\
-E[\partial^2\ln L/\partial P^2] &= E[\Sigma_i\,\psi'(P)] &&= n\psi'(P).
\end{aligned}$$

Since the expectations of the cross partials with respect to $\boldsymbol{\beta}$ and the other parameters are all zero, it follows that the asymptotic covariance matrix for the MLE of $\boldsymbol{\beta}$ is simply

$$\text{Asy.Var}[\hat{\boldsymbol{\beta}}_{MLE}] = \{-E[\partial^2\ln L/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}']\}^{-1} = [(P-2)/\lambda^2](\mathbf{X}'\mathbf{X})^{-1}.$$

Recall, the asymptotic covariance matrix of the ordinary least squares estimator is

$$\text{Asy.Var}[\mathbf{b}] = [P/\lambda^2](\mathbf{X}'\mathbf{X})^{-1}.$$

(Note that the MLE is ill defined if P is less than 2.) Thus, the ratio of the variance of the MLE of any element of $\boldsymbol{\beta}$ to that of the corresponding element of **b** is $(P-2)/P$ which is the result claimed in Example 4.9.

Applications

1. a. For both probabilities, the symmetry implies that $1 - F(t) = F(-t)$. In either model, then,

$$\text{Prob}(y=1) = F(\text{t}) \text{ and } \text{Prob}(y=0) = 1 - F(\text{t}) = F(-t).$$

These are combined in $\text{Prob}(Y=y) = \text{F}[(2y_i-1)t_i]$ where $t_i = \mathbf{x}_i'\beta$. Therefore,

$$\ln L = \Sigma_i \ln F[(2y_i-1)\mathbf{x}_i'\beta]$$

b. $$\partial \text{lnL}/\partial\beta = \sum_{i=1}^{n} \frac{f[(2y_i-1)\mathbf{x}_i'\beta]}{F[(2y_i-1)\mathbf{x}_i'\beta]}(2y_i-1)\mathbf{x}_i = \mathbf{0}$$

where $f[(2y_i-1)\mathbf{x}_i'\beta]$ is the density function. For the logit model, $f = F(1-F)$. So, for the logit model,

$$\partial \text{lnL}/\partial\beta = \sum_{i=1}^{n} \{1 - F[(2y_i-1)\mathbf{x}_i'\beta]\}(2y_i-1)\mathbf{x}_i = \mathbf{0}$$

Evaluating this expression for $y_i = 0$, we get simply $-F(\mathbf{x}_i'\beta)\mathbf{x}_i$. When $y_i = 1$, the term is $[1- F(\mathbf{x}_i'\beta)]\mathbf{x}_i$. It follows that both cases are $[y_i - F(\mathbf{x}_i'\beta)]\mathbf{x}_i$, so the likelihood equations for the logit model are

$$\partial \text{lnL}/\partial\beta = \sum_{i=1}^{n} [y_i - \Lambda(\mathbf{x}_i'\beta)]\mathbf{x}_i = \mathbf{0}.$$

For the probit model, $F[(2y_i-1)\mathbf{x}_i'\beta] = \Phi[(2y_i-1)\mathbf{x}_i'\beta]$ and $\text{f}[(2y_i-1)\mathbf{x}_i'\beta] = \phi[(2y_i-1)\mathbf{x}_i'\beta]$, which does not simplify further, save for that the term 2y_i inside may be dropped since $\phi(\text{t}) = \phi(\text{-t})$. Therefore,

$$\partial \text{lnL}/\partial\beta = \sum_{i=1}^{n} \frac{\phi[(2y_i-1)\mathbf{x}_i'\beta]}{\Phi[(2y_i-1)\mathbf{x}_i'\beta]}(2y_i-1)\mathbf{x}_i = \mathbf{0}$$

c. For the logit model, the result is very simple.

$$\partial^2 \text{lnL}/\partial\beta\partial\beta' = \sum_{i=1}^{n} -\Lambda(\mathbf{x}_i'\beta)[1-\Lambda(\beta)]\mathbf{x}_i\mathbf{x}_i'.$$

For the probit model, the result is more complicated. We will use the result that

$$\text{d}\phi(\text{t})/\text{dt} = -\text{t}\phi(\text{t}).$$

It follows, then, that $\text{d}[\phi(\text{t})/\Phi(\text{t})]/\text{dt} = [-\phi(\text{t})/\Phi(\text{t})][\text{t} + \phi(\text{t})/\Phi(\text{t})]$. Using this result directly, it follows that

$$\partial^2 \text{lnL}/\partial\beta\partial\beta' = \sum_{i=1}^{n} -\left(\frac{\phi[(2y_i-1)\mathbf{x}_i'\beta]}{\Phi[(2y_i-1)\mathbf{x}_i'\beta]}\right)\left((2y_i-1)\mathbf{x}_i'\beta + \frac{\phi[(2y_i-1)\mathbf{x}_i'\beta]}{\Phi[(2y_i-1)\mathbf{x}_i'\beta]}\right)(2y_i-1)^2\mathbf{x}_i\mathbf{x}_i' = \mathbf{0}$$

This actually simplifies somewhat because $(2\text{y}_i-1)^2 = 1$ for both values of y_i and $\phi[(2y_i-1)\mathbf{x}_i'\beta] = \phi(\mathbf{x}_i'\beta)$

d. Denote by **H** the actual second derivatives matrix derived in the previous part. Then, Newton's method is

$$\hat{\beta}(j+1)=\hat{\beta}(j)-\left\{\mathbf{H}\left[\hat{\beta}(j)\right]\right\}^{-1}\left[\frac{\partial \ln L[\hat{\beta}(j)]}{\partial\hat{\beta}(j)}\right]$$

where the terms on the right hand side indicate first and second derivatives evaluated at the "previous" estimate of β.

e. The method of scoring uses the expected Hessian instead of the actual Hessian in the iterations. The methods are the same for the logit model, since the Hessian does not involve y_i. The methods are different for the probit model, since the expected Hessian does not equal the actual one. For the logit model

$$-[E(\mathbf{H})]^{-1} = \left\{\sum_{i=1}^{n} \Lambda(\mathbf{x}_i'\beta)[1-\Lambda(\beta)]\mathbf{x}_i\mathbf{x}_i'\right\}^{-1}$$

For the probit model, we need first to obtain the expected value. Do obtain this, we take the expected value, with Prob(y=0) = 1 - Φ and Prob(y=1) = Φ. The expected value of the ith term in the negative hessian is the expected value of the term,

$$\left(\frac{\phi[(2y_i-1)\mathbf{x}_i'\beta]}{\Phi[(2y_i-1)\mathbf{x}_i'\beta]}\right)\left((2y_i-1)\mathbf{x}_i'\beta+\frac{\phi[(2y_i-1)\mathbf{x}_i'\beta]}{\Phi[(2y_i-1)\mathbf{x}_i'\beta]}\right)\mathbf{x}_i\mathbf{x}_i'$$

This is

$$\Phi[-\mathbf{x}_i'\beta]\left(\frac{\phi[\mathbf{x}_i'\beta]}{\Phi[-\mathbf{x}_i'\beta]}\right)\left(-\mathbf{x}_i'\beta+\frac{\phi[\mathbf{x}_i'\beta]}{\Phi[-\mathbf{x}_i'\beta]}\right)\mathbf{x}_i\mathbf{x}_i' + \Phi[\mathbf{x}_i'\beta]\left(\frac{\phi[\mathbf{x}_i'\beta]}{\Phi[\mathbf{x}_i'\beta]}\right)\left(\mathbf{x}_i'\beta+\frac{\phi[\mathbf{x}_i'\beta]}{\Phi[\mathbf{x}_i'\beta]}\right)\mathbf{x}_i\mathbf{x}_i'$$

$$=\phi[\mathbf{x}_i'\beta]\left(-\mathbf{x}_i'\beta+\frac{\phi[\mathbf{x}_i'\beta]}{\Phi[-\mathbf{x}_i'\beta]}+\mathbf{x}_i'\beta+\frac{\phi[\mathbf{x}_i'\beta]}{\Phi[\mathbf{x}_i'\beta]}\right)\mathbf{x}_i\mathbf{x}_i'$$

$$=\phi[\mathbf{x}_i'\beta]\left(\frac{\phi[\mathbf{x}_i'\beta]}{\Phi[-\mathbf{x}_i'\beta]}+\frac{\phi[\mathbf{x}_i'\beta]}{\Phi[\mathbf{x}_i'\beta]}\right)\mathbf{x}_i\mathbf{x}_i'$$

$$=\left(\phi[\mathbf{x}_i'\beta]\right)^2\left(\frac{1}{\Phi[-\mathbf{x}_i'\beta]}+\frac{1}{\Phi[\mathbf{x}_i'\beta]}\right)\mathbf{x}_i\mathbf{x}'$$

$$=\left(\phi[\mathbf{x}_i'\beta]\right)^2\left(\frac{\Phi[\mathbf{x}_i'\beta]+\Phi[-\mathbf{x}_i'\beta]}{\Phi[-\mathbf{x}_i'\beta]\Phi[\mathbf{x}_i'\beta]}\right)\mathbf{x}_i\mathbf{x}'$$

$$=\left(\frac{\left(\phi[\mathbf{x}_i'\beta]\right)^2}{[1-\Phi(\mathbf{x}_i'\beta)]\Phi(\mathbf{x}_i'\beta)}\right)\mathbf{x}_i\mathbf{x}'$$

e.

```
?======================================================
? Application 16.1
?======================================================
Namelist ; x = one,age,educ,hsat,female,married $
LOGIT ; Lhs = Doctor ; Rhs = X $
Calc ; L1 = logl $
+---------------------------------------------+
```

```
| Binary Logit Model for Binary Choice        |
| Dependent variable               DOCTOR     |
| Number of observations            27326     |
| Log likelihood function       -16405.94     |
| Number of parameters                  6     |
| Info. Criterion: AIC =          1.20120     |
| Info. Criterion: BIC =          1.20300     |
| Restricted log likelihood     -18019.55     |
+---------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
---------+Characteristics in numerator of Prob[Y = 1]
 Constant|    1.82207669       .10763712    16.928   .0000
 AGE     |     .01235692       .00124643     9.914   .0000   43.5256898
 EDUC    |    -.00569371       .00578743     -.984   .3252   11.3206310
 HSAT    |    -.29276744       .00686076   -42.673   .0000   6.78542607
 FEMALE  |     .58376753       .02717992    21.478   .0000    .47877479
 MARRIED |     .03550015       .03173886     1.119   .2634    .75861817
```

f.

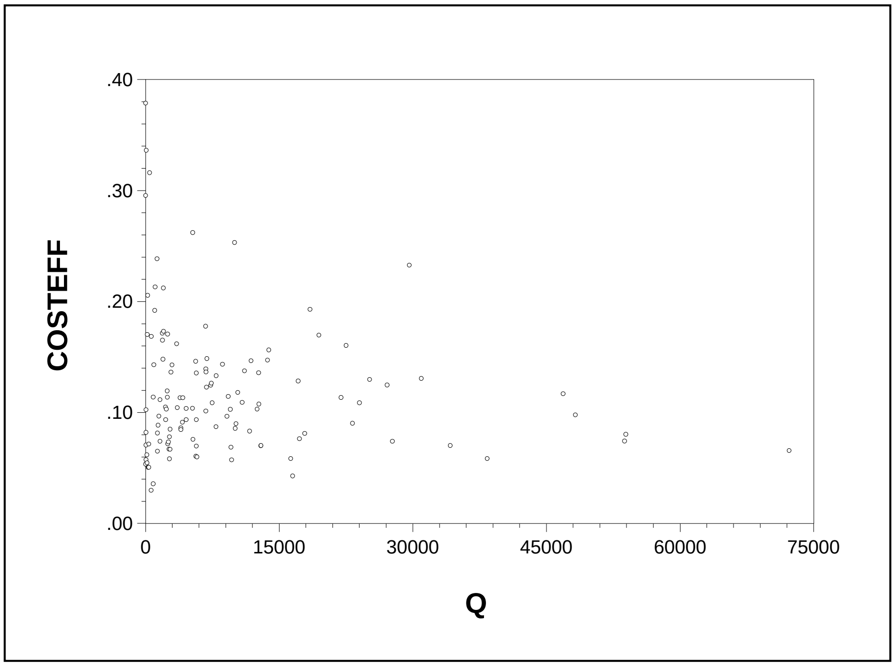
```
Matr ; bw = b(5:6) ; vw = varb(5:6,5:6) $
Matrix ; list ; WaldStat = bw'<vw>bw $
Calc ; list ; ctb(.95,2) $
LOGIT ; Lhs = Doctor ; Rhs = One,age,educ,hsat $
Calc ; L0 = logl $
Calc ; List ; LRStat = 2*(l1-l0) $
Matrix WALDSTAT has  1 rows and  1 columns.
               1
        +--------------
       1|  461.43784
--> Calc ; list ; ctb(.95,2) $
+------------------------------------+
| Listed Calculator Results          |
+------------------------------------+
 Result  =      5.991465
--> Calc ; L0 = logl $
--> Calc ; List ; LRStat = 2*(l1-l0) $
+------------------------------------+
| Listed Calculator Results          |
+------------------------------------+
 LRSTAT  =    467.336374
Logit ; Lhs = Doctor ; Rhs = X ; Start = b,0,0 ; Maxit = 0 $
+---------------------------------------------+
| Binary Logit Model for Binary Choice        |
| Maximum Likelihood Estimates                |
| Model estimated: May 17, 2007 at 11:49:42PM.|
| Dependent variable               DOCTOR     |
| Weighting variable                 None     |
| Number of observations            27326     |
| Iterations completed                  1     |
| LM Stat. at start values       466.0288     |
| LM statistic kept as scalar    LMSTAT       |
| Log likelihood function       -16639.61     |
| Number of parameters                  6     |
| Info. Criterion: AIC =          1.21830     |
|   Finite Sample: AIC =          1.21830     |
| Info. Criterion: BIC =          1.22010     |
| Info. Criterion:HQIC =          1.21888     |
| Restricted log likelihood     -18019.55     |
| McFadden Pseudo R-squared      .0765802     |
| Chi squared                    2759.883     |
| Degrees of freedom                    5     |
```

```
| Prob[ChiSqd > value] =         .0000000     |
| Hosmer-Lemeshow chi-squared =  23.44388     |
| P-value=  .00284 with deg.fr. =      8     |
+--------------------------------------------+
```

g. The restricted log likelihood given with the initial results equals -18019.55. This is the log likelihood for a model that contains only a constant term. The log likelihood for the model is -16405.94. Twice the difference is about 3,200, which vastly exceeds the critical chi squared with 5 degrees of freedom. The hypothesis would be rejected.

2. 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
```



.40
.30
.20
.10
.00
COSTEFF
0
15000
30000
45000
60000
75000
Q

Chapter 17

Simulation Based Estimation and Inference

Exercises

1. Exponential: The pdf is f(x) = θexp(-θx). The CDF is

$$F(x) = \int_0^x \theta \exp(-\theta t)dt = \theta\left[-\frac{1}{\theta}\exp(-\theta x) - \left(-\frac{1}{\theta}\exp(-\theta 0)\right)\right] = 1 - \exp(-\theta x).$$

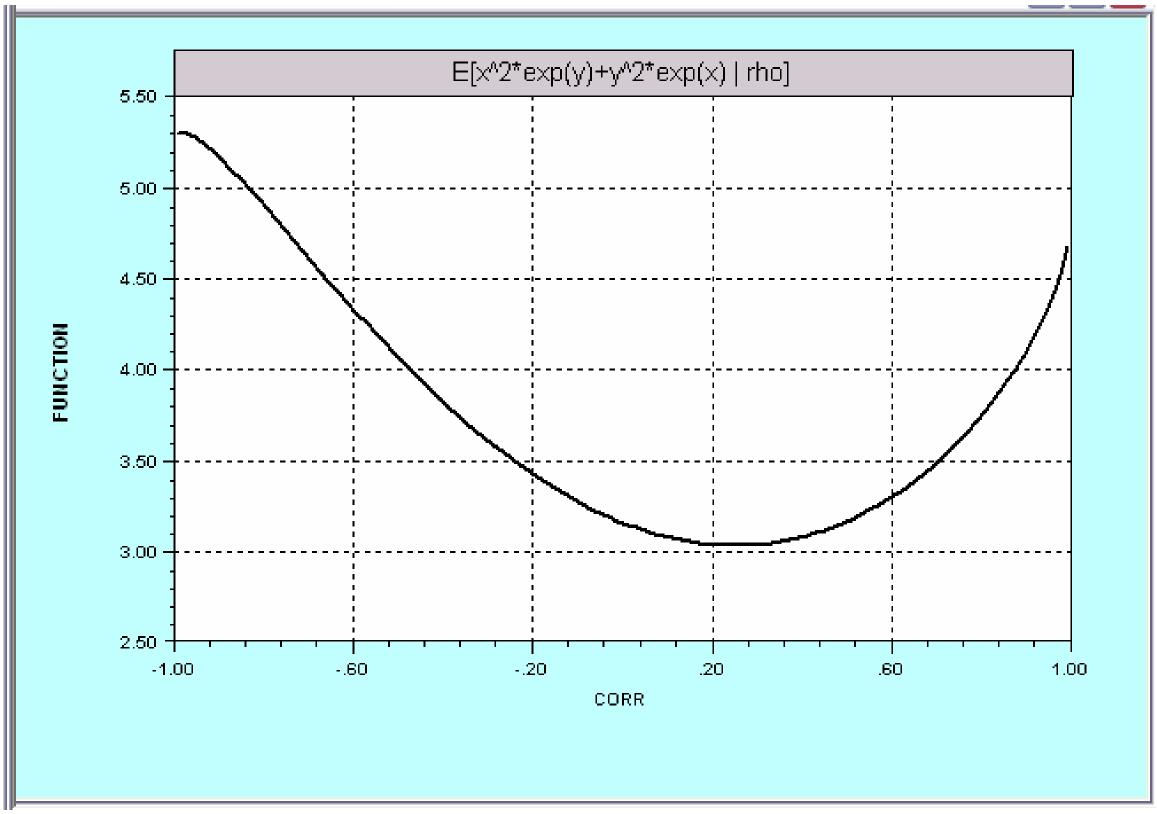
We would draw observations from the U(0,1) population, say F_i, and equate these to $F(x_i)$. Inverting the function, we find that $1-F_i = \exp(-\theta x_i)$, or $-(1/\theta)\ln(1-F_i) = x_i$. If x_i has an exponential density, then the density of $y_i = x_i^P$ is

Weibull. If the survival function is $S(x) = \lambda p\exp[-(\lambda x)^p]$, then we may equate random draws from the uniform distribution, S_i to this function (a draw of S_i is the same as a draw of $F_i = 1-S_i$). Solving for x_i, we find

$\ln S_i = \ln(\lambda p) - (\lambda x)^p$, so $x_i = (1/\lambda)[\ln(\lambda p) - \ln S_i]^{1/p}$.

2. We will need a bivariate sample on x and y to compute the random variable, then average the draws on it. The precise method of using a Gibbs sampler to draw this bivaraite sample is shown in Example 18.5. Once the bivariate sample of (x,y) is drawn, a large number of observations on $[x^2\exp(y)+y^2\exp(x)]$ is computed and averaged. As noted there, the Gibbs sampler is not much of a simplification for this particular problem. It is simple to draw a sample dircectly from a bivariate normal distribution. Here is a program that does the simulation and plots the estimate of the function

```
Calc   ; Ran(12345) $
Sample ; 1-1000$
Create ; xf=rnn(0,1) ; yfb=rnn(0,1) $
Matrix ; corr=init(100,1,0) ; function=corr $
Calc   ; i=0 $
Proc
Calc   ; i=i+1 $
Matrix ; corr(i)=ro $
Matrix ; c=[1/ro,1] ; c=chol(c) $
Create ; yf = c(2,1)*xf + c(2,2)*yfb $
Create ; fr=xf^2*exp(yf)+yf^2*exp(xf) $
Calc   ; ef = xbr(fr) ; ro=ro+.02 $
Matrix ; function(i)=ef $
Endproc $
Calc   ; ro=-.99 $
Execute; n=100 $
Mplot  ; Lhs = corr ; Rhs = Function ; Fill
       ; Grid ; Endpoints = -1,1
       ; Title=E[x^2*exp(y)+y^2*exp(x) | rho] $
```



Application

```
?================================================================
? Application 17.1.  Monte Carlo Simulation
?================================================================
? Set seed of RNG for replicability
Calc ; Ran(123579) $
? Sample size is 50.  Generate x(i) and z(i) held fixed
Sample ; 1 - 50 $
Create ; xi = rnn(0,1) ; zi = rnn(0,1) $
Namelist ; X = one,xi,zi ; X0 = one,xi $
? Moment Matrices
Matrix ; XXinv = <X'X> ; X0X0inv = <X0'X0> $
Matrix ; Waldi = init(1000,1,0) $
Matrix ; LMi = init(1000,1,0) $

?*****************************************************************
? Procedure studies the LM statistic
?*****************************************************************
Proc = LM (c) $
? Three kinds of disturbances
Create ?; Eps = Rnt(5) ? Nonnormal distribution
       ; vi=exp(.2*xi) ; eps = vi*rnn(0,1) ? Heteroscedasticity
       ?;eps= Rnn(0,1) ? Standard normal distribution
       ; y = 0 + xi + c*zi +eps $
Matrix ; b0 = X0X0inv*X0'y $
Create ; e0 = y - X0'b0 $
Matrix ; g = X'e0 $
Calc   ; lmstat = qfr(g,xxinv)/(e0'e0/n) ; i = i + 1 $
Matrix ; Lmi (i) = lmstat $
EndProc $
```

```
Calc ; i = 0 ; gamma = -1 $
Exec ; Proc=LM(gamma) ; n = 1000 $
samp;1-1000$
create;LMv=lmi $
create;reject=lmv>3.84$
Calc  ; List ; Type1 = xbr(reject) ; pwr = 1-Type1 $

?****************************************************************
? Procedure studies the Wald statistic
?****************************************************************
Proc = Wald(c) $
Create ; if(type=1)Eps = Rnn(0,1) ? Standard normal distribution
       ; if(type=2)vi=exp(.2*xi)   ? eps = vi*rnn(0,1) ? Heteroscedasticity
       ; if(type=3)eps= Rnt(5)     ? Nonnormal distribution
       ; y = 0 + xi + c*zi +eps $
Matrix ; b0=XXinv*X'y $
Create ; e0=y-X'b0$
Calc   ; ss0 = e0'e0/(47)
       ; v0 = ss0*xxinv(3,3)
       ; wald0=(b0(3))^2/v0
       ; i=i+1 $
Matrix ; Waldi(i)=Wald0 $
EndProc $
? Set the values for the simulation
Calc ; i = 0 ; gamma = 0 ; type=1 $
Sample ; 1-50 $
Exec ; Proc=Wald(gamma) ; n = 1000 $
samp;1-1000$
create;Waldv=Waldi $
create;reject=Waldv > 3.84$
Calc  ; List ; Type1 = xbr(reject) ; pwr = 1-Type1 $
```

To carry out the simulation, execute the procedure for different values of “gamma” and “type.” Summarize the results with a table or plot of the rejection probabilities as a function of gamma.

Chapter 18

Bayesian Estimation and Inference

Exercise

a. The likelihood function is

$$L(\mathbf{y}|\lambda) = \prod_{i=1}^{n} f(y_i \mid \lambda) = \prod_{i=1}^{n} \frac{\exp(-\lambda)\lambda^{y_i}}{\Gamma(y_i+1)} = \exp(-n\lambda)\lambda^{\Sigma_i y_i} \prod_{i=1}^{n} \frac{1}{\Gamma(y_i+1)}.$$

b. The posterior is

$$p(\lambda \mid y_1,...,y_n) = \frac{p(y_1,...,y_n \mid \lambda)\, p(\lambda)}{\int_0^\infty p(y_1,...,y_n \mid \lambda)\, p(\lambda) d\lambda}.$$

The product of factorials will fall out. This leaves

$$\begin{aligned} p(\lambda \mid y_1,...,y_n) &= \frac{\exp(-n\lambda)\lambda^{\Sigma_i y_i}(1/\lambda)}{\int_0^\infty \exp(-n\lambda)\lambda^{\Sigma_i y_i}(1/\lambda)d\lambda} \\ &= \frac{\exp(-n\lambda)\lambda^{(\Sigma_i y_i)-1}}{\int_0^\infty \exp(-n\lambda)\lambda^{(\Sigma_i y_i)-1}d\lambda} \\ &= \frac{\exp(-n\lambda)\lambda^{n\overline{y}-1}}{\int_0^\infty \exp(-n\lambda)\lambda^{n\overline{y}-1}d\lambda} \\ &= \frac{n^{n\overline{y}}\exp(-n\lambda)\lambda^{n\overline{y}-1}}{\Gamma(n\overline{y})}. \end{aligned}$$

where we have used the gamma integral at the last step. The posterior defines a two parameter gamma distribution, G(n, $n\overline{y}$).

c. The estimator of λ is the mean of the posterior. There is no need to do the integration. This falls simply out of the posterior density, E[λ|**y**] = $n\overline{y}/n = \overline{y}$.

d. The posterior variance also drops out simply; it is $n\overline{y}/n^2 = \overline{y}/n$.

Application

a. $p(F_i|K_i,\theta) = \binom{K_i}{F_i}\theta^{F_i}(1-\theta)^{K_i-F_i}$ so the log likelihood function is

$$\ln L(\theta\,|\,\mathbf{y}) = \sum_{i=1}^{n} \ln\binom{K_i}{F_i} + F_i \ln\theta + (K_i - F_i)\ln(1-\theta)$$

The MLE is obtained by setting $\partial \ln L(\theta|y)/\partial\theta = \Sigma_i\,[F_i/\theta - (K_i-F_i)/(1-\theta)] = 0$. Multiply both sides by $\theta(1-\theta)$ to obtain

$\Sigma_i\,[(1-\theta)F_i - \theta\,(K_i-F_i)] = 0$

A line of algebra reveals that the solution is $\theta = (\Sigma_i F_i)/(\Sigma_i K_i) = 0.651596$.

b. The posterior density is $$\frac{\left[\prod_{i=1}^{n}\binom{K_i}{F_i}\theta^{F_i}(1-\theta)^{K_i-F_i}\right]\dfrac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}}{\int_0^1\left[\prod_{i=1}^{n}\binom{K_i}{F_i}\theta^{F_i}(1-\theta)^{K_i-F_i}\right]\dfrac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}d\theta}$$

This simplifies considerably. The combinatorials and gamma functions fall out, leaving

$$p(\theta\,|\,\mathbf{y}) = \frac{\left[\prod_{i=1}^{n}\theta^{F_i}(1-\theta)^{K_i-F_i}\right]\theta^{a-1}(1-\theta)^{b-1}}{\int_0^1\left[\prod_{i=1}^{n}\theta^{F_i}(1-\theta)^{K_i-F_i}\right]\theta^{a-1}(1-\theta)^{b-1}d\theta} = \frac{\left[\theta^{\Sigma_i F_i}(1-\theta)^{\Sigma_i(K_i-F_i)}\right]\theta^{a-1}(1-\theta)^{b-1}}{\int_0^1\left[\theta^{\Sigma_i F_i}(1-\theta)^{\Sigma_i(K_i-F_i)}\right]\theta^{a-1}(1-\theta)^{b-1}d\theta}$$
$$= \frac{\left[\theta^{(\Sigma_i F_i)+(a-1)}(1-\theta)^{[\Sigma_i(K_i-F_i)]+(b-1)}\right]}{\int_0^1\left[\theta^{(\Sigma_i F_i)+(a-1)}(1-\theta)^{\Sigma_i(K_i-F_i)]+(b-1)}\right]d\theta}$$

The denominator is a beta integral, so the posterior density is
$$p(\theta\,|\,\mathbf{y}) = \frac{\Gamma[(\Sigma_i F_i)+(a-1)]\Gamma[(\Sigma_i(K_i-F_i))+(b-1)]}{\Gamma[(\Sigma_i F_i)+(a-1)+(\Sigma_i(K_i-F_i))+(b-1)]}\left[\theta^{(\Sigma_i F_i)+(a-1)}(1-\theta)^{[\Sigma_i(K_i-F_i)]+(b-1)}\right]$$

The denominator simplifies slightly;

$$p(\theta\,|\,\mathbf{y}) = \frac{\Gamma[(\Sigma_i F_i)+(a-1)]\Gamma[(\Sigma_i(K_i-F_i))+(b-1)]}{\Gamma[(\Sigma_i K_i)+(a-1)+(b-1)]}\left[\theta^{(\Sigma_i F_i)+(a-1)}(1-\theta)^{[\Sigma_i(K_i-F_i)]+(b-1)}\right]$$
$$= \frac{\Gamma[(a+\Sigma_i F_i)-1)]\Gamma[(b+\Sigma_i(K_i-F_i))-1)]}{\Gamma[(a+b)+(\Sigma_i K_i)-1-1)]}\left[\theta^{(a+\Sigma_i F_i)-1}(1-\theta)^{[b+\Sigma_i(K_i-F_i)]-1}\right]$$

c-e. The posterior distribution is a beta distribution with parameters $a^*=(a+\Sigma_i F_i)$ and $b^*=[b+\Sigma_i(K_i-F_i)]$. The mean of this beta random variable is $a^*/(a^*+b^*) = (a+\Sigma_i F_i)/(a+b+\Sigma_i K_i)$. In the data, $\Sigma_i = 49$ and $\Sigma_i K_i = 75$. For the values given, the posterior means are

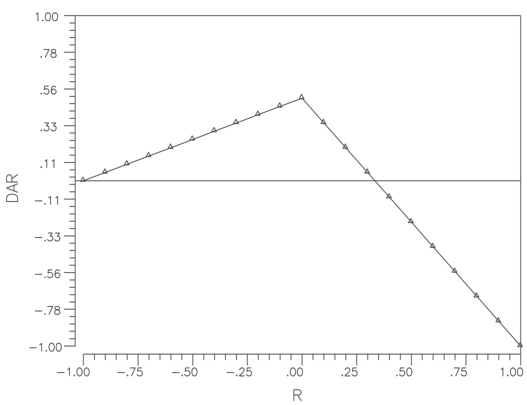
```
(a=1,b=1): Result  =      .647668
(a=2,b=2): Result  =      .643939
(a=1,b=2): Result  =      .639386
```

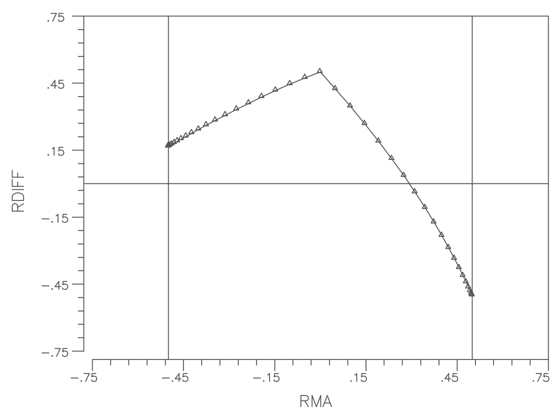
Chapter 19

Serial Correlation

Exercises

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 $y_t = \boldsymbol{\beta}'\mathbf{x}_t + \varepsilon_t$, $\varepsilon_t = \rho\varepsilon_{t-1} + \mathrm{u}_\mathbf{t} - \lambda \mathrm{u}_{t-1}$. 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

$$\varepsilon_t = u_t + \rho u_{t-1} + \rho^2 u_{t-2} + \rho^3 u_{t-3} + ... - \lambda u_{t-1} - \rho\lambda u_{t-2} - \rho^2\lambda u_{t-3} - ...$$
$$= u_t + (\rho-\lambda)u_{t-1} + \rho(\rho-\lambda)u_{t-2} + \rho^2(\rho-\lambda)u_{t-3} + ...$$

Therefore,

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

To evaluate this expression, write

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

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. Since the regression contains a lagged dependent variable, we cannot use the Durbin-Watson statistic directly. The h statistic in (15-34) would be $h = (1 - 1.21/2)[21 / (1 - 21(.18^2)]^{1/2} = 3.201$. 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 $c_1 = \theta_1c_0 + \theta_2c_1$ 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 $\rho_1 = \text{Cov}[\varepsilon_t,\varepsilon_{t-1}]/\text{Var}[\varepsilon_t] = -\lambda\text{Var}[u_{t-1}]/\{\text{Var}[u_t] + \lambda^2\text{Var}[u_{t-1}]\} = -\lambda/(1 + \lambda^2)$.

Applications

1. Phillips Curve

```
--> 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.

2. (Improved Phillips curve model)

```
--> 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
```

3. (GARCH Models)

.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 20

Models with Lagged Variables

Exercises

1. 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. 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.

3. 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.

4. 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. ~

5. 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. ~

Applications

1. 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,

$$E\left[\begin{matrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{matrix} \middle| X_1\right] = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix} + \begin{bmatrix} 1 & r & r^2 & r^3 & r^4 & r^5 \\ r & 1 & r & r^2 & r^3 & r^4 \\ r^2 & r & 1 & r & r^2 & r^3 \\ r^3 & r^2 & r & 1 & r & r^2 \\ r^4 & r^3 & r^2 & r & 1 & r \\ r^5 & r^4 & r^3 & r^2 & r & 1 \end{bmatrix}^{-1} \begin{bmatrix} r^6 \\ r^5 \\ r^4 \\ r^3 \\ r^2 \\ r \end{bmatrix} \beta_6$$

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,

$$E\left[\begin{matrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{matrix} \middle| X_1\right] = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ r \end{bmatrix} \beta_6.$$

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.

2. 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)

Chapter 21

Time Series Models

There are no exercises or applications in Chapter 21.

Chapter 22

Nonstationary Data

Exercise

1. 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 in Section 21.2.3 that the partial autocorrelations for the MA(2) process mirror tha autocorrelations for an AR(2). Thus, the results in Section 21.2.3 for the AR(2) can be used directly.

Applications

1. ADF 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
```

The unit root hypothesis is definitely not rejected.

2. Macroeconomic Model

```
--> 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 Result   has  2 rows and  2 columns.
               1             2
        +---------------------------
       1|     1.0688     .0000000D+00
       2|    19.8378     .0000000D+00
```

Short run multiplier is $\beta = .07869$. Long run is $\beta/(1-\gamma_1 - \gamma_2) = 12.669$. (Not very plausible.)

3. ADF Test. 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.

4. Reestimated model in example 13.1.

```
--> 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
--> 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

+---------------------------------------------------------------------+
| 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 23
Models for Discrete Choice

Exercises

1. 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.

					λ_1	λ_0	
y	d	$\alpha+\beta d$	Φ	ϕ	ϕ/Φ	$-\phi/(1-\Phi)$	$\lambda_0\lambda_1$
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 $-E[\mathbf{H}]$.

$$-\hat{\mathbf{H}} = 24(.636)\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix} + 28(.636)\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix} + 32(.597)\begin{bmatrix}1 & 1\\1 & 1\end{bmatrix} + 16(.597)\begin{bmatrix}1 & 1\\1 & 1\end{bmatrix}.\text{ Then,}$$

$$\left[-\hat{\mathbf{H}}\right]^{-1} = \begin{bmatrix}61.728 & 28.656\\28.656 & 28.656\end{bmatrix}^{-1} = \begin{bmatrix}.03024 & -.03024\\-.03024 & .06513\end{bmatrix}.$$ 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

$$\begin{aligned}
\text{Prob}[y = 0 , d = 0] &= P_{00} = 1/(1 + e^{.156}) &&= .462\\
\text{Prob}[y = 1 , d = 0] &= P_{10} = 1 - P_{00} &&= .538\\
\text{Prob}[y = 0 , d = 1] &= P_{01} = 1/(1 + e^{-.431}) &&= .667\\
\text{Prob}[y = 1 , d = 1] &= P_{11} = 1 - P_{01} &&= .333.
\end{aligned}$$

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. 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. The model was estimated using Newton's method as described in the text. The estimated coefficients and their standard are shown below:

$$\hat{y}^* = -.51274 + .15964X$$
$$(1.042) \qquad (.202)$$

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. The derivatives of the log-likelihood are given in (23-18)-(23-21). 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 $P = \dfrac{\exp(\hat{\alpha})}{1+\exp(\hat{\alpha})}$, or $\hat{\alpha} = \text{logit}(P)$. The LM statistic based on the BHHH estimator of the covariance matrix of the first derivatives would be $LM = [\Sigma_i\mathbf{g}_i]'[\Sigma_i\mathbf{g}_i\mathbf{g}_i']^{-1}[\Sigma_i\mathbf{g}_i]$ where $\mathbf{g}_i = \Sigma_i(y_i - P)\mathbf{x}_i$. In full, the statistic is

$$LM = [\Sigma_i(y_i - P)\mathbf{x}_i]'[\Sigma_i(y_i - P)^2\mathbf{x}_i\mathbf{x}_i']^{-1}[\Sigma_i(y_i - P)\mathbf{x}_i].$$

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

$$LM = [\mathbf{X}'(\mathbf{y} - P\mathbf{i})]'[(\mathbf{X}'\mathbf{X})^{-1}][\mathbf{X}'(\mathbf{y} - P\mathbf{i})]/P(1 - P) = \mathbf{e}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}/P(1 - P)$$

In the preceding, $\mathbf{e}'\mathbf{e} = \Sigma_i(y_i - P)^2 = nP(1 - P)$. Therefore, $LM = n[\mathbf{e}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}/\mathbf{e}'\mathbf{e}]$, 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. Since there is no regressor, we may write the log-likelihood as

$$\ln L = 50\ln\Phi(-\alpha) + 40\ln[\Phi(\mu_1-\alpha) - \Phi(-\alpha)] + 45\ln[\Phi(\mu_2-\alpha) - \Phi(\mu_1-\alpha)] +$$

$$80\ln[\Phi(\mu_3-\alpha) - \Phi(\mu_2-\alpha)] + 35\ln[1 - \Phi(\mu_3-\alpha)].$$

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,

$$\begin{aligned}
90\pi_0 - 50\pi_1 &= 0\\
45\pi_0 - 85\pi_1 + 40\pi_2 &= 0\\
80\pi_1 - 125\pi_2 + 45\pi_3 &= 0\\
- 35\pi_2 + 115\pi_3 &= 80
\end{aligned}$$

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. 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. 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. 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, $n_0f(\alpha)/[1-F(\alpha)] = n_1f(\alpha)/F(\alpha)$. 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 (23-28) follows immediately.

9. 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.

Applications

1. Binary Choice for Extramarital Affairs using Redbook data

```
?============================================================
? Application 23.1
?============================================================
?
Create ; A = (Yrb > 0) $
Namelist ; X = one,v1,v2,v5,v6 $
Probit ; Lhs = A ; Rhs = X ; marginal Effects $
Logit  ; Lhs = A ; Rhs = X ; marginal Effects $
+---------------------------------------------+
| Binomial Probit Model                       |
| Maximum Likelihood Estimates                |
| Dependent variable                   A      |
| Number of observations            6366      |
| Log likelihood function      -3547.865      |
| Number of parameters                 5      |
| Info. Criterion: AIC =         1.11620      |
| Info. Criterion: BIC =         1.12151      |
| Restricted log likelihood    -4002.530      |
+---------------------------------------------+
+--------+-------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+-------------+----------------+--------+--------+----------+
---------+Index function for probability
```

```
 Constant|    1.43453507       .15493583     9.259   .0000
 V1      |    -.42595261       .01807583   -23.565   .0000   4.10964499
 V2      |     .02797013       .00254409    10.994   .0000   29.0828621
 V5      |    -.20942202       .02015534   -10.390   .0000   2.42617028
 V6      |    -.03522668       .00801808    -4.393   .0000   14.2098649
+----------------------------------------+
| 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]|Elasticity|
+--------+-------------+--------------+--------+--------+----------+
---------+
 Constant|     .27876593       .01081795    25.769   .0000
 V1      |    -.14911732       .00634679   -23.495   .0000  -2.01181601
 V2      |     .00979177       .00088860    11.019   .0000    .93487672
 V5      |    -.07331438       .00703451   -10.422   .0000   -.58393740
 V6      |    -.01233214       .00280535    -4.396   .0000   -.57528664
+--------------------------------------------+
| Binary Logit Model for Binary Choice        |
| Maximum Likelihood Estimates                |
| Dependent variable                    A     |
| Number of observations             6366     |
| Log likelihood function       -3549.741     |
| Number of parameters                  5     |
| Info. Criterion: AIC =          1.11679     |
| Info. Criterion: BIC =          1.12210     |
| Restricted log likelihood     -4002.530     |
+--------------------------------------------+
+--------+-------------+--------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+-------------+--------------+--------+--------+----------+
---------+Characteristics in numerator of Prob[Y = 1]
 Constant|    2.41622262       .26160831     9.236   .0000
 V1      |    -.70802698       .03091557   -22.902   .0000   4.10964499
 V2      |     .04624150       .00426119    10.852   .0000   29.0828621
 V5      |    -.35139771       .03413337   -10.295   .0000   2.42617028
 V6      |    -.05899324       .01354756    -4.355   .0000   14.2098649
+----------------------------------------+
| Partial derivatives of probabilities with |
| respect to the vector of characteristics. |
| They are computed at the means of the Xs. |
| Observations used are All Obs.            |
+----------------------------------------+
+--------+-------------+--------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]|Elasticity|
+--------+-------------+--------------+--------+--------+----------+
---------+Marginal effect for variable in probability
 Constant|     .50898166       .05554126     9.164   .0000
 V1      |    -.14914716       .00650799   -22.918   .0000  -2.03205673
 V2      |     .00974086       .00089378    10.898   .0000    .93918419
 V5      |    -.07402256       .00714156   -10.365   .0000   -.59539053
 V6      |    -.01242703       .00285019    -4.360   .0000   -.58542862
```

2. Ordered Choice For Self Reported Marriage Rating

```
+--------------------------------------------+
| Ordered Probability Model                   |
| Maximum Likelihood Estimates                |
| Dependent variable             MARRIAGE     |
| Weighting variable                 None     |
```

```
| Number of observations             6366     |
| Iterations completed                 15     |
| Log likelihood function       -7720.145     |
| Number of parameters                 12     |
| Info. Criterion: AIC =          2.42920     |
| Info. Criterion: BIC =          2.44194     |
| Restricted log likelihood     -7926.487     |
| Underlying probabilities based on Normal    |
+---------------------------------------------+
+---------------------------------------------+
| Ordered Probability Model                   |
| Cell frequencies for outcomes               |
|  Y Count Freq  Y Count Freq  Y Count Freq   |
|  0    99 .015  1   348 .054  2   993 .155   |
|  3  2242 .352  4  2684 .421                 |
+---------------------------------------------+
+--------+-------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+-------------+----------------+--------+--------+----------+
---------+Index function for probability
 Constant|    1.87997564       .12760529    14.733   .0000
 YRB     |    -.09669427       .00649907   -14.878   .0000    .70537389
 V2      |    -.00624520       .00471646    -1.324   .1855   29.0828621
 V3      |    -.00952932       .00506534    -1.881   .0599   9.00942507
 V4      |    -.05879586       .01520251    -3.868   .0001   1.39687402
 V5      |     .10524384       .01624338     6.479   .0000   2.42617028
 V6      |     .02526318       .00727002     3.475   .0005   14.2098649
 V7      |     .02069865       .01614318     1.282   .1998   3.42412818
 V8      |     .02725715       .01072244     2.542   .0110   3.85014138
---------+Threshold parameters for index
 Mu(1)   |     .71088354       .02219910    32.023   .0000
 Mu(2)   |    1.47186849       .01737814    84.697   .0000
 Mu(3)   |    2.46392113       .01923976   128.064   .0000

+--------------------------------------------------------------------+
| Summary of Marginal Effects for Ordered Probability Model (probit)      |
+--------------------------------------------------------------------+
Variable|    Y=00    Y=01    Y=02    Y=03    Y=04    Y=05    Y=06    Y=07 |
--------------------------------------------------------------------+
YRB         .0031   .0087   .0167   .0093  -.0377
V2          .0002   .0006   .0011   .0006  -.0024
V3          .0003   .0009   .0016   .0009  -.0037
V4          .0019   .0053   .0101   .0056  -.0229
V5         -.0033  -.0095  -.0182  -.0101   .0411
V6         -.0008  -.0023  -.0044  -.0024   .0099
V7         -.0007  -.0019  -.0036  -.0020   .0081
V8         -.0009  -.0025  -.0047  -.0026   .0106
+----------------------------------------------------------------------+
|   Cross tabulation of predictions. Row is actual, column is predicted.    |
|   Model = Probit    .  Prediction is number of the most probable cell.    |
+-------+-------+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| Actual|Row Sum|  0  |  1  |  2  |  3  |  4  |  5  |  6  |  7  |  8  |  9  |
+-------+-------+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|      0|     99|    0|    0|    0|   68|   31|
|      1|    348|    2|    0|    5|  170|  171|
|      2|    993|    7|    0|    7|  453|  526|
|      3|   2242|    3|    0|   10|  674| 1555|
|      4|   2684|    2|    0|    5|  593| 2084|
+-------+-------+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|Col Sum|   6366|   14|    0|   27| 1958| 4367|    0|    0|    0|    0|    0|
+-------+-------+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
```

Chapter 24

Truncation, Censoring and Sample Selection

Exercises

1. 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. The log-likelihood for the Tobit model is given in (24-13). 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. The log-likelihood for the truncated regression with only a constant term is

$$\ln L = (-n/2)[\ln(2\pi) + \ln\sigma^2] - (1/(2\sigma^2))\Sigma_1(y_i - \mu)^2 - \Sigma_i\ln\Phi(\mu/\sigma)$$

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 (24-6). 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. Using Theorem 24.5, 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. The conditional mean function is $E[y|\mathbf{x}] = \Phi(\boldsymbol{\beta}'\mathbf{x}_i/\sigma_i)\boldsymbol{\beta}'\mathbf{x}_i + \sigma_i\Phi(\boldsymbol{\beta}'\mathbf{x}_i/\sigma_i)$ using the equation before (24-12). Suppose that $\sigma_i = \sigma\exp(\boldsymbol{\alpha}'\mathbf{x}_i)$ for the same vector $\mathbf{x}_i$. (We'll relax that assumption shortly.) Now, differentiate this expression with respect to $\mathbf{x}$. We differentiate the two parts, first with respect to $\boldsymbol{\beta}'\mathbf{x}$ then with respect to σ_i.

$$\frac{\partial E[y_i|\mathbf{x}_i]}{\partial \mathbf{x}_i} = \Phi\left(\frac{\boldsymbol{\beta}'\mathbf{x}_i}{\sigma_i}\right)\boldsymbol{\beta} + \left(\boldsymbol{\beta}'\mathbf{x}_i\right)\phi\left(\frac{\boldsymbol{\beta}'\mathbf{x}_i}{\sigma_i}\right)\frac{1}{\sigma_i}\boldsymbol{\beta} + \sigma_i\left[-\left(\frac{\boldsymbol{\beta}'\mathbf{x}_i}{\sigma_i}\right)\phi\left(\frac{\boldsymbol{\beta}'\mathbf{x}_i}{\sigma_i}\right)\right]\frac{1}{\sigma_i}\boldsymbol{\beta}$$

$$+\left(\boldsymbol{\beta}'\mathbf{x}_i\right)\phi\left(\frac{\boldsymbol{\beta}'\mathbf{x}_i}{\sigma_i}\right)\left(\frac{-1}{\sigma_i}\right)\left(\frac{\boldsymbol{\beta}'\mathbf{x}_i}{\sigma_i}\right)\sigma_i\boldsymbol{\alpha} + \phi\left(\frac{\boldsymbol{\beta}'\mathbf{x}_i}{\sigma_i}\right)\sigma_i\boldsymbol{\alpha} + \sigma_i\left[-\left(\frac{\boldsymbol{\beta}'\mathbf{x}_i}{\sigma_i}\right)\phi\left(\frac{\boldsymbol{\beta}'\mathbf{x}_i}{\sigma_i}\right)\right]\left(\frac{-1}{\sigma_i}\right)\left(\frac{\boldsymbol{\beta}'\mathbf{x}_i}{\sigma_i}\right)\sigma_i\boldsymbol{\alpha}$$

After collecting the terms, we obtain $\partial E[y_i|\mathbf{x}_i]/\partial\mathbf{x}_i = \Phi(a_i)\boldsymbol{\beta} + \sigma_i\phi(a_i)\boldsymbol{\alpha}$ where $a_i = \boldsymbol{\beta}'\mathbf{x}_i/\sigma_i$. Thus, the marginal effect has two parts. one for $\boldsymbol{\beta}$ and one for α. Now, if a variable appears in σ_i but not in $\mathbf{x}_i$, then only the second term appears while if a variable appears only in $\mathbf{x}_i$ and not in σ_i, then only the first term appears in the marginal effect.

6. The transformed log likelihood function is

$$\log L = \Sigma_{y>0}\,(-1/2)[\log 2\pi - \log\theta^2 + (\theta y - \mathbf{x}'\boldsymbol{\gamma})^2] + \Sigma_{y=0}\log[1-\Phi(\mathbf{x}'\boldsymbol{\gamma})]$$

It will be convenient to define $a_i = \mathbf{x}_i'\boldsymbol{\gamma}$. Note also that $1 - \Phi(a_i) = \Phi(-a_i)$. The first derivatives and Hessian in the transformed parameters are

$$\frac{\partial \log L}{\partial \theta} = \sum\nolimits_{y_i>0}\ (1/\theta) - y_i\left(\theta y_i - a_i\right)$$

$$\frac{\partial \log L}{\partial \boldsymbol{\gamma}} = \sum\nolimits_{y_i>0}\ \mathbf{x}_i\left(\theta y_i - a_i\right) + \sum\nolimits_{y_i=0}\ \left[\phi(-a_i)/\Phi(-a_i)\right](-\mathbf{x}_i)$$

$$\frac{\partial^2 \log L}{\partial \theta^2} = \sum\nolimits_{y_i>0}\ -1/\theta^2 - y_i^{\,2}$$

$$\frac{\partial^2 \log L}{\partial \boldsymbol{\gamma}\partial \boldsymbol{\gamma}'} = \sum\nolimits_{y_i>0}\ -\mathbf{x}_i\mathbf{x}_i' + \sum\nolimits_{y_i=0}\ -\left[\phi(-a_i)/\Phi(-a_i)\right]\{-a_i + \left[\phi(-a_i)/\Phi(-a_i)\right]\}\mathbf{x}_i\mathbf{x}_i'$$

$$\frac{\partial^2 \log L}{\partial \boldsymbol{\gamma}\partial \theta} = \sum\nolimits_{y_i>0}\ -\mathbf{x}_i y_i$$

The second derivatives can be collected in a matrix format:

$$\frac{\partial \log L}{\partial\begin{pmatrix}\boldsymbol{\gamma}\\ \theta\end{pmatrix}\partial\begin{pmatrix}\boldsymbol{\gamma}\\ \theta\end{pmatrix}'} = \sum\nolimits_{y>0}\ \left[-\begin{pmatrix}\mathbf{x}_i\\ -y_i\end{pmatrix}\begin{pmatrix}\mathbf{x}_i\\ -y_i\end{pmatrix}' - \begin{pmatrix}0\\ \theta\end{pmatrix}\begin{pmatrix}0\\ \theta\end{pmatrix}'\right] + \sum\nolimits_{y=0}\ \delta_i\begin{pmatrix}\mathbf{x}_i\\ 0\end{pmatrix}\begin{pmatrix}\mathbf{x}_i\\ 0\end{pmatrix}'$$

where δ_i is the last scalar term in $\partial^2\log L/\partial\delta\partial\boldsymbol{\gamma}'$. By Theorem 22.2 (see (24-4)), 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.

Applications

1. Tobit model for Redbook data

```
?================================================================
? Applications in Chapter 24
?================================================================
? 1.  Tobit, Scaled Tobit, Probit and Truncated Regression.
?     In principle, all are estimating the same paramter.
? For consistency and convenience, we are going to use the
? sample with YRB <= 5 only.
?================================================================
Sample ; All $
Reject ; YRB > 5 $
Namelist ; X = one,v1,v2,v3,v4,v5$
Tobit ; Lhs = yrb ; Rhs = x ; marginal $
Matrix ; list ;  scaled_b = 1/s * b $
Probit ; Lhs = a ; Rhs = x $
reject ; yrb <= 0 $
Truncation ; Lhs = yrb ; Rhs = x $

+---------------------------------------------+
| Limited Dependent Variable Model - CENSORED |
| Maximum Likelihood Estimates                |
| Dependent variable                  YRB     |
| Weighting variable                 None     |
| Number of observations             6217     |
| Iterations completed                  6     |
| Log likelihood function       -6118.089     |
| Number of parameters                  7     |
| Info. Criterion: AIC =          1.97043     |
|   Finite Sample: AIC =          1.97044     |
| Info. Criterion: BIC =          1.97802     |
| Info. Criterion:HQIC =          1.97306     |
| Threshold values for the model:             |
| Lower=     .0000     Upper=+infinity        |
| LM test [df] for tobit=    622.887[  6]     |
| Normality Test, LM    =    150.850[  2]     |
| ANOVA  based fit measure =    .293201       |
| DECOMP based fit measure =    .438743       |
+---------------------------------------------+
+--------+-------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+-------------+----------------+--------+--------+----------+
---------+Primary Index Equation for Model
 Constant|    4.13828429       .31908252    12.969   .0000
 V1      |    -.80415431       .03782416   -21.260   .0000   4.12272800
 V2      |    -.06923599       .01229186    -5.633   .0000   29.1829661
 V3      |     .10402446       .01325380     7.849   .0000   9.12329098
 V4      |    -.02190617       .03898707     -.562   .5742   1.41499115
 V5      |    -.43110692       .04356398    -9.896   .0000   2.43670581
---------+Disturbance standard deviation
 Sigma   |    2.27697641       .04212836    54.049   .0000
+-----------------------------------------+
| Partial derivatives of expected val. with |
| respect to the vector of characteristics. |
| They are computed at the means of the Xs. |
| Observations used for means are All Obs.  |
| Conditional Mean at Sample Point    .3941 |
| Scale Factor for Marginal Effects   .2796 |
+-----------------------------------------+
+--------+-------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
```

```
+--------+-------------+---------------+--------+--------+----------+
 Constant|    1.15697490       .09110678    12.699   .0000
 V1      |    -.22482418       .01048093   -21.451   .0000   4.12272800
 V2      |    -.01935689       .00342807    -5.647   .0000   29.1829661
 V3      |     .02908299       .00367661     7.910   .0000   9.12329098
 V4      |    -.00612449       .01090115     -.562   .5742   1.41499115
 V5      |    -.12052818       .01207702    -9.980   .0000   2.43670581
 Sigma   |       .000000    ......(Fixed Parameter).......
Matrix SCALED_B has  6 rows and  1 columns.
               1
        +--------------
       1|    1.81745
       2|    -.35317
       3|    -.03041
       4|     .04569
       5|    -.00962
       6|    -.18933
+---------------------------------------------+
| Binomial Probit Model                       |
| Maximum Likelihood Estimates                |
| Dependent variable                  A       |
| Weighting variable               None       |
| Number of observations           6217       |
| Iterations completed                5       |
| Log likelihood function       -3310.310     |
| Number of parameters                6       |
| Info. Criterion: AIC =          1.06685     |
| Info. Criterion: BIC =          1.07335     |
| Restricted log likelihood     -3830.126     |
+---------------------------------------------+
+--------+-------------+---------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+-------------+---------------+--------+--------+----------+
---------+Index function for probability
 Constant|    2.03641060       .15678428    12.989   .0000
 V1      |    -.41449474       .01860450   -22.279   .0000   4.12272800
 V2      |    -.03568737       .00593540    -6.013   .0000   29.1829661
 V3      |     .07215336       .00640693    11.262   .0000   9.12329098
 V4      |    -.00241124       .01891503     -.127   .8986   1.41499115
 V5      |    -.21212886       .02089864   -10.150   .0000   2.43670581
+---------------------------------------------+
| Limited Dependent Variable Model - TRUNCATE |
| Maximum Likelihood Estimates                |
| Dependent variable                YRB       |
| Weighting variable               None       |
| Number of observations           1904       |
| Iterations completed                8       |
| Log likelihood function       -2437.473     |
| Number of parameters                7       |
| Info. Criterion: AIC =          2.56772     |
| Info. Criterion: BIC =          2.58813     |
| Threshold values for the model:             |
| Lower=     .0000     Upper=+infinity        |
| Observations after truncation      1904     |
+---------------------------------------------+
+--------+-------------+---------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+-------------+---------------+--------+--------+----------+
---------+Primary Index Equation for Model
 Constant|    5.22651388       .94010948     5.559   .0000
 V1      |    -.45753380       .10715203    -4.270   .0000   3.65388655
 V2      |    -.04779763       .03766086    -1.269   .2044   30.9776786
 V3      |    -.25376184       .04622853    -5.489   .0000   11.6919643
```

```
 V4      |    -.37961397       .12878071    -2.948   .0032   1.81407563
 V5      |    -.22780476       .13328147    -1.709   .0874   2.28308824
---------+Disturbance standard deviation
 Sigma   |    2.38479704       .13327563    17.894   .0000
```

2. Two part Model.

The three estimated models appear above. The test statistic is

```
+------------------------------------+
| Listed Calculator Results          |
+------------------------------------+
 TEST2   =    740.610758
```

This is much larger than the chi squared critical value for 5 degrees of freedom. We conclude that the participation equation (probit) is different from the intensity equation (yrb).

Chapter 25

Models for Event Counts and Duration

Exercises

1. a. Conditional variance in the ZIP model. The essential ingredients that are needed for this derivation are

$$E[y^* \mid y^* > 0, \mathbf{x}_i] = \frac{\lambda_i}{1-\exp(-\lambda_i)} \;=\; E_i^*$$

and

$$Var[y^* \mid y^* > 0, \mathbf{x}_i] = \left(\frac{\lambda_i}{1-\exp(-\lambda_i)}\right)\left(1-\frac{\lambda_i}{\exp(\lambda_i)-1}\right) = E_i{}^*\left(1-\frac{\lambda_i}{\exp(\lambda_i)-1}\right) = E_i{}^* V_i{}^*$$

[See, e.g., Winkelmann (2003, pp. 33-34).]. We found the conditional mean in the text to be

$$E[y_i|x_i,w_i] \;=\; \frac{F_i\lambda_i}{1-\exp(-\lambda_i)} \;=\; F_i\, E_i^*$$

To obtain the variance, we will use the variance decomposition,

$$Var[y_i|x_i,w_i] = E_z[Var[y_i|x_i,z]] + Var_z[E[y_i|x_i,z]].$$

The expectation of the conditional variance is

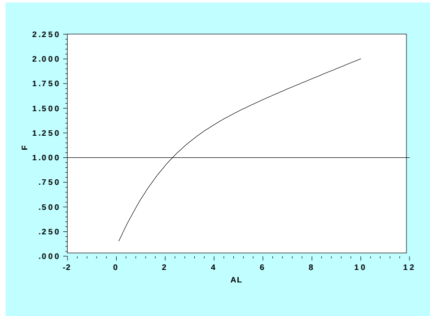
$$E_z[Var[y_i|x_i,z]] \;=\; (1-F_i)\times 0 + F_i\times\left(\frac{\lambda_i}{1-\exp(-\lambda_i)}\right)\left(1-\frac{\lambda_i}{\exp(\lambda_i)-1}\right) = F_i \times E_i^* \times V_i^*$$

The variance of the conditional mean is

$$(1-F_i)\times\left(0-\frac{F_i\lambda_i}{1-\exp(-\lambda_i)}\right)^2 + F_i\left(\frac{\lambda_i}{1-\exp(-\lambda_i)}-\frac{F_i\lambda_i}{1-\exp(-\lambda_i)}\right)^2 \;=\; F_i(1-F_i)\left(\frac{\lambda_i}{1-\exp(-\lambda_i)}\right)^2$$

$$=\; F_i(1-F_i)E_i^{*2}.$$

The unconditional variance is thus, $F_i\, E_i^*\, [V_i^* + (1-F_i)E_i^*]$. To obtain τ_i we divide by the conditional mean, which is $F_i\, E_i^*$, so $\tau_i = [V_i^* + (1-F_i)E_i^*]$. Is this greater than E_i^*? Not necessarily. The figure below plots $F_i(1-F_i)E_i^{*2}$ for $F_i = .9$ and various values of λ from .1 to about 12. There is a large range over which the function is less than one.



b. Partial Effects. The mean is $F_i E_i^*$. We suppose that w_i and x_i are the same for the moment.

$$\partial E_i/\partial x_i = E_i^*\partial F_i/\partial x_i + F_i\, \partial E_i^*/\partial x_i.$$

The first term is $E_i^*\times f_i\times\gamma$. The second term is $F_i\, \partial E_i^*/\partial\lambda_i\, \lambda_i\beta$. The missing element is

$$\partial E_i^*/\partial\lambda_i = \lambda_i/[1-\exp(-\lambda_i)] \times [1 - \exp(-\lambda_i)/[1-\exp(-\lambda_i)].$$

Comnbining terms produces the marginal effects.

2. Let y* denote the unobserved random variable that is distributed as Poisson with probability

$$\text{Prob}(y^* = j|x) = P(j) = \exp(-\lambda)\lambda^j/j!.$$

The observed random variable before the censoring is is $y = y^*|y^*>0$. The probabilities are

$$\text{Prob}(y = j|x) = P(j)/[1 - P(0)].$$

Let yc = the censored random variable. Then, yc = y for y = 1,2,3,4. yc = 5 when $y \geq 5$. The probabilities associated with the observed yc are

$$\text{Prob}(yc = 1|x) = \text{Prob}(y = 1|x) = P(1)/[1-P(0)]$$
$$\text{Prob}(yc = 2|x) = \text{Prob}(y = 2|x) = P(2)/[1-P(0)]$$
$$\text{Prob}(yc = 3|x) = \text{Prob}(y = 3|x) = P(3)/[1-P(0)]$$
$$\text{Prob}(yc = 4|x) = \text{Prob}(y = 4|x) = P(4)/[1-P(0)]$$
$$\text{Prob}(yc = 5|x) = \text{Prob}(y = 5|x) + \text{Prob}(y = 6|x) + \text{Prob}(y = 7|x) + \ldots$$

The last term is an infinite sum. But,

$$\text{Prob}(y = 5|x) + \text{Prob}(y = 6|x) + \text{Prob}(y = 7|x) + \ldots$$
$$= 1 - \text{Prob}(y = 1|x) - \text{Prob}(y = 2|x) - \text{Prob}(y = 3|x) - \text{Prob}(y = 4|x)$$

Therefore,

$$\text{Prob}(yc = 5|x) = [1 - P(1) - P(2) - P(3) - P(4)]/[1 - P(0)].$$

These are the probabilities used to construct the log likelihood function for the observed values of yc, 1,2,3,4,5.

3. The hazard function is easily obtained as $h(t) = -d\ln S(t)/dt$. For the Weibull model, $\ln S(t) = -(\lambda t)^P$ to the hazard function is $(\lambda p)(\lambda t)^{P-1}$. The median survival time occurs where the survival function equals .5. Thus,

$$\exp(-(\lambda t)^P) = .5$$
$$-(\lambda t)^P = \ln .5$$
$$(\lambda t)^P = -\ln.5 = \ln 2$$
$$P^*\ln(\lambda) + P \ln t = \ln\ln 2$$
$$P \ln t = \ln\ln 2 - P \ln\lambda$$
$$\ln t = (1/P)[\ln\ln 2 - P \ln\lambda]$$
$$t = \exp[(1/P)[\ln\ln 2 - P \ln \lambda].$$

Applications

1.

```
?====================================================
? Application 25.1
?====================================================
Namelist ;x = age,educ,hhninc,hsat $
Poisson ; Lhs = HospVis ; Rhs = One,X
        ; Marginal effects $
Calc    ; Lp = logl $
Regress ; Lhs = HospVis ; Rhs = One,X $
Negbin  ; Lhs = HospVis ; Rhs = One,X
        ; Marginal effects $
Calc    ; Ln = logl $
Calc    ; List ; LRstat = 2*(ln - lp) $

?====================================================
? Application 25.2
?====================================================
Sample  ; All $
Regress ; Lhs = one ; Rhs = one ; Str = ID ; Panel $
Poisson ; Lhs = HospVis ; Rhs = One,X
        ; Marginal effects
        ; Pds = _Groupti $
Poisson ; Lhs = HospVis ; Rhs = One,X
        ; Marginal effects
        ; Pds = _Groupti ; Random $

+---------------------------------------------+
| Poisson Regression                          |
| Maximum Likelihood Estimates                |
| Dependent variable              HOSPVIS     |
| Weighting variable                 None     |
| Number of observations            27326     |
| Iterations completed                  9     |
| Log likelihood function       -12636.40     |
| Number of parameters                  5     |
| Info. Criterion: AIC =           .92523     |
| Info. Criterion: BIC =           .92673     |
| Restricted log likelihood     -13433.21     |
+---------------------------------------------+
+---------------------------------------------+
| Poisson Regression                          |
| Chi- squared =124476.35621  RsqP=   .1947   |
| G  - squared = 20025.66932  RsqD=   .0737   |
| Overdispersion tests: g=mu(i)  :  5.279     |
| Overdispersion tests: g=mu(i)^2:  5.468     |
+---------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|     .12613692       .12567036     1.004   .3155
 AGE     |    -.00340754       .00149685    -2.276   .0228   43.5256898
 EDUC    |    -.05295428       .00834958    -6.342   .0000   11.3206310
 HHNINC  |     .39889043       .08982355     4.441   .0000    .35208362
 HSAT    |    -.24901310       .00634000   -39.277   .0000   6.78542607
+-------------------------------------------+
| Partial derivatives of expected val. with |
| respect to the vector of characteristics. |
| Effects are averaged over individuals.    |
| Observations used for means are All Obs.  |
```

```
| Conditional Mean at Sample Point    .1383 |
| Scale Factor for Marginal Effects   .1383 |
+-------------------------------------------+
+--------+-------------+---------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+-------------+---------------+--------+--------+----------+
 Constant|     .01743926       .02183573      .799   .4245
 AGE     |    -.00047111       .00025979    -1.813   .0698   43.5256898
 EDUC    |    -.00732128       .00149415    -4.900   .0000   11.3206310
 HHNINC  |     .05514924       .01579375     3.492   .0005    .35208362
 HSAT    |    -.03442771       .00220148   -15.638   .0000   6.78542607
+----------------------------------------------------+
| Ordinary    least squares regression               |
| LHS=HOSPVIS  Mean                  =   .1382566     |
|              Standard deviation    =   .8843390     |
| WTS=none     Number of observs.    =      27326     |
| Model size   Parameters            =          5     |
|              Degrees of freedom    =      27321     |
| Residuals    Sum of squares        =   21121.96     |
|              Standard error of e   =   .8792630     |
| Fit          R-squared             =   .1159150E-01 |
|              Adjusted R-squared    =   .1144679E-01 |
| Model test   F[  4, 27321] (prob) =  80.10 (.0000) |
+----------------------------------------------------+
+--------+-------------+---------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+-------------+---------------+--------+--------+----------+
 Constant|     .49839670       .04097910    12.162   .0000
 AGE     |    -.00064393       .00048945    -1.316   .1883   43.5256898
 EDUC    |    -.00619390       .00241633    -2.563   .0104   11.3206310
 HHNINC  |     .04936160       .03122845     1.581   .1140    .35208362
 HSAT    |    -.04117251       .00240443   -17.124   .0000   6.78542607
+---------------------------------------------+
| Negative Binomial Regression                 |
| Dependent variable                HOSPVIS    |
| Number of observations              27326    |
| Iterations completed                    9    |
| Log likelihood function       -10044.46    |
| Number of parameters                    6    |
| Info. Criterion: AIC =            .73560    |
| Info. Criterion: BIC =            .73740    |
| Restricted log likelihood     -12636.40    |
+---------------------------------------------+
+--------+-------------+---------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+-------------+---------------+--------+--------+----------+
 Constant|     .10394982       .12631220      .823   .4105
 AGE     |    -.00369348       .00143149    -2.580   .0099   43.5256898
 EDUC    |    -.05795593       .00826247    -7.014   .0000   11.3206310
 HHNINC  |     .38542430       .09259876     4.162   .0000    .35208362
 HSAT    |    -.23323713       .00651715   -35.788   .0000   6.78542607
---------+Dispersion parameter for count data model
 Alpha   |    6.70461029       .17537071    38.231   .0000
+-------------------------------------------+
| Partial derivatives of expected val. with |
| respect to the vector of characteristics. |
| Effects are averaged over individuals.    |
| Observations used for means are All Obs.  |
| Conditional Mean at Sample Point    .1367 |
| Scale Factor for Marginal Effects   .1367 |
+-------------------------------------------+
+--------+-------------+---------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
```

```
+--------+--------------+----------------+--------+--------+----------+
 Constant|     .01421398       .02120646      .670   .5027
 AGE     |    -.00050504       .00024071    -2.098   .0359   43.5256898
 EDUC    |    -.00792483       .00146645    -5.404   .0000   11.3206310
 HHNINC  |     .05270247       .01588312     3.318   .0009    .35208362
 HSAT    |    -.03189257       .00226820   -14.061   .0000   6.78542607
+-------------------------------------+
| Listed Calculator Results           |
+-------------------------------------+
 LRSTAT  =   5183.862874
```

2.

```
+----------------------------------------------+
| Panel Model with Group Effects               |
| Dependent variable              HOSPVIS      |
| Weighting variable                 None      |
| Number of observations            27326      |
| Log likelihood function       -4198.145      |
| Number of parameters                  4      |
| Info. Criterion: AIC =           .30756      |
| Info. Criterion: BIC =           .30876      |
| Unbalanced panel has    7293 individuals.    |
| Missing or sumY=0, Skipped  5640 groups.     |
| Poisson      Regression -- Fixed Effects     |
+----------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 AGE     |    -.00020613       .00705126     -.029   .9767   43.5256898
 EDUC    |    -.04033708       .09220144     -.437   .6618   11.3206310
 HHNINC  |     .49927712       .18484588     2.701   .0069    .35208362
 HSAT    |    -.16686419       .01027579   -16.239   .0000   6.78542607
+--------------------------------------------+
| Partial derivatives of expected val. with |
| respect to the vector of characteristics. |
| They are computed at the means of the Xs. |
| Observations used for means are All Obs.  |
| Conditional Mean at Sample Point    .1383 |
| Scale Factor for Marginal Effects   .1383 |
+--------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 AGE     |   -.284995D-04      .00097488     -.029   .9767   1.00000000
 EDUC    |    -.00557687       .01274746     -.437   .6618   43.5256898
 HHNINC  |     .06902836       .02555616     2.701   .0069   11.3206310
 HSAT    |    -.02307008       .00142070   -16.239   .0000    .35208362
+----------------------------------------------+
| Panel Model with Group Effects               |
| Dependent variable              HOSPVIS      |
| Number of observations            27326      |
| Log likelihood function       -10200.91      |
| Number of parameters                  6      |
| Info. Criterion: AIC =           .74705      |
| Info. Criterion: BIC =           .74885      |
| Unbalanced panel has    7293 individuals.    |
| Poisson      Regression -- Random Effects    |
+----------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|    -.22178663       .13617622    -1.629   .1034
 AGE     |    -.00170639       .00145901    -1.170   .2422   43.5256898
```

```
 EDUC    |    -.05399730       .01001912    -5.389   .0000   11.3206310
 HHNINC  |     .40499179       .06938275     5.837   .0000    .35208362
 HSAT    |    -.20075292       .00400154   -50.169   .0000   6.78542607
 Alpha   |    3.59227655       .11685254    30.742   .0000
+-------------------------------------------+
| Partial derivatives of expected val. with |
| respect to the vector of characteristics. |
| They are computed at the means of the Xs. |
| Observations used for means are All Obs.  |
| Conditional Mean at Sample Point    .1383 |
| Scale Factor for Marginal Effects   .1383 |
+-------------------------------------------+
+--------+-------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+-------------+----------------+--------+--------+----------+
 Constant|    -.03066347       .01882726    -1.629   .1034
 AGE     |    -.00023592       .00020172    -1.170   .2422   43.5256898
 EDUC    |    -.00746548       .00138521    -5.389   .0000   11.3206310
 HHNINC  |     .05599279       .00959262     5.837   .0000    .35208362
 HSAT    |    -.02775542       .00055324   -50.169   .0000   6.78542607
```

3. Ship Accidents

```
Create ; logmth = log(months) $
Name ; X=logmth,one,ta,tb,tc,td,t6064,t6569,t7074,o6074$
Reject ; acc < 0 $
Pois ; lhs = acc ; Rhs = x $
Pois ; lhs = acc ; Rhs = x ; Rst = 1,9_b $
Negb ; lhs = acc ; Rhs = x ; Rst = 1,9_b,alpha $
+---------------------------------------------+
| Poisson Regression                          |
| Dependent variable                  ACC     |
| Number of observations               34     |
| Log likelihood function       -67.99930     |
| Number of parameters                 10     |
| Info. Criterion: AIC =          4.58819     |
| Info. Criterion: BIC =          5.03712     |
| Restricted log likelihood     -356.2029     |
+---------------------------------------------+
+---------------------------------------------+
| Poisson Regression                          |
| Chi- squared =    39.70580  RsqP=   .9491   |
| G  - squared =    38.13211  RsqD=   .9380   |
| Overdispersion tests: g=mu(i)  :   .853     |
| Overdispersion tests: g=mu(i)^2:  -.760     |
+---------------------------------------------+
+--------+-------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+-------------+----------------+--------+--------+----------+
 LOGMTH  |     .90617018       .10174566     8.906   .0000   7.04925451
 Constant|   -4.61752968       .72938865    -6.331   .0000
 TA      |    -.26966656       .24189066    -1.115   .2649    .20588235
 TB      |    -.62826604       .32582681    -1.928   .0538    .20588235
 TC      |   -1.03179604       .34039236    -3.031   .0024    .20588235
 TD      |    -.40106977       .30540945    -1.313   .1891    .20588235
 T6064   |    -.36146212       .24726698    -1.462   .1438    .23529412
 T6569   |     .30035782       .21325393     1.408   .1590    .29411765
 T7074   |     .39874282       .20053445     1.988   .0468    .29411765
 O6074   |    -.36986273       .11821010    -3.129   .0018    .41176471
+---------------------------------------------+
| Poisson Regression                          |
| Maximum Likelihood Estimates                |
| Dependent variable                  ACC     |
```

```
| Number of observations               34     |
| Log likelihood function       -68.41456     |
| Number of parameters                  9     |
| Info. Criterion: AIC =          4.55380     |
| Info. Criterion: BIC =          4.95783     |
| Restricted log likelihood     -356.2029     |
+---------------------------------------------+
+---------------------------------------------+
| Poisson Regression                          |
| Chi- squared =    42.44145  RsqP=   .9456   |
| G  - squared =    38.96262  RsqD=   .9366   |
| Overdispersion tests: g=mu(i)  :   .934     |
| Overdispersion tests: g=mu(i)^2:  -.613     |
+---------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 LOGMTH  |    1.00000000    ......(Fixed Parameter).......
 Constant|   -5.25351861       .24642858   -21.319   .0000
 TA      |    -.32052881       .23575203    -1.360   .1740    .20588235
 TB      |    -.86524026       .19852119    -4.358   .0000    .20588235
 TC      |   -1.00929327       .33950071    -2.973   .0030    .20588235
 TD      |    -.39483795       .30680184    -1.287   .1981    .20588235
 T6064   |    -.44497064       .23323916    -1.908   .0564    .23529412
 T6569   |     .25087485       .20875483     1.202   .2295    .29411765
 T7074   |     .37248476       .19930193     1.869   .0616    .29411765
 O6074   |    -.38385913       .11826046    -3.246   .0012    .41176471
```

There is no evidence of overdispersion. The tests from the Poisson model are both insignificant, and the estimate of α in the negative binomial model is essentially zero.

```
+---------------------------------------------+
| Negative Binomial Regression                |
| Dependent variable                  ACC     |
| Weighting variable                 None     |
| Number of observations               34     |
| Log likelihood function       -68.42007     |
| Number of parameters                 10     |
| Info. Criterion: AIC =          4.61295     |
|   Finite Sample: AIC =          4.89428     |
| Info. Criterion: BIC =          5.06188     |
| Info. Criterion:HQIC =          4.76604     |
| NegBin form 2; Psi(i) = theta               |
+---------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 LOGMTH  |    1.00000000    ......(Fixed Parameter).......
 Constant|   -5.25074235       .26830333   -19.570   .0000
 TA      |    -.32296435       .39695609     -.814   .4159    .20588235
 TB      |    -.86731524       .20092395    -4.317   .0000    .20588235
 TC      |   -1.01171406       .24980570    -4.050   .0001    .20588235
 TD      |    -.39875463       .23889734    -1.669   .0951    .20588235
 T6064   |    -.44585250       .31679943    -1.407   .1593    .23529412
 T6569   |     .25060358       .27552926      .910   .3631    .29411765
 T7074   |     .37073607       .25504806     1.454   .1461    .29411765
 O6074   |    -.38364155       .15800844    -2.428   .0152    .41176471
---------+Dispersion parameter for count data model
 Alpha   |    .648724D-04      .02406424      .003   .9978
```

4. Strikes. There are 9 years of data. The number of strikes is 8,6,11,3,3,2,19,2,9. The Poisson regression is shown below. It does appear that the number of strikes is significantly related to the PROD variable. However, with only 9 observations, use of the asymptotic distribution for the test is probably overly optimistic. The result is probably borderline.

```
+---------------------------------------------+
| Poisson Regression                          |
| Dependent variable            _GROUPTI      |
| Weighting variable                None      |
| Number of observations               9      |
| Log likelihood function      -28.99317      |
| Number of parameters                 2      |
| Info. Criterion: AIC =         6.88737      |
| Info. Criterion: BIC =         6.93120      |
| Restricted log likelihood    -31.19884      |
+---------------------------------------------+
+---------------------------------------------+
| Poisson Regression                          |
| Chi- squared =    25.08061  RsqP=   .2317   |
| G  - squared =    26.13767  RsqD=   .1444   |
| Overdispersion tests: g=mu(i)  :  1.954     |
| Overdispersion tests: g=mu(i)^2:  2.618     |
+---------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|    1.90854253       .12998621    14.683   .0000
 PROD    |    5.16576744      2.51306610     2.056   .0398   -.00302000
```

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, $\mathbf{x}_i$ $i = 1,...,n$, and some nonzero vector, **a**,

$$\sum_{i=1}^{n}(\mathbf{x}_i - \mathbf{a})(\mathbf{x}_i - \mathbf{a})' = \mathbf{X}'\mathbf{M}^0\mathbf{X} + n(\overline{\mathbf{x}} - \mathbf{a})(\overline{\mathbf{x}} - \mathbf{a})'.$$

Write $\mathbf{x}_i - \mathbf{a}$ as $[(\mathbf{x}_i - \overline{\mathbf{x}}) + (\overline{\mathbf{x}} - \mathbf{a})]$. Then, the sum is

$$\begin{aligned}\sum_{i=1}^{n} [(\mathbf{x}_i - \overline{\mathbf{x}}) + (\overline{\mathbf{x}} - \mathbf{a})]\,[(\mathbf{x}_i - \overline{\mathbf{x}}) + (\overline{\mathbf{x}} - \mathbf{a})]' =& \\ \sum_{i=1}^{n} (\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})' &+ \sum_{i=1}^{n} (\overline{\mathbf{x}} - \mathbf{a})(\overline{\mathbf{x}} - \mathbf{a})' \\ + \sum_{i=1}^{n} (\mathbf{x}_i - \overline{\mathbf{x}})(\overline{\mathbf{x}} - \mathbf{a})' &+ \sum_{i=1}^{n} (\overline{\mathbf{x}} - \mathbf{a})(\mathbf{x}_i - \overline{\mathbf{x}})'\end{aligned}$$

Since $(\overline{\mathbf{x}} - \mathbf{a})$ is a vector of constants, it may be moved out of the summations. Thus, the fourth term is $(\overline{\mathbf{x}} - \mathbf{a}) \sum_{i=1}^{n} (\mathbf{x}_i - \overline{\mathbf{x}})' = \mathbf{0}$. The third term is likewise. The first term is $\mathbf{X}'\mathbf{M}^0\mathbf{X}$ by the definition while the second is $n(\overline{\mathbf{x}} - \mathbf{a})(\overline{\mathbf{x}} - \mathbf{a})'$. □

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 $\mathbf{A}\,\mathbf{b}_i$ 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 $|\mathbf{A}|$, $tr(\mathbf{A})$ 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

$$\mathbf{A}^{-1} = \frac{-1}{18}\begin{bmatrix} \det\begin{pmatrix} 2 & 5 \\ 2 & 8 \end{pmatrix} & -\det\begin{pmatrix} 4 & 7 \\ 2 & 8 \end{pmatrix} & \det\begin{pmatrix} 4 & 7 \\ 2 & 5 \end{pmatrix} \\ -\det\begin{pmatrix} 3 & 5 \\ 5 & 8 \end{pmatrix} & \det\begin{pmatrix} 1 & 7 \\ 5 & 8 \end{pmatrix} & -\det\begin{pmatrix} 1 & 7 \\ 3 & 5 \end{pmatrix} \\ \det\begin{pmatrix} 3 & 2 \\ 5 & 2 \end{pmatrix} & -\det\begin{pmatrix} 1 & 4 \\ 5 & 2 \end{pmatrix} & \det\begin{pmatrix} 1 & 4 \\ 3 & 2 \end{pmatrix} \end{bmatrix} = \begin{bmatrix} -6/18 & 18/18 & -6/18 \\ -1/18 & 27/18 & -16/18 \\ 4/18 & -18/18 & 10/18 \end{bmatrix}. \quad \square$$

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 $\mu_{12} = 7/\sqrt{13} = 1.941$. Finally, $25 = \mu_{11}^2 + \mu_{12}^2$ so $\mu_{11}^2 = 25 - 49/13 = 21.23$, or $\mu_{11} = 4.61$.

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.

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 $\partial\mathbf{x}'\mathbf{Ax}/\partial z$? Now, suppose each element of **x** is also a function of z. Once again, what is $\partial\mathbf{x}'\mathbf{Ax}/\partial z$?

The quadratic form is $\sum_i\sum_j x_ix_ja_{ij}$, so

$$\partial\mathbf{x}'\mathbf{A}(z)\mathbf{x}/\partial z = \sum_i\sum_j x_ix_j(\partial a_{ij}/\partial z) = \mathbf{x}'(\partial\mathbf{A}(z)/\partial z)\mathbf{x}$$ where $\partial\mathbf{A}(z)/\partial z$ is a matrix of partial derivatives.

Now, if each element of **x** is also a function of z, then,

$$\partial\mathbf{x}'\mathbf{Ax}/\partial z = \sum_i\sum_j x_ix_j(\partial a_{ij}/\partial z) + \sum_i\sum_j(\partial x_i/\partial z)x_ja_{ij} + \sum_i\sum_j x_i(\partial x_j/\partial z)a_{ij}$$
$$= \mathbf{x}'(\partial\mathbf{A}(z)/\partial z)\mathbf{x} + (\partial\mathbf{x}(z)/\partial z)'\mathbf{A}(z)\mathbf{x}(z) + \mathbf{x}(z)'\mathbf{A}(z)(\partial\mathbf{x}(z)/\partial z)$$

If **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 $|\mathbf{B}^{-1}\mathbf{A} - \mu\mathbf{I}| = 0$?

Since **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 **A** is nonzero if **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 **A** and **B** are symmetric and positive definite. As such, it can be shown tat the solution is the same as that of a third determinantal equation involving a symmetric matrix.

18. Using the matrix **A** in Exercise 9, find the vector **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?

$$y_1 = x_1/x_2,$$

$$\ln y_2 = \ln x_1 - \ln x_2 + \ln x_3,$$

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 $\mathbf{Y} = \begin{bmatrix} 1 & -1 & 0 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \end{bmatrix}\mathbf{X} = \mathbf{J}\mathbf{X}$. 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 = 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 $\partial[(\mathbf{x}'\mathbf{A}\mathbf{x})/(\mathbf{x}'\mathbf{B}\mathbf{x})]/z$?

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,

$$\begin{aligned}\partial(a/b)/\partial z &= [b(\partial a/\partial z) - a(\partial b/\partial z)]/b^2 \\ &= [\mathbf{x}'\mathbf{B}\mathbf{x}(2\mathbf{x}'\mathbf{A}\mathbf{g}) - \mathbf{x}'\mathbf{A}\mathbf{x}(2\mathbf{x}'\mathbf{B}\mathbf{g})] / (\mathbf{x}'\mathbf{B}\mathbf{x})^2 = 2[\mathbf{x}'\mathbf{A}\mathbf{x}/\mathbf{x}'\mathbf{B}\mathbf{x}][\mathbf{x}'\mathbf{A}\mathbf{g}/\mathbf{x}'\mathbf{A}\mathbf{x} - \mathbf{x}'\mathbf{B}\mathbf{g}/\mathbf{x}'\mathbf{B}\mathbf{x}].\end{aligned}$$

22. Suppose **y** is an $n\times1$ vector and **X** is an $n\times K$ matrix. The projection of **y** into the column space of **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 **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 **y** and its projection into the column space of **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 $\mathbf{P}'\mathbf{X}'(c\mathbf{y}) = \mathbf{P}'\mathbf{X}'\mathbf{X}\mathbf{P}\mathbf{b}^*$. Since **P** is nonsingular, $\mathbf{P}'$ has an inverse. Premultiplying the equation by $(\mathbf{P}')^{-1}$, we have $c\mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{X}(\mathbf{P}\mathbf{b}^*)$ or $\mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{X}[(1/c)\mathbf{P}\mathbf{b}^*]$. Therefore, in terms of the original **y** and **X**, we see that $\mathbf{b} = (1/c)\mathbf{P}\mathbf{b}^*$ 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 **y** into the space of **X**. This makes some sense, since, if **P** is a nonsingular matrix, the column space of $\mathbf{X}^*$ is exactly the same as the same as that of **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 **y** and **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 **M** and **P** are idempotent.)

(a) Compute the **P** and **M** based on **XQ** instead of **X**.
(b) What are the characteristic roots of **M** and **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 **M** and **P** for **XQ**, they are the same. Proof: The counterpart to **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 **M** matrix would be the same as well. This is an application of the result found in the previous exercise. The **P** matrix is the projection matrix, and, as we found, the projection into the space of **X** is the same as the projection into the space of **XQ**.

(b) Since **M** and **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 **M** and **P** are 2, so each has 2 unit roots and 2 zero roots.

24. Suppose that **A** is an $n\times n$ matrix of the form $\mathbf{A} = (1-\rho\mathbf{I}) + \rho\mathbf{ii}'$, where **i** is a column of 1s and $0 < \rho < 1$. Write out the format of **A** explicitly for $n = 4$. Find all of the characteristic roots and vectors of **A**. (**Hint:** There are only two distinct characteristic roots, which occur with multiplicity 1 and n-1. Every **c** of a certain type is a characteristic vector of **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 $[(1-\rho)\mathbf{I} + \rho\mathbf{ii}']\mathbf{c} = \lambda\mathbf{c}$. Multiply this out to obtain $(1-\rho)\mathbf{c} + \rho\mathbf{ii}'\mathbf{c} = \lambda\mathbf{c}$ or $\rho\mathbf{ii}'\mathbf{c} = [\lambda - (1-\rho)]\mathbf{c}$. Let $\mu = \lambda - (1-\rho)$, 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 $\lambda = \mu + (1-\rho)$. 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 $0 + (1-\rho) = (1-\rho)$. 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 $(1-\rho+n\rho)$ is the same, or **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 **D** be any arbitrary set of n-1 vectors containing n elements. Transform all columns of **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 $\mathbf{C} = \mathbf{F}(\mathbf{F}'\mathbf{F})^{-2}$. **C** is a linear combination of the columns of **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 (A-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 extension, prove that every matrix in the sequence of matrices in (E-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

$$\mathbf{x}'\mathbf{H}_{i+1}\mathbf{x} \ = \ \mathbf{x}'\mathbf{x} \ + \ \sum_{j=1}^{i}(\mathbf{x}'\mathbf{d}_j)(\mathbf{d}_j'\mathbf{x}) \ = \mathbf{x}'\mathbf{x} \ + \ \sum_{j=1}^{i}(\mathbf{x}'\mathbf{d}_j)^{\mathbf{2}}$$

This is obviously positive for all $\mathbf{x}$. A simple way to establish this for the matrix in (E-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 $|\mathbf{P}-\lambda\mathbf{I}| = 0$, 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 B.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 2.4.1,

(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.)

$$\text{The matrix is} \quad \begin{bmatrix} 15.000 & 120.00 & 19.310 & 111.79 & 99.770\\ 120.00 & 1240.0 & 164.30 & 1035.9 & 875.60\\ 19.310 & 164.30 & 25.218 & 148.98 & 131.22\\ 111.79 & 1035.9 & 148.98 & 943.86 & 799.02\\ 99.770 & 875.60 & 131.22 & 799.02 & 716.67 \end{bmatrix}$$

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 $\mathbf{V} = \mathbf{D}^{-2}\mathbf{X}'\mathbf{X}\mathbf{D}^{-2}$.

$$\text{The resulting matrix is} \begin{bmatrix} 1 & .8798823 & .992845 & .939515 & .962265\\ .879883 & 1 & .929119 & .957532 & .928828\\ .992845 & .929119 & 1 & .965648 & .976079\\ .939515 & .957532 & .965648 & 1 & .971503\\ .962265 & .928828 & .976079 & .971503 & 1 \end{bmatrix}.$$

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. $\square$

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,

(a) Prob[$|x| > 2$] = 1 - Prob[$|x| \leq 2$]
= 1 - Prob[$-2 \leq x \leq 2$]
= 1 - Prob[$(-2-1)/3 \leq z \leq (2-1)/3$]
= 1 - [F(1/3) - F(-1)] = 1 - .6306 + .1587 = .5281

(b) Prob[$x > -1 \mid x < 1.5$] = Prob[$-1 < x < 1.5$] / Prob[$x < 1.5$]
Prob[$-1 < x < 1.5$] = Prob[$(-1-1)/3 < z < (1.5-1)/3$)]
= Prob[$z < 1/6$] - Prob[$z < -2/3$]
= .5662 - .2525 = .3137.

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 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	X = 1	X = 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(12).11+1(22).19+2(12).15+2(22).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

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

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
Var[$Y*X$=2] = $1.700 - 1.167^2 = .6749 = .3381$
E[Var[$Y*X$]] = .34(.3751)+.36(.6749)+.30(.3381) = .4719
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[Var[$Y*X$]] + Var[E[$Y*X$]] = .4719 + .0012 = .4731 = 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 $E[y - E[y|x]]^2 \le E[y - a - bx]^2$. The result is fundamental in least squares theory. Verify that the a and b which you found satisfy (3-68) and (3-69).

$$
\begin{array}{lllll}
 & & (x=0) & (x=1) & (x=2) \\
E[y - \mathrm{E}[y|x]]^2 = & (y=0) & .05(0 - 1.088)^2 & + .10(0 - 1.139)^2 & + .03(0 - 1.167)^2 \\
 & (y=1) & + .21(1 - 1.088)^2 & + .11(1 - 1.139)^2 & + .19(1 - 1.167)^2 \\
 & (y=2) & + .08(2 - 1.088)^2 & + .15(2 - 1.139)^2 & + .08(2 - 1.167)^2
\end{array}
$$

$= .4719 = E[\mathrm{Var}[y|x]]$.

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

$$E[y] - a - bE[x] = 0$$
$$E[xy] - aE[x] - b\mathrm{E}[x^2] = 0.$$

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 $\quad b = \mathrm{Cov}[x,y] / \mathrm{Var}[x] = -.0708 / .4731 = -.150$

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

$$
\begin{array}{lllll}
 & & (x=0) & (x=1) & (x=2) \\
\mathrm{E}[y - a - bx]^2 = & (y=0) & .05(0 - 1.274)^2 & + .10(0 - 1.124)^2 & + .03(0 - .974)^2 \\
 & (y=1) & + .21(1 - 1.274)^2 & + .11(1 - 1.124)^2 & + .19(1 - .974)^2 \\
 & (y=2) & + .08(2 - 1.274)^2 & + .15(2 - 1.124)^2 & + .08(2 - .974)^2
\end{array}
$$

$= .4950 > .4719.$

9. Suppose x has an exponential distribution, $f(x) = \theta e^{-\theta x}$, $x \ge 0$. Find the mean, variance, skewness, and kurtosis of x. The Gamma integral 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.

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

$$c = 26.217 + \frac{(1/11.3 - 1/12)}{(1/11 - 1/12)}(24.725 - 26.217) = 25.2009. \square$$

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 (C-30). The density of this z is given in (C-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 (C-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 $x_1 = x_1(y_1)$, ... $x_n = x_n(y_n)$.

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 (C-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 B.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 (C-28). The function based on the ith derivative of the density given by $H_i = [(-1)^i d^i\phi\,(x)/dx^{\text{i}}]/\phi(x)$, $i = 0,1,2,...$ 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

$$d^r\phi(x)/dx^r + xd^{r\text{-}1}\phi(x)/dx^{r\text{-}1} + (r\text{-}1)d^{r\text{-}2}\phi(x)/dx^{r\text{-}2} = 0.$$

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 $d\phi(x)/\text{d}x = -x\phi(x)$. Therefore,

$$d^2\phi(x)/dx^2 = (x^2 - 1)\phi(x)$$

and $\quad d^3\phi(x)/dx^3 = (3x - x^3)\phi(x)$.

The polynomials are $\quad H_1 = x,\ H_2 = x^2 - 1$, and $H_3 = x^3 - 3x$.

For part (b), we solve for $\quad d^r\phi(x)/dx^r = -xd^{r\text{-}1}\phi(x)/dx^{r\text{-}1} - (r\text{-}1)d^{r\text{-}2}\phi(x)/dx^{r\text{-}2}$

Therefore, $\quad d^4\phi(x)/dx^4 = -x(3x - x^3)\phi(x) - 3(x^2 - 1)\phi(x) = (x^4 - 6x^2 + 3)\phi(x)$

and $\quad d^5\phi(x)/dx^5 = (-x^5 + 10x^3 - 15x)\phi(x)$.

Thus, $\quad H_4 = x^4 - 6x^2 + 3$ 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 $\text{i} \neq \text{j}$. 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 $\mathbf{J\Sigma J}$N are $\quad (1,1) = \dfrac{\sigma_x^2}{\mu_y^2} + \dfrac{\sigma_y^2\mu_2^x}{\mu_y^4} - \dfrac{2\sigma_{xy}\mu_x}{\mu_y^3}$

$$(1,2) = \sigma_x^2 - \sigma_y^2\ \mu_x^2 /\ \mu_y^4$$

$$(2,2) = \sigma_x^2\ \mu_y^4 + \sigma_y^2\ \mu_x^2 + 2\sigma_{xy}\mu_x\mu_y.$$

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

$$\begin{aligned} E[x] &= \textstyle\sum_{x=0}^{\infty} xf(x) = \sum_{x=0}^{\infty} xe^{-\lambda}\lambda^x / x! \\ &= \textstyle\sum_{x=1}^{\infty} e^{-\lambda}\lambda^{x-1} / (x-1)! \\ &= \lambda \textstyle\sum_{y=0}^{\infty} e^{-\lambda}\lambda^y / y! \\ &= \lambda, \end{aligned}$$

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,

$$E[x(x\text{-}1)] = \sum_{x=0}^{\infty} x(x-1)f(x) = \sum_{x=0}^{\infty} x(x-1)e^{-\lambda}\lambda^x / x!$$

$$= \sum_{x=2}^{\infty} e^{-\lambda}\lambda^{x-2} / (x-2)!$$

$$= \lambda^2 \sum_{y=0}^{\infty} e^{-\lambda}\lambda^{y} / y!$$

$$= \lambda^2$$

$$= \mathrm{E}[x^2] \; - \; \mathrm{E}[x]$$

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

$$E[x(x\text{-}1)(x\text{-}2)] = E[x^3] - 3E[x^2] + 2E[x].$$

$$= \lambda^3.$$

Therefore, $$E[x^3] = \lambda^3 + 3(\lambda + \lambda^2) - 2\lambda$$

$$= \lambda^3 + 3\lambda^2 + \lambda.$$

Then, $$E[x - E[x]]^3 = E[x^3] - 3\lambda E[x^2] + 3\lambda^2 E[x] - \lambda^3$$

$$= \lambda. \quad \square$$

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

$$f(z) = \frac{1}{\sigma\sqrt{2\pi}} \frac{1}{\sqrt{z}} \frac{1}{2\sqrt{z}} e^{-\frac{1}{2}\left[\left(\frac{1}{2}\ln z - \mu\right)/\sigma\right]^2}, z > 0$$

$$= \frac{1}{(2\sigma)z\sqrt{2\pi}} e^{-\frac{1}{2}\left[(\ln z - 2\mu)/(2\sigma)\right]^2}, z > 0.$$

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 $\boldsymbol{\mu}\mathbf{N} = [1, 2, 4]$

and covariance matrix $\boldsymbol{\Sigma} = \begin{bmatrix} 2 & 3 & 1 \\ 3 & 5 & 2 \\ 1 & 2 & 6 \end{bmatrix}$

(a) Compute the intercept and slope in the function $E[y*x_1]$, $\mathrm{Var}[y*x_1]$, 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 $\mathrm{Var}[y*x_1{=}2.5,x_2{=}3.3]$?

First, for normally distributed variables, we have from (3-102),

$$E[y*\mathbf{x}] = \mu_y + \mathrm{Cov}[y,\mathbf{x}]\{\mathrm{Var}[\mathbf{x}]\}^{-1}(\mathbf{x} - :_x)$$

and $$\mathrm{Var}[y*\mathbf{x}] = \mathrm{Var}[y] - \mathrm{Cov}[y,\mathbf{x}]\{\mathrm{Var}[\mathbf{x}]\}^{-1}\mathrm{Cov}[\mathbf{x},y]$$

and $$COD = \mathrm{Var}[E[y*\mathbf{x}]] / \mathrm{Var}[y]$$

$$= \mathrm{Cov}[y,\mathbf{x}]\{\mathrm{Var}[\mathbf{x}]\}^{-1}\mathrm{Cov}[\mathbf{x},y] / \mathrm{Var}[y].$$

We may just insert the figures above to obtain the results.

$$E[y*x_1] = 1 + (3/5)(x_1 - 2) = -.2 + .6x_1,$$

$$\mathrm{Var}[y*x_1] = 2 - 3(1/5)3 = 1/5 = .2$$

COD = $.6^2(5)/2 = .9$

E[y*x_1,x_2] = $1 + \begin{bmatrix} 3 & 1 \end{bmatrix} \begin{bmatrix} 5 & 2 \\ 2 & 6 \end{bmatrix}^{-1} \begin{bmatrix} 3 \\ 1 \end{bmatrix}$

= $-.4615 + .6154x_1 - .03846x_2$,

Var[y*x_1,x_2] = 2 - (.6154,-.03846)(3,1)N = .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,

$$f(y) = \frac{(1/2)^{n/2}}{\Gamma(n/2)} e^{-\frac{1}{2y}} \left(\frac{1}{y}\right)^{\frac{n}{2}+1}, y > 0. \quad \square$$

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 C.4.5.

The density of x is $f(x) = \frac{\lambda^P}{\Gamma(P)} e^{-\lambda x} x^{P-1}, x > 0.$ 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 $f(y) = \frac{\lambda^P}{\Gamma(P)} \frac{1}{y^2} e^{-\lambda/y} \left(\frac{1}{y}\right)^{P-1}, y > 0.$ The mean is $E(y) = \int_0^\infty y \frac{\lambda^P}{\Gamma(P)} \frac{1}{y^2} e^{-\lambda/y} \left(\frac{1}{y}\right)^{P-1} dy$. This is a gamma integral (see Section 5.2.4b). Combine terms to obtain $E(y) = \int_0^\infty \frac{\lambda^P}{\Gamma(P)} e^{-\lambda/y} \left(\frac{1}{y}\right)^{P} dy$. 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 $E(y) = \int_0^\infty \frac{\lambda^P}{\Gamma(P)} e^{-\lambda z} z^P \left(\frac{1}{z^2}\right) dz = \int_0^\infty \frac{\lambda^P}{\Gamma(P)} e^{-\lambda z} z^{P-2} dz$. Now, we can use the gamma integral directly, to find $E(y) = \frac{\lambda^P}{\Gamma(P)} \times \frac{\Gamma(P-1)}{\lambda^{P-1}} = \frac{\lambda}{P-1}$. 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, $E(y^2) = \int_0^\infty y^2 \frac{\lambda^P}{\Gamma(P)} \frac{1}{y^2} e^{-\lambda/y} \left(\frac{1}{y}\right)^{P-1} dy = \int_0^\infty \frac{\lambda^P}{\Gamma(P)} e^{-\lambda/y} \left(\frac{1}{y}\right)^{P-1} dy$. 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 $E(y^2) = \int_0^\infty \frac{\lambda^P}{\Gamma(P)} e^{-\lambda z} z^{P-1} \left(\frac{1}{z^2}\right) dz = \int_0^\infty \frac{\lambda^P}{\Gamma(P)} e^{-\lambda z} z^{P-3} dz$. This is $\frac{\lambda^P}{\Gamma(P)} \times \frac{\Gamma(P-2)}{\lambda^{P-2}} = \frac{\lambda^2}{(P-1)(P-2)}$.

Now, Var[y] = $E[y^2] - E^2[y]$ = $\frac{\lambda^3}{(P-1)^2(P-2)}, P > 2.$

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 C.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?

$$\begin{aligned} E[t^x] &= \sum_{x=0}^{\infty} t^x (1-\pi)^{x-1}\pi \\ &= \frac{\pi}{(1-\pi)} \sum_{x=0}^{\infty} [t(1-\pi)]^x \\ &= \frac{\pi}{(1-\pi)} \frac{1}{1-t(1-\pi)}. \quad \square \end{aligned}$$

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 (C-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

$$M(t) \quad = k\int_0^\infty e^{tx} e^{-x/2} x^{(n/2)-1} dx$$

$$= k\int_0^\infty e^{-x(1/2-t)} x^{(n/2)-1} dx\ .$$

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

$$M_{\sum_i x_i} = \exp\left[\sum_i \mu_i - \frac{1}{2}\left(\sum_i \sigma_i^2\right)t^2\right].\ \square$$

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, 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 $(n-1)s^2/\sigma^2$ 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 $\mu = 1.8$, 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,

$$\bar{x}_1 \sim N[\mu_1, \sigma_1^2 / n_1] \text{ and } \bar{x}_2 \sim N[\mu_2, \sigma_2^2 / n_2],$$

$$\bar{x}_1 - \bar{x}_2 \sim N[\mu_1 - \mu_2, \sigma^2\{(1/n_1) + (1/n_2)\}],$$

$$(n_1\text{-}1)s_1^2/\sigma^2 \sim \chi^2[n_1\text{-}1] \text{ and } (n_2\text{-}1)s_2^2/\sigma^2 \sim \chi^2[n_2\text{-}1]$$

$$(n_1\text{-}1)s_1^2/\sigma^2 + (n_2\text{-}1)s_2^2/\sigma^2 \sim \chi^2[n_1 + n_2 - 2]$$

Thus, the statistic

$$t = \frac{\left\{\left(\bar{x}_1 - \bar{x}_2\right) - \left(\mu_1 - \mu_2\right)\right\} / \sqrt{\sigma^2\left[(1/n_1) + (1/n_2)\right]}}{\sqrt{\left\{(n_1 - 1)s_1^2 / \sigma^2 + (n_2 - 1)s_2^2 / \sigma^2\right\} / (n_1 + n_2 - 2)}}$$

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

$$t = \frac{\left(\bar{x}_1 - \bar{x}_2\right) / \sqrt{(1/n_1) + (1/n_2)}}{\sqrt{\left\{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2\right\} / (n_1 + n_2 - 2)}}$$

The sample statistics are

$$n_1 = 10,\ \bar{x}_1 = 1.52,\ s_1^2 = .9418$$

$$n_2 = 10,\ \bar{x}_2 = 1.62,\ s_2^2 = 2.0907$$

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 $\lambda = -2[-31.4876 - (-30.713077)] = 1.541$. 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 $1.52 - 2.262[.97/\sqrt{10}] \le \mu \le 1.52 + 2.262[.97/\sqrt{10}]$ or $.826 \le \mu \le 2.214$. Using the procedure in Example 4.30, we obtain a 95% confidence for σ^2 of $9(.941)/19.02 \le \sigma^2 \le 9(.941)/2.70$ or $.445 \le \sigma^2 \le 3.137$. Taking square roots gives the confidence interval for σ, $.667 \le \sigma \le 1.771$.

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 $34 - 1.96(13.3^2\pi/130)^2 \le \mu \le 34 + 1.96(13.3^2\pi/130)^2$ or $29.95 \le \mu \le 38.052$. 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 $f(x) = 1/\theta$, $0 \le x \le \theta$. 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

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

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 $z = \theta$. □

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 $[\mathbf{GS}^{-1}\mathbf{G}']^{-1} = \begin{bmatrix} 0.10178 & -0.12492 \\ -0.12492 & 0.16973 \end{bmatrix}$. 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.

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

$\mathbf{H} = \dfrac{1}{n}\begin{bmatrix} 1+\mu^2/(2\sigma^2) & 2\mu/\sigma+(\mu/\sigma)^3 \\ 2\mu/\sigma+(\mu/\sigma)^3 & 4\mu^2/\sigma^2+2\mu^4/\sigma^4 \end{bmatrix}$.

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 $\lambda = 2$.

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 $\theta = 1$. 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.

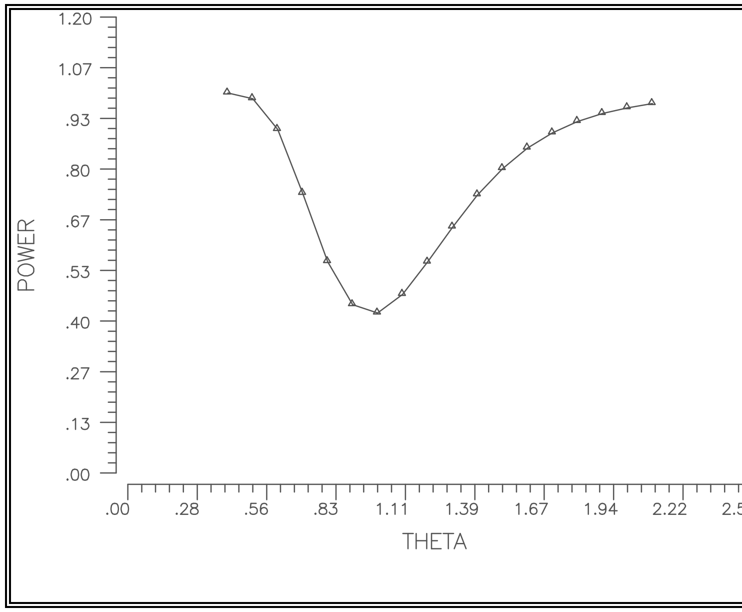
(a) Using the asymptotic distribution of $\bar{x}$ graph the asymptotic approximation to the true power function.
(b) Using the result discussed in Example 4.17, describe how to obtain the true power function for this test.

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 θ = 1, but at about θ = .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 **JVJ/N** 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 $\pi = \theta/(\beta+\theta)$. (This is the distribution of the number of tries until the first success of independent trials each with success probability 1-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

$$\log L = n\log\alpha + n\log\beta + (\beta\text{-}1)\sum_{i=1}^{n}\log x_i - \alpha\sum_{i=1}^{n} x_i^{\beta}$$

$$\partial\log L/\partial\alpha = n/\alpha - \sum_{i=1}^{n} x_i^{\beta}$$

$$\partial\log L/\partial\beta = n/\beta + \sum_{i=1}^{n}\log x_i - \alpha\sum_{i=1}^{n}(\log x_i)x_i^{\beta}$$

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 $\qquad 1/(\alpha\beta) = \text{Cov}[\ln x_i, x_i^{\beta}]$. ~

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 $\beta = 1$.

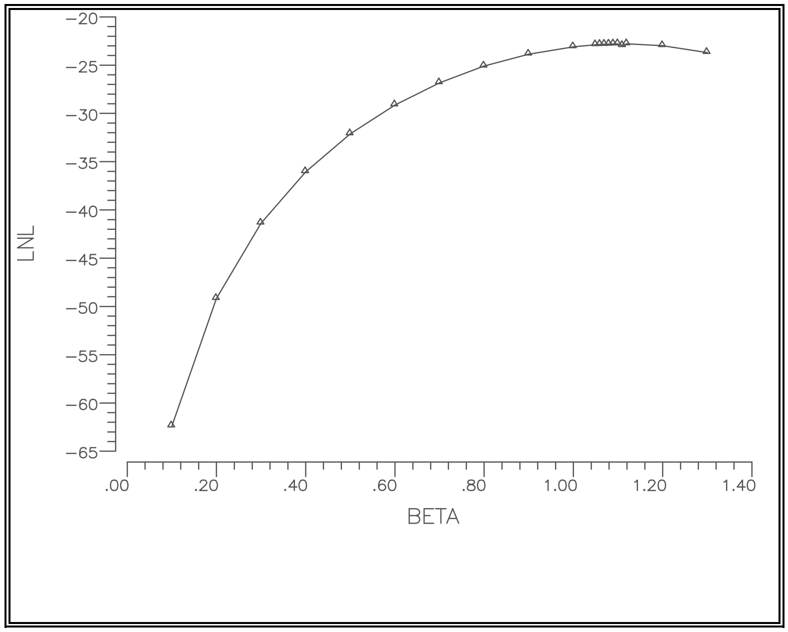
(d) Using the results of a. and c. carry out a likelihood ratio test of the hypothesis that $\beta = 1$.

(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.

$\beta \qquad \log L$

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 $\alpha = .88496$ and $\beta = 1$. 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 $\text{Prob}[c_{lower} < \chi^2[n-1] < c_{upper}] = 1 - \alpha$.

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 $(1-\alpha)$. 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

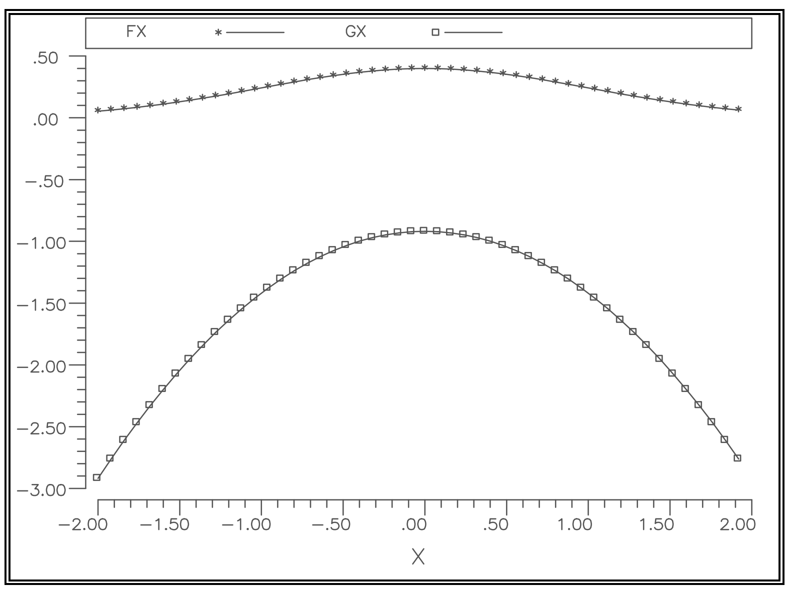
$$f(\beta) \;=\; \frac{1}{\sqrt{2\pi}}e^{-(\beta-c)^2/2}$$

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

$$df(\beta)/d\beta \;=\; \frac{1}{\sqrt{2\pi}}e^{-(\beta-c)^2/2}\,[-(\beta - c)] \;=\; 0 \;=\; -(\beta - c)\mathrm{f}(\beta).$$

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*(β) = log*f*(β) = $-(1/2)\ln(2\pi) - (1/2)(\beta - c)^2$. 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

$$\log L = \sum_{i=1}^{n} [-\lambda_i + y_i\ln\lambda_i - \ln y_i!] = -\sum_{i=1}^{n} e^{\boldsymbol{\beta}'\mathbf{x}_i} + \sum_{i=1}^{n} y_i\left(\boldsymbol{\beta}'\mathbf{x}_i\right) - \sum_{i=1}^{n}\log y_i!$$

$$= -\sum_{i=1}^{n} e^{\boldsymbol{\beta}'\mathbf{x}_i} + \boldsymbol{\beta}'\sum_{i=1}^{n}\mathbf{x}_i y_i - \sum_{i=1}^{n}\log y_i!$$

The necessary condition is $\text{M}\ln L/\text{M}\boldsymbol{\beta} = -\sum_{i=1}^{n}\mathbf{x}_i e^{\boldsymbol{\beta}'\mathbf{x}_i} + \sum_{i=1}^{n}\mathbf{x}_i y_i = \mathbf{0}$ or $\mathbf{X}\text{N}\mathbf{y} = \sum_{i=1}^{n}\mathbf{x}_i\lambda_i$. It is useful to note, since $E[y_i*\mathbf{x}_i] = \lambda_i = e^{\boldsymbol{\beta}\text{N}\mathbf{x}_i}$, 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 $\mathbf{X}\text{N}\mathbf{y} = \mathbf{X}\text{N}E[\mathbf{y}]$, which makes sense. We may write the first order condition as $\text{M}\ln L/\text{M}\boldsymbol{\beta} = \sum_{i=1}^{n}\mathbf{x}_i(y_i - \lambda_i) = \mathbf{0}$

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 $\partial\ln L/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}' = -\sum_{i=1}^{n}\left(e^{\boldsymbol{\beta}'\mathbf{x}_i}\right)\mathbf{x}_i\mathbf{x}_i' = -\sum_{i=1}^{n}\lambda_i\mathbf{x}_i\mathbf{x}_i'$. 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, $\partial\ln L/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}' = -\mathbf{Z}'\mathbf{Z}$ 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 $\hat{\beta}_{t+1} = \hat{\beta}_t + \left[\sum_{i=1}^{n} \hat{\lambda}_i \mathbf{x}_i \mathbf{x}_i'\right]^{-1} \left[\sum_{i=1}^{n} \mathbf{x}_i (y_i - \hat{\lambda}_i)\right]$.

(3) Each time we compute $\hat{\beta}_{t+1}$, we should check for convergence. Some possibilities are

(a) Gradient: Are the elements of $\partial \ln L/\partial \beta$ 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	$\ln L$	$\partial \ln L/\partial \alpha$	$\partial \ln L/\partial \beta$	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

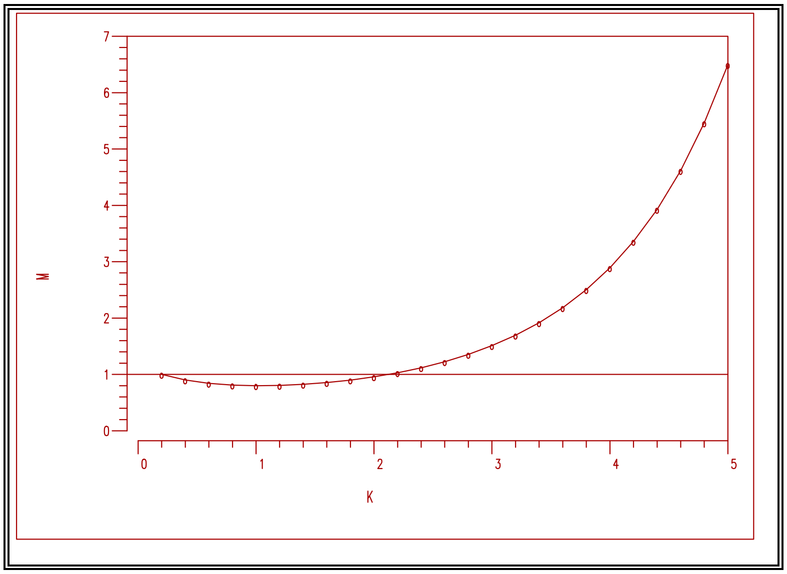
$$\left[\sum_{i=1}^{n} \hat{\lambda}_i \mathbf{x}_i \mathbf{x}_i'\right]^{-1} = \begin{bmatrix} .151044 & -.095961 \\ -.095961 & .0664665 \end{bmatrix}.$$

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 E.13, suppose that $E[y_i] = \mu$, 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 $E[\mathbf{H}_{\mu\mu}] = \mathbf{H}_{\mu\mu} = -\sum_{i=1}^{n}\frac{1}{\sigma_i^2}$. 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

$$\begin{bmatrix}\mu\\ \gamma\end{bmatrix}_{(t+1)} = \begin{bmatrix}\mu\\ \gamma\end{bmatrix}_{(t)} + \begin{bmatrix}\sum_{i=1}^{n}\frac{1}{\sigma_i^2(t)} & 0\\ 0 & \frac{1}{2}\mathbf{Z'Z}\end{bmatrix}^{-1}\begin{bmatrix}\sum_{i=1}^{n}\frac{y_i-\mu(t)}{\sigma_i^2(t)}\\ \frac{1}{2}\sum_{i=1}^{n}\mathbf{z}_i\left(\frac{(y_i-\mu(t))^2}{\sigma_i^2(t)}-1\right)\end{bmatrix}.$$

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.

	(Constrained model, $\mu = \bar{y}$)		(Unconstrained model)	
Iteration	log likelihood	δ	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).