Springer Texts in Business and Economics

Badi H. Baltagi

Solutions Manual for *Econometrics*

Third Edition

Springer

Springer Texts in Business and Economics

More information about this series at
http://www.springer.com/series/10099

Badi H. Baltagi

Solutions Manual for *Econometrics*

Third Edition

Springer

Badi H. Baltagi
Center for Policy Research
and Department of Economics
Syracuse University
Syracuse
NY, USA

ISSN 2192-4333 ISSN 2192-4341 (electronic)
ISBN 978-3-642-54547-4 ISBN 978-3-642-54548-1 (eBook)
DOI 10.1007/978-3-642-54548-1
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014948208

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

This manual provides solutions to selected exercises from each chapter of the fifth edition of *Econometrics* by Badi H. Baltagi.[1] Eviews and Stata as well as SAS® programs are provided for the empirical exercises. Some of the problems and solutions are obtained from *Econometric Theory* (ET) and these are reprinted with the permission of Cambridge University Press. I would like to thank Peter C.B. Phillips, and past editors of the Problems and Solutions section, Alberto Holly, Juan Dolado and Paolo Paruolo for their useful service to the econometrics profession. I would also like to thank my colleague (from Texas A&M) James M. Griffin for providing many empirical problems and data sets. I have also used three empirical data sets from Lott and Ray (1992). The reader is encouraged to apply these econometric techniques to their own data sets and to replicate the results of published articles. Instructors and students are encouraged to get other data sets from the Internet or journals that provide backup data sets to published articles. The *Journal of Applied Econometrics* and the *American Economic Review* are two such journals. In fact, the *Journal of Applied Econometrics* has a replication section for which I am serving as an editor. In my course I require my students to replicate an empirical paper.

I would like to thank my students Wei-Wen Xiong, Ming-Jang Weng, Kiseok Nam, Dong Li, Gustavo Sanchez, Long Liu and Liu Tian who solved several of the exercises. I would also like to thank Martina Bihn at Springer for her continuous support and professional editorial help.

Please report any errors, typos or suggestions to: Badi H. Baltagi, Center for Policy Research and Department of Economics, Syracuse University, Syracuse, New York 13244-1020, Telephone (315) 443-1630, Fax (315) 443-1081, or send Email to bbaltagi@ maxwell.syr.edu. My home page is www-cpr.maxwell.syr.edu/faculty/baltagi.

Data

The data sets used in this text can be downloaded from the Springer website. The address is: http://www.springer.com/978-3-642-54547-4. Please check the link "Samples, Supplements, Data Sets" from the right-hand column. There is also a readme file that describes the contents of each data set and its source.

[1] Baltagi, Badi H., *Econometrics*, 5th Ed., Springer-Verlag, Berlin, Heidelberg, 2011.

Reference

Lott, W.F. and S. Ray (1992), *Econometrics Problems and Data Sets* (Harcourt, Brace Jovanovich: San Diego, CA).

Contents

CHAPTER 1
What is Econometrics?

This chapter emphasizes that an econometrician has to be a competent mathematician and statistician who is an economist by training. It is the unification of statistics, economic theory and mathematics that constitutes econometrics. Each view point, by itself is necessary but not sufficient for a real understanding of quantitative relations in modern economic life, see Frisch (1933).

Econometrics aims at giving empirical content to economic relationships. The three key ingredients are economic theory, economic data, and statistical methods. Neither 'theory without measurement', nor 'measurement without theory' are sufficient for explaining economic phenomena. It is as Frisch emphasized their union that is the key for success in the future development of econometrics.

Econometrics provides tools for *testing* economic laws, such as purchasing power parity, the life cycle hypothesis, the wage curve, etc. These economic laws or hypotheses are testable with economic data. As Hendry (1980) emphasized "*The three golden rules of econometrics are test, test and test*."

Econometrics also provides quantitative *estimates* of price and income elasticities of demand, returns to scale in production, technical efficiency in cost functions, wage elasticities, etc. These are important for policy decision making. Raising the tax on a pack of cigarettes by 10%, how much will that reduce consumption of cigarettes? How much will it generate in tax revenues? What is the effect of raising minimum wage by $1 per hour on unemployment? What is the effect of raising beer tax on motor vehicle fatality?

Econometrics also provides *predictions* or *forecasts* about future interest rates, unemployment, or GNP growth. As Klein (1971) emphasized: "*Econometrics should give a base for economic prediction beyond experience if it is to be useful*."

Data in economics are not generated under ideal experimental conditions as in a physics laboratory. This data cannot be replicated and is most likely *measured with error*. Most of the time the data collected are not ideal for the economic question at hand. Griliches (1986, p. 1466) describes economic data as the world that we want to explain, and at the same time the source of all our trouble. The data's imperfections makes the econometrician's job difficult and sometimes impossible, yet these imperfections are what gives econometricians their legitimacy.

B.H. Baltagi, *Solutions Manual for Econometrics*, Springer Texts in Business and Economics, DOI 10.1007/978-3-642-54548-1_1,

Even though economists are increasingly getting involved in collecting their data and measuring variables more accurately and despite the increase in data sets and data storage and computational accuracy, some of the warnings given by Griliches (1986, p. 1468) are still valid today:

> econometricians want too much from the data and hence tend to be disappointed by the answers, because the data are incomplete and imperfect. In part it is our fault, the appetite grows with eating. As we get larger samples, we keep adding variables and expanding our models, until on the margin, we come back to the same insignificance levels.

Pesaran (1990, pp. 25–26) also summarizes some of the limitations of econometrics:

> There is no doubt that econometrics is subject to important limitations, which stem largely from the incompleteness of the economic theory and the non-experimental nature of economic data. But these limitations should not distract us from recognizing the fundamental role that econometrics has come to play in the development of economics as a scientific discipline. It may not be possible conclusively to reject economic theories by means of econometric methods, but it does not mean that nothing useful can be learned from attempts at testing particular formulations of a given theory against (possible) rival alternatives. Similarly, the fact that econometric modelling is inevitably subject to the problem of specification searches does not mean that the whole activity is pointless. Econometric models are important tools for forecasting and policy analysis, and it is unlikely that they will be discarded in the future. The challenge is to recognize their limitations and to work towards turning them into more reliable and effective tools. There seem to be no viable alternatives.

Econometrics have experienced phenomenal growth in the past 50 years. There are six volumes of the *Handbook of Econometrics*, most of it dealing with post 1960s research. A lot of the recent growth reflects the rapid advances in computing technology. The broad availability of micro data bases is a major advance which facilitated the growth of panel data methods (see Chap. 12) and microeconometric methods especially on sample selection and discrete choice (see Chap. 13) and that also lead to the award of the Nobel Prize in Economics to James Heckman and Daniel McFadden in 2000. The explosion in research in time series econometrics which lead to the development of ARCH and GARCH and cointegration (see Chap. 14) which also lead to the award of the Nobel Prize in Economics to Clive Granger and Robert Engle in 2003.

The challenge for the twenty-first century is to narrow the gap between theory and practice. Many feel that this gap has been widening with theoretical research growing more and more abstract and highly mathematical without an application in sight or a motivation for practical use. Heckman (2001) argues that econometrics is useful only if it helps economists conduct and interpret *empirical research* on economic data. He warns that the gap between econometric theory and empirical practice has grown over the past two decades. Theoretical econometrics becoming more closely tied to mathematical statistics. Although he finds nothing wrong, and much potential value, in using methods and ideas from other fields to improve empirical work in economics, he does warn of the risks involved in uncritically adopting the methods and mind set of the statisticians:

> Econometric methods uncritically adapted from statistics are not useful in many research activities pursued by economists. A theorem-proof format is poorly suited for analyzing economic data, which requires skills of synthesis, interpretation and empirical investigation. Command of statistical methods is only a part, and sometimes a very small part, of what is required to do first class empirical research.

Geweke et al. (2008) in the *The New Palgrave Dictionary* provide the following recommendations for the future:

> Econometric theory and practice seek to provide information required for informed decision-making in public and private economic policy. This process is limited not only by the adequacy of econometrics, but also by the development of economic theory and the adequacy of data and other information. Effective progress, in the future as in the past, will come from simultaneous improvements in econometrics, economic theory, and data. Research that specifically addresses the effectiveness of the interface between any two of these three in improving policy — to say nothing of all of them — necessarily transcends traditional subdisciplinary boundaries within economics. But it is precisely these combinations that hold the greatest promise for the social contribution of academic economics.

For a world wide ranking of econometricians as well as academic institutions in the field of econometrics, see Baltagi (2007).

References

Baltagi, B.H. (2007), “Worldwide Econometrics Rankings: 1989–2005,” *Econometric Theory*, 23: 952–1012.

Frisch, R. (1933), “Editorial,” *Econometrica*, 1: 1–14.

Geweke, J., J. Horowitz, and M. H. Pesaran (2008), “Econometrics,” *The New Palgrave Dictionary of Economics*. Second Edition. Eds. Steven N. Durlauf and Lawrence E. Blume. Palgrave Macmillan.

Griliches, Z. (1986), “Economic Data Issues,” in Z. Griliches and M.D. Intriligator (eds), *Handbook of Econometrics Vol. III* (North Holland: Amsterdam).

Heckman, J.J. (2001), “Econometrics and Empirical Economics,” *Journal of Econometrics*, 100: 3–5.

Hendry, D.F. (1980), “Econometrics - Alchemy or Science?” *Economica*, 47: 387–406.

Klein, L.R. (1971), “Whither Econometrics?” *Journal of the American Statistical Association*, 66: 415–421.

Pesaran, M.H. (1990), “Econometrics,” in J. Eatwell, M. Milgate and P. Newman; *The New Palgrave: Econometrics* (W.W. Norton and Company: New York).

CHAPTER 2
A Review of Some Basic Statistical Concepts

2.1 *Variance and Covariance of Linear Combinations of Random Variables.*

a. Let $Y = a + bX$, then $E(Y) = E(a + bX) = a + bE(X)$. Hence,

$$\begin{aligned}\mathrm{var}(Y) &= E[Y - E(Y)]^2 = E[a + bX - a - bE(X)]^2 = E[b(X - E(X))]^2 \\ &= b^2E[X - E(X)]^2 = b^2\,\mathrm{var}(X).\end{aligned}$$

Only the multiplicative constant b matters for the variance, not the additive constant a.

b. Let $Z = a + bX + cY$, then $E(Z) = a + bE(X) + cE(Y)$ and

$$\begin{aligned}\mathrm{var}(Z) &= E[Z - E(Z)]^2 = E[a + bX + cY - a - bE(X) - cE(Y)]^2 \\ &= E[b(X - E(X)) + c(Y - E(Y))]^2 \\ &= b^2E[X-E(X)]^2 + c^2E[Y-E(Y)]^2 + 2bc\,E[X-E(X)][Y-E(Y)] \\ &= b^2\mathrm{var}(X) + c^2\mathrm{var}(Y) + 2bc\,\mathrm{cov}(X, Y).\end{aligned}$$

c. Let $Z = a+bX+cY$, and $W = d+eX+fY$, then $E(Z) = a+bE(X)+cE(Y)$

$$E(W) = d + eE(X) + fE(Y)$$

and

$$\begin{aligned}\mathrm{cov}(Z, W) &= E[Z - E(Z)][W - E(W)] \\ &= E[b(X-E(X))+c(Y-E(Y))][e(X-E(X))+f(Y-E(Y))] \\ &= be\,\mathrm{var}(X) + cf\,\mathrm{var}(Y) + (bf + ce)\,\mathrm{cov}(X, Y).\end{aligned}$$

2.2 *Independence and Simple Correlation.*

a. Assume that X and Y are continuous random variables. The proof is similar if X and Y are discrete random variables and is left to the reader. If X and Y are independent, then $f(x, y) = f_1(x)f_2(y)$ where $f_1(x)$ is the marginal probability density function (p.d.f.) of X and $f_2(y)$ is the marginal p.d.f. of Y. In this case,

$$\begin{aligned}E(XY) &= \iint xyf(x, y)dxdy = \iint xyf_1(x)f_2(y)dxdy \\ &= \left(\int xf_1(x)dx\right)\left(\int yf_2(y)dy\right) = E(X)E(Y)\end{aligned}$$

B.H. Baltagi, *Solutions Manual for Econometrics*, Springer Texts in Business and Economics, DOI 10.1007/978-3-642-54548-1_2,

Hence,

$$\text{cov}(X, Y) = E[X - E(X)][Y - E(Y)] = E(XY) - E(X)E(Y)$$

$$= E(X)E(Y) - E(X)E(Y) = 0.$$

b. If $Y = a + bX$, then $E(Y) = a + bE(X)$ and $\text{cov}(X, Y) = E[X - E(X)][Y - E(Y)] = E[X - E(X)][a + bX - a - bE(X)] = b\,\text{var}(X)$ which takes the sign of b since var(X) is always positive. Hence,

$$\text{correl}(X, y) = \rho_{xy} = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)\text{var}(Y)}} = \frac{b\,\text{var}(X)}{\sqrt{\text{var}(X)\text{var}(Y)}}$$

but $\text{var}(Y) = b^2\,\text{var}(X)$ from problem 2.1a. Hence, $\rho_{XY} = \frac{b\,\text{var}(X)}{\sqrt{b^2(\text{var}(X))^2}} = \pm 1$ depending on the sign of b.

2.3 Zero Covariance Does Not Necessarily Imply Independence.

X	P(X)
−2	1/5
−1	1/5
0	1/5
1	1/5
2	1/5

$$E(X) = \sum_{X=-2}^{2} XP(X) = \frac{1}{5}[(-2) + (-1) + 0 + 1 + 2] = 0$$

$$E(X^2) = \sum_{X=-2}^{2} X^2P(X) = \frac{1}{5}[4 + 1 + 0 + 1 + 4] = 2$$

and $\text{var}(X) = 2$. For $Y = X^2$, $E(Y) = E(X^2) = 2$ and

$$E(X^3) = \sum_{X=-2}^{2} X^3P(X) = \frac{1}{5}[(-2)^3 + (-1)^3 + 0 + 1^3 + 2^3] = 0$$

In fact, any odd moment of X is zero. Therefore,

$$E(YX) = E(X^2.X) = E(X^3) = 0$$

and

$$\text{cov}(Y, X) = E(X - E(X))(Y - E(Y)) = E(X - 0)(Y - 2)$$
$$= E(XY) - 2E(X) = E(XY) = E(X^3) = 0$$

Hence, $\rho_{XY} = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)\text{var}(Y)}} = 0.$

2.4 The *Binomial Distribution*.

a. $\Pr[X = 5 \text{ or } 6] = \Pr[X = 5] + \Pr[X = 6]$
$= b(n = 20, X = 5, \theta = 0.1) + b(n = 20, X = 6, \theta = 0.1)$
$= \binom{20}{5}(0.1)^5(0.9)^{15} + \binom{20}{6}(0.1)^6(0.9)^{14}$
$= 0.0319 + 0.0089 = 0.0408.$

This can be easily done with a calculator, on the computer or using the Binomial tables, see Freund (1992).

b. $\binom{n}{n-X} = \frac{n!}{(n-X)!(n-n+X)!} = \frac{n!}{(n-X)!X!} = \binom{n}{X}$

Hence,

$$b(n, n - X, 1 - \theta) = \binom{n}{n-X}(1 - \theta)^{n-X}(1 - 1 + \theta)^{n-n+X}$$
$$= \binom{n}{X}(1 - \theta)^{n-X}\theta^X = b(n, X, \theta).$$

c. Using the MGF for the Binomial distribution given in problem 2.14a, we get

$M_X(t) = [(1 - \theta) + \theta e^t]^n.$

Differentiating with respect to t yields $M'_X(t) = n[(1 - \theta) + \theta e^t]^{n-1}\theta e^t$.

Therefore, $M'_X(0) = n\theta = E(X)$.

Differentiating $M'_X(t)$ again with respect to t yields

$M''_X(t) = n(n - 1)[(1 - \theta) + \theta e^t]^{n-2}(\theta e^t)^2 + n[(1 - \theta) + \theta e^t]^{n-1}\theta e^t.$

Therefore $M''_X(0) = n(n - 1)\theta^2 + n\theta = E(X^2)$.

Hence $\mathrm{var}(X) = E(X^2) - (E(X))^2 = n\theta + n^2\theta^2 - n\theta^2 - n^2\theta^2 = n\theta(1-\theta)$.

An alternative proof for $E(X) = \sum_{X=0}^{n} Xb(n, X, \theta)$. This entails factorial moments and the reader is referred to Freund (1992).

d. The likelihood function is given by

$$L(\theta) = f(X_1, .., X_n; \theta) = \theta^{\sum_{i=1}^{n} X_i}(1-\theta)^{n-\sum_{i=1}^{n} X_i}$$

so that $\log L(\theta) = \left(\sum_{i=1}^{n} X_i\right)\log\theta + \left(n - \sum_{i=1}^{n} X_i\right)\log(1-\theta)$

$$\frac{\partial \log L(\theta)}{\partial\theta} = \frac{\sum_{i=1}^{n} X_i}{\theta} - \frac{\left(n - \sum_{i=1}^{n} X_i\right)}{(1-\theta)} = 0.$$

Solving for θ one gets

$$\sum_{i=1}^{n} X_i - \theta\sum_{i=1}^{n} X_i - \theta n + \theta\sum_{i=1}^{n} X_i = 0$$

so that $\hat{\theta}_{mle} = \sum_{i=1}^{n} X_i/n = \bar{X}$.

e. $E(\bar{X}) = \sum_{i=1}^{n} E(X_i)/n = n\theta/n = \theta$.

Hence, $\bar{X}$ is unbiased for θ. Also, $\mathrm{var}(\bar{X}) = \mathrm{var}(X_i)/n = \theta(1-\theta)/n$ which goes to zero as $n \to \infty$. Hence, the sufficient condition for $\bar{X}$ to be consistent for θ is satisfied.

f. The joint probability function in part d can be written as

$$f(X_1, .., X_n; \theta) = \theta^{n\bar{X}}(1-\theta)^{n-n\bar{X}} = h(\bar{X}, \theta)g(X_1, \ldots, X_n)$$

where $h(\bar{X}, \theta) = \theta^{n\bar{X}}(1-\theta)^{n-n\bar{X}}$ and $g(X_1, .., X_n) = 1$ for all X_i's. The latter function is independent of θ in form and domain. Hence, by the factorization theorem, $\bar{X}$ is sufficient for θ.

g. $\bar{X}$ was shown to be MVU for θ for the Bernoulli case in Example 2 in the text.

h. From part (d), $L(0.2) = (0.2)^{\sum_{i=1}^{n} X_i}(0.8)^{n-\sum_{i=1}^{n} X_i}$ while $L(0.6) = (0.6)^{\sum_{i=1}^{n} X_i}(0.4)^{n-\sum_{i=1}^{n} X_i}$ with the likelihood ratio $\frac{L(0.2)}{L(0.6)} = \left(\frac{1}{3}\right)^{\sum_{i=1}^{n} X_i} 2^{n-\sum_{i=1}^{n} X_i}$

The uniformly most powerful critical region C of size $\alpha \leq 0.05$ is given by $\left(\frac{1}{3}\right)^{\sum_{i=1}^{n} X_i} 2^{n-\sum_{i=1}^{n} X_i} \leq k$ inside C. Taking logarithms of both sides

$$-\sum_{i=1}^{n} X_i(\log\ 3) + \left(n - \sum_{i=1}^{n} X_i\right)\log 2 < \log k$$

solving $\quad -\left(\sum_{i=1}^{n} X_i\right)\log\ 6 \leq K' \quad$ or $\quad \sum_{i=1}^{n} X_i \geq K$

where K is determined by making the size of $C = \alpha \leq 0.05$. In this case, $\sum_{i=1}^{n} X_i \sim b(n,\theta)$ and under H_o ; $\theta = 0.2$. Therefore, $\sum_{i=1}^{n} X_i \sim b(n = 20, \theta = 0.2)$. Hence, $\alpha = \Pr[b(n = 20, \theta = 0.2) \geq K] \leq 0.05$.

From the Binomial tables for $n = 20$ and $\theta = 0.2$, $K = 7$ gives $\Pr[b(n = 20, \theta = 0.2) \geq 7] = 0.0322$. Hence, $\sum_{i=1}^{n} X_i \geq 7$ is our required critical region.

i. The likelihood ratio test is

$$\frac{L(0.2)}{L(\hat{\theta}_{mle})} = \frac{(0.2)^{\sum_{i=1}^{n} X_i}(0.8)^{n-\sum_{i=1}^{n} X_i}}{(\bar{X})^{\sum_{i=1}^{n} X_i}\left(1-\bar{X}\right)^{n-\sum_{i=1}^{n} X_i}}$$

so that $LR = -2\log L(0.2) + 2\log L(\hat{\theta}_{mle})$

$$= -2\left[\sum_{i=1}^{n} X_i(\log 0.2 - \log \bar{X})\right] - 2\left[\left(n - \sum_{i=1}^{n} X_i\right)(\log 0.8 - \log(1-\bar{X}))\right].$$

This is given in Example 5 in the text for a general θ_o. The Wald statistic is given by $W = \dfrac{(\bar{X}-0.2)^2}{\bar{X}(1-\bar{X})/n}$

and the LM statistic is given by $LM = \dfrac{(\bar{X}-0.2)^2}{(0.2)(0.8)/n}$

Although, the three statistics, LR, LM and W look different, they are all based on $|\bar{X} - 2| \geq k$ and for a finite n, the same exact critical value could be obtained from the binomial distribution.

2.5 d. *The Wald, LR, and LM Inequality*. This is based on Baltagi (1994). The likelihood is given by Eq. (2.1) in the text.

$$L\left(\mu, \sigma^2\right) = (1/2\pi\sigma^2)^{n/2} e^{-(1/2\sigma^2)\sum_{i=1}^{n}(X_i-\mu)^2} \tag{1}$$

It is easy to show that the score is given by

$$S\left(\mu,\sigma^2\right) = \begin{pmatrix} \frac{n(\bar{X}-\mu)}{\sigma^2} \\ \frac{\sum_{i=1}^{n}(X_i-\mu)^2 - n\sigma^2}{2\sigma^4} \end{pmatrix}, \tag{2}$$

and setting $S(\mu,\sigma^2) = 0$ yields $\hat{\mu} = \bar{X}$ and $\hat{\sigma}^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2/n$. Under H_o, $\tilde{\mu} = \mu_0$ and

$$\tilde{\sigma}^2 = \sum_{i=1}^{n}(X_i - \mu_0)^2/n.$$

Therefore,

$$\log L\left(\tilde{\mu},\tilde{\sigma}^2\right) = -\frac{n}{2}\log\tilde{\sigma}^2 - \frac{n}{2}\log\ 2\pi - \frac{n}{2} \tag{3}$$

and

$$\log L\left(\hat{\mu},\hat{\sigma}^2\right) = -\frac{n}{2}\log\hat{\sigma}^2 - \frac{n}{2}\log 2\pi - \frac{n}{2}. \tag{4}$$

Hence,

$$LR = n\log\left[\frac{\sum_{i=1}^{n}(X_i-\mu_0)^2}{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}\right]. \tag{5}$$

It is also known that the information matrix is given by

$$I\begin{pmatrix}\mu \\ \sigma_2\end{pmatrix} = \begin{bmatrix}\frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4}\end{bmatrix}. \tag{6}$$

Therefore,

$$W = (\hat{\mu}-\mu_0)^2\,\hat{I}_{11} = \frac{n^2\left(\bar{X}-\mu_0\right)^2}{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}, \tag{7}$$

where $\hat{I}_{11}$ denotes the (1,1) element of the information matrix evaluated at the unrestricted maximum likelihood estimates. It is easy to show from (1) that

$$\log L\left(\tilde{\mu},\hat{\sigma}^2\right) = -\frac{n}{2}\log\hat{\sigma}^2 - \frac{n}{2}\log 2\pi - \frac{\sum_{i=1}^{n}(X_i-\mu_0)^2}{2\hat{\sigma}^2}. \tag{8}$$

Hence, using (4) and (8), one gets

$$-2\log\left[L\left(\tilde{\mu},\hat{\sigma}^2\right)/L(\hat{\mu},\hat{\sigma}^2)\right] = \frac{\sum_{i=1}^{n}(X_i-\mu_0)^2 - n\hat{\sigma}^2}{\hat{\sigma}^2} = W, \tag{9}$$

and the last equality follows from (7). Similarly,

$$LM = S^2\left(\tilde{\mu},\tilde{\sigma}^2\right)\tilde{I}^{11} = \frac{n^2(\bar{X}-\mu_0)^2}{\sum_{i=1}^{n}(X_i-\mu_0)^2}, \tag{10}$$

where $\tilde{I}^{11}$ denotes the (1,1) element of the inverse of the information matrix evaluated at the restricted maximum likelihood estimates. From (1), we also get

$$\log L\left(\hat{\mu},\tilde{\sigma}^2\right) = -\frac{n}{2}\log\tilde{\sigma}^2 - \frac{n}{2}\log 2\pi - \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{2\tilde{\sigma}^2} \tag{11}$$

Hence, using (3) and (11), one gets

$$-2\log\left[L\left(\tilde{\mu},\tilde{\sigma}^2\right)/L\left(\hat{\mu},\tilde{\sigma}^2\right)\right] = n - \frac{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}{\tilde{\sigma}^2} = LM, \tag{12}$$

where the last equality follows from (10). $L(\tilde{\mu},\tilde{\sigma}^2)$ is the restricted maximum; therefore, $\log L(\tilde{\mu},\hat{\sigma}^2) \le \log L(\tilde{\mu},\tilde{\sigma}^2)$, from which we deduce that $W \ge LR$. Also, $L(\hat{\mu},\hat{\sigma}^2)$ is the unrestricted maximum; therefore $\log L(\hat{\mu},\hat{\sigma}^2) \ge \log L(\hat{\mu},\tilde{\sigma}^2)$, from which we deduce that $LR \ge LM$.

An alternative derivation of this inequality shows first that

$$\frac{LM}{n} = \frac{W/n}{1+(W/n)} \quad \text{and} \quad \frac{LR}{n} = \log\left(1+\frac{W}{n}\right).$$

Then one uses the fact that $y \ge \log(1+y) \ge y/(1+y)$ for $y = W/n$.

2.6 *Poisson Distribution*.

a. Using the MGF for the Poisson derived in problem 2.14c one gets

$M_x(t) = e^{\lambda(e^t-1)}$.

Differentiating with respect to t yields

$M'_x(t) = e^{\lambda(e^t-1)}\lambda e^t$.

Evaluating $M'_X(t)$ at $t = 0$, we get

$M'_x(0) = E(X) = \lambda$.

Similarly, differentiating $M'_X(t)$ once more with respect to t, we get

$M''_x(t) = e^{\lambda(e^t-1)}\left(\lambda e^t\right)^2 + e^{\lambda(e^t-1)}\lambda e^t$

evaluating it at $t = 0$ gives

$M''_x(0) = \lambda^2 + \lambda = E(X^2)$

so that

$\text{var}(X) = E(X^2) - (E(X)^2) = \lambda^2 + \lambda - \lambda^2 = \lambda$.

Hence, the mean and variance of the Poisson are both equal to λ.

b. The likelihood function is

$$L(\lambda) = \frac{e^{-n\lambda}\lambda^{\sum\limits_{i=1}^{n} X_i}}{X_1!X_2!..X_n!}$$

so that

$$\log L(\lambda) = -n\lambda + \left(\sum_{i=1}^{n} X_i\right)\log\lambda - \sum_{i=1}^{n}\log X_i!$$

$$\frac{\partial \log L(\lambda)}{\partial\lambda} = -n + \frac{\sum\limits_{i=1}^{n} X_i}{\lambda} = 0.$$

Solving for λ, yields $\hat{\lambda}_{mle} = \bar{X}$.

c. The method of moments equates $E(X)$ to $\bar{X}$ and since $E(X) = \lambda$ the solution is $\hat{\lambda} = \bar{X}$, same as the ML method.

d. $E(\bar{X}) = \sum\limits_{i=1}^{n} E(X_i)/n = n\lambda/n = \lambda$. Therefore $\bar{X}$ is unbiased for λ. Also, $\text{var}(\bar{X}) = \frac{\text{var}(X_i)}{n} = \frac{\lambda}{n}$ which tends to zero as $n \to \infty$. Therefore, the sufficient condition for $\bar{X}$ to be consistent for λ is satisfied.

e. The joint probability function can be written as

$$f(X_1,..,X_n;\lambda) = e^{-n\lambda}\lambda^{n\bar{X}}\frac{1}{X_1!..X_n!} = h\left(\bar{X},\lambda\right)\ g\left(X_1,..,X_n\right)$$

where $h(\bar{X},\lambda) = e^{-n\lambda}\lambda^{n\bar{X}}$ and $g(X_1,..,X_n) = \frac{1}{X_1!..X_n!}$. The latter is independent of λ in form and domain. Therefore, $\bar{X}$ is a sufficient statistic for λ.

f. $\log f(X;\lambda) = -\lambda + X\log\lambda - \log X!$

and

$$\frac{\partial \log f(X;\lambda)}{\partial\lambda} = -1 + \frac{X}{\lambda}$$

$$\frac{\partial^2 \log f(X;\lambda)}{\partial\lambda^2} = \frac{-X}{\lambda^2}.$$

The Cramér–Rao lower bound for any unbiased estimator $\hat{\lambda}$ of λ is given by

$$\mathrm{var}\left(\hat{\lambda}\right) \geq -\frac{1}{nE\left(\frac{\partial^2 \log\ f(X;\lambda)}{\partial\lambda^2}\right)} = \frac{\lambda^2}{n\,E(X)} = \frac{\lambda}{n}.$$

But $\mathrm{var}(\bar{X}) = \lambda/n$, see part (d). Hence, $\bar{X}$ attains the Cramér–Rao lower bound.

g. The likelihood ratio is given by

$$\frac{L(2)}{L(4)} = \frac{e^{-2n}2^{\sum\limits_{i=1}^{n} X_i}}{e^{-4n}4^{\sum\limits_{i=1}^{n} X_i}}$$

The uniformly most powerful critical region C of size $\alpha \leq 0.05$ is given by

$$e^{2n}\left(\frac{1}{2}\right)^{\sum\limits_{i=1}^{n} X_i} \leq k \text{ inside C}$$

Taking logarithms of both sides and rearranging terms, we get

$$-\left(\sum_{i=1}^{n} X_i\right)\log 2 \leq K'$$

or

$$\sum_{i=1}^{n} X_i \geq K$$

where K is determined by making the size of $C = \alpha \le 0.05$. In this case, $\sum_{i=1}^{n} X_i \sim \text{Poisson}(n\lambda)$ and under H_o; $\lambda = 2$. Therefore, $\sum_{i=1}^{n} X_i \sim$ Poisson ($\lambda = 18$). Hence $\alpha = \Pr[\text{Poisson}(18) \ge K] \le 0.05$.

From the Poisson tables, for $\lambda = 18$, $K = 26$ gives $\Pr[\text{Poisson}(18) \ge 26] = 0.0446$. Hence, $\sum_{i=1}^{n} X_i \ge 26$ is our required critical region.

h. The likelihood ratio test is

$$\frac{L(2)}{L(\bar{X})} = \frac{e^{-2n} 2^{\sum_{i=1}^{n} X_i}}{e^{-n\bar{X}} (\bar{X})^{\sum_{i=1}^{n} X_i}} = e^{n(\bar{X}-2)} \left(\frac{2}{\bar{X}}\right)^{\sum_{i=1}^{n} X_i}$$

so that

$$LR = -2 \log L(2) + 2 \log L(\bar{X}) = -2n(\bar{X}-2) - 2\sum_{i=1}^{n} X_i \left[\log 2 - \log \bar{X}\right].$$

In this case,

$$C(\lambda) = \left|\frac{\partial^2 \log L(\lambda)}{\partial \lambda^2}\right| = \left|-\frac{\sum_{i=1}^{n} X_i}{\lambda^2}\right|$$

and

$$I(\lambda) = -E\left[\frac{\partial^2 \log L(\lambda)}{\partial \lambda^2}\right] = \frac{n\lambda}{\lambda^2} = \frac{n}{\lambda}.$$

The Wald statistic is based upon

$$W = (\bar{X}-2)^2 I\left(\hat{\lambda}_{mle}\right) = (\bar{X}-2)^2 \cdot \frac{n}{\bar{X}}$$

using the fact that $\hat{\lambda}_{mle} = \bar{X}$. The LM statistic is based upon

$$LM = S^2(2) I^{-1}(2) = \frac{n^2(\bar{X}-2)^2}{4} \cdot \frac{2}{n} = \frac{n(\bar{X}-2)^2}{2}.$$

Note that all three test statistics are based upon $|\bar{X} - 2| \ge K$, and for finite n the same exact critical value could be obtained using the fact that $\sum_{i=1}^{n} X_i$ has a Poisson distribution, see part (g).

2.7 *The Geometric Distribution.*

a. Using the MGF for the Geometric distribution derived in problem 2.14d, one gets

$$M_x(t) = \frac{\theta e^t}{[1-(1-\theta)e^t]}.$$

Differentiating it with respect to t yields

$$M_X'(t) = \frac{\theta e^t\left[1-(1-\theta)e^t\right] + (1-\theta)e^t\theta e^t}{[1-(1-\theta)e^t]^2} = \frac{\theta e^t}{[1-(1-\theta)e^t]^2}$$

evaluating $M_X'(t)$ at $t = 0$, we get

$$M_x'(0) = E(X) = \frac{\theta}{\theta^2} = \frac{1}{\theta}.$$

Similarly, differentiating $M_X'(t)$ once more with respect to t, we get

$$M_x''(t) = \frac{\theta e^t\left[1-(1-\theta)e^t\right]^2 + 2\left[1-(1-\theta)e^t\right](1-\theta)e^t\theta e^t}{[1-(1-\theta)e^t]^4}$$

evaluating $M_X''(t)$ at $t = 0$, we get

$$M_x''(0) = E(X^2) = \frac{\theta^3 + 2\theta^2(1-\theta)}{\theta^4} = \frac{2\theta^2-\theta^3}{\theta^4} = \frac{2-\theta}{\theta^2}$$

so that

$$\text{var}(X) = E(X^2) - (E(X))^2 = \frac{2-\theta}{\theta^2} - \frac{1}{\theta^2} = \frac{1-\theta}{\theta^2}.$$

b. The likelihood function is given by

$$L(\theta) = \theta^n(1-\theta)^{\sum_{i=1}^{n} X_i - n}$$

so that

$$\log L(\theta) = n\log\theta + \left(\sum_{i=1}^{n} X_i - n\right)\log(1-\theta)$$

$$\frac{\partial \log L(\theta)}{\partial\theta} = \frac{n}{\theta} - \frac{\sum_{i=1}^{n} X_i - n}{(1-\theta)} = 0$$

solving for θ one gets

$$n(1-\theta) - \theta\sum_{i=1}^{n} X_i + n\theta = 0$$

or

$$n = \theta \sum_{i=1}^{n} X_i$$

which yields

$$\hat{\theta}_{mle} = n / \sum_{i=1}^{n} X_i = \frac{1}{\bar{X}}.$$

The method of moments estimator equates

$$E(X) = \bar{X}$$

so that

$$\frac{1}{\hat{\theta}} = \bar{X} \quad \text{or} \quad \hat{\theta} = \frac{1}{\bar{X}}$$

which is the same as the MLE.

2.8 The *Uniform Density*.

a. $E(X) = \int_0^1 x\,dx = \frac{1}{2}[x^2]_0^1 = \frac{1}{2}$
$E(X^2) = \int_0^1 x^2 dx = \frac{1}{3}\left[x^3\right]_0^1 = \frac{1}{3}$
so that $\text{var}(X) = E(X^2) - (E(X))^2 = \frac{1}{3} - \frac{1}{4} = \frac{4-3}{12} = \frac{1}{12}$.

b. $\Pr[0.1 < X < 0.3] = \int_{0.1}^{0.3} dx = 0.3 - 0.1 = 0.2$. It does not matter if we include the equality signs $\Pr[0.1 \le X \le 0.3]$ since this is a continuous random variable. Note that this integral is the area of the rectangle, for X between 0.1 and 0.3 and height equal to 1. This is just the length of this rectangle, i.e., $0.3 - 0.1 = 0.2$.

2.9 The *Exponential Distribution*.

a. Using the MGF for the exponential distribution derived in problem 2.14e, we get

$$M_x(t) = \frac{1}{(1 - \theta t)}.$$

Differentiating with respect to t yields

$$M_x'(t) = \frac{\theta}{(1-\theta t)^2}.$$

Therefore

$$M_X{}'(0) = \theta = E(X).$$

Differentiating $M_X'(t)$ with respect to t yields

$$M_X''(t) = \frac{2\theta^2\,(1-\theta t)}{(1-\theta t)^4} = \frac{2\theta^2}{(1-\theta t)^3}.$$

Therefore

$$M_X''(0) = 2\theta^2 = E(X^2).$$

Hence

$$\mathrm{var}(X) = E(X^2) - (E(X))^2 = 2\theta^2 - \theta^2 = \theta^2.$$

b. The likelihood function is given by

$$L(\theta) = \left(\frac{1}{\theta}\right)^n e^{-\sum_{i=1}^{n} X_i/\theta}$$

so that

$$\log L(\theta) = -n\log\theta - \frac{\sum_{i=1}^{n} X_i}{\theta}$$

$$\frac{\partial \log L(\theta)}{\partial\theta} = \frac{-n}{\theta} + \frac{\sum_{i=1}^{n} X_i}{\theta^2} = 0$$

solving for θ one gets

$$\sum_{i=1}^{n} X_i - n\theta = 0$$

so that

$$\hat{\theta}_{mle} = \bar{X}.$$

c. The method of moments equates $E(X) = \bar{X}$. In this case, $E(X) = \theta$, hence $\hat{\theta} = \bar{X}$ is the same as MLE.

d. $E(\bar{X}) = \sum_{i=1}^{n} E(X_i)/n = n\theta/n = \theta$. Hence, $\bar{X}$ is unbiased for θ. Also, $\text{var}(\bar{X}) = \text{var}(X_i)/n = \theta^2/n$ which goes to zero as $n \to \infty$. Hence, the sufficient condition for $\bar{X}$ to be consistent for θ is satisfied.

e. The joint p.d.f. is given by

$$f(X_1, .., X_n; \theta) = \left(\frac{1}{\theta}\right)^n e^{-\sum_{i=1}^{n} X_i/\theta} = e^{-n\bar{X}/\theta}\left(\frac{1}{\theta}\right)^n = h(\bar{X};\theta)g(X_1,\ldots,X_n)$$

where $h(\bar{X};\theta) = e^{-n\bar{X}/\theta}\left(\frac{1}{\theta}\right)^n$ and $g(X_1,\ldots,X_n) = 1$ independent of θ in form and domain. Hence, by the factorization theorem, $\bar{X}$ is a sufficient statistic for θ.

f. $\log f(X;\theta) = -\log\theta - \frac{X}{\theta}$

and

$$\frac{\partial \log f(X;\theta)}{\partial\theta} = \frac{-1}{\theta} + \frac{X}{\theta^2} = \frac{X-\theta}{\theta^2}$$

$$\frac{\partial^2 \log f(X;\theta)}{\partial\theta^2} = \frac{1}{\theta^2} - \frac{2X\theta}{\theta^4} = \frac{\theta - 2X}{\theta^3}$$

The Cramér–Rao lower bound for any unbiased estimator $\hat{\theta}$ of θ is given by

$$\text{var}\left(\hat{\theta}\right) \geq \frac{-1}{nE\left(\frac{\partial^2 \log f(X;\theta)}{\partial\theta^2}\right)} = \frac{-\theta^3}{nE\left(\theta - 2X\right)} = \frac{\theta^2}{n}.$$

But $\text{var}(\bar{X}) = \theta^2/n$, see part (d). Hence, $\bar{X}$ attains the Cramér–Rao lower bound.

g. The likelihood ratio is given by

$$\frac{L(1)}{L(2)} = \frac{e^{-\sum_{i=1}^{n} X_i}}{2^{-n}e^{-\sum_{i=1}^{n} X_i/2}} = 2^n e^{-\sum_{i=1}^{n} x_i/2}.$$

The uniformly most powerful critical region C of size $\alpha \leq 0.05$ is given by $2^n\, e^{-\sum_{i=1}^{n} X_i/2} \leq k$ inside C.

Taking logarithms of both sides and rearranging terms, we get

$$n\log 2 - \left(\sum_{i=1}^{n} X/2\right) \leq K'$$

or

$$\sum_{i=1}^{n} X_i \geq K$$

where K is determined by making the size of $C = \alpha \leq 0.05$. In this case, $\sum_{i=1}^{n} X_i$ is distributed as a Gamma p.d.f. with $\beta = \theta$ and $\alpha = n$. Under H_o; $\theta = 1$. Therefore,

$$\sum_{i=1}^{n} X_i \sim Gamma(\alpha = n, \beta = 1).$$

Hence, $\Pr[\text{Gamma}(\alpha = n,\ \beta = 1) \geq K] \leq 0.05$

K should be determined from the integral $\int_K^\infty \frac{1}{\Gamma(n)} x^{n-1} e^{-x}\, dx = 0.05$ for $n = 20$.

h. The likelihood ratio test is

$$\frac{L(1)}{L(\bar{X})} = \frac{e^{-\sum_{i=1}^{n} X_i}}{\left(\frac{1}{\bar{X}}\right)^n e^{-n}}$$

so that

$$LR = -2\log L(1) + 2\log L(\bar{X}) = 2\sum_{i=1}^{n} X_i - 2n\log \bar{X} - 2n.$$

In this case,

$$C(\theta) = \left|\frac{\partial^2 \log L(\theta)}{\partial \theta^2}\right| = \left|\frac{n}{\theta^2} - \frac{2\sum_{i-1}^{n} X_i}{\theta^3}\right| = \left|\frac{n\theta - 2\sum_{i=1}^{n} X_i}{\theta^3}\right|$$

and

$$I(\theta) = -E\left[\frac{\partial^2 \ell n L(\theta)}{\partial \theta^2}\right] = \frac{n}{\theta^2}.$$

The Wald statistic is based upon

$$W = (\bar{X} - 1)^2\, I\left(\hat{\theta}_{mle}\right) = (\bar{X} - 1)^2 \frac{n}{\bar{X}^2}$$

using the fact that $\hat{\theta}_{mle} = \bar{X}$. The LM statistic is based upon

$$LM = S^2(1) I^{-1}(1) = \left(\sum_{i=1}^{n} X_i - n\right)^2 \frac{1}{n} = n(\bar{X} - 1)^2.$$

All three test statistics are based upon $|\bar{X}-1| \geq k$ and, for finite n, the same exact critical value could be obtained using the fact that $\sum_{i=1}^{n} X_i$ is Gamma ($\alpha = n$, and $\beta = 1$) under H_o, see part (g).

2.10 The *Gamma Distribution*.

a. Using the MGF for the Gamma distribution derived in problem 2.14f, we get

$M_X(t) = (1-\beta t)^{-\alpha}$.

Differentiating with respect to t yields

$M_X'(t) = -\alpha(1-\beta t)^{-\alpha-1}(-\beta) = \alpha\beta(1-\beta t)^{-\alpha-1}$.

Therefore

$M_x'(0) = \alpha\beta = E(X)$.

Differentiating $M_X{}'(t)$ with respect to t yields

$M_x''(t) = -\alpha\beta(\alpha+1)(1-\beta t)^{-\alpha-2}(-\beta) = \alpha\beta^2(\alpha+1)(1-\beta t)^{-\alpha-2}$.

Therefore

$M_x''(0) = \alpha^2\beta^2 + \alpha\beta^2 = E\left(X^2\right)$.

Hence

$\mathrm{var}(X) = E\left(X^2\right) - (E(X))^2 = \alpha\beta^2$.

b. The method of moments equates

$E(X) = \bar{X} = \alpha\beta$

and

$$E(X^2) = \sum_{i=1}^{n} X_i^2/n = \alpha^2\beta^2 + \alpha\beta^2.$$

These are two non-linear equations in two unknowns. Substitute $\alpha = \bar{X}/\beta$ into the second equation, one gets

$$\sum_{i=1}^{n} X_i^2/n = \bar{X}^2 + \bar{X}\beta$$

Hence,

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n\bar{X}}$$

and

$$\hat{\alpha} = \frac{n\bar{X}^2}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}.$$

c. For $\alpha = 1$ and $\beta = \theta$, we get

$$f(X; \alpha = 1, \beta = \theta) = \frac{1}{\Gamma(1)\theta} X^{1-1} e^{-X/\theta} \quad \text{for} \quad X > 0 \quad \text{and} \quad \theta > 0,$$

$$= \frac{1}{\theta} e^{-X/\theta}$$

which is the exponential p.d.f.

d. For $\alpha = r/2$ and $\beta = 2$, the Gamma $(\alpha = r/2,\ \beta = 2)$ is χ_r^2. Hence, from part (a), we get

$$E(X) = \alpha\beta = (r/2)(2) = r$$

and

$$\text{var}(X) = \alpha\beta^2 = (r/2)(4) = 2r.$$

The expected value of a χ_r^2 is r and its variance is 2r.

e. The joint p.d.f. for $\alpha = r/2$ and $\beta = 2$ is given by

$$f(X_1, .., X_n; \alpha = r/2, \beta = 2) = \left(\frac{1}{\Gamma(r/2)2^{r/2}}\right)^n (X_1 \ldots X_n)^{\frac{r}{2}-1} e^{-\sum_{i=1}^{n} X_i/2}$$
$$= h(X_1, .., X_n; r) g(X_1, .., X_n)$$

where

$$h(X_1,..,X_n;r) = \left(\frac{1}{\Gamma(r/2)\,2^{r/2}}\right)^n (X_1\ldots X_n)^{\frac{r}{2}-1}$$

and $g(X_1,..,X_n) = e^{-\sum\limits_{i=1}^{n} X_i/2}$ independent of r in form and domain. Hence, by the factorization theorem $(X_1\ldots X_n)$ is a sufficient statistic for r.

f. Let $X_1,..,X_m$ denote independent N(0, 1) random variables. Then, $X_1^2,..,X_m^2$ will be independent χ_1^2 random variables and $Y = \sum\limits_{i=1}^{m} X_i^2$ will be χ_m^2. The sum of m independent χ_1^2 random variables is a χ_m^2 random variable.

2.12 The *t-distribution with r Degrees of Freedom.*

a. If $X_1,..,X_n$ are $IIN(\mu,\sigma^2)$, then $\bar{X} \sim N(\mu,\sigma^2/n)$ and $z = \frac{(\bar{X}-\mu)}{\sigma/\sqrt{n}}$ is N(0, 1).

b. $(n-1)s^2/\sigma^2 \sim \chi_{n-1}^2$. Dividing our N(0, 1) random variables z in part (a), by the square-root of our χ_{n-1}^2 random variable in part (b), divided by its degrees of freedom, we get

$$t = \frac{(\bar{X}-\mu)/\sigma/\sqrt{n}}{\sqrt{\frac{(n-1)s^2}{\sigma^2}/(n-1)}} = \frac{(\bar{X}-\mu)}{s/\sqrt{n}}.$$

Using the fact that $\bar{X}$ is independent of s^2, this has a t-distribution with (n − 1) degrees of freedom.

c. The 95% confidence interval for μ would be based on the t-distribution derived in part (b) with (n − 1) = 15 degrees of freedom.

$$t = \frac{\bar{X}-\mu}{s/\sqrt{n}} = \frac{20-\mu}{2/\sqrt{16}} = \frac{20-\mu}{1/2} = 40-2\mu$$

$\Pr[-t_{\alpha/2} < t < t_{\alpha/2}] = 1-\alpha = 0.95$

From the t-tables with 15 degrees of freedom, $t_{0.025} = 2.131$. Hence

$\Pr[-2.131 < 40-2\mu < 2.131] = 0.95.$

rearranging terms, one gets

$\Pr[37.869/2 < \mu < 42.131/2] = 0.95$

$\Pr[18.9345 < \mu < 21.0655] = 0.95.$

2.13 The *F-distribution*.

$$(n_1 - 1)\, s_1^2/\sigma_1^2 = 24(15.6)/\sigma_1^2 \sim \chi^2_{24}$$

also

$$(n_2 - 1)s_2^2/\sigma_2^2 = 30\,(18.9)\,/\sigma_2^2 \sim \chi^2_{30}.$$

Therefore, under H_o; $\sigma_1^2 = \sigma_2^2$

$$F = s_2^2/s_1^2 = \frac{18.9}{15.6} = 1.2115 \sim F_{30,24}.$$

Using the F-tables with 30 and 24 degrees of freedom, we find $F_{.05,30,24} = 1.94$. Since the observed F-statistic 1.2115 is less than 1.94, we do not reject H_o that the variance of the two shifts is the same.

2.14 *Moment Generating Function (MGF).*

a. For the *Binomial Distribution*,

$$\begin{aligned} M_{x(t)} = E(e^{Xt}) &= \sum_{X=0}^{n} \binom{n}{X} e^{Xt}\theta^X (1-\theta)^{n-X} \\ &= \sum_{X=0}^{n} \binom{n}{X} \left(\theta e^t\right)^X (1-\theta)^{n-X} \\ &= \left[(1-\theta) + \theta e^t\right]^n \end{aligned}$$

where the last equality uses the binomial expansion $(a+b)^n = \sum_{X=0}^{n} \binom{n}{X} a^X b^{n-X}$ with $a = \theta e^t$ and $b = (1-\theta)$. This is the fundamental relationship underlying the binomial probability function and what makes it a proper probability function.

b. For the *Normal Distribution*,

$$\begin{aligned} M_X(t) = E\left(e^{Xt}\right) &= \int_{-\infty}^{+\infty} e^{Xt} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(X-\mu)^2} dx \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2\sigma^2}\left\{X^2 - 2\mu X + \mu^2 - Xt2\sigma^2\right\}} dx \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2\sigma^2}\left\{X^2 - 2(\mu + t\sigma^2)X + \mu^2\right\}} dx \end{aligned}$$

completing the square

$$M_x(t) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\frac{1}{2\sigma^2}\left\{\left[x-\left(\mu+t\sigma^2\right)\right]^2-\left(\mu+t\sigma^2\right)^2+\mu^2\right\}}dx$$
$$= e^{-\frac{1}{2\sigma^2}\left[\mu^2-\mu^2-2\mu t\sigma^2-t^2\sigma^4\right]}$$

The remaining integral integrates to 1 using the fact that the Normal density is proper and integrates to one. Hence $M_x(t) = e^{\mu t+\frac{1}{2}\sigma^2t^2}$ after some cancellations.

c. For the *Poisson Distribution*,

$$M_x(t) = E(e^{Xt}) = \sum_{X=0}^{\infty} e^{Xt}\frac{e^{-\lambda}\lambda^X}{X!} = \sum_{X=0}^{\infty}\frac{e^{-\lambda}(\lambda e^t)^X}{X!}$$
$$= e^{-\lambda}\sum_{X=0}^{\infty}\frac{(\lambda e^t)^X}{X!} = e^{\lambda}e^t - \lambda = e^{\lambda(e^t-1)}$$

where the fifth equality follows from the fact that $\sum\limits_{X=0}^{\infty}\frac{a^X}{X!} = e^a$ and in this case $a = \lambda e^t$. This is the fundamental relationship underlying the Poisson distribution and what makes it a proper probability function.

d. For the *Geometric Distribution*,

$$M_X(t) = E(e^{Xt}) = \sum_{X=1}^{\infty}\theta(1-\theta)^{X-1}e^{Xt} = \theta\sum_{X=1}^{\infty}(1-\theta)^{X-1}e^{(X-1)t}e^t$$
$$= \theta e^t\sum_{X=1}^{\infty}\left[(1-\theta)\,e^t\right]^{X-1} = \frac{\theta e^t}{1-(1-\theta)e^t}$$

where the last equality uses the fact that $\sum\limits_{X=1}^{\infty}a^{X-1} = \frac{1}{1-a}$ and in this case $a = (1-\theta)e^t$. This is the fundamental relationship underlying the Geometric distribution and what makes it a proper probability function.

e. For the *Exponential Distribution*,

$$M_X(t) = E(e^{Xt}) = \int_0^\infty \frac{1}{\theta} e^{-X/\theta} e^{Xt} dx$$

$$= \frac{1}{\theta} \int_0^\infty e^{-X[\frac{1}{\theta}-t]} dx$$

$$= \frac{1}{\theta} \int_0^\infty e^{-X[\frac{1-\theta t}{\theta}]} dx$$

$$= \frac{1}{\theta} \cdot \frac{-\theta}{(1-\theta t)} \left[e^{-X\left(\frac{1-\theta t}{\theta}\right)}\right]_0^\infty = (1-\theta t)^{-1}$$

f. For the *Gamma Distribution*,

$$M_X(t) = E(e^{Xt}) = \int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} X^{\alpha-1} e^{-X/\beta} e^{Xt}\, dx$$

$$= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty X^{\alpha-1} e^{-X\left(\frac{1}{\beta}-t\right)} dx$$

$$= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty X^{\alpha-1} e^{-X\left(\frac{1-\beta t}{\beta}\right)} dx$$

The Gamma density is proper and integrates to one using the fact that $\int_0^\infty X^{\alpha-1} e^{-X/\beta} dx = \Gamma(\alpha)\beta^\alpha$. Using this fundamental relationship for the last integral, we get

$$M_x(t) = \frac{1}{\Gamma(\alpha)\beta^\alpha} \cdot \Gamma(\alpha) \left(\frac{\beta}{1-\beta t}\right)^\alpha = (1-\beta t)^{-\alpha}$$

where we substituted $\beta/(1-\beta t)$ for the usual β. The χ_r^2 distribution is Gamma with $\alpha = \frac{r}{2}$ and $\beta = 2$. Hence, its MGF is $(1-2t)^{-r/2}$.

g. This was already done in the solutions to problems 5, 6, 7, 9 and 10.

2.15 *Moment Generating Function Method.*

a. If $X_1, .., X_n$ are independent Poisson distributed with parameters (λ_i) respectively, then from problem 2.14c, we have

$$M_{X_i}(t) = e^{\lambda_i(e^t-1)} \quad \text{for } i = 1, 2, \ldots, n$$

$Y = \sum_{i=1}^{n} X_i$ has $M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t)$ since the X_i's are independent. Hence

$$M_Y(t) = e^{\sum_{i=1}^{n} \lambda_i (e^t - 1)}$$

which we recognize as a Poisson with parameter $\sum_{i=1}^{n} \lambda_i$.

b. If $X_1, .., X_n$ are IIN (μ_i, σ_i^2), then from problem 2.14b, we have

$$M_{X_i(t)} = e^{\mu_i t + \frac{1}{2}\sigma_i^2 t^2} \quad \text{for } i = 1, 2, .., n$$

$Y = \sum_{i=1}^{n} X_i$ has $M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t)$ since the X_i's are independent. Hence

$$M_Y(t) = e^{\left(\sum_{i=1}^{n} \mu_i\right)t + \frac{1}{2}\left(\sum_{i=1}^{n} \sigma_i^2\right)t^2}$$

which we recognize as Normal with mean $\sum_{i=1}^{n} \mu_i$ and variance $\sum_{i=1}^{n} \sigma_i^2$.

c. If $X_1, .., X_n$ are IIN(μ, σ^2), then $Y = \sum_{i=1}^{n} X_i$ is $N(n\mu, n\sigma^2)$ from part b using the equality of means and variances. Therefore, $\bar{X} = Y/n$ is $N(\mu, \sigma^2/n)$.

d. If $X_1, .., X_n$ are independent χ^2 distributed with parameters (r_i) respectively, then from problem 2.14f, we get

$$M_{X_i}(t) = (1 - 2t)^{-r_i/2} \quad \text{for } i = 1, 2, .., n$$

$Y = \sum_{i=1}^{n} X_i$ has $M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t)$ since the X_i's are independent. Hence,

$$M_Y(t) = (1 - 2t)^{-\sum_{i=1}^{n} r_i/2}$$

which we recognize as χ^2 with degrees of freedom $\sum_{i=1}^{n} r_i$.

2.16 *Best Linear Prediction*. This is based on Amemiya (1994).

a. The mean squared error predictor is given by $MSE = E(Y - \alpha - \beta X)^2 = E(Y^2) + \alpha^2 + \beta^2 E(X^2) - 2\alpha E(Y) - 2\beta E(XY) + 2\alpha\beta E(X)$ minimizing this

MSE with respect to α and β yields the following first-order conditions:

$$\frac{\partial \text{MSE}}{\partial \alpha} = 2\alpha - 2E(Y) + 2\beta E(X) = 0$$

$$\frac{\partial \text{MSE}}{\partial \beta} = 2\beta E(X^2) - 2E(XY) + 2\alpha E(X) = 0.$$

Solving these two equations for α and β yields $\hat{\alpha} = \mu_Y - \hat{\beta}\mu_X$ from the first equation, where $\mu_Y = E(Y)$ and $\mu_X = E(X)$. Substituting this in the second equation one gets

$$\hat{\beta}E(X^2) - E(XY) + \mu_Y\mu_X - \hat{\beta}\mu_X^2 = 0$$

$$\hat{\beta}\text{var}(X) = E(XY) - \mu_X\mu_Y = \text{cov}(X, Y).$$

Hence, $\hat{\beta} = \text{cov}(X, Y)/\text{var}(X) = \sigma_{XY}/\sigma_X^2 = \rho\sigma_Y/\sigma_X$, since $\rho = \sigma_{XY}/\sigma_X\sigma_Y$. The *best* predictor is given by $\hat{Y} = \hat{\alpha} + \hat{\beta}X$.

b. Substituting $\hat{\alpha}$ into the *best* predictor one gets

$$\hat{Y} = \hat{\alpha} + \hat{\beta}X = \mu_Y + \hat{\beta}(X - \mu_X) = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(X - \mu_X)$$

one clearly deduces that $E(\hat{Y}) = \mu_Y$ and $\text{var}(\hat{Y}) = \rho^2\frac{\sigma_Y^2}{\sigma_X^2}\text{var}(X) = \rho^2\sigma_Y^2$. The prediction error $\hat{u} = Y - \hat{Y} = (Y - \mu_Y) - \rho\frac{\sigma_Y}{\sigma_X}(X - \mu_X)$ with $E(\hat{u}) = 0$ and $\text{var}(\hat{u}) = E(\hat{u}^2) = \text{var}(Y) + \rho^2\frac{\sigma_Y^2}{\sigma_X^2}\text{var}(X) - 2\rho\frac{\sigma_Y}{\sigma_X}\sigma_{XY} = \sigma_Y^2 + \rho^2\sigma_Y^2 - 2\rho^2\sigma_Y^2 = \sigma_Y^2(1 - \rho^2)$.

This is the proportion of the var(Y) that is not explained by the *best linear predictor* $\hat{Y}$.

c. $\text{cov}(\hat{Y}, \hat{u}) = \text{cov}(\hat{Y}, Y - \hat{Y}) = \text{cov}(\hat{Y}, Y) - \text{var}(\hat{Y})$

But

$$\text{cov}(\hat{Y}, Y) = E\left(\hat{Y} - \mu_Y\right)(Y - \mu_Y) = E\left(\rho\frac{\sigma_Y}{\sigma_X}(X - \mu_X)(Y - \mu_Y)\right)$$

$$= \rho\frac{\sigma_Y}{\sigma_X}\text{cov}(X, Y) = \rho^2\frac{\sigma_X\sigma_Y^2}{\sigma_X} = \rho^2\sigma_Y^2$$

Hence, $\text{cov}(\hat{Y}, \hat{u}) = \rho^2\sigma_Y^2 - \rho^2\sigma_Y^2 = 0$.

2.17 The *Best Predictor*.

a. The problem is to minimize $E[Y - h(X)]^2$ with respect to h(X). Add and subtract E(Y/X) to get

$$E\{[Y - E(Y/X)] + [E(Y/X) - h(X)]\}^2$$
$$= E[Y - E(Y/X)]^2 + E[E(Y/X) - h(X)]^2$$

and the cross-product term $E[Y - E(Y/X)][E(Y/X) - h(X)]$ is zero because of the law of iterated expectations, see the Appendix to Chapter 2 Amemiya (1994). In fact, this says that expectations can be written as

$$E = E_X E_{Y/X}$$

and the cross-product term given above $E_{Y/X}[Y - E(Y/X)][E(Y/X) - h(X)]$ is clearly zero. Hence, $E[Y - h(X)]^2$ is expressed as the sum of two positive terms. The first term is not affected by our choice of h(X). The second term however is zero for $h(X) = E(Y/X)$. Clearly, this is the *best predictor* of Y based on X.

b. In the Appendix to Chapter 2, we considered the bivariate Normal distribution and showed that $E(Y/X) = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(X - \mu_X)$. In part (a), we showed that this is the *best predictor* of Y based on X. But, in this case, this is exactly the form for the *best linear predictor* of Y based on X derived in problem 2.16. Hence, for the bivariate Normal density, the *best predictor* is identical to the *best linear predictor* of Y based on X.

References

Amemiya, T. (1994), Introduction to Statistics and Econometrics (Harvard University Press: Cambridge).

Baltagi, B.H. (1994), "The Wald, LR, and LM Inequality", *Econometric Theory*, Problem 94.1.2, 10: 223–224.

Freund, J.E. (1992), Mathematical Statistics (Prentice-Hall: New Jersey).

CHAPTER 3

Simple Linear Regression

3.1 For least squares, the first-order conditions of minimization, given by Eqs. (3.2) and (3.3), yield immediately the first two numerical properties of OLS estimates, i.e., $\sum_{i=1}^{n} e_i = 0$ and $\sum_{i=1}^{n} e_iX_i = 0$. Now consider $\sum_{i=1}^{n} e_i\hat{Y}_i = \hat{\alpha}\sum_{i=1}^{n} e_i + \hat{\beta}\sum_{i=1}^{n} e_iX_i = 0$ where the first equality uses $\hat{Y}_i = \hat{\alpha} + \hat{\beta}X_i$ and the second equality uses the first two numerical properties of OLS. Using the fact that $e_i = Y_i - \hat{Y}_i$, we can sum both sides to get $\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} \hat{Y}_i$, but $\sum_{i=1}^{n} e_i = 0$, therefore we get $\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \hat{Y}_i$. Dividing both sides by n, we get $\overline{Y} = \overline{\hat{Y}}$.

3.2 Minimizing $\sum_{i=1}^{n}(Y_i - \alpha)^2$ with respect to α yields $-2\sum_{i=1}^{n}(Y_i - \alpha) = 0$. Solving for α yields $\hat{\alpha}_{ols} = \overline{Y}$. Averaging $Y_i = \alpha + u_i$ we get $\overline{Y} = \alpha + \bar{u}$. Hence $\hat{\alpha}_{ols} = \alpha + \bar{u}$ with $E(\hat{\alpha}_{ols}) = \alpha$ since $E(\bar{u}) = \sum_{i=1}^{n} E(u_i)/n = 0$ and $\text{var}(\hat{\alpha}_{ols}) = E(\hat{\alpha}_{ols} - \alpha)^2 = E(\bar{u})^2 = \text{var}(\bar{u}) = \sigma^2/n$. The residual sum of squares is $\sum_{i=1}^{n}(Y_i - \hat{\alpha}_{ols})^2 = \sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2 = \sum_{i=1}^{n} y_i{}^2$.

3.3 **a.** Minimizing $\sum_{i=1}^{n}(Y_i - \beta X_i)^2$ with respect to β yields $-2\sum_{i=1}^{n}(Y_i - \beta X_i)X_i = 0$. Solving for β yields $\hat{\beta}_{ols} = \sum_{i=1}^{n} Y_iX_i\Big/\sum_{i=1}^{n} X_i{}^2$. Substituting $Y_i = \beta X_i + u_i$ yields $\hat{\beta}_{ols} = \beta + \sum_{i=1}^{n} X_iu_i\Big/\sum_{i=1}^{n} X_i{}^2$ with $E\left(\hat{\beta}_{ols}\right) = \beta$ since X_i is nonstochastic and $E(u_i) = 0$. Also, $\text{var}\left(\hat{\beta}_{ols}\right) = E\left(\hat{\beta}_{ols} - \beta\right)^2 = E\left(\sum_{i=1}^{n} X_iu_i\Big/\sum_{i=1}^{n} X_i{}^2\right)^2 = \sigma^2\sum_{i=1}^{n} X_i{}^2\Big/\left(\sum_{i=1}^{n} X_i{}^2\right)^2 = \sigma^2\Big/\sum_{i=1}^{n} X_i{}^2$.

b. From the first-order condition in part (a), we get $\sum_{i=1}^{n} e_iX_i = 0$, where $e_i = Y_i - \hat{\beta}_{ols}X_i$. However, $\sum_{i=1}^{n} e_i$ is not necessarily zero. Therefore $\sum_{i=1}^{n} Y_i$ is

B.H. Baltagi, *Solutions Manual for Econometrics*, Springer Texts in Business and Economics, DOI 10.1007/978-3-642-54548-1_3,

not necessarily equal to $\sum_{i=1}^{n} \hat{Y}_i$, see problem 3.1. $\sum_{i=1}^{n} e_i\hat{Y}_i = \hat{\beta}_{ols} \sum_{i=1}^{n} e_iX_i = 0$, using $\hat{Y}_i = \hat{\beta}_{ols}X_i$ and $\sum_{i=1}^{n} e_iX_i = 0$. Therefore, only two of the numerical properties considered in problem 3.1 hold for this regression without a constant.

3.4 $E\left(\sum_{i=1}^{n} x_iu_i\right)^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} x_ix_jE\left(u_iu_j\right) = \sum_{i=1}^{n} x_i{}^2 var(u_i) + \sum\sum_{i\neq j} x_ix_j cov(u_i, u_j) = \sigma^2 \sum_{i=1}^{n} x_i{}^2$ since $cov(u_i, u_j) = 0$ for $i \neq j$ and $var(u_i) = \sigma^2$ using assumptions 2 and 3.

3.5 **a.** From Eq. (3.4), we know that $\hat{\alpha}_{ols} = \overline{Y} - \hat{\beta}_{ols}\overline{X}$, substituting $\overline{Y} = \alpha + \beta\overline{X} + \overline{u}$ we get $\hat{\alpha}_{ols} = \alpha + \left(\beta - \hat{\beta}_{ols}\right)\overline{X} + \overline{u}$. Therefore, $E\left(\hat{\alpha}_{ols}\right) = \alpha$ since $E\left(\hat{\beta}_{ols}\right) = \beta$, see Eq. (3.5), and $E(\overline{u}) = 0$.

b. $var\left(\hat{\alpha}_{ols}\right) = E(\hat{\alpha}_{ols} - \alpha)^2 = E\left[\left(\beta - \hat{\beta}_{ols}\right)\overline{X} + \overline{u}\right]^2$.

$= var\left(\hat{\beta}_{ols}\right)\overline{X}^2 + var(\overline{u}) + 2\overline{X}cov\left(\hat{\beta}_{ols}, \overline{u}\right)$

But from (3.6), $var\left(\hat{\beta}_{ols}\right) = \sigma^2 / \sum_{i=1}^{n} x_i{}^2$. Also, $var(\overline{u}) = \sigma^2/n$ and

$$cov\left(\hat{\beta}_{ols}, \overline{u}\right) = E\left(\frac{\sum_{i=1}^{n} x_iu_i}{\sum_{i=1}^{n} x_i{}^2} \cdot \frac{\sum_{i=1}^{n} u_i}{n}\right) = \sigma^2 \sum_{i=1}^{n} x_i / n \sum_{i=1}^{n} x_i{}^2 = 0$$

where the first equality follows from Eq. (3.5). The second equality uses the fact that $cov(u_i, u_j) = 0$ for $i \neq j$ and $var(u_i) = \sigma^2$ from assumptions 2 and 3. The last equality follows from the fact that $\sum_{i=1}^{n} x_i = 0$. Therefore,

$$var\left(\hat{\alpha}_{ols}\right) = \sigma^2 \left[(1/n) + \left(\overline{X}^2 / \sum_{i=1}^{n} x_i{}^2\right)\right] = \sigma^2 \sum_{i=1}^{n} X_i{}^2 / \left(n \sum_{i=1}^{n} x_i{}^2\right).$$

c. $\hat{\alpha}_{ols}$ is unbiased for α and $var\left(\hat{\alpha}_{ols}\right) \rightarrow 0$ as $n \rightarrow \infty$, since $\sum_{i=1}^{n} x_i{}^2/n$ has a finite limit as $n \rightarrow \infty$ by assumption 4. Therefore, $\hat{\alpha}_{ols}$ is consistent for α.

d. Using (3.5) and part (a), we get

$$\operatorname{cov}\left(\hat{\alpha}_{ols}, \hat{\beta}_{ols}\right) = E\left(\hat{\alpha}_{ols} - \alpha\right)\left(\hat{\beta}_{ols} - \beta\right) = -\overline{X}\operatorname{var}\left(\hat{\beta}_{ols}\right) + \operatorname{cov}\left(\overline{u}, \hat{\beta}_{ols}\right)$$

$$= -\sigma^2\overline{X}\Big/\sum_{i=1}^{n} x_i^2$$

The last equality uses the fact that $\operatorname{cov}(\overline{u}, \hat{\beta}_{ols}) = 0$ from part (b).

3.6 a. $\hat{\alpha}_{ols} = \overline{Y} - \hat{\beta}_{ols}\overline{X} = \sum_{i=1}^{n} Y_i/n - \overline{X}\sum_{i=1}^{n} w_iY_i = \sum_{i=1}^{n} \lambda_iY_i$ where $\lambda_i = (1/n) - \overline{X}w_i$, $\hat{\beta}_{ols} = \sum_{i=1}^{n} w_iY_i$ and w_i is defined above (3.7).

b. $\sum_{i=1}^{n} \lambda_i = 1 - \overline{X}\sum_{i=1}^{n} w_i = 1$ where $\sum_{i=1}^{n} w_i = 0$ from (3.7), and

$$\sum_{i=1}^{n} \lambda_iX_i = \sum_{i=1}^{n} X_i/n - \overline{X}\sum_{i=1}^{n} X_iw_i = \overline{X} - \overline{X} = 0$$

using $\sum_{i=1}^{n} w_iX_i = 1$ from (3.7).

c. Any other linear estimator of α, can be written as

$$\tilde{\alpha} = \sum_{i=1}^{n} b_iY_i = \alpha\sum_{i=1}^{n} b_i + \beta\sum_{i=1}^{n} b_iX_i + \sum_{i=1}^{n} b_iu_i$$

where the last equality is obtained by substituting Eq. (3.1) for Y_i. For $E(\tilde{\alpha})$ to equal α, we must have $\sum_{i=1}^{n} b_i = 1$, $\sum_{i=1}^{n} b_iX_i = 0$. Hence, $\tilde{\alpha} = \alpha + \sum_{i=1}^{n} b_iu_i$.

d. Let $b_i = \lambda_i + f_i$, then $1 = \sum_{i=1}^{n} b_i = \sum_{i=1}^{n} \lambda_i + \sum_{i=1}^{n} f_i$. From part (b), $\sum_{i=1}^{n} \lambda_i = 1$, therefore $\sum_{i=1}^{n} f_i = 0$. Similarly, $0 = \sum_{i=1}^{n} b_iX_i = \sum_{i=1}^{n} \lambda_iX_i + \sum_{i=1}^{n} f_iX_i$. From part (b), $\sum_{i=1}^{n} \lambda_iX_i = 0$, therefore $\sum_{i=1}^{n} f_iX_i = 0$.

e. $\operatorname{var}(\tilde{\alpha}) = E(\tilde{\alpha} - \alpha)^2 = E\left(\sum_{i=1}^{n} b_iu_i\right)^2 = \sigma^2\sum_{i=1}^{n} b_i^2$ using the expression of $\tilde{\alpha}$ from part (c) and assumptions 2 and 3 with $\operatorname{var}(u_i) = \sigma^2$ and $\operatorname{cov}(u_i, u_j) = 0$ for $i \neq j$. Substituting, $b_i = \lambda_i + f_i$ from part (d), we get

$$\operatorname{var}(\tilde{\alpha}) = \sigma^2\left(\sum_{i=1}^{n} \lambda_i^2 + \sum_{i=1}^{n} f_i^2 + 2\sum_{i=1}^{n} \lambda_if_i\right)$$

For $\lambda_i = (1/n) - \overline{X}w_i$ from part (a), we get $\sum_{i=1}^{n} \lambda_i f_i = \sum_{i=1}^{n} f_i/n - \overline{X}\sum_{i=1}^{n} w_i f_i = 0$.

This uses $\sum_{i=1}^{n} f_i = 0$ and $\sum_{i=1}^{n} f_i X_i = 0$ from part (d). Recall, $w_i = x_i/\sum_{i=1}^{n} x_i{}^2$ and so $\sum_{i=1}^{n} w_i f_i = 0$. Hence, $\text{var}(\tilde{\alpha}) = \sigma^2\left(\sum_{i=1}^{n} \lambda_i^2 + \sum_{i=1}^{n} f_i^2\right) = \text{var}(\hat{\alpha}_{ols}) + \sigma^2\sum_{i=1}^{n} f_i^2$. Any linear unbiased estimator $\tilde{\alpha}$ of α, will have a larger variance than $\hat{\alpha}_{ols}$ as long as $\sum_{i=1}^{n} f_i^2 \neq 0$. $\sum_{i=1}^{n} f_i^2 = 0$ when $f_i = 0$ for all i, which means that $b_i = \lambda_i$ for all i and $\tilde{\alpha} = \sum_{i=1}^{n} b_i Y_i = \sum_{i=1}^{n} \lambda_i Y_i = \hat{\alpha}_{ols}$. This proves that $\hat{\alpha}_{ols}$ is BLUE for α.

3.7 a. $\dfrac{\partial \text{logL}}{\partial\alpha} = 0$ from (3.9) leads to $-\dfrac{1}{2\sigma^2}\dfrac{\partial \text{RSS}}{\partial\alpha} = 0$ where $\text{RSS} = \sum_{i=1}^{n}(Y_i - \alpha - \beta X_i)^2$.

Solving this first-order condition yields $\hat{\alpha}_{ols}$. Therefore, $\hat{\alpha}_{mle} = \hat{\alpha}_{ols}$. Similarly, $\dfrac{\partial \text{logL}}{\partial\beta} = 0$ from (3.9) leads to $-\dfrac{1}{2\sigma^2}\dfrac{\partial \text{RSS}}{\partial\beta} = 0$. Solving for β yields $\hat{\beta}_{ols}$. Hence, $\hat{\beta}_{mle} = \hat{\beta}_{ols}$.

b. $\dfrac{\partial \log L}{\partial\sigma^2} = -\dfrac{n}{2}\cdot\dfrac{1}{\sigma^2} + \dfrac{\sum_{i=1}^{n}(Y_i - \alpha - \beta X_i)^2}{2\sigma^4}$

Setting $\dfrac{\partial \text{logL}}{\partial\sigma^2} = 0$, yields $\hat{\sigma}^2_{mle} = \sum_{i=1}^{n}\left(Y_i - \hat{\alpha} - \hat{\beta}X_i\right)^2/n = \sum_{i=1}^{n} e_i{}^2/n$ where $\hat{\alpha}$ and $\hat{\beta}$ are the MLE estimates, which in this case are identical to the OLS estimates.

3.9 a. $$R^2 = \frac{\sum_{i=1}^{n}\hat{y}_i^2}{\sum_{i=1}^{n} y_i{}^2} = \frac{\hat{\beta}^2_{ols}\sum_{i=1}^{n} x_i{}^2}{\sum_{i=1}^{n} y_i{}^2} = \frac{\left(\sum_{i=1}^{n} x_i y_i\right)^2\left(\sum_{i=1}^{n} x_i{}^2\right)}{\left(\sum_{i=1}^{n} x_i{}^2\right)^2\left(\sum_{i=1}^{n} y_i{}^2\right)} = \frac{\left(\sum_{i=1}^{n} x_i y_i\right)^2}{\sum_{i=1}^{n} x_i{}^2\sum_{i=1}^{n} y_i{}^2} = r^2_{xy}$$
where the second equality substitutes $\hat{y}_i = \hat{\beta}_{ols}x_i$ and the third equality substitutes $\hat{\beta}_{ols} = \sum_{i=1}^{n} x_i y_i/\sum_{i=1}^{n} x_i{}^2$.

b. Multiply both sides of $y_i = \hat{y}_i + e_i$ by $\hat{y}_i$ and sum, one gets $\sum_{i=1}^{n} y_i\hat{y}_i = \sum_{i=1}^{n}\hat{y}_i^2 + \sum_{i=1}^{n}\hat{y}_i e_i$. The last term is zero by the numerical properties of least squares. Hence, $\sum_{i=1}^{n} y_i\hat{y}_i = \sum_{i=1}^{n} \hat{y}_i^2$. Therefore

$$r_{y\hat{y}}^2 = \frac{\left(\sum_{i=1}^{n} y_i\hat{y}_i\right)^2}{\sum_{i=1}^{n} {y_i}^2 \sum_{i=1}^{n} \hat{y}_i^2} = \frac{\left(\sum_{i=1}^{n} \hat{y}_i^2\right)^2}{\sum_{i=1}^{n} {y_i}^2 \sum_{i=1}^{n} \hat{y}_i^2} = \frac{\sum_{i=1}^{n} \hat{y}_i^2}{\sum_{i=1}^{n} {y_i}^2} = R^2.$$

3.11 *Optimal Weighting of Unbiased Estimators*. This is based on Baltagi (1995). All three estimators of β can be put in the linear form $\hat{\beta} = \sum_{i=1}^{n} w_i Y_i$ where

$$\begin{aligned} w_i &= X_i / \sum_{i=1}^{n} X_i^2 &&\text{for } \hat{\beta}_1, \\ &= \left(1 \Big/ \sum_{i=1}^{n} X_i\right) &&\text{for } \hat{\beta}_2, \\ &= (X_i - \bar{X}) \Big/ \sum_{i=1}^{n} (X_i - \bar{X})^2 &&\text{for } \hat{\beta}_3. \end{aligned}$$

All satisfy $\sum_{i=1}^{n} w_i X_i = 1$, which is necessary for $\hat{\beta}_i$ to be unbiased for $i = 1, 2, 3$. Therefore

$$\hat{\beta}_i = \beta + \sum_{i=1}^{n} w_i u_i \qquad \text{for } i = 1, 2, 3$$

with $E(\hat{\beta}_i) = \beta$ and $\text{var}(\hat{\beta}_i) = \sigma^2 \sum_{i=1}^{n} {w_i}^2$ since the u_i's are $\text{IID}(0, \sigma^2)$. Hence,

$$\text{var}(\hat{\beta}_1) = \sigma^2 / \sum_{i=1}^{n} {X_i}^2, \text{var}(\hat{\beta}_2) = \sigma^2/n\bar{X}^2 \text{ and } \text{var}(\hat{\beta}_3) = \sigma^2 / \sum_{i=1}^{n} (X_i - \bar{X})^2.$$

a. $$\begin{aligned} \text{cov}(\hat{\beta}_1, \hat{\beta}_2) &= E\left(\sum_{i=1}^{n} u_i \Big/ \sum_{i=1}^{n} X_i\right)\left(\sum_{i=1}^{n} X_i u_i \Big/ \sum_{i=1}^{n} {X_i}^2\right) \\ &= \sigma^2 \sum_{i=1}^{n}\left(X_i \Big/ \sum_{i=1}^{n} {X_i}^2\right)\left(1 \Big/ \sum_{i=1}^{n} X_i\right) = \sigma^2 / \sum_{i=1}^{n} {X_i}^2 = \text{var}(\hat{\beta}_1) > 0. \end{aligned}$$

Also,

$$\begin{aligned} \rho_{12} &= \text{cov}(\hat{\beta}_1, \hat{\beta}_2) / \left[\text{var}(\hat{\beta}_1)\text{var}(\hat{\beta}_2)\right]^{1/2} \\ &= \left[\text{var}(\hat{\beta}_1)/\text{var}(\hat{\beta}_2)\right]^{1/2} = \left(n\bar{X}^2 \Big/ \sum_{i=1}^{n} {X_i}^2\right)^{1/2} \end{aligned}$$

with $0 < \rho_{12} \leq 1$. Samuel-Cahn (1994) showed that whenever the correlation coefficient between two unbiased estimators is equal to the square root of the ratio of their variances, the optimal combination of these two unbiased estimators is the one that weights the estimator with the smaller variance by 1. In this example, $\hat{\beta}_1$ is the best linear unbiased estimator (BLUE). Therefore, as expected, when we combine $\hat{\beta}_1$ with any other unbiased estimator of β, the optimal weight α^* turns out to be 1 for $\hat{\beta}_1$ and zero for the other linear unbiased estimator.

b. Similarly,

$$\begin{aligned}\mathrm{cov}(\hat{\beta}_1, \hat{\beta}_3) &= \mathrm{E}\left(\sum_{i=1}^{n} X_i u_i \Big/ \sum_{i=1}^{n} X_i^2\right)\left[\sum_{i=1}^{n}(X_i - \overline{X})u_i \Big/ \sum_{i=1}^{n}(X_i - \overline{X})^2\right] \\ &= \sigma^2 \sum_{i=1}^{n}\left(X_i \Big/ \sum_{i=1}^{n} X_i^2\right)\left[(X_i - \overline{X}) \Big/ \sum_{i=1}^{n}(X_i - \overline{X})^2\right] \\ &= \sigma^2 \Big/ \sum_{i=1}^{n} X_i^2 = \mathrm{var}(\hat{\beta}_1) > 0. \text{ Also}\end{aligned}$$

$$\begin{aligned}\rho_{13} &= \mathrm{cov}(\hat{\beta}_1, \hat{\beta}_3) \Big/ \left[\mathrm{var}(\hat{\beta}_1)\mathrm{var}(\hat{\beta}_3)\right]^{1/2} = \left[\mathrm{var}(\hat{\beta}_1)/\mathrm{var}(\hat{\beta}_3)\right]^{1/2} \\ &= \left[\sum_{i=1}^{n}(X_i - \overline{X})^2 \Big/ \sum_{i=1}^{n} X_i^2\right]^{1/2} = \left(1 - \rho_{12}^2\right)^{1/2}\end{aligned}$$

with $0 < \rho_{13} \leq 1$. For the same reasons given in part (a), the optimal combination of $\hat{\beta}_1$ and $\hat{\beta}_3$ is that for which $\alpha^* = 1$ for $\hat{\beta}_1$ and $1 - \alpha^* = 0$ for $\hat{\beta}_3$.

c. Finally, $\mathrm{cov}(\hat{\beta}_2, \hat{\beta}_3) = \mathrm{E}\left(\sum\limits_{i=1}^{n} u_i \Big/ \sum\limits_{i=1}^{n} X_i\right)\left[\sum\limits_{i=1}^{n}(X_i - \overline{X})\, u_i \Big/ \sum\limits_{i=1}^{n}(X_i - \overline{X})^2\right]$

$$= \sigma^2 \sum_{i=1}^{n}(X_i - \overline{X}) \Big/ \left(\sum_{i=1}^{n} X_i\right)\left[\sum_{i=1}^{n}(X_i - \overline{X})^2\right] = 0.$$

Hence $\rho_{23} = 0$. In this case, it is a standard result, see Samuel-Cahn (1994), that the optimal combination of $\hat{\beta}_2$ and $\hat{\beta}_3$ weights these unbiased estimators in proportion to $\mathrm{var}(\hat{\beta}_3)$ and $\mathrm{var}(\hat{\beta}_2)$, respectively. In other words

$$\begin{aligned}\hat{\beta} = &\left\{\mathrm{var}(\hat{\beta}_2) \Big/ \left[\mathrm{var}(\hat{\beta}_2) + \mathrm{var}(\hat{\beta}_3)\right]\right\}\hat{\beta}_3 \\ &+ \left\{\mathrm{var}(\hat{\beta}_3) \Big/ \left[\mathrm{var}(\hat{\beta}_2) + \mathrm{var}(\hat{\beta}_3)\right]\right\}\hat{\beta}_2\end{aligned}$$

$$= \left(1-\rho_{12}^2\right)\hat{\beta}_3 + \rho_{12}^2\hat{\beta}_2$$

since $\mathrm{var}(\hat{\beta}_3)/\mathrm{var}(\hat{\beta}_2) = \rho_{12}^2/\left(1-\rho_{12}^2\right)$ and

$\rho_{12}^2 = n\overline{X}^2/\sum_{i=1}^{n} X_i^2$ while $1-\rho_{12}^2 = \sum_{i=1}^{n}(X_i-\overline{X})^2/\sum_{i=1}^{n} X_i^2$. Hence,

$$\hat{\beta} = \left[\sum_{i=1}^{n}(X_i-\overline{X})^2/\sum_{i=1}^{n}X_i^2\right]\hat{\beta}_3 + \left[n\overline{X}^2/\sum_{i=1}^{n}X_i^2\right]\hat{\beta}_2$$

$$= \left(1/\sum_{i=1}^{n}X_i^2\right)\left[\sum_{i=1}^{n}(X_i-\overline{X})\,Y_i + n\overline{X}\overline{Y}\right] = \sum_{i=1}^{n}X_iY_i/\sum_{i=1}^{n}X_i^2 = \hat{\beta}_1.$$

See also the solution by Trenkler (1996).

3.12 *Efficiency as Correlation*. This is based on Zheng (1994). Since $\hat{\beta}$ and $\tilde{\beta}$ are linear unbiased estimators of β, it follows that $\hat{\beta}+\lambda(\hat{\beta}-\tilde{\beta})$ for any λ is a linear unbiased estimator of β. Since $\hat{\beta}$ is the BLU estimator of β,

$$\mathrm{var}\left[\hat{\beta}+\lambda(\hat{\beta}-\tilde{\beta})\right] = \mathrm{E}\left[\hat{\beta}+\lambda(\hat{\beta}-\tilde{\beta})\right]^2 - \beta^2$$

is minimized at $\lambda = 0$. Setting the derivative of $\mathrm{var}\left[\hat{\beta}+\lambda(\hat{\beta}-\tilde{\beta})\right]$ with respect to λ at $\lambda = 0$, we have $2\mathrm{E}\left[\hat{\beta}(\hat{\beta}-\tilde{\beta})\right] = 0$, or $\mathrm{E}(\hat{\beta}^2) = \mathrm{E}(\hat{\beta}\tilde{\beta})$. Thus, the squared correlation between $\hat{\beta}$ and $\tilde{\beta}$ is

$$\frac{\left[\mathrm{cov}(\hat{\beta},\tilde{\beta})\right]^2}{\mathrm{var}(\hat{\beta})\mathrm{var}(\tilde{\beta})} = \frac{\left[\mathrm{E}\left[(\hat{\beta}-\beta)\,(\tilde{\beta}-\beta)\right]\right]^2}{\mathrm{var}(\hat{\beta})\mathrm{var}(\tilde{\beta})} = \frac{\left[\mathrm{E}(\hat{\beta}\tilde{\beta})-\beta^2\right]^2}{\mathrm{var}(\hat{\beta})\mathrm{var}(\tilde{\beta})}$$

$$= \frac{\left[\mathrm{E}(\hat{\beta}^2)-\beta^2\right]^2}{\mathrm{var}(\hat{\beta})\mathrm{var}\left(\tilde{\beta}\right)} = \frac{\left[\mathrm{var}\left(\hat{\beta}\right)\right]^2}{\mathrm{var}(\hat{\beta})\mathrm{var}(\tilde{\beta})} = \frac{\mathrm{var}(\hat{\beta})}{\mathrm{var}(\tilde{\beta})},$$

where the third equality uses the result that $\mathrm{E}(\hat{\beta}^2) = \mathrm{E}(\hat{\beta}\tilde{\beta})$. The final equality gives $\mathrm{var}(\hat{\beta})/\mathrm{var}(\tilde{\beta})$ which is the relative efficiency of $\hat{\beta}$ and $\tilde{\beta}$.

3.13 **a.** Adding 5 to each observation of X_i, adds 5 to the sample average $\overline{X}$ and it is now 12.5. This means that $x_i = X_i - \overline{X}$ is unaffected. Hence $\sum_{i=1}^{n} x_i^2$ is the same and since Y_i, is unchanged, we conclude that $\hat{\beta}_{ols}$ is still the same at

0.8095. However, $\hat{\alpha}_{ols} = \bar{Y} - \hat{\beta}_{ols}\bar{X}$ is changed because $\bar{X}$ is changed. This is now $\hat{\alpha}_{ols} = 6.5 - (0.8095)(12.5) = -3.6188$. It has decreased by $5\hat{\beta}_{ols}$ since $\bar{X}$ increased by 5 while $\hat{\beta}_{ols}$ and $\bar{Y}$ remained unchanged. It is easy to see that $\hat{Y}_i = \hat{\alpha}_{ols} + \hat{\beta}_{ols}X_i$ remains the same. When X_i increased by 5, with $\hat{\beta}_{ols}$ the same, this increases $\hat{Y}_i$ by $5\hat{\beta}_{ols}$. But $\hat{\alpha}_{ols}$ decreases $\hat{Y}_i$ by $-5\hat{\beta}_{ols}$. The net effect on $\hat{Y}_i$ is zero. Since Y_i is unchanged, this means $e_i = Y_i - \hat{Y}_i$ is unchanged. Hence $s^2 = \sum_{i=1}^{n} e_i^2/n-2$ is unchanged. Since x_i is unchanged $s^2/\sum_{i=1}^{n} x_i^2$ is unchanged and $se(\hat{\beta}_{ols})$ and the t-statistic for $H_o^a;\ \beta = 0$ are unchanged. The

$$\text{var}(\hat{\alpha}_{ols}) = s^2\left[\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n} x_i^2}\right] = 0.311905\left[\frac{1}{10} + \frac{(12.5)^2}{52.5}\right] = 0.95697$$

with its square root given by $\hat{se}(\hat{\alpha}_{ols}) = 0.97825$. Hence the t-statistic for $H_o^b;\ \alpha = 0$ is $-3.6189/0.97825 = -3.699$. This is significantly different from zero. $R^2 = 1-\left(\sum_{i=1}^{n} e_i^2/\sum_{i=1}^{n} y_i^2\right)$ is unchanged. Hence, only $\hat{\alpha}_{ols}$ and its standard error and t-statistic are affected by an additive constant of 5 on X_i.

b. Adding 2 to each observation of Y_i, adds 2 to the sample average $\bar{Y}$ and it is now 8.5. This means that $y_i = Y_i - \bar{Y}$ is unaffected. Hence $\hat{\beta}_{ols}$ is still the same at 0.8095. However, $\hat{\alpha}_{ols} = \bar{Y} - \hat{\beta}_{ols}\bar{X}$ is changed because $\bar{Y}$ is changed. This is now $\hat{\alpha}_{ols} = 8.5 - (0.8095)(7.5) = 2.4286$ two more than the old $\hat{\alpha}_{ols}$ given in the numerical example. It is easy to see that $\hat{Y}_i = \hat{\alpha}_{ols} + \hat{\beta}_{ols}X_i$ has increased by 2 for each $i = 1, 2, .., 10$ since $\hat{\alpha}_{ols}$ increased by 2 while $\hat{\beta}_{ols}$ and X_i are the same. Both Y_i and $\hat{Y}_i$ increased by 2 for each observation. Hence, $e_i = Y_i - \hat{Y}_i$ is unchanged. Also, $s^2 = \sum_{i=1}^{n} e_i^2/n - 2$ and $s^2/\sum_{i=1}^{n} x_i^2$ are unchanged. This means that $se(\hat{\beta}_{ols})$ and the t-statistic for $H_o^a;\ \beta = 0$ are unchanged. The

$$\text{var}(\hat{\alpha}_{ols}) = s^2\left[\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n} x_i^2}\right]$$

is unchanged and therefore se $(\hat{\alpha}_{ols})$ is unchanged. However, the t-statistic for H_o^b; $\alpha = 0$ is now $2.4286/0.60446 = 4.018$ which is now statistically significant. $R^2 = 1 - \left(\sum_{i=1}^{n} e_i^2 / \sum_{i=1}^{n} y_i^2\right)$ is unchanged. Again, only $\hat{\alpha}_{ols}$ is affected by an additive constant of 2 on Y_i.

c. If each X_i is multiplied by 2, then the old $\overline{X}$ is multiplied by 2 and it is now 15. This means that each $x_i = X_i - \overline{X}$ is now double what it was in Table 3.1. Since y_i is the same, this means that $\sum_{i=1}^{n} x_i y_i$ is double what it was in Table 3.1 and $\sum_{i=1}^{n} x_i^2$ is four times what it was in Table 3.1. Hence,

$$\hat{\beta}_{ols} = \sum_{i=1}^{n} x_i y_i / \sum_{i=1}^{n} x_i^2 = 2\,(42.5)\,/4\,(52.5) = 0.8095/2 = 0.40475.$$

In this case, $\hat{\beta}_{ols}$ is half what it was in the numerical example. $\hat{\alpha}_{ols} = \overline{Y} - \hat{\beta}_{ols}\overline{X}$ is the same since $\hat{\beta}_{ols}$ is half what it used to be while $\overline{X}$ is double what it used to be. Also, $\hat{Y}_i = \hat{\alpha}_{ols} + \hat{\beta}_{ols} X_i$ is the same, since X_i is doubled while $\hat{\beta}_{ols}$ is half what it used to be and $\hat{\alpha}_{ols}$ is unchanged. Therefore, $e_i = Y_i - \hat{Y}_i$ is unchanged and $s^2 = \sum_{i=1}^{n} e_i^2/n - 2$ is also unchanged.

Now, $s^2/\sum_{i=1}^{n} x_i^2$ is one fourth what it used to be and $se(\hat{\beta}_{ols})$ is half what it used to be. Since, $\hat{\beta}_{ols}$ and $se(\hat{\beta}_{ols})$ have been both reduced by half, the t-statistic for H_o^a; $\beta = 0$ remains unchanged. The

$$\text{var}\,(\hat{\alpha}) = s^2 \left[\frac{1}{n} + \frac{\overline{X}^2}{\sum_{i=1}^{n} x_i^2} \right]$$

is unchanged since $\overline{X}^2$ and $\sum_{i=1}^{n} x_i^2$ are now both multiplied by 4. Hence, $se(\hat{\alpha}_{ols})$ and the t-statistic for H_o^b ; $\alpha = 0$ are unchanged. $R^2 = 1 - \left(\sum_{i=1}^{n} e_i^2 / \sum_{i=1}^{n} y_i^2\right)$ is also unchanged. Hence, only $\hat{\beta}$ is affected by a multiplicative constant of 2 on X_i.

3.14 a. Dependent Variable: LNC

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob > F
Model	1	0.04693	0.04693	1.288	0.2625
Error	44	1.60260	0.03642		
C Total	45	1.64953			

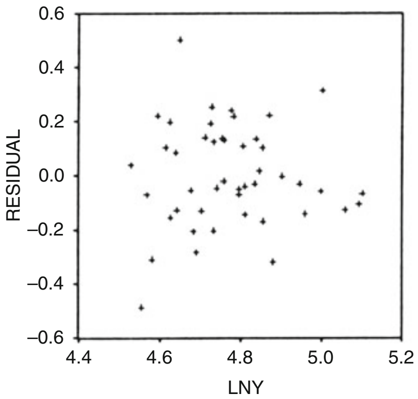
Root MSE	0.19085	R-square	0.0285
Dep Mean	4.84784	Adj R-sq	0.0064
C.V.	3.93675		

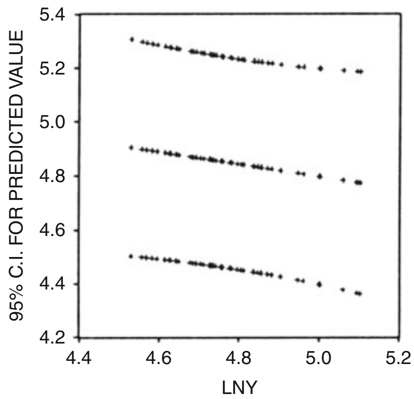
Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob > |T|
INTERCEP	1	5.931889	0.95542530	6.209	0.0001
LNY	1	-0.227003	0.19998321	-1.135	0.2625

The income elasticity is -0.227 which is negative! Its standard error is (0.1999) and the t-statistic for testing this income elasticity is zero is -1.135 which is insignificant with a p-value of 0.26. Hence, we cannot reject the null hypothesis. $R^2 = 0.0285$ and $s = 0.19085$. This regression is not very useful. The income variable is not significant and the R^2 indicates that the regression explains only 2.8% of the variation in consumption.

b. Plot of Residuals, and the 95% confidence interval for the predicted value.





SAS Program for 3.14

```
Data CIGARETT;
   Input OBS STATE $ LNC LNP LNY;
   Cards;
   Proc reg data=cigarett;
        model lnc=lny;
        *plot residual. *lny='*';
        *plot (U95. L95.)*lny='-' p.*lny /overlay
        symbol='*';
   output out=out1 r=resid p=pred u95=upper95 195=lower95;
   proc plot data=out1 vpercent=75 hpercent=100;
        plot resid*lny='*';
   proc plot data=out1 vpercent=95 hpercent=100;
        plot (Upper95 Lower95)*lny='-' Pred*lny='*'
   /overlay;
   run;
```

3.15 Theil's *minimum mean square* estimator of β. $\tilde{\beta}$ can be written as:

a. $\tilde{\beta} = \sum_{i=1}^{n} X_i Y_i \Big/ \left[\sum_{i=1}^{n} X_i^2 + \left(\sigma^2/\beta^2\right)\right]$

Substituting $Y_i = \beta X_i + u_i$ we get

$$\tilde{\beta} = \frac{\beta \sum_{i=1}^{n} X_i{}^2 + \sum_{i=1}^{n} X_i u_i}{\sum_{i=1}^{n} X_i{}^2 + (\sigma^2/\beta^2)}$$

and since $E(X_i u_i) = 0$, we get

$$E(\tilde{\beta}) = \frac{\beta \sum_{i=1}^{n} X_i{}^2}{\sum_{i=1}^{n} X_i{}^2 + (\sigma^2/\beta^2)} = \beta\left(\frac{1}{1+c}\right)$$

where $c = \sigma^2/\beta^2 \sum_{i=1}^{n} X_i{}^2 > 0$.

b. Therefore, $\text{Bias}(\tilde{\beta}) = E(\tilde{\beta}) - \beta = \beta\left(\frac{1}{1+c}\right) - \beta = -[c/(1+c)]\beta$. This bias is positive (negative) when β is negative (positive). This also means that $\tilde{\beta}$ is biased towards zero.

c. $\text{Bias}^2(\tilde{\beta}) = [c^2/(1+c)^2]\beta^2$ and

$$\text{var}(\tilde{\beta}) = E\left(\tilde{\beta} - E(\tilde{\beta})\right)^2 = E\left(\frac{\sum_{i=1}^{n} X_i u_i}{\sum_{i=1}^{n} X_i^2 + (\sigma^2/\beta^2)}\right)^2$$

$$= \frac{\sigma^2 \sum_{i=1}^{n} X_i^2}{\left[\sum_{i=1}^{n} X_i^2 + (\sigma^2/\beta^2)\right]^2}$$

using $\text{var}(u_i) = \sigma^2$ and $\text{cov}(u_i, u_j) = 0$ for $i \neq j$. This can also be written as $\text{var}(\tilde{\beta}) = \frac{\sigma^2}{\sum_{i=1}^{n} X_i^2(1+c)^2}$. Therefore,

$$\text{MSE}(\tilde{\beta}) = \text{Bias}^2(\tilde{\beta}) + \text{var}(\tilde{\beta})$$

$$\text{MSE}(\tilde{\beta}) = \frac{c^2}{(1+c)^2}\beta^2 + \frac{\sigma^2}{\sum_{i=1}^{n} X_i^2 (1+c)^2} = \frac{\beta^2 c^2 \sum_{i=1}^{n} X_i^2 + \sigma^2}{\sum_{i=1}^{n} X_i^2 (1+c)^2}$$

But $\beta^2 \sum_{i=1}^{n} X_i^2 c = \sigma^2$ from the definition of c. Hence,

$$\text{MSE}(\tilde{\beta}) = \frac{\sigma^2(1+c)}{\sum_{i=1}^{n} X_i^2 (1+c)^2} = \frac{\sigma^2}{\sum_{i=1}^{n} X_i^2 (1+c)} = \frac{\sigma^2}{\sum_{i=1}^{n} X_i^2 + (\sigma^2/\beta^2)}$$

The $\text{Bias}(\hat{\beta}_{\text{ols}}) = 0$ and $\text{var}(\hat{\beta}_{\text{ols}}) = \sigma^2/\sum_{i=1}^{n} X_i^2$. Hence

$$\text{MSE}(\hat{\beta}_{\text{ols}}) = \text{var}\left(\hat{\beta}_{\text{ols}}\right) = \sigma^2/\sum_{i=1}^{n} X_i^2.$$

This is larger than $MSE(\tilde{\beta})$ since the latter has a positive constant (σ^2/β^2) in the denominator.

3.16 a. Dependent Variable: LNEN

Analysis of Variance

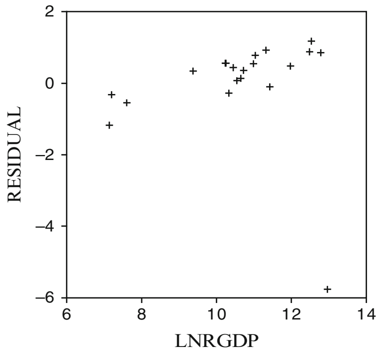
Source	DF	Sum of Squares	Mean Square	F Value	Prob > F
Model	1	30.82384	30.82384	13.462	0.0018
Error	18	41.21502	2.28972		
C Total	19	72.03886			

Root MSE	1.51318	R-square	0.4279
Dep Mean	9.87616	Adj R-sq	0.3961
C.V.	15.32158		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob > |T|
INTERCEP	1	1.988607	2.17622656	0.914	0.3729
LNRGDP	1	0.743950	0.20276460	3.669	0.0018

b. Plot of Residual *LNRGDP



As clear from this plot, the W. Germany observation has a large residual.

c. For $H_o; \beta = 1$ we get $t = \left(\hat{\beta} - 1\right)/\text{s.e.}\left(\hat{\beta}\right) = (0.744 - 1)/0.203 = -1.26$. We do not reject H_o.

e. Dependent Variable: LNEN

Analysis of Variance

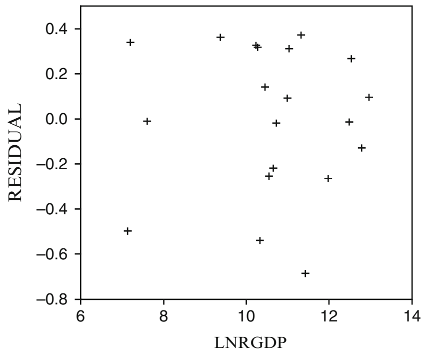
Source	DF	Sum of Squares	Mean Square	F Value	Prob > F
Model	1	59.94798	59.94798	535.903	0.0001
Error	18	2.01354	0.11186		
C Total	19	61.96153			

Root MSE	0.33446	R-square	0.9675
Dep Mean	10.22155	Adj R-sq	0.9657
C.V.	3.27211		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob > |T|
INTERCEP	1	-0.778287	0.48101294	-1.618	0.1230
LNRGDP	1	1.037499	0.04481721	23.150	0.0001

Plot of Residual*LNRGDP



SAS PROGRAM

```
Data Rawdata;
    Input Country $ RGDP EN;
    Cards;
    Data Energy; Set Rawdata;
    LNRGDP=log(RGDP);     LNEN=log(EN);
    Proc reg data=energy;
        Model LNEN=LNRGDP;
```

```
    Output out=OUT1 R=RESID;
Proc Plot data=OUT1 hpercent=85 vpercent=60;
        Plot RESID*LNRGDP='*';
run;
```

3.17 b. Dependent Variable: LNRGDP

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob > F
Model	1	53.88294	53.88294	535.903	0.0001
Error	18	1.80983	0.10055		
C Total	19	55.69277			

Root MSE	0.31709	R-square	0.9675
Dep Mean	10.60225	Adj R-sq	0.9657
C.V.	2.99078		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob > |T|
INTERCEP	1	1.070317	0.41781436	2.562	0.0196
LNEN	1	0.932534	0.04028297	23.150	0.0001

e. Log–log specification

Dependent Variable: LNEN1

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob > F
Model	1	59.94798	59.94798	535.903	0.0001
Error	18	2.01354	0.11186		
C Total	19	61.96153			

Root MSE	0.33446	R-square	0.9675
Dep Mean	14.31589	Adj R-sq	0.9657
C.V.	2.33628		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob > |T|
INTERCEP	1	3.316057	0.48101294	6.894	0.0001
LNRGDP	1	1.037499	0.04481721	23.150	0.0001

Linear Specificiation

Dependent Variable: EN1

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob > F
Model	1	6.7345506E14	6.7345506E14	386.28	0.0001
Error	18	3.1381407E13	1.7434115E12	6	
C Total	19	7.0483646E14			

Root MSE	1320383.09457	R-square	0.9555
Dep Mean	4607256.0000	Adj R-sq	0.9530
C.V.	28.65877		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob > |T|
INTERCEP	1	-190151	383081.08995	-0.496	0.6256
RGDP	1	46.759427	2.37911166	19.654	0.0001

Linear Specification before the multiplication by 60

Dependent Variable: EN

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob > F
Model	1	187070848717	187070848717	386.286	0.0001
Error	18	8717057582.2	484280976.79		
C Total	19	195787906299			

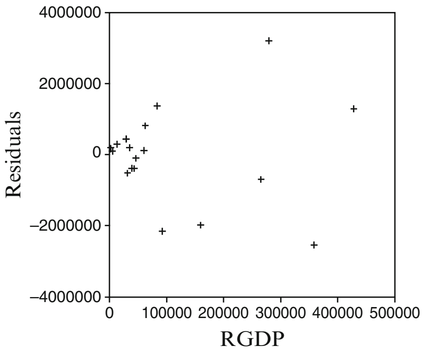
Root MSE	22006.38491	R-square	0.9555
Dep Mean	76787.60000	Adj R-sq	0.9530
C.V.	28.6587		

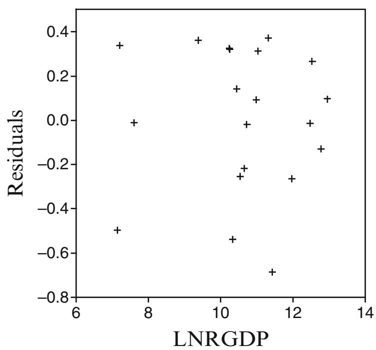
Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob > |T|
INTERCEP	1	-3169.188324	6384.6848326	-0.496	0.6256
RGDP	1	0.779324	0.03965186	19.654	0.0001

What happens when we multiply our energy variable by 60? For the linear model specification, both $\hat{\alpha}^*$ and $\hat{\beta}^*$ are multiplied by 60, their standard errors are also multiplied by 60 and their t-statistics are the same.

For the log–log model specification, $\hat{\beta}$ is the same, but $\hat{\alpha}$ is equal to the old $\hat{\alpha} + \log 60$. The intercept therefore is affected but not the slope. Its standard error is the same, but its t-statistic is changed.





g. Plot of residuals for both linear and log–log models

SAS PROGRAM

```
Data Rawdata;
   Input Country $ RGDP EN;
   Cards;
   Data Energy; Set Rawdata;
   LNRGDP=log (RGDP); LNEN=log(EN);
   EN1=EN*60;    LNEN1=log(EN1);
   Proc reg data=energy; Model LNRGDP=LNEN;
   Proc reg data=energy; Model LNEN1=LNRGDP/CLM, CLI;
      Output out=OUT1 R=LN_RESID;
   Proc reg data=energy; Model EN1=RGDP;
      Output out=OUT2 R=RESID;
```

```
Proc reg data=energy; Model EN=RGDP;
data Resid; set out1 (keep=lnrgdp ln_resid);
     set out2(keep=rgdp resid);
Proc plot data=resid vpercent=60 hpercent=85;
     Plot ln_resid*lnrgdp='*';
     Plot resid*rgdp='*';
run;
```

3.18 For parts (b) and (c), SAS will automatically compute confidence intervals for the mean (CLM option) and for a specific observation (CLI option), see the SAS program in 3.17.

95% CONFIDENCE PREDICTION INTERVAL

COUNTRY	Dep Var LNEN1	Predict Value	Std Err Predict	Lower95% Mean	Upper95% Mean	Lower95% Predict	Upper95% Predict
AUSTRIA	14.4242	14.4426	0.075	14.2851	14.6001	13.7225	15.1627
BELGIUM	15.0778	14.7656	0.077	14.6032	14.9279	14.0444	15.4868
CYPRUS	11.1935	11.2035	0.154	10.8803	11.5268	10.4301	11.9770
DENMARK	14.2997	14.1578	0.075	14.0000	14.3156	13.4376	14.8780
FINLAND	14.2757	13.9543	0.076	13.7938	14.1148	13.2335	14.6751
FRANCE	16.4570	16.5859	0.123	16.3268	16.8449	15.8369	17.3348
GREECE	14.0038	14.2579	0.075	14.1007	14.4151	13.5379	14.9780
ICELAND	11.1190	10.7795	0.170	10.4222	11.1368	9.9912	11.5678
IRELAND	13.4048	13.0424	0.093	12.8474	13.2375	12.3132	13.7717
ITALY	16.2620	16.2752	0.113	16.0379	16.5125	15.5335	17.0168
MALTA	10.2168	10.7152	0.173	10.3526	11.0778	9.9245	11.5059
NETHERLAND	15.4379	15.0649	0.081	14.8937	15.2361	14.3417	15.7882
NORWAY	14.2635	13.9368	0.077	13.7760	14.0977	13.2160	14.6577
PORTUGAL	13.4937	14.0336	0.076	13.8744	14.1928	13.3131	14.7540
SPAIN	15.4811	15.7458	0.097	15.5420	15.9495	15.0141	16.4774
SWEDEN	14.8117	14.7194	0.077	14.5581	14.8808	13.9985	15.4404
SWZERLAND	14.1477	14.3665	0.075	14.2094	14.5237	13.6465	15.0866
TURKEY	14.4870	15.1736	0.083	14.9982	15.3489	14.4494	15.8978
UK	16.5933	16.3259	0.115	16.0852	16.5667	15.5832	17.0687
W. GERMANY	16.8677	16.7713	0.130	16.4987	17.0440	16.0176	17.5251

References

Baltagi, B.H. (1995), "Optimal Weighting of Unbiased Estimators," *Econometric Theory*, Problem 95.3.1, 11:637.

Samuel-Cahn, E. (1994), "Combining Unbiased Estimators," The American Statistician, 48: 34–36.

Trenkler, G. (1996), "Optimal Weighting of Unbiased Estimators," *Econometric Theory*, Solution 95.3.1, 12: 585.

Zheng, J. X. (1994), "Efficiency as Correlation," *Econometric Theory*, Solution 93.1.3, 10:228.

CHAPTER 4
Multiple Regression Analysis

4.1 The regressions for parts (a), (b), (c), (d) and (e) are given below.

a. Regression of LNC on LNP and LNY

Dependent Variable: LNC

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	0.50098	0.25049	9.378	0.0004
Error	43	1.14854	0.02671		
C Total	45	1.64953			

Root MSE	0.16343	R-square	0.3037
Dep Mean	4.84784	Adj R-sq	0.2713
C.V.	3.37125		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob>|T|
INTERCEP	1	4.299662	0.90892571	4.730	0.0001
LNP	1	-1.338335	0.32460147	-4.123	0.0002
LNY	1	0.172386	0.19675440	0.876	0.3858

b. Regression of LNC on LNP

Dependent Variable: LNC

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	0.48048	0.48048	18.084	0.0001
Error	44	1.16905	0.02657		
C Total	45	1.64953			

Root MSE	0.16300	R-square	0.2913
Dep Mean	4.84784	Adj R-sq	0.2752
C.V.	3.36234		

B.H. Baltagi, *Solutions Manual for Econometrics*, Springer Texts in Business and Economics, DOI 10.1007/978-3-642-54548-1_4,

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob>|T|
INTERCEP	1	5.094108	0.06269897	81.247	0.0001
LNP	1	-1.198316	0.28178857	-4.253	0.0001

c. Regression of LNY on LNP

Dependent Variable: LNY

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob > F
Model	1	0.22075	0.22075	14.077	0.0005
Error	44	0.68997	0.01568		
C Total	45	0.91072			

Root MSE	0.12522	R-square	0.2424
Dep Mean	4.77546	Adj R-sq	0.2252
C.V.	2.62225		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob>|T|
INTERCEP	1	4.608533	0.04816809	95.676	0.0001
LNP	1	0.812239	0.21648230	3.752	0.0005

d. Regression of LNC on the residuals of part (c).

Dependent Variable: LNC

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	0.02050	0.02050	0.554	0.4607
Error	44	1.62903	0.03702		
C Total	45	1.64953			

Root MSE	0.19241	R-square	0.0124
Dep Mean	4.84784	Adj R-sq	-0.0100
C.V.	3.96907		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob>|T|
INTERCEP	1	4.847844	0.02836996	170.879	0.0001
RESID_C	1	0.172386	0.23164467	0.744	0.4607

e. Regression of Residuals from part (b) on those from part (c).

Dependent Variable: RESID_B

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	0.02050	0.02050	0.785	0.3803
Error	44	1.14854	0.02610		
C Total	45	1.16905			

Root MSE	0.16157	R-square	0.0175
Dep Mean	-0.00000	Adj R-sq	-0.0048
C.V.	-2.463347E16		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob > |T|
INTERCEP	1	-6.84415 E-16	0.02382148	-0.000	1.0000
RESID_C	1	0.172386	0.19450570	0.886	0.3803

f. The multiple regression coefficient estimate of real income in part (a) is equal to the slope coefficient of the regressions in parts (d) and (e). This demonstrates the residualing out interpretation of multiple regression coefficients.

4.2 *Simple Versus Multiple Regression Coefficients*. This is based on Baltagi (1987).

The OLS residuals from $Y_i = \gamma + \delta_2\hat{\nu}_{2i} + \delta_3\hat{\nu}_{3i} + w_i$, say $\hat{w}_i$, satisfy the following conditions:

$$\sum_{i=1}^{n}\hat{w}_i = 0 \qquad \sum_{i=1}^{n}\hat{w}_i\hat{\nu}_{2i} = 0 \qquad \sum_{i=1}^{n}\hat{w}_i\hat{\nu}_{3i} = 0$$

with $Y_i = \hat{\gamma} + \hat{\delta}_2\hat{\nu}_{2i} + \hat{\delta}_3\hat{\nu}_{3i} + \hat{w}_i$.

Multiply this last equation by $\hat{\nu}_{2i}$ and sum, we get $\sum_{i=1}^{n} Y_i\hat{\nu}_{2i} = \hat{\delta}_2 \sum_{i=1}^{n} \hat{\nu}_{2i}^2 + \hat{\delta}_3 \sum_{i=1}^{n} \hat{\nu}_{2i}\hat{\nu}_{3i}$ since $\sum_{i=1}^{n} \hat{\nu}_{2i} = 0$ and $\sum_{i=1}^{n} \hat{\nu}_{2i}\hat{w}_i = 0$. Substituting $\hat{\nu}_{3i} = X_{3i} - \hat{c} - \hat{d}X_{2i}$ and using the fact that $\sum_{i=1}^{n} \hat{\nu}_{2i}X_{3i} = 0$, we get $\sum_{i=1}^{n} Y_i\hat{\nu}_{2i} = \hat{\delta}_2 \sum_{i=1}^{n} \hat{\nu}_{2i}^2 - \hat{\delta}_3\hat{d} \sum_{i=1}^{n} X_{2i}\hat{\nu}_{2i}$. From $X_{2i} = \hat{a} + \hat{b}X_{3i} + \hat{\nu}_{2i}$, we get that $\sum_{i=1}^{n} X_{2i}\hat{\nu}_{2i} = \sum_{i=1}^{n} \hat{\nu}_{2i}^2$.
Hence,

$$\sum_{i=1}^{n} Y_i\hat{\nu}_{2i} = \left(\hat{\delta}_2 - \hat{\delta}_3\hat{d}\right) \sum_{i=1}^{n} \hat{\nu}_{2i}^2 \text{ and } \hat{\beta}_2 = \sum_{i=1}^{n} Y_i\hat{\nu}_{2i} \Big/ \sum_{i=1}^{n} \hat{\nu}_{2i}^2 = \hat{\delta}_2 - \hat{\delta}_3\hat{d}.$$

Similarly, one can show that $\hat{\beta}_3 = \hat{\delta}_3 - \hat{\delta}_2\hat{b}$. Solving for $\hat{\delta}_2$ and $\hat{\delta}_3$ we get

$$\hat{\delta}_2 = \left(\hat{\beta}_2 + \hat{\beta}_3\hat{d}\right)/(1 - \hat{b}\hat{d}) \qquad \text{and} \qquad \hat{\delta}_3 = \left(\hat{\beta}_3 + \hat{b}\hat{\beta}_2\right)/(1 - \hat{b}\hat{d}).$$

4.3 a. Regressing X_i on a constant we get $\hat{X}_i = \overline{X}$ and $\hat{\nu}_i = X_i - \hat{X}_i = X_i - \overline{X} = x_i$.

Regressing Y_i on $\hat{\nu}_i$ we get

$$\hat{\beta}_{ols} = \sum_{i=1}^{n} Y_i\hat{\nu}_i \Big/ \sum_{i=1}^{n} \hat{\nu}_i^2 = \sum_{i=1}^{n} Y_i x_i \Big/ \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \Big/ \sum_{i=1}^{n} x_i^2.$$

b. Regressing a constant 1 on X_i we get

$$b = \sum_{i=1}^{n} X_i \Big/ \sum_{i=1}^{n} X_i^2 \text{ with residuals } \hat{w}_i = 1 - \left(n\overline{X} \Big/ \sum_{i=1}^{n} X_i^2\right) X_i$$

so that, regressing Y_i on $\hat{w}_i$ yields $\hat{\alpha} = \sum_{i=1}^{n} \hat{w}_i Y_i / \sum_{i=1}^{n} \hat{w}_i^2$.

But $\sum_{i=1}^{n} \hat{w}_i Y_i = n\overline{Y} - n\overline{X} \sum_{i=1}^{n} X_i Y_i / \sum_{i=1}^{n} X_i^2 = \dfrac{n\overline{Y} \sum_{i=1}^{n} X_i^2 - n\overline{X} \sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2}$ and

$$\sum_{i=1}^{n} \hat{w}_i^2 = n + \frac{n^2\overline{X}^2}{\sum_{i=1}^{n} X_i^2} - \frac{2n\overline{X} \sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} X_i^2} = \frac{n \sum_{i=1}^{n} X_i^2 - n^2\overline{X}^2}{\sum_{i=1}^{n} X_i^2} = \frac{n \sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n} X_i^2}.$$

$$\text{Therefore, } \hat{\alpha} = \frac{\overline{Y}\sum_{i=1}^{n} X_i{}^2 - \overline{X}\sum_{i=1}^{n} X_iY_i}{\sum_{i=1}^{n} x_i{}^2} = \frac{\overline{Y}\sum_{i=1}^{n} x_i{}^2 + n\overline{X}^2\overline{Y} - \overline{X}\sum_{i=1}^{n} X_iY_i}{\sum_{i=1}^{n} x_i{}^2}$$

$$= Y - \frac{\overline{X}\left(\sum_{i=1}^{n} X_iY_i - n\overline{XY}\right)}{\sum_{i=1}^{n} x_i{}^2} = \overline{Y} - \hat{\beta}_{ols}\overline{X}$$

where $\hat{\beta}_{ols} = \sum_{i=1}^{n} x_iy_i \Big/ \sum_{i=1}^{n} x_i{}^2$.

c. From part (a), $\text{var}(\hat{\beta}_{ols}) = \sigma^2 / \sum_{i=1}^{n} \hat{\nu}_i^2 = \sigma^2 / \sum_{i=1}^{n} x_i{}^2$ as it should be. From part (b), $\text{var}\,(\hat{\alpha}_{ols}) = \sigma^2 / \sum_{i=1}^{n} \hat{w}_i^2 = \dfrac{\sigma^2 \sum_{i=1}^{n} X_i^2}{n \sum_{i=1}^{n} x_i{}^2}$ as it should be.

4.4 *Effect of Additional Regressors on R^2*

a. Least Squares on the $K = K_1 + K_2$ regressors minimizes the sum of squared error and yields $SSE_2 = \min \sum_{i=1}^{n} (Y_i - \alpha - \beta_2X_{2i} - .. - \beta_{K_1} X_{K_1i} - .. - \beta_K X_{Ki})^2$ Let us denote the corresponding estimates by $(a, b_2, .., b_{K_1}, .., b_K)$. This implies that $SSE_2^* = \sum_{i=1}^{n} (Y_i - \alpha^* - \beta_2^* X_{2i} - .. - \beta_{K_1}^* X_{K_1i} - .. - \beta_K^* X_{Ki})^2$ based on arbitrary $(\alpha^*, \beta_2^*, .., \beta_{K_1}^*, .., \beta_K^*)$ satisfies $SSE_2^* \geq SSE_2$. In particular, substituting the least squares estimates using only K_1 regressors say $\hat{\alpha}, \hat{\beta}_2, .., \hat{\beta}_{K_1}$ and $\hat{\beta}_{K_1+1} = 0, .., \hat{\beta}_K = 0$ satisfy the above inequality. Hence, $SSE_1 \geq SSE_2$. Since $\sum_{i=1}^{n} y_i{}^2$ is fixed, this means that $R_2^2 \geq R_1^2$. This is based on the solution by Rao and White (1988).

b. From the definition of $\overline{R}^2$, we get $(1 - \overline{R}^2) = \left[\sum_{i=1}^{n} e_i{}^2 / (n - K)\right] \Big/ \left[\sum_{i=1}^{n} y_i{}^2 / (n-1)\right]$. But $R^2 = 1 - \left(\sum_{i=1}^{n} e_i{}^2 \Big/ \sum_{i=1}^{n} y_i{}^2\right)$, hence $(1 - \overline{R}^2) = \dfrac{\Sigma e_i{}^2}{\Sigma y_i{}^2} \cdot \dfrac{n-1}{n-k} = (1 - R^2) \cdot \dfrac{n-1}{n-k}$ as required in (4.16).

4.5 This regression suffers from perfect multicollinearity. $X_2 + X_3$ is perfectly collinear with X_2 and X_3. Collecting terms in X_2 and X_3 we get

$$Y_i = \alpha + (\beta_2 + \beta_4)X_{2i} + (\beta_3 + \beta_4)X_{3i} + \beta_5 X_{2i}^2 + \beta_6 X_{3i}^2 + u_i$$

so

$(\beta_2 + \beta_4)$, $(\beta_3 + \beta_4)$, β_5 and β_6 are estimable by OLS.

4.6 **a.** If we regress e_t on X_{2t} and X_{3t} and a constant, we get zero regression coefficients and therefore zero predicted values, i.e., $\hat{e}_t = 0$. The residuals are therefore equal to $e_t - \hat{e}_t = e_t$ and the residual sum of squares is equal to the total sum of squares. Therefore, $R^2 = 1 - \dfrac{RSS}{TSS} = 1 - 1 = 0$.

b. For $Y_t = a + b\hat{Y}_t + \nu_t$, OLS yields $\hat{b} = \sum_{t=1}^{T} \hat{y}_t y_t \Big/ \sum_{t=1}^{T} \hat{y}_t^2$ and $\hat{a} = \overline{Y} - \hat{b}\overline{\hat{Y}}$. Also, $Y_t = \hat{Y}_t + e_t$ which gives $\overline{Y} = \overline{\hat{Y}}$ since $\sum_{t=1}^{T} e_t = 0$. Also, $y_t = \hat{y}_t + e_t$. Therefore, $\sum_{t=1}^{T} y_t\hat{y}_t = \sum_{t=1}^{T} \hat{y}_t^2$ since $\sum_{t=1}^{T} \hat{y}_t e_t = 0$. Hence,

$$\hat{b} = \sum_{t=1}^{T} \hat{y}_t^2 \Big/ \sum_{t=1}^{T} \hat{y}_t^2 = 1 \qquad \text{and} \qquad \hat{a} = \overline{Y} - 1 \cdot \overline{Y} = 0$$

$$\hat{Y}_t = 0 + 1 \cdot \hat{Y}_t = \hat{Y}_t \qquad \text{and} \qquad Y_t - \hat{Y}_t = e_t$$

$\sum_{t=1}^{T} e_t^2$ is the same as the original regression, $\sum_{t=1}^{T} y_t^2$ is still the same, therefore, R^2 is still the same.

c. For $Y_t = a + be_t + \nu_t$, OLS yields $\hat{b} = \sum_{t=1}^{T} e_t y_t \Big/ \sum_{t=1}^{T} e_t^2$ and $\hat{a} = \overline{Y} - \hat{b}\overline{e} = \overline{Y}$ since $\overline{e} = 0$. But $y_t = \hat{y}_t + e_t$, therefore, $\sum_{t=1}^{T} e_t y_t = \sum_{t=1}^{T} e_t\hat{y}_t + \sum_{t=1}^{T} e_t^2 = \sum_{t=1}^{T} e_t^2$ since $\sum_{t=1}^{T} e_t\hat{y}_t = 0$. Hence, $\hat{b} = \sum_{t=1}^{T} e_t^2 \Big/ \sum_{t=1}^{T} e_t^2 = 1$.
Also, the predicted value is now $\hat{Y}_t = \hat{a} + \hat{b}e_t = \overline{Y} + e_t$ and the new residual is $= Y_t - \hat{Y}_t = y_t - e_t = \hat{y}_t$. Hence, the new RSS = old regression sum of

squares $= \sum_{t=1}^{T} \hat{y}_t^2$, and

$$(\text{newR}^2) = 1 - \frac{\text{new RSS}}{\text{TSS}} = 1 - \frac{\text{Old Reg.SS}}{\text{TSS}} = \frac{\sum_{t=1}^{T} e_t^2}{\sum_{t=1}^{T} y_t^2} = (1 - (\text{old } R^2)).$$

4.7 For the Cobb–Douglas production given by Eq. (4.18) one can test H_o; $\alpha + \beta + \gamma + \delta = 1$ using the following t-statistic $t = \frac{(\hat{\alpha}+\hat{\beta}+\hat{\gamma}+\hat{\delta})-1}{\text{s.e.}(\hat{\alpha}+\hat{\beta}+\hat{\gamma}+\hat{\delta})}$ where the estimates are obtained from the unrestricted OLS regression given by (4.18). The $\text{var}(\hat{\alpha} + \hat{\beta} + \hat{\gamma} + \hat{\delta}) = \text{var}(\hat{\alpha}) + \text{var}(\hat{\beta}) + \text{var}(\hat{\gamma}) + \text{var}(\hat{\delta}) + 2\text{cov}(\hat{\alpha}, \hat{\beta}) + 2\text{cov}(\hat{\alpha}, \hat{\gamma}) + 2\text{cov}(\hat{\alpha}, \hat{\delta}) + 2\text{cov}(\hat{\beta}, \hat{\gamma}) + 2\text{cov}(\hat{\beta}, \hat{\delta}) + 2\text{cov}(\hat{\gamma}, \hat{\delta})$. These variance–covariance estimates are obtained from the unrestricted regression. The observed t-statistic is distributed as t_{n-5} under H_o.

4.8 **a.** The restricted regression for H_o; $\beta_2 = \beta_4 = \beta_6$ is given by $Y_i = \alpha + \beta_2(X_{2i} + X_{4i} + X_{6i}) + \beta_3 X_{3i} + \beta_5 X_{5i} + \beta_7 X_{7i} + \beta_8 X_{8i} + .. + \beta_K X_{Ki} + u_i$ obtained by substituting $\beta_2 = \beta_4 = \beta_6$ in Eq. (4.1). The unrestricted regression is given by (4.1) and the F-statistic in (4.17) has two restrictions and is distributed $F_{2,n-K}$ under H_o.

b. The restricted regression for H_o; $\beta_2 = -\beta_3$ and $\beta_5 - \beta_6 = 1$ is given by $Y_i + X_{6i} = \alpha + \beta_2(X_{2i} - X_{3i}) + \beta_4 X_{4i} + \beta_5(X_{5i} + X_{6i}) + \beta_7 X_{7i} + .. + \beta_K X_{Ki} + u_i$ obtained by substituting both restrictions in (4.1). The unrestricted regression is given by (4.1) and the F-statistic in (4.17) has two restrictions and is distributed $F_{2,n-K}$ under H_o.

4.10 **a.** For the data underlying Table 4.1, the following computer output gives the mean of log(wage) for females and for males. Out of 595 individuals observed, there were 528 Males and 67 Females. The corresponding means of log(wage) for Males and Females being $\overline{Y}_M = 7.004$ and $\overline{Y}_F = 6.530$, respectively. The regression of log(wage) on FEMALE and MALE without a constant yields coefficient estimates $\hat{\alpha}_F = \overline{Y}_F = 6.530$ and $\hat{\alpha}_M = \overline{Y}_M = 7.004$, as expected.

Dependent Variable: LWAGE

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	28759.48792	14379.74396	84576.019	0.0001
Error	593	100.82277	0.17002		
U Total	595	28860.31068			

Root MSE	0.41234	R-square	0.9965
Dep Mean	6.95074	Adj R-sq	0.9965
C.V.	5.93227		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob>|T|
FEM	1	6.530366	0.05037494	129.635	0.0001
M	1	7.004088	0.01794465	390.316	0.0001

FEM=0

Variable	N	Mean	Std Dev	Minimum	Maximum
LWAGE	528	7.0040880	0.4160069	5.6767500	8.5370000

FEM=1

Variable	N	Mean	Std Dev	Minimum	Maximum
LWAGE	67	6.5303664	0.3817668	5.7493900	7.2793200

b. Running log(wage) on a constant and the FEMALE dummy variable yields

$\hat{\alpha} = 7.004 = \overline{Y}_M = \hat{\alpha}_M$ and $\hat{\beta} = -0.474 = \left(\widehat{\alpha_F - \alpha_M}\right)$. But $\hat{\alpha}_M = 7.004$.

Therefore, $\hat{\alpha}_F = \hat{\beta} + \hat{\alpha}_M = 7.004 - 0.474 = 6.530 = \overline{Y}_F = \hat{\alpha}_F$.

Dependent Variable: LWAGE

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	13.34252	13.34252	78.476	0.0001
Error	593	100.82277	0.17002		
C Total	594	114.16529			

Root MSE	0.41234	R-square	0.1169
Dep Mean	6.95074	Adj R-sq	0.1154
C.V.	5.93227		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob>|T|
INTERCEP	1	7.004088	0.01794465	390.316	0.0001
FEM	1	-0.473722	0.05347565	-8.859	0.0001

4.12 a. The unrestricted regression is given by (4.28). This regression runs EARN on a constant, FEMALE, EDUCATION and (FEMALE $\times$ EDUCATION). The URSS $=$ 76.63525. The restricted regression for equality of slopes and intercepts for Males and Females, tests the restriction H_o; $\alpha_F = \gamma = 0$. This regression runs EARN on a constant and EDUC. The RRSS $=$ 90.36713. The SAS regression output is given below. There are two restrictions and the F-test given by (4.17) yields

$$F = \frac{(90.36713 - 76.63525)/2}{76.63525/591} = 52.94941.$$

This is distributed as $F_{2,591}$ under H_o. The null hypothesis is rejected.

Unrestricted Model (with FEM and FEM*EDUC)

Dependent Variable: LWAGE

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	3	37.53004	12.51001	96.475	0.0001
Error	591	76.63525	0.12967		
C Total	594	114.16529			

Root MSE	0.36010	R-square	0.3287
Dep Mean	6.95074	Adj R-sq	0.3253
C.V.	5.18071		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob>|T|
INTERCEP	1	6.122535	0.07304328	83.821	0.0001
FEM	1	-0.905504	0.24132106	-3.752	0.0002
ED	1	0.068622	0.00555341	12.357	0.0001
F_EDC	1	0.033696	0.01844378	1.827	0.0682

Restricted Model (without FEM and FEM*EDUC)

Dependent Variable:˜LWAGE

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	23.79816	23.79816	156.166	0.0001
Error	593	90.36713	0.15239		
C Total	594	114.16529			

Root MSE	0.39037	R-square	0.2085
Dep Mean	6.95074	Adj R-sq	0.2071
C.V.	5.61625		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob>|T|
INTERCEP	1	6.029192	0.07546051	79.899	0.0001
ED	1	0.071742	0.00574089	12.497	0.0001

b. The unrestricted regression is given by (4.27). This regression runs EARN on a constant, FEMALE and EDUCATION. The URSS = 77.06808. The restricted regression for the equality of intercepts given the same slopes for Males and Females, tests the restriction H_o; $\alpha_F = 0$ given that $\gamma = 0$. This is the same restricted regression given in part (a), running EARN on a constant and EDUC. The RRSS = 90.36713. The F-test given by (4.17) tests one restriction and yields

$$F = \frac{(90.36713 - 77.06808)/1}{77.06808/592} = 102.2.$$

This is distributed as $F_{1,592}$ under H_o. Note that this observed F-statistic is the square of the observed t-statistic of -10.107 for $\alpha_F = 0$ in the unrestricted regression. The SAS regression output is given below.

Unrestricted Model (with FEM)

Dependent Variable: LWAGE

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	37.09721	18.54861	142.482	0.0001
Error	592	77.06808	0.13018		
C Total	594	114.16529			

Root MSE	0.36081	R-square	0.3249
Dep Mean	6.95074	Adj R-sq	0.3227
C.V.	5.19093		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob>|T|
INTERCEP	1	6.083290	0.06995090	86.965	0.0001
FEM	1	-0.472950	0.04679300	-10.107	0.0001
ED	1	0.071676	0.00530614	13.508	0.0001

c. The unrestricted regression is given by (4.28), see part (a). The restricted regression for the equality of intercepts allowing for different slopes for Males and Females, tests the restriction H_o; $\alpha_F = 0$ given that $\gamma \neq 0$. This regression runs EARN on a constant, EDUCATION and (FEMALE × EDUCATION). The RRSS $= 78.46096$. The SAS regression output is given below. The F-test given by (4.17), tests one restriction and yields:

$$F = \frac{(78.46096 - 76.63525)/1}{76.63525/591} = 14.0796.$$

This is distributed as $F_{1,591}$ under H_o. The null hypothesis is rejected. Note that this observed F-statistic is the square of the t-statistic (-3.752) on $\alpha_F = 0$ in the unrestricted regression.

Restricted Model (without FEM)

Dependent Variable: LWAGE

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	35.70433	17.85216	134.697	0.0001
Error	592	78.46096	0.13254		
C Total	594	114.16529			

Root MSE	0.36405	R-square	0.3127
Dep Mean	6.95074	Adj R-sq	0.3104
C.V.	5.23763		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob>|T|
INTERCEP	1	6.039577	0.07038181	85.812	0.0001
ED	1	0.074782	0.00536347	13.943	0.0001
F_EDC	1	-0.034202	0.00360849	-9.478	0.0001

4.13 a. Dependent Variable: LWAGE

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	12	52.48064	4.37339	41.263	0.0001
Error	582	61.68465	0.10599		
C Total	594	114.16529			

Root MSE	0.32556	R-square	0.4597
Dep Mean	6.95074	Adj R-sq	0.4485
C.V.	4.68377		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob>|T|
INTERCEP	1	5.590093	0.19011263	29.404	0.0001
EXP	1	0.029380	0.00652410	4.503	0.0001
EXP2	1	-0.000486	0.00012680	-3.833	0.0001

WKS	1	0.003413	0.00267762	1.275	0.2030
OCC	1	-0.161522	0.03690729	-4.376	0.0001
IND	1	0.084663	0.02916370	2.903	0.0038
SOUTH	1	-0.058763	0.03090689	-1.901	0.0578
SMSA	1	0.166191	0.2955099	5.624	0.0001
MS	1	0.095237	0.04892770	1.946	0.0521
FEM	1	-0.324557	0.06072947	-5.344	0.0001
UNION	1	0.106278	0.03167547	3.355	0.0008
ED	1	0.057194	0.00659101	8.678	0.0001
BLK	1	-0.190422	0.05441180	-3.500	0.0005

b. H_o : $EARN = \alpha + u$.

If you run EARN on an intercept only, you would get $\hat{\alpha} = 6.9507$ which is average log wage or average earnings $= \bar{y}$. The total sum of squares $=$ the residual sum of squares $= \sum_{i=1}^{n}(y_i - \bar{y})^2 = 114.16529$ and this is the restricted residual sum of squares (RRSS) needed for the F-test. The unrestricted model is given in Table 4.1 or part (a) and yields URSS $= 61.68465$. Hence, the joint significance for all slopes using (4.20) yields

$$F = \frac{(114.16529 - 61.68465)/12}{61.68465/582} = 41.26 \quad \text{also}$$

$$F = \frac{R^2}{1-R^2} \cdot \frac{n-K}{K-1} = \frac{0.4597}{1-0.4597} \cdot \frac{582}{12} = 41.26.$$

This F-statistic is distributed as $F_{12,582}$ under the null hypothesis. It has a p-value of 0.0001 as shown in Table 4.1 and we reject H_o. The Analysis of Variance table in the SAS output given in Table 4.1 always reports this F-statistic for the significance of all slopes for any regression.

c. The restricted model excludes FEM and BLACK. The SAS regression output is given below. The RRSS $= 66.27893$. The unrestricted model is given in Table 4.1 with URSS $= 61.68465$. The F-statistic given in (4.17) tests two restrictions and yields

$$F = \frac{(66.27893 - 61.68465)/2}{61.68465/582} = 21.6737.$$

This is distributed as $F_{2,582}$ under the null hypothesis. We reject H_o.

Model:Restricted Model (w/o FEMALE & BLACK)

Dependent Variable: LWAGE

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	10	47.88636	4.78864	42.194	0.0001
Error	584	66.27893	0.11349		
C Total	594	114.16529			

Root MSE	0.33688	R-square	0.4194
Dep Mean	6.95074	Adj R-sq	0.4095
C.V.	4.84674		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob>|T|
INTERCEP	1	5.316110	0.19153698	27.755	0.0001
EXP	1	0.028108	0.00674771	4.165	0.0001
EXP2	1	-0.000468	0.00013117	-3.570	0.0004
WKS	1	0.004527	0.00276523	1.637	0.1022
OCC	1	-0.162382	0.03816211	-4.255	0.0001
IND	1	0.102697	0.03004143	3.419	0.0007
SOUTH	1	-0.073099	0.03175589	-2.302	0.0217
SMSA	1	0.142285	0.03022571	4.707	0.0001
MS	1	0.298940	0.03667049	8.152	0.0001
UNION	1	0.112941	0.03271187	3.453	0.0006
ED	1	0.059991	0.00680032	8.822	0.0001

d. The restricted model excludes MS and UNION. The SAS regression output is given below. This yields RRSS = 63.37107. The unrestricted model is given in Table 4.1 and yields URSS = 61.68465. The F-test given in (4.17) tests two restrictions and yields

$$F = \frac{(63.37107 - 61.68465)/2}{61.68465/582} = 7.9558.$$

This is distributed as $F_{2,582}$ under the null hypothesis. We reject H_o.

Restricted Model (without MS & UNION)

Dependent Variable: LWAGE

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	10	50.79422	5.07942	46.810	0.0001
Error	584	63.37107	0.10851		
C Total	594	114.16529			

Root MSE	0.32941	R-square	0.4449
Dep Mean	6.95074	Adj R-sq	0.4354
C.V.	4.73923		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob>|T|
INTERCEP	1	5.766243	0.18704262	30.828	0.0001
EXP	1	0.031307	0.00657565	4.761	0.0001
EXP2	1	-0.000520	0.00012799	-4.064	0.0001
WKS	1	0.001782	0.00264789	0.673	0.5013
OCC	1	-0.127261	0.03591988	-3.543	0.0004
IND	1	0.089621	0.02948058	3.040	0.0025
SOUTH	1	-0.077250	0.03079302	-2.509	0.0124
SMSA	1	0.172674	0.02974798	5.805	0.0001
FEM	1	-0.425261	0.04498979	-9.452	0.0001
ED	1	0.056144	0.00664068	8.454	0.0001
BLK	1	-0.197010	0.5474680	-3.599	0.0003

e. From Table 4.1, using the coefficient estimate on Union, $\hat{\beta}_u = 0.106278$, we obtain $\hat{g}_u = e^{\hat{\beta}_u} - 1 = e^{0.106278} - 1 = 0.112131$ or (11.2131%). If the disturbances are log normal, Kennedy's (1981) suggestion yields $\tilde{g}_u = e^{\hat{\beta}_u - 0.5 \cdot \mathrm{var}(\hat{\beta}_u)} - 1 = e^{0.106278 - 0.5(0.001003335)} - 1 = 0.111573$ or (11.1573%).

f. From Table 4.1, using the coefficient estimate on MS, $\hat{\beta}_{MS} = 0.095237$, we obtain $\hat{g}_{MS} = e^{\hat{\beta}_{MS}} - 1 = e^{0.095237} - 1 = 0.09992$ or (9.992%).

4.14 Crude Quality

a. Regression of POIL on GRAVITY and SULPHUR.

Dependent Variable: POIL

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	249.21442	124.60721	532.364	0.0001
Error	96	22.47014	0.23406		
C Total	98	271.68456			

Root MSE	0.48380	R-square	0.9173
Dep Mean	15.33727	Adj R-sq	0.9156
C.V.	3.15442		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob>|T|
INTERCEP	1	12.354268	0.23453113	52.676	0.0001
GRAVITY	1	0.146640	0.00759695	19.302	0.0001
SULPHUR	1	-0.414723	0.04462224	-9.294	0.0001

b. Regression of GRAVITY on SULPHUR.

Dependent Variable: GRAVITY

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	2333.89536	2333.89536	55.821	0.0001
Error	97	4055.61191	41.81043		
C Total	98	6389.50727			

Root MSE	6.46610	R-square	0.3653
Dep Mean	24.38182	Adj R-sq	0.3587
C.V.	26.52017		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob>|T|
INTERCEP	1	29.452116	0.93961195	31.345	0.0001
SULPHUR	1	-3.549926	0.47513923	-7.471	0.0001

Regression of POIL on the Residuals from the Previous Regression

Dependent Variable: POIL

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	87.20885	87.20885	45.856	0.0001
Error	97	184.47571	1.90181		
C Total	98	271.68456			

Root MSE	1.37906	R-square	0.3210
Dep Mean	15.33727	Adj R-sq	0.3140
C.V.	8.99157		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob>|T|
INTERCEP	1	15.337273	0.13860093	110.658	0.0001
RESID_V	1	0.146640	0.02165487	6.772	0.0001

c. Regression of POIL on SULPHUR

Dependent Variable: POIL

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	162.00557	162.00557	143.278	0.0001
Error	97	109.67900	1.13071		
C Total	98	271.68456			

Root MSE	1.06335	R-square	0.5963
Dep Mean	15.33727	Adj R-sq	0.5921
C.V.	6.93310		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob>|T|
INTERCEP	1	16.673123	0.15451906	107.903	0.0001
SULPHUR	1	-0.935284	0.07813658	-11.970	0.0001

Regression of Residuals in part (c) on those in part (b).

Dependent Variable: RESID_W

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	87.20885	87.20885	376.467	0.0001
Error	97	22.47014	0.23165		
C Total	98	109.67900			

Root MSE	0.48130	R-square	0.7951
Dep Mean	0.00000	Adj R-sq	0.7930
C.V.	2.127826E16		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob>|T|
INTERCEP	1	3.082861E-15	0.04837260	0.000	1.0000
RESID_V	1	0.146640	0.00755769	19.403	0.0001

d. Regression based on the first 25 crudes.

Dependent Variable: POIL

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	37.68556	18.84278	169.640	0.0001
Error	22	2.44366	0.11108		
C Total	24	40.12922			

Root MSE	0.33328	R-square	0.9391
Dep Mean	15.65560	Adj R-sq	0.9336
C.V.	2.12882		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob>|T|
INTERCEP	1	11.457899	0.34330283	33.375	0.0001
GRAVITY	1	0.166174	0.01048538	15.848	0.0001
SULPHUR	1	0.110178	0.09723998	1.133	0.2694

e. Deleting all crudes with sulphur content outside the range of 1–2%.

Dependent Variable: POIL

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	28.99180	14.49590	128.714	0.0001
Error	25	2.81553	0.11262		
C Total	27	31.80732			

Root MSE	0.33559	R-square	0.9115
Dep Mean	15.05250	Adj R-sq	0.9044
C.V.	2.22947		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob>|T|
INTERCEP	1	11.090789	0.37273724	29.755	0.0001
GRAVITY	1	0.180260	0.01123651	16.042	0.0001
SULPHUR	1	0.176138	0.18615220	0.946	0.3531

SAS PROGRAM

```
Data Crude;
Input POIL GRAVITY SULPHUR;
Cards;

Proc reg data=CRUDE;
        Model POIL=GRAVITY SULPHUR;

Proc reg data=CRUDE;
        Model GRAVITY=SULPHUR;
    Output out=OUT1 R=RESID_V;
run;

Data CRUDE1; set crude; set OUT1(keep=RESID_V);

Proc reg data=CRUDE1;
```

```
        Model POIL=RESID_V;

Proc reg data=CRUDE1;
        Model POIL=SULPHUR;
    Output out=OUT2 R=RESID_W;

Proc reg data=OUT2;
        Model RESID_W=RESID_V;

Data CRUDE2; set CRUDE(firstobs=1 obs=25);

Proc reg data=CRUDE2;
        Model POIL=GRAVITY SULPHUR;
run;

data CRUDE3; set CRUDE;
        if SULPHUR < 1 then delete;
        if SULPHUR > 2 then delete;

Proc reg data=CRUDE3;
        Model POIL=GRAVITY SULPHUR;
run;
```

4.15 a. MODEL1 (1950–1972)

Dependent Variable: LNQMG

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	5	1.22628	0.24526	699.770	0.0001
Error	17	0.00596	0.00035		
C Total	22	1.23224			

Root MSE	0.01872	R-square	0.9952
Dep Mean	17.96942	Adj R-sq	0.9937
C.V.	0.10418		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob>|T|
INTERCEP	1	1.680143	2.79355393	0.601	0.5555
LNCAR	1	0.363533	0.51515166	0.706	0.4899
LNPOP	1	1.053931	0.90483097	1.165	0.2602
LNRGNP	1	-0.311388	0.16250458	-1.916	0.0723
LNPGNP	1	0.124957	0.15802894	0.791	0.4400
LNPMG	1	1.048145	0.26824906	3.907	0.0011

MODEL2(1950–1972)

Dependent Variable: QMG_CAR

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	3	0.01463	0.00488	4.032	0.0224
Error	19	0.02298	0.00121		
C Total	22	0.03762			

Root MSE	0.03478	R-square	0.3890
Dep Mean	-0.18682	Adj R-sq	0.2925
C.V.	-18.61715		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob>|T|
INTERCEP	1	-0.306528	2.37844176	-0.129	0.8988
RGNP_POP	1	-0.139715	0.23851425	-0.586	0.5649
CAR_POP	1	0.054462	0.28275915	0.193	0.8493
PMG_PGNP	1	0.185270	0.27881714	0.664	0.5144

b. From model (2) we have

$$\log \mathrm{QMG} - \log \mathrm{CAR} = \gamma_1 + \gamma_2 \log \mathrm{RGNP} - \gamma_2 \log \mathrm{POP} + \gamma_3 \log \mathrm{CAR}$$

$$- \gamma_3 \log \mathrm{POP} + \gamma_4 \log \mathrm{PMG} - \gamma_4 \log \mathrm{PGNP} + \mathrm{v}$$

For this to be the same as model (1), the following restrictions must hold:

$\beta_1 = \gamma_1$, $\beta_2 = \gamma_3 + 1$, $\beta_3 = -(\gamma_2 + \gamma_3)$, $\beta_4 = \gamma_2$,

$\beta_5 = -\gamma_4$, and $\beta_6 = \gamma_4$.

d. Correlation Analysis

Variables: LNCAR LNPOP LNRGNP LNPGNP LNPMG

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
LNCAR	23	18.15623	0.26033	417.59339	17.71131	18.57813
LNPOP	23	12.10913	0.10056	278.50990	11.93342	12.25437
LNRGNP	23	7.40235	0.24814	170.25415	6.99430	7.81015
LNPGNP	23	3.52739	0.16277	81.12989	3.26194	3.86073
LNPMG	23	-1.15352	0.09203	-26.53096	-1.30195	-0.94675

Pearson Correlation Coefficients/Prob > |R| under Ho: Rho = 0/N = 23

	LNCAR	LNPOP	LNRGNP	LNPGNP	LNPMG
LNCAR	1.00000	0.99588	0.99177	0.97686	0.94374
	0.0	0.0001	0.0001	0.0001	0.0001
LNPOP	0.99588	1.00000	0.98092	0.96281	0.91564
	0.0001	0.0	0.0001	0.0001	0.0001
LNRGNP	0.99177	0.98092	1.00000	0.97295	0.94983
	0.0001	0.0001	0.0	0.0001	0.0001
LNPGNP	0.97686	0.96281	0.97295	1.00000	0.97025
	0.0001	0.0001	0.0001	0.0	0.0001
LNPMG	0.94374	0.91564	0.94983	0.97025	1.00000
	0.0001	0.0001	0.0001	0.0001	0.0

This indicates the presence of multicollinearity among the regressors.

f. MODEL1 (1950–1987)

Dependent Variable: LNQMG

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	5	3.46566	0.69313	868.762	0.0001
Error	32	0.02553	0.00080		
C Total	37	3.49119			

Root MSE	0.02825	R-square	0.9927
Dep Mean	18.16523	Adj R-sq	0.9915
C.V.	0.15550		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob>|T|
INTERCEP	1	9.986981	2.62061670	3.811	0.0006
LNCAR	1	2.559918	0.22812676	11.221	0.0001
LNPOP	1	-2.878083	0.45344340	-6.347	0.0001
LNRGNP	1	-0.429270	0.14837889	-2.893	0.0068
LNPGNP	1	-0.178866	0.06336001	-2.823	0.0081
LNPMG	1	-0.141105	0.04339646	-3.252	0.0027

MODEL2 (1950–1987)

Dependent Variable: QMG_CAR

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	3	0.35693	0.11898	32.508	0.0001
Error	34	0.12444	0.00366		
C Total	37	0.48137			

Root MSE	0.06050	R - square	0.7415
Dep Mean	-0.25616	Adj R - sq	0.7187
C.V.	-23.61671		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob>|T|
INTERCEP	1	-5.853977	3.10247647	-1.887	0.0677
RGNP_POP	1	-0.690460	0.29336969	-2.354	0.0245
CAR_POP	1	0.288735	0.27723429	1.041	0.3050
PMG_PGNP	1	-0.143127	0.07487993	-1.911	0.0644

g. Model 2: SHOCK74 ($= 1$ if year ≥ 1974)

Dependent Variable: QMG_CAR

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	4	0.36718	0.09179	26.527	0.0001
Error	33	0.11419	0.00346		
C Total	37	0.48137			

Root MSE	0.05883	R - square	0.7628
Dep Mean	-0.25616	Adj R - sq	0.7340
C.V.	-22.96396		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	for HO: Parameter=0	Prob>|T|
INTERCEP	1	-6.633509	3.05055820	-2.175	0.0369
SHOCK74	1	-0.073504	0.04272079	-1.721	0.0947
RGNP_POP	1	-0.733358	0.28634866	-2.561	0.0152
CAR_POP	1	0.441777	0.28386742	1.556	0.1292
PMG_PGNP	1	-0.069656	0.08440816	-0.825	0.4152

The t-statistic on SHOCK 74 yields -1.721 with a p-value of 0.0947. This is insignificant. Therefore, we cannot reject that gasoline demand per car did not permanently shift after 1973.

h. Model 2: DUMMY74 ($=$SCHOCK74 $\times$ PMG_PGNP)

Dependent Variable: QMG_CAR

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	4	0.36706	0.09177	26.492	0.0001
Error	33	0.11431	0.00346		
C Total	37	0.48137			

Root MSE	0.05886	R - square	0.7625
Dep Mean	-0.25616	Adj R - sq	0.7337
C.V.	-22.97560		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob>|T|
INTERCEP	1	-6.606761	3.05019318	-2.166	0.0376
RGNP_POP	1	-0.727422	0.28622322	-2.541	0.0159
CAR_POP	1	0.431492	0.28233413	1.528	0.1360
PMG_PGNP	1	-0.083217	0.08083459	-1.029	0.3107
DUMMY74	1	0.015283	0.00893783	1.710	0.0967

The interaction dummy named DUMMY74 has a t-statistic of 1.71 with a p-value of 0.0967. This is insignificant and we cannot reject that the price elasticity did not change after 1973.

SAS PROGRAM

```
Data RAWDATA;
Input Year CAR QMG PMG POP RGNP PGNP;
Cards;

Data USGAS; set RAWDATA;

LNQMG=LOG(QMG);
LNCAR=LOG(CAR);
LNPOP=LOG(POP);
LNRGNP=LOG(RGNP);
LNPGNP=LOG(PGNP);
LNPMG=LOG(PMG);

QMG_CAR=LOG(QMG/CAR);
RGNP_POP=LOG(RGNP/POP);
CAR_POP=LOG(CAR/POP);
PMG_PGNP=LOG(PMG/PGNP);

Data USGAS1; set USGAS;
If YEAR>1972 then delete;

Proc reg data=USGAS1;
   Model LNQMG=LNCAR LNPOP LNRGNP LNPGNP LNPMG;
   Model QMG_CAR=RGNP_POP CAR_POP PMG_PGNP;

Proc corr data=USGAS1;
   Var LNCAR LNPOP LNRGNP LNPGNP LNPMG;
run;

Proc reg data=USGAS;
   Model LNQMG=LNCAR LNPOP LNRGNP LNPGNP LNPMG;
   Model QMG_CAR=RGNP_POP CAR_POP PMG_PGNP;
run;
```

```
data DUMMY1; set USGAS;
If Year>1974 then SHOCK74=0; else SHOCK74=1;
DUMMY74=PMG_PGNP*SHOCK74;

Proc reg data=DUMMY1;
   Model QMG_CAR=SHOCK74 RGNP_POP CAR_POP PMG_PGNP;
   Model QMG_CAR=RGNP_POP CAR_POP PMG_PGNP DUMMY74;
run;
```

4.16 a. MODEL1

Dependent Variable: LNCONS

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	5	165.96803	33.19361	70.593	0.0001
Error	132	62.06757	0.47021		
C Total	137	228.03560			

Root MSE	0.68572	R - square	0.7278
Dep Mean	11.89979	Adj R - sq	0.7175
C.V.	5.76243		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob>|T|
INTERCEP	1	-53.577951	4.53057139	-11.826	0.0001
LNPG	1	-1.425259	0.31539170	-4.519	0.0001
LNPE	1	0.131999	0.50531143	0.261	0.7943
LNPO	1	0.237464	0.22438605	1.058	0.2919
LNHDD	1	0.617801	0.10673360	5.788	0.0001
LNPI	1	6.554349	0.48246036	13.585	0.0001

b. The following plot show that states with low level of consumption are over predicted, while states with high level of consumption are under predicted. One can correct for this problem by either running a separate regression for each state, or use dummy variables for each state to allow for a varying intercept.

c. Model 2

Dependent Variable: LNCONS

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	10	226.75508	22.67551	2248.917	0.0001
Error	127	1.28052	0.01008		
C Total	137	228.03560			

Root MSE	0.10041	R - square	0.9944
Dep Mean	11.89979	Adj R - sq	0.9939
C.V.	0.84382		

```
                 Plot of PRED*LNCONS.  Symbol is value of STATE.

  14 +
     |                                                       N
     |                                                      N
     |
     |                                                      NN
     |                                                     NNN    CCCC
P 13 +                                                     MMM    CC C
r    |                                                     NNMM  CCCC
e    |                                                     NN     CC
d    |                                                     NNM   C
i    |                                                      N     C
c    |                                                      N
t 12 +                                                      M
e    |                           UU                          M
d    |                       U   U  UU                TT   TM
     |                          U                     TTTT
V    |                        UUU  U                  TT T
a    |                          U  U                 TTT
l 11 +        F F               U                     TT
u    |         F F            U    U
e    |        FFF F            U   U                 T
     |               F             U
o    |         FFF
f    |         F  F
  10 +          F
L    |       FFF
N    |      F F
C    |
O    |      F
N    |
S  9 +
     |   F
     |
     |
     |
     |
   8 +
     --+-------------+-------------+-------------+-------------+-------------+--
       9            10            11            12            13            14

                                               LNCONS
```

NOTE: 42 obs hidden.

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob>|T|
INTERCEP	1	6.260803	1.65972981	3.772	0.0002
LNPG	1	-0.125472	0.05512350	-2.276	0.0245
LNPE	1	-0.121120	0.08715359	-1.390	0.1670
LNPO	1	0.155036	0.03706820	4.182	0.0001
LNHDD	1	0.359612	0.07904527	4.549	0.0001
LNPI	1	0.416600	0.16626018	2.506	0.0135
DUMMY_NY	1	-0.702981	0.07640346	-9.201	0.0001
DUMMY_FL	1	-3.024007	0.11424754	-26.469	0.0001
DUMMY_MI	1	-0.766215	0.08491262	-9.024	0.0001
DUMMY_TX	1	-0.679327	0.04838414	-14.040	0.0001
DUMMY_UT	1	-2.597099	0.09925674	-26.165	0.0001

f. Dummy Variable Regression without an Intercept

NOTE: No intercept in model. R-square is redefined.

Dependent Variable: LNCONS

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	11	19768.26019	1797.11456	178234.700	0.0001
Error	127	1.28052	0.01008		
U Total	138	19769.54071			

Root MSE	0.10041	R - square	0.9999
Dep Mean	11.89979	Adj R - sq	0.9999
C.V.	0.84382		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob>|T|
LNPG	1	-0.125472	0.05512350	-2.276	0.0245
LNPE	1	-0.121120	0.08715359	-1.390	0.1670
LNPO	1	0.155036	0.03706820	4.182	0.0001
LNHDD	1	0.359612	0.07904527	4.549	0.0001
LNPI	1	0.416600	0.16626018	2.506	0.0135
DUMMY_CA	1	6.260803	1.65972981	3.772	0.0002
DUMMY_NY	1	5.557822	1.67584667	3.316	0.0012
DUMMY_FL	1	3.236796	1.59445076	2.030	0.0444
DUMMY_MI	1	5.494588	1.67372513	3.283	0.0013
DUMMY_TX	1	5.581476	1.62224148	3.441	0.0008
DUMMY_UT	1	3.663703	1.63598515	2.239	0.0269

SAS PROGRAM

```
Data NATURAL;
Input STATE $ SCODE YEAR Cons Pg Pe Po LPgas HDD Pi;
Cards;
Data NATURAL1; SET NATURAL;
LNCONS=LOG(CONS);
LNPG=LOG(PG) ;
LNPO=LOG(PO) ;
LNPE=LOG(PE) ;
LNHDD=LOG(HDD) ;
LNPI=LOG(PI) ;

******** PROB16.a **********;
********************************;
Proc reg data=NATURAL1;
        Model LNCONS=LNPG LNPE LNPO LNHDD LNPI;
    Output out=OUT1 R=RESID P=PRED;
Proc plot data=OUT1 vpercent=75 hpercent=100;
        plot PRED*LNCONS=STATE;
Data NATURAL2; set NATURAL1;
If STATE='NY' THEN DUMMY_NY=1; ELSE DUMMY_NY=0;
If STATE='FL' THEN DUMMY_FL=1; ELSE DUMMY_FL=0;
If STATE='MI' THEN DUMMY_MI=1; ELSE DUMMY_MI=0;
If STATE='TX' THEN DUMMY_TX=1; ELSE DUMMY_TX=0;
If STATE='UT' THEN DUMMY_UT=1; ELSE DUMMY_UT=0;
If STATE='CA' THEN DUMMY_CA=1; ELSE DUMMY_CA=0;

******** PROB16.c **************;
************************************;
Proc reg data=NATURAL2;
        Model LNCONS=LNPG LNPE LNPO LNHDD LNPI DUMMY_NY
DUMMY_FL DUMMY_MI DUMMY_TX              DUMMY_UT;
******** PROB16.f ************;
********************************;
```

```
Proc reg data=NATURAL2;
     Model LNCONS=LNPG LNPE LNPO LNHDD LNPI
                    DUMMY_CA DUMMY_NY DUMMY_FL DUMMY_MI
DUMMY_TX DUMMY_UT/NOINT;
run;
```

References

Baltagi, B.H. (1987), "Simple versus Multiple Regression Coefficients," *Econometric Theory*, Problem 87.1.1, 3: 159.

Kennedy, P.E. (1981), "Estimation with Correctly Interpreted Dummy Variables in Semilogarithmic Equations," American Economic Review, 71: 802.

Rao, U.L.G. and P.M. White (1988), "Effect of an Additional Regressor on R^2," *Econometric Theory*, Solution 86.3.1, 4: 352.

CHAPTER 5
Violations of the Classical Assumptions

5.1 *s^2 is Biased Under Heteroskedasticity*. From Chap. 3 we have shown that

$$e_i = Y_i - \hat{\alpha}_{ols} - \hat{\beta}_{ols}X_i = y_i - \hat{\beta}_{ols}x_i = \left(\beta - \hat{\beta}_{ols}\right)x_i + (u_i - \bar{u})$$

for $i = 1, 2, .., n$.

The second equality substitutes $\hat{\alpha}_{ols} = \bar{Y} - \hat{\beta}_{ols}\bar{X}$ and the third equality substitutes $y_i = \beta x_i + (u_i - \bar{u})$. Hence,

$$\sum_{i=1}^{n} e_i^2 = \left(\hat{\beta}_{ols} - \beta\right)^2 \sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n} (u_i - \bar{u})^2 - 2\left(\hat{\beta}_{ols} - \beta\right)\sum_{i=1}^{n} x_i(u_i - \bar{u}) \text{ and}$$

$$E\left(\sum_{i=1}^{n} e_i^2\right) = \sum_{i=1}^{n} x_i^2 \text{var}\left(\hat{\beta}_{ols}\right) + E\left[\sum_{t=1}^{T}(u_i - \bar{u})^2\right] - 2E\left(\sum_{i=1}^{n} x_i u_i\right)^2 \Big/ \sum_{i=1}^{n} x_i^2.$$

But $E\left(\sum_{i=1}^{n} x_i u_i\right)^2 = \sum_{i=1}^{n} x_i{}^2\sigma_i^2$ since the u_i's are uncorrelated and heteroskedastic and

$$E\left[\sum_{i=1}^{n}(u_i - \bar{u})^2\right] = E\left(\sum_{i=1}^{n} u_i^2\right) + nE\left(\bar{u}^2\right) - 2E\left(\bar{u}\sum_{i=1}^{n} u_i\right)$$

$$= \sum_{i=1}^{n} E\left(u_i^2\right) - nE\left(\bar{u}^2\right)$$

$$= \sum_{i=1}^{n} \sigma_i^2 - \frac{1}{n}E\left(\sum_{i=1}^{n} u_i^2\right) = \sum_{i=1}^{n} \sigma_i^2 - \frac{1}{n}\sum_{i=1}^{n} \sigma_i^2$$

$$= \left(\frac{n-1}{n}\right)\sum_{i=1}^{n} \sigma_i^2.$$

B.H. Baltagi, *Solutions Manual for Econometrics*, Springer Texts in Business and Economics, DOI 10.1007/978-3-642-54548-1_5,

$$\text{Hence, } E(s^2) = \frac{1}{n-2} E\left(\sum_{i=1}^{n} e_i^2\right)$$

$$= \frac{1}{n-2}\left[\frac{\sum\limits_{i=1}^{n} x_i{}^2\sigma_i^2}{\sum\limits_{i=1}^{n} x_i{}^2} + \left(\frac{n-1}{n}\right)\sum_{i=1}^{n}\sigma_i^2 - 2\,\frac{\sum\limits_{i=1}^{n} x_i{}^2\sigma_i^2}{\sum\limits_{i=1}^{n} x_i{}^2}\right]$$

$$= \frac{1}{n-2}\left[-\frac{\sum\limits_{i=1}^{n} x_i{}^2\sigma_i^2}{\sum\limits_{i=1}^{n} x_i{}^2} + \left(\frac{n-1}{n}\right)\sum_{i=1}^{n}\sigma_i^2\right].$$

Under homoskedasticity this reverts back to $E(s^2) = \sigma^2$.

5.2 *OLS Variance is Biased Under Heteroskedasticity.*

$E\left[\widehat{\text{var}}\left(\hat{\beta}_{ols}\right)\right] = E\left(\frac{s^2}{\sum\limits_{i=1}^{n} x_i{}^2}\right) = \frac{E(s^2)}{\sum\limits_{i=1}^{n} x_i{}^2}$. Using the results in problem 5.1, we

$$\text{get } E\left[\widehat{\text{var}}\left(\hat{\beta}_{ols}\right)\right] = \frac{1}{n-2}\left[-\frac{\sum\limits_{i=1}^{n} x_i{}^2\sigma_i^2}{\left(\sum\limits_{i=1}^{n} x_i{}^2\right)^2} + \left(\frac{n-1}{n}\right)\frac{\sum\limits_{i=1}^{n}\sigma_i^2}{\sum\limits_{i=1}^{n} x_i{}^2}\right]$$

$$E\left[\widehat{\text{var}}\left(\hat{\beta}_{ols}\right)\right] - \text{var}\left(\hat{\beta}_{ols}\right) = \frac{(n-1)}{n(n-2)}\frac{\sum\limits_{i=1}^{n}\sigma_i^2}{\sum\limits_{i=1}^{n} x_i{}^2} - \frac{\sum\limits_{i=1}^{n} x_i{}^2\sigma_i^2}{\left(\sum\limits_{i=1}^{n} x_i{}^2\right)^2}\left(\frac{1}{n-2} + 1\right)$$

Substituting $\sigma_i^2 = bx_i^2$ where $b > 0$, we get

$$E\left[\widehat{\text{var}}\left(\hat{\beta}_{ols}\right)\right] - \text{var}\left(\hat{\beta}_{ols}\right) = \frac{b\sum\limits_{i=1}^{n} x_i{}^2\sum\limits_{i=1}^{n} x_i{}^2 - nb\sum\limits_{i=1}^{n} x_i{}^2x_i{}^2}{n\left(\sum\limits_{i=1}^{n} x_i{}^2\right)^2}\cdot\frac{n-1}{n-2}$$

$$= -\frac{n-1}{n-2}\cdot\frac{b\sum\limits_{i=1}^{n}\left(x_i{}^2 - \sum\limits_{i=1}^{n} x_i{}^2\Big/n\right)^2}{\left(\sum\limits_{i=1}^{n} x_i{}^2\right)^2} < 0$$

$$\text{for} \quad b > 0.$$

This means that, on the average, the estimated standard error of $\hat{\beta}_{ols}$ understates the true standard error. Hence, the t-statistic reported by the regression package for H_o; $\beta = 0$ is overblown.

5.3 *Weighted Least Squares*. This is based on Kmenta (1986).

a. From the first equation in (5.11), one could solve for $\tilde{\alpha}$

$$\tilde{\alpha}\sum_{i=1}^{n}\left(1/\sigma_i^2\right) = \sum_{i=1}^{n}\left(Y_i/\sigma_i^2\right) - \tilde{\beta}\sum_{i=1}^{n}\left(X_i/\sigma_i^2\right).$$

Dividing both sides by $\sum_{i=1}^{n}\left(1/\sigma_i^2\right)$ one gets

$$\begin{aligned}\tilde{\alpha} &= \left[\sum_{i=1}^{n}\left(Y_i/\sigma_i^2\right)\Big/\sum_{i=1}^{n}\left(1/\sigma_i^2\right)\right] - \tilde{\beta}\left[\sum_{i=1}^{n}\left(X_i/\sigma_i^2\right)\Big/\sum_{i=1}^{n}\left(1/\sigma_i^2\right)\right] \\ &= \overline{Y}^* - \tilde{\beta}\overline{X}^*.\end{aligned}$$

Substituting $\tilde{\alpha}$ in the second equation of (5.11) one gets

$$\begin{aligned}\sum_{i=1}^{n} Y_iX_i/\sigma_i^2 = {} & \left(\sum_{i=1}^{n} X_i/\sigma_i^2\right)\left[\sum_{i=1}^{n}\left(Y_i/\sigma_i^2\right)\Big/\sum_{i=1}^{n}\left(1/\sigma_i^2\right)\right] \\ & - \tilde{\beta}\left[\sum_{i=1}^{n}\left(X_i/\sigma_i^2\right)\right]^2\Big/\sum_{i=1}^{n}\left(1/\sigma_i^2\right) + \tilde{\beta}\sum_{i=1}^{n}\left(X_i^2/\sigma_i^2\right).\end{aligned}$$

Multiplying both sides by $\sum_{i=1}^{n}\left(1/\sigma_i^2\right)$ and solving for $\tilde{\beta}$ one gets (5.12b)

$$\tilde{\beta} = \frac{\left[\sum_{i=1}^{n}\left(1/\sigma_i^2\right)\right]\left[\sum_{i=1}^{n}\left(Y_iX_i/\sigma_i^2\right)\right] - \left[\sum_{i=1}^{n}\left(X_i/\sigma_i^2\right)\right]\left[\sum_{i=1}^{n}\left(Y_i/\sigma_i^2\right)\right]}{\left[\sum_{i=1}^{n} X_i^2/\sigma_i^2\right]\left[\sum_{i=1}^{n}\left(1/\sigma_i^2\right)\right] - \left[\sum_{i=1}^{n}\left(X_i/\sigma_i^2\right)\right]^2}$$

b. From the regression equation $Y_i = \alpha + \beta X_i + u_i$ one can multiply by w_i^* and sum to get $\sum_{i=1}^{n} w_i^*Y_i = \alpha\sum_{i=1}^{n} w_i^* + \beta\sum_{i=1}^{n} w_i^*X_i + \sum_{i=1}^{n} w_i^*u_i$. Now divide by $\sum_{i=1}^{n} w_i^*$ and use the definitions of $\bar{Y}^*$ and $\bar{X}^*$ to get $\bar{Y}^* = \alpha + \beta\bar{X}^* + \bar{u}^*$ where $\bar{u}^* = \sum_{i=1}^{n} w_i^*u_i\Big/\sum_{i=1}^{n} w_i^*$.

Subtract this equation from the original regression equation to get $Y_i - \bar{Y}^* = \beta(X_i - \bar{X}^*) + (u_i - \bar{u}^*)$. Substitute this in the expression for $\tilde{\beta}$ in (5.12b), we get

$$\tilde{\beta} = \beta + \frac{\sum_{i=1}^{n} w_i{}^*\left(X_i - \overline{X}^*\right)\left(u_i - \bar{u}^*\right)}{\sum_{i=1}^{n} w_i{}^*\left(X_i - \overline{X}^*\right)^2} = \beta + \frac{\sum_{i=1}^{n} w_i{}^*\left(X_i - \overline{X}^*\right)u_i}{\sum_{i=1}^{n} w_i{}^*\left(X_i - \overline{X}^*\right)^2}$$

where the second equality uses the fact that

$$\sum_{i=1}^{n} w_i^*\left(X_i - \overline{X}^*\right) = \sum_{i=1}^{n} w_i^* X_i - \left(\sum_{i=1}^{n} w_i^*\right)\left(\sum_{i=1}^{n} w_i^* X_i\right) \Big/ \sum_{i=1}^{n} w_i^* = 0.$$

Therefore, $E(\tilde{\beta}) = \beta$ as expected, and

$$\mathrm{var}\left(\tilde{\beta}\right) = E\left(\tilde{\beta} - \beta\right)^2 = E\left(\frac{\sum_{i=1}^{n} w_i{}^*\left(X_i - \overline{X}^*\right)u_i}{\sum_{i=1}^{n} w_i{}^*\left(X_i - \overline{X}^*\right)^2}\right)^2$$

$$= \frac{\sum_{i=1}^{n} w_i{}^{*2}\left(X_i - \overline{X}^*\right)^2\sigma_i^2}{\left[\sum_{i=1}^{n} w_i{}^*\left(X_i - \overline{X}^*\right)^2\right]^2} = \frac{\sum_{i=1}^{n} w_i{}^*\left(X_i - \overline{X}^*\right)^2}{\left[\sum_{i=1}^{n} w_i{}^*\left(X_i - \overline{X}^*\right)^2\right]^2}$$

$$= \frac{1}{\sum_{i=1}^{n} w_i{}^*\left(X_i - \overline{X}^*\right)^2}$$

where the third equality uses the fact that the u_i's are not serially correlated and heteroskedastic and the fourth equality uses the fact that $w_i^* = \left(1/\sigma_i^2\right)$.

5.4 *Relative Efficiency of OLS Under Heteroskedasticity*

a. From Eq. (5.9) we have

$$\mathrm{var}\left(\hat{\beta}_{ols}\right) = \sum_{i=1}^{n} x_i^2\sigma_i^2 \Big/ \left(\sum_{i=1}^{n} x_i^2\right)^2 = \sigma^2 \sum_{i=1}^{n} x_i^2 X_i^\delta \Big/ \left(\sum_{i=1}^{n} x_i^2\right)^2$$

where $x_i = X_i - \bar{X}$. For $X_i = 1, 2, .., 10$ and $\delta = 0.5$, 1, 1.5 and 2. This is tabulated below.

b. From problem 5.3 part (b) we get

$$\mathrm{var}(\tilde{\beta}_{BLUE}) = \frac{1}{\sum_{i=1}^{n} w_i{}^*\left(X_i - \overline{X}^*\right)^2}$$

where $w_i^* = (1/\sigma_i^2) = (1/\sigma^2 X_i^\delta)$ and $\overline{X}^* = \sum_{i=1}^{n} w_i^* X_i \Big/ \sum_{i=1}^{n} w_i^* = \dfrac{\sum_{i=1}^{n}\left(X_i/X_i{}^\delta\right)}{\sum_{i=1}^{n}\left(1/X_i{}^\delta\right)}$.

For $X_i = 1, 2, .., 10$ and $\delta = 0.5, 1, 1.5$ and 2, this variance is tabulated below.

c. The relative efficiency of OLS is given by the ratio of the two variances computed in parts (a) and (b). This is tabulated below for various values of δ.

$\boldsymbol{\delta}$	**var**$(\hat{\boldsymbol{\beta}}_{\mathbf{ols}})$	**var**$(\tilde{\boldsymbol{\beta}}_{BLUE})$	**var**$(\tilde{\boldsymbol{\beta}}_{BLUE})/$**var**$(\hat{\boldsymbol{\beta}}_{\mathbf{ols}})$
0.5	$0.0262\ \sigma^2$	$0.0237\ \sigma^2$	0.905
1	$0.0667\ \sigma^2$	$0.04794\ \sigma^2$	0.719
1.5	$0.1860\ \sigma^2$	$0.1017\ \sigma^2$	0.547
2	$0.5442\ \sigma^2$	$0.224\ \sigma^2$	0.412

As δ increases from 0.5 to 2, the relative efficiency of OLS with respect to BLUE decreases from 0.9 for mild heteroskedasticity to 0.4 for more serious heteroskedasticity.

5.6 *The AR(1) model.* From (5.26), by continuous substitution just like (5.29), one could stop at u_{t-s} to get

$$u_t = \rho^s u_{t-s} + \rho^{s-1}\varepsilon_{t-s+1} + \rho^{s-2}\varepsilon_{t-s+2} + .. + \rho\varepsilon_{t-1} + \varepsilon_t \qquad \text{for } t > s.$$

Note that the power of ρ and the subscript of ε always sum to t. Multiplying both sides by u_{t-s} and taking expected value, one gets

$$E\left(u_t u_{t-s}\right) = \rho^s E\left(u_{t-s}^2\right) + \rho^{s-1} E\left(\varepsilon_{t-s+1} u_{t-s}\right) + .. + \rho E\left(\varepsilon_{t-1} u_{t-s}\right) + E\left(\varepsilon_t u_{t-s}\right)$$

using (5.29), u_{t-s} is a function of ε_{t-s}, past values of ε_{t-s} and u_o. Since u_o is independent of the ε's, and the ε's themselves are not serially correlated, then u_{t-s} is independent of $\varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_{t-s+1}$. Hence, all the terms on the right hand side of $E(u_t u_{t-s})$ except the first are zero. Therefore, $\text{cov}(u_t, u_{t-s}) = E(u_t u_{t-s}) = \rho^s \sigma_u^2$ for $t > s$.

5.7 *Relative Efficiency of OLS Under the AR(1) Model.*

a. $\hat{\beta}_{ols} = \sum_{t=1}^{T} x_t y_t \Big/ \sum_{t=1}^{T} x_t^2 = \beta + \sum_{t=1}^{T} x_t u_t \Big/ \sum_{t=1}^{T} x_t^2$ with $E(\hat{\beta}_{ols}) = \beta$ since x_t and u_t are independent. Also,

$$\text{var}\left(\hat{\beta}_{ols}\right) = E\left(\hat{\beta}_{ols} - \beta\right)^2 = E\left(\sum_{t=1}^{T} x_t u_t \Big/ \sum_{t=1}^{T} x_t^2\right)^2 = \sum_{t=1}^{T} x_t^2 E\left(u_t^2\right) \Big/ \left(\sum_{t=1}^{T} x_t^2\right)^2$$

$$+ E\left[\sum\sum_{s \neq t} x_t x_s u_t u_s \Big/ \left(\sum_{t=1}^{T} x_t^2\right)^2\right]$$

$$= \frac{\sigma_u^2}{\sum_{t=1}^{T} x_t^2}\left[1 + 2\rho \frac{\sum_{t=1}^{T-1} x_t x_{t+1}}{\sum_{t=1}^{T} x_t^2} + 2\rho^2 \frac{\sum_{t=1}^{T-2} x_t x_{t+2}}{\sum_{t=1}^{T} x_t^2} + .. + 2\rho^{T-1} \frac{x_1 x_T}{\sum_{t=1}^{T} x_t^2}\right]$$

using the fact that $E(u_t u_s) = \rho^{|t-s|}\sigma_u^2$ as shown in problem 6.

Alternatively, one can use matrix algebra, see Chap. 9. For the AR(1) model, $\Omega = E(uu')$ is given by Eq. (9.9) of Chap. 9. So

$$x'\Omega x = (x_1, \ldots, x_T)\begin{bmatrix} 1 & \rho & \rho^2 & \ldots & \rho^{T-1} \\ \rho & 1 & \rho & \ldots & \rho^{T-2} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \ldots & 1 \end{bmatrix}\begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_T \end{pmatrix}$$

$$= \left(x_1^2 + \rho x_1 x_2 + .. + \rho^{T-1} x_T x_1\right) + \left(\rho x_1 x_2 + x_2^2 + .. + \rho^{T-2} x_T x_2\right)$$

$$+ \left(\rho^2 x_1 x_3 + \rho x_2 x_3 + .. + \rho^{T-3} x_T x_3\right) + ..$$

$$+ \left(\rho^{T-1} x_1 x_T + \rho^{T-2} x_2 x_T + .. + x_T^2\right)$$

collecting terms, we get

$$x'\Omega x = \sum_{t=1}^{T} x_t^2 + 2\rho \sum_{t=1}^{T-1} x_t x_{t+1} + 2\rho^2 \sum_{t=1}^{T-2} x_t x_{t+2} + .. + 2\rho^{T-1} x_1 x_T$$

Using Eq. (9.5) of Chap. 9, we get $\mathrm{var}(\hat{\beta}_{ols}) = \sigma_u^2 \frac{x'\Omega x}{(x'x)^2}$

$$= \frac{\sigma_u^2}{\sum_{t=1}^{T} x_t^2}\left[1 + 2\rho \frac{\sum_{t=1}^{T-1} x_t x_{t+1}}{\sum_{t=1}^{T} x_t^2} + 2\rho^2 \frac{\sum_{t=1}^{T-2} x_t x_{t+2}}{\sum_{t=1}^{T} x_t^2} + .. + 2\rho^{T-1} \frac{x_1 x_T}{\sum_{t=1}^{T} x_t^2}\right].$$

Similarly, Ω^{-1} for the AR(1) model is given by Eq. (9.10) of Chap. 9, and $\mathrm{var}\left(\hat{\beta}_{BLUE}\right) = \sigma^2(X'\Omega^{-1}X)^{-1}$ below equation (9.4). In this case,

$$x'\Omega^{-1}x = \frac{1}{1-\rho^2}(x_1, \ldots, x_T)\begin{bmatrix} 1 & -\rho & 0 & \ldots & 0 & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & \ldots & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \ldots & -\rho & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & \ldots & 0 & -\rho & 1 \end{bmatrix}\begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_T \end{pmatrix}$$

$$x'\Omega^{-1}x = \frac{1}{1-\rho^2}\left[x_1^2 - \rho x_1 x_2 - \rho x_1 x_2 + (1+\rho^2)x_2^2 - \rho x_3 x_2 - \rho x_2 x_3 \right.$$
$$\left. + (1+\rho^2)x_3^2 - \rho x_4 x_3 + .. + x_T^2 - \rho x_{T-1} x_T\right].$$

Collecting terms, we get

$$x'\Omega^{-1}x = \frac{1}{1-\rho^2}\left[\sum_{t=1}^{T} x_t^2 - 2\rho \sum_{t=1}^{T-1} x_t x_{t+1} + \rho^2 \sum_{t=1}^{T-1} x_t^2\right]$$

For $T \to \infty$, $\sum_{t=2}^{T-1} x_t^2$ is equivalent to $\sum_{t=1}^{T} x_t^2$, hence

$$\mathrm{var}\left(\hat{\beta}_{PW}\right) = \sigma_u^2(x'\Omega^{-1}x)^{-1} = \frac{\sigma_u^2}{\sum_{t=1}^{T} x_t^2}\left[\frac{1-\rho^2}{1+\rho^2 - 2\rho \sum_{t=1}^{T-1} x_t x_{t+1} \Big/ \sum_{t=1}^{T} x_t^2}\right].$$

b. For x_t following itself an AR(1) process: $x_t = \lambda x_{t-1} + v_t$, we know that λ is the correlation of x_t and x_{t-1}, and from problem 5.6, $\mathrm{correl}(x_t, x_{t-s}) = \lambda^s$. As $T \to \infty$, λ is estimated well by $\sum_{t=1}^{T-1} x_t x_{t+1} \Big/ \sum_{t=1}^{T} x_t^2$. Also, λ^2 is estimated well by $\sum_{t=1}^{T-2} x_t x_{t+2} \Big/ \sum_{t=1}^{T} x_t^2$, etc. Hence

$$\text{asy eff}\left(\hat{\beta}_{ols}\right) = \lim_{T\to\infty} \frac{\text{var}\left(\hat{\beta}_{PW}\right)}{\text{var}\left(\hat{\beta}_{ols}\right)} = \frac{1-\rho^2}{(1+\rho^2-2\rho\lambda)\,(1+2\rho\lambda+2\rho^2\lambda^2+..)}$$

$$= \frac{(1-\rho^2)(1-\rho\lambda)}{(1+\rho^2-2\rho\lambda)(1+\rho\lambda)}$$

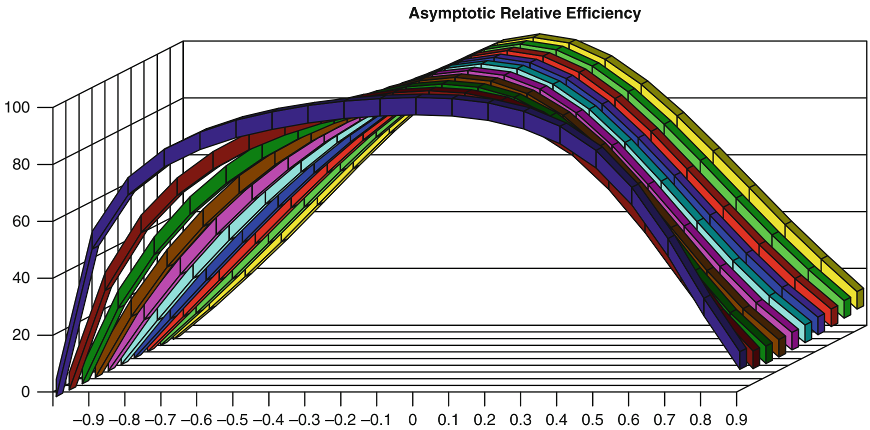
where the last equality uses the fact that $(1+\rho\lambda)/(1-\rho\lambda) = (1+2\rho\lambda+2\rho^2\lambda^2+..)$. For $\lambda = 0$, or $\rho = \lambda$, this asy eff$(\hat{\beta}_{ols})$ is equal to $(1-\rho^2)/(1+\rho^2)$.

c. The asy eff$(\hat{\beta}_{ols})$ derived in part (b) is tabulated below for various values of ρ and λ. A similar table is given in Johnston (1984, p. 312). For $\rho > 0$, loss in efficiency is big as ρ increases. For a fixed λ, this asymptotic efficiency drops from the 90 to a 10% range as ρ increases from 0.2 to 0.9. Variation in λ has minor effects when $\rho > 0$. For $\rho < 0$, the efficiency loss is still big as the absolute value of ρ increases, for a fixed λ. However, now variation in λ has a much stronger effect. For a fixed negative ρ, the loss in efficiency decreases with λ. In fact, for $\lambda = 0.9$, the loss in efficiency drops from 99% to 53 as ρ goes from -0.2 to -0.9. This is in contrast to say $\lambda = 0.2$ where the loss in efficiency drops from 93 to 13% as ρ goes from -0.2 to -0.9.

(Asymptotic Relative Efficiency of $\hat{\beta}_{ols}$) x 100

ρ

λ	−0.9	−0.8	−0.7	−0.6	−0.5	−0.4	−0.3	−0.2	−0.1	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0	10.5	22.0	34.2	47.1	60.0	72.4	83.5	92.3	98.0	100	98.0	92.3	83.5	72.4	60.0	47.1	34.2	22.0	10.5
0.1	11.4	23.5	36.0	48.8	61.4	73.4	84.0	92.5	98.1	100	98.0	92.2	83.2	71.8	59.0	45.8	32.8	20.7	9.7
0.2	12.6	25.4	38.2	50.9	63.2	74.7	84.8	92.9	98.1	100	98.1	92.3	83.2	71.6	58.4	44.9	31.8	19.8	9.1
0.3	14.1	27.7	40.9	53.5	65.5	76.4	85.8	93.3	98.2	100	98.1	92.5	83.5	71.7	58.4	44.5	31.1	19.0	8.6
0.4	16.0	30.7	44.2	56.8	68.2	78.4	87.1	93.9	98.4	100	98.3	92.9	84.1	72.4	58.8	44.6	30.8	18.5	8.2
0.5	18.5	34.4	48.4	60.6	71.4	80.8	88.6	94.6	98.6	100	98.4	93.5	85.1	73.7	60.0	45.3	31.1	18.4	7.9
0.6	22.0	39.4	53.6	65.4	75.3	83.6	90.3	95.5	98.8	100	98.6	94.3	86.6	75.7	62.1	47.1	32.0	18.6	7.8
0.7	27.3	46.2	60.3	71.2	79.9	86.8	92.3	96.4	99.0	100	98.9	95.3	88.7	78.8	65.7	50.3	34.2	19.5	7.8
0.8	35.9	56.2	69.3	78.5	85.4	90.6	94.6	97.5	99.3	100	99.2	96.6	91.4	83.2	71.4	56.2	38.9	22.0	8.4
0.9	52.8	71.8	81.7	87.8	92.0	94.9	97.1	98.7	99.6	100	99.6	98.1	95.1	89.8	81.3	68.3	50.3	29.3	10.5



d. Ignoring autocorrelation, $s^2 \Big/ \sum_{t=1}^{T} x_t^2$ estimates $\sigma_u^2 \Big/ \sum_{t=1}^{T} x_t^2$, but

$$\text{asy.var}\left(\hat{\beta}_{ols}\right) = \left(\sigma_u^2 \Big/ \sum_{t=1}^{T} x_t^2\right)(1-\rho\lambda)/(1+\rho\lambda)$$

so the asy.bias in estimating the var$(\hat{\beta}_{ols})$ is

$$\text{asy.var}\left(\hat{\beta}_{ols}\right) - \sigma_u^2 \Big/ \sum_{t=1}^{T} {x_t}^2 = \frac{\sigma_u^2}{\sum_{t=1}^{T} {x_t}^2}\left[\frac{1-\rho\lambda}{1+\rho\lambda} - 1\right] = \frac{\sigma_u^2}{\sum_{t=1}^{T} {x_t}^2}\left[\frac{-2\rho\lambda}{1+\rho\lambda}\right]$$

and asy.proportionate bias $= -2\rho\lambda/(1+\rho\lambda)$.

Percentage Bias in estimating var($\hat{\beta}_{ols}$)

ρ

λ	−0.9	−0.8	−0.7	−0.6	−0.5	−0.4	−0.3	−0.2	0	0.1	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.1	19.8	17.4	15.1	12.8	10.5	8.3	6.2	4.1	0	–2.0	–5.8	–7.7	–9.5	–11.3	–13.1	–14.8	–16.5
0.2	43.9	38.1	32.6	27.3	22.2	17.4	12.8	8.3	0	–3.9	–11.3	–14.8	–18.2	–21.4	–24.6	–27.6	–30.5
0.3	74.0	63.2	53.2	43.9	35.3	27.3	19.8	12.8	0	–5.8	–16.5	–21.4	–26.1	–30.5	–34.7	–38.7	–42.5
0.4	112.5	94.1	77.8	63.2	50.0	38.1	27.3	17.4	0	–7.7	–21.4	–27.6	–33.3	–38.7	–43.8	–48.5	–52.9
0.5	163.6	133.3	107.7	85.7	66.7	50.0	35.3	22.2	0	–9.5	–26.1	–33.3	–40.0	–46.2	–51.9	–57.1	–62.1
0.6	234.8	184.6	144.8	112.5	85.7	63.2	43.9	27.3	0	–11.3	–30.5	–38.7	–46.2	–52.9	–59.2	–64.9	–70.1
0.7	340.5	254.5	192.2	144.8	107.7	77.8	53.2	32.6	0	–13.1	–34.7	–43.8	–51.9	–59.2	–65.8	–71.8	–77.3
0.8	514.3	355.6	254.5	184.6	133.3	94.1	63.2	38.1	0	–14.8	–38.7	–48.5	–57.1	–64.9	–71.8	–78.0	–83.7
0.9	852.6	514.3	340.5	234.8	163.6	112.5	74.0	43.9	0	–16.5	–42.5	–52.9	–62.1	–70.1	–77.3	–83.7	–89.5

This is tabulated for various values of ρ and λ. A similar table is given in Johnston (1984, p. 312).

For ρ and λ positive, var $\left(\hat{\beta}_{ols}\right)$ is underestimated by the conventional formula. For $\rho = \lambda = 0.9$, this underestimation is almost 90%. For $\rho < 0$, the var($\hat{\beta}_{ols}$) is overestimated by the conventional formula. For $\rho = -0.9$ and $\lambda = 0.9$, this overestimation is of magnitude 853%.

e. $e_t = y_t - \hat{y}_t = (\beta - \hat{\beta}_{ols})x_t + u_t$

Hence, $\sum_{t=1}^{T} e_t^2 = \left(\hat{\beta}_{ols} - \beta\right)^2 \sum_{t=1}^{T} x_t^2 + \sum_{t=1}^{T} u_t^2 - 2\left(\hat{\beta}_{ols} - \beta\right) \sum_{t=1}^{T} x_t u_t$ and

$$E\left(\sum_{t=1}^{T} e_t^2\right) = \sum_{t=1}^{T} x_t^2 \operatorname{var}\left(\hat{\beta}_{ols}\right) + T\sigma_u^2 - 2E\left(\sum_{s=1}^{T} x_s u_s\right)\left(\sum_{t=1}^{T} x_t u_t\right) \Big/ \left(\sum_{t=1}^{T} x_t^2\right)$$

$$= -\sum_{t=1}^{T} x_t^2 \operatorname{var}\left(\hat{\beta}_{ols}\right) + T\sigma_u^2$$

$$= -\sigma_u^2 \left[1 + 2\rho \frac{\sum_{t=1}^{T-1} x_t x_{t+1}}{\sum_{t=1}^{T} x_t^2} + 2\rho^2 \frac{\sum_{t=1}^{T-2} x_t x_{t+2}}{\sum_{t=1}^{T} x_t^2} + .. + 2\rho^{T-1} \frac{x_1 x_T}{\sum_{t=1}^{T} x_t^2}\right] + T\sigma_u^2$$

So that $E(s^2) = E\left(\sum_{t=1}^{T} e_t^2/(T-1)\right)$

$$= \sigma_u^2 \left\{ T - \left(1 + 2\rho \frac{\sum_{t=1}^{T-1} x_t x_{t+1}}{\sum_{t=1}^{T} x_t^2} + 2\rho^2 \frac{\sum_{t=1}^{T-2} x_t x_{t+2}}{\sum_{t=1}^{T} x_t^2} + .. + 2\rho^{T-1} \frac{x_1 x_T}{\sum_{t=1}^{T} x_t^2} \right) \right\} \Big/ (T-1)$$

If $\rho = 0$, then $E(s^2) = \sigma_u^2$. If x_t follows an AR(1) model with parameter λ, then for large T, $E(s^2) = \sigma_u^2 \left(T - \frac{1+\rho\lambda}{1-\rho\lambda}\right) / (T-1)$.
For $T = 101$, $E(s^2) = \sigma_u^2 \left(101 - \frac{1+\rho\lambda}{1-\rho\lambda}\right) / 100$.
This can be tabulated for various values of ρ and λ. For example, when $\rho = \lambda = 0.9$, $E(s^2) = 0.915\, \sigma_u^2$.

5.10 *Regressions with Non-zero Mean Disturbances.*

a. For the gamma distribution, $E(u_i) = \theta$ and $var(u_i) = \theta$. Hence, the disturbances of the simple regression have non-zero mean but constant variance. Adding and subtracting θ on the right hand side of the regression we get $Y_i = (\alpha + \theta) + \beta X_i + (u_i - \theta) = \alpha^* + \beta X_i + u_i^*$ where $\alpha^* = \alpha + \theta$ and $u_i^* = u_i - \theta$ with $E\,(u_i^*) = 0$ and $var\,(u_i^*) = \theta$. OLS yields the BLU estimators of α^* and β since the u_i^* disturbances satisfy all the requirements of the Gauss–Markov Theorem, including the zero mean assumption. Hence, $E(\hat{\alpha}_{ols}) = \alpha^* = \alpha + \theta$ which is biased for α by the mean of the disturbances θ. But $E(s^2) = var\left(u_i^*\right) = var\,(u_i) = \theta$. Therefore,

$$E\left(\hat{\alpha}_{ols} - s^2\right) = E\left(\hat{\alpha}_{ols}\right) - E(s^2) = \alpha + \theta - \theta = \alpha.$$

b. Similarly, for the χ_ν^2 distribution, we have $E(u_i) = \nu$ and $var(u_i) = 2\nu$. In this case, the same analysis applies except that $E(\hat{\alpha}_{ols}) = \alpha + \nu$ and $E(s^2) = 2\nu$. Hence, $E(\hat{\alpha}_{ols} - s^2/2) = \alpha$.

c. Finally, for the exponential distribution, we have $E(u_i) = \theta$ and $var(u_i) = \theta^2$. In this case, plim $\hat{\alpha}_{ols} = \alpha + \theta$ and plim $s^2 = \theta^2$. Hence, plim $(\hat{\alpha}_{ols} - s) =$ plim $\hat{\alpha}_{ols} -$ plim $s = \alpha$.

5.11 *The Heteroskedastic Consequences of an Arbitrary Variance for the Initial Disturbance of an AR(1) Model.* Parts (a), (b), and (c) are based on the solution given by Kim (1991).

a. Continual substitution into the process $u_t = \rho u_{t-1} + \varepsilon_t$ yields

$$u_t = \rho^t u_o + \sum_{j=0}^{t-1} \rho^j \varepsilon_{t-j}.$$

Exploiting the independence conditions, $E(u_o \varepsilon_j) = 0$ for all j, and $E(\varepsilon_i \varepsilon_j) = 0$ for all $i \neq j$, the variance of u_t is

$$\sigma_t^2 = \text{var}(u_t) = \rho^{2t}\,\text{var}(u_o) + \sum_{j=0}^{t-1} \rho^{2j}\,\text{var}\left(\varepsilon_{t-j}\right) = \rho^{2t}\sigma_\varepsilon^2/\tau + \sigma_\varepsilon^2 \sum_{j=0}^{t-1} \rho^{2j}$$

$$= \rho^{2t}\sigma_\varepsilon^2/\tau + \sigma_\varepsilon^2[(1-\rho^{2t})/(1-\rho^2)]$$

$$= \left[\left\{1/(1-\rho^2) + [1/\tau - 1/(1-\rho^2)]\rho^{2t}\right\}\sigma_\varepsilon^2\right. .$$

This expression for σ_t^2 depends on t. Hence, for an arbitrary value of τ, the disturbances are rendered heteroskedastic. If $\tau = 1 - \rho^2$ then σ_t^2 does not depend on t and σ_t^2 becomes homoskedastic.

b. Using the above equation for σ_t^2, define

$$a = \sigma_t^2 - \sigma_{t-1}^2 = \left[(1/\tau) - 1/(1-\rho^2)\right]\rho^{2t}(1-\rho^{-2})\sigma_\varepsilon^2 = bc$$

where $b = (1/\tau) - 1/(1-\rho^2)$ and $c = \rho^{2t}(1-\rho^{-2})\sigma_\varepsilon^2$. Note that $c < 0$. If $\tau > 1 - \rho^2$, then $b < 0$ implying that $a > 0$ and the variance is increasing. While if $\tau < 1 - \rho^2$, then $b > 0$ implying that $a < 0$ and the variance is decreasing. If $\tau = 1 - \rho^2$ then $a = 0$ and the process becomes homoskedastic.

Alternatively, one can show that $\dfrac{\partial \sigma_t^2}{\partial t} = \dfrac{2\sigma_\varepsilon^2 \rho^{2t} \log \rho}{\tau(1-\rho^2)}[(1-\rho^2) - \tau]$ where $\partial a^{bx}/\partial x = ba^{bx} \log a$, has been used. Since $\log \rho < 0$, the right hand side of the above equation is positive (σ_t^2 is increasing) if $\tau > 1 - \rho^2$, and is negative otherwise.

c. Using $u_t = \rho^t u_o + \sum_{j=0}^{t-1} \rho^j \varepsilon_{t-j}$, and noting that $E(u_t) = 0$ for all t, finds, for $t \geq s$,

$$\mathrm{cov}(u_t,\ u_{t-s}) = E(u_t u_{t-s}) = \left(\rho^t u_o + \sum_{i=0}^{t-1} \rho^i \varepsilon_{t-i}\right)\left(\rho^{t-s} u_o + \sum_{j=0}^{t-s-1} \rho^j \varepsilon_{t-s-j}\right)$$

$$= \rho^{2t-s}\mathrm{var}(u_o) + \rho^s \sum_{i=0}^{t-s-1} \rho^{2i}\,\mathrm{var}(\varepsilon_{t-s-i})$$

$$= \rho^{2t-s}\sigma_\varepsilon^2/\tau + \sigma_\varepsilon^2\left[\rho^s(1-\rho^{2(t-s)})/(1-\rho^2)\right]$$

$$= \rho^s\left\{1/(1-\rho^2) + \left[(1/\tau) - 1/(1-\rho^2)\right]\rho^{2(t-s)}\right\}\sigma_\varepsilon^2 = \rho^s\sigma_{t-s}^2$$

d. *The Bias of the Standard Errors of OLS Process with an Arbitrary Variance on the Initial Observation*. This is based on the solution by Koning (1992). Consider the time-series model

$$y_t = \beta x_t + u_t$$

$$u_t = \rho u_{t-1} + \varepsilon_t$$

with $0 < \rho < 1,\ \varepsilon_t \sim \mathrm{IIN}\left(0, \sigma_\varepsilon^2\right),\ u_o \sim \mathrm{IIN}\left(0, \left(\sigma_\varepsilon^2/\tau\right)\right),\ \tau > 0$. The x's are positively autocorrelated. Note that

$$\mathrm{var}(u_t) = \sigma_t^2 = \sigma_\varepsilon^2\left(\frac{\rho^{2t}}{\tau} + \sum_{s=1}^{t-1}\rho^{2s}\right) = \sigma_\varepsilon^2\left(\frac{\rho^{2t}}{\tau} + \frac{1-\rho^{2t}}{1-\rho^2}\right).$$

The population variance of the OLS estimator for β is

$$\mathrm{var}\left(\hat{\beta}_{ols}\right) = \frac{\sum_{t=1}^{T}\sum_{s=1}^{T} x_s x_t \mathrm{cov}(u_t, u_s)}{\left(\sum_{t=1}^{T} x_t^2\right)^2} = \frac{\sum_{t=1}^{T}\sigma_t^2 x_t^2 + 2\sum_{t=1}^{T}\sum_{s>t}^{T} x_s x_t \rho^{s-t}\sigma_t^2}{\left(\sum_{t=1}^{T} x_t^2\right)^2}.$$

Also, note that in the stationary case ($\tau = 1-\rho^2$), we have:

$$\mathrm{var}\left(\hat{\beta}_{ols}\right) = \frac{\sigma_\varepsilon^2\Big/(1-\rho^2)}{\sum_{t=1}^{T} x_t^2} + \frac{(2\sigma_\varepsilon^2/(1-\rho^2))\sum_{t=1}^{T}\sum_{s>t}^{T} x_s x_t \rho^{s-t}}{\left(\sum_{t=1}^{T} x_t^2\right)^2} \geq \frac{\sigma_\varepsilon^2/(1-\rho^2)}{\sum_{t=1}^{T} x_t^2},$$

since each term in the double summation is non-negative. Hence, the true variance of $\hat{\beta}_{ols}$ (the left hand side of the last equation) is greater than the estimated variance of $\hat{\beta}_{ols}$ (the right hand side of the last equation). Therefore, the true t-ratio corresponding to $H_o;\ \beta = 0$ versus $H_1;\ \beta \neq 0$ is lower than the estimated t-ratio. In other words, OLS rejects too often.

From the equation for σ_t^2 it is easily seen that $\partial\sigma_t^2/\partial\tau < 0$, and hence, that the true variance of $\hat{\beta}_{ols}$ is larger than the estimated $\text{var}(\hat{\beta}_{ols})$ if $\tau < 1 - \rho^2$. Hence, the true t-ratio decreases and compared to the stationary case, OLS rejects more often. The opposite holds for $\tau > 1 - \rho^2$.

5.12 *ML Estimation of Linear Regression Model with AR(1) Errors and Two Observations*. This is based on Baltagi and Li (1995).

a. The OLS estimator of β is given by

$$\hat{\beta}_{ols} = \sum_{i=1}^{2} x_i y_i \Big/ \sum_{i=1}^{2} x_i^2 = (y_1x_1 + y_2x_2)/\left(x_1^2 + x_2^2\right).$$

b. The log-likelihood function is given by $\log L = -\log 2\pi - \log\sigma^2 - (1/2)\log(1-\rho^2) - \left(u_2^2 - 2\rho u_1u_2 + u_1^2\right)/2\sigma^2(1-\rho^2)$; setting $\partial\log L/\partial\sigma^2 = 0$ gives $\hat{\sigma}^2 = \left(u_2^2 - 2\rho u_1u_2 + u_1^2\right)/2(1-\rho^2)$; setting $\partial\log L/\partial\rho = 0$ gives $\rho\sigma^2(1-\rho^2) + u_1u_2 + \rho^2u_1u_2 - \rho u_2^2 - \rho u_1^2 = 0$; substituting $\hat{\sigma}^2$ in this last equation, we get $\hat{\rho} = 2u_1u_2/\left(u_1^2 + u_2^2\right)$; setting $\partial\log L/\partial\beta = 0$ gives $u_2(x_2 - \rho x_1) + u_1(x_1 - \rho x_2) = 0$; substituting $\hat{\rho}$ in this last equation, we get $(u_1x_1 - u_2x_2)\left(u_1^2 - u_2^2\right) = 0$. Note that $u_1^2 = u_2^2$ implies a $\hat{\rho}$ of ± 1, and this is ruled out by the stationarity of the AR(1) process. Solving $(u_1x_1 - u_2x_2) = 0$ gives the required MLE of β:

$$\hat{\beta}_{mle} = (y_1x_1 - y_2x_2)/\left(x_1^2 - x_2^2\right).$$

c. By substituting $\hat{\beta}_{mle}$ into u_1 and u_2, one gets $\hat{u}_1 = x_2(x_1y_2 - x_2y_1)/\left(x_1^2 - x_2^2\right)$ and $\hat{u}_2 = x_1(x_1y_2 - x_2y_1)/\left(x_1^2 - x_2^2\right)$, which upon substitution in the expression for $\hat{\rho}$ give the required estimate of ρ : $\hat{\rho} = 2x_1x_2/\left(x_1^2 + x_2^2\right)$.

d. If $x_1 \to x_2$, with $x_2 \neq 0$, then $\hat{\rho} \to 1$. For $\hat{\beta}_{mle}$, we distinguish between two cases.

(i) For $y_1 = y_2$, $\hat{\beta}_{mle} \to y_2/(2x_2)$, which is half the limit of $\hat{\beta}_{ols} \to y_2/x_2$. The latter is the slope of the line connecting the origin to the observation (x_2, y_2).

(ii) For $y_1 \neq y_2$, $\hat{\beta}_{mle} \to \pm\infty$, with the sign depending on the sign of x_2, $(y_1 - y_2)$, and the direction from which x_1 approaches x_2. In this case, $\hat{\beta}_{ols} \to \bar{y}/x_2$, where $\bar{y} = (y_1 + y_2)/2$. This is the slope of the line connecting the origin to $(x_2, \bar{y})$.

Similarly, if $x_1 \to -x_2$, with $x_2 \neq 0$, then $\hat{\rho} \to -1$. For $\hat{\beta}_{mle}$, we distinguish between two cases:

(i) For $y_1 = -y_2$, $\hat{\beta}_{mle} \to y_2/(2x_2)$, which is half the limit of $\hat{\beta}_{ols} \to y_2/x_2$.

(ii) For $y_1 \neq -y_2$, $\hat{\beta}_{mle} \to \pm\infty$, with the sign depending on the sign of x_2, $(y_1 + y_2)$, and the direction from which x_1 approaches $-x_2$. In this case, $\hat{\beta}_{ols} \to (y_2 - y_1)/2x_2 = (y_2 - y_1)/(x_2 - x_1)$, which is the standard formula for the slope of a straight line based on two observations. In conclusion, $\hat{\beta}_{ols}$ is a more reasonable estimator of β than $\hat{\beta}_{mle}$ for this two-observation example.

5.13 The backup regressions are given below: These are performed using SAS.

OLS REGRESSION OF LNC ON CONSTANT, LNP, AND LNY

Dependent Variable: LNC

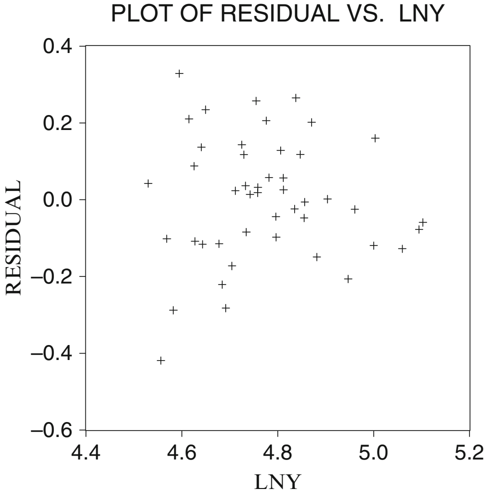
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	0.50098	0.25049	9.378	0.0004
Error	43	1.14854	0.02671		
C Total	45	1.64953			

Root MSE	0.16343	R-square	0.3037
Dep Mean	4.84784	Adj R-sq	0.2713
C.V.	3.37125		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	4.299662	0.90892571	4.730	0.0001
LNP	1	-1.338335	0.32460147	-4.123	0.0002
LNY	1	0.172386	0.19675440	0.876	0.3858



b. Regression for Glejser Test (1969)

Dependent Variable: ABS_E

MODEL: Z1=LNY^(−1)

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	0.04501	0.04501	5.300	0.0261
Error	44	0.37364	0.00849		
C Total	45	0.41865			

Root MSE	0.09215	R-square	0.1075
Dep Mean	0.12597	Adj R-sq	0.0872
C.V.	73.15601		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>F
INTERCEP	1	-0.948925	0.46709261	-2.032	0.0483
Z1(LNY^ -1)	1	5.128691	2.22772532	2.302	0.0261

MODEL: Z2=LNY^(-0.5)

Dependent Variable: ABS_E

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	0.04447	0.04447	5.229	0.0271
Error	44	0.37418	0.00850		
C Total	45	0.41865			

Root MSE	0.09222	R-square	0.1062
Dep Mean	0.12597	Adj R-sq	0.0859
C.V.	73.20853		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	-2.004483	0.93172690	-2.151	0.0370
Z2 (LNY^ $-.5$)	1	4.654129	2.03521298	2.287	0.0271

MODEL: Z3=LNY^(0.5)

Dependent Variable: ABS_E

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	0.04339	0.04339	5.087	0.0291
Error	44	0.37526	0.00853		
C Total	45	0.41865			

Root MSE	0.09235	R-square	0.1036
Dep Mean	0.12597	Adj R-sq	0.0833
C.V.	73.31455		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	2.217240	0.92729847	2.391	0.0211
Z3 (LNY^.5)	1	-0.957085	0.42433823	-2.255	0.0291

MODEL: Z4=LNY^1

Dependent Variable: ABS_E

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	0.04284	0.04284	5.016	0.0302
Error	44	0.37581	0.00854		
C Total	45	0.41865			

Root MSE	0.09242	R-square	0.1023
Dep Mean	0.12597	Adj R-sq	0.0819
C.V.	73.36798		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	1.161694	0.46266689	2.511	0.0158
Z4 (LNY^1)	1	-0.216886	0.09684233	-2.240	0.0302

c. Regression for Goldfeld and Quandt Test (1965) with first 17 obervations

Dependent Variable: LNC

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	0.22975	0.11488	2.354	0.1315
Error	14	0.68330	0.04881		
C Total	16	0.91305			

Root MSE	0.22092	R-square	0.2516
Dep Mean	4.85806	Adj R-sq	0.1447
C.V.	4.54756		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	3.983911	4.51225092	0.883	0.3922
LNP	1	-1.817254	0.86970957	-2.089	0.0554
LNY	1	0.248409	0.96827122	0.257	0.8013

REGRESSION FOR GOLDFELD AND QUANDT TEST (1965) w/last 17 obs

Dependent Variable: LNC

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	0.17042	0.08521	5.482	0.0174
Error	14	0.21760	0.01554		
C Total	16	0.38803			

Root MSE	0.12467	R-square	0.4392
Dep Mean	4.78796	Adj R-sq	0.3591
C.V.	2.60387		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	6.912881	1.57090447	4.401	0.0006
LNP	1	-1.248584	0.39773565	-3.139	0.0072
LNY	1	-0.363625	0.31771223	-1.145	0.2716

d. Data for Spearman Rank Correlation Test

OBS	STATE	LNC	RANKY	ABS_E	RANKE	D
1	MS	4.93990	1	0.04196	11	10
2	UT	4.40859	2	0.41867	46	44
3	WV	4.82454	3	0.10198	21	18
4	NM	4.58107	4	0.28820	44	40
5	AR	5.10709	5	0.32868	45	40
6	LA	4.98602	6	0.21014	38	32
7	SC	5.07801	7	0.08730	19	12
8	OK	4.72720	8	0.10845	22	14
9	AL	4.96213	9	0.13671	30	21
10	ID	4.74902	10	0.11628	24	14

11	KY	5.37906	11	0.23428	40	29
12	SD	4.81545	12	0.11470	23	11
13	AZ	4.66312	13	0.22128	39	26
14	ND	4.58237	14	0.28253	43	29
15	MT	4.73313	15	0.17266	34	19
16	WY	5.00087	16	0.02320	5	-11
17	TN	5.04939	17	0.14323	31	14
18	IN	5.11129	18	0.11673	25	7
19	GA	4.97974	19	0.03583	10	-9
20	TX	4.65398	20	0.08446	18	-2
21	IA	4.80857	21	0.01372	3	-18
22	ME	4.98722	22	0.25740	41	19
23	WI	4.83026	23	0.01754	4	-19
24	OH	4.97952	24	0.03201	9	-15
25	VT	5.08799	25	0.20619	36	11
26	MO	5.06430	26	0.05716	15	-11
27	KS	4.79263	27	0.04417	12	-15
28	NE	4.77558	28	0.09793	20	-8
29	MI	4.94744	29	0.12797	28	-1
30	FL	4.80081	30	0.05625	14	-16
31	MN	4.69589	31	0.02570	8	-23
32	PA	4.80363	32	0.02462	6	-26
33	NV	4.96642	33	0.26506	42	9
34	RI	4.84693	34	0.11760	26	-8
35	VA	4.93065	35	0.04776	13	-22
36	WA	4.66134	36	0.00638	2	-34
37	DE	5.04705	37	0.20120	35	-2
38	CA	4.50449	38	0.14953	32	-6
39	IL	4.81445	39	0.00142	1	-38
40	MD	4.77751	40	0.20664	37	-3
41	NY	4.66496	41	0.02545	7	-34
42	MA	4.73877	42	0.12018	27	-15
43	NH	5.10990	43	0.15991	33	-10
44	DC	4.65637	44	0.12810	29	-15
45	CT	4.66983	45	0.07783	17	-28
46	NJ	4.70633	46	0.05940	16	-30

SPEARMAN RANK CORRELATION TEST

OBS	R	T
1	-0.28178	1.94803

e. Harvey's Multiplicative Heteroskedasticity Test (1976)

Dependent Variable: LNE_SQ

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	14.36012	14.36012	2.992	0.0907
Error	44	211.14516	4.79875		
C Total	45	225.50528			

Root MSE	2.19061	R-square	0.0637
Dep Mean	-4.97462	Adj R-sq	0.0424
C.V.	-44.03568		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	24.852752	17.24551850	1.441	0.1566
LLNY	1	-19.082690	11.03125044	-1.730	0.0907

Variable	N	Mean	Std Dev	Minimum	Maximum
HV_TEMP	46	25.0589810	22.8713299	-2.4583422	113.8840204
LNE_SQ	46	-4.9746160	2.2385773	-13.1180671	-1.7413218

HARVEY'S MULTIPLICATIVE HETEROSKEDASTICITY TEST (1976)

OBS	HV_TEST
1	2.90997

f. Regression for Breusch and Pagan Test (1979)

Dependent Variable: X2

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	10.97070	10.97070	6.412	0.0150
Error	44	75.28273	1.71097		
C Total	45	86.25344			

Root MSE	1.30804	R-square	0.1272
Dep Mean	1.00001	Adj R-sq	0.1074
C.V.	130.80220		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	17.574476	6.54835118	2.684	0.0102
LNY	1	-3.470761	1.37065694	-2.532	0.0150

BREUSCH & PAGAN TEST (1979)

OBS	RSSBP
1	5.48535

g. Regression for White Test (1979)

Dependent Variable: E_SQ

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	5	0.01830	0.00366	4.128	0.0041
Error	40	0.03547	0.00089		
C Total	45	0.05377			

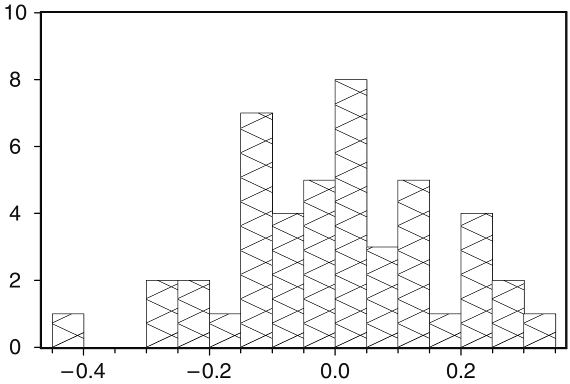
Root MSE	0.02978	R-square	0.3404
Dep Mean	0.02497	Adj R-sq	0.2579
C.V.	119.26315		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	18.221989	5.37406002	3.391	0.0016
LNP	1	9.506059	3.30257013	2.878	0.0064
LNY	1	-7.893179	2.32938645	-3.389	0.0016
LNP_SQ	1	1.281141	0.65620773	1.952	0.0579
LNPY	1	-2.078635	0.72752332	-2.857	0.0068
LNY_SQ	1	0.855726	0.25304827	3.382	0.0016

Normality Test (Jarque-Bera)

This chart was done with EViews.



Series:RESID	
Sample 1 46	
Observations 4 6	
Mean	–9.95E-16
Median	0.007568
Maximum	0.328677
Minimum	–0.418675
Std.Dev.	0.159760
Skewness	–0.181935
Kurtosis	2.812520
Jarque-Bera	0.321137
Probability	0.851659

SAS PROGRAM

```
Data CIGAR;
Input OBS STATE $ LNC LNP LNY;
CARDS;

Proc reg data=CIGAR;
        Model LNC=LNP LNY;
    Output OUT=OUT 1 R=RESID;

Proc Plot data=OUT1 hpercent=85 vpercent=60;
        Plot RESID*LNY=‘*’;
run;

***** GLEJSER’S TEST (1969) *****;
****************************************;
Data GLEJSER; set OUT1;
    ABS_E=ABS(RESID);
    Z1=LNY**-1;
    Z2=LNY**-.5;
    Z3=LNY**.5;
    Z4=LNY;
```

```
Proc reg data=GLEJSER;
        Model ABS_E=Z1;
        Model ABS_E=Z2;
        Model ABS_E=Z3;
        Model ABS_E=Z4;
TITLE 'REGRESSION FOR GLEJSER TEST (1969)';
LABEL Z1='LNY^^(-1)'
        Z2='LNY^^(-0.5)'
        Z3='LNY^^(0.5)'
        Z4='LNY^^(1)';
run;

***** GOLDFELD & QUANDT TEST (1965) *****;
*****************************************************;
Proc sort data=CIGAR out=GOLDFELD;
        By LNY;

Data GQTEST1; set GOLDFELD;
        If _N_<18; OBS=_N_;

Data GQTEST2; set GOLDFELD;
        If _N_>29; OBS=_N_-29;

Proc reg data=GQTEST1;
        Model LNC=LNP LNY;
    Output out=GQ_OUT1 R=GQ_RES1;
TITLE 'REGRESSION FOR GOLDFELD AND QUANDT TEST (1965)
        w/ first 17 obs';

Proc reg data=GQTEST2;
        Model LNC=LNP LNY;
    Output out=GQ_OUT2 R=GQ_RES2;
TITLE 'REGRESSION FOR GOLDFELD AND QUANDT TEST (1965)
        w/last 17 obs';
run;
```

```
***** SPEARMAN'S RANK CORRELATION TEST *****;
*************************************************************;
Data SPEARMN1; set GOLDFELD;
    RANKY=_N_;

Proc sort data=GLEJSER out=OUT2;
    By ABS_E;

Data TEMP1; set OUT2;
    RANKE=_N_;

Proc sort data=TEMP1 out=SPEARMN2;
    By LNY;

Data SPEARMAN;
    Merge SPEARMN1 SPEARMN2;

        By LNY;
    D=RANKE-RANKY;
* Difference b/w Ranking of —RES— and Ranking of LNY;
    D_SQ=D**2;

Proc means data=SPEARMAN NOPRINT;
    Var D_SQ;
    Output out=OUT3 SUM=SUM_DSQ;

Data SPTEST; Set OUT3;
    R=1-((6*SUM_DSQ)/(46**3-46));
    T=SQRT (R**2*(46-2)/(1-R**2));

Proc print data=SPEARMAN;
    Var STATE LNC RANKY ABS_E RANKE D;
TITLE 'DATA FOR SPEARMAN RANK CORRELATION TEST';

Proc print data=SPTEST;
```

```
    Var R T;
TITLE 'SPEARMAN RANK CORRELATION TEST';
run;

*HARVEY'S MULTIPLICATIVE HETEROSKEDASTICITY TEST (1976) **;
********************************************************************************;
Data HARVEY; set OUT1;
    E_SQ=RESID**2;
    LNE_SQ=LOG(E_SQ);
    LLNY=LOG(LNY);

Proc reg data=HARVEY;
        Model LNE_SQ=LLNY;
    Output out=OUT4 R=RLNESQ;
TITLE 'HARVEY' 'S MULTIPLICATIVE HETEROSKEDASTICITY TEST (1976)';

Data HARVEY1; set OUT4;
    HV_TEMP=LNE_SQ**2-RLNESQ**2;

Proc means data=HARVEY1;
    Var HV_TEMP LNE_SQ;
    Output out=HARVEY2 SUM=SUMTMP SUMLNESQ;

Data HARVEY3; set HARVEY2;
    HV_TEST=(SUMTMP-46* (SUMLNESQ/46)**2)/4.9348;

Proc print data=HARVEY3;
    Var HV_TEST;
TITLE 'HARVEY' 'S MULTIPLICATIVE HETEROSKEDASTICITY TEST (1976)';

***** BREUSCH & PAGAN TEST (1979) *****;
***********************************************;
Proc means data=HARVEY;
    Var E_SQ;
```

```
    Output out=OUT5 MEAN=S2HAT;
    TITLE 'BREUSCH & PAGAN TEST (1979)';

Proc print data=OUT5;
    Var S2HAT;

Data BPTEST; set HARVEY;
    X2=E_SQ/0.024968;

Proc reg data=BPTEST;
        Model X2=LNY;
    Output out=OUT6 R=BP_RES;
    TITLE 'REGRESSION FOR BREUSCH & PAGAN TEST (1979)';

Data BPTEST1; set OUT6;
    BP_TMP=X2**2-BP_RES**2;

Proc means data=BPTEST1;
    Var BP_TMP X2;
    Output out=OUT7 SUM=SUMBPTMP SUMX2;

Data BPTEST2; set OUT7;
    RSSBP=(SUMBPTMP-SUMX2**2/46)/2;

Proc print data=BPTEST2;
    Var RSSBP;

***** WHITE'S TEST (1980) *****;
************************************;
Data WHITE; set HARVEY;
    LNP_SQ=LNP**2;
    LNY_SQ=LNY**2;
    LNPY=LNP*LNY;

Proc reg data=WHITE;
```

```
    Model E_SQ=LNP LNY LNP_SQ LNPY LNY_SQ;
  TITLE 'REGRESSION FOR WHITE TEST (1979)';
run;
```

5.15 The backup regressions are given below.

a. OLS regression of consumption on a constant and Income using EViews

Dependent Variable: Consumption

Method: Least Squares

Sample: 1959 2007

Included observations: 49

	Coefficient	Std. Error	t-Statistic	Prob.
C	-1343.314	219.5614	-6.118168	0.0000
Y	0.979228	0.011392	85.96093	0.0000

R-squared	0.993680	Mean dependent var	16749.10
Adjusted R-squared	0.993545	S.D. dependent var	5447.060
S.E. of regression	437.6277	Akaike info criterion	15.04057
Sum squared resid	9001348.	Schwarz criterion	15.11779
Log likelihood	-366.4941	Hannan-Quinn criter.	15.06987
F-statistic	7389.281	Durbin-Watson stat	0.180503
Prob(F-statistic)	0.000000		

b. The Breusch–Godfrey Serial Correlation LM Test for serial correlation of the first order is obtained below using EViews. An F-statistic as well as the LM statistic which is computed as T* R-squared are reported, both of which are significant. The back up regression is also shown below these statistics. This regression runs the OLS residuals on their lagged values and the regressors in the original model. We cannot reject first order serial correlation.

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	168.9023	Prob. F(1,46)	0.0000
Obs*R-squared	38.51151	Prob. Chi-Square(1)	0.0000

Test Equation:
Dependent Variable: RESID
Method: Least Squares

Sample: 1959 2007
Included observations: 49
Presample missing value lagged residuals set to zero.

	Coefficient	Std. Error	t-Statistic	Prob.
C	-54.41017	102.7650	-0.529462	0.5990
Y	0.003590	0.005335	0.673044	0.5043
RESID(-1)	0.909272	0.069964	12.99624	0.0000

R-squared	0.785949	Mean dependent var	-5.34E-13
Adjusted R-squared	0.776643	S.D. dependent var	433.0451
S.E. of regression	204.6601	Akaike info criterion	13.53985
Sum squared resid	1926746.	Schwarz criterion	13.65567
Log likelihood	-328.7263	Hannan-Quinn criter.	13.58379
F-statistic	84.45113	Durbin-Watson stat	2.116362
Prob(F-statistic)	0.000000		

c. Cochrane-Orcutt AR(1) regression—twostep estimates using Stata

. prais c y, corc two

Iteration 0: rho = 0.0000

Iteration 1: rho = 0.9059

Cochrane-Orcutt AR(1) regression – twostep estimates

Source	SS	df	MS			
				Number of obs	=	48
				F(1, 46)	=	519.58
Model	17473195	1	17473195	Prob > F	=	0.0000
Residual	1546950.74	46	33629.364	R-squared	=	0.9187
				Adj R-squared	=	0.9169
Total	19020145.7	47	404683.951	Root MSE	=	183.38

c	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
y	.9892295	.0433981	22.79	0.000	.9018738	1.076585
_cons	-1579.722	1014.436	-1.56	0.126	-3621.676	462.2328
rho	.9059431					

Durbin-Watson statistic (original) 0.180503

Durbin-Watson statistic (transformed) 2.457550

Cochrane-Orcutt AR(1) regression – iterated estimates using Stata 11

. prais c y, corc

Iteration 0: rho = 0.0000

Iteration 1: rho = 0.9059

Iteration 2: rho = 0.8939

Iteration 3: rho = 0.8893

Iteration 4: rho = 0.8882

Iteration 5: rho = 0.8880

Iteration 6: rho = 0.8879

Iteration 7: rho = 0.8879

Iteration 8: rho = 0.8879

Iteration 9: rho = 0.8879

Cochrane-Orcutt AR(1) regression – iterated estimates

Source	SS	df	MS			
				Number of obs	=	48
				F(1, 46)	=	689.89
Model	23168597	1	23168597	Prob > F	=	0.0000
Residual	1544819.27	46	33583.0277	R-squared	=	0.9375
				Adj R-squared	=	0.9361
Total	24713416.3	47	525817.367	Root MSE	=	183.26

c	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
y	.996136	.0379253	26.27	0.000	.9197964	1.072476
_cons	-1723.689	859.2143	-2.01	0.051	-3453.198	5.819176
rho	.8879325					

Durbin-Watson statistic (original) 0.180503

Durbin-Watson statistic (transformed) 2.447750

d. Prais-Winsten AR(1) regression—iterated estimates using Stata 11

. prais c y

Iteration 0: rho = 0.0000

Iteration 1: rho = 0.9059

Iteration 2: rho = 0.9462

Iteration 3: rho = 0.9660

Iteration 4: rho = 0.9757

Iteration 5: rho = 0.9794

Iteration 6: rho = 0.9805

Iteration 7: rho = 0.9808

Iteration 8: rho = 0.9808

Iteration 9: rho = 0.9808

Iteration 10: rho = 0.9809

Iteration 11: rho = 0.9809

Iteration 12: rho = 0.9809

Prais-Winsten AR(1) regression – iterated estimates

Source	SS	df	MS			
				Number of obs	=	49
				F(1, 47)	=	119.89
Model	3916565.48	1	3916565.48	Prob > F	=	0.0000
Residual	1535401.45	47	32668.1159	R-squared	=	0.7184
				Adj R-squared	=	0.7124
Total	5451966.93	48	113582.644	Root MSE	=	180.74

c	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
y	.912147	.047007	19.40	0.000	.8175811	1.006713
_cons	358.9638	1174.865	0.31	0.761	-2004.56	2722.488
rho	.9808528					

Durbin-Watson statistic (original) 0.180503

Durbin-Watson statistic (transformed) 2.314703

e. The Newey-West HAC Standard Errors using EViews are shown below:

Dependent Variable: Consum

Method: Least Squares

Sample: 1959 2007

Included observations: 49

Newey-West HAC Standard Errors & Covariance (lag truncation = 3)

	Coefficient	Std. Error	t-Statistic	Prob.
C	-1343.314	422.2947	-3.180987	0.0026
Y	0.979228	0.022434	43.64969	0.0000

R-squared	0.993680	Mean dependent var	16749.10
Adjusted R-squared	0.993545	S.D. dependent var	5447.060
S.E. of regression	437.6277	Akaike info criterion	15.04057
Sum squared resid	9001348.	Schwarz criterion	15.11779
Log likelihood	-366.4941	Hannan-Quinn criter.	15.06987
F-statistic	7389.281	Durbin-Watson stat	0.180503
Prob(F-statistic)	0.000000		

5.16 Using EViews, Q_{t+1} is simply Q(1) and one can set the sample range from 1954–1976.

a. The OLS regression over the period 1954–1976 yields

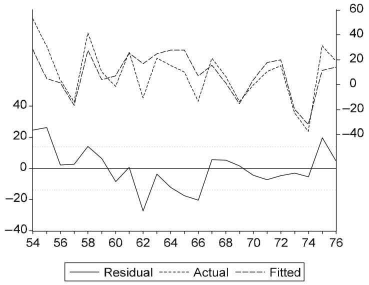
$$RS_t = \underset{(8.53)}{-6.14} + \underset{(1.44)}{6.33}\ Q_{t+1} - \underset{(1.37)}{1.67}\, P_t$$

with $R^2 = 0.62$ and D.W. $= 1.07$. The t-statistic for $\gamma = 0$ yields $t = -1.67/1.37 = -1.21$ which is insignificant with a p-value of 0.24. Therefore, the inflation rate is insignificant in explaining real stock returns.

LS // Dependent Variable is RS
Sample: 1954 1976
Included observations: 23

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-6.137282	8.528957	-0.719582	0.4801
Q(1)	6.329580	1.439842	4.396024	0.0003
P	-1.665309	1.370766	-1.214875	0.2386

R-squared	0.616110	Mean dependent var	8.900000
Adjusted R-squared	0.577721	S.D. dependent var	21.37086
S.E. of regression	13.88743	Akaike info criterion	5.383075
Sum squared resid	3857.212	Schwarz criterion	5.531183
Log likelihood	-91.54095	F-statistic	16.04912
Durbin-Watson stat	1.066618	Prob(F-statistic)	0.000070



b. The D.W. $=1.07$. for $n=23$ and two slope coefficients, the 5% critical values of the D.W. are $d_L=1.17$ and $d_U=1.54$. Since $1.07<d_L$, this indicates the presence of positive serial correlation.

c. The Breusch and Godfrey test for first-order serial correlation runs the regression of OLS residuals e_t on the regressors in the model and e_{t-1}. This yields

$$e_t = \underset{(8.35)}{-4.95} + \underset{(1.44)}{1.03}\ Q_{t+1} + \underset{(1.30)}{0.49}\ P_t + \underset{(0.22)}{0.45}\ e_{t-1}$$

with $R^2=0.171$ and $n=23$. The LM statistic is nR^2 which yields 3.94. This distributed as χ_1^2 under the null hypothesis and has a p-value of 0.047. This is significant at the 5% level and indicates the presence of first-order serial correlation.

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	3.922666	Probability	0.062305
Obs*R-squared	3.935900	Probability	0.047266

Test Equation:

LS // Dependent Variable is RESID

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-4.953720	8.350094	-0.593253	0.5600
Q(1)	1.030343	1.442030	0.714509	0.4836
P	0.487511	1.303845	0.373903	0.7126
RESID (−1)	0.445119	0.224743	1.980572	0.0623

R-squared	0.171126	Mean dependent var	4.98E-15
Adjusted R-squared	0.040251	S.D. dependent var	13.24114
S.E. of regression	12.97192	Akaike info criterion	5.282345
Sum squared resid	3197.143	Schwarz criterion	5.479822
Log likelihood	-89.38255	F-statistic	1.307555
Durbin-Watson stat	1.818515	Prob(F-statistic)	0.301033

d. The Cochrane-Orcutt yields the following regression

$$RS_t^* = -\underset{(9.34)}{14.19} + \underset{(1.20)}{7.47}\ Q_{t+1}^* - \underset{(1.63)}{0.92}\ P_t^*$$

where $RS_t^* = RS_t - \hat{\rho}RS_{t-1}$, $Q_{t+1}^* = Q_{t+1} - \hat{\rho}Q_t$ and $P_t{}^* = P_t - \hat{\rho}P_{t-1}$ with $\hat{\rho}_{CO} = 0.387$.

e. The AR(1) options on EViews yields the following results:

$$RS_t = \underset{(7.92)}{-7.32} + \underset{(1.36)}{5.91}\ Q_{t+1} - \underset{(1.28)}{1.25}\ P_t$$

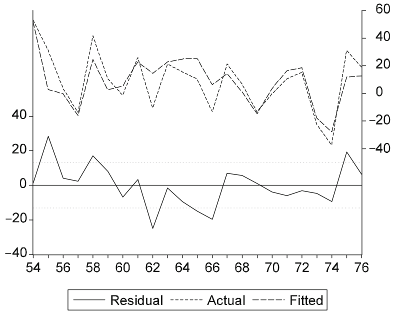
with $R^2 = 0.68$. The estimate of ρ is $\hat{\rho} = -0.027$ with a standard error of 0.014 and a t-statistic for $\rho = 0$ of -1.92. This has a p-value of 0.07. Note that even after correcting for serial correlation, P_t remains insignificant while Q_{t+1} remains significant. The estimates as well as their standard errors are affected by the correction for serial correlation. Compare with part (a).

PRAIS-WINSTEN PROCEDURE

LS // Dependent Variable is RS
Sample: 1954 1976
Included observations: 23
Convergence achieved after 4 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-7.315299	7.921839	-0.923435	0.3674
Q(1)	5.905362	1.362572	4.333981	0.0004
P	-1.246799	1.277783	-0.975752	0.3414
AR(1)	-0.027115	0.014118	-1.920591	0.0699

R-squared	0.677654	Mean dependent var	8.900000
Adjusted R-squared	0.626757	S.D. dependent var	21.37086
S.E. of regression	13.05623	Akaike info criterion	5.295301
Sum squared resid	3238.837	Schwarz criterion	5.492779
Log likelihood	-89.53155	F-statistic	13.31429
Durbin-Watson stat	1.609639	Prob (F-statistic)	0.000065
Inverted AR Roots	-.03		



5.18 The back up regressions are given below. These are performed using SAS.

a. Dependent Variable: EMP

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	2770902.9483	2770902.9483	4686.549	0.0001
Error	73	43160.95304	591.24593		
C Total	74	2814063.9013			

Root MSE	24.31555	R-square	0.9847
Dep Mean	587.94141	Adj R-sq	0.9845
C.V.	4.13571		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	-670.591096	18.59708022	-36.059	0.0001
RGNP_1	1	1.008467	0.01473110	68.458	0.0001

Durbin-Watson D	0.349
(For Number of Obs.)	75
1st Order Autocorrelation	0.788

c. Dependent Variable: RESID

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	26929.25956	26929.25956	148.930	0.0001
Error	73	13199.74001	180.81836		
U Total	74	40128.99957			

Root MSE	13.44687	R-square	0.6711
Dep Mean	-0.74410	Adj R-sq	0.6666
C.V.	-1807.13964		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
RESID_	1	0.791403	0.06484952	12.204	0.0001

COCHRANE-ORCUTT(1949) METHOD

Dependent Variable: EMP_STAR

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	1358958.4052	679479.20261	3767.845	0.0001
Error	72	12984.21373	180.33630		
U Total	74	1371942.619			

Root MSE	13.42894	R-square	0.9905
Dep Mean	129.70776	Adj R-sq	0.9903
C.V.	10.35322		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
C_STAR	1	-628.267447	50.28495095	-12.494	0.0001
RGNPSTAR	1	0.972263	0.03867418	25.140	0.0001

d. Prais-Winsten(1954, Yule-Walker) 2-Step Method

Autoreg Procedure
Dependent Variable = EMP

Ordinary Least Squares Estimates

SSE	53280.82	DFE	74
MSE	720.0111	Root MSE	26.83302
SBC	722.3376	AIC	717.6761
Reg Rsq	0.9816	Total Rsq	0.9816
Durbin's t	13.43293	PROB>t	0.0001
Durbin-Watson	0.3023		

Variable	DF	B Value	Std Error	t Ratio	Approx Prob
Intercept	1	-672.969905	20.216	-33.289	0.0001
RGNP	1	1.003782	0.016	62.910	0.0001

Estimates of Autocorrelations

Lag	Covariance	Correlation	−1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
0	701.0635	1.000000 |	|****************|
1	574.9163	0.820063 |	|********** |

Preliminary MSE = 229.5957

Estimates of the Autoregressive Parameters

Lag	Coefficient	Std Error	t Ratio
1	-0.82006318	0.06697949	-12.243497

Yule-Walker Estimates

SSE	14259.84	DFE	73
MSE	195.3403	Root MSE	13.97642
SBC	627.6068	AIC	620.6146
Reg Rsq	0.8919	Total Rsq	0.9951
Durbin-Watson	2.2216		

Variable	DF	B Value	Std Error	t Ratio	Approx Prob
Intercept	1	-559.809933	47.373	-11.817	0.0001
RGNP	1	0.914564	0.037	24.539	0.0001

e. Breusch and Godfrey (1978) LM Test

Dependent Variable: RESID Residual

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	27011.59888	13505.79944	73.331	0.0001
Error	71	13076.42834	184.17505		
C Total	73	40088.02723			

Root MSE	13.57111	R-square	0.6738
Dep Mean	-0.74410	Adj R-sq	0.6646
C.V.	-1823.83627		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	-7.177291	10.64909636	-0.674	0.5025
RESID_1	1	0.790787	0.06546691	12.079	0.0001
RGNP_1	1	0.005026	0.00840813	0.598	0.5519

SAS PROGRAM

```
Data ORANGE;
Input DATE EMP RGNP;
Cards;
Data ORANGE1; set ORANGE;
RGNP_1=LAG(RGNP);

RGNP_2=LAG2(RGNP);
EMP_1=LAG(EMP);

Proc reg data=ORANGE1;
        Model EMP=RGNP_1/DW;
    Output out=OUT1 R=RESID;

Data TEMP; set OUT1;
RESID_1=LAG(RESID);

Proc reg data=TEMP;
        Model RESID=RESID_1/noint;
run;

***** COCHRANE-ORCUTT(1949) METHOD *****;
*********************************************************;

Data CO_DATA; set ORANGE1;
EMP_STAR=EMP-0.791403^*EMP_1;        *** RHO=0.791403 ***;

RGNPSTAR=RGNP_1-0.791403^*RGNP_2;
C_STAR=1-0.791403;

Proc reg data=CO_DATA;
        Model EMP_STAR=C_STAR RGNPSTAR/noint;
TITLE 'COCHRANE-ORCUTT(1949) METHOD';
```

```
***** PRAIS-WINSTEN (1954, YULE-WALKER) METHOD *****;
*********************************************************************;

Proc autoreg data=ORANGE1;
      Model EMP=RGNP /DW=1 DWPROB LAGDEP NLAG=1 METHOD=YW;
TITLE 'PRAIS-WINSTEN(1954, YULE-WALKER) 2-STEP METHOD';

****BREUSCH & GODFREY (1978) LM TEST FOR AUTOCORRELATION***;
*********************************************************************************;

Proc reg data=TEMP;
      Model RESID=RESID_1 RGNP_1;
Title 'BREUSCH AND GODFREY (1978) LM TEST';
run;
```

5.21 a. Replication of TABLE VIII, Wheeler(2003, p. 90) entitled: County population Growth Robustness Check. We will only replicate the first and last columns of that table for p = 1 and p = 5.

For p = 1:

```
. reg dpopgr collrate mfgrate ur pcinc educsh hwsh polsh nwrate logpop, vce(r);
```

Linear regression

Number of obs	=	3102
F(9, 3092)	=	39.69
Prob > F	=	0.0000
R-squared	=	0.1275
Root MSE	=	.11041

dpopgr	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
collrate	.532155	.0658555	8.08	0.000	.40303	.66128
mfgrate	.0529082	.0202971	2.61	0.009	.0131111	.0927054
ur	-.0242001	.0790394	-0.31	0.759	-.1791751	.1307749
pcinc	1.17e-06	3.59e-06	0.33	0.743	-5.86e-06	8.21e-06
educsh	.1104783	.0197946	5.58	0.000	.0716664	.1492901
hwsh	-.0928439	.0420438	-2.21	0.027	-.1752805	-.0104073
polsh	.1669871	.1764535	0.95	0.344	-.1789909	.512965
nwrate	-.1077442	.0161301	-6.68	0.000	-.139371	-.0761173
logpop	.011208	.0026897	4.17	0.000	.0059341	.0164818
_cons	-.2330741	.027877	-8.36	0.000	-.2877334	-.1784147

The higher the proportion of the adult resident population (i.e. of age 25 or older) with a bachelor's degree or more (collrate); and the higher the proportion of total employment in manufacturing (mfgrate) the higher the County population growth rate (over the period 1980–1990). The unemployment rate (ur) and Per capita income (pcinc) have the right sign but are not significant. The higher the share of local government expenditures going to education (educsh) and police protection (polsh), the higher the County population growth rate, whereas the higher the share of local government expenditures going to roads and highways (hwsh), the lower is the County population growth rate. Except for (polsh), these local government expenditures shares are significant. The higher the proportion of the resident population that are non-white (nwrate), the lower is the County population growth rate. The larger the size of the county as measured by the log of total resident population (logpop), the higher is the County population growth rate. This is significant.

```
. test logpop=0;
(1) logpop = 0

F(1, 3092) = 17.36
Prob > F = 0.0000
```

Adding a polynomial of degree 5 in size as measured by logpop, we get for $p = 5$:

```
. reg dpopgr collrate mfgrate ur pcinc educsh hwsh polsh nwrate logpop logpop2
logpop3 logpop4 logpop5, vce(r);
```

Linear regression

Number of obs	=	3102
F(9, 3092)	=	.
Prob > F	=	.
R-squared	=	0.1584
Root MSE	=	.1085

dpopgr	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
collrate	.5625516	.0654039	8.60	0.000	.4343121	.6907911
mfgrate	.0134073	.0207008	0.65	0.517	-.0271813	.053996
ur	-.0814677	.0801316	-1.02	0.309	-.2385844	.0756489
pcinc	2.12e-06	3.25e-06	0.65	0.514	-4.26e-06	8.50e-06
educsh	.1078933	.0193862	5.57	0.000	.0698822	.1459044
hwsh	-.0492989	.0415435	-1.19	0.235	-.1307546	.0321568
polsh	.2709651	.1762385	1.54	0.124	-.0745914	.6165217
nwrate	-.0963006	.0156924	-6.14	0.000	-.1270692	-.0655321
logpop	-1.570676	1.056792	-1.49	0.137	-3.642762	.5014103
logpop2	.2330949	.2197	1.06	0.289	-.197678	.6638678
logpop3	-.014022	.0222803	-0.63	0.529	-.0577077	.0296637
logpop4	.0002559	.0011035	0.23	0.817	-.0019078	.0024196
logpop5	2.34e-06	.0000214	0.11	0.913	-.0000396	.0000443
_cons	3.515146	1.989501	1.77	0.077	-.3857338	7.416025

The results are similar to the $p = 1$ regression, but we lost the significance of mfgrate and hwsh. The joint test for the fifth degree polynomial in logpop is significant, but the last term is not, suggesting that the fourth degree polynomial is a good place to stop.

. test (logpop = 0) (logpop2 = 0) (logpop3 = 0) (logpop4 = 0) (logpop5 = 0);

(1) logpop = 0

(2) logpop2 = 0

(3) logpop3 = 0

(4) logpop4 = 0

(5) 7 logpop5 = 0

F(5, 3088) = 25.88

Prob > F = 0.0000

. test logpop5 = 0;

(1) logpop5 = 0

F(1, 3088) = 0.01

Prob > F = 0.9127

Replication of TABLE IX, Wheeler(2003, p. 91) entitled: County employment Growth Robustness Check. We will only replicate the first and last columns of that table for p = 1 and p = 5.

For p=1:

```
. reg dempgr collrate mfgrate ur pcinc educsh hwsh polsh nwrate logemp, vce(r);
```

```
Linear regression                                  Number of obs  =    3102
                                                   F(9, 3092)     =   36.11
                                                   Prob > F       =  0.0000
                                                   R-squared      =  0.1108
                                                   Root MSE       =  .13223
```

dpopgr	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
collrate	.6710405	.0836559	8.02	0.000	.5070137	.8350673
mfgrate	.0500175	.0248382	2.01	0.044	.0013163	.0987186
ur	.1106634	.0858513	1.29	0.197	-.057668	.2789947
pcinc	-3.31e-06	5.00e-06	-0.66	0.508	-.0000131	6.49e-06
educsh	.1659616	.0234637	7.07	0.000	.1199556	.2119675
hwsh	-.0704289	.0500245	-1.41	0.159	-.1685135	.0276557
polsh	.1989802	.1944894	1.02	0.306	-.1823613	.5803217
nwrate	-.1593102	.019703	-8.09	0.000	-.1979426	-.1206779
logemp	.0088257	.0033238	2.66	0.008	.0023087	.0153427
_cons	-.2242782	.0310394	-7.23	0.000	-.2851382	-.1634183

The higher the proportion of the adult resident population (i.e. of age 25 or older) with a bachelor's degree or more (collrate); and the higher the proportion of total employment in manufacturing (mfgrate) the higher the County employment growth rate (over the period 1980–1990). The unemployment rate (ur) and Per capita income (pcinc) are not significant. The higher the share of local government expenditures going to education (educsh) and police protection (polsh), the higher the County employment growth rate, whereas the higher the share of local government expenditures going to roads and highways (hwsh), the lower is the County employment growth rate. Except for (educsh), these local government expenditures shares are not significant. The higher the proportion of the resident population that are non-white (nwrate), the lower is the County educsh growth rate. The larger the size of the county as measured by the log of total resident educsh (logemp), the higher is the County employment growth rate. This is significant.

```
. test logemp = 0;
(1)  logemp = 0
     F(1, 3092) = 7.05
          Prob > F = 0.0080
```

Adding a polynomial of degree 5 in size as measured by logemp, we get for p $= 5$:

```
. reg dempgr collrate mfgrate ur pcinc educsh hwsh polsh nwrate logemp logemp2
logemp3 logemp4 logemp5, vce(r);
```

Linear regression

Number of obs	=	3102
F(13, 3088)	=	33.47
Prob > F	=	0.0000
R-squared	=	0.1295
Root MSE	=	.13091

dempgr	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf. Interval]	
collrate	.7127472	.0798073	8.93	0.000	.5562665	.8692279
mfgrate	.0162264	.0256496	0.63	0.527	-.0340656	.0665184
ur	.0654624	.0829589	0.79	0.430	-.0971977	.2281226
pcinc	-5.60e-06	3.67e-06	-1.52	0.128	-.0000128	1.60e-06
educsh	.1677939	.0231205	7.26	0.000	.1224607	.2131271
hwsh	-.034885	.049391	-0.71	0.480	-.1317276	.0619577
polsh	.2933336	.1922457	1.53	0.127	-.0836087	.6702759
nwrate	-.1540241	.0175901	-8.76	0.000	-.1885135	-.1195347
logemp	-2.760358	.8853777	-3.12	0.002	-4.496347	-1.024369
logemp2	.5122333	.1992707	2.57	0.010	.1215167	.9029498
logemp3	-.044739	.0218277	-2.05	0.040	-.0875373	-.0019406
logemp4	.0018485	.0011644	1.59	0.113	-.0004346	.0041316
logemp5	-.000029	.0000242	-1.20	0.231	-.0000765	.0000185
_cons	5.430814	1.540746	3.52	0.000	2.409824	8.451804

The results are similar to the p = 1 regression, but we lost the significance of mfgrate. The joint test for the fifth degree polynomial in logpop is significant, but the last term is not, suggesting that the fourth degree polynomial is a good place to stop.

. test (logemp = 0) (logemp2 = 0) (logemp3 = 0) (logemp4 = 0) (logemp5 = 0);

(1) logemp = 0
(2) logemp2 = 0
(3) logemp3 = 0
(4) logemp4 = 0
(5) logemp5 = 0

F(5, 3088) = 18.27
Prob > F = 0.0000

. test logemp5 = 0;

(1) logemp5 = 0

F(1, 3088) = 1.44
Prob > F = 0.2307

b. Breusch–Pagan test for heteroskedasticity for the specification with a fourth degree polynomial in log(size) in TABLE VIII, Wheeler(2003, p. 90).

. reg dpopgr collrate mfgrate ur pcinc educsh hwsh polsh nwrate logpop logpop2 logpop3 logpop4;

Source	SS	df	MS			
				Number of obs	=	3102
				F(12,3089)	=	48.43
Model	6.84020955	12	.570017462	Prob > F	=	0.0000
Residual	36.3559406	3089	.011769485	R-squared	=	0.1584
				Adj R-squared	=	0.1551
Total	43.1961502	3101	.013929749	Root MSE	=	.10849

dpopgr	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
collrate	.5632503	.0516707	10.90	0.000	.4619378	.6645628
mfgrate	.013464	.0198056	0.68	0.497	-.0253694	.0522974
ur	-.0817679	.0672396	-1.22	0.224	-.2136069	.050071
pcinc	2.07e-06	2.51e-06	0.83	0.409	-2.85e-06	6.99e-06
educsh	.10792	.0179056	6.03	0.000	.072812	.143028
hwsh	-.0493372	.0421236	-1.17	0.242	-.1319303	.0332559
polsh	.2712608	.1304948	2.08	0.038	.0153953	.5271262
nwrate	-.0964934	.0151612	-6.36	0.000	-.1262205	-.0667663
logpop	-1.684932	.3836672	-4.39	0.000	-2.437201	-.9326635
logpop2	.2569229	.057085	4.50	0.000	.1449945	.3688513
logpop3	-.0164455	.0037225	-4.42	0.000	-.0237443	-.0091467
logpop4	.0003763	.0000898	4.19	0.000	.0002003	.0005523
_cons	3.7286	.9577641	3.89	0.000	1.850681	5.606519

. estat hettest logpop logpop2 logpop3 logpop4, rhs normal;

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity

Ho: Constant variance

Variables: logpop logpop2 logpop3 logpop4 collrate mfgrate ur pcinc educsh hwsh polsh nwrate

chi2(12) = 589.14

Prob > chi2 = 0.0000

Breusch-Pagan test for heteroskedasticity for the specification with a 4$^{\text{th}}$ degree polynomial in log(size) in TABLE IX, Wheeler(2003, p. 91).

. reg dempgr collrate mfgrate ur pcinc educsh hwsh polsh nwrate logemp logemp2 logemp3 logemp4 ;

Source	SS	df	MS			
				Number of obs	=	3102
				F(12,3089)	=	38.23
Model	7.861795	12	.655149583	Prob > F	=	0.0000
Residual	52.9333023	3089	.017136064	R-squared	=	0.1293
				Adj R-squared	=	0.1259
Total	60.7950973	3101	.019604998	Root MSE	=	.1309

dempgr	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
collrate	.7040415	.0626979	11.23	0.000	.5811078	.8269753
mfgrate	.0156912	.0242125	0.65	0.517	-.0317831	.0631654
ur	.0691242	.0802564	0.86	0.389	-.0882371	.2264855
pcinc	-4.95e-06	3.06e-06	-1.61	0.106	-.000011	1.06e-06
educsh	.1674244	.0216266	7.74	0.000	.1250204	.2098283
hwsh	-.034799	.0507792	-0.69	0.493	-.1343634	.0647654
polsh	.2889395	.1573668	1.84	0.066	-.0196147	.5974937
nwrate	-.1516358	.0182735	-8.30	0.000	-.1874653	-.1158063
logemp	-1.730393	.3710692	-4.66	0.000	-2.45796	-1.002826
logemp2	.2785811	.0599367	4.65	0.000	.1610613	.3961009
logemp3	-.0189696	.0042326	-4.48	0.000	-.0272686	-.0106706
logemp4	.000464	.0001103	4.21	0.000	.0002478	.0006802
_cons	3.666787	.8526094	4.30	0.000	1.995049	5.338526

```
. estat hettest logemp logemp2 logemp3 logemp4, rhs normal;

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity

        Ho: Constant variance

        Variables: logemp logemp2 logemp3 logemp4 collrate mfgrate ur pcinc
educsh hwsh polsh nwrate

        chi2(12) = 425.82

        Prob > chi2 = 0.0000
```

References

Baltagi, B. and Q. Li (1995), “ML Estimation of Linear Regression Model with AR(1) Errors and Two Observations,” *Econometric Theory*, Solution 93.3.2, 11: 641–642.

Johnston, J. (1984), Econometric Methods, 3rd. Ed., (McGraw-Hill: New York).

Kim, J.H. (1991), “The Heteroscedastic Consequences of an Arbitrary Variance for the Initial Disturbance of an AR(1) Model,” *Econometric Theory*, Solution 90.3.1, 7: 544–545.

Kmenta, J. (1986), Elements of Econometrics (Macmillan: New York).

Koning, R.H. (1992), “The Bias of the Standard Errors of OLS for an AR(1) Process with an Arbitrary Variance on the Initial Observations,” *Econometric Theory*, Solution 92.1.4, 9: 149–150.

CHAPTER 6
Distributed Lags and Dynamic Models

6.1 a. Using the Linear Arithmetic lag given in Eq. (6.2), a 6 year lag on income gives a regression of consumption on a constant and $Z_t = \sum_{i=0}^{6} (7 - i)\, X_{t-i}$ where X_t denotes income. In this case,

$Z_t = 7X_t + 6X_{t-1} + .. + X_{t-6}$,

The Stata regression output is given below:

. gen z_6=7*ly+6*l.ly+5*l2.ly+4*l3.ly+3*l4.ly+2*l5.ly+l6.ly

(6 missing values generated)

. reg lc z_6

Source	SS	df	MS			
				Number of obs	=	43
				F(1, 41)	=	3543.62
Model	3.26755259	1	3.26755259	Prob > F	=	0.0000
Residual	.037805823	41	.000922093	R-squared	=	0.9886
				Adj R-squared	=	0.9883
Total	3.30535842	42	.07869901	Root MSE	=	.03037

lc	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
z_6	.0373029	.0006266	59.53	0.000	.0360374	.0385684
_cons	-.4950913	.1721567	-2.88	0.006	-.8427688	-.1474138

From Eq. (6.2)

$\beta_i = [(s + 1) - i]\beta$ for $i = 0, .., 6$

with β estimated as the coefficient of Z_t (which is z_6 in the regression). This estimate is 0.037 and is statistically significant.

Now we generate the regressors for an Almon lag `first-degree` polynomial with a `far end point` constraint using Stata:

. gen Z0= ly+l.ly+l2.ly+l3.ly+l4.ly+l5.ly+l6.ly

(6 missing values generated)

B.H. Baltagi, *Solutions Manual for Econometrics*, Springer Texts in Business and Economics, DOI 10.1007/978-3-642-54548-1_6,

```
. gen Z1=0*ly+l.ly+2*l2.ly+3*l3.ly+4*l4.ly+5*l5.ly+6*l6.ly
(6 missing values generated)
. gen Z=Z1-7*Z0
(6 missing values generated)
. reg lc Z
```

Source	SS	df	MS
Model	3.26755293	1	3.26755293
Residual	.037805483	41	.000922085
Total	3.30535842	42	.07869901

Number of obs	=	43
F(1, 41)	=	3543.66
Prob > F	=	0.0000
R-squared	=	0.9886
Adj R-squared	=	0.9883
Root MSE	=	.03037

lc	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
Z	-.0373029	.0006266	-59.53	0.000	-.0385684	-.0360374
_cons	-.4950919	.1721559	-2.88	0.006	-.8427679	-.147416

The EViews output for PDL(Y,6,1,2) which is a sixth order lag, *first degree* polynomial, with a *far end point* constraint is given by:

Dependent Variable: LNC

Method: Least Squares

Sample (adjusted): 1965 2007

Included observations: 43 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.495091	0.172156	-2.875817	0.0064
PDL01	0.149212	0.002507	59.52843	0.0000

R-squared	0.988562	Mean dependent var	9.749406
Adjusted R-squared	0.988283	S.D. dependent var	0.280533
S.E. of regression	0.030366	Akaike info criterion	-4.105595
Sum squared resid	0.037806	Schwarz criterion	-4.023678
Log likelihood	90.27028	Hannan-Quinn criter.	-4.075386
F-statistic	3543.634	Durbin-Watson stat	0.221468
Prob(F-statistic)	0.000000		

Lag Distribution of LNY	i	Coefficient	Std. Error	t-Statistic
. *|	0	0.26112	0.00439	59.5284
. * |	1	0.22382	0.00376	59.5284
. * |	2	0.18651	0.00313	59.5284
. * |	3	0.14921	0.00251	59.5284
. * |	4	0.11191	0.00188	59.5284
. * |	5	0.07461	0.00125	59.5284
.* |	6	0.03730	0.00063	59.5284
	Sum of Lags	1.04448	0.01755	59.5284

b. Using an Almon-lag second degree polynomial described in Eq. (6.4), a 6 year lag on income gives a regression of consumption on a constant, $Z_0 = \sum_{i=0}^{6} X_{t-i}, Z_1 = \sum_{i=0}^{6} iX_{t-i}$ and $Z_2 = \sum_{i=0}^{6} i^2X_{t-i}$. This yields the Almon-lag without near or far end-point constraints. A near end-point constraint imposes $\beta_{-1} = 0$ in Eq. (6.1) which yields $a_0 - a_1 + a_2 = 0$ in Eq. (6.4). Substituting for a_0 in Eq. (6.4) yields the regression in (6.5).

The following Stata code, generates the variables needed to estimate an Almon lag `second-degree` polynomial with a `near end point` constraint:

```
. gen Z2=0*ly+l.ly+2^2*l2.ly+3^2*l3.ly+4^2*l4.ly+5^2*l5.ly+6^2*l6.ly
(6 missing values generated)
. gen Z01= Z1+Z0
(6 missing values generated)
. gen Z02= Z2-Z0
(6 missing values generated)
. reg lc Z01 Z02
```

Source	SS	df	MS			
				Number of obs	=	43
				F(2, 40)	=	1701.72
Model	3.26696244	2	1.63348122	Prob > F	=	0.0000
Residual	.038395978	40	.000959899	R-squared	=	0.9884
				Adj R-squared	=	0.9878
Total	3.30535842	42	.07869901	Root MSE	=	.03098

lc	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Z01	.1708636	.0260634	6.56	0.000	.1181875	.2235397
Z02	-.0441775	.0085125	-5.19	0.000	-.0613819	-.026973
_cons	-.8139917	.2306514	-3.53	0.001	-1.280156	-.3478277

The EViews output for PDL(Y,6,2,1) which is a sixth order lag, second degree polynomial, with a *near end point* constraint is given by:

Dependent Variable: LNC

Method: Least Squares

Sample (adjusted): 1965 2007

Included observations: 43 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.813961	0.230649	-3.529009	0.0011
PDL01	0.259216	0.043085	6.016333	0.0000
PDL02	-0.044177	0.008512	-5.189671	0.0000

R-squared	0.988384	Mean dependent var	9.749406
Adjusted R-squared	0.987803	S.D. dependent var	0.280533
S.E. of regression	0.030982	Akaike info criterion	-4.043589
Sum squared resid	0.038396	Schwarz criterion	-3.920714
Log likelihood	89.93716	Hannan-Quinn criter.	-3.998276
F-statistic	1701.718	Durbin-Watson stat	0.377061
Prob(F-statistic)	0.000000		

Lag Distribution of LNY	i	Coefficient	Std. Error	t-Statistic
. * |	0	0.21504	0.03457	6.21968
. * |	1	0.34172	0.05213	6.55566
. *|	2	0.38006	0.05266	7.21676
. * |	3	0.33003	0.03621	9.11501
. * |	4	0.19166	0.00402	47.7225
*. |	5	-0.03508	0.04810	0.72915
* . |	6	-0.35016	0.11562	3.02843
	Sum of Lags	1.07327	0.02249	47.7225

c. The far end-point constraint imposes $\beta_7 = 0$. This translates into the following restriction $a_0 + 7a_1 + 49a_2 = 0$. Substituting for a_0 in (6.4) yields the regression in (6.6) with $s = 6$, i.e., the regression of consumption on a constant, $(Z_1 - 7Z_0)$ and $(Z_2 - 49Z_0)$.

The following Stata code, generates the variables needed to estimate an Almon lag `second-degree` polynomial with a `far end point` constraint:

```
. gen Z10_far=Z1-7*Z0
(6 missing values generated)

. gen Z20_far=Z2-7^2*Z0
(6 missing values generated)

. reg lc Z10_far Z20_far
```

Source	SS	df	MS			
				Number of obs	=	43
				F(2, 40)	=	3536.82
Model	3.28677238	2	1.64338619	Prob > F	=	0.0000
Residual	.018586036	40	.000464651	R-squared	=	0.9944
				Adj R-squared	=	0.9941
Total	3.30535842	42	.07869901	Root MSE	=	.02156

lc	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Z10_far	-.3833962	.0538147	-7.12	0.000	-.4921598	-.2746326
Z20_far	.0381843	.0059371	6.43	0.000	.0261849	.0501837
_cons	-1.237493	.1681062	-7.36	0.000	-1.577249	-.8977381

The EViews output for PDL(Y,6,2,2) which is a sixth order lag, second degree polynomial, with a *far end point* constraint is given by:

Dependent Variable: LNC

Method: Least Squares

Sample (adjusted): 1965 2007

Included observations: 43 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.237555	0.168112	-7.361483	0.0000
PDL01	0.006206	0.022306	0.278216	0.7823
PDL02	-0.154297	0.018196	-8.479577	0.0000

R-squared	0.994377	Mean dependent var	9.749406
Adjusted R-squared	0.994096	S.D. dependent var	0.280533
S.E. of regression	0.021556	Akaike info criterion	-4.769135
Sum squared resid	0.018586	Schwarz criterion	-4.646261
Log likelihood	105.5364	Hannan-Quinn criter.	-4.723823
F-statistic	3536.826	Durbin-Watson stat	0.377278
Prob(F-statistic)	0.000000		

Lag Distribution of LNY	i	Coefficient	Std. Error	t-Statistic
. *|	0	0.81278	0.08583	9.46951
. * |	1	0.46755	0.03799	12.3071
. * |	2	0.19869	0.00292	68.0281
* |	3	0.00621	0.02231	0.27822
* . |	4	-0.10990	0.03451	-3.18430
* . |	5	-0.14964	0.03488	-4.29038
* . |	6	-0.11301	0.02338	-4.83449
	Sum of Lags	1.11266	0.01636	-68.0281

d. The following Stata code, generates the variables needed to estimate an Almon lag *second-degree* polynomial with *both near and far end point* constraints:

```
. gen Z_NF=-47* Z0-6* Z1+ Z2
(6 missing values generated)


. reg lc Z_NF
```

Source	SS	df	MS			
				Number of obs	=	43
				F(1, 41)	=	2663.90
Model	3.25525695	1	3.25525695	Prob > F	=	0.0000
Residual	.050101463	41	.001221987	R-squared	=	0.9848
				Adj R-squared	=	0.9845
Total	3.30535842	42	.07869901	Root MSE	=	.03496

lc	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Z_NF	-.0028049	.0000543	-51.61	0.000	-.0029147	-.0026952
_cons	-.2411805	.1936405	-1.25	0.220	-.6322455	.1498845

The EViews output for PDL(Y,6,2,3) which is a sixth order lag, second degree polynomial, with *both near and far end point* constraints is given by:

Dependent Variable: LNC

Method: Least Squares

Sample (adjusted): 1965 2007

Included observations: 43 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.233622	0.197176	-1.184841	0.2429
PDL01	0.097163	0.001918	50.64935	0.0000

R-squared	0.984269	Mean dependent var	9.749406
Adjusted R-squared	0.983886	S.D. dependent var	0.280533
S.E. of regression	0.035612	Akaike info criterion	-3.786892
Sum squared resid	0.051996	Schwarz criterion	-3.704975
Log likelihood	83.41817	Hannan-Quinn criter.	-3.756683
F-statistic	2565.357	Durbin-Watson stat	0.214478
Prob(F-statistic)	0.000000		

Lag Distribution of LNY	i	Coefficient	Std. Error	t-Statistic
. * |	0	0.08502	0.00168	50.6494
. * |	1	0.14574	0.00288	50.6494
. * |	2	0.18218	0.00360	50.6494
. *|	3	0.19433	0.00384	50.6494
. * |	4	0.18218	0.00360	50.6494
. * |	5	0.14574	0.00288	50.6494
. * |	6	0.08502	0.00168	50.6494
	Sum of Lags	1.02021	0.02014	50.6494

e. The RRSS for the Chow test for the *arithmetic lag restrictions* is given by the residual sum of squares of the regression in part (a), i.e., .037805823. The URSS is obtained from running consumption on a constant and six lags on income. The corresponding Stata regression is given below:

```
. reg lc ly l.ly l2.ly l3.ly l4.ly l5.ly l6.ly
```

Source	SS	df	MS			
				Number of obs	=	43
				F(7, 35)	=	1093.59
Model	3.29031477	7	.470044967	Prob > F	=	0.0000
Residual	.015043647	35	.000429818	R-squared	=	0.9954
				Adj R-squared	=	0.9945
Total	3.30535842	42	.07869901	Root MSE	=	.02073

lc	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ly						
–.	1.237818	.2192865	5.64	0.000	.7926427	1.682993
L1.	.2504519	.310222	0.81	0.425	-.3793323	.8802361
L2.	-.203472	.3005438	-0.68	0.503	-.8136084	.4066644
L3.	-.0279364	.3041055	-0.09	0.927	-.6453034	.5894306
L4.	.0312238	.3049614	0.10	0.919	-.5878808	.6503284
L5.	-.0460776	.3048432	-0.15	0.881	-.6649422	.572787
L6.	-.1270834	.2028801	-0.63	0.535	-.5389519	.2847851
_cons	-1.262225	.1667564	-7.57	0.000	-1.600758	-.9236913

URSS = .015043647. The number of restrictions given in (6.2) is 6. Hence, the Chow F-statistic can be computed as follows:

```
. display (.037805483-.015043647)*35/(6*.015043647)
8.8261428
```

and this is distributed as F(6,35) under the null hypothesis. This rejects the arithmetic lag restrictions.

f. Similarly, the Chow test for the Almon lag *second-degree* polynomial with a *near end point* constraint can be computed using RRSS = .038395978 from part (b). URSS = .015043647 from the last regression in part (e), and the number of restrictions is 5:

```
. display (.038395978-.015043647)*35/(5*.015043647)
10.866136
```

and this is distributed as F(5,35) under the null hypothesis. This rejects the Almon lag *second-degree* polynomial with a *near end point* constraint.

g. The Chow test for the Almon lag *second-degree* polynomial with a *far end point* constraint can be computed using RRSS = .018586036 from part (c). URSS = .015043647 from the last regression in part (e), and the number of restrictions is 5:

```
. display (.018586036-.015043647)*35/(5*.015043647)
1.6483186
```

and this is distributed as F(5,35) under the null hypothesis. This does not reject the Almon lag *second-degree* polynomial with a *far end point* constraint.

Finally, The Chow test for the Almon lag *second-degree* polynomial with *both near and far end point* constraints can be computed using RRSS = .050101463 from part (d). URSS = .015043647 from the last regression in part (e), and the number of restrictions is 6:

```
. display (.050101463-.015043647)*35/(6*.015043647)
13.594039
```

and this is distributed as F(6,35) under the null hypothesis. This rejects the Almon lag *second-degree* polynomial with *both near and far end point* constraints.

6.2 a. For the Almon-lag third degree polynomial

$$\beta_i = a_0 + a_1 i + a_2 i^2 + a_3 i^3 \text{ for } i = 0, 1, .., 5.$$

In this case, (6.1) reduces to

$$Y_i = \alpha + \sum_{i=1}^{5} \left(a_0 + a_1 i + a_2 i^2 + a_3 i^3\right) X_{t-i} + u_t$$

$$= \alpha + a_0 + \sum_{i=0}^{5} X_{t-i} + a_1 \sum_{i=0}^{5} iX_{t-i} + a_2 \sum_{i=0}^{5} i^2 X_{t-i} + a_3 \sum_{i=0}^{5} i^3 X_{t-i} + u_t,$$

Now α, a_0, a_1, a_2 and a_3 can be estimated from the regression of Y_t on a constant, $Z_0 = \sum_{i=0}^{5} X_{t-i}$, $Z_1 = \sum_{i=0}^{5} iX_{t-i}$, $Z_2 = \sum_{i=0}^{5} i^2 X_{t-i}$ and $Z_3 = \sum_{i=0}^{5} i^3 X_{t-i}$. The following Stata code generates the variables to run the OLS regression:

```
. gen Z5_0= ly+l.ly+l2.ly+l3.ly +l4.ly +l5.ly
(5 missing values generated)
. gen Z5_1=l.ly+2*l2.ly +3*l3.ly +4*l4.ly +5*l5.ly
(5 missing values generated)
. gen Z5_2=l.ly+2^2*l2.ly +3^2*l3.ly +4^2*l4.ly +5^2*l5.ly
(5 missing values generated)
```

```
. gen Z5_3=l.ly+2^3*l2.ly +3^3*l3.ly +4^3*l4.ly +5^3*l5.ly
(5 missing values generated)
. reg lc Z5_0 Z5_1 Z5_2 Z5_3
```

Source	SS	df	MS		
				Number of obs =	44
				F(4, 39) =	2063.28
Model	3.58974185	4	.897435462	Prob > F =	0.0000
Residual	.016963297	39	.000434956	R-squared =	0.9953
				Adj R-squared =	0.9948
Total	3.60670515	43	.083876864	Root MSE =	.02086

lc	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Z5_0	1.27317	.1911863	6.66	0.000	.8864591	1.659881
Z5_1	-1.623359	.5714614	-2.84	0.007	-2.779249	-.4674696
Z5_2	.5693275	.2955923	1.93	0.061	-.0285644	1.167219
Z5_3	-.0599977	.0390623	-1.54	0.133	-.1390087	.0190133
_cons	-1.128979	.1533499	-7.36	0.000	-1.439158	-.8187994

b. The estimate of β_3 is $\hat{\beta}_3 = \hat{a}_0 + 3\hat{a}_1 + 9\hat{a}_2 + 27\hat{a}_3$ with

$$\mathrm{var}\left(\hat{\beta}_3\right) = \mathrm{var}(\hat{a}_0) + 9\,\mathrm{var}(\hat{a}_1) + 81\,\mathrm{var}(\hat{a}_2) + 27^2\mathrm{var}(\hat{a}_3)$$

$$+ 2.3.\ \mathrm{cov}(\hat{a}_0, \hat{a}_1) + 2.9.\ \mathrm{cov}(\hat{a}_0, \hat{a}_2) + 2.27.\mathrm{cov}(\hat{a}_0, \hat{a}_3)$$

$$+ 2.3.9.\ \mathrm{cov}(\hat{a}_1, \hat{a}_2) + 2.3.27\ \mathrm{cov}(\hat{a}_1, \hat{a}_3) + 2.9.27\ \mathrm{cov}(\hat{a}_2, \hat{a}_3).$$

The estimate of β_3 can be computed from this regression results as follows:

```
. *beta3
. display 1.27317+3*(-1.623359)+9*.5693275+27*(-.0599977)
-.0928974
```

The var-cov matrix of the regression estimates are given by:

```
. matrix list e(V)
symmetric e(V)[5,5]
```

	Z5_0	Z5_1	Z5_2	Z5_3	_cons
Z5_0	.0365522				
Z5_1	-.09755986	.32656818			
Z5_2	.04577699	-.16633866	.08737482		
Z5_3	-.00565526	.0214819	-.0114859	.00152587	
_cons	-.01320609	.02236579	-.00962469	.00120366	.02351619

The EViews output for PDL(Y,5,3) which is a fifth order lag, *third degree* polynomial, with *no end point* constraint is given by:

Dependent Variable: LNC

Method: Least Squares

Sample (adjusted): 1964 2007

Included observations: 44 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.129013	0.153388	-7.360517	0.0000
PDL01	-0.176078	0.122030	-1.442904	0.1570
PDL02	-0.066325	0.162878	-0.407206	0.6861
PDL03	0.209243	0.066943	3.125704	0.0033
PDL04	-0.059926	0.039088	-1.533091	0.1333

R-squared	0.995295	Mean dependent var	9.736786
Adjusted R-squared	0.994813	S.D. dependent var	0.289615
S.E. of regression	0.020859	Akaike info criterion	-4.795464
Sum squared resid	0.016968	Schwarz criterion	-4.592715
Log likelihood	110.5002	Hannan-Quinn criter.	-4.720275
F-statistic	2062.698	Durbin-Watson stat	0.393581
Prob(F-statistic)	0.000000		

Lag Distribution of LNY	i	Coefficient	Std. Error	t-Statistic
. *|	0	1.27295	0.19130	6.65411
. * |	1	0.15942	0.15649	1.01868
* . |	2	-0.17608	0.12203	-1.44290
*. |	3	-0.09309	0.11943	-0.77943
.* |	4	0.04883	0.15599	0.31306
*. |	5	-0.10987	0.17870	-0.61485
	Sum of Lags	1.10217	0.01492	73.8706

c. Imposing the near end-point constraint $\beta_{-1} = 0$ yields the following restriction on the third degree polynomial in a's:

$a_0 - a_1 + a_2 - a_3 = 0.$

solving for a_0 and substituting above yields the following constrained regression:

$Y_t = \alpha + a_1(Z_1 + Z_0) + a_2(Z_2 - Z_0) + a_3(Z_1 + Z_3) + u_t$

The corresponding Stata regression is reported below.

```
. gen ZZ01= Z5_0+Z5_1
(5 missing values generated)
. gen ZZ03= Z5_0+Z5_3
(5 missing values generated)
. gen ZZ02= Z5_2-Z5_0
(5 missing values generated)
. reg lc ZZ01 ZZ03 ZZ02
```

Source	SS	df	MS
Model	3.58503644	3	1.19501215
Residual	.021668704	40	.000541718
Total	3.60670515	43	.083876864

Number of obs	=	44
F(3, 40)	=	2205.97
Prob > F	=	0.0000
R-squared	=	0.9940
Adj R-squared	=	0.9935
Root MSE	=	.02327

lc	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
ZZ01	.2551128	.0259506	9.83	0.000	.2026646	.307561
ZZ03	.0632448	.0123507	5.12	0.000	.0382831	.0882065
ZZ02	-.3852599	.0629406	-6.12	0.000	-.5124676	-.2580523
_cons	-.9865419	.1641747	-6.01	0.000	-1.318351	-.6547324

d. Test the near end point constraint with a Chow test. The URSS = .016963297 from part (a) and RRSS = .021668704 from part (c) and there is one restriction.

```
. display (.021668704-.016963297)*39/(1*.016963297)
10.818114
```

and this is distributed as F(1,39) under the null hypothesis. This rejects the *near end point* constraint.

e. The Chow test for the Almon 5 year lag *third-degree* polynomial with *no end point* constraints can be computed using RRSS = .016963297 from part (a). URSS = .016924337 from the unrestricted 5 year lag regression given below:

```
. reg lc ly l.ly l2.ly l3.ly l4.ly l5.ly

Source   |      SS          df        MS            Number of obs  =      44
-----------------------------------------           F(6, 37)       = 1308.00
Model    |  3.58978081       6    .598296802        Prob > F       =  0.0000
Residual |  .016924337      37    .000457415        R-squared      =  0.9953
-----------------------------------------           Adj R-squared  =  0.9945
Total    |  3.60670515      43    .083876864        Root MSE       =  .02139

------------------------------------------------------------------------------
      lc |    Coef.      Std. Err.      t     P>|t|     [95% Conf. Interval]
------------------------------------------------------------------------------
      ly |
      --. |  1.303457    .2224286     5.86   0.000     .852774      1.75414
      L1. |   .090694    .3099447     0.29   0.771    -.5373136     .7187016
      L2. | -.1401868    .3081789    -0.45   0.652    -.7646165      .484243
      L3. | -.0460399    .3104643    -0.15   0.883    -.6751003     .5830204
      L4. | -.0255622    .3129629    -0.08   0.935    -.6596853     .6085608
      L5. | -.0800021    .2087812    -0.38   0.704    -.5030331     .3430288
   _cons |  -1.13104     .1574744    -7.18   0.000    -1.450114    -.8119665
------------------------------------------------------------------------------
```

and the number of restrictions is 3:

```
. display (.016963297- .016924337)*37/(3*.016924337)
.02839146
```

and this is distributed as F(3,37) under the null hypothesis. This does not reject the Almon 5 year lag *third-degree* polynomial restrictions.

6.3 a. From (6.18), $Y_t = \beta Y_{t-1} + \nu_t$. Therefore, $Y_{t-1} = \beta Y_{t-2} + \nu_{t-1}$ and

$\rho Y_{t-1} = \rho\beta Y_{t-2} + \rho\nu_{t-1}$.

Subtracting this last equation from (6.18) and re-arranging terms, one gets

$Y_t = (\beta + \rho)\, Y_{t-1} - \rho\beta Y_{t-2} + \varepsilon_t$.

Multiply both sides by Y_{t-1} and sum, we get

$$\sum_{t=2}^{T} Y_t Y_{t-1} = (\beta + \rho) \sum_{t=2}^{T} Y_{t-1}^2 - \rho\beta \sum_{t=2}^{T} Y_{t-1} Y_{t-2} + \sum_{t=2}^{T} Y_{t-1}\varepsilon_t.$$

Divide by $\sum_{t=2}^{T} Y_{t-1}^2$ and take probability limits, we get

$$\text{plim} \sum_{t=2}^{T} Y_t Y_{t-1} \Big/ \sum_{t=2}^{T} Y_{t-1}^2 = (\beta + \rho) - \rho\beta \,\text{plim} \left(\sum_{t=2}^{T} Y_{t-1} Y_{t-2} \Big/ \sum_{t=2}^{T} Y_{t-1}^2 \right) + \text{plim} \left(\sum_{t=2}^{T} Y_{t-1}\varepsilon_t \Big/ \sum_{t=2}^{T} Y_{t-1}^2 \right)$$

The last term has zero probability limits since it is the $\text{cov}(Y_{t-1}, \varepsilon_t)/\text{var}(Y_{t-1})$. The numerator is zero whereas the denominator is finite and positive. As $T \to \infty$, plim $\sum_{t=2}^{T} Y_{t-1}Y_{t-2}$ is almost identical to plim $\sum_{t=2}^{T} Y_t Y_{t-1}$. Hence, as $T \to \infty$, one can solve for

$$\text{plim} \sum_{t=2}^{T} Y_t Y_{t-1} / \sum_{t=2}^{T} Y_{t-1}^2 = (\beta + \rho) / (1 + \rho\beta).$$

Therefore,

$$\text{plim}\hat{\beta}_{ols} = (\beta + \rho) / (1 + \rho\beta)$$

and

$$\text{plim} \left(\hat{\beta}_{ols} - \beta \right) = \rho \left(1 - \beta^2 \right) / (1 + \rho\beta).$$

b. One can tabulate the asymptotic bias of $\hat{\beta}_{ols}$ derived in (a) for various values of $|\rho| < 1$ and $|\beta| < 1$.

c. For $\hat{\rho} = \sum_{t=2}^{T} \hat{\nu}_t \hat{\nu}_{t-1} / \sum_{t=2}^{T} \hat{\nu}_{t-1}^2$ with $\hat{\nu}_t = Y_t - \hat{\beta}_{ols} Y_{t-1}$ we compute

$$\sum_{t=2}^{T} \hat{\nu}_{t-1}^2 / T = \sum_{t=2}^{T} Y_t^2 / T - 2\hat{\beta}_{ols} \left(\sum_{t=2}^{T} Y_t Y_{t-1} / T \right) + \hat{\beta}_{ols}^2 \left(\sum_{t=2}^{T} Y_{t-1}^2 / T \right).$$

Defining the covariances of Y_t by γ_s, and using the fact that $\text{plim}(\Sigma Y_t Y_{t-s}/T) = \gamma_s$ for $s = 0, 1, 2, ..$, we get

$$\text{plim} \sum_{t=2}^{T} \hat{\nu}_{t-1}^2 / T = \gamma_0 - 2 \left(\text{plim}\hat{\beta}_{ols} \right) \gamma_1 + \left(\text{plim}\hat{\beta}_{ols} \right)^2 \gamma_0.$$

But

$$\text{plim}\hat{\beta}_{ols} = \text{plim} \left[\frac{\sum_{t=2}^{T} Y_t Y_{t-1} / T}{\sum_{t=2}^{T} Y_{t-1}^2 / T} \right] = \gamma_1 / \gamma_0.$$

Hence,

$$\operatorname{plim} \sum_{t=2}^{T} \hat{\nu}_{t-1}^2 \, /T = \gamma_0 - 2\gamma_1^2/\gamma_0 - \gamma_1^2/\gamma_0 = \gamma_0 - \left(\gamma_1^2/\gamma_0\right) = \left(\gamma_0^2 - \gamma_1^2\right) /\gamma_0.$$

Also, $\hat{\nu}_{t-1} = Y_{t-1} - \hat{\beta}_{ols} Y_{t-2}$. Multiply this equation by $\hat{\nu}_t$ and sum, we get

$$\sum_{t=2}^{T} \hat{\nu}_t \hat{\nu}_{t-1} = \sum_{t=2}^{T} Y_{t-1} \hat{\nu}_t - \hat{\beta}_{ols} \sum_{t=2}^{T} Y_{t-2} \hat{\nu}_{t-1}.$$

But, by the property of least squares $\sum_{t=2}^{T} Y_{t-1}\hat{\nu}_t = 0$, hence

$$\sum_{t=2}^{T} \hat{\nu}_t \hat{\nu}_{t-1} \, /T = -\hat{\beta}_{ols} \sum_{t=2}^{T} Y_{t-2} \hat{\nu}_t \, /T = -\hat{\beta}_{ols} \sum_{t=2}^{T} Y_{t-2} Y_t/T + \hat{\beta}_{ols}^2 \sum_{t=2}^{T} Y_{t-2} Y_{t-1}/T$$

and

$$\operatorname{plim} \sum_{t=2}^{T} \hat{\nu}_t \hat{\nu}_{t-1} \, /T = -\frac{\gamma_1}{\gamma_0} \cdot \gamma_2 + \frac{\gamma_1^2}{\gamma_0^2} \cdot \gamma_1 = \frac{\gamma_1 \left(\gamma_1^2 - \gamma_0 \gamma_2\right)}{\gamma_0^2}.$$

Hence,

$$\operatorname{plim} \hat{\rho} = \operatorname{plim} \left(\sum_{t=2}^{T} \hat{\nu}_t \hat{\nu}_{t-1} \, /T \right) /\operatorname{plim} \left(\sum_{t=2}^{T} \hat{\nu}_{t-1}^2 \, /T \right) = \frac{\gamma_1}{\gamma_0} \frac{\left(\gamma_1^2 - \gamma_0 \gamma_2\right)}{\left(\gamma_0^2 - \gamma_1^2\right)}.$$

From part (a), we know that

$$Y_1 = (\beta + \rho) Y_{t-1} - \rho\beta Y_{t-2} + \varepsilon_t$$

multiply both sides by Y_{t-2} and sum and take plim after dividing by T, we get

$$\gamma_2 = (\beta + \rho) \gamma_1 - \rho\beta\gamma_0$$

so that

$$\gamma_0\gamma_2 = (\beta + \rho) \gamma_1\gamma_0 + \rho\beta\gamma_0^2$$

and

$$\gamma_1^2 - \gamma_0\gamma_2 = \gamma_1^2 - (\beta + \rho) \gamma_1\gamma_0 + \rho\beta\gamma_0^2.$$

But from part (a), $\text{plim}\hat{\beta}_{\text{ols}} = \gamma_1/\gamma_0 = (\beta + \rho)/(1 + \rho\beta)$. Substituting

$(\beta + \rho)\gamma_0 = (1 + \rho\beta)\gamma_1$

above we get

$$\gamma_1^2 - \gamma_0\gamma_2 = \gamma_1^2 - (1 + \rho\beta)\,\gamma_1^2 + \rho\beta\gamma_0^2 = \rho\beta\left(\gamma_0^2 - \gamma_1^2\right).$$

Hence,

$$\text{plim}\hat{\rho} = \frac{\gamma_1}{\gamma_0}\cdot\rho\beta = \rho\beta\,(\beta + \rho)\,/\,(1 + \rho\beta)$$

and

$$\begin{aligned}\text{plim}\,(\hat{\rho} - \rho) &= \left(\rho\beta^2 + \rho^2\beta - \rho - \rho^2\beta\right)/\,(1 + \rho\beta) = \rho\,\left(\beta^2 - 1\right)\,/\,(1 + \rho\beta)\\ &= -\text{plim}\left(\hat{\beta}_{\text{ols}} - \beta\right).\end{aligned}$$

The asymptotic bias of $\hat{\rho}$ is negative that of $\hat{\beta}_{\text{ols}}$.

d. Since $d = \sum_{t=2}^{T}(\hat{\nu}_t - \hat{\nu}_{t-1})^2 \,/\, \sum_{t=2}^{T}\hat{\nu}_t^2$ and as $T \to \infty$, $\sum_{t=2}^{T}\hat{\nu}_t^2$ is almost identical to $\sum_{t=2}^{T}\hat{\nu}_{t-1}^2$, then $\text{plim}\ d \approx 2(1 - \text{plim}\hat{\rho})$ where $\hat{\rho}$ was defined in (c). But $\text{plim}\hat{\rho} = \rho - \rho(1-\beta^2)/(1+\rho\beta) = (\rho^2\beta+\rho\beta^2)/(1+\rho\beta) = \rho\beta(\rho+\beta)/(1+\rho\beta)$. Hence, $\text{plim}\ d = 2\left[1 - \frac{\rho\beta(\beta+\rho)}{1+\beta\rho}\right]$.

e. Knowing the true disturbances, the Durbin-Watson statistic would be

$$d^* = \sum_{t=2}^{T}(\nu_t - \nu_{t-1})^2 \,/\, \sum_{t=2}^{T}\nu_t^2$$

and its $\text{plim}\ d^* = 2(1 - \rho)$. This means that from part (d)

$$\begin{aligned}\text{plim}\left(d - d^*\right) &= 2\,(1 - \text{plim}\hat{\rho}) - 2\,(1 - \rho) = 2\,[\rho - \beta\rho\,(\rho + \beta)\,/\,(1 + \rho\beta)]\\ &= 2\left[\rho + \rho^2\beta - \beta\rho^2 - \beta^2\rho\right]/\,(1 + \rho\beta) = 2\rho\left(1 - \beta^2\right)/\,(1 + \rho\beta)\\ &= 2\text{plim}(\hat{\beta}_{\text{ols}} - \beta)\end{aligned}$$

from part (a). The asymptotic bias of the D.W. statistic is twice that of $\hat{\beta}_{\text{ols}}$. The plim d and plim d* and the asymptotic bias in d can be tabulated for various values of ρ and β.

6.4 a. From (6.18) with MA(1) disturbances, we get

$$Y_t = \beta Y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1} \text{ with } |\beta| < 1.$$

In this case,

$$\hat{\beta}_{ols} = \sum_{t=2}^{T} Y_t Y_{t-1} / \sum_{t=2}^{T} Y_{t-1}^2 = \beta + \sum_{t=2}^{T} Y_{t-1}\varepsilon_t / \sum_{t=2}^{T} Y_{t-1}^2$$
$$+ \theta \sum_{t=2}^{T} Y_{t-1}\varepsilon_{t-1} / \sum_{t=2}^{T} Y_{t-1}^2$$

so that

$$\text{plim}\left(\hat{\beta}_{ols} - \beta\right) = \text{plim}\left(\sum_{t=2}^{T} Y_{t-1}\varepsilon_t / T\right) / \text{plim}\left(\sum_{t=2}^{T} Y_{t-1}^2 / T\right)$$
$$+ \theta \text{plim}\left(\sum_{t=2}^{T} Y_{t-1}\varepsilon_{t-1} / T\right) / \text{plim}\left(\sum_{t=2}^{T} Y_{t-1}^2 / T\right).$$

Now the above model can be written as

$$(1 - \beta L) Y_t = (1 + \theta L) \varepsilon_t$$

or

$$Y_t = (1 + \theta L) \sum_{i=0}^{\infty} \beta^i L^i \varepsilon_t$$

$$Y_t = (1 + \theta L) \left(\varepsilon_t + \beta\varepsilon_{t-1} + \beta^2\varepsilon_{t-2} + ..\right)$$

$$Y_t = \varepsilon_t + (\theta + \beta) \left[\varepsilon_{t-1} + \beta\varepsilon_{t-2} + \beta^2\varepsilon_{t-3} + ..\right]$$

From the last expression, it is clear that $E(Y_t) = 0$ and

$$\text{var}(Y_t) = \sigma_\varepsilon^2[1 + (\theta + \beta)^2/(1 - \beta^2)] = \text{plim} \sum_{t=2}^{T} Y_{t-1}^2/T.$$

Also,

$$\sum_{t=2}^{T} Y_{t-1}\varepsilon_t/T = \sum_{t=2}^{T} \varepsilon_t \left[\varepsilon_{t-1} + (\theta + \beta)(\varepsilon_{t-2} + \beta\varepsilon_{t-3} + ..)\right] / T$$

Since the ε_t's are not serially correlated, each term on the right hand side has zero plim. Hence, $\text{plim} \sum_{t=2}^{T} Y_{t-1}\varepsilon_t/T = 0$ and the first term on the right hand side of plim $\left(\hat{\beta}_{ols} - \beta\right)$ is zero. Similarly,

$$\sum_{t=2}^{T} Y_{t-1}\varepsilon_{t-1}/T = \sum_{t=2}^{T} \varepsilon_{t-1}^2/T + (\theta + \beta)$$
$$\left[\sum_{t=2}^{T} \varepsilon_{t-1}\varepsilon_{t-2}/T + \beta \sum_{t=2}^{T} \varepsilon_{t-1}\varepsilon_{t-3}/T + ..\right]$$

which yields plim $\sum_{t=2}^{T} Y_{t-1}\varepsilon_{t-1}/T = \sigma_\varepsilon^2$ since the second term on the right hand side has zero plim. Therefore,

$$\begin{aligned}\text{plim}\left(\hat{\beta}_{ols} - \beta\right) &= \theta\sigma_\varepsilon^2 / \sigma_\varepsilon^2 \left[1 + (\theta + \beta)^2 / \left(1 - \beta^2\right)\right] \\ &= \theta\left(1 - \beta^2\right) / \left(1 - \beta^2 + \theta^2 + \beta^2 + 2\theta\beta\right) \\ &= \theta\left(1 - \beta^2\right) / \left(1 + \theta^2 + 2\theta\beta\right) = \delta\left(1 - \beta^2\right) / (1 + 2\beta\delta)\end{aligned}$$

where $\delta = \theta/(1 + \theta^2)$.

b. The asymptotic bias of $\hat{\beta}_{ols}$ derived in part (a) can be tabulated for various values of β and $0 < \theta < 1$.

c. Let $\hat{\nu}_t = Y_t - \hat{\beta}_{ols}Y_{t-1} = \beta Y_{t-1} - \hat{\beta}_{ols}Y_{t-1} + \nu_t = \nu_t - \left(\hat{\beta}_{ols} - \beta\right) Y_{t-1}$. But

$$\hat{\beta}_{ols} - \beta = \sum_{t=2}^{T} Y_{t-1}\nu_t \Big/ \sum_{t=2}^{T} Y_{t-1}^2.$$

Therefore,

$$\begin{aligned}\sum_{t=2}^{T} \hat{\nu}_t^2 &= \sum_{t=2}^{T} \nu_t^2 + \left(\sum_{t=2}^{T} Y_{t-1}\nu_t\right)^2 \Big/ \sum_{t=2}^{T} Y_{t-1}^2 - 2\left(\Sigma Y_{t-1}\nu_t\right)^2 \Big/ \sum_{t=2}^{T} Y_{t-1}^2 \\ &= \sum_{t=2}^{T} \nu_t^2 - \left(\sum_{t=2}^{T} Y_{t-1}\nu_t\right)^2 \Big/ \sum_{t=2}^{T} Y_{t-1}^2.\end{aligned}$$

But, $\nu_t = \varepsilon_t + \theta\varepsilon_{t-1}$ with $\text{var}(\nu_t) = \sigma_\varepsilon^2(1 + \theta^2)$. Hence,

$$\text{plim} \sum_{t=2}^{T} \nu_t^2/T = \sigma_\varepsilon^2 + \theta^2\sigma_\varepsilon^2 = \sigma_\varepsilon^2(1 + \theta^2).$$

Also,

$$\text{plim}\sum_{t=2}^{T} Y_{t-1}\nu_t/T = \text{plim}\sum_{t=2}^{T} Y_{t-1}\varepsilon_t/T + \theta\text{plim}\sum_{t=2}^{T} Y_{t-1}\varepsilon_t/T$$
$$= 0 + \theta\sigma_\varepsilon^2 = \theta\sigma_\varepsilon^2$$

from part (a). Therefore,

$$\text{plim}\sum_{t=2}^{T} \hat{\nu}_t^2/T = \sigma_\varepsilon^2\left(1+\theta^2\right) - \theta^2\sigma_\varepsilon^4/\sigma_\varepsilon^2\left[1+(\theta+\beta)^2/\left(1-\beta^2\right)\right]$$
$$= \sigma_\varepsilon^2\left[1+\theta^2-\theta^2\left(1-\beta^2\right)/\left(1-\theta^2+2\theta\beta\right)\right]$$
$$= \sigma_\varepsilon^2\left[1+\theta^2-\theta\delta\left(1-\beta^2\right)/(1+2\delta\beta)\right] = \sigma_\varepsilon^2\left[1+\theta^2-\theta\theta^*\right]$$

where $\delta = \frac{\theta}{(1+\theta^2)}$ and $\theta^* = \delta(1-\beta^2)/(1+2\beta\delta)$.

6.5 a. The stata commands to generate the Durbin h statistic from scratch are the following:

```
. gen lc_lag=l.lc
(1 missing value generated)
. reg lc ly lc_lag
```

Source	SS	df	MS		
Model	5.02839276	2	2.51419638	Number of obs =	48
Residual	.008947057	45	.000198823	F(2, 45) =	12645.37
Total	5.03733982	47	.107177443	Prob > F =	0.0000
				R-squared =	0.9982
				Adj R-squared =	0.9981
				Root MSE =	.0141

lc	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ly	.3091105	.0816785	3.78	0.000	.1446015	.4736194
lc_lag	.7045042	.0765288	9.21	0.000	.5503673	.8586411
_cons	-.1475238	.0870474	-1.69	0.097	-.3228462	.0277986

```
. display .0765288^2
.00585666
. predict miu, resid
(1 missing value generated)
. reg miu l.miu, noconstant
```

Source	SS	df	MS			
				Number of obs	=	47
				F(1, 46)	=	15.78
Model	.002284571	1	.002284571	Prob > F	=	0.0002
Residual	.006660987	46	.000144804	R-squared	=	0.2554
				Adj R-squared	=	0.2392
Total	.008945558	47	.000190331	Root MSE	=	.01203

miu	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
miu						
L1.	.5065964	.1275411	3.97	0.000	.2498695	.7633234

```
. * Durbin's h
. display .5065964*(48*(1-.00585666))^0.5
3.4995099
```

This is asymptotically distributed as N(0,1) under the null hypothesis of $\rho = 0$. This rejects H_o indicating the presence of serial correlation.

b. The EViews output for the Breusch and Godfrey test for *first*-order serial correlation is given below. The back up regression is given below the test statistics. This regresses the OLS residuals on their lagged value and the regressors in the model including the lagged dependent variable. This yields an $R^2 = 0.278576$. The number of observations is 48. Therefore, the LM statistic $= TR^2 = 48(0.278576) = 13.372$. This is asymptotically distributed as Chi-Square(1) under the null hypothesis of $\rho = 0$. The p-value is 0.0003, and we reject the null hypothesis.

Breusch–Godfrey Serial Correlation LM Test:

F-statistic	16.99048	Prob. F(1,44)	0.0002
Obs*R-squared	13.37165	Prob. Chi-Square(1)	0.0003

Test Equation:

Dependent Variable: RESID

Method: Least Squares

Sample: 1960 2007

Included observations: 48

Presample missing value lagged residuals set to zero.

	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.061130	0.076227	-0.801941	0.4269
LNC(-1)	-0.081685	0.068658	-1.189747	0.2405
LNY	0.086867	0.073256	1.185810	0.2421
RESID(-1)	0.552693	0.134085	4.121951	0.0002

R-squared	0.278576	Mean dependent var	-5.00E-16
Adjusted R-squared	0.229388	S.D. dependent var	0.013797
S.E. of regression	0.012112	Akaike info criterion	-5.909614
Sum squared resid	0.006455	Schwarz criterion	-5.753680
Log likelihood	145.8307	Hannan-Quinn criter.	-5.850686
F-statistic	5.663494	Durbin-Watson stat	1.773229
Prob(F-statistic)	0.002272		

c. The EViews output for the Breusch and Godfrey test for *second*-order serial correlation is given below. The back up regression is given below the test statistics. This regresses the OLS residuals e_t on e_{t-1} and e_{t-2} and the regressors in the model including the lagged dependent variable. This yields an $R^2 = 0.284238$. The number of observations is 48. Therefore, the LM statistic $= TR^2 = 48(0.284238) = 13.643$.

This is asymptotically distributed as Chi-Square(2) under the null hypothesis of no second-order serial correlation. The p-value is 0.0011, and we reject the null hypothesis.

Breusch–Godfrey Serial Correlation LM Test:

F-statistic	8.537930	Prob. F(2,43)	0.0008
Obs*R-squared	13.64344	Prob. Chi-Square(2)	0.0011

Test Equation:

Dependent Variable: RESID

Method: Least Squares

Sample: 1960 2007

Included observations: 48

Presample missing value lagged residuals set to zero.

	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.049646	0.079289	-0.626136	0.5345
LNC(-1)	-0.067455	0.073355	-0.919582	0.3629
LNY	0.071648	0.078288	0.915192	0.3652
RESID(-1)	0.591123	0.150313	3.932602	0.0003
RESID(-2)	-0.092963	0.159390	-0.583241	0.5628

R-squared	0.284238	Mean dependent var	-5.00E-16
Adjusted R-squared	0.217656	S.D. dependent var	0.013797
S.E. of regression	0.012204	Akaike info criterion	-5.875827
Sum squared resid	0.006404	Schwarz criterion	-5.680910
Log likelihood	146.0198	Hannan-Quinn criter.	-5.802167
F-statistic	4.268965	Durbin-Watson stat	1.867925
Prob(F-statistic)	0.005364		

6.6 U.S. Gasoline Data, Table 4.2.

a. STATIC MODEL

Dependent Variable: QMG_CAR

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	3	0.35693	0.11898	32.508	0.0001
Error	34	0.12444	0.00366		
C Total	37	0.48137			

Root MSE	0.06050	R-square	0.7415
Dep Mean	-0.25616	Adj R-sq	0.7187
C.V.	-23.61671		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob > |T|
INTERCEP	1	-5.853977	3.10247647	-1.887	0.0677
RGNP_POP	1	-0.690460	0.29336969	-2.354	0.0245
CAR_POP	1	0.288735	0.27723429	1.041	0.3050
PMG_PGNP	1	-0.143127	0.07487993	-1.911	0.0644

DYNAMIC MODEL

Autoreg Procedure

Dependent Variable = QMG_CAR

Ordinary Least Squares Estimates

SSE	0.014269	DFE	32
MSE	0.000446	Root MSE	0.021117
SBC	-167.784	AIC	-175.839
Reg Rsq	0.9701	Total Rsq	0.9701
Durbin h	2.147687	PROB>h	0.0159

Godfrey's Serial Correlation Test

Alternative	LM	Prob>LM
AR(+ 1)	4.6989	0.0302
AR(+ 2)	5.1570	0.0759

Variable	DF	B Value	Std Error	t Ratio	Approx Prob
Intercept	1	0.523006	1.1594	0.451	0.6550
RGNP_POP	1	0.050519	0.1127	0.448	0.6571
CAR_POP	1	-0.106323	0.1005	-1.058	0.2981
PMG_PGNP	1	-0.072884	0.0267	-2.733	0.0101
LAG_DEP	1	0.907674	0.0593	15.315	0.0001

c. LM Test for AR(1) by BREUSCH & GODFREY

Dependent Variable: RESID

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	5	0.00178	0.00036	0.863	0.5172
Error	30	0.01237	0.00041		
C Total	35	0.01415			

Root MSE	0.02031	R-square	0.1257
Dep Mean	-0.00030	Adj R-sq	-0.0200
C.V.	-6774.59933		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob > |T|
INTERCEP	1	-0.377403	1.15207945	-0.328	0.7455
RESID_1	1	0.380254	0.18431559	2.063	0.0479
RGNP_POP	1	-0.045964	0.11158093	-0.412	0.6833
CAR_POP	1	0.031417	0.10061545	0.312	0.7570
PMG_PGNP	1	0.007723	0.02612020	0.296	0.7695
LAG_DEP	1	-0.034217	0.06069319	-0.564	0.5771

SAS PROGRAM

```
Data RAWDATA;
Input Year CAR QMG PMG POP RGNP PGNP;
Cards;
Data USGAS; set RAWDATA;
LNQMG=LOG(QMG);
LNCAR=LOG(CAR);
LNPOP=LOG(POP);
LNRGNP=LOG(RGNP);
LNPGNP=LOG(PGNP);
LNPMG=LOG(PMG);
QMG_CAR=LOG(QMG/CAR);
RGNP_POP=LOG(RGNP/POP);
CAR_POP=LOG(CAR/POP);
PMG_PGNP=LOG(PMG/PGNP);
LAG_DEP=LAG(QMG_CAR);

Proc reg data=USGAS;
   Model QMG_CAR=RGNP_POP CAR_POP PMG_PGNP;
   TITLE ' STATIC MODEL';

Proc autoreg data=USGAS;
      Model QMG_CAR=RGNP_POP CAR_POP PMG_PGNP
                    LAG_DEP/LAGDEP=LAG_DEP godfrey=2;
   OUTPUT OUT=MODEL2 R=RESID;
   TITLE ' DYNAMIC MODEL';
RUN;

DATA DW_DATA;    SET MODEL2;
RESID_1=LAG(RESID);
PROC REG DATA=DW_DATA;
        MODEL RESID=RESID_1 RGNP_POP CAR_POP PMG_PGNP LAG_DEP;
TITLE 'LM Test for AR(1) by BREUSCH & GODFREY';
RUN;
```

6.7 a. Unrestricted Model

Autoreg Procedure

Dependent Variable = QMG_CAR

Ordinary Least Squares Estimates

SSE	0.031403	DFE	22
MSE	0.001427	Root MSE	0.037781
SBC	-96.1812	AIC	-110.839
Reg Rsq	0.9216	Total Rsq	0.9216
Durbin-Watson	0.5683		

Variable	DF	B Value	Std Error	t Ratio	Approx Prob
Intercept	1	-7.46541713	3.0889	-2.417	0.0244
RGNP_POP	1	-0.58684334	0.2831	-2.073	0.0501
CAR_POP	1	0.24215182	0.2850	0.850	0.4046
PMG_PGNP	1	-0.02611161	0.0896	-0.291	0.7734
PMPG_1	1	-0.15248735	0.1429	-1.067	0.2975
PMPG_2	1	-0.13752842	0.1882	-0.731	0.4726
PMPG_3	1	0.05906629	0.2164	0.273	0.7875
PMPG_4	1	-0.21264747	0.2184	-0.974	0.3408
PMPG_5	1	0.22649780	0.1963	1.154	0.2609
PMPG_6	1	-0.41142284	0.1181	-3.483	0.0021

b. Almon Lag $(S = 6, P = 2)$

PDLREG Procedure

Dependent Variable = QMG_CAR

Ordinary Least Squares Estimates

SSE	0.04017	DFE	26
MSE	0.001545	Root MSE	0.039306
SBC	-102.165	AIC	-110.96
Reg Rsq	0.8998	Total Rsq	0.8998
Durbin-Watson	0.5094		

Variable	DF	B Value	Std Error	t Ratio	Approx Prob
Intercept	1	-5.06184299	2.9928	-1.691	0.1027
RGNP_POP	1	-0.35769028	0.2724	-1.313	0.2006
CAR_POP	1	0.02394559	0.2756	0.087	0.9314
PMG_PGNP**0	1	-0.24718333	0.0340	-7.278	0.0001
PMG_PGNP**1	1	-0.05979404	0.0439	-1.363	0.1847
PMG_PGNP**2	1	-0.10450923	0.0674	-1.551	0.1331

Variable	Parameter Value	Std Error	t Ratio	Approx Prob
PMG_PGNP(0)	-0.11654	0.045	-2.58	0.0159
PMG_PGNP(1)	-0.07083	0.020	-3.46	0.0019
PMG_PGNP(2)	-0.04792	0.024	-1.98	0.0584
PMG_PGNP(3)	-0.04781	0.028	-1.74	0.0944
PMG_PGNP(4)	-0.07052	0.021	-3.33	0.0026
PMG_PGNP(5)	-0.11603	0.021	-5.40	0.0001
PMG_PGNP(6)	-0.18434	0.054	-3.42	0.0021

Estimate of Lag Distribution

```
Variable       -0.184                                    0

PMG_PGNP(0)    |            **************************|
PMG_PGNP(1)    |                      ****************|
PMG_PGNP(2)    |                           ***********|
PMG_PGNP(3)    |                           ***********|
PMG_PGNP(4)    |                      ****************|
PMG_PGNP(5)    |            **************************|
PMG_PGNP(6)    | *************************************|
```

c. Almon Lag (S=4,P=2)

PDLREG Procedure

Dependent Variable = QMG_CAR

Ordinary Least Squares Estimates

SSE	0.065767	DFE	28
MSE	0.002349	Root MSE	0.048464
SBC	-94.7861	AIC	-103.944
Reg Rsq	0.8490	Total Rsq	0.8490
Durbin-Watson	0.5046		

Variable	DF	B Value	Std Error	t Ratio	Approx Prob
Intercept	1	-6.19647990	3.6920	-1.678	0.1044
RGNP_POP	1	-0.57281368	0.3422	-1.674	0.1053
CAR_POP	1	0.21338192	0.3397	0.628	0.5351
PMG_PGNP**0	1	-0.19423745	0.0414	-4.687	0.0001
PMG_PGNP**1	1	-0.06534647	0.0637	-1.026	0.3138
PMG_PGNP**2	1	0.03085234	0.1188	0.260	0.7970

Variable	Parameter Value	Std Error	t Ratio	Approx Prob
PMG_PGNP(0)	-0.02905	0.070	-0.41	0.6834
PMG_PGNP(1)	-0.07445	0.042	-1.78	0.0851
PMG_PGNP(2)	-0.10336	0.061	-1.70	0.0999
PMG_PGNP(3)	-0.11578	0.033	-3.51	0.0015
PMG_PGNP(4)	-0.11170	0.092	-1.22	0.2329

Estimate of Lag Distribution

```
Variable       -0.116                                        0

PMG_PGNP(0)    |                               **********|
PMG_PGNP(1)    |               **************************|
PMG_PGNP(2)    |    *************************************|
PMG_PGNP(3)    |*****************************************|
PMG_PGNP(4)    | ****************************************|
```

ALMON LAG(S=8,P=2)

PDLREG Procedure

Dependent Variable = QMG_CAR

Ordinary Least Squares Estimates

SSE	0.020741	DFE	24
MSE	0.000864	Root MSE	0.029398
SBC	-112.761	AIC	-121.168
Reg Rsq	0.9438	Total Rsq	0.9438
Durbin-Watson	0.9531		

Variable	DF	B Value	Std Error	t Ratio	Approx Prob
Intercept	1	-7.71363805	2.3053	-3.346	0.0027
RGNP_POP	1	-0.53016065	0.2041	-2.597	0.0158
CAR_POP	1	0.17117375	0.2099	0.815	0.4229
PMG_PGNP**0	1	-0.28572518	0.0267	-10.698	0.0001
PMG_PGNP**1	1	-0.09282151	0.0417	-2.225	0.0358
PMG_PGNP**2	1	-0.12948786	0.0512	-2.527	0.0185

Variable	Parameter Value	Std Error	t Ratio	Approx Prob
PMG_PGNP(0)	-0.11617	0.028	-4.09	0.0004
PMG_PGNP(1)	-0.07651	0.016	-4.73	0.0001
PMG_PGNP(2)	-0.05160	0.015	-3.34	0.0027
PMG_PGNP(3)	-0.04145	0.018	-2.30	0.0301
PMG_PGNP(4)	-0.04605	0.017	-2.63	0.0146
PMG_PGNP(5)	-0.06541	0.013	-4.85	0.0001
PMG_PGNP(6)	-0.09953	0.012	-8.10	0.0001
PMG_PGNP(7)	-0.14841	0.025	-5.97	0.0001
PMG_PGNP(8)	-0.21204	0.047	-4.53	0.0001

Estimate of Lag Distribution

```
Variable       -0.212                                      0

PMG_PGNP(0)    |                   **********************|
PMG_PGNP(1)    |                          ***************|
PMG_PGNP(2)    |                               **********|
PMG_PGNP(3)    |                                 ********|
PMG_PGNP(4)    |                                *********|
PMG_PGNP(5)    |                            *************|
PMG_PGNP(6)    |                      *******************|
PMG_PGNP(7)    |            *****************************|
PMG_PGNP(8)    |*****************************************|
```

d. Third Degree Polynomial Almon Lag$(\mathrm{S}=6, \mathrm{P}=3)$

PDLREG Procedure

Dependent Variable = QMG_CAR

Ordinary Least Squares Estimates

SSE	0.034308	DFE	25
MSE	0.001372	Root MSE	0.037045
SBC	-103.747	AIC	-114.007
Reg Rsq	0.9144	Total Rsq	0.9144
Durbin-Watson	0.6763		

Variable	DF	B Value	Std Error	t Ratio	Approx Prob
Intercept	1	-7.31542415	3.0240	-2.419	0.0232
RGNP_POP	1	-0.57343614	0.2771	-2.069	0.0490
CAR_POP	1	0.23462358	0.2790	0.841	0.4084
PMG_PGNP**0	1	-0.24397597	0.0320	-7.613	0.0001
PMG_PGNP**1	1	-0.07041380	0.0417	-1.690	0.1035
PMG_PGNP**2	1	-0.11318734	0.0637	-1.778	0.0876
PMG_PGNP**3	1	-0.19730731	0.0955	-2.067	0.0493

Variable	Parameter Value	Std Error	t Ratio	Approx Prob
PMG_PGNP(0)	-0.03349	0.059	-0.57	0.5725
PMG_PGNP(1)	-0.14615	0.041	-3.54	0.0016
PMG_PGNP(2)	-0.12241	0.043	-2.87	0.0082
PMG_PGNP(3)	-0.04282	0.026	-1.64	0.1130
PMG_PGNP(4)	-0.01208	0.045	0.27	0.7890
PMG_PGNP(5)	-0.03828	0.043	-0.90	0.3788
PMG_PGNP(6)	-0.27443	0.067	-4.10	0.0004

Estimate of Lag Distribution

```
Variable        -0.274                              0.0121

PMG_PGNP(0)     |                             **** |    |
PMG_PGNP(1)     |             ******************** |    |
PMG_PGNP(2)     |                 **************** |    |
PMG_PGNP(3)     |                            ***** |    |
PMG_PGNP(4)     |                                  | ** |
PMG_PGNP(5)     |                              ****|    |
PMG_PGNP(6)     |**********************************|    |
```

e. Almon Lag$(S = 6, P = 2)$ with Near End-Point Restriction

PDLREG Procedure

Dependent Variable = QMG_CAR

Ordinary Least Squares Estimates

SSE	0.046362	DFE	27
MSE	0.001717	Root MSE	0.041438
SBC	-101.043	AIC	-108.372
Reg Rsq	0.8843	Total Rsq	0.8843
Durbin-Watson	0.5360		

Variable	DF	B Value	Std Error	t Ratio	Approx Prob
Intercept	1	-3.81408793	3.0859	-1.236	0.2271
RGNP_POP	1	-0.28069982	0.2843	-0.988	0.3322
CAR_POP	1	-0.05768233	0.2873	-0.201	0.8424
PMG_PGNP**0	1	-0.21562744	0.0317	-6.799	0.0001
PMG_PGNP**1	1	-0.07238330	0.0458	-1.581	0.1255
PMG_PGNP**2	1	0.02045576	0.0268	0.763	0.4519

Restriction	DF	L Value	Std Error	t Ratio	Approx Prob
PMG_PGNP(−1)	-1	0.03346081	0.0176	1.899	0.0683

Variable	Parameter Value	Std Error	t Ratio	Approx Prob
PMG_PGNP(0)	-0.02930	0.013	-2.31	0.0286
PMG_PGNP(1)	-0.05414	0.020	-2.75	0.0105
PMG_PGNP(2)	-0.07452	0.021	-3.49	0.0017
PMG_PGNP(3)	-0.09043	0.018	-4.91	0.0001
PMG_PGNP(4)	-0.10187	0.015	-6.80	0.0001
PMG_PGNP(5)	-0.10886	0.022	-4.88	0.0001
PMG_PGNP(6)	-0.11138	0.042	-2.66	0.0130

Estimate of Lag Distribution

```
Variable       -0.111                                    0

PMG_PGNP(0)    |                             ***********|
PMG_PGNP(1)    |                    ********************|
PMG_PGNP(2)    |             ***************************|
PMG_PGNP(3)    |       *********************************|
PMG_PGNP(4)    |   *************************************|
PMG_PGNP(5)    | ***************************************|
PMG_PGNP(6)    |****************************************|
```

ALMON LAG(S=6,P=2) with FAR END-POINT RESTRICTION

PDLREG Procedure

Dependent Variable = QMG_CAR

Ordinary Least Squares Estimates

SSE	0.050648	DFE	27
MSE	0.001876	Root MSE	0.043311
SBC	-98.2144	AIC	-105.543
Reg Rsq	0.8736	Total Rsq	0.8736
Durbin-Watson	0.5690		

Variable	DF	B Value	Std Error	t Ratio	Approx Prob
Intercept	1	-6.02848892	3.2722	-1.842	0.0764
RGNP_POP	1	-0.49021381	0.2948	-1.663	0.1079
CAR_POP	1	0.15878127	0.2982	0.532	0.5988
PMG_PGNP**0	1	-0.20851840	0.0337	-6.195	0.0001
PMG_PGNP**1	1	-0.00242944	0.0418	-0.058	0.9541
PMG_PGNP**2	1	0.06159671	0.0240	2.568	0.0161

Restriction	DF	L Value	Std Error	t Ratio	Approx Prob
PMG_PGNP(7)	-1	0.03803694	0.0161	2.363	0.0256

Variable	Parameter Value	Std Error	t Ratio	Approx Prob
PMG_PGNP(0)	-0.04383	0.039	-1.12	0.2730
PMG_PGNP(1)	-0.07789	0.022	-3.49	0.0017
PMG_PGNP(2)	-0.09852	0.016	-6.20	0.0001
PMG_PGNP(3)	-0.10570	0.018	-5.90	0.0001
PMG_PGNP(4)	-0.09943	0.020	-5.01	0.0001
PMG_PGNP(5)	-0.07973	0.018	-4.43	0.0001
PMG_PGNP(6)	-0.04659	0.011	-4.05	0.0004

```
                    Estimate of Lag Distribution

Variable       -0.106                                    0

PMG_PGNP(0)    |                       ******************|
PMG_PGNP(1)    |           ******************************|
PMG_PGNP(2)    |   **************************************|
PMG_PGNP(3)    |*****************************************|
PMG_PGNP(4)    |  ***************************************|
PMG_PGNP(5)    |          *******************************|
PMG_PGNP(6)    |                       ******************|
```

SAS PROGRAM

```
Data RAWDATA;
Input Year CAR QMG PMG POP RGNP PGNP;
Cards;
Data USGAS; set RAWDATA;
LNQMG=LOG(QMG);
LNCAR=LOG(CAR);
LNPOP=LOG(POP);
LNRGNP=LOG(RGNP);
LNPGNP=LOG(PGNP);
LNPMG=LOG(PMG);
QMG_CAR=LOG(QMG/CAR);
RGNP_POP=LOG(RGNP/POP);
CAR_POP=LOG(CAR/POP);
PMG_PGNP=LOG(PMG/PGNP);
PMPG_1=LAG1(PMG_PGNP);
PMPG_2=LAG2(PMG_PGNP);
PMPG_3=LAG3(PMG_PGNP);
PMPG_4=LAG4(PMG_PGNP);
PMPG_5=LAG5(PMG_PGNP);
```

```
PMPG_6=LAG6(PMG_PGNP);

Proc autoreg data=USGAS;
 Model QMG_CAR=RGNP_POP CAR_POP PMG_PGNP PMPG_1 PMPG_2 PMPG_3
                        PMPG_4 PMPG_5 PMPG_6;
    TITLE 'UNRESTRICTED MODEL';

PROC PDLREG DATA=USGAS;
       MODEL QMG_CAR=RGNP_POP CAR_POP PMG_PGNP(6,2);
TITLE 'ALMON LAG(S=6,P=2)';

PROC PDLREG DATA=USGAS;
       MODEL QMG_CAR=RGNP_POP CAR_POP PMG_PGNP(4,2);
TITLE 'ALMON LAG(S=4,P=2)';

PROC PDLREG DATA=USGAS;
       MODEL QMG_CAR=RGNP_POP CAR_POP PMG_PGNP(8,2);
TITLE 'ALMON LAG(S=8,P=2)';

PROC PDLREG DATA=USGAS;
       MODEL QMG_CAR=RGNP_POP CAR_POP PMG_PGNP(6,3);
TITLE 'Third Degree Polynomial ALMON LAG(S=6,P=3)';

PROC PDLREG DATA=USGAS;
       MODEL QMG_CAR=RGNP_POP CAR_POP PMG_PGNP(6,2,,FIRST);
TITLE 'ALMON LAG(S=6,P=2) with NEAR END-POINT RESTRICTION';

PROC PDLREG DATA=USGAS;
       MODEL QMG_CAR=RGNP_POP CAR_POP PMG_PGNP(6,2,,LAST);
TITLE 'ALMON LAG(S=6,P=2) with FAR END-POINT RESTRICTION';
RUN;
```

CHAPTER 7
The General Linear Model: The Basics

7.1 *Invariance of the fitted values and residuals to non-singular transformations of the independent variables.*

The regression model in (7.1) can be written as $y = XCC^{-1}\beta + u$ where C is a non-singular matrix. Let $X^* = XC$, then $y = X^*\beta^* + u$ where $\beta^* = C^{-1}\beta$.

a. $P_{X^*} = X^*(X^{*\prime}X^*)^{-1}X^{*\prime} = XC[C'X'XC]^{-1}C'X' = XCC^{-1}(X'X)^{-1}C'^{-1}C'X' = P_X$.

Hence, the regression of y on X^* yields

$$\hat{y} = X^*\hat{\beta}^*_{ols} = P_{X^*}y = P_Xy = X\hat{\beta}_{ols}$$

which is the same fitted values as those from the regression of y on X. Since the dependent variable y is the same, the residuals from both regressions will be the same.

b. Multiplying each X by a constant is equivalent to post-multiplying the matrix X by a diagonal matrix C with a typical k-th element c_k. Each X_k will be multiplied by the constant c_k for $k = 1, 2, .., K$. This diagonal matrix C is non-singular. Therefore, using the results in part (a), the fitted values and the residuals will remain the same.

c. In this case, $X = [X_1, X_2]$ is of dimension nx2 and

$$X^* = [X_1 - X_2,\ X_1 + X_2] = [X_1, X_2]\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} = XC$$

where $C = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$ is non-singular with $C^{-1} = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$. Hence, the results of part (a) apply, and we get the same fitted values and residuals when we regress y on $(X_1 - X_2)$ and $(X_1 + X_2)$ as in the regression of y on X_1 and X_2.

B.H. Baltagi, *Solutions Manual for Econometrics*, Springer Texts in Business and Economics, DOI 10.1007/978-3-642-54548-1_7,

7.2 *The FWL Theorem.*

a. The inverse of a partitioned matrix

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

is given by

$$A^{-1} = \begin{bmatrix} A_{11}^{-1} + A_{11}^{-1}A_{12}B_{22}A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}B_{22} \\ -B_{22}A_{21}A_{11}^{-1} & B_{22} \end{bmatrix}$$

where $B_{22} = \left(A_{22} - A_{21}A_{11}^{-1}A_{12}\right)^{-1}$. From (7.9), we get

$$B_{22} = \left(X_2'X_2 - X_2'X_1\left(X_1'X_1\right)^{-1}X_1'X_2\right)^{-1} = \left(X_2'X_2 - X_2'P_{X_1}X_2\right)^{-1}$$
$$= \left(X_2'\overline{P}_{X_1}X_2\right)^{-1}.$$

Also, $-B_{22}A_{21}A_{11}^{-1} = -\left(X_2'\overline{P}_{X_1}X_2\right)^{-1}X_2'X_1\left(X_1'X_1\right)^{-1}$. Hence, from (7.9), we solve for $\hat{\beta}_{2,ols}$ to get

$$\hat{\beta}_{2,ols} = -B_{22}A_{21}A_{11}^{-1}X_1'y + B_{22}X_2'y$$

which yields

$$\hat{\beta}_{2,ols} = -\left(X_2'\overline{P}_{X_1}X_2\right)^{-1}X_2'P_{X_1}y + \left(X_2'\overline{P}_{X_1}X_2\right)^{-1}X_2'y$$
$$= \left(X_2'\overline{P}_{X_1}X_2\right)^{-1}X_2'\overline{P}_{X_1}y$$

as required in (7.10).

b. Alternatively, one can write (7.9) as

$$\left(X_1'X_1\right)\hat{\beta}_{1,ols} + \left(X_1'X_2\right)\hat{\beta}_{2,ols} = X_1'y$$
$$\left(X_2'X_1\right)\hat{\beta}_{1,ols} + \left(X_2'X_2\right)\hat{\beta}_{2,ols} = X_2'y.$$

Solving for $\hat{\beta}_{1,ols}$ in terms of $\hat{\beta}_{2,ols}$ by multiplying the first equation by $\left(X_1'X_1\right)^{-1}$ we get

$$\hat{\beta}_{1,ols} = \left(X_1'X_1\right)^{-1}X_1'y - \left(X_1'X_1\right)^{-1}X_1'X_2\hat{\beta}_{2,ols}$$
$$= \left(X_1'X_1\right)^{-1}X_1'\left(y - X_2\hat{\beta}_{2,ols}\right).$$

Substituting $\hat{\beta}_{1,ols}$ in the second equation, we get

$$X_2'X_1\left(X_1'X_1\right)^{-1}X_1'y - X_2'P_{X_1}X_2\hat{\beta}_{2,ols} + \left(X_2'X_2\right)\hat{\beta}_{2,ols} = X_2'y.$$

Collecting terms, we get $\left(X_2'\overline{P}_{X_1}X_2\right)\hat{\beta}_{2,ols} = X_2'\overline{P}_{X_1}y$.

Hence, $\hat{\beta}_{2,ols} = \left(X_2'\overline{P}_{X_1}X_2\right)^{-1}X_2'\overline{P}_{X_1}y$ as given in (7.10).

c. In this case, $X = [\iota_n,\ X_2]$ where ι_n is a vector of ones of dimension n. $P_{X_1} = \iota_n\left(\iota_n'\iota_n\right)^{-1}\iota_n' = \iota_n\iota_n'/n = J_n/n$ where $J_n = \iota_n\iota_n'$ is a matrix of ones of dimension n. But $\iota_n'y = \sum_{i=1}^{n} y_i$ and $\iota_n'y/n = \bar{y}$. Hence, $\overline{P}_{X_1} = I_n - P_{X_1} = I_n - J_n/n$ and $\overline{P}_{X_1}y = (I_n - J_n/n)y$ has a typical element $(y_i - \bar{y})$. From the FWL Theorem, $\hat{\beta}_{2,ols}$ can be obtained from the regression of $(y_i - \bar{y})$ on the set of variables in X_2 expressed as deviations from their respective means, i.e., $\overline{P}_{X_1}X_2 = (I_n - J_n/n)\,X_2$.

From part (b),

$$\hat{\beta}_{1,ols} = \left(X_1'X_1\right)^{-1}X_1'\left(y - X_2\hat{\beta}_{2,ols}\right) = \left(\iota_n'\iota_n\right)^{-1}\iota_n'\left(y - X_2\hat{\beta}_{2,ols}\right)$$
$$= \frac{\iota_n'}{n}\left(y - X_2\hat{\beta}_{2,ols}\right) = \bar{y} - \overline{X}_2'\hat{\beta}_{2,ols}$$

where $\overline{X}_2' = \iota_n'X_2/n$ is the vector of sample means of the independent variables in X_2.

7.3 $D_i = (0, 0, .., 1, 0, .., 0)'$ where all the elements of this nx1 vector are zeroes except for the i-th element which takes the value 1. In this case, $P_{D_i} = D_i\left(D_i'D_i\right)^{-1}D_i' = D_iD_i'$ which is a matrix of zeroes except for the i-th diagonal element which takes the value 1. Hence, $I_n - P_{D_i}$ is an identity matrix except for the i-th diagonal element which takes the value zero. Therefore, $(I_n - P_{D_i})y$ returns the vector y except for the i-th element which is zero. Using the FWL Theorem, the OLS regression

$$y = X\beta + D_i\gamma + u$$

yields the same estimates as $(I_n - P_{D_i})y = (I_n - P_{D_i})X\beta + (I_n - P_{D_i})u$ which can be rewritten as $\tilde{y} = \tilde{X}\beta + \tilde{u}$ with $\tilde{y} = (I_n - P_{D_i})y$, $\tilde{X} = (I_n - P_{D_i})X$.

The OLS normal equations yield $(\tilde{X}'\tilde{X})\hat{\beta}_{ols} = \tilde{X}'\tilde{y}$ and the i-th OLS normal equation can be ignored since it gives $0'\hat{\beta}_{ols} = 0$. Ignoring the i-th observation equation yields $(X^{*\prime}X^{*})\hat{\beta}_{ols} = X^{*\prime}y^{*}$ where X^{*} is the matrix X without the i-th observation and y^{*} is the vector y without the i-th observation. The FWL Theorem also states that the residuals from $\tilde{y}$ on $\tilde{X}$ are the same as those from y on X and D_i. For the i-th observation, $\tilde{y}_i = 0$ and $\tilde{x}_i = 0$. Hence the i-th residual must be zero. This also means that the i-th residual in the original regression with the dummy variable D_i is zero, i.e., $y_i - x_i'\hat{\beta}_{ols} - \hat{\gamma}_{ols} = 0$. Rearranging terms, we get $\hat{\gamma}_{ols} = y_i - x_i'\hat{\beta}_{ols}$. In other words, $\hat{\gamma}_{ols}$ is the forecasted OLS residual for the i-th observation from the regression of y^{*} on X^{*}. The i-th observation was excluded from the estimation of $\hat{\beta}_{ols}$ by the inclusion of the dummy variable D_i.

7.5 If $u \sim N(0, \sigma^2 I_n)$ then $(n-K)s^2/\sigma^2 \sim \chi^2_{n-K}$. In this case,

a. $E[(n-K)s^2/\sigma^2] = E\left(\chi^2_{n-K}\right) = n-K$ since the expected value of a χ^2 random variable with $(n-K)$ degrees of freedom is $(n-K)$. Hence,

$$\left[(n-K)/\sigma^2\right]E(s^2) = (n-K) \text{ or } E(s^2) = \sigma^2.$$

b. $\mathrm{var}[(n-K)s^2/\sigma^2] = \mathrm{var}\left(\chi^2_{n-K}\right) = 2(n-K)$ since the variance of a χ^2 random variable with $(n-K)$ degrees of freedom is $2(n-K)$. Hence,

$$\left[(n-K)^2/\sigma^4\right]\ \mathrm{var}(s^2) = 2(n-K) \text{ or } \mathrm{var}(s^2) = 2\sigma^4/(n-K).$$

7.6 a. Using the results in problem 7.4, we know that $\hat{\sigma}^2_{mle} = e'e/n = (n-K)s^2/n$. Hence,

$$E\left(\hat{\sigma}^2_{mle}\right) = (n-K)\ E(s^2)/n = (n-K)\sigma^2/n.$$

This means that $\hat{\sigma}^2_{mle}$ is biased for σ^2, but asymptotically unbiased. The bias is equal to $-K\sigma^2/n$ which goes to zero as $n \to \infty$.

b. $\text{var}\left(\hat{\sigma}^2_{mle}\right) = (n-K)^2\,\text{var}(s^2)/n^2 = (n-K)^2 2\sigma^4/n^2(n-K) = 2(n-K)\sigma^4/n^2$

and

$$\text{MSE}\left(\hat{\sigma}^2_{mle}\right) = \text{Bias}^2\left(\hat{\sigma}^2_{mle}\right) + \text{var}\left(\hat{\sigma}^2_{mle}\right) = K^2\sigma^4/n^2 + 2(n-K)\sigma^4/n^2$$
$$= \left(K^2 + 2n - 2K\right)\sigma^4/n^2.$$

c. Similarly, $\tilde{\sigma}^2 = e'e/r = (n-K)s^2/r$ with $E(\tilde{\sigma}^2) = (n-K)\sigma^2/r$ and

$$\text{var}(\tilde{\sigma}^2) = (n-K)^2\,\text{var}(s^2)/r^2 = (n-K)^2 2\sigma^4/r^2(n-K) = 2(n-K)\sigma^4/r^2$$

$$\text{MSE}(\tilde{\sigma}^2) = \text{Bias}^2(\tilde{\sigma}^2) + \text{var}(\tilde{\sigma}^2) = (n-K-r)^2\sigma^4/r^2 + 2(n-K)\sigma^4/r^2.$$

Minimizing $\text{MSE}(\tilde{\sigma}^2)$ with respect to r yields the first-order condition

$$\frac{\partial\text{MSE}\left(\tilde{\sigma}^2\right)}{\partial r} = \frac{-2(n-K-r)\sigma^4 r^2 - 2r(n-K-r)^2\sigma^4}{r^4} - \frac{4r(n-K)\sigma^4}{r^4} = 0$$

which yields

$$(n-K-r)r + (n-K-r)^2 + 2(n-K) = 0$$

$$(n-K-r)(r+n-K-r) + 2(n-K) = 0$$

$$(n-K)(n-K-r+2) = 0$$

since $n > K$, this is zero for $r = n-K+2$. Hence, the Minimum MSE is obtained at $\tilde{\sigma}^2_* = e'e/(n-K+2)$ with

$$\text{MSE}\left(\tilde{\sigma}^2_*\right) = 4\sigma^4/(n-K+2)^2 + 2(n-K)\sigma^4/(n-K+2)^2$$
$$= 2\sigma^4(n-K+2)/(n-K+2)^2 = 2\sigma^4/(n-K+2).$$

Note that $s^2 = e'e/(n-K)$ with $\text{MSE}(s^2) = \text{var}(s^2) = 2\sigma^4/(n-K) > \text{MSE}\left(\tilde{\sigma}^2_*\right)$.

Also, it can be easily verified that $\text{MSE}\left(\tilde{\sigma}^2_{mle}\right) = (K^2 - 2K + 2n)\sigma^4/n^2 \geq \text{MSE}\left(\tilde{\sigma}^2_*\right)$ for $2 \leq K < n$, with the equality holding for $K = 2$.

7.7 *Computing Forecasts and Forecast Standard Errors Using a Regression Package*. This is based on Salkever (1976). From (7.23) one gets

$$X^{*\prime}X^{*} = \begin{bmatrix} X' & X_o' \\ 0 & I_{T_o} \end{bmatrix} \begin{bmatrix} X & 0 \\ X_o & I_{T_o} \end{bmatrix} = \begin{bmatrix} X'X + X_o'X_o & X_o' \\ X_o & I_{T_o} \end{bmatrix}$$

and

$$X^{*\prime}y^{*} = \begin{bmatrix} X'y + X_o'y_o \\ y_o \end{bmatrix}.$$

a. The OLS normal equations yield

$$X^{*\prime}X^{*} \begin{bmatrix} \hat{\beta}_{ols} \\ \hat{\gamma}_{ols} \end{bmatrix} = X^{*\prime}y^{*}$$

or $(X'X)\hat{\beta}_{ols} + \left(X_o'X_o\right)\hat{\beta}_{ols} + X_o'\hat{\gamma}_{ols} = X'y + X_o'y_o$

and $X_o\hat{\beta}_{ols} + \hat{\gamma}_{ols} = y_o$

From the second equation, it is obvious that $\hat{\gamma}_{ols} = y_o - X_o\hat{\beta}_{ols}$. Substituting this in the first equation yields

$$(X'X)\hat{\beta}_{ols} + (X_o'X_o)\hat{\beta}_{ols} + X_o'y_o - X_o'X_o\hat{\beta}_{ols} = X'y + X_o'y_o$$

which upon cancellations gives $\hat{\beta}_{ols} = (X'X)^{-1}X'y$.

Alternatively, one could apply the FWL Theorem using $X_1 = \begin{bmatrix} X \\ X_o \end{bmatrix}$ and $X_2 = \begin{bmatrix} 0 \\ I_{T_0} \end{bmatrix}$. In this case, $X_2'X_2 = I_{T_0}$ and

$$P_{X_2} = X_2\left(X_2'X_2\right)^{-1}X_2' = X_2X_2' = \begin{bmatrix} 0 & 0 \\ 0 & I_{T_o} \end{bmatrix}.$$

This means that

$$\overline{P}_{X_2} = I_{n+T_0} - P_{X_2} = \begin{bmatrix} I_n & 0 \\ 0 & 0 \end{bmatrix}.$$

Premultiplying (7.23) by $\overline{P}_{X_2}$ is equivalent to omitting the last T_o observations. The resulting regression is that of y on X which yields $\hat{\beta}_{ols} = (X'X)^{-1}X'y$ as obtained above. Also, premultiplying by $\overline{P}_{X_2}$, the last T_o observations yield zero residuals because the observations on both the dependent and independent variables are zero. For this to be true in the original regression, we must have $y_o - X_o\hat{\beta}_{ols} - \hat{\gamma}_{ols} = 0$. This means that $\hat{\gamma}_{ols} = y_o - X_o\hat{\beta}_{ols}$ as required.

b. The OLS residuals of (7.23) yield the usual least squares residuals

$$e_{ols} = y - X\hat{\beta}_{ols}$$

for the first n observations and zero residuals for the next T_o observations. This means that $e^{*\prime} = \left(e'_{ols}, 0'\right)$ and $e^{*\prime}e^* = e'_{ols}e_{ols}$ with the same residual sum of squares. The number of observations in (7.23) is $n+T_o$ and the number of parameters estimated is $K+T_o$. Hence the new degrees of freedom in (7.23) is $(n+T_o)-(K+T_o) = (n-K) =$ the old degrees of freedom in the regression of y on X. Hence, $s^{*2} = e^{*\prime}e^*/(n-K) = e'_{ols}e_{ols}/(n-K) = s^2$.

c. Using partitioned inverse formulas on $(X^{*\prime}X^*)$ one gets

$$\left(X^{*\prime}X^*\right)^{-1} = \begin{bmatrix} (X'X)^{-1} & -(X'X)^{-1}X'_o \\ -X_o(X'X)^{-1} & I_{T_o} + X_o(X'X)^{-1}X'_o \end{bmatrix}.$$

This uses the fact that the inverse of

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \quad \text{is} \quad A^{-1} = \begin{bmatrix} B_{11} & -B_{11}A_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{21}B_{11} & A_{22}^{-1} + A_{22}^{-1}A_{21}B_{11}A_{12}A_{22}^{-1} \end{bmatrix}$$

where $B_{11} = \left(A_{11} - A_{12}A_{22}^{-1}A_{21}\right)^{-1}$. Hence, $s^{*2}(X^{*\prime}X^*)^{-1} = s^2(X^{*\prime}X^*)^{-1}$ and is given by (7.25).

d. If we replace y_o by 0 and I_{T_o} by $-I_{T_o}$ in (7.23), we get

$$\begin{bmatrix} y \\ 0 \end{bmatrix} = \begin{bmatrix} X & 0 \\ X_o & -I_{T_o} \end{bmatrix}\begin{bmatrix} \beta \\ \gamma \end{bmatrix} + \begin{bmatrix} u \\ u_o \end{bmatrix}$$

or $y^* = X^*\delta + u^*$. Now $X^{*\prime}X^* = \begin{bmatrix} X' & X'_o \\ 0 & -I_{T_o} \end{bmatrix}\begin{bmatrix} X & 0 \\ X_o & -I_{T_o} \end{bmatrix} = \begin{bmatrix} X'X + X'_oX_o & -X'_o \\ -X_o & I_{T_o} \end{bmatrix}$ and $X^{*\prime}y^* = \begin{bmatrix} X'y \\ 0 \end{bmatrix}$. The OLS normal equations yield $(X'X)\hat{\beta}_{ols} + (X'_oX_o)\hat{\beta}_{ols} - X'_o\hat{\gamma}_{ols} = X'y$ and $-X_o\hat{\beta}_{ols} + \hat{\gamma}_{ols} = 0$.

From the second equation, it immediately follows that $\hat{\gamma}_{ols} = X_o\hat{\beta}_{ols} = \hat{y}_o$ the forecast of the T_o observations using the estimates from the first n observations. Substituting this in the first equation yields

$$(X'X)\hat{\beta}_{ols} + \left(X_o'X_o\right)\hat{\beta}_{ols} - X_o'X_o\hat{\beta}_{ols} = X'y$$

which gives $\hat{\beta}_{ols} = (X'X)^{-1}X'y$.

Alternatively, one could apply the FWL Theorem using $X_1 = \begin{bmatrix} X \\ X_o \end{bmatrix}$ and $X_2 = \begin{bmatrix} 0 \\ -I_{T_o} \end{bmatrix}$. In this case, $X_2'X_2 = I_{T_o}$ and $P_{X_2} = X_2X_2' = \begin{bmatrix} 0 & 0 \\ 0 & I_{T_o} \end{bmatrix}$ as before. This means that $\overline{P}_{X_2} = I_{n+T_o} - P_{X_2} = \begin{bmatrix} I_n & 0 \\ 0 & 0 \end{bmatrix}$.

As in part (a), premultiplying by $\overline{P}_{X_2}$ omits the last T_o observations and yields $\hat{\beta}_{ols}$ based on the regression of y on X from the first n observations only. The last T_o observations yield zero residuals because the dependent and independent variables for these T_o observations have zero values. For this to be true in the original regression, it must be true that $0 - X_o\hat{\beta}_{ols} + \hat{\gamma}_{ols} = 0$ which yields $\hat{\gamma}_{ols} = X_o\hat{\beta}_{ols} = \hat{y}_o$ as expected. The residuals are still $\left(e_{ols}', 0'\right)$ and $s^{*2} = s^2$ for the same reasons given in part (b). Also, using partitioned inverse as in part (c) above, we get

$$\left(X^{*\prime}X^*\right)^{-1} = \begin{bmatrix} (X'X)^{-1} & (X'X)^{-1}X_o' \\ X_o(X'X)^{-1} & I_{T_o} + X_o(X'X)^{-1}X_o' \end{bmatrix}.$$

Hence, $s^{*2}(X^{*\prime}X^*)^{-1} = s^2(X^{*\prime}X^*)^{-1}$ and the diagonal elements are as given in (7.25).

7.8 **a.** $\text{cov}\left(\hat{\beta}_{ols}, e\right) = E\left(\hat{\beta}_{ols} - \beta\right)e' = E\left[(X'X)^{-1}X'uu'\overline{P}_X\right] = \sigma^2(X'X)^{-1}X'\overline{P}_X = 0$ where the second equality uses the fact that $e = \overline{P}_X u$ and $\hat{\beta}_{ols} = \beta + (X'X)^{-1}X'u$. The third equality uses the fact that $E(uu') = \sigma^2 I_n$ and the last equality uses the fact that $\overline{P}_X X = 0$. But $e \sim N(0, \sigma^2\overline{P}_X)$ and $\hat{\beta}_{ols} \sim N(\beta, \sigma^2(X'X)^{-1})$, therefore zero covariance and normality imply independence of $\hat{\beta}_{ols}$ and e.

b. $\hat{\beta}_{ols} - \beta = (X'X)^{-1}X'u$ is linear in u, and $(n - K)s^2 = e'e = u'\overline{P}_X u$ is quadratic in u. A linear and quadratic forms in normal random variables

$u \sim N(0, \sigma^2 I_n)$ are independent if $(X'X)^{-1}X'\overline{P}_X = 0$, see Graybill (1961), Theorem 4.17. This is true since $\overline{P}_X X = 0$.

7.9 a. Replacing R by c' in (7.29) one gets $\left(c'\hat{\beta}_{ols} - c'\beta\right)'[c'(X'X)^{-1}c]^{-1}\left(c'\hat{\beta}_{ols} - c'\beta\right)/\sigma^2$. Since $c'(X'X)^{-1}c$ is a scalar, this can be rewritten as

$$\left(c'\hat{\beta}_{ols} - c'\beta\right)^2/\sigma^2 c'(X'X)^{-1}c$$

which is exactly the square of z_{obs} in (7.26). Since $z_{obs} \sim N(0,1)$ under the null hypothesis, its square is χ_1^2 under the null hypothesis.

b. Dividing the statistic given in part (a) by $(n-K)s^2/\sigma^2 \sim \chi^2_{n-K}$ divided by its degrees of freedom (n-K) results in replacing σ^2 by s^2, i.e.,

$$\left(c'\hat{\beta}_{ols} - c'\beta\right)^2/s^2 c'(X'X)^{-1}c.$$

This is the square of the t-statistic given in (7.27). But, the numerator is $z^2_{obs} \sim \chi^2_1$ and the denominator is $\chi^2_{n-K}/(n-K)$. Hence, if the numerator and denominator are independent, the resulting statistic is distributed as F(1,n-K) under the null hypothesis.

7.10 a. The quadratic form $u'Au/\sigma^2$ in (7.30) has

$$A = X(X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}R(X'X)^{-1}X'.$$

This is symmetric, and idempotent since

$$\begin{aligned} A^2 &= X(X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}R(X'X)^{-1}X'X(X'X)^{-1} \\ &\quad R'\left[R(X'X)^{-1}R'\right]^{-1}R(X'X)^{-1}X' \\ &= X(X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}R(X'X)^{-1}X' = A \end{aligned}$$

and rank (A) = tr(A) = tr $\left(R(X'X)^{-1}(X'X)(X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}\right)$

$= \mathrm{tr}(I_g) = g$ since R is gxK.

b. From lemma 1, $u'Au/\sigma^2 \sim \chi^2_g$ since A is symmetric and idempotent of rank g and $u \sim N(0, \sigma^2 I_n)$.

7.11 **a.** The two quadratic forms $s^2 = u'\overline{P}_X u/(n-K)$ and $u'Au/\sigma^2$ given in (7.30) are independent if and only if $\overline{P}_X A = 0$, see Graybill (1961), Theorem 4.10. This is true since $\overline{P}_X X = 0$.

b. $(n-K)s^2/\sigma^2$ is χ^2_{n-K} and $u'Au/\sigma^2 \sim \chi^2_g$ and both quadratic forms are independent of each other. Hence, dividing χ^2_g by g we get $u'Au/g\sigma^2$. Also, χ^2_{n-K} by (n-K) we get s^2/σ^2. Dividing $u'Au/g\sigma^2$ by s^2/σ^2 we get $u'Au/gs^2$ which is another way of writing (7.31). This is distributed as F(g,n-K) under the null hypothesis.

7.12 *Restricted Least Squares*

a. From (7.36), taking expected value we get

$$E(\hat{\beta}_{rls}) = E(\hat{\beta}_{ols}) + (X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}\left(r\text{-}RE(\hat{\beta}_{ols})\right)$$

$$= \beta + (X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}(r\text{-}R\beta)$$

since $E\left(\hat{\beta}_{ols}\right) = \beta$. It is clear that $\hat{\beta}_{rls}$ is in general biased for β unless $r = R\beta$ is satisfied, in which case the second term above is zero.

b. $\text{var}(\hat{\beta}_{rls}) = E[\hat{\beta}_{rls} - E(\hat{\beta}_{rls})][\hat{\beta}_{rls} - E(\hat{\beta}_{rls})]'$

But from (7.36) and part (a), we have

$$\hat{\beta}_{rls} - E(\hat{\beta}_{rls}) = (\hat{\beta}_{ols} - \beta) + (X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}R(\beta - \hat{\beta}_{ols})$$

using $\hat{\beta}_{ols} - \beta = (X'X)^{-1}X'u$, one gets $\hat{\beta}_{rls} - E(\hat{\beta}_{rls}) = A(X'X)^{-1}X'u$ where $A = I_K - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R$. It is obvious that A is not symmetric, i.e., $A \neq A'$. However, $A^2 = A$, since

$$A^2 = I_K - (X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}R - (X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}R$$
$$+ (X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}R(X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}R$$
$$= I_K - (X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}R = A.$$

Therefore,

$$\begin{aligned}
\mathrm{var}\left(\hat{\beta}_{rls}\right) &= E\left[A(X'X)^{-1}X'uu'X(X'X)^{-1}A'\right] = \sigma^2 A(X'X)^{-1}A' \\
&= \sigma^2\left[(X'X)^{-1} - (X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}R(X'X)^{-1}\right. \\
&\quad - (X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}R(X'X)^{-1} \\
&\quad + (X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}R(X'X)^{-1}R' \\
&\quad \left.\left[R(X'X)^{-1}R'\right]^{-1}R(X'X)^{-1}\right] \\
&= \sigma^2\left[(X'X)^{-1} - (X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}R(X'X)^{-1}\right]
\end{aligned}$$

c. Using part (b) and the fact that $\mathrm{var}\left(\hat{\beta}_{ols}\right) = \sigma^2(X'X)^{-1}$ gives $\mathrm{var}\left(\hat{\beta}_{ols}\right) - \mathrm{var}\left(\hat{\beta}_{rls}\right) = \sigma^2(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}$ and this is positive semi-definite, since $R(X'X)^{-1}R'$ is positive definite.

7.13 *The Chow Test*

a. OLS on (7.47) yields

$$\begin{pmatrix}\hat{\beta}_{1,ols}\\ \hat{\beta}_{2,ols}\end{pmatrix} = \begin{bmatrix}X_1'X_1 & 0\\ 0 & X_2'X_2\end{bmatrix}^{-1}\begin{bmatrix}X_1'y_1\\ X_2'y_2\end{bmatrix}$$
$$= \begin{bmatrix}\left(X_1'X_1\right)^{-1} & 0\\ 0 & \left(X_2'X_2\right)^{-1}\end{bmatrix}\begin{pmatrix}X_1'y_1\\ X_2'y_2\end{pmatrix} = \begin{pmatrix}\left(X_1'X_1\right)^{-1}X_1'y_1\\ \left(X_2'X_2\right)^{-1}X_2'y_2\end{pmatrix}$$

which is OLS on each equation in (7.46) separately.

b. The vector of OLS residuals for (7.47) can be written as $e' = \left(e_1', e_2'\right)$ where $e_1 = y_1 - X_1\hat{\beta}_{1,ols}$ and $e_2 = y_2 - X_2\hat{\beta}_{2,ols}$ are the vectors of OLS residuals from the two equations in (7.46) separately. Hence, the residual sum of squares $= e'e = e_1'e_1 + e_2'e_2 =$ sum of the residual sum of squares from running y_i on X_i for $i = 1, 2$.

c. From (7.47), one can write

$$\begin{pmatrix}y_1\\ y_2\end{pmatrix} = \begin{bmatrix}X_1\\ 0\end{bmatrix}\beta_1 + \begin{bmatrix}0\\ X_2\end{bmatrix}\beta_2 + \begin{pmatrix}u_1\\ u_2\end{pmatrix}.$$

This has the same OLS fitted values and OLS residuals as $\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}\gamma_1 + \begin{bmatrix} 0 \\ X_2 \end{bmatrix}\gamma_2 + \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$ where $\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ X_2 \end{bmatrix}$. This follows from problem 7.1 with $X = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix}$ and $C = \begin{bmatrix} I_K & 0 \\ I_K & I_K \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \otimes I_K$ where C is non-singular since $C^{-1} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \otimes I_K = \begin{bmatrix} I_K & 0 \\ -I_K & I_K \end{bmatrix}$. Hence, the X matrix in (7.47) is related to that in (7.49) as follows:

$$\begin{bmatrix} X_1 & 0 \\ X_2 & X_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} C$$

and the coefficients are therefore related as follows:

$$\begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} = C^{-1}\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{bmatrix} I_K & 0 \\ -I_K & I_K \end{bmatrix}\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \beta_2 - \beta_1 \end{pmatrix}$$

as required in (7.49). Hence, (7.49) yields the same URSS as (7.47). The RRSS sets $(\beta_2 - \beta_1) = 0$ which yields the same regression as (7.48).

7.14 a. The FWL Theorem states that $\hat{\beta}_{2,ols}$ from $\bar{P}_{X_1}y = \bar{P}_{X_1}X_2\beta_2 + \bar{P}_{X_1}u$ will be identical to the estimate of β_2 obtained from Eq. (7.8). Also, the residuals from both regressions will be identical. This means that the RSS from (7.8) given by $y'\bar{P}_X y = y'y - y'P_X y$ is identical to that from the above regression. The latter is similarly obtained as

$$\begin{aligned} &y'\bar{P}_{X_1}y - y'\bar{P}_{X_1}\bar{P}_{X_1}X_2\left(X_2'\bar{P}_{X_1}X_2\right)^{-1}X_2'\bar{P}_{X_1}\bar{P}_{X_1}y \\ &\quad = y'\bar{P}_{X_1}y - y'\bar{P}_{X_1}X_2\left(X_2'\bar{P}_{X_1}X_2\right)^{-1}X_2'\bar{P}_{X_1}y \end{aligned}$$

b. For testing $H_o;\ \beta_2 = 0$, the RRSS $= y'\bar{P}_{X_1}y$ and the URSS is given in part (a). Hence, the numerator of the Chow F-statistics given in (7.45) is given by

$$(\text{RRSS-URSS})/\,k_2 = y'\bar{P}_{X_1}X_2\left(X_2'\bar{P}_{X_1}X_2\right)^{-1}X_2'\bar{P}_{X_1}y/k_2$$

Substituting $y = X_1\beta_1 + u$ under the null hypothesis, yields

$u'\bar{P}_{X_1}X_2\left(X_2'\bar{P}_{X_1}X_2\right)^{-1}X_2'\bar{P}_{X_1}u/k_2$ since $\bar{P}_{X_1}X_1 = 0$.

c. Let $v = X_2'\bar{P}_{X_1}u$. Given that $u \sim N\left(0, \sigma^2 I_n\right)$, then v is Normal with mean zero and $\text{var}(v) = X_2'\bar{P}_{X_1}\text{var}(u)\bar{P}_{X_1}X_2 = \sigma^2 X_2'\bar{P}_{X_1}X_2$ since $\bar{P}_{X_1}$ is idempotent. Hence, $v \sim N\left(0, \sigma^2 X_2'\bar{P}_{X_1}X_2\right)$. Therefore, the numerator of the Chow F-statistic given in part (b) when divided by σ^2 can be written as $v'[\text{var}(v)]^{-1}v/k_2$. This is distributed as $\chi^2_{k_2}$ divided by its degrees of freedom k_2 under the null hypothesis. In fact, from part (b), $A = \bar{P}_{X_1}X_2\left(X_2'\bar{P}_{X_1}X_2\right)^{-1}X_2'\bar{P}_{X_1}$ is symmetric and idempotent and of rank equal to its trace equal to k_2. Hence, by lemma 1, $u'Au/\sigma^2$ is $\chi^2_{k_2}$ under the null hypothesis.

d. The numerator $u'Au/k_2$ is independent of the denominator $(n-k)s^2/(n-k) = u'\bar{P}_X u/(n-k)$ provided $\bar{P}_X A = 0$ as seen in problem 7.11. This is true because $\bar{P}_X\bar{P}_{X_1} = \bar{P}_X\left(I_n - P_{X_1}\right) = \bar{P}_X - \bar{P}_X X_1\left(X_1'X_1\right)^{-1}X_1' = \bar{P}_X$ since $\bar{P}_X X_1 = 0$. Hence,

$$\bar{P}_X A = \bar{P}_X\bar{P}_{X_1}X_2\left(X_2'\bar{P}_{X_1}X_2\right)^{-1}X_2'\bar{P}_{X_1} = \bar{P}_X X_2\left(X_2'\bar{P}_{X_1}X_2\right)^{-1}X_2'\bar{P}_{X_1} = 0$$

since $\bar{P}_X X_2 = 0$. Recall, $\bar{P}_X X = \bar{P}_X[X_1, X_2] = [\bar{P}_X X_1, \bar{P}_X X_2] = 0$.

e. The Wald statistic for H_o; $\beta_2 = 0$ given in (7.41) boils down to replacing R by $[0, I_{k_2}]$ and r by 0. Also, $\hat{\beta}_{mle}$ by $\hat{\beta}_{ols}$ from the unrestricted model given in (7.8) and σ^2 is replaced by its estimate $s^2 = URSS/(n-k)$ to make the Wald statistic feasible. This yields $W = \hat{\beta}_2'[R(X'X)^{-1}R']^{-1}\hat{\beta}_2/s^2$. From problem 7.2, we showed that the partitioned inverse of $X'X$ yields $B_{22} = \left(X_2'\bar{P}_{X_1}X_2\right)^{-1}$ for its second diagonal $(k_2 x k_2)$ block. Hence,

$$R(X'X)^{-1}R' = [0, I_{k_2}]\,(X'X)^{-1}\begin{bmatrix} 0 \\ I_{k_2} \end{bmatrix} = B_{22} = \left(X_2'\bar{P}_{X_1}X_2\right)^{-1}.$$

Also, from problem 7.2, $\hat{\beta}_{2,ols} = \left(X_2'\bar{P}_{X_1}X_2\right)^{-1}X_2'\bar{P}_{X_1}y = \left(X_2'\bar{P}_{X_1}X_2\right)^{-1}X_2'\bar{P}_{x_1}u$ after substituting $y = X_1\beta_1 + u$ under the null hypothesis and using

$\bar{P}_{X_1}X_1 = 0$. Hence,

$$s^2W = u'\bar{P}_{X_1}X_2\left(X_2'\bar{P}_{X_1}X_2\right)^{-1}\left(X_2'\bar{P}_{X_1}X_2\right)\left(X_2'\bar{P}_{X_1}X_2\right)^{-1}X_2'\bar{P}_{X_1}u$$

$$= u'\bar{P}_{X_1}X_2\left(X_2'\bar{P}_{X_1}X_2\right)^{-1}X_2'\bar{P}_{X_1}u$$

which is exactly k_2 times the expression in part (b), i.e., the numerator of the Chow F-statistic.

f. The restricted MLE of β is $\left(\hat{\beta}'_{1,rls}, 0'\right)$ since $\beta_2 = 0$ under the null hypothesis. Hence the score form of the LM test given in (7.44) yields

$$\left(y - X_1\hat{\beta}_{1,rls}\right)'X(X'X)^{-1}X'\left(y - X_1\hat{\beta}_{1,rls}\right)\Big/\sigma^2.$$

In order to make this feasible, we replace σ^2 by $\tilde{s}^2 = \text{RRSS}/(n - k_1)$ where RRSS is the restricted residual sum of squares from running y on X_1. But this expression is exactly the regression sum of squares from running $(y - X_1\hat{\beta}_{1,rls})/\tilde{s}$ on the matrix X. In order to see this, the regression sum of squares of y on X is usually $y'P_xy$. Here, y is replaced by $(y - X_1\hat{\beta}_{1,rls})/\tilde{s}$.

7.15 *Iterative Estimation in Partitioned Regression Models*. This is based on Baltagi (1996).

a. The least squares residuals of y on X_1 are given by $\bar{P}_{X_1}y$, where $\bar{P}_{X_1} = I - P_{X_1}$ and $P_{X_1} = X_1\left(X_1'X_1\right)^{-1}X_1'$. Regressing these residuals on x_2 yields $b_2^{(1)} = \left(x_2'x_2\right)^{-1}x_2'\bar{P}_{x_1}y$. Substituting for y from (7.8) and using $\bar{P}_{X_1}X_1 = 0$ yields $b_2^{(1)} = \left(x_2'x_2\right)^{-1}x_2'\bar{P}_{X_1}(x_2\beta_2 + u)$ with

$$E\left(b_2^{(1)}\right) = \left(x_2'x_2\right)^{-1}x_2'\bar{P}_{X_1}x_2\beta_2 = \beta_2 - \left(x_2'x_2\right)^{-1}x_2'P_{X_1}x_2\beta_2 = (1 - a)\beta_2$$

where $a = \left(x_2'P_{X_1}x_2\right)/\left(x_2'x_2\right)$ is a scalar, with $0 \le a < 1$. $a \ne 1$ as long as x_2 is linearly independent of X_1. Therefore, the bias $\left(b_2^{(1)}\right) = -a\beta_2$.

b. $b_1^{(1)} = \left(X_1'X_1\right)^{-1} X_1' \left(y - x_2 b_2^{(1)}\right) = \left(X_1'X_1\right)^{-1} X_1' \left(I - P_{x_2}\overline{P}_{X_1}\right) y$ and

$$b_2^{(2)} = \left(x_2'x_2\right)^{-1} x_2' \left(y - X_1 b_1^{(1)}\right) = \left(x_2'x_2\right)^{-1} x_2' \left[I - P_{X_1}\left(I - P_{x_2}\overline{P}_{X_1}\right)\right] y$$
$$= \left(x_2'x_2\right)^{-1} x_2' \left[\overline{P}_{X_1} + P_{X_1}P_{x_2}\overline{P}_{X_1}\right] y = (1+a)b_2^{(1)}.$$

Similarly,

$$b_1^{(2)} = \left(X_1'X_1\right)^{-1} X_1' \left(y - x_2 b_2^{(2)}\right) = \left(X_1'X_1\right)^{-1} X_1' \left(y - (1+a)x_2 b_2^{(1)}\right)$$
$$b_2^{(3)} = \left(x_2'x_2\right)^{-1} x_2' \left(y - X_1 b_1^{(2)}\right)$$
$$= \left(x_2'x_2\right)^{-1} x_2' \left[y - P_{X_1}\left(y - (1+a)x_2 b_2^{(1)}\right)\right]$$
$$= b_2^{(1)} + (1+a)\left(x_2'x_2\right)^{-1} x_2' P_{X_1} x_2 b_2^{(1)} = (1 + a + a^2) b_2^{(1)}.$$

By induction, one can infer that

$$b_2^{(j+1)} = (1 + a + a^2 + .. + a^j) b_2^{(1)} \quad \text{for} \quad j = 0, 1, 2, \ldots$$

Therefore,

$$E\left(b_2^{(j+1)}\right) = (1 + a + a^2 + .. + a^j) E\left(b_2^{(1)}\right)$$
$$= (1 + a + a^2 + .. + a^j)(1-a)\beta_2 = (1 - a^{j+1})\beta_2$$

and the bias $\left(b_2^{(j+1)}\right) = -a^{j+1}\beta_2$ tends to zero as $j \to \infty$, since $|a| < 1$.

c. Using the Frisch–Waugh–Lovell Theorem, least squares on the original model yields

$$\hat{\beta}_2 = \left(x_2'\overline{P}_{X_1}x_2\right)^{-1} x_2'\overline{P}_{X_1} y = \left(x_2'x_2 - x_2'P_{X_1}x_2\right)^{-1} x_2'\overline{P}_{X_1} y$$
$$= (1-a)^{-1}\left(x_2'x_2\right)^{-1} x_2'\overline{P}_{X_1} y = b_2^{(1)}/(1-a).$$

As $j \to \infty$, $\lim\ b_2^{(j+1)} = \left(\sum_{j=0}^{\infty} a^j\right) b_2^{(1)} = b_2^{(1)}/(1-a) = \hat{\beta}_2.$

7.16 *Maddala (1992, pp. 120–127).*

a. For $H_o;\ \beta = 0$, the RRSS is based on a regression of y_i on a constant. This yields $\hat{\alpha} = \bar{y}$ and the RRSS $= \sum_{i=1}^{n}(y_i - \bar{y})^2 =$ usual TSS. The URSS is the usual least squares residual sum of squares based on estimating α and β. The log-likelihood in this case is given by

$$\log L(\alpha,\beta,\sigma^2) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\alpha-\beta X_i)^2$$

with the unrestricted MLE of α and β yielding $\hat{\alpha}_{ols} = \bar{y} - \hat{\beta}_{ols}\bar{X}$ and

$$\hat{\beta}_{ols} = \sum_{i=1}^{n}\left(X_i-\bar{X}\right)y_i \Big/ \sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2$$

and $\hat{\sigma}^2_{mle} = \text{URSS}/n$. In this case, the unrestricted log-likelihood yields

$$\begin{aligned}\log L\left(\hat{\alpha}_{mle},\hat{\beta}_{mle},\hat{\sigma}^2_{mle}\right) &= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\hat{\sigma}^2_{mle} - \frac{\text{URSS}}{(2\text{URSS}/n)}\\ &= -\frac{n}{2}\log 2\pi - \frac{n}{2} - \frac{n}{2}\log(\text{URSS}/n).\end{aligned}$$

Similarly, the restricted MLE yields $\hat{\alpha}_{rmle} = \bar{y}$ and $\hat{\beta}_{rmle} = 0$ and

$$\hat{\sigma}^2_{rmle} = \text{RRSS}/n = \sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2/n.$$

The restricted log-likelihood yields

$$\begin{aligned}\log L\left(\hat{\alpha}_{rmle},\hat{\beta}_{rmle},\hat{\sigma}^2_{rmle}\right) &= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\hat{\sigma}^2_{rmle} - \frac{\text{RRSS}}{(2\text{RRSS}/n)}\\ &= -\frac{n}{2}\log 2\pi - \frac{n}{2} - \frac{n}{2}\log(\text{RRSS}/n).\end{aligned}$$

Hence, the LR test is given by

$$\begin{aligned}\text{LR} &= n(\log\ \text{RRSS-logURSS}) = n\ \log(\text{RRSS}/\text{URSS})\\ &= n\log(\text{TSS}/\text{RSS}) = n\ \log(1/1-r^2)\end{aligned}$$

where TSS and RSS are the total and residual sum of squares from the unrestricted regression. By definition $R^2 = 1 - (\text{RSS}/\text{TSS})$ and for the simple regression $r^2_{XY} = R^2$ of that regression, see Chap. 3.

b. The Wald statistic for H_o; $\beta = 0$ is based upon $r\left(\hat{\beta}_{mle}\right) = (\hat{\beta}_{mle} - 0)$ and $R\left(\hat{\beta}_{mle}\right) = 1$ and from (7.40), we get $W = \hat{\beta}^2_{mle}/\text{var}\left(\hat{\beta}_{mle}\right) = \hat{\beta}^2_{ols}/\text{var}(\hat{\beta}_{ols})$.

This is the square of the usual t-statistic for $\beta = 0$ with $\hat{\sigma}^2_{mle}$ used instead of s^2 in estimating σ^2. Using the results in Chap. 3, we get

$$W = \frac{\left(\sum_{i=1}^{n} x_i y_i\right)^2 \Big/ \left(\sum_{i=1}^{n} x_i^2\right)^2}{\hat{\sigma}^2_{mle} \Big/ \left(\sum_{i=1}^{n} x_i^2\right)} = \frac{\left(\sum_{i=1}^{n} x_i y_i\right)^2}{\hat{\sigma}^2_{mle} \sum_{i=1}^{n} x_i^2}$$

with $\hat{\sigma}^2_{mle} = \mathrm{URSS}/n = \mathrm{TSS}(1 - R^2)/n$ from the definition of R^2. Hence,

$$W = \frac{n\left(\sum_{i=1}^{n} x_i y_i\right)^2}{\left(\sum_{i=1}^{n} y_i^2\right)\left(\sum_{i=1}^{n} x_i^2\right)(1 - R^2)} = \frac{nr^2}{1 - r^2}$$

using the definition of $r^2_{XY} = r^2 = R^2$ for the simple regression.

c. The LM statistic given in (7.43) is based upon $LM = \hat{\beta}^2_{ols}/\hat{\sigma}^2_{rmle}\left(\sum_{i=1}^{n} x_i^2\right)$. This is the t-statistic on $\beta = 0$ using $\hat{\sigma}^2_{rmle}$ as an estimate for σ^2. In this case, $\hat{\sigma}^2_{rmle} = \mathrm{RRSS}/n = \sum_{i=1}^{n} y_i^2/n$. Hence, $LM = n\left(\sum_{i=1}^{n} x_i y_i\right)^2 \Big/ \sum_{i=1}^{n} y_i^2 \sum_{i=1}^{n} x_i^2 = nr^2$ from the definition of $r^2_{XY} = r^2$.

d. Note that from part (b), we get $W/n = r^2/(1 - r^2)$ and $1 + (W/n) = 1/(1 - r^2)$. Hence, from part (a), we have $(LR/n) = \log(1/1 - r^2) = \log[1 + (W/n)]$.

From part (c), we get $(LM/n) = (W/n)/[1 + (W/n)]$. Using the inequality

$x \geq \log(1 + x) \geq x/(1 + x)$ with $x = W/n$ we get $W \geq LR \geq LM$.

e. From Chap. 3, R^2 of the regression of logC on logP is 0.2913 and $n = 46$. Hence, $W = nr^2/1 - r^2 = (46)(0.2913)/(0.7087) = 18.91$,

$LM = nr^2 = (46)(0.2913) = 13.399$

and $LR = n\log(1/1 - r^2) = 46\ \log(1/0.7087) = 46\ \log(1.4110342) = 15.838$. It is clear that $W > LR > LM$ in this case.

7.17 *Engle (1984, pp. 785–786).*

a. For the Bernoulli random variable y_t with probability of success θ, the log-likelihood function is given by $\log L(\theta) = \sum_{t=1}^{T}[y_t \log\theta + (1-y_t)\log(1-\theta)]$.

The score is given by

$$\begin{aligned} S(\theta) = \partial \log L(\theta)/\partial\theta &= \sum_{t=1}^{T} y_t/\theta - \sum_{t=1}^{T}(1-y_t)/(1-\theta) \\ &= \left[\sum_{t=1}^{T} y_t - \theta\sum_{t=1}^{T} y_t - \theta T + \theta\sum_{t=1}^{T} y_t\right] \Big/ \theta(1-\theta) \\ &= \sum_{t=1}^{T}(y_t-\theta)/\theta(1-\theta) \end{aligned}$$

The MLE is given by setting $S(\theta) = 0$ giving $\hat{\theta}_{mle} = \sum_{t=1}^{T} y_t/T = \bar{y}$.

Now $\partial^2 \log L(\theta)/\partial\theta^2 = \left[-T\theta(1-\theta) - \sum_{t=1}^{T}(y_t-\theta)(1-2\theta)\right]/\theta^2(1-\theta)^2$.

Therefore, the information matrix is

$$I(\theta) = -E\left[\frac{\partial^2 \log L(\theta)}{\partial\theta^2}\right] = \left(T\theta(1-\theta) + (1-2\theta)\sum_{t=1}^{T}[E(y_t)-\theta]\right)/$$

$$\theta^2(1-\theta^2) = T/\theta(1-\theta).$$

b. For testing $H_o; \theta = \theta_o$ versus $H_A; \theta \neq \theta_o$, the Wald statistic given in (7.40) has $r(\theta) = \theta - \theta_o$ and $R(\theta) = 1$ with $I^{-1}\left(\hat{\theta}_{mle}\right) = \hat{\theta}_{mle}(1-\hat{\theta}_{mle})/T$. Hence,

$$W = T\left(\hat{\theta}_{mle} - \theta_o\right)^2 \Big/ \hat{\theta}_{mle}\left(1-\hat{\theta}_{mle}\right) = T(\bar{y}-\theta_o)^2/\bar{y}(1-\bar{y}).$$

The LM statistic given in (7.42) has $S(\theta_o) = T(\bar{y}-\theta_o)/\theta_o(1-\theta_o)$ and $I^{-1}(\theta_o) = \theta_o(1-\theta_o)/T$. Hence,

$$LM = \frac{T^2(\bar{y}-\theta_o)^2}{[\theta_o(1-\theta_o)]^2} \cdot \frac{\theta_o(1-\theta_o)}{T} = \frac{T(\bar{y}-\theta_o)^2}{\theta_o(1-\theta_o)}.$$

The unrestricted log-likelihood is given by

$$\log L\left(\hat{\theta}_{mle}\right) = \log L(\bar{y}) = \sum_{t=1}^{T} [y_t \log \bar{y} + (1 - y_t)\log(1 - \bar{y})]$$
$$= T\bar{y}\log\bar{y} + T(1 - \bar{y})\log(1 - \bar{y}).$$

The restricted log-likelihood is given by

$$\log L(\theta_o) = \sum_{t=1}^{T} [y_t \log \theta_o + (1 - y_t)\log(1 - \theta_o)]$$
$$= T\bar{y}\log\theta_o + T(1 - \bar{y})\log(1 - \theta_o).$$

Hence, the likelihood ratio test gives

$$LR = 2T\bar{y}(\log\bar{y} - \log\theta_o) + 2T(1 - \bar{y})[\log(1 - \bar{y}) - \log(1 - \theta_o)]$$
$$= 2T\bar{y}\log(\bar{y}/\theta_o) + 2T(1 - \bar{y})\log[(1 - \bar{y})/(1 - \theta_o)].$$

All three statistics have a limiting χ_1^2 distribution under H_o. Each statistic will reject when $(\bar{y} - \theta_o)^2$ is large. Hence, for finite sample exact results one can refer to the binomial distribution and compute exact critical values.

7.18 For the regression model

$$y = X\beta + u \quad \text{with} \quad u \sim N(0, \sigma^2 I_T).$$

a. $L(\beta, \sigma^2) = (1/2\pi\sigma^2)^{T/2} \exp\left\{-(y - X\beta)'(y - X\beta)/2\sigma^2\right\}$

$\log L = -(T/2)(\log 2\pi + \log\sigma^2) - (y - X\beta)'(y - X\beta)/2\sigma^2$

$\partial \log L/\partial\beta = -(-2X'y + 2X'X\beta)/2\sigma^2 = 0$

so that $\hat{\beta}_{mle} = (X'X)^{-1}X'y$ and $\partial \log L/\partial\sigma^2 = -T/2\sigma^2 + \hat{u}'\hat{u}/2\sigma^4 = 0.$

Hence, $\hat{\sigma}^2_{mle} = \hat{u}'\hat{u}/T$ (where $\hat{u} = y - X\hat{\beta}_{mle}$).

b. The score for β is $S(\beta) = \partial \log L(\beta)/\partial\beta = (X'y - X'X\beta)/\sigma^2$.

The Information matrix is given by

$$I(\beta, \sigma^2) = -E\left[\frac{\partial^2 \log L}{\partial(\beta, \sigma^2)\partial(\beta, \sigma^2)'}\right] = -E\begin{bmatrix} -\frac{X'X}{\sigma^2} & -\frac{X'y - X'X\beta}{\sigma^4} \\ -\frac{X'y - X'X\beta}{\sigma^4} & \frac{T}{2\sigma^4} - \frac{(y - X\beta)'(y - X\beta)}{\sigma^6} \end{bmatrix}$$

since $E(X'y - X'X\beta) = X'E(y - X\beta) = 0$ and $E(y - X\beta)'(y - X\beta) = T\sigma^2$

$$I(\beta, \sigma^2) = \begin{bmatrix} \frac{X'X}{\sigma^2} & 0 \\ 0 & \frac{T}{2\sigma^4} \end{bmatrix}$$

which is block-diagonal and also given in (7.19).

c. The Wald statistic given in (7.41) needs

$$r(\beta) = \beta_1 - \beta_1^0$$

$$R(\beta) = [I_{k_1}, 0]$$

$$W = \left(\hat{\beta}_1 - \beta_1^0\right)' \left[(I_{k_1}, 0)\, \hat{\sigma}^2 (X'X)^{-1} \begin{pmatrix} I_{k_1} \\ 0 \end{pmatrix}\right]^{-1} \left(\hat{\beta}_1 - \beta_1^0\right)$$

with

$$(X'X)^{-1} = \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix}^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

and

$$a = \left(X_1'X_1 - X_1'X_2 \left(X_2'X_2\right)^{-1} X_2'X_1\right)^{-1}$$
$$= \left[X_1' \left[I - X_2 \left(X_2'X_2\right)^{-1} X_2'\right] X_1\right]^{-1} = \left[X_1'\bar{P}_{X_2}X_1\right]^{-1}$$

by partitioned inverse. Therefore,

$$W = \left(\beta_1^0 - \hat{\beta}_1\right)' \left[X_1'\bar{P}_{X_2}X_1\right] \left(\beta_1^0 - \hat{\beta}_1\right) \Big/ \hat{\sigma}^2$$

as required. For the LR statistic $LR = -2\left(\log L_r^* - \log L_u^*\right)$ where L_r^* is the restricted likelihood and L_u^* is the unrestricted likelihood.

$$LR = -2\left[\left(-\frac{T}{2} - \frac{T}{2}\log 2\pi - \frac{T}{2}\log\left(\frac{RRSS}{T}\right)\right) - \left(-\frac{T}{2} - \frac{T}{2}\log 2\pi - \frac{T}{2}\log\left(\frac{URSS}{T}\right)\right)\right]$$
$$= T\ \log(RRSS/URSS) = T\log\left[\frac{\tilde{u}'\tilde{u}}{\hat{u}'\hat{u}}\right]$$

For the LM statistic, the score version is given in (7.42) as

$$LM = S(\tilde{\beta})'\, I^{-1}\, (\tilde{\beta})\; S\, (\tilde{\beta})$$

where

$$\tilde{\sigma}^2 S\left(\tilde{\beta}_2\right) = \begin{bmatrix} \tilde{\sigma}^2 S_1\left(\tilde{\beta}_2\right) \\ \tilde{\sigma}^2 S_2\left(\tilde{\beta}_2\right) \end{bmatrix} = \begin{bmatrix} X_1'\left(y - X\tilde{\beta}\right) \\ X_2'\left(y - X\tilde{\beta}\right) \end{bmatrix}.$$

The restriction on β_1 is $\left(\beta_1 = \beta_1^0\right)$, but there are no restrictions on β_2. Note that $\tilde{\beta}_2$ can be obtained from the regression of $\left(y - X_1\beta_1^0\right)$ on X_2. This yields

$$\tilde{\beta}_2 = \left(X_2'X_2\right)^{-1} X_2'\left(y - X_1\beta_1^0\right).$$

Therefore,

$$\begin{aligned}\tilde{\sigma}^2 S_2\left(\tilde{\beta}_2\right) &= X_2'\left(y - X\tilde{\beta}\right) = X_2'y - X_2'X_1\beta_1^0 - X_2'X_2\,\tilde{\beta}_2 \\ &= X_2'y - X_2'X_1\beta_1^0 - X_2'y + X_2'X_1\beta_1^0 = 0.\end{aligned}$$

Hence, $LM = S_1\left(\tilde{\beta}\right)' I^{11}\left(\tilde{\beta}\right) S_1\left(\tilde{\beta}\right)$ where $I^{11}(\beta)$ is obtained from the partitioned inverse of $I^{-1}(\beta)$. Since $S_1\left(\tilde{\beta}\right) = X_1'\left(y - X\tilde{\beta}\right)/\tilde{\sigma}^2 = X_1'\tilde{u}/\tilde{\sigma}^2$ and $I^{-1}(\tilde{\beta}) = \tilde{\sigma}^2(X'X)^{-1}$ with $I^{11}\left(\tilde{\beta}\right) = \tilde{\sigma}^2\left[X_1'\bar{P}_{X_2}X_1\right]^{-1}$ we get

$$LM = \tilde{u}'X_1\left[X_1'\bar{P}_{X_2}X_1\right]^{-1} X_1'\tilde{u}/\tilde{\sigma}^2$$

d. $$\begin{aligned}W &= \left(\beta_1^o - \hat{\beta}_1\right)'\left[X_1'\bar{P}_{X_2}X_1\right]\left(\beta_1^0 - \hat{\beta}_1\right)/\hat{\sigma}^2 \\ &= \left(\beta_1^0 - \hat{\beta}_1\right)'\left[R\left(X'X\right)^{-1}R'\right]^{-1}\left(\beta_1^0 - \hat{\beta}_1\right)/\hat{\sigma}^2 \\ &= \left(r - R\hat{\beta}\right)'\left[R\left(X'X\right)^{-1}R'\right]^{-1}\left(r - R\hat{\beta}\right)/\hat{\sigma}^2.\end{aligned}$$

From (7.39) we know that $\tilde{u}'\tilde{u} - \hat{u}'\hat{u} = \left(r - R\hat{\beta}\right)'\left[R(X'X)^{-1}R'\right]^{-1}\left(r - R\hat{\beta}\right)$. Also, $\hat{\sigma}^2 = \hat{u}'\hat{u}/T$. Therefore, $W = T(\tilde{u}'\tilde{u} - \hat{u}'\hat{u})/\hat{u}'\hat{u}$ as required.

$$LM = \tilde{u}'X_1\left[X_1'\bar{P}_{X_2}X_1\right]^{-1} X_1'\tilde{u}/\tilde{\sigma}^2$$

From (7.43) we know that $LM = (r - R\tilde{\beta})'[R(X'X)^{-1}R']^{-1}(r - R\tilde{\beta})/\tilde{\sigma}^2$ and $\tilde{\sigma}^2 = \tilde{u}'\tilde{u}/T$. Using (7.39) we can rewrite this as $LM = T\,(\tilde{u}'\tilde{u} - \hat{u}'\hat{u})/\tilde{u}'\tilde{u}$ as required. Finally,

$$LR = T\ \log\left(\tilde{u}'\tilde{u}/\hat{u}'\hat{u}\right) = T\ \log\left(1 + \frac{(\tilde{u}'\tilde{u} - \hat{u}'\hat{u})}{\hat{u}'\hat{u}}\right) = T\ \log(1 + W/T).$$

Also, $W/LM = \frac{\tilde{u}'\tilde{u}}{\hat{u}'\hat{u}} = 1 + W/T$. Hence, $LM = W/(1 + W/T)$ and from (7.45)

$$((T-k)/T) \cdot W/k_1 = \frac{(\tilde{u}'\tilde{u} - \hat{u}'\hat{u})/k_1}{\hat{u}'\hat{u}/T-k} \sim F_{k_1,T-k}$$

Using the inequality $x \geq \log(1+x) \geq x/(1+x)$ with $x = W/T$ we get $(W/T) \geq \log(1 + W/T) \geq (W/T)/(1 + W/T)$ or $(W/T) \geq (LR/T) \geq (LM/T)$ or $W \geq LR \geq LM$. However, it is important to note that all the statistics are monotonic functions of the F-statistic and exact tests for each would produce identical critical regions.

e. For the cigarette consumption data given in Table 3.2 the following test statistics were computed for $H_o; \beta = -1$

$\text{Wald} = 1.16 \geq LR = 1.15 \geq LM = 1.13$

and the SAS program that produces these results is given below.

f. The Wald statistic for $H_0^A; \beta = -1$ yields 1.16, for $H_0^B; \beta^5 = -1$ yields 0.43 and for $H_0^C; \beta^{-5} = -1$ yields 7.89. The SAS program that produces these results is given below.

SAS PROGRAM

```
Data CIGARETT;
Input OBS STATE $ LNC LNP LNY;
Cards;

Proc IML; Use CIGARETT;
Read all into Temp;

N=NROW(TEMP);            ONE=Repeat(1,N,1);
Y=Temp[,2];    X=ONE||Temp[,3]||Temp[,4];
BETA_U=INV(X`*X)*X`*yY;
R={0 1 0};
Ho=BETA_U[2,]+1;
BETA_R=BETA_U+INV(X`*X)*R`*INV(R*yINV(X`*X)*R`)*Ho;
ET_U=Y-X*BETA_U;
```

```
ET_R=Y-X*BETA_R;
SIG_U=(ET_U`*ET_U)/N;
SIG_R=(ET_R`*ET_R)/N;
X1=X[,2];
X2=X[,1]||X[,3];
Q_X2=I(N)-X2*INV(X2`*X2)*X2`;
VAR_D=SIG_U*INV(X1`*Q_X2*X1);
WALD=Ho `*INV(VAR_D)*Ho;
LR=N*LOG(1+(WALD/N));
LM=(ET_R`*X1*INV(X1`*Q_X2*X1)*X1`*ET_R)/SIG_R;

   *WALD=N* (ET_R`*ET_R-ET_U`*ET_U)/(ET_U`*ET_U);
   *LR=N*Log(ET_R`*ET_R/(ET_U`*ET_U));
   *LM=N*(ET_R`*ET_R-ET_U`*ET_U)/(ET_R`*ET_R);

PRINT ‘Chapter˜7 Problem˜18. (e)’,, WALD;
PRINT LR;
PRINT LM;

BETA=BETA_U[2,];
H1=BETA+1;
H2=BETA**5+1;
H3=BETA**(-5)+1;

VAR_D1=SIG_U*INV(X1`*Q_X2*X1);
VAR_D2=(5*BETA**4)*VAR_D1*(5*BETA**4);
VAR_D3=(-5*BETA**(-6))*VAR_D1*(-5*BETA**(-6));

WALD1=H1`*INV(VAR_D1)*H1;
WALD2=H2`*INV(VAR_D2)*H2;
WALD3=H3`*INV(VAR_D3)*H3;

PRINT ‘Chapter˜7 Problem˜18.(f)’,, WALD1;
PRINT WALD2;
PRINT WALD3;
```

7.19 *Gregory and Veall (1985).*

a. For $H^A: \beta_1 - 1/\beta_2 = 0$, we have $r^A(\beta) = \beta_1 - 1/\beta_2$ and $\beta' = (\beta_o, \beta_1, \beta_2)$. In this case, $R^A(\beta) = \left(0, 1, 1/\beta_2^2\right)$ and the unrestricted MLE is OLS on (7.50) with variance–covariance matrix $\hat{V}(\hat{\beta}_{ols}) = \hat{\sigma}^2(X'X)^{-1}$ where $\hat{\sigma}^2_{mle} =$ URSS/n. Let v_{ij} denote the corresponding elements of $\hat{V}(\hat{\beta}_{ols})$ for $i, j =$ 0,1,2. Therefore,

$$W^A = \left(\hat{\beta}_1 - 1/\hat{\beta}_2\right)\left[\left(0, 1, 1/\hat{\beta}_2^2\right)\ \hat{V}\left(\hat{\beta}_{ols}\right)\left(0, 1, 1/\hat{\beta}_2^2\right)'\right]^{-1}\left(\hat{\beta}_1 - 1/\hat{\beta}_2\right)$$

$$= \left(\hat{\beta}_1\hat{\beta}_2 - 1\right)^2 \Big/ \left(\hat{\beta}_2^2\, v_{11} + 2v_{12} + v_{22}/\hat{\beta}_2^2\right)$$

as required in (7.52). Similarly, for $H^B; \beta_1\beta_2 - 1 = 0$, we have $r^B(\beta) = \beta_1\beta_2 - 1$.

In this case, $R^B(\beta) = (0, \beta_2, \beta_1)$ and

$$W^B = \left(\hat{\beta}_1\hat{\beta}_2 - 1\right)\left[\left(0, \hat{\beta}_2, \hat{\beta}_1\right)\ \hat{V}\left(\hat{\beta}_{ols}\right)\left(0, \hat{\beta}_2, \hat{\beta}_1\right)'\right]^{-1}\left(\hat{\beta}_1\hat{\beta}_2 - 1\right)$$

$$= \left(\hat{\beta}_1\hat{\beta}_2 - 1\right)^2 \Big/ \left(\hat{\beta}_2^2 v_{11} + 2\hat{\beta}_1\hat{\beta}_2\, v_{12} + \hat{\beta}_1^2\, v_{22}\right)$$

as required in (7.53).

7.20 *Gregory and Veall (1986).*

a. From (7.51), we get $W = r\left(\hat{\beta}_{ols}\right)'\left[R\left(\hat{\beta}_{ols}\right)\hat{\sigma}^2(X'X)^{-1}R\left(\hat{\beta}_{ols}\right)'\right]^{-1}r\left(\hat{\beta}_{ols}\right)$. For $H^A; \beta_1\rho + \beta_2 = 0$, we have $r(\beta) = \beta_1\rho + \beta_2$ and $R(\beta) = (\beta_1, \rho, 1)$ where $\beta' = (\rho, \beta_1, \beta_2)$. Hence,

$$W^A = \left(\hat{\beta}_1\hat{\rho} + \hat{\beta}_2\right)\left[\left(\hat{\beta}_1, \hat{\rho}, 1\right)\hat{\sigma}^2(X'X)^{-1}\left(\hat{\beta}_1, \hat{\rho}, 1\right)'\right]^{-1}\left(\hat{\beta}_1\hat{\rho} + \hat{\beta}_2\right)$$

Where the typical element of the matrix are $[y_{t-1}, x_t, x_{t-1}]$. For $H^B;\ \beta_1 + (\beta_2/\rho) = 0$, we have $r(\beta) = \beta_1 + (\beta_2/\rho)$ and

$$R(\beta) = \left(-\frac{\beta_2}{\rho^2}, 1, \frac{1}{\rho}\right).$$ Hence,

$$W^B = \left(\hat{\beta}_1 + \hat{\beta}_2/\hat{\rho}\right)\left[\left(-\frac{\hat{\beta}_2}{\hat{\rho}^2}, 1, \frac{1}{\hat{\rho}}\right)\hat{\sigma}^2(X'X)^{-1}\left(-\frac{\hat{\beta}_2}{\hat{\rho}^2}, 1, \frac{1}{\hat{\rho}}\right)'\right]^{-1}$$
$$\left(\hat{\beta}_1 + \hat{\beta}_2/\hat{\rho}\right).$$

For H^C; $\rho + (\beta_2/\beta_1) = 0$, we have

$$r(\beta) = \rho + \beta_2/\beta_1 \qquad \text{and} \qquad R(\beta) = \left(1, -\frac{\beta_2}{\beta_1^2}, \frac{1}{\beta_1}\right).$$

Hence,

$$W^C = \left(\hat{\rho} + \hat{\beta}_2/\hat{\beta}_1\right)\left[\left(1, -\frac{\hat{\beta}_2}{\hat{\beta}_1^2}, \frac{1}{\hat{\beta}_1}\right)\hat{\sigma}^2(X'X)^{-1}\left(1, -\frac{\hat{\beta}_2}{\hat{\beta}_1^2}, \frac{1}{\hat{\beta}_1}\right)'\right]^{-1}$$
$$\left(\hat{\rho} + \frac{\hat{\beta}_2}{\hat{\beta}_1}\right).$$

for H^D, $(\beta_1\rho/\beta_2) + 1 = 0$, we have

$$r(\beta) = (\beta_1\rho/\beta_2) + 1 \qquad \text{and} \qquad R(\beta) = \left(\frac{\beta_1}{\beta_2}, \frac{\rho}{\beta_2}, -\frac{\beta_1\rho}{\beta_2^2}\right).$$

Hence,

$$W^D = \left(\frac{\hat{\beta}_1\hat{\rho}}{\hat{\beta}_2} + 1\right)\left[\left(\frac{\hat{\beta}_1}{\hat{\beta}_2}, \frac{\hat{\rho}}{\hat{\beta}_2}, -\frac{\hat{\beta}_1\hat{\rho}}{\hat{\beta}_2^2}\right)\hat{\sigma}^2(X'X)^{-1}\left(\frac{\hat{\beta}_1}{\hat{\beta}_2}, \frac{\hat{\rho}}{\hat{\beta}_2}, -\frac{\hat{\beta}_1\hat{\rho}}{\hat{\beta}_2^2}\right)'\right]^{-1}$$
$$\left(\frac{\hat{\beta}_1\hat{\rho}}{\hat{\beta}_2} + 1\right).$$

b. Apply these four Wald statistics to the equation relating real per-capita consumption to real per-capita disposable income in the U.S. over the post World War II period 1959–2007. The SAS program that generated these Wald statistics is given below

SAS PROGRAM

```
Data CONSUMP;
Input YEAR Y C;
cards;

PROC IML; USE CONSUMP;
READ ALL VAR {Y C};

Yt=Y[2:NROW(Y)];
YLAG=Y[1:NROW(Y)-1];
Ct=C[2:NROW(C)];
CLAG=C[1:NROW(C)-1];
X=CLAG || Yt || YLAG;
BETA=INV(X`*X)*X`*Ct;
RH0=BETA[1];
BT1=BETA[2];
BT2=BETA[3];

Px=X*INV(X`*X)*X`;
Qx=I(NROW(X))-Px;
et_U=Qx*Ct;
SIG_U=SSQ(et_U)/NROW(X);

Ha=BT1*RHO+BT2;
Hb=BT1+BT2/RHO;
Hc=RHO+BT2/BT1;
Hd=BT1*RHO/BT2+1;
Ra=BT1 || RHO || {1};
Rb=(-BT2/RHO**2) || {1} || (1/RHO);
Rc={1} || (-BT2/BT1**2) || (1/BT1);
Rd=(BT1/BT2) || (RHO/BT2) || (-BT1*RHO/BT2**2);

VAR_a=Ra*SIG_U*INV(X`*X)*Ra`;
VAR_b=Rb*SIG_U*INV(X`*X)*Rb`;
```

```
VAR_c=Rc*SIG_U*INV(X`*X)*Rc`;
VAR_d=Rd*SIG_U*INV(X`*X)*Rd`;

WALD_a=Ha`*INV(VAR_a)*Ha;
WALD_b=Hb`*INV(VAR_b)*Hb;
WALD_c=Hc`*INV(VAR_c)*Hc;
WALD_d=Hd`*INV(VAR_d)*Hd;

PRINT 'Chapter˜7 Problem20. (b)',,WALD_a;
PRINT WALD_b;
PRINT WALD_c;
PRINT WALD_d;
```

7.21 *Effect of Additional Regressors on R^2*. For the regression equation $y = X\beta + u$ the OLS residuals are given by $e = y - X\hat{\beta}_{ols} = \bar{P}_X y$ where $\bar{P}_X = I_n - P_X$, and $P_x = X(X'X)^{-1}X'$ is the projection matrix. Therefore, the SSE for this regression is $e'e = y'\bar{P}_X y$. In particular, $SSE_1 = y'\bar{P}_{X_1}y$, for $X = X_1$ and $SSE_2 = y'\bar{P}_X y$ for $X = (X_1, X_2)$. Therefore,

$$SSE_1 - SSE_2 = y'\left(\bar{P}_{X_1} - \bar{P}_X\right)y = y'\left(P_X - P_{X_1}\right)y = y'Ay$$

where $A = P_X - P_{X_1}$. This difference in the residual sums of squares is non-negative for any vector y because $y'Ay$ is positive semi-definite. The latter result holds because A is symmetric and idempotent. In fact, A is the difference between two idempotent matrices that also satisfy the following property: $P_X P_{X_1} = P_{X_1} P_X = P_{X_1}$. Hence,

$$A^2 = P_X^2 - P_{X_1} - P_{X_1} + P_{X_1}^2 = P_X - P_{X_1} = A.$$

$R^2 = 1$ - (SSE/TSS) where TSS is the total sum of squares to be explained by the regression and this depends only on the y's. TSS is fixed for both regressions. Hence $R_2^2 \geq R_1^2$, since $SSE_1 \geq SSE_2$.

References

Baltagi, B. H. (1996), “Iterative Estimation in Partitioned Regression Models,” *Econometric Theory*, Solutions 95.5.1, 12:869–870.

Engle, R.F. (1984), “Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics,” In: Griliches, Z. and M.D. Intrilligator (eds) Handbook of Econometrics (North-Holland: Amsterdam).

Graybill, F.A. (1961), An Introduction to Linear Statistical Models, Vol. 1 (McGraw-Hill: New York).

Gregory, A.W. and M.R. Veall (1985), “Formulating Wald Tests of Nonlinear Restrictions,” Econometrica, 53: 1465–1468.

Gregory, A.W. and M.R. Veall (1986), “Wald Tests of Common Factor Restrictions,” Economics Letters, 22: 203–208.

Maddala, G.S. (1992), Introduction to Econometrics (Macmillan: New York).

Salkever, D. (1976), “The Use of Dummy Variables to Compute Predictions, Prediction Errors, and Confidence Intervals,” Journal of Econometrics, 4: 393–397.

CHAPTER 8
Regression Diagnostics and Specification Tests

8.1 Since $H = P_X$ is idempotent, it is positive semi-definite with $b'H\,b \geq 0$ for any arbitrary vector b. Specifically, for $b' = (1, 0, .., 0)$ we get $h_{11} \geq 0$. Also, $H^2 = H$. Hence,

$$h_{11} = \sum_{j=1}^{n} h_{1j}{}^2 \geq h_{11}^2 \geq 0.$$

From this inequality, we deduce that $h_{11}^2 - h_{11} \leq 0$ or that $h_{11}(h_{11} - 1) \leq 0$. But $h_{11} \geq 0$, hence $0 \leq h_{11} \leq 1$. There is nothing particular about our choice of h_{11}. The same proof holds for h_{22} or h_{33} or in general h_{ii}. Hence, $0 \leq h_{ii} \leq 1$ for $i = 1, 2, .., n$.

8.2 *A Simple Regression With No Intercept*. Consider

$$y_i = x_i\beta + u_i \text{ for } i = 1, 2, .., n$$

a. $H = P_x = x(x'x)^{-1}x' = xx'/x'x$ since $x'x$ is a scalar. Therefore, $h_{ii} = x_i^2/\sum_{i=1}^{n} x_i^2$ for $i = 1, 2, .., n$. Note that the x_i's are not in deviation form as in the case of a simple regression with an intercept. In this case $tr(H) = tr(P_x) = tr(xx')/x'x = tr(x'x)/x'x = x'x/x'x = 1$. Hence, $\sum_{i=1}^{n} h_{ii} = 1$.

b. From (8.13), $\hat{\beta} - \hat{\beta}_{(i)} = \dfrac{(x'x)^{-1}x_ie_i}{1 - h_{ii}} = \dfrac{x_ie_i}{\sum_{j=1}^{n} x_j{}^2 - x_i{}^2} = \dfrac{x_ie_i}{\sum_{j\neq i} x_j{}^2}$.

From (8.18), (n-2) $s_{(i)}{}^2 = (n-1)s^2 - \frac{e_i^2}{1-h_{ii}} = (n-1)s^2 - e_i^2\left(\sum_{i=1}^{n} x_i{}^2/\sum_{j\neq i} x_j{}^2\right)$.

From (8.19), $\text{DFBETAS}_i = (\hat{\beta} - \hat{\beta}_{(i)})/s_{(i)}\sqrt{(x'x)^{-1}} = \dfrac{x_ie_i}{\sum_{j\neq i} x_j{}^2} \cdot \left(\sum_{i=1}^{n} x_i{}^2\right)^{1/2} \Big/ s_{(i)}$.

c. From (8.21), $\text{DFFIT}_i = \hat{y}_i - \hat{y}_{(i)} = x_i'[\hat{\beta} - \hat{\beta}_{(i)}] = \dfrac{h_{ii}e_i}{(1 - h_{ii})} = \dfrac{x_i^2e_i}{\sum_{j\neq i} x_j^2}$.

B.H. Baltagi, *Solutions Manual for Econometrics*, Springer Texts in Business and Economics, DOI 10.1007/978-3-642-54548-1_8,

From (8.22),

$$\mathrm{DFFITS}_i = \left(\frac{h_{ii}}{1-h_{ii}}\right)^{1/2}\frac{e_i}{s_{(i)}\sqrt{1-h_{ii}}} = \left(\frac{x_i^2}{\sum\limits_{j\neq i}x_j^2}\right)^{1/2}\frac{e_i}{s_{(i)}\left(\sum\limits_{j\neq i}x_j^2\Big/\sum\limits_{i=1}^{n}x_i^2\right)^{1/2}}$$

$$= \frac{x_ie_i\left(\sum\limits_{i=1}^{n}x_i^2\right)^{1/2}}{s_{(i)}\sum\limits_{j\neq i}x_j^2}$$

d. From (8.24), $D_i^2(s) = \dfrac{(\hat{\beta}-\hat{\beta}_{(i)})^2\sum\limits_{i=1}^{n}x_i^2}{ks^2} = \dfrac{x_i^2e_i^2}{\left(\sum\limits_{j\neq i}x_j^2\right)^2}\cdot\dfrac{\sum\limits_{i=1}^{n}x_i^2}{ks^2}$ from part (b).

From (8.25), $D_i^2(s) = \dfrac{e_i^2}{ks^2}\left(\dfrac{x_i^2\sum\limits_{j=1}^{n}x_j^2}{\left(\sum\limits_{j\neq i}x_j^2\right)^2}\right)$.

e. $\det\left[X_{(i)}'X_{(i)}\right] = x_{(i)}'x_{(i)} = \sum\limits_{j\neq i}x_j^2$, while

$$\det\left[X'X\right] = x'x = \sum_{i=1}^{n}x_i^2 \text{ and } (1-h_{ii}) = 1-\left(x_i^2\Big/\sum_{i=1}^{n}x_i^2\right) = \sum_{j\neq i}x_j^2\Big/\sum_{i=1}^{n}x_i^2.$$

The last term is the ratio of the two determinants. Rearranging terms, one can easily verify (8.27). From (8.26),

$$\mathrm{COVRATIO}_i = \frac{s_{(i)}^2}{s^2}\left[\frac{\sum\limits_{i=1}^{n}x_i^2}{\sum\limits_{j\neq i}x_j^2}\right].$$

8.3 From (8.17) $s_{(i)}^2 = \frac{1}{n-k-1}\sum\limits_{t\neq i}\left(y_t - x_t'\hat{\beta}_{(i)}\right)^2$ substituting (8.13), one gets

$$s_{(i)}^2 = \frac{1}{n-k-1}\sum_{t\neq i}\left(y_t - x_t'\hat{\beta} + \frac{x_t'(X'X)^{-1}x_ie_i}{1-h_{ii}}\right)^2$$

$$\text{or}\quad (n-k-1)s_{(i)}^2 = \sum_{t\neq i}\left(e_t + \frac{h_{it}e_i}{1-h_{ii}}\right)^2$$

where $h_{it} = x_t'(X'X)^{-1}x_i$. Adding and subtracting the i-th term of this summation, yields

$$(n-k-1)s_{(i)}^2 = \sum_{t=1}^{n}\left(e_t + \frac{h_{it}e_i}{1-h_{ii}}\right)^2 - \left(e_i + \frac{h_{ii}e_i}{1-h_{ii}}\right)^2$$

$$= \sum_{t=1}^{n}\left(e_t + \frac{h_{it}e_i}{1-h_{ii}}\right)^2 - \frac{e_i^2}{(1-h_{ii})^2}$$

which is the first equation in (8.18). The next two equations in (8.18) simply expand this square and substitute $\sum_{t=1}^{n} e_t h_{it} = 0$ and $\sum_{t=1}^{n} {h_{it}}^2 = h_{ii}$ which follow from the fact that $He = 0$ and $H^2 = H$.

8.4 Obtaining $e_i{}^*$ from an Augmented Regression

a. In order to get $\hat{\beta}^*$ from the augmented regression given in (8.5), one can premultiply by $\overline{P}_{d_i}$ as described in (8.14) and perform OLS. The Frisch–Waugh–Lovell Theorem guarantees that the resulting estimate of β^* is the same as that from (8.5). The effect of $\overline{P}_{d_i}$ is to wipe out the i-th observation from the regression and hence $\hat{\beta}^* = \hat{\beta}_{(i)}$ as required.

b. In order to get $\hat{\varphi}$ from (8.5), one can premultiply by $\overline{P}_x$ and perform OLS on the transformed equation $\overline{P}_x y = \overline{P}_x d_i \varphi + \overline{P}_x u$. The Frisch–Waugh–Lovell Theorem guarantees that the resulting estimate of φ is the same as that from (8.5). OLS yields $\hat{\varphi} = \left(d_i'\overline{P}_x d_i\right)^{-1} d_i'\overline{P}_x y$. Using the fact that $e = \overline{P}_x y$, $d_i'e = e_i$ and $d_i'P_x d_i = d_i'Hd_i = h_{ii}$ one gets $\hat{\varphi} = e_i/(1-h_{ii})$ as required.

c. The Frisch–Waugh–Lovell Theorem also states that the residuals from (8.5) are the same as those obtained from (8.14). This means that the i-th observation residual is zero. This also gives us the fact that $\hat{\varphi} = y_i - x_i'\hat{\beta}_{(i)} =$ the forecasted residual for the i-th observation, see below (8.14). Hence the RSS from (8.5) $= \sum_{t \neq i}\left(y_t - x_t'\hat{\beta}_{(i)}\right)^2$ since the i-th observation contributes a zero residual. As in (8.17) and (8.18) and problem 8.3 one can substitute (8.13) to get

$$\sum_{t\neq i}\left(y_t - x_t'\hat{\beta}_{(i)}\right)^2 = \sum_{t\neq i}\left(y_t - x_t'\hat{\beta} + \frac{x_t'(X'X)^{-1}x_i e_i}{1-h_{ii}}\right)^2 = \sum_{t\neq i}\left(e_t + \frac{h_{it}e_i}{1-h_{ii}}\right)^2$$

Using the same derivation as in the solution to problem 8.3, one gets $\sum_{t\neq i}\left(y_t - x_t'\hat{\beta}_{(i)}\right)^2 = \sum_{t=1}^{n} e_t^2 - \frac{e_i^2}{1-h_{ii}}$ which is an alternative way of writing (8.18).

d. Under Normality of the disturbances, $u \sim N(0, \sigma^2 I_n)$, we get the standard t-statistic on $\hat{\varphi}$ as $t = \hat{\varphi}/\text{s.e.}(\hat{\varphi}) \sim t_{n-k-1}$ under the null hypothesis that $\varphi = 0$. But $\hat{\varphi} = \left(d_i'\overline{P}_x d_i\right)^{-1} d_i'\overline{P}_x y = \varphi + \left(d_i'\overline{P}_x d_i\right)^{-1} d_i'\overline{P}_x u$ with $E(\hat{\varphi}) = \varphi$ and $\text{var}(\hat{\varphi}) = \sigma^2\left(d_i'\overline{P}_x d_i\right)^{-1} = \sigma^2/(1-h_{ii})$. This is estimated by $\widehat{\text{var}}(\hat{\varphi}) = s_{(i)}^2/(1-h_{ii})$. Hence, $t = \hat{\varphi}/\text{s.e.}(\hat{\varphi}) = \frac{e_i}{1-h_{ii}} \cdot \left(\frac{1-h_{ii}}{s_{(i)}^2}\right)^{1/2} = \frac{e_i}{s_{(i)}\sqrt{1-h_{ii}}} = e_i^*$ as in (8.3).

8.5 a. Applying OLS on this regression equation yields

$$\begin{pmatrix}\hat{\beta}^* \\ \hat{\varphi}^*\end{pmatrix} = \begin{bmatrix} X'X & 0 \\ 0 & D_p'\overline{P}_x D_p \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ D_p'\overline{P}_x y \end{bmatrix}$$

using the fact that $\overline{P}_x X = 0$. Hence $\hat{\beta}^* = (X'X)^{-1}X'y = \hat{\beta}_{ols}$ and

b. $\hat{\varphi}^* = \left(D_p'\overline{P}_x D_p\right)^{-1} D_p'\overline{P}_x y = \left(D_p'\overline{P}_x D_p\right)^{-1} D_p'e = \left(D_p'\overline{P}_x D_p\right)^{-1} e_p$ since $e = \overline{P}_x y$ and $D_p'e = e_p$

c. The residuals are given by

$$y - X\hat{\beta}_{ols} - \overline{P}_x D_p\left(D_p'\overline{P}_x D_p\right)^{-1} e_p = e - \overline{P}_x D_p\left(D_p'\overline{P}_x D_p\right)^{-1} D_p'e$$

so that the residuals sum of squares is

$$e'e + e_p'(D_p'\overline{P}_x D_p)^{-1} D_p'\overline{P}_x D_p\left(D_p'\overline{P}_x D_p\right)^{-1} e_p - 2e'\overline{P}_x D_p\left(D_p'\overline{P}_x D_p\right)^{-1} e_p$$
$$= e'e - e_p'\left(D_p'\overline{P}_x D_p\right)^{-1} e_p$$

since $\overline{P}_x e = e$ and $e'D_p = e_p'$. From the Frisch–Waugh–Lovell Theorem, (8.6) has the same residuals sum of squares as (8.6) premultiplied by $\overline{P}_x$, i.e., $\overline{P}_x y = \overline{P}_x D_p \varphi^* + \overline{P}_x u$. This has also the same residuals sum of squares as the augmented regression in problem 8.5, since this also yields the above regression when premultiplied by $\overline{P}_x$.

d. From (8.7), the denominator residual sum of squares is given in part (c). The numerator residual sum of squares is $e'e$—the RSS obtained in part (c). This yields $e'_p\left(D'_p\overline{P}_xD_p\right)^{-1}e_p$ as required.

e. For problem 8.4, consider the augmented regression $y = X\beta^* + \overline{P}_x d_i\varphi + u$. This has the same residual sum of squares by the Frisch–Waugh–Lovell Theorem as the following regression:

$$\overline{P}_x y = \overline{P}_x d_i\varphi + \overline{P}_x u$$

This last regression also have the same residual sum of squares as

$$y = X\beta^* + d_i\varphi + u$$

Hence, using the first regression and the results in part (c) we get Residual Sum of Squares = (Residual Sum of Squares with d_i deleted)

$$-e'd_i\left(d'_i\overline{P}_xd_i\right)^{-1}d'_ie$$

when D_p is replaced by d_i. But $d'_ie = e_i$ and $d'_i\overline{P}_xd_i = 1 - h_{ii}$, hence this last term is $e_i^2/(1-h_{ii})$ which is exactly what we proved in problem 8.4, part (c). The F-statistic for $\varphi = 0$, would be exactly the square of the t-statistic in problem 8.4, part (d).

8.6 Let $A = X'X$ and $a = b = x_i'$. Using the updated formula

$$(A - a'b)^{-1} = A^{-1} + A^{-1}a'\left(I - bA^{-1}a'\right)^{-1}bA^{-1}$$

$$\left(X'X - x_ix_i'\right)^{-1} = \left(X'X\right)^{-1} + \left(X'X\right)^{-1}x_i\left(1 - x_i'\left(X'X\right)^{-1}x_i\right)^{-1}x_i'\left(X'X\right)^{-1}$$

Note that $X' = [x_1, .., x_n]$ where x_i is kxl, so that

$$X'X = \sum_{j=1}^{n} x_jx_j' \quad \text{and} \quad X'_{(i)}X_{(i)} = \sum_{j\neq i} x_jx_j' = X'X - x_ix_i'.$$

Therefore,

$$\left(X'_{(i)}X_{(i)}\right)^{-1} = \left(X'X\right)^{-1} + \frac{\left(X'X\right)^{-1}x_ix_i'\left(X'X\right)^{-1}}{1 - h_{ii}}$$

where $h_{ii} = x_i'(X'X)^{-1}x_i$. This verifies (8.12). Note that $X'y = \sum_{j=1}^{n} x_j y_j$ and

$$X'_{(i)}y_{(i)} = \sum_{j \neq i} x_j y_j = X'y - x_i y_i.$$

Therefore, post-multiplying $\left(X'_{(i)}X_{(i)}\right)^{-1}$ by $X'_{(i)}y_{(i)}$ we get

$$\begin{aligned}\hat{\beta}_{(i)} &= \left(X'_{(i)}X_{(i)}\right)^{-1} X'_{(i)}y_{(i)} = \hat{\beta} - \left(X'X\right)^{-1} x_i y_i + \frac{(X'X)^{-1} x_i x_i'\hat{\beta}}{1-h_{ii}} \\ &\quad - \frac{(X'X)^{-1} x_i x_i' (X'X)^{-1} x_i y_i}{1-h_{ii}}\end{aligned}$$

and

$$\begin{aligned}\hat{\beta} - \hat{\beta}_{(i)} &= -\frac{(X'X)^{-1}x_i x_i'\hat{\beta}}{1-h_{ii}} + \left(\frac{1-h_{ii}+h_{ii}}{1-h_{ii}}\right)\left(X'X\right)^{-1} x_i y_i \\ &= \frac{(X'X)^{-1} x_i \left(y - x_i'\hat{\beta}\right)}{1-h_{ii}} = \frac{(X'X)^{-1} x_i e_i}{1-h_{ii}}\end{aligned}$$

which verifies (8.13).

8.7 From (8.22),

$$\text{DFFITS}_i(\sigma) = \left(\frac{h_{ii}}{1-h_{ii}}\right)^{1/2} \frac{e_i}{\sigma\sqrt{1-h_{ii}}} = \frac{\sqrt{h_{ii}}e_i}{\sigma(1-h_{ii})}$$

while from (8.25)

$$\sqrt{k}D_i(\sigma) = \frac{e_i}{\sigma}\left(\frac{\sqrt{h_{ii}}}{1-h_{ii}}\right)$$

which is identical.

8.8 $\det\left[X'_{(i)}X_{(i)}\right] = \det\left[X'X - x_i x_i'\right] = \det\left[\left\{I_k - x_i x_i'(X'X)^{-1}\right\}X'X\right]$. Let $a = x_i$ and $b' = x_i'(X'X)^{-1}$. Using $\det\left[I_k - ab'\right] = 1 - b'a$ one gets $\det\left[I_k - x_i x_i'(X'X)^{-1}\right] = 1 - x_i'(X'X)^{-1}x_i = 1 - h_{ii}$. Using $\det(AB) = \det(A)\det(B)$, one gets $\det\left[X'_{(i)}X_{(i)}\right] = (1-h_{ii})\det(X'X)$ which verifies (8.27).

8.9 The cigarette data example given in Table 3.2.

a. Tables 8.1 and 8.2 were generated by SAS with PROC REG asking for all the diagnostic options allowed by this procedure.

b. For the New Hampshire (NH) observation number 27 in Table 8.2, the leverage $h_{NH} = 0.13081$ which is larger than $2\bar{h} = 2k/n = 0.13043$. The internally studentized residual $\tilde{e}_{NH}$ is computed from (8.1) as follows:

$$\tilde{e}_{NH} = \frac{e_{NH}}{s\sqrt{1-h_{NH}}} = \frac{0.15991}{0.16343\sqrt{1-0.13081}} = 1.0495$$

The externally studentized residual e^*_{NH} is computed from (8.3) as follows:

$$e^*_{(NH)} = \frac{e_{NH}}{s_{(NH)}\sqrt{1-h_{NH}}} = \frac{0.15991}{0.163235\sqrt{1-0.13081}} = 1.0508$$

where $s^2_{(NH)}$ is obtained from (8.2) as follows:

$$\begin{aligned} s^2_{(NH)} &= \frac{(n-k)s^2 - e^2_{NH}/(1-h_{NH})}{(n-k-1)} \\ &= \frac{(46-3)(0.16343)^2 - (0.15991)^2/(1-0.13081)}{(46-3-1)} \\ &= \frac{1.14854 - 0.02941945}{42} = 0.0266457 \end{aligned}$$

both $\tilde{e}_{NH}$ and $e^*_{(NH)}$ are less than 2 in absolute value. From (8.13), the change in the regression coefficients due to the omission of the NH observation is given by $\hat{\beta} - \hat{\beta}_{(NH)} = (X'X)^{-1}x_{NH}e_{NH}/(1-h_{NH})$ where $(X'X)^{-1}$ is given in the empirical example and $x'_{NH} = (1, 0.15852, 5.00319)$ with $e_{NH} = 0.15991$ and $h_{NH} = 0.13081$. This gives $(\hat{\beta} - \hat{\beta}_{(NH)})' = (-0.3174, -0.0834, 0.0709)$. In order to assess whether this change is large or small, we compute DFBETAS given in (8.19). For the NH observation, these are given by

$$\text{DFBETAS}_{NH,1} = \frac{\hat{\beta}_1 - \hat{\beta}_{1,(NH)}}{s_{(NH)}\sqrt{(X'X)^{-1}_{11}}} = \frac{-0.3174}{0.163235\sqrt{30.9298169}} = -0.34967$$

Similarly, $\text{DFBETAS}_{NH,2} = -0.2573$ and $\text{DFBETAS}_{NH,3} = 0.3608$. These are not larger than 2 in absolute value. However, $\text{DFBETAS}_{NH,1}$ and $\text{DFBETAS}_{NH,3}$ are both larger than $2/\sqrt{n} = 2/\sqrt{46} = 0.2949$ in absolute value.

The change in the fit due to omission of the NH observation is given by (8.21). In fact,

$$\begin{aligned}\mathrm{DFFIT}_{\mathrm{NH}} &= \hat{y}_{\mathrm{NH}} - \hat{y}_{(\mathrm{NH})} = x'_{\mathrm{NH}}[\hat{\beta} - \hat{\beta}_{(\mathrm{NH})}]\\ &= (1,\ 0.15952,\ 5.00319)\begin{pmatrix}-0.3174\\ -0.0834\\ 0.0709\end{pmatrix} = 0.02407.\end{aligned}$$

or simply

$$\mathrm{DFFIT}_{\mathrm{NH}} = \frac{h_{\mathrm{NH}}e_{\mathrm{NH}}}{1 - h_{\mathrm{NH}}} = \frac{(0.13081)(0.15991)}{(1 - 0.13081)} = 0.02407.$$

Scaling it by the variance of $\hat{y}_{(\mathrm{NH})}$ we get from (8.22)

$$\begin{aligned}\mathrm{DFFIT}_{\mathrm{NH}} &= \left(\frac{h_{\mathrm{NH}}}{(1 - h_{\mathrm{NH}})}\right)^{1/2} e^{*}_{\mathrm{NH}}\\ &= \left(\frac{0.13081}{1 - 0.13081}\right)^{1/2}(1.0508) = 0.40764.\end{aligned}$$

This is not larger than the size adjusted cutoff of $2/\sqrt{k/n} = 0.511$. Cook's distance measure is given by (8.25) and for NH can be computed as follows:

$$\begin{aligned}D^2_{\mathrm{NH}}(s) &= \frac{e^2_{\mathrm{NH}}}{ks^2}\left(\frac{h_{\mathrm{NH}}}{(1 - h_{\mathrm{NH}})^2}\right) = \left(\frac{(0.15991)^2}{3(0.16343)^2}\right)\left(\frac{0.13081}{(1 - 0.13081)^2}\right)\\ &= 0.05526\end{aligned}$$

COVRATIO omitting the NH observation can be computed from (8.28) as follows:

$$\begin{aligned}\mathrm{COVRATIO}_{\mathrm{NH}} &= \left(\frac{s^2_{(\mathrm{NH})}}{s^2}\right)^{k}\frac{1}{1 - h_{\mathrm{NH}}} = \left(\frac{0.0266457}{(0.16343)^2}\right)^{3}\left(\frac{1}{(1 - 0.13081)}\right)\\ &= 1.1422\end{aligned}$$

which means that $|\mathrm{COVRATIO}_{\mathrm{NH}} - 1| = 0.1422$ is less than $3k/n = 0.1956$. Finally, FVARATIO omitting the NH observation can be computed from (8.29) as

$$\mathrm{FVARATIO}_{\mathrm{NH}} = \frac{s^2_{(\mathrm{NH})}}{s^2(1 - h_{\mathrm{NH}})} = \frac{0.0266457}{(0.16343)^2(1 - 0.13081)} = 1.14775$$

c. Similarly, the same calculations can be obtained for the observations of the states of AR, CT, NJ and UT.

d. Also, the states of NV, ME, NM and ND.

8.10 The Consumption—Income data given in Table 5.3.

The following Stata output runs the ols regression of C on Y and generates the influence diagnostics one by one. In addition, it highlights the observations with the diagnostic higher than the cut off value recommended in the text of the chapter.

```
. reg c y

Source   |        SS         df          MS             Number of obs  =       49
----------------------------------------------          F(1, 47)       =  7389.28
Model    |   1.4152e+09       1     1.4152e+09          Prob > F       =   0.0000
Residual |   9001347.76      47     191518.037          R-squared      =   0.9937
----------------------------------------------          Adj R-squared  =   0.9935
Total    |   1.4242e+09      48     29670457.8          Root MSE       =   437.63

------------------------------------------------------------------------------
       c |      Coef.    Std. Err.       t     P> |t|     [95% Conf. Interval]
------------------------------------------------------------------------------
       y |    .979228     .0113915    85.96    0.000     .9563111     1.002145
   _cons |  -1343.314     219.5614    -6.12    0.000    -1785.014    -901.6131
------------------------------------------------------------------------------
. predict h, hat

. predict e if e(sample), residual

. predict est1 if e(sample), rstandard

. predict est2 if e(sample), rstudent

. predict dfits, dfits

. predict covr, covratio

. predict cook if e(sample), cooksd

. predict dfbet, dfbeta(y)

. list e est1 est2 cook h dfits covr, divider
```

	e	est1	est2	cook	h	dfits	covr
1.	635.4909	1.508036	1.529366	.0892458	.0727745	.4284583	1.019567
2.	647.5295	1.536112	1.55933	.0917852	.0721806	.4349271	1.015024
3.	520.9777	1.234602	1.241698	.0575692	.0702329	.3412708	1.051163
4.	498.7493	1.179571	1.184622	.0495726	.0665165	.3162213	1.053104
5.	517.4854	1.222239	1.228853	.0510749	.064003	.3213385	1.045562
6.	351.0732	.8263996	.8235661	.0208956	.0576647	.2037279	1.075873
7.	321.1447	.7538875	.7503751	.0157461	.0525012	.1766337	1.075311
8.	321.9283	.7540409	.7505296	.0144151	.0482588	.1690038	1.070507
9.	139.071	.3251773	.3220618	.0024888	.0449572	.0698759	1.08818
10.	229.1068	.5347434	.5306408	.0061962	.0415371	.1104667	1.07598
11.	273.736	.6382437	.6341716	.0083841	.0395361	.1286659	1.068164
12.	17.04481	.0396845	.0392607	.0000301	.0367645	.0076702	1.083723
13.	-110.857	-.25773	-.2551538	.0011682	.0339782	-.047853	1.077618
14.	.7574701	.0017584	.0017396	4.95e-08	.0310562	.0003114	1.077411
15.	-311.9394	-.7226375	-.7189135	.0072596	.0270514	-.1198744	1.049266
16.	-289.1532	-.6702327	-.6662558	.006508	.0281591	-.1134104	1.053764
17.	-310.0611	-.7183715	-.7146223	.0072371	.0272825	-.119681	1.049793
18.	-148.776	-.3443774	-.3411247	.0015509	.0254884	-.0551685	1.065856
19.	-85.67493	-.1981783	-.1961407	.0004859	.0241443	-.0308519	1.067993
20.	-176.7102	-.4084195	-.4047702	.0019229	.0225362	-.0614609	1.060452
21.	-205.995	-.4759794	-.4720276	.0025513	.022026	-.0708388	1.057196
22.	-428.808	-.9908097	-.9906127	.0110453	.0220072	-.1485998	1.023316
23.	-637.0542	-1.471591	-1.490597	.0237717	.0214825	-.2208608	.9708202
24.	-768.879	-1.77582	-1.818907	.034097	.0211669	-.267476	.928207
25.	-458.3625	-1.058388	-1.059773	.0118348	.0206929	-.1540507	1.015801
26.	-929.7892	-2.146842	-2.23636	.0484752	.020602	-.3243518	.8671094
27.	-688.1302	-1.589254	-1.616285	.0272016	.0210855	-.2372122	.9548985
28.	-579.1982	-1.338157	-1.349808	.019947	.0217933	-.2014737	.9874383
29.	-317.75	-.734245	-.7305942	.0061013	.0221336	-.1099165	1.043229
30.	-411.8743	-.9525952	-.9516382	.0111001	.0238806	-.1488479	1.028592
31.	-439.981	-1.01826	-1.018669	.0133716	.0251442	-.1635992	1.02415
32.	-428.6367	-.9923085	-.9921432	.0130068	.0257385	-.1612607	1.027102
33.	-479.2094	-1.109026	-1.111808	.0158363	.0251048	-.1784142	1.015522
34.	-549.0905	-1.271857	-1.280483	.0222745	.0268017	-.2124977	1.000132
35.	-110.233	-.2553027	-.2527474	.0008897	.0265748	-.041761	1.069479
36.	66.3933	.1538773	.1522699	.0003404	.0279479	.0258193	1.072884
37.	32.47656	.0753326	.0745313	.0000865	.0295681	.0130097	1.075499
38.	100.64	.2336927	.2313277	.0008919	.031631	.0418084	1.075547
39.	122.4207	.2847169	.2819149	.001456	.0346758	.0534312	1.077724
40.	-103.4364	-.2414919	-.2390574	.0012808	.0420747	-.0501011	1.087101

41.	339.5579	.7941921	.7910236	.0150403	.0455199	.1727454	1.064579
42.	262.4189	.6163763	.6122634	.010752	.0535693	.145664	1.085279
43.	395.0927	.9290681	.9276894	.025475	.0557367	.2253859	1.065336
44.	282.3303	.6658444	.6618519	.0144581	.0612288	.1690281	1.091159
45.	402.1435	.9501951	.9491935	.0312551	.0647518	.2497567	1.073755
46.	385.8501	.9156712	.9140675	.0329419	.0728531	.2562288	1.086167
47.	799.5297	1.898972	1.955164	.1449345	.0744023	.5543262	.9614341
48.	663.9663	1.584488	1.611164	.1138252	.0831371	.4851599	1.020218
49.	642.6846	1.539513	1.562965	.117268	.0900456	.4916666	1.034467

. list year c est2 h dfbet if dfbet>(2/7)

	year	c	est2	h	dfbet
47.	2005	26290	1.955164	.0744023	.4722213
48.	2006	26835	1.611164	.0831371	.4214261
49.	2007	27319	1.562965	.0900456	.4323753

. list year c est2 dfbet h if abs(est2)>2

	year	c	est2	dfbet	h
26.	1984	16343	-2.23636	-.0314578	.020602

. list year c est2 dfbet h if h>(4/49)

	year	c	est2	dfbet	h
48.	2006	26835	1.611164	.4214261	.0831371
49.	2007	27319	1.562965	.4323753	.0900456

. list year c est2 dfbet h if abs(est2)>2

	year	c	est2	dfbet	h
26.	1984	16343	-2.23636	-.0314578	.020602

. list year c est2 dfbet h dfits if dfits>(2*sqrt(2/49))

	year	c	est2	dfbet	h	dfits
1.	1959	8776	1.529366	-.3634503	.0727745	.4284583
2.	1960	8837	1.55933	-.3683456	.0721806	.4349271
47.	2005	26290	1.955164	.4722213	.0744023	.5543262
48.	2006	26835	1.611164	.4214261	.0831371	.4851599
49.	2007	27319	1.562965	.4323753	.0900456	.4916666

```
. list year c est2 dfbet h covr if abs(covr-1)>(3*(2/49))
```

	year	c	est2	dfbet	h	covr
26.	1984	16343	-2.23636	-.0314578	.020602	.8671094

8.12 The Gasoline data used in Chap. 10. The following SAS output gives the diagnostics for two countries: Austria and Belgium.

a. AUSTRIA:

Obs	Dep Var Y	Predict Value	Std Err Predict	Lower95% Mean	Upper95% Mean	Lower95% Predict	Upper95% Predict
1	4.1732	4.1443	0.026	4.0891	4.1996	4.0442	4.2445
2	4.1010	4.1121	0.021	4.0680	4.1563	4.0176	4.2066
3	4.0732	4.0700	0.016	4.0359	4.1041	3.9798	4.1602
4	4.0595	4.0661	0.014	4.0362	4.0960	3.9774	4.1548
5	4.0377	4.0728	0.012	4.0469	4.0987	3.9854	4.1603
6	4.0340	4.0756	0.013	4.0480	4.1032	3.9877	4.1636
7	4.0475	4.0296	0.014	3.9991	4.0601	3.9407	4.1185
8	4.0529	4.0298	0.016	3.9947	4.0650	3.9392	4.1205
9	4.0455	4.0191	0.019	3.9794	4.0587	3.9266	4.1115
10	4.0464	4.0583	0.015	4.0267	4.0898	3.9690	4.1475
11	4.0809	4.1108	0.016	4.0767	4.1450	4.0206	4.2011
12	4.1067	4.1378	0.022	4.0909	4.1847	4.0420	4.2336
13	4.1280	4.0885	0.015	4.0561	4.1210	3.9989	4.1782
14	4.1994	4.1258	0.022	4.0796	4.1721	4.0303	4.2214
15	4.0185	4.0270	0.017	3.9907	4.0632	3.9359	4.1180
16	4.0290	3.9799	0.016	3.9449	4.0149	3.8893	4.0704
17	3.9854	4.0287	0.017	3.9931	4.0642	3.9379	4.1195
18	3.9317	3.9125	0.024	3.8606	3.9643	3.8141	4.0108
19	3.9227	3.9845	0.019	3.9439	4.0250	3.8916	4.0773

Obs	Residual	Std Err Residual	Student Residual	-2-1-0 1 2	Cook's D	Rstudent
1	0.0289	0.029	0.983	*	0.188	0.9821
2	-0.0111	0.033	-0.335		0.011	-0.3246
3	0.00316	0.036	0.088		0.000	0.0853
4	-0.00661	0.037	-0.180		0.001	-0.1746
5	-0.0351	0.037	-0.942	*	0.024	-0.9387
6	-0.0417	0.037	-1.126	**	0.039	-1.1370
7	0.0179	0.036	0.491		0.009	0.4787
8	0.0231	0.036	0.649	*	0.023	0.6360
9	0.0264	0.034	0.766	*	0.043	0.7554
10	-0.0119	0.036	-0.328		0.004	-0.3178
11	-0.0300	0.036	-0.838	*	0.035	-0.8287
12	-0.0310	0.032	-0.957	*	0.105	-0.9541
13	0.0395	0.036	1.093	**	0.053	1.1008
14	0.0735	0.033	2.254	****	0.563	2.6772
15	-0.00846	0.035	-0.239		0.003	-0.2318
16	0.0492	0.036	1.381	**	0.101	1.4281
17	-0.0433	0.035	-1.220	**	0.082	-1.2415
18	0.0192	0.031	0.625	*	0.061	0.6118
19	-0.0617	0.034	-1.802	***	0.250	-1.9663

Obs	Hat Diag H	Cov Ratio	Dffits	INTERCEP Dfbetas	X1 Dfbetas	X2 Dfbetas	X3 Dfbetas
1	0.4377	1.7954	0.8666	0.2112	0.4940	-0.0549	-0.5587
2	0.2795	1.7750	-0.2022	-0.0321	-0.0859	-0.0141	0.1009
3	0.1665	1.5779	0.0381	-0.0095	-0.0042	0.0134	0.0008
4	0.1279	1.4980	-0.0669	0.0284	0.0188	-0.0202	-0.0118
5	0.0959	1.1418	-0.3057	0.1381	0.0735	-0.0382	-0.0387
6	0.1091	1.0390	-0.3980	0.2743	0.1728	0.0437	-0.1219
7	0.1331	1.4246	0.1876	-0.1200	-0.1390	0.0742	0.1286
8	0.1767	1.4284	0.2947	-0.2174	-0.2359	0.0665	0.2176
9	0.2254	1.4500	0.4075	-0.2985	-0.3416	0.0910	0.3220
10	0.1424	1.4930	-0.1295	0.0812	0.0693	0.0280	-0.0607
11	0.1669	1.3061	-0.3710	0.0513	-0.0795	0.2901	0.1027
12	0.3150	1.4954	-0.6470	0.0762	-0.1710	0.5599	0.2111
13	0.1508	1.1134	0.4639	0.0284	0.1127	-0.3194	-0.1143
14	0.3070	0.3639	1.7819	0.9050	1.3184	-1.3258	-1.2772
15	0.1884	1.5990	-0.1117	-0.0945	-0.0574	-0.0153	0.0407
16	0.1754	0.9276	0.6587	0.3450	-0.0416	0.3722	0.1521
17	0.1810	1.0595	-0.5836	-0.4710	-0.2819	-0.0069	0.1923
18	0.3854	1.9295	0.4845	0.1093	-0.2126	0.3851	0.2859
19	0.2357	0.6501	-1.0918	-0.8004	-0.2221	-0.4001	0.0226

Sum of Residuals	0
Sum of Squared Residuals	0.0230
Predicted Resid SS (Press)	0.0399

b. BELGIUM:

Obs	Dep Var Y	Predict Value	Std Err Predict	Lower95% Mean	Upper95% Mean	Lower95% Predict	Upper95% Predict
1	4.1640	4.1311	0.019	4.0907	4.1715	4.0477	4.2144
2	4.1244	4.0947	0.017	4.0593	4.1301	4.0137	4.1757
3	4.0760	4.0794	0.015	4.0471	4.1117	3.9997	4.1591
4	4.0013	4.0412	0.013	4.0136	4.0688	3.9632	4.1192
5	3.9944	4.0172	0.012	3.9924	4.0420	3.9402	4.0942
6	3.9515	3.9485	0.015	3.9156	3.9814	3.8685	4.0285
7	3.8205	3.8823	0.017	3.8458	3.9189	3.8008	3.9639
8	3.9069	3.9045	0.012	3.8782	3.9309	3.8270	3.9820
9	3.8287	3.8242	0.019	3.7842	3.8643	3.7410	3.9074
10	3.8546	3.8457	0.014	3.8166	3.8748	3.7672	3.9242
11	3.8704	3.8516	0.011	3.8273	3.8759	3.7747	3.9284
12	3.8722	3.8537	0.011	3.8297	3.8776	3.7769	3.9304
13	3.9054	3.8614	0.014	3.8305	3.8922	3.7822	3.9405
14	3.8960	3.8874	0.013	3.8606	3.9142	3.8097	3.9651
15	3.8182	3.8941	0.016	3.8592	3.9289	3.8133	3.9749
16	3.8778	3.8472	0.013	3.8185	3.8760	3.7689	3.9256
17	3.8641	3.8649	0.016	3.8310	3.8988	3.7845	3.9453
18	3.8543	3.8452	0.014	3.8163	3.8742	3.7668	3.9237
19	3.8427	3.8492	0.028	3.7897	3.9086	3.7551	3.9433

Obs	Residual	Std Err Residual	Student Residual	-2-1-0 1 2	Cook's D	Rstudent
1	0.0329	0.028	1.157	| |**	0.148	1.1715
2	0.0297	0.030	0.992	| |*	0.076	0.9911
3	-0.00343	0.031	-0.112	| |	0.001	-0.1080
4	-0.0399	0.032	-1.261	|** |	0.067	-1.2890
5	-0.0228	0.032	-0.710	|* |	0.016	-0.6973
6	0.00302	0.031	0.099	| |	0.001	0.0956
7	-0.0618	0.030	-2.088	|**** |	0.366	-2.3952
8	0.00236	0.032	0.074	| |	0.000	0.0715
9	0.00445	0.029	0.156	| |	0.003	0.1506
10	0.00890	0.031	0.284	| |	0.004	0.2748
11	0.0188	0.032	0.583	| |*	0.011	0.5700
12	0.0186	0.032	0.575	| |*	0.010	0.5620
13	0.0440	0.031	1.421	| |**	0.110	1.4762
14	0.00862	0.032	0.271	| |	0.003	0.2624
15	-0.0759	0.030	-2.525	|***** |	0.472	-3.2161
16	0.0305	0.031	0.972	| |*	0.043	0.9698
17	-0.00074	0.030	-0.025	| |	0.000	-0.0237
18	0.00907	0.031	0.289	| |	0.004	0.2798
19	-0.00645	0.020	-0.326	| |	0.053	-0.3160

Obs	Hat Diag H	Cov Ratio	Dffits	INTERCEP Dfbetas	X1 Dfbetas	X2 Dfbetas	X3 Dfbetas
1	0.3070	1.3082	0.7797	-0.0068	0.3518	0.0665	-0.4465
2	0.2352	1.3138	0.5497	0.0273	0.2039	0.1169	-0.2469
3	0.1959	1.6334	-0.0533	0.0091	-0.0199	0.0058	0.0296
4	0.1433	0.9822	-0.5272	0.1909	-0.0889	0.1344	0.2118
5	0.1158	1.3002	-0.2524	0.0948	-0.0430	0.0826	0.1029
6	0.2039	1.6509	0.0484	-0.0411	-0.0219	-0.0356	0.0061
7	0.2516	0.4458	-1.3888	0.0166	0.9139	-0.6530	-1.0734
8	0.1307	1.5138	0.0277	-0.0028	-0.0169	0.0093	0.0186
9	0.3019	1.8754	0.0990	-0.0335	-0.0873	0.0070	0.0884
10	0.1596	1.5347	0.1198	-0.0551	-0.0959	-0.0243	0.0886
11	0.1110	1.3524	0.2014	-0.0807	-0.1254	-0.0633	0.1126
12	0.1080	1.3513	0.1956	-0.0788	-0.0905	-0.0902	0.0721
13	0.1792	0.9001	0.6897	0.5169	0.0787	0.5245	0.1559
14	0.1353	1.4943	0.1038	0.0707	0.0596	0.0313	-0.0366
15	0.2284	0.1868	-1.7498	-1.4346	-1.1919	-0.7987	0.7277
16	0.1552	1.2027	0.4157	0.3104	0.1147	0.2246	0.0158
17	0.2158	1.6802	-0.0124	-0.0101	-0.0071	-0.0057	0.0035
18	0.1573	1.5292	0.1209	0.0564	0.0499	0.0005	-0.0308
19	0.6649	3.8223	-0.4451	0.1947	-0.0442	0.3807	0.1410

Sum of Residuals	0
Sum of Squared Residuals	0.0175
Predicted Resid SS (Press)	0.0289

SAS PROGRAM

```
Data GASOLINE;
Input COUNTRY $ YEAR Y X1 X2 X3;
CARDS;

DATA AUSTRIA; SET GASOLINE;
IF COUNTRY='AUSTRIA';

Proc reg data=AUSTRIA;
    Model Y=X1 X2 X3 / influence p r cli clm;
RUN;

DATA BELGIUM; SET GASOLINE;
IF COUNTRY='BELGIUM';%
```

```
Proc reg data=BELGIUM;
    Model Y=X1 X2 X3 / influence p r cli clm;
RUN;
```

8.13 *Independence of Recursive Residual*. This is based on Johnston (1984, p. 386).

a. Using the updating formula given in (8.11) with $A = \left(X_t'X_t\right)$ and $a = -b = x_{t+1}'$, we get

$$\left(X_t'X_t + x_{t+1}x_{t+1}'\right)^{-1} = \left(X_t'X_t\right)^{-1}$$
$$- \left(X_t'X_t\right)^{-1} x_{t+1} \left(1 + x_{t+1}' \left(X_t'X_t\right)^{-1} x_{t+1}\right)^{-1} x_{t+1}' \left(X_t'X_t\right)^{-1}$$

But $X_{t+1} = \begin{pmatrix} X_t \\ x_{t+1}' \end{pmatrix}$, hence $X_{t+1}'X_{t+1} = X_t'X_t + x_{t+1}x_{t+1}'$. Substituting this expression we get

$$\left(X_{t-1}'X_{t+1}\right)^{-1} = \left(X_t'X_t\right)^{-1} - \frac{\left(X_t'X_t\right)^{-1} x_{t+1}x_{t+1}' \left(X_t'X_t\right)^{-1}}{1 + x_{t+1}' \left(X_t'X_t\right)^{-1} x_{t+1}}$$

which is exactly (8.31).

b. $\hat{\beta}_{t+1} = \left(X_{t+1}'X_{t+1}\right)^{-1} X_{t+1}' Y_{t+1}$

Replacing X_{t+1}' by $\left(X_t', x_{t+1}\right)$ and Y_{t+1} by $\begin{pmatrix} Y_t \\ y_{t+1} \end{pmatrix}$ yields

$$\hat{\beta}_{t+1} = \left(X_{t+1}'X_{t+1}\right)^{-1} \left(X_t'Y_t + x_{t+1}y_{t+1}\right)$$

Replacing $\left(X_{t+1}'X_{t+1}\right)^{-1}$ by its expression in (8.31) we get

$$\hat{\beta}_{t+1} = \hat{\beta}_t + \left(X_t'X_t\right)^{-1} x_{t+1}y_{t+1} - \frac{\left(X_t'X_t\right)^{-1} x_{t+1}x_{t+1}'\hat{\beta}_t}{f_{t+1}}$$
$$- \frac{\left(X_t'X_t\right)^{-1} x_{t+1}x_{t+1}' \left(X_t'X_t\right)^{-1} x_{t+1}y_{t+1}}{f_{t+1}}$$
$$= \hat{\beta}_t - \frac{\left(X_t'X_t\right)^{-1} x_{t+1}x_{t+1}'\hat{\beta}_t}{f_{t+1}} + \left(\frac{f_{t+1} - f_{t+1} + 1}{f_{t+1}}\right) \left(X_t'X_t\right)^{-1} x_{t+1}y_{t+1}$$
$$= \hat{\beta}_t - \left(X_t'X_t\right)^{-1} x_{t+1} \left(y_{t+1} - x_{t+1}'\hat{\beta}_t\right) / f_{t+1}$$

where we used the fact that $x_{t+1}' \left(X_t'X_t\right)^{-1} x_{t+1} = f_{t+1} - 1$.

c. Using (8.30), $w_{t+1} = \left(y_{t+1} - x'_{t+1}\hat{\beta}_t\right)/\sqrt{f_{t+1}}$ where $f_{t+1} = 1 + x'_{t+1}\left(X'_tX_t\right)^{-1}x_{t+1}$. Defining $v_{t+1} = \sqrt{f_{t+1}}\,w_{t+1}$ we get $v_{t+1} = y_{t+1} - x'_{t+1}\hat{\beta}_t = x'_{t+1}(\beta - \hat{\beta}_t) + u_{t+1}$ for $t = k,..,T-1$. Since $u_t \sim \text{IIN}(0,\sigma^2)$, then w_{t+1} has zero mean and $\text{var}(w_{t+1}) = \sigma^2$. Furthermore, w_{t+1} are linear in the y's. Therefore, they are themselves normally distributed. Given normality of the w_{t+1}'s it is sufficient to show that $\text{cov}(w_{t+1}, w_{s+1}) = 0$ for $t \neq s$; $t = s = k,..,T-1$. But f_{t+1} is fixed. Therefore, it suffices to show that $\text{cov}(v_{t+1}, v_{s+1}) = 0$ for $t \neq s$.

Using the fact that $\hat{\beta}_t = \left(X'_tX_t\right)^{-1}X'_tY_t = \beta + \left(X'_tX_t\right)^{-1}X_t{}'u_t$ where $u'_t = (u_1,..,u_t)$, we get $v_{t+1} = u_{t+1} - x'_{t+1}\left(X'_tX_t\right)^{-1}X'_tu_t$. Therefore,

$$\begin{aligned}
E\left(v_{t+1}v_{s+1}\right) &= E\left\{\left[u_{t+1} - x'_{t+1}(X'_tX_t)^{-1}X'_tu_t\right]\left[u_{s+1} - x'_{s+1}\left(X'_sX_s\right)^{-1}X'_su_s\right]\right\} \\
&= E\left(u_{t+1}u_{s+1}\right) - x'_{t+1}\left(X'_tX_t\right)^{-1}X'_tE\left(u_tu_{s+1}\right) \\
&\quad - E\left(u_{t+1}u'_s\right)X_s\left(X'_sX_s\right)^{-1}x_{s+1} \\
&\quad + x'_{t+1}\left(X'_tX'_t\right)^{-1}X'_tE\left(u_tu'_s\right)X_s\left(X'_sX_s\right)^{-1}x_{s+1}.
\end{aligned}$$

Assuming, without loss of generality, that $t < s$, we get $E(u_{t+1}u_{s+1}) = 0$, $E(u_{t+1}u_{s+1}) = 0$ since $t < s < s+1$, $E\left(u_{t+1}u'_s\right) = (0,..,\sigma^2,..0)$ where the σ^2 is in the $(t+1)$-th position, and

$$E\left(u_tu'_s\right) = E\begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_t \end{pmatrix}(u_1,..,u_t,u_{t+1},..,u_s) = \sigma^2\left(I_t, 0\right)$$

Substituting these covariances in $E(v_{t+1}v_{s+1})$ yields immediately zero for the first two terms and what remains is

$$E\left(v_{t+1}v_{s+1}\right) = -\sigma^2x_{t+1}\left(X'_sX_s\right)^{-1}x_{s+1} + \sigma^2x'_{t+1}\left(X'_tX_t\right)^{-1}X'_t\,[I_t,0]\,X_s\left(X'_sX_s\right)^{-1}x_{s+1}$$

But $[I_t,0]X_s = X_t$ for $t < s$. Hence

$$E(v_{t+1}v_{s+1}) = -\sigma^2x_{t+1}(X'_sX_s)^{-1}x_{s+1} + \sigma^2x_{t+1}(X'_sX_s)^{-1}x_{s+1} = 0 \qquad \text{for } t \neq s.$$

A simpler proof using the C matrix defined in (8.34) is given in problem 8.14.

8.14 *Recursive Residuals are Linear Unbiased With Scalar Covariance Matrix (LUS).*

a. The first element of $w = Cy$ where C is defined in (8.34) is

$$w_{k+1} = -x'_{k+1}(X'_k X_k)^{-1} X'_k Y_k/\sqrt{f_{k+1}} + y_{k+1}/\sqrt{f_{k+1}}$$

where $Y'_k = (y_1, .., y_k)$. Hence, using the fact that $\hat{\beta}_k = \left(X'_k X_k\right)^{-1} X'_k Y_k$, we get $w_{k+1} = \left(y_{k+1} - x'_{k+1}\hat{\beta}_k\right)/\sqrt{f_{k+1}}$ which is (8.30) for $t = k$. Similarly, the t-th element of $w = Cy$ from (8.34) is $w_t = -x'_t\left(X'_{t-1}X_{t-1}\right)^{-1}$ $X'_{t-1}Y_{t-1}/\sqrt{f_t} + y_t/\sqrt{f_t}$ where $Y'_{t-1} = (y_1, .., y_{t-1})$. Hence, using the fact that $\hat{\beta}_{t-1} = \left(X'_{t-1}X_{t-1}\right)^{-1} X'_{t-1}Y_{t-1}$, we get $w_t = \left(y_t - x'_t\hat{\beta}_{t-1}\right)/\sqrt{f_t}$ which is (8.30) for $t = t - 1$. Also, the last term can be expressed in the same way.

$$w_T = -x'_T(X'_{T-1}X_{T-1})^{-1}X'_{T-1}Y_{T-1}/\sqrt{f_T} + y_T/\sqrt{f_T} = (y_T - x'_T\hat{\beta}_{T-1})/\sqrt{f_T}$$

which is (8.30) for $t = T - 1$.

b. The first row of CX is obtained by multiplying the first row of C by

$$X = \begin{bmatrix} X_k \\ x'_{k+1} \\ \vdots \\ x'_T \end{bmatrix}.$$

This yields $-x'_{k+1}\left(X'_k X_k\right)^{-1} X'_k X_k/\sqrt{f_{k+1}} + x'_{k+1}/\sqrt{f_{k+1}} = 0$.

Similarly, the t-th row of CX is obtained by multiplying the t-th row of C by

$$X = \begin{bmatrix} X_{t-1} \\ x'_t \\ \vdots \\ x'_T \end{bmatrix}.$$

This yields $-x'_t\left(X'_{t-1}X_{t-1}\right)^{-1} X'_{t-1}X_{t-1}/\sqrt{f_t} + x'_t/\sqrt{f_t} = 0$.

Also, the last row of CX is obtained by multiplying the T-th row of C by $X = \begin{bmatrix} X_{T-1} \\ x'_T \end{bmatrix}$.

This yields $-x'_T\left(X'_{T-1}X_{T-1}\right)^{-1} X'_{T-1}X_{T-1}/\sqrt{f_T} + x'_T/\sqrt{f_T} = 0$.

This proves that $CX = 0$. This means that $w = Cy = CX\beta + Cu = Cu$ since $CX = 0$.

Hence, $E(w) = CE(u) = 0$. Note that

$$C' = \begin{bmatrix} \frac{-X_k(X_k'X_k)^{-1}x_{k+1}}{\sqrt{f_{k+1}}} & .. & \frac{-X_{t-1}(X_{t-1}'X_{t-1})^{-1}x_t}{\sqrt{f_t}} & .. & \frac{-X_{T-1}(X_{T-1}'X_{T-1})^{-1}x_T}{\sqrt{f_T}} \\ \frac{1}{\sqrt{f_{k+1}}} & & \frac{1}{\sqrt{f_t}} & .. & . \\ 0 & & 0 & .. & . \\ \vdots & & \vdots & & \vdots \\ 0 & & 0 & .. & \frac{1}{\sqrt{f_T}} \end{bmatrix}$$

so that the first diagonal element of CC' is

$$\frac{x_{k+1}'(X_k'X_k)^{-1}X_k'X_k(X_k'X_k)^{-1}x_{k+1}}{f_{k+1}} + \frac{1}{f_{k+1}} = \frac{1 + x_{k+1}'(X_k'X_k)^{-1}x_{k+1}}{f_{k+1}}$$

$$= \frac{f_{k+1}}{f_{k+1}} = 1$$

similarly, the t-th diagonal element of CC' is

$$\frac{x_t'(X_{t-1}'X_{t-1})^{-1}X_{t-1}'X_{t-1}(X_{t-1}'X_{t-1})^{-1}x_t}{f_t} + \frac{1}{f_t} = \frac{f_t}{f_t} = 1$$

and the T-th diagonal element of CC' is

$$\frac{x_T'(X_{T-1}'X_{T-1})^{-1}X_{T-1}'X_{T-1}(X_{T-1}'X_{T-1})^{-1}x_T}{f_T} + \frac{1}{f_T} = \frac{f_T}{f_T} = 1$$

Using the fact that

$$X_{t-1} = \begin{bmatrix} X_k \\ x_{k+1}' \\ \vdots \\ x_{t-1}' \end{bmatrix},$$

one gets the (1,t) element of CC' by multiplying the first row of C by the t-th column of C'. This yields

$$\frac{x_{k+1}'(X_k'X_k)^{-1}X_k'X_k(X_{t-1}'X_{t-1})^{-1}x_t}{\sqrt{f_{k+1}}\sqrt{f_t}} - \frac{x_{k+1}'(X_{t-1}'X_{t-1})^{-1}x_t}{\sqrt{f_{k+1}}\sqrt{f_t}} = 0.$$

Similarly, the (1,T) element of CC' is obtained by multiplying the first row of C by the T-th column of C'. This yields

$$\frac{x_{k+1}'(X_k'X_k)^{-1}X_k'X_k(X_{T-1}'X_{T-1})^{-1}x_T}{\sqrt{f_{k+1}}\sqrt{f_T}} - \frac{x_{k+1}'(X_{T-1}'X_{T-1})^{-1}x_T}{\sqrt{f_{k+1}}\sqrt{f_T}} = 0.$$

Similarly, one can show that the (t,1)-th element, the (t,T)-th element, the (T,1)-th element and the (T,t)-th element of CC' are all zero. This proves that $CC' = I_{T-k}$. Hence, w is linear in y, unbiased with mean zero, and $\text{var}(w) = \text{var}(Cu) = C\,E(uu')C' = \sigma^2 CC' = \sigma^2 I_{T-k}$,

i.e., the variance–covariance matrix is a scalar times an identity matrix. Using Theil's (1971) terminology, these recursive residuals are (LUS) linear unbiased with a scalar covariance matrix. A further result holds for all LUS residuals. In fact, $C'C = \overline{P}_x$ see Theil (1971, p. 208). Hence,

$$\sum_{t=k+1}^{T} w_t^2 = w'w = y'C'Cy = y'\overline{P}_x y = e'e = \sum_{t=1}^{T} e_t{}^2$$

Therefore, the sum of squares of the $(T-k)$ recursive residuals is equal to the sum of squares of the T least squares residuals.

c. If $u \sim N(0, \sigma^2 I_T)$ then $w = Cu$ is also Normal since it is a linear function of normal random variables. From part (b) we proved that w has mean zero and variance $\sigma^2 I_{T-k}$. Hence, $w \sim N(0, \sigma^2 I_{T-k})$ as required.

d. Let us express the $(t+1)$ residuals as follows:

$$Y_{t+1} - X_{t+1}\hat{\beta}_{t+1} = Y_{t+1} - X_{t+1}\hat{\beta}_t - X_{t+1}(\hat{\beta}_{t+1} - \hat{\beta}_t)$$

so that

$$\begin{aligned}\text{RSS}_{t+1} &= (Y_{t+1} - X_{t+1}\hat{\beta}_{t+1})'(Y_{t+1} - X_{t+1}\hat{\beta}_{t+1}) \\ &= (Y_{t+1} - X_{t+1}\hat{\beta}_t)'(Y_{t+1} - X_{t+1}\hat{\beta}_t) + (\hat{\beta}_{t+1} - \hat{\beta}_t)'X'_{t+1}X_{t+1} \\ &\quad (\hat{\beta}_{t+1} - \hat{\beta}_t) - 2(\hat{\beta}_{t+1} - \hat{\beta}_t)'X'_{t+1}(Y_{t+1} - X_{t+1}\hat{\beta}_t)\end{aligned}$$

Partition $Y_{t+1} = \begin{pmatrix} Y_t \\ y_{t+1} \end{pmatrix}$ and $X_{t+1} = \begin{pmatrix} X_t \\ x'_{t+1} \end{pmatrix}$ so that the first term of RSS_{t+1} becomes $(Y_{t+1} - X_{t+1}\hat{\beta}_t)'(Y_{t+1} - X_{t+1}\hat{\beta}_t)$

$$= [(Y_t - X_t\hat{\beta}_t)', (y_{t+1} - x'_{t+1}\hat{\beta}_t)'] \begin{pmatrix} Y_t - X_t\hat{\beta}_t \\ y_{t+1} - x'_{t+1}\hat{\beta}_t \end{pmatrix}$$

$$= \text{RSS}_t + f_{t+1}\, w_{t+1}^2$$

where $f_{t+1} = 1 + x'_{t+1}\left(X'_tX_t\right)^{-1}x_{t+1}$ is defined in (8.32) and w_{t+1} is the recursive residual defined in (8.30). Next, we make the substitution for $(\hat{\beta}_{t+1} - \hat{\beta}_t)$ from (8.32) into the second term of RSS_{t+1} to get $(\hat{\beta}_{t+1} - \hat{\beta}_t)'X'_{t+1}X_{t+1}(\hat{\beta}_{t+1} - \hat{\beta}_t) = (y_{t+1} - x'_{t+1}\hat{\beta}_t)'x'_{t+1}(X'_tX_t)^{-1}(X'_{t+1}X_{t+1})(X'_tX_t)^{-1}x_{t+1}(y_{t+1} - x'_{t+1}\hat{\beta}_t)$

Postmultiplying (8.31) by x_{t+1} one gets

$$(X'_{t+1}X_{t+1})^{-1}x_{t+1} = \left(X'_tX_t\right)^{-1}x_{t+1} - \frac{\left(X'_tX_t\right)^{-1}x_{t+1}c}{1+c}$$

where $c = x'_{t+1}\left(X'_tx_t\right)^{-1}x_{t+1}$. Collecting terms one gets

$$\left(X'_{t+1}X_{t+1}\right)^{-1}x_{t+1} = \left(X'_tX_t\right)^{-1}x_{t+1}/f_{t+1}$$

where $f_{t+1} = 1 + c$.

Substituting this above, the second term of RSS_{t+1} reduces to $x'_{t+1}\left(X'_tX_t\right)^{-1}x_{t+1}w^2_{t+1}$.

For the third term of RSS_{t+1}, we observe that $X'_{t+1}\left(Y_{t+1} - X'_{t+1}\hat{\beta}_t\right) = \left[X'_t, x_{t+1}\right]\begin{bmatrix} y_t - X_t\hat{\beta}_t \\ y_{t+1} - x'_{t+1}\hat{\beta}_t \end{bmatrix} = X'_t(Y_t - X_t\hat{\beta}_t) + x_{t+1}(y_{t+1} - x'_{t+1}\hat{\beta}_t)$.

The first term is zero since X_t and its least squares residuals are orthogonal. Hence, $X'_{t+1}(Y_{t+1} - X_{t+1}\hat{\beta}_t) = x_{t+1}w_{t+1}\sqrt{f_{t+1}}$ and the third term of RSS_{t+1} becomes $-2(\hat{\beta}_{t+1} - \hat{\beta}_t)'x_{t+1}w_{t+1}\sqrt{f_{t+1}}$ using (8.32), this reduces to $-2x'_{t+1}\left(X'_tX_t\right)^{-1}x_{t+1}w^2_{t+1}$. Adding all three terms yields

$$\begin{aligned} RSS_{t+1} &= RSS_t + f_{t+1}w^2_{t+1} + x'_{t+1}\left(X'_tX_t\right)^{-1}x_{t+1}w^2_{t+1} \\ &\quad - 2x'_{t+1}\left(X'_tX_t\right)^{-1}x_{t+1}w^2_{t+1} \\ &= RSS_t + f_{t+1}w^2_{t+1} - (f_{t+1} - 1)w^2_{t+1} = RSS_t + w^2_{t+1} \text{ as required.} \end{aligned}$$

8.15 *The Harvey and Collier (1977) Misspecification t-test as a Variable Additions Test*. This is based on Wu (1993).

a. The Chow F-test for $H_o;\ \gamma = 0$ in (8.44) is given by

$$F = \frac{RRSS - URSS}{URSS/(T-k-1)} = \frac{y'\overline{P}_Xy - y'\overline{P}_{[X,z]}y}{y'\overline{P}_{[X,z]}y/(T-k-1)}.$$

Using the fact that $z = C'\iota_{T-k}$ where C is defined in (8.34) and ι_{T-k} is a vector of ones of dimension $(T-k)$. From (8.35) we know that $CX = 0$, $CC' = I_{T-k}$ and $C'C = \overline{P}_x$. Hence, $z'X = \iota'_{T-k}CX = 0$ and

$$P_{[X,z]} = [X, z]\begin{bmatrix} X'X & 0 \\ 0 & z'z \end{bmatrix}^{-1}\begin{pmatrix} X' \\ z' \end{pmatrix} = P_X + P_z$$

Therefore, $URSS = y'\overline{P}_{[X,z]}y = y'y - y'P_xy - y'P_zy$ and $RRSS = y'\overline{P}_Xy = y'y - y'P_xy$ and the F-statistic reduces to

$$F = \frac{y'P_zY}{y'(\overline{P}_x - P_z)y/(T-k-1)}$$

which is distributed as $F(1, T-k-1)$ under the null hypothesis $H_o;\ \gamma = 0$.

b. The numerator of the F-statistic is

$$y'P_zy = y'z(z'z)^{-1}z'y = y'C'\iota_{T-k}\left(\iota'_{T-K}CC'\iota_{T-k}\right)^{-1}\iota'_{T-k}Cy$$

But $CC' = I_{T-k}$ and $\iota'_{T-k}CC'\iota_{T-k} = \iota'_{T-k}\iota_{T-k} = (T-k)$. Therefore $y'P_zy = y'C'\iota_{T-k}\iota'_{T-k}Cy/(T-k)$. The recursive residuals are constructed as $w = Cy$, see below (8.34). Hence

$$y'P_zy = w'\iota_{T-k}\iota'_{T-k}w/(T-k) = \left(\sum_{t=k+1}^{T} w_t\right)^2/(T-k) = (T-k)\overline{w}^2$$

where $\overline{w} = \sum\limits_{t=k+1}^{T} w_t/(T-k)$. Using $\overline{P}_X = C'C$, we can write

$$y'\overline{P}_xy = y'C'Cy = w'w = \sum_{t=k+1}^{T} w_t^2$$

Hence, the denomentator of the F-statistic is $y'(\overline{P}_x - P_z)y/(T-k-1) = \left[\sum\limits_{t=k+1}^{T} w_t^2 - (T-k)\overline{w}^2\right]/(T-k-1) = \sum\limits_{t=k+1}^{T}(w_t - \overline{w})^2/(T-k-1) = s_w^2$ where $s_w^2 = \sum\limits_{t=k+1}^{T}(w_t - \overline{w})^2/(T-k-1)$ was given below (8.43). Therefore, the F-statistic in part (a) is $F = (T-k)\overline{w}^2/s_w^2$ which is the square of the Harvey and Collier (1977) t_{T-k-1}-statistic given in (8.43).

8.16 The Gasoline data model given in Chap. 10.

a. Using Eviews, and the data for AUSTRIA one can generate the regression of y on X_1, X_2 and X_3 and the corresponding recursive residuals.

LS // Dependent Variable is Y

Sample: 1960 1978

Included observations: 19

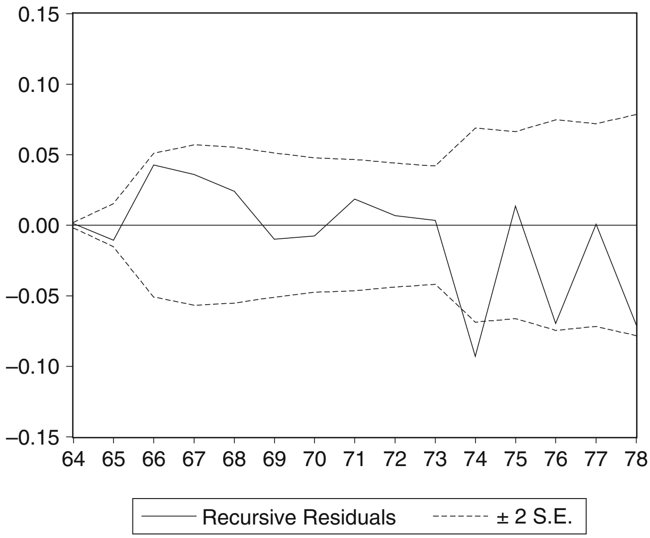
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.726605	0.373018	9.990422	0.0000
X1	0.760721	0.211471	3.597289	0.0026
X2	-0.793199	0.150086	-5.284956	0.0001
X3	-0.519871	0.113130	-4.595336	0.0004

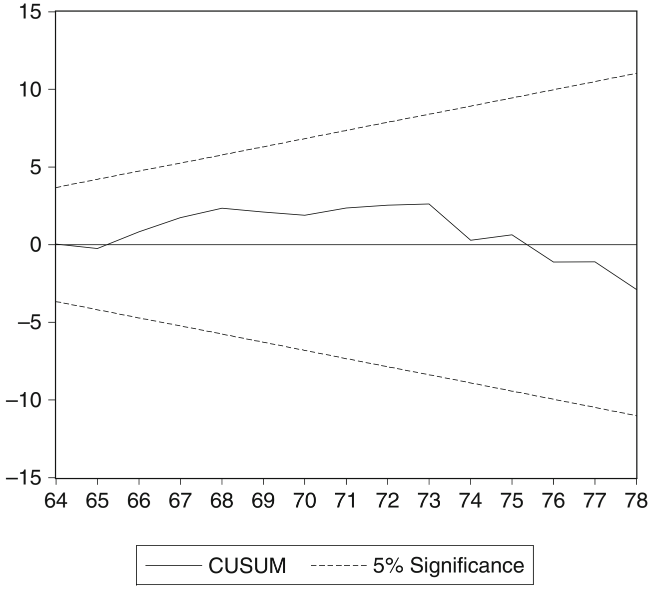
R-squared	0.733440	Mean dependent var	4.056487
Adjusted R-squared	0.680128	S.D. dependent var	0.069299
S.E. of regression	0.039194	Akaike info criterion	-6.293809
Sum squared resid	0.023042	Schwarz criterion	-6.094980
Log likelihood	36.83136	F-statistic	13.75753
Durbin-Watson stat	1.879634	Prob(F-statistic)	0.000140

RECURSIVE RESIDUALS:

0.001047
-0.010798
0.042785
0.035901
0.024071
-0.009944
-0.007583
0.018521
0.006881
0.003415
-0.092971
0.013540
-0.069668
0.000706
-0.070614

b. EViews also gives as an option a plot of the recursive residuals plus or minus twice their standard error. Next, the CUSUM plot is given along with 5% upper and lower lines.





d. Chow Forecast Test: Forecast from 1978 to 1978

F-statistic	3.866193	Probability	0.069418
Log likelihood ratio	4.633206	Probability	0.031359

Test Equation:

LS // Dependent Variable is Y

Sample: 1960 1977

Included observations: 18

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.000172	0.369022	10.83991	0.0000
X1	0.803752	0.194999	4.121830	0.0010
X2	-0.738174	0.140340	-5.259896	0.0001
X3	-0.522216	0.103666	-5.037483	0.0002

R-squared	0.732757	Mean dependent var	4.063917
Adjusted R-squared	0.675490	S.D. dependent var	0.063042
S.E. of regression	0.035913	Akaike info criterion	-6.460203
Sum squared resid	0.018056	Schwarz criterion	-6.262343
Log likelihood	36.60094	F-statistic	12.79557
Durbin-Watson stat	2.028488	Prob(F-statistic)	0.000268

8.17 *The Differencing Test in a Regression with Equicorrelated Disturbances*. This is based on Baltagi (1990).

a. For the equicorrelated case Ω can be written as $\Omega = \sigma^2(1-\rho)[E_T + \theta\bar{J}_T]$ where $E_T = I_T - \bar{J}_T$, $\bar{J}_T = J_T/T$ and $\theta = [1 + (T-1)\rho]/(1-\rho)$.

$$\Omega^{-1} = [E_T + (1/\theta)\bar{J}_T]/\sigma^2(1-\rho)$$

$$\iota'\Omega^{-1} = \iota'/\theta\sigma^2(1-\rho)$$

and $L = E_T/\sigma^2(1-\rho)$. Hence $\hat{\beta} = (X'E_TX)^{-1}X'E_TY$, which is the OLS estimator of β, since E_T is a matrix that transforms each variable into deviations from its mean. That GLS is equivalent to OLS for the equicorrelated case, is a standard result in the literature.

Also, $D\Omega = \sigma^2(1-\rho)D$ since $D\iota = 0$ and $D\Omega D' = \sigma^2(1-\rho)DD'$. Therefore, $M = P_{D'}/\sigma^2(1-\rho)$ where $P_{D'} = D'(DD')^{-1}D$. In order to show that $M = L$, it remains to show that $P_{D'} = E_T$ or equivalently that $P_{D'} + \bar{J}_T = I_T$ from the definition of E_T. Note that both $P_{D'}$ and $\bar{J}_T$ are symmetric idempotent matrices which are orthogonal to each other. ($D\bar{J}_T = 0$ since $D\iota = 0$). Using Graybill's (1961) Theorem 1.69, these two properties of $P_{D'}$ and $\bar{J}_T$ imply a third: Their sum is idempotent with rank equal to the sum of the ranks. But rank of $P_{D'}$ is (T-1) and rank of $\bar{J}_T$ is 1, hence their sum is idempotent of rank T, which could only be I_T. This proves the Maeshiro and Wichers (1989) result, i.e., $\hat{\beta} = \tilde{\beta}$, which happens to be the OLS estimator from (1) in this particular case.

b. The Plosser, Schwert and White differencing test is based on the difference between the OLS estimator from (1) and the OLS estimator from (2). But OLS on (1) is equivalent to GLS on (1). Also, part (a) proved that GLS on (1) is in fact GLS from (2). Hence, the differencing test can be based upon the difference between the OLS and GLS estimators from the differenced equation. An alternative solution is given by Koning (1992).

8.18 a. Stata performs Ramsey's RESET test by issuing the command estat ovtest after running the regression as follows:

```
. reg lwage wks south smsa ms exp exp2 occ ind union fem blk ed
```

(We suppress the regression output since it is the same as Table 4.1 in the text and solution 4.13).

```
. estat ovtest

Ramsey RESET test using powers of the fitted values of lwage
       Ho: model has no omitted variables
                 F(3, 579)  =  0.79
                   Prob>F  =  0.5006
```

This does not reject the null of no regression misspecification.

b. Stata also performs White's (1982) Information matrix test by issuing the command estat imtest after running the regression as follows:

```
. estat imtest
```

Cameron & Trivedi's decomposition of IM-test

Source	chi2	df	p
Heteroskedasticity	103.28	81	0.082
Skewness	5.37	12	0.9444
Kurtosis	1.76	1	0.1844
Total	110.41	94	0.1187

This does not reject the null, even though heteroskedasticity seems to be a problem.

8.20 a. Stata performs Ramsey's RESET test by issuing the command estat ovtest after running the regression as follows:

```
. reg lnc lnrp lnrdi
```

(We suppress the regression output since it is the same as that in solution 5.13).

```
. estat ovtest
```

Ramsey RESET test using powers of the fitted values of lnc

Ho: model has no omitted variables

F(3, 40) = 3.11

Prob>F = 0.0369

This rejects the null of no regression misspecification at the 5% level.

b. Stata also performs White's (1982) Information matrix test by issuing the command estat imtest after running the regression as follows:

```
. estat imtest
```

Cameron & Trivedi's decomposition of IM-test

Source	chi2	df	p
Heteroskedasticity	15.66	5	0.0079
Skewness	3.59	2	0.1664
Kurtosis	0.06	1	0.8028
Total	19.31	8	0.0133

This rejects the null, and seems to indicate that heteroskedasticity seems to be the problem. This was confirmed using this data set in problem 5.13.

References

Baltagi, B.H. (1990), "The Differencing Test in a Regression with Equicorrelated Disturbances," Econometric Theory, Problem 90.4.5, 6: 488.

Graybill, F.A. (1961), An Introduction to Linear Statistical Models, Vol. 1 (McGraw-Hill: New York).

Harvey, A.C. and P. Collier (1977), "Testing for Functional Misspecification in Regression Analysis," Journal of Econometrics, 6: 103–119.

Johnston, J. (1984), Econometric Methods (McGraw-Hill: New York).

Koning, R.H. (1992), "The Differencing Test in a Regression with Equicorrelated Disturbances," Econometric Theory, Solution 90.4.5, 8: 155–156.

Maeshiro, A. and R. Wichers (1989), "On the Relationship Between the Estimates of Level Models and Difference Models," American Journal of Agricultural Economics, 71: 432–434.

Theil, H. (1971), Principles of Econometrics (Wiley: New York).

White, H. (1982), "Maximum Likelihood Estimation of Misspecified Models," Econometrica, 50: 1–25.

Wu, P. (1993), "Variable Addition Test," Econometric Theory, Problem 93.1.2, 9: 145–146.

CHAPTER 9
Generalized Least Squares

9.1 *GLS Is More Efficient than OLS.*

a. Equation (7.5) of Chap. 7 gives $\hat{\beta}_{ols} = \beta + (X'X)^{-1}X'u$ so that $E(\hat{\beta}_{ols}) = \beta$ as long as X and u are uncorrelated and u has zero mean. Also,

$$\mathrm{var}(\hat{\beta}_{ols}) = E(\hat{\beta}_{ols} - \beta)(\hat{\beta}_{ols} - \beta)' = E[(X'X)^{-1}X'uu'X(X'X)^{-1}]$$
$$= (X'X)^{-1}X'\,E(uu')X(X'X)^{-1} = \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1}.$$

b. $\mathrm{var}(\hat{\beta}_{ols}) - \mathrm{var}(\hat{\beta}_{GLS}) = \sigma^2[(X'X)^{-1}X'\Omega X(X'X)^{-1} - (X'\Omega^{-1}X)^{-1}]$

$$= \sigma^2[(X'X)^{-1}X'\Omega X(X'X)^{-1} - (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\Omega\Omega^{-1}$$
$$X(X'\Omega^{-1}X)^{-1}]$$
$$= \sigma^2[(X'X)^{-1}X' - (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}]\Omega[X(X'X)^{-1}$$
$$- \Omega^{-1}X(X'\Omega^{-1}X)^{-1}]$$
$$= \sigma^2\,A\Omega A'$$

where $A = [(X'X)^{-1}X' - (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}]$. The second equality post multiplies $(X'\Omega^{-1}X)^{-1}$ by $(X'\Omega^{-1}X)(X'\Omega^{-1}X)^{-1}$ which is an identity of dimension K. The third equality follows since the cross-product terms give $-2(X'\Omega^{-1}X)^{-1}$. The difference in variances is positive semi-definite since Ω is positive definite.

9.2 a. From Chap. 7, we know that $s^2 = e'e/(n-K) = u'\bar{P}_X u/(n-K)$ or $(n-K)s^2 = u'\bar{P}_X u$. Hence,

$$(n-K)E(s^2) = E(u'\bar{P}_X u) = E[\mathrm{tr}(u'\bar{P}_X u)]$$
$$= \mathrm{tr}[E(uu')\bar{P}_X] = \mathrm{tr}(\Sigma\bar{P}_X) = \sigma^2\mathrm{tr}(\Omega\bar{P}_X)$$

and $E(s^2) = \sigma^2\mathrm{tr}(\Omega\bar{P}_X)/(n-K)$ which in general is not equal to σ^2.

B.H. Baltagi, *Solutions Manual for Econometrics*, Springer Texts in Business and Economics, DOI 10.1007/978-3-642-54548-1_9,

b. From part (a),

$$(n-K)E(s^2) = tr(\Sigma\bar{P}_X) = tr(\Sigma) - tr(\Sigma P_X)$$

but, both Σ and P_X are non-negative definite. Hence, $tr(\Sigma P_X) \geq 0$ and

$$(n-K)E(s^2) \leq tr(\Sigma)$$

which upon rearranging yields $E(s^2) \leq tr(\Sigma)/(n-K)$. Also, Σ and $\bar{P}_X$ are non-negative definite. Hence, $tr(\Sigma\bar{P}_X) \geq 0$. Therefore, $E(s^2) \geq 0$. This proves the bound derived by Dufour (1986):

$$0 \leq E(s^2) \leq tr(\Sigma)/(n-K)$$

where $tr(\Sigma) = \sum_{i=1}^{n} \sigma_i^2$. Under homoskedasticity $\sigma_i^2 = \sigma^2$ for $i = 1, 2, .., n$. Hence, $tr(\Sigma) = n\sigma^2$ and the upper bound becomes $n\sigma^2/(n-K)$. A useful bound for $E(s^2)$ has been derived by Sathe and Vinod (1974) and Neudecker (1977, 1978). This is given by $0 \leq$ mean of $(n-K)$ smallest characteristic roots of $\Sigma \leq E(s^2) \leq$ mean of $(n-K)$ largest characteristic roots of $\Sigma \leq tr(\Sigma)/(n-K)$.

c. Using $s^2 = u'\bar{P}_X u/(n-K) = u'u/(n-K) - u'P_X u/(n-K)$ we have $\text{plim}\, s^2 = \text{plim}\, u'u/(n-K) - \text{plim}\, u'P_X u/(n-K)$. By assumption plim $u'u/n = \sigma^2$. Hence, the first term tend in plim to σ^2 as $n \to \infty$. The second term has expectation $\sigma^2 tr(P_X\Omega)/(n-K)$. But, $P_X\Omega$ has rank K and therefore exactly K non-zero characteristic roots each of which cannot exceed λ_{max}. This means that

$$E[u'P_X u/(n-K)] \leq \sigma^2 K\lambda_{max}/(n-K).$$

Using the condition that $\lambda_{max}/n \to 0$ as $n \to \infty$ proves that

$$\lim E[u'P_X u/(n-K)] \to 0$$

as $n \to \infty$. Hence, $\text{plim}\, [u'P_X u/(n-K)] \to 0$ as $n \to \infty$ and $\text{plim}\, s^2 = \sigma^2$. Therefore, a sufficient condition for s^2 to be consistent for σ^2 irrespective of X is that $\lambda_{max}/n \to 0$ and $\text{plim}\, (u'u/n) = \sigma^2$ as $n \to \infty$, see Krämer and Berghoff (1991).

d. From (9.6), $s^{*2} = e^{*\prime}e^{*}/(n-K)$ where $e^{*} = y^{*} - X^{*}\hat{\beta}_{GLS} = y^{*} - X^{*}(X^{*\prime}X^{*})^{-1}X^{*\prime}y^{*} = \bar{P}_{X^{*}}y^{*}$ using (9.4), where $\bar{P}_{X^{*}} = I_n - P_{X^{*}}$ and $P_{X^{*}} = X^{*}(X^{*\prime}X^{*})^{-1}X^{*\prime}$. Substituting y^{*} from (9.3), we get $e^{*} = \bar{P}_{X^{*}}u^{*}$ where $\bar{P}_{X^{*}}X^{*} = 0$. Hence, $(n-K)s^{*2} = e^{*\prime}e^{*} = u^{*\prime}\bar{P}_{X^{*}}u^{*}$ with

$$\begin{aligned}(n-K)E(s^{*2}) &= E\left(u^{*\prime}\bar{P}_{X^{*}}u^{*}\right) = E\left[\mathrm{tr}\left(u^{*}u^{*\prime}\bar{P}_{X^{*}}\right)\right]\\ &= \mathrm{tr}\left[E\left(u^{*}u^{*\prime}\right)\bar{P}_{X^{*}}\right] = \mathrm{tr}\left(\sigma^{2}\bar{P}_{X^{*}}\right) = \sigma^{2}(n-K)\end{aligned}$$

from the fact that $\mathrm{var}(u^{*}) = \sigma^{2}I_n$. Hence, $E(s^{*2}) = \sigma^{2}$ and s^{*2} is unbiased for σ^{2}.

9.3 *The AR(1) Model.*

a. From (9.9) and (9.10) we get

$$\Omega\Omega^{1} = \left[\frac{1}{1-\rho^{2}}\right]\begin{bmatrix} 1 & \rho & \rho^{2} & \ldots & \rho^{T-1} \\ \rho & 1 & \rho & \ldots & \rho^{T-2} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \ldots & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & -\rho & 0 & \ldots & 0 & 0 & 0 \\ -\rho & 1+\rho^{2} & -\rho & \ldots & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \ldots & -\rho & 1+\rho^{2} & -\rho \\ 0 & 0 & 0 & \ldots & 0 & -\rho & 1 \end{bmatrix}$$

$$= \left[\frac{1}{1-\rho^{2}}\right]\begin{bmatrix} (1-\rho^{2}) & 0 & 0 & \ldots & 0 \\ 0 & (1-\rho^{2}) & 0 & \ldots & 0 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & 0 & \ldots & (1-\rho^{2}) \end{bmatrix} = I_T$$

The multiplication is tedious but simple. The (1,1) element automatically gives $(1-\rho^{2})$. The (1,2) element gives $-\rho + \rho(1+\rho^{2}) - \rho\rho^{2} = -\rho +$

$\rho + \rho^3 - \rho^3 = 0$. The (2,2) element gives $-\rho^2 + (1 + \rho^2) - \rho\rho = 1 - \rho^2$ and so on.

b. For P^{-1} defined in (9.11), we get

$$P^{-1\prime} P^{-1} = \begin{bmatrix} \sqrt{1-\rho^2} & -\rho & 0 & \ldots & 0 & 0 \\ 0 & 1 & -\rho & \ldots & 0 & 0 \\ 0 & 0 & 1 & \ldots & 0 & 0 \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ 0 & 0 & 0 & \ldots & 1 & -\rho \\ 0 & 0 & 0 & \ldots & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & \ldots & 0 & 0 & 0 \\ -\rho & 1 & 0 & \ldots & 0 & 0 & 0 \\ 0 & -\rho & 1 & \ldots & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \ldots & -\rho & 1 & 0 \\ 0 & 0 & 0 & \ldots & 0 & -\rho & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & -\rho & 0 & \ldots & 0 & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & \ldots & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \ldots & -\rho & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & \ldots & 0 & -\rho & 1 \end{bmatrix} = (1-\rho^2)\Omega^{-1}$$

Again, the multiplication is simple but tedious. The (1,1) element gives

$$\sqrt{1-\rho^2}\sqrt{1-\rho^2} - \rho(-\rho) = (1-\rho^2) + \rho^2 = 1,$$

the (1,2) element gives $\sqrt{1-\rho^2}.0 - \rho.1 = -\rho$, the (2,2) element gives $1 - \rho(-\rho) = 1 + \rho^2$ and so on.

c. From part (b) we verified that $P^{-1\prime}P^{-1} = (1-\rho^2)\Omega^{-1}$. Hence, $\Omega/(1-\rho^2) = PP'$ or $\Omega = (1-\rho^2)PP'$. Therefore,

$$\begin{aligned} \mathrm{var}(P^{-1}u) &= P^{-1}\mathrm{var}(u)P^{-1\prime} = \sigma_u^2 P^{-1}\Omega P^{-1\prime} \\ &= \sigma_u^2(1-\rho^2)P^{-1}PP'P^{-1\prime} = \sigma_\varepsilon^2 I_T \end{aligned}$$

since $\sigma_u^2 = \sigma_\varepsilon^2/(1-\rho^2)$.

9.4 *Restricted GLS*. From Chap. 7, restricted least squares is given by $\hat{\beta}_{rls} = \hat{\beta}_{ols} + (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(r - R\hat{\beta}_{ols})$. Applying the same analysis to the transformed model in (9.3) we get that $\hat{\beta}^*_{ols} = (X^{*\prime}X^*)^{-1}X^{*\prime}y^* = \hat{\beta}_{GLS}$. From (9.4) and the above restricted estimator, we get

$$\hat{\beta}_{RGLS} = \hat{\beta}_{GLS} + (X^{*\prime}X^*)^{-1}R'[R(X^{*\prime}X^*)^{-1}R']^{-1}(r - R\,\hat{\beta}_{GLS})$$

where X^* now replaces X. But $X^{*\prime}X^* = X'\Omega^{-1}X$, hence,

$$\hat{\beta}_{RGLS} = \hat{\beta}_{GLS} + (X'\Omega^{-1}X)^{-1}R'[R(X'\Omega^{-1}X)^{-1}R']^{-1}(r - R\,\hat{\beta}_{GLS}).$$

9.5 *Best Linear Unbiased Prediction*. This is based on Goldberger (1962).

a. Consider linear predictors of the scalar y_{T+s} given by $\hat{y}_{T+s} = c'y$. From (9.1) we get $\hat{y}_{T+s} = c'X\beta + c'u$ and using the fact that $y_{T+s} = x'_{T+s}\beta + u_{T+s}$, we get

$$\hat{y}_{T+s} - y_{T+s} = (c'X - x'_{T+s})\beta + c'u - u_{T+s}.$$

The unbiased condition is given by $E(\hat{y}_{T+s} - y_{T+s}) = 0$. Since $E(u) = 0$ and $E(u_{T+s}) = 0$, this requires that $c'X = x'_{T+s}$ for this to hold for every β. Therefore, an unbiased predictor will have prediction error

$$\hat{y}_{T+s} - y_{T+s} = c'u - u_{T+s}.$$

b. The prediction variance is given by

$$\begin{aligned}\text{var}(\hat{y}_{T+s}) &= E(\hat{y}_{T+s} - y_{T+s})(\hat{y}_{T+s} - y_{T+s})' = E(c'u - u_{T+s})(c'u - u_{T+s})' \\ &= c'E(uu')c + \text{var}(u_{T+s}) - 2c'E(u_{T+s}u) = c'\Sigma c + \sigma^2_{T+s} - 2c'\omega\end{aligned}$$

using the definitions $\sigma^2_{T+s} = \text{var}(u_{T+s})$ and $\omega = E(u_{T+s}u)$.

c. Minimizing $\text{var}(\hat{y}_{T+s})$ subject to $c'X = x'_{T+s}$ sets up the following Lagrangian function

$$\psi(c, \lambda) = c'\Sigma c - 2c'\omega - 2\lambda'(X'c - x_{T+s})$$

where σ^2_{T+s} is fixed and where λ denotes the Kx1 vector of Lagrangian multipliers. The first order conditions of ψ with respect to c and λ yield $\partial\psi/\partial c = 2\Sigma c - 2\omega - 2X\lambda = 0$ and $\partial\psi/\partial\lambda = 2X'c - 2x_{T+s} = 0$.

In matrix form, these two equations become

$$\begin{bmatrix} \Sigma & X \\ X' & 0 \end{bmatrix} \begin{pmatrix} \hat{c} \\ -\hat{\lambda} \end{pmatrix} = \begin{pmatrix} \omega \\ x_{T+s} \end{pmatrix}.$$

Using partitioned inverse matrix formulas one gets

$$\begin{pmatrix} \hat{c} \\ -\hat{\lambda} \end{pmatrix} = \begin{bmatrix} \Sigma^{-1}[I_T - X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}] & \Sigma^{-1}X(X'\Sigma^{-1}X)^{-1} \\ (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1} & -(X'\Sigma^{-1}X)^{-1} \end{bmatrix} \begin{pmatrix} \omega \\ x_{T+s} \end{pmatrix}$$

so that

$$\hat{c} = \Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}x_{T+s} + \Sigma^{-1}[I_T - X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}]\omega.$$

Therefore, the BLUP is given by

$$\begin{aligned} \hat{y}_{T+s} = \hat{c}'y &= x'_{T+s}(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y + \omega'\Sigma^{-1}y \\ &\quad - \omega'\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y \\ &= x'_{T+s}\hat{\beta}_{GLS} + \omega'\Sigma^{-1}y - \omega'\Sigma^{-1}X\hat{\beta}_{GLS} \\ &= x'_{T+s}\hat{\beta}_{GLS} + \omega'\Sigma^{-1}(y - X\hat{\beta}_{GLS}) \\ &= x'_{T+s}\hat{\beta}_{GLS} + \omega'\Sigma^{-1}e_{GLS} \end{aligned}$$

where $e_{GLS} = y - X\hat{\beta}_{GLS}$. For $\Sigma = \sigma^2\Omega$, this can also be written as

$$\hat{y}_{T+s} = x'_{T+s}\hat{\beta}_{GLS} + \omega'\Omega^{-1}e_{GLS}/\sigma^2.$$

d. For the stationary AR(1) case

$$u_t = \rho u_{t-1} + \varepsilon \quad \text{with} \quad \varepsilon_t \sim \text{IID}\left(0, \sigma_\varepsilon^2\right)$$

$|\rho| < 1$ and $\text{var}(u_t) = \sigma_u^2 = \sigma_\varepsilon^2/(1-\rho^2)$. In this case, $\text{cov}(u_t, u_{t-s}) = \rho^s\sigma_u^2$. Therefore, for s periods ahead forecast, we get

$$\omega = E(u_{T+s}u) = \begin{pmatrix} E(u_{T+s}u_1) \\ E(u_{T+s}u_2) \\ \vdots \\ E(u_{T+s}u_T) \end{pmatrix} = \sigma_u^2 \begin{pmatrix} \rho^{T+s-1} \\ \rho^{T+s-2} \\ \vdots \\ \rho^s \end{pmatrix}.$$

From Ω given in (9.9) we can deduce that $\omega = \rho^s\sigma_u^2$ (last column of Ω). But, $\Omega^{-1}\Omega = I_T$. Hence, Ω^{-1} (last column of Ω) = (last column of I_T) = $(0, 0, .., 1)'$. Substituting for the last column of Ω the expression $\left(\omega/\rho^s\sigma_u^2\right)$ yields

$$\Omega^{-1}\ \omega/\rho^s\sigma_u^2 = (0, 0, .., 1)'$$

which can be transposed and rewritten as

$$\omega'\Omega^{-1}/\sigma_u^2 = \rho^s(0, 0, .., 1).$$

Substituting this expression in the BLUP for y_{T+s} in part (c) we get

$$\begin{aligned}\hat{y}_{T+s} &= x'_{T+s}\hat{\beta}_{GLS} + \omega'\Omega^{-1}e_{GLS}/\sigma^2 = x'_{T+s}\hat{\beta}_{GLS} + \rho^s(0, 0, .., 1)e_{GLS}\\ &= x'_{T+s}\hat{\beta}_{GLS} + \rho^s e_{T,GLS}\end{aligned}$$

where $e_{T,GLS}$ is the T-th GLS residual. For $s = 1$, this gives

$$\hat{y}_{T+1} = x'_{T+1}\hat{\beta}_{GLS} + \rho e_{T,GLS}$$

as shown in the text.

9.6 *The W, LR and LM Inequality*. From Eq. (9.27), the Wald statistic W can be interpreted as a LR statistic conditional on $\hat{\Sigma}$, the unrestricted MLE of Σ, i.e., $W = -2\log[\max\limits_{R\beta=r} L(\beta/\hat{\Sigma})/\max\limits_{\beta} L(\beta/\hat{\Sigma})]$. But, from (9.34), we know that the likelihood ratio statistic $LR = -2\log[\max\limits_{R\beta=r,\Sigma} L(\beta, \Sigma)/\max\limits_{\beta,\Sigma} L(\beta, \Sigma)]$. Using (9.33), $\max\limits_{R\beta=r} L\left(\beta/\hat{\Sigma}\right) \leq \max\limits_{R\beta=r,\Sigma} L(\beta, \Sigma)$. The right hand side term is an unconditional maximum over all Σ whereas the left hand side is a conditional maximum based on $\hat{\Sigma}$ under the null hypothesis H_o; $R\beta = r$. Also, from (9.32) $\max\limits_{\beta,\Sigma} L(\beta, \Sigma) = \max\limits_{\beta} L(\beta/\hat{\Sigma})$. Therefore, $W \geq LR$. Similarly, from Eq. (9.31), the Lagrange Multiplier statistic can be interpreted as a LR statistic conditional on $\tilde{\Sigma}$, the restricted maximum likelihood of Σ, i.e., $LM = -2\log[\max\limits_{R\beta=r} L(\beta/\tilde{\Sigma})/\max\limits_{\beta} L(\beta/\tilde{\Sigma})]$. Using (9.33), $\max\limits_{R\beta=r} L(\beta/\tilde{\Sigma}) = \max\limits_{R\beta=r,\Sigma} L(\beta, \Sigma)$ and from (9.32), we get $\max\limits_{\beta} L(\beta/\tilde{\Sigma}) \leq \max\limits_{\beta,\Sigma} L(\beta, \Sigma)$ because the latter is an unconditional maximum over all Σ. Hence, LR $\geq$ LM. Therefore, W $\geq$ LR $\geq$ LM.

9.7 The W, LR and LM for this simple regression with H_o; $\beta = 0$ were derived in problem 7.16 in Chap. 7. Here, we follow the alternative derivation proposed by Breusch (1979) and considered in problem 9.6. From (9.34), the LR is given by $LR = -2\log\left[L\left(\tilde{\alpha},\ \tilde{\beta} = 0,\ \tilde{\sigma}^2\right)/L\left(\hat{\alpha}_{mle}, \hat{\beta}_{mle}, \hat{\sigma}^2_{mle}\right)\right]$

$$\text{where}\quad \tilde{\alpha} = \bar{y},\ \tilde{\beta} = 0,\ \tilde{\sigma}^2 = \sum_{i=1}^{n}(y_i - \bar{y})^2/n \text{ and}$$

$$\hat{\alpha}_{mle} = \hat{\alpha}_{ols} = \bar{y} - \hat{\beta}_{ols}\bar{X},$$

$$\hat{\beta}_{mle} = \hat{\beta}_{ols} = \sum_{i=1}^{n} x_i y_i / \sum_{i=1}^{n} x_i{}^2,\ \hat{\sigma}^2_{mle} = \sum_{i=1}^{n} e_i{}^2/n$$

and $e_i = y_i - \hat{\alpha}_{ols} - \hat{\beta}_{ols}X_i$, see the solution to problem 7.16. But,

$$\log L\left(\alpha, \beta, \sigma^2\right) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \sum_{i=1}^{n}(y_i - \alpha - \beta X_i)^2/2\sigma^2.$$

Therefore,

$$\log L\left(\tilde{\alpha},\ \tilde{\beta} = 0,\ \tilde{\sigma}^2\right) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\tilde{\sigma}^2 - \sum_{i=1}^{n}(y_i - \bar{y})^2/2\tilde{\sigma}^2$$
$$= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\tilde{\sigma}^2 - \frac{n}{2}$$

and

$$\log L\left(\hat{\alpha}_{mle}, \hat{\beta}_{mle}, \hat{\sigma}^2_{mle}\right) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\hat{\sigma}^2_{mle} - \sum_{i=1}^{n} e_i^2/2\hat{\sigma}^2_{mle}$$
$$= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\hat{\sigma}^2_{mle} - \frac{n}{2}.$$

Therefore,

$$LR = -2\left[-\frac{n}{2}\log\tilde{\sigma}^2 + \frac{n}{2}\log\hat{\sigma}^2_{mle}\right] = n\log\left(\tilde{\sigma}^2/\hat{\sigma}^2_{mle}\right)$$
$$= n\log(TSS/RSS) = n\log(1/1 - R^2)$$

where TSS = total sum of squares, and RSS = residual sum of squares for the simple regression. Of course, $R^2 = 1-$ (RSS/TSS).

Similarly, from (9.31) we have

$$LM = -2\log\left[\max_{\alpha,\beta=0} L\left(\alpha,\beta/\tilde{\sigma}^2\right) / \max_{\alpha,\beta} L(\alpha,\beta/\tilde{\sigma}^2)\right].$$

But, maximization of $L(\alpha,\beta/\tilde{\sigma}^2)$ gives $\hat{\alpha}_{ols}$ and $\hat{\beta}_{ols}$. Therefore,

$$\max_{\alpha,\beta} L(\alpha,\beta/\tilde{\sigma}^2) = L\left(\hat{\alpha},\hat{\beta},\tilde{\sigma}^2\right)$$

with

$$\log L\left(\hat{\alpha},\hat{\beta},\tilde{\sigma}^2\right) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\tilde{\sigma}^2 - \sum_{i=1}^{n} e_i^2/2\tilde{\sigma}^2.$$

Also, restricted maximization of $L(\alpha,\beta/\tilde{\sigma}^2)$ under H_o; $\beta = 0$ gives $\tilde{\alpha} = \bar{y}$ and $\tilde{\beta} = 0$. Therefore, $\max_{\alpha,\beta=0} L(\alpha,\beta/\tilde{\sigma}^2) = L(\tilde{\alpha},\tilde{\beta},\tilde{\sigma}^2)$. From this, we conclude that

$$LM = -2\left[-\frac{n}{2} + \frac{\sum_{i=1}^{n} e_i{}^2}{2\tilde{\sigma}^2}\right] = n - \left(\sum_{i=1}^{n} e_i^2 \Big/ \tilde{\sigma}^2\right)$$

$$= n - \left(n\sum_{i=1}^{n} e_i^2 \Big/ \sum_{i=1}^{n} y_i^2\right) = n[1-(RSS/TSS)] = nR^2.$$

Finally, from (9.27), we have $W = -2\log\left[\max_{\alpha,\beta=0} L\left(\alpha,\beta/\hat{\sigma}^2_{mle}\right)/\max_{\alpha,\beta} L\left(\alpha,\beta/\hat{\sigma}^2_{mle}\right)\right]$. The maximization of $L(\alpha,\beta/\hat{\sigma}^2)$ gives $\hat{\alpha}_{ols}$ and $\hat{\beta}_{ols}$. Therefore,

$$\max_{\alpha\beta} L(\alpha,\beta/\hat{\sigma}^2) = L(\hat{\alpha},\hat{\beta},\hat{\sigma}^2).$$

Also, restricted maximization of $L(\alpha,\beta/\hat{\sigma}^2)$ under $\beta = 0$ gives $\tilde{\alpha} = \bar{y}$ and $\tilde{\beta} = 0$. Therefore, $\max_{\alpha,\beta=0} L(\alpha,\beta/\tilde{\sigma}^2) = L(\tilde{\alpha},\tilde{\beta} = 0,\hat{\sigma}^2)$ with

$$\log L(\tilde{\alpha},\tilde{\beta} = 0,\hat{\sigma}^2) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\hat{\sigma}^2 - \sum_{i=1}^{n}(y_i - \bar{y})^2/2\hat{\sigma}^2_{mle}.$$

Therefore,

$$W = -2\left[-\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{2\hat{\sigma}^2_{mle}}+\frac{n}{2}\right] = \frac{TSS}{\hat{\sigma}^2_{mle}} - n = \frac{nTSS}{RSS} - n$$

$$= n\left(\frac{TSS-RSS}{RSS}\right) = n\left(\frac{R^2}{1-R^2}\right).$$

This is exactly what we got in problem 7.16, but now from Breusch's (1979) alternative derivation. From problem 9.6, we infer using this LR interpretation of all three statistics that $W \geq LR \geq LM$.

9.8 *Sampling Distributions and Efficiency Comparison of OLS and GLS.* This is based on Baltagi (1992).

a. From the model it is clear that $\sum_{t=1}^{2} x_t^2 = 5,\ y_1 = 2+u_1,\ y_2 = 4+u_2$, and

$$\hat{\beta}_{ols} = \frac{\sum_{t=1}^{2} x_t y_t}{\sum_{t=1}^{2} x_t{}^2} = \beta + \frac{\sum_{t=1}^{2} x_t u_t}{\sum_{t=1}^{2} x_t{}^2} = 2 + 0.2u_1 + 0.4u_2$$

Let $u' = (u_1, u_2)$, then it is easy to verify that $E(u) = 0$ and

$$\Omega = var(u) = \begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix}.$$

The disturbances have zero mean, are heteroskedastic and serially correlated with a correlation coefficient $\rho = -0.5$.

b. Using the joint probability function $P(u_1, u_2)$ and $\hat{\beta}_{ols}$ from part (a), one gets

$\hat{\beta}_{ols}$	Probability
1	1/8
1.4	3/8
2.6	3/8
3	1/8

Therefore, $E(\hat{\beta}_{ols}) = \beta = 2$ and $var(\hat{\beta}_{ols}) = 0.52$. These results can be also verified from $\hat{\beta}_{ols} = 2 + 0.2u_1 + 0.4u_2$. In fact, $E(\hat{\beta}_{ols}) = 2$ since

$E(u_1) = E(u_2) = 0$ and

$$\mathrm{var}\left(\hat{\beta}_{ols}\right) = 0.04\,\mathrm{var}(u_1) + 0.16\,\mathrm{var}(u_2) + 0.16\,\mathrm{cov}(u_1, u_2)$$

$$= 0.04 + 0.64 - 0.16 = 0.52.$$

Also,

$$\Omega^{-1} = \frac{1}{3}\begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix} \text{ and } (x'\Omega^{-1}x)^{-1} = 1/4 = \mathrm{var}\left(\tilde{\beta}_{GLS}\right)$$

In fact, $\tilde{\beta}_{GLS} = (x'\Omega^{-1}x)^{-1}x'\Omega^{-1}y = 1/4(2y_1 + y_2)$ which can be rewritten as

$$\tilde{\beta}_{GLS} = 2 + 1/4[2u_1 + u_2]$$

Using $P(u_1, u_2)$ and this equation for $\tilde{\beta}_{GLS}$, one gets

$\tilde{\beta}_{GLS}$	Probability
1	1/8
2	3/4
3	1/8

Therefore, $E(\tilde{\beta}_{GLS}) = \beta = 2$ and $\mathrm{var}(\tilde{\beta}_{GLS}) = 0.25$. This can also be verified from $\tilde{\beta}_{GLS} = 2 + 1/4[2u_1 + u_2]$. In fact, $E(\tilde{\beta}_{GLS}) = 2$ since $E(u_1) = E(u_2) = 0$ and

$$\mathrm{var}\left(\tilde{\beta}_{GLS}\right) = \frac{1}{16}[4\mathrm{var}(u_1) + \mathrm{var}(u_2) + 4\mathrm{cov}(u_1, u_2)] = \frac{1}{16}[4 + 4 - 4] = \frac{1}{4}.$$

This variance is approximately 48% of the variance of the OLS estimator.

c. The OLS predictions are given by $\hat{y}_t = \hat{\beta}_{ols}x_t$ which means that $\hat{y}_1 = \hat{\beta}_{ols}$ and $\hat{y}_2 = 2\hat{\beta}_{ols}$. The OLS residuals are given by $\hat{e}_t = y_t - \hat{y}_t$ and their probability function is given by

$(\hat{e}_1, \hat{e}_2)$	Probability
(0,0)	1/4
(1.6, −0.8)	3/8
(−1.6, 0.8)	3/8

Now the estimated $\mathrm{var}(\hat{\beta}_{ols}) = S(\hat{\beta}_{ols}) = \dfrac{s^2}{\sum_{t=1}^{2} x_t{}^2} = \dfrac{\Sigma\hat{e}_t^2}{5}$, and this has probability function

$S(\hat{\beta}_{ols})$	Probability
0	1/4
0.64	3/4

with $E\left[S\left(\hat{\beta}_{ols}\right)\right] = 0.48 \neq \mathrm{var}\left(\hat{\beta}_{ols}\right) = 0.52.$

Similarly, the GLS predictions are given by $\tilde{y}_t = \tilde{\beta}_{GLS}x_t$ which means that $\tilde{y}_1 = \tilde{\beta}_{GLS}$ and $\tilde{y}_2 = 2\tilde{\beta}_{GLS}$. The GLS residuals are given by $\tilde{e}_t = y_t - \tilde{y}_t$ and their probability function is given by

$(\tilde{e}_1, \tilde{e}_2)$	Probability
(0,0)	1/4
$(1, -2)$	3/8
$(-1, 2)$	3/8

The MSE of the GLS regression is given by $\tilde{s}^2 = \tilde{e}'\Omega^{-1}\tilde{e} = 1/3\left[4\tilde{e}_1^2 + 2\tilde{e}_1\tilde{e}_2 + \tilde{e}_2^2\right]$ and this has a probability function

$\tilde{s}^2$	Probability
0	1/4
4/3	3/4

with $E(\tilde{s}^2) = 1$. An alternative solution of this problem is given by Im and Snow (1993).

9.9 *Equi-correlation.*

a. For the regression with equi-correlated disturbances, OLS is equivalent to GLS as long as there is a constant in the regression model. Note that

$$\Omega = \begin{bmatrix} 1 & \rho & \rho & \dots & \rho \\ \rho & 1 & \rho & \dots & \rho \\ \vdots & \vdots & \vdots & & \vdots \\ \rho & \rho & \rho & \dots & 1 \end{bmatrix}$$

so that u_t is homoskedastic and has constant serial correlation. In fact, correlation $(u_t, u_{t-s}) = \rho$ for $t \neq s$. Therefore, this is called equi-correlated. Zyskind's (1967) condition given in (9.8) yields

$$P_X\Omega = \Omega P_X.$$

In this case,

$$P_X\Omega = (1-\rho)P_X + \rho P_X\iota_T\iota_T'$$

and

$$\Omega P_X = (1-\rho)P_X + \rho\iota_T\iota_T'P_X.$$

But, we know that X contains a constant, i.e., a column of ones denoted by ι_T. Therefore, using $P_XX = X$ we get $P_X\iota_T = \iota_T$ since ι_T is a column of X. Substituting this in $P_X\Omega$ we get

$$P_X\Omega = (1-\rho)P_X + \rho\iota_T\iota_T'.$$

Similarly, substituting $\iota_T'P_X = \iota_T'$ in ΩP_X we get $\Omega P_X = (1-\rho)P_X + \rho\iota_T\iota_T'$. Hence, $\Omega P_X = P_X\Omega$ and OLS is equivalent to GLS for this model.

b. We know that $(T-K)s^2 = u'\bar{P}_Xu$, see Chap. 7. Also that

$$E(u'\bar{P}_Xu) = E[tr(uu'\bar{P}_X)] = tr[E(uu'\bar{P}_X)] = tr(\sigma^2\Omega\bar{P}_X)$$
$$= \sigma^2 tr[(1-\rho)\bar{P}_X + \rho\iota_T\iota_T'\bar{P}_X] = \sigma^2(1-\rho)tr(\bar{P}_X)$$

since $\iota_T'\bar{P}_X = \iota_T' - \iota_T'P_X = \iota_T' - \iota_T' = 0$ see part (a). But, $tr(\bar{P}_X) = T-K$, hence, $E(u'\bar{P}_Xu) = \sigma^2(1-\rho)(T-K)$ and $E(s^2) = \sigma^2(1-\rho)$.

Now for Ω to be positive semi-definite, it should be true that for every arbitrary non-zero vector a we have $a'\Omega a \geq 0$. In particular, for $a = \iota_T$, we get

$$\iota_T'\Omega\iota_T = (1-\rho)\iota_T'\iota_T + \rho\iota_T'\iota_T\iota_T'\iota_T = T(1-\rho) + T^2\rho.$$

This should be non-negative for every ρ. Hence, $(T^2 - T)\rho + T \geq 0$ which gives $\rho \geq -1/(T-1)$. But, we know that $|\rho| \leq 1$. Hence, $-1/(T-1) \leq$

$\rho \leq 1$ as required. This means that $0 \leq E(s^2) \leq [T/(T-1)]\sigma^2$ where the lower and upper bounds for $E(s^2)$ are attained at $\rho = 1$ and $\rho = -1/(T-1)$, respectively. These bounds were derived by Dufour (1986).

9.10 **a.** The model can be written in vector form as: $y = \alpha\iota_n + u$ where $y' = (y_1, .., y_n)$, ι_n is a vector of ones of dimension n, and $u' = (u_1, .., u_n)$. Therefore, $\hat{\alpha}_{ols} = \left(\iota_n'\iota_n\right)^{-1} \iota_n' y = \sum_{i=1}^{n} y_i/n = \bar{y}$ and

$$\Sigma = \text{var}(u) = E(uu') = \sigma^2 \begin{bmatrix} 1 & \rho & .. & \rho \\ \rho & 1 & .. & \rho \\ \vdots & \vdots & .. & \vdots \\ \rho & \rho & .. & 1 \end{bmatrix} = \sigma^2[(1-\rho)I_n + \rho J_n]$$

where I_n is an identity matrix of dimension n and J_n is a matrix of ones of dimension n. Define $E_n = I_n - \bar{J}_n$ where $\bar{J}_n = J_n/n$, one can rewrite Σ as $\Sigma = \sigma^2[(1-\rho)E_n + (1+\rho(n-1))\bar{J}_n] = \sigma^2\Omega$ with

$$\Omega^{-1} = \left[\frac{1}{1-\rho}E_n + \frac{1}{1+\rho(n-1)}\bar{J}_n\right]$$

Therefore,

$$\hat{\alpha}_{GLS} = (\iota_n'\Sigma^{-1}\iota_n)^{-1}\iota_n'\Sigma^{-1}y = \left(\frac{\iota_n'\bar{J}_n\iota_n}{1+\rho(n-1)}\right)^{-1} \frac{\iota_n'\bar{J}_n y}{1+\rho(n-1)}$$

$$= \frac{\iota_n'\bar{J}_n y}{n} = \frac{\iota_n' y}{n} = \bar{y}$$

b. $s^2 = e'e/(n-1)$ where e is the vector of OLS residuals with typical element $e_i = y_i - \bar{y}$ for $i = 1, .., n$. In vector form, $e = E_n y$ and

$$s^2 = y'E_n y/(n-1) = u'E_n u/(n-1)$$

since $E_n\,\iota_n = 0$. But,

$$E(u'E_n u) = \text{tr}(\Sigma E_n) = \sigma^2 \text{tr}[(1-\rho)E_n] = \sigma^2(1-\rho)(n-1)$$

since $E_n\bar{J}_n = 0$ and $\text{tr}(E_n) = (n-1)$. Hence, $E(s^2) = \sigma^2(1-\rho)$ and $E(s^2) - \sigma^2 = -\rho\sigma^2$.

This bias is negative if $0 < \rho < 1$ and positive if $-1/(n-1) < \rho < 0$.

c. $s_*^2 = e'_{GLS}\Omega^{-1}e_{GLS}/(n-1) = e'\Omega^{-1}e/(n-1)$ where e_{GLS} denotes the vector of GLS residuals which in this case is identical to the OLS residuals. Substituting for $e = E_n y$ we get

$$s_*^2 = \frac{y'E_n\Omega^{-1}E_n y}{n-1} = \frac{u'E_n\Omega^{-1}E_n u}{(n-1)}$$

since $E_n\iota_n = 0$. Hence,

$$E(s_*^2) = \sigma^2 tr(\Omega E_n\Omega^{-1}E_n)/(n-1) = \frac{\sigma^2}{(n-1)}tr[(1-\rho)E_n]\left[\frac{1}{1-\rho}E_n\right]$$

$$= \frac{\sigma^2}{(n-1)}tr[E_n] = \sigma^2$$

d. true var $(\hat{\alpha}_{ols}) = (\iota'_n\iota_n)^{-1}\ \iota'_n\Sigma\iota_n(\iota'_n\iota_n)^{-1} = \iota'_n\Sigma\iota_n/n^2$

$$= \sigma^2[(1+\rho(n-1))\iota'_n\bar{J}_n\iota_n]/n^2 = \sigma^2[1+\rho(n-1)]/n \quad (9.1)$$

which is equal to var $(\hat{\alpha}_{GLS}) = \left(\iota'_n\Sigma^{-1}\iota_n\right)^{-1}$ as it should be.

estimated var$((\hat{\alpha}_{ols}) = s^2\left(\iota'_n\iota_n\right)^{-1} = s^2/n$ so that

$$E[\text{estimated var}(\hat{\alpha}_{ols}) - \text{true var}(\hat{\alpha}_{ols})] = E(s^2)/n - \sigma^2[1+\rho(n-1)]/n$$

$$= \sigma^2[1-\rho-1-\rho(n-1)]/n$$

$$= -\rho\sigma^2.$$

9.15 *Neighborhood Effects and Housing Demand*

a. This replicates the first three columns of Table VII in Ioannides and Zabel (2003, p. 569) generating descriptive statistics on key variables, by year:

```
. by year, sort: sum price pincom1 highschool changehand white npersons married
```

-> year = 1985

Variable	Obs	Mean	Std. Dev.	Min	Max
price	1947	81.92058	25.0474	44.89161	146.1314
pincom1	1947	28.55038	15.47855	3.557706	90.00319
highschool	1947	.8361582	.3702271	0	1
changehand	1947	.2891628	.4534901	0	1
white	1947	.8798151	.3252612	0	1
npersons	1947	2.850539	1.438622	1	11
married	1947	.7134052	.4522867	0	1

-> year = 1989

Variable	Obs	Mean	Std. Dev.	Min	Max
price	2318	116.7232	49.82718	48.3513	220.3118
pincom1	2318	47.75942	30.3148	4.444763	174.0451
highschool	2318	.8597929	.3472767	0	1
changehand	2318	.3170837	.4654407	0	1
white	2318	.8658326	.3409056	0	1
npersons	2318	2.768335	1.469969	1	11
married	2318	.6535807	.4759314	0	1

-> year = 1993

Variable	Obs	Mean	Std. Dev.	Min	Max
price	2909	115.8608	44.73127	53.93157	240.2594
pincom1	2909	50.07294	29.95046	6.201	184.7133
highschool	2909	.8697147	.3366749	0	1
changehand	2909	.2781024	.4481412	0	1
white	2909	.8480578	.3590266	0	1
npersons	2909	2.738398	1.435682	1	9
married	2909	.6452389	.4785231	0	1

This replicates the last column of Table VII in Ioannides and Zabel (2003, p.569) generating descriptive statistics on key variables for the pooled data:

. sum price pincom1 highschool changehand white npersons married

Variable	Obs	Mean	Std. Dev.	Min	Max
price	7174	106.9282	44.90505	44.89161	240.2594
pincom1	7174	43.48427	28.45273	3.557706	184.7133
highschool	7174	.8574017	.3496871	0	1
changehand	7174	.2936995	.4554877	0	1
white	7174	.8624198	.3444828	0	1
npersons	7174	2.778506	1.448163	1	11
married	7174	.6664343	.4715195	0	1

b. This replicates column 1 of Table VIII of Ioannides and Zabel(2003, p. 577) estimating mean of neighbors housing demand. The estimates are close but do not match.

. reg lnhdemm lnprice d89 d93 lnpincomem highschoolm changehandm whitem npersonsm marriedm hagem hage2m fullbathsm bedroomsm garagem

Source	SS	df	MS
Model	882.202688	14	63.0144777
Residual	572.231228	7159	.079931726
Total	1454.43392	7173	.20276508

Number of obs	=	7174
F(14, 7159)	=	788.35
Prob > F	=	0.0000
R-squared	=	0.6066
Adj R-squared	=	0.6058
Root MSE	=	.28272

lnhdemm	Coef.	Std. Err.	t	P> |t|	[95% Conf. Interval]	
lnprice	-.2429891	.0266485	-9.12	0.000	-.295228	-.1907503
d89	-.0967419	.0102894	-9.40	0.000	-.1169122	-.0765715
d93	-.1497614	.0109956	-13.62	0.000	-.1713159	-.1282068
lnpincomem	.3622927	.0250064	14.49	0.000	.3132728	.4113126
highschoolm	.1185672	.0263588	4.50	0.000	.066896	.1702383
changehandm	.0249327	.0179043	1.39	0.164	-.0101651	.0600305
whitem	.2402858	.0144223	16.66	0.000	.2120139	.2685577
npersonsm	-.0692484	.0069556	-9.96	0.000	-.0828834	-.0556134
marriedm	.1034179	.0236629	4.37	0.000	.0570315	.1498042
hagem	.0074906	.0009053	8.27	0.000	.0057159	.0092652
hage2m	-.008222	.0010102	-8.14	0.000	-.0102023	-.0062417
fullbathsm	.2544969	.0085027	29.93	0.000	.2378291	.2711647
bedroomsm	.1770101	.009258	19.12	0.000	.1588616	.1951586
garagem	.2081956	.0135873	15.32	0.000	.1815604	.2348308
_cons	2.861724	.1188536	24.08	0.000	2.628735	3.094712

This replicates column 2 of Table VIII of Ioannides and Zabel(2003, p.577) estimating a standard housing demand with no neighborhood effects. The estimates do not match. This may be because the authors included only one observation per cluster. Here, all observations are used.

. reg lnhdem lnprice lnpincome highschool changehand white npersons married d89 d93, vce(cluster neigh)

Linear regression

Number of obs	=	7174
F(9, 364)	=	24.53
Prob > F	=	0.0000
R-squared	=	0.1456
Root MSE	=	.49666

(Std. Err. adjusted for 365 clusters in neigh)

lnhdemm	Coef.	Robust Std. Err.	t	P> |t|	[95% Conf. Interval]	
lnprice	-.1982708	.059528	-3.33	0.001	-.3153328	-.0812088
lnpincome	.3376063	.0346802	9.73	0.000	.2694076	.405805
highschool	.0702029	.0285124	2.46	0.014	.0141332	.1262727
changehand	.0009962	.0210258	0.05	0.962	-.040351	.0423434

white	.1964783	.0563134	3.49	0.001	.0857379	.3072186
npersons	-.0038604	.0077642	-0.50	0.619	-.0191287	.0114078
married	-.000329	.0227977	-0.01	0.988	-.0451607	.0445026
d89	-.0882059	.0223177	-3.95	0.000	-.1320937	-.0443181
d93	-.1181923	.0240871	-4.91	0.000	-.1655596	-.0708249
_cons	4.144696	.2843044	14.58	0.000	3.585611	4.703781

This replicates column 3 of Table VIII of Ioannides and Zabel(2003, p.577) estimating a reduced form housing demand. The estimates do not match.

```
. reg lnhdem lnprice lnpincome highschool changehand white npersons married d89
> d93 lnpincomem highschoolm changehandm whitem npersonsm marriedm hagem
hage2m fullbathsm bedroomsm garagem, vce(cluster neigh)
```

Linear regression

Number of obs	=	7174
F(20, 364)	=	34.27
Prob > F	=	0.0000
R-squared	=	0.4220
Root MSE	=	.40883

(Std. Err. adjusted for 365 clusters in neigh)

lnhdemm	Coef.	Robust Std. Err.	t	P> |t|	[95% Conf.	Interval]
lnprice	-.2850123	.1280493	-2.23	0.027	-.5368216	-.0332029
lnpincome	.1042973	.0192407	5.42	0.000	.0664603	.1421342
highschool	.0027566	.0172223	0.16	0.873	-.031111	.0366242
changehand	.0154469	.0110693	1.40	0.164	-.006321	.0372148
white	.0189483	.0208229	0.91	0.363	-.022	.0598967
npersons	.0044466	.0042704	1.04	0.298	-.0039512	.0128443
married	.0079301	.0150311	0.53	0.598	-.0216286	.0374889
d89	-.1015519	.0288299	-3.52	0.000	-.158246	-.0448578
d93	-.1571002	.0365114	-4.30	0.000	-.2289	-.0853004
lnpincomem	.2963244	.1108145	2.67	0.008	.0784074	.5142413
highschoolm	.1294341	.0832116	1.56	0.121	-.0342018	.2930701
changehandm	.0029356	.0550799	0.05	0.958	-.1053792	.1112504
whitem	.207916	.0682558	3.05	0.002	.0736908	.3421412
npersonsm	-.0710319	.0216564	-3.28	0.001	-.1136194	-.0284444
marriedm	.1154993	.0812373	1.42	0.156	-.0442541	.2752526
hagem	.0079615	.0039834	2.00	0.046	.0001281	.0157949
hage2m	-.0085303	.0043061	-1.98	0.048	-.0169983	-.0000623
fullbathsm	.2520568	.0343649	7.33	0.000	.1844781	.3196356
bedroomsm	.1514635	.0385524	3.93	0.000	.0756501	.2272769
garagem	.2012601	.0555207	3.62	0.000	.0920784	.3104418
_cons	2.715015	.6149303	4.42	0.000	1.505753	3.924277

References

Baltagi, B.H. (1992), “Sampling Distributions and Efficiency Comparisons of OLS and GLS in the Presence of Both Serial Correlation and Heteroskedasticity,” Econometric Theory, Problem 92.2.3, 8: 304–305.

Breusch, T.S. (1979), “Conflict Among Criteria for Testing Hypotheses: Extensions and Comments,” Econometrica, 47: 203–207.

Dufour, J.M. (1986), “Bias of s^2 in Linear Regressions with Dependent Errors,” The American Statistician, 40: 284–285.

Goldberger, A.S. (1962), “Best Linear Unbiased Prediction in the Generalized Linear Regression Model,” Journal of the American Statistical Association, 57: 369–375.

Ioannides, Y.M. and J.E Zabel (2003), “Neighbourhood Effects and Housing Demand,” Journal of Applied Econometrics 18: 563–584.

Krämer, W. and S. Berghoff (1991), “Consistency of s^2 in the Linear Regression Model with Correlated Errors,” Empirical Economics, 16: 375–377.

Neudecker, H. (1977), “Bounds for the Bias of the Least Squares Estimator of s^2 in Case of a First-Order Autoregressive Process (positive autocorrelation),” Econometrica, 45: 1257–1262.

Neudecker, H. (1978), ”Bounds for the Bias of the LS Estimator in the Case of a First-Order (positive) Autoregressive Process Where the Regression Contains a Constant Term,” Econometrica, 46: 1223–1226.

Sathe, S.T. and H.D. Vinod (1974), “Bounds on the Variance of Regression Coefficients Due to Heteroscedastic or Autoregressive Errors,” Econometrica, 42: 333–340.

Zyskind, G. (1967), “On Canonical Forms, Non-Negative Covariance Matrices and Best and Simple Least Squares Linear Estimators in Linear Models,” The Annals of Mathematical Statistics, 38: 1092–1109.

CHAPTER 10
Seemingly Unrelated Regressions

10.1 *When Is OLS as Efficient as Zellner's SUR?*

a. From (10.2), OLS on this system gives

$$\hat{\beta}_{ols} = \begin{pmatrix} \hat{\beta}_{1,ols} \\ \hat{\beta}_{2,ols} \end{pmatrix} = \begin{bmatrix} X_1'X_1 & 0 \\ 0 & X_2'X_2 \end{bmatrix}^{-1} \begin{bmatrix} X_1'y_1 \\ X_2'y_2 \end{bmatrix}$$

$$= \begin{bmatrix} (X_1'X_1)^{-1} & 0 \\ 0 & (X_2'X_2)^{-1} \end{bmatrix} \begin{pmatrix} X_1'y_1 \\ X_2'y_2 \end{pmatrix} = \begin{bmatrix} (X_1'X_1)^{-1} X_1'y_1 \\ (X_2'X_2)^{-1} X_2'y_2 \end{bmatrix}$$

This is OLS on each equation taken separately. For (10.2), the estimated $\text{var}(\hat{\beta}_{ols})$ is given by

$$s^2 \begin{bmatrix} (X_1'X_1)^{-1} & 0 \\ 0 & (X_2'X_2)^{-1} \end{bmatrix}$$

where $s^2 = RSS/(2T - (K_1 + K_2))$ and RSS denotes the residual sum of squares of this system. In fact, the $RSS = e_1'e_1 + e_2e_2 = RSS_1 + RSS_2$ where

$$e_i = y_i - X_i\hat{\beta}_{i,ols} \qquad \text{for } i = 1, 2.$$

If OLS was applied on each equation separately, then

$$\text{var}\left(\hat{\beta}_{1,ols}\right) = s_1^2 \left(X_1'X_1\right)^{-1} \qquad \text{with} \qquad s_1^2 = RSS_1/(T - K_1)$$

and

$$\text{var}\left(\hat{\beta}_{2,ols}\right) = s_2^2 \left(X_2'X_2\right)^{-1} \qquad \text{with} \qquad s_2^2 = RSS_2/(T - K_2).$$

Therefore, the estimates of the variance–covariance matrix of OLS from the system of two equations differs from OLS on each equation separately by a scalar, namely s^2 rather than s_i^2 for $i = 1, 2$.

B.H. Baltagi, *Solutions Manual for Econometrics*, Springer Texts in Business and Economics, DOI 10.1007/978-3-642-54548-1_10,

b. For the system of equations given in (10.2), $X' = \text{diag}\left[X_i'\right], \Omega^{-1} = \Sigma^{-1} \otimes I_T$ and

$$X'\Omega^{-1} = \begin{bmatrix} X_1' & 0 \\ 0 & X_2' \end{bmatrix} \begin{bmatrix} \sigma^{11} I_T & \sigma^{12} I_T \\ \sigma^{21} I_T & \sigma^{22} I_T \end{bmatrix} = \begin{bmatrix} \sigma^{11} X_1' & \sigma^{12} X_1' \\ \sigma^{21} X_2' & \sigma^{22} X_2' \end{bmatrix}.$$

Also, $X'X = \text{diag}\left[X_i'X_i\right]$ with $(X'X)^{-1} = \text{diag}\left[(X_i'X_i)^{-1}\right]$. Therefore,

$$P_X = \text{diag}[P_{X_i}] \qquad \text{and} \qquad \bar{P}_X = \text{diag}\left[\bar{P}_{X_i}\right].$$

Hence,

$$X'\Omega^{-1}\bar{P}_X = \begin{bmatrix} \sigma^{11} X_1'\bar{P}_{X_1} & \sigma^{12} X_1'\bar{P}_{X_2} \\ \sigma^{21} X_2'\bar{P}_{X_1} & \sigma^{22} X_2'\bar{P}_{X_2} \end{bmatrix}.$$

But, $X_i'\bar{P}_{X_i} = 0$ for $i = 1, 2$. Hence, $X'\Omega^{-1}\bar{P}_X = \begin{bmatrix} 0 & \sigma^{12} X_1'\bar{P}_{X_2} \\ \sigma^{21} X_2'\bar{P}_{X_1} & 0 \end{bmatrix}$ and this is zero if $\sigma^{ij} X_i'\bar{P}_{X_j} = 0$ for $i \neq j$ with $i, j = 1, 2$.

c. (i) If $\sigma_{ij} = 0$ for $i \neq j$, then, Σ is diagonal. Hence, Σ^{-1} is diagonal and $\sigma^{ij} = 0$ for $i \neq j$. This automatically gives $\sigma^{ij} X_i'\bar{P}_{X_j} = 0$ for $i \neq j$ from part (b).

(ii) If all the X_i's are the same, then $X_1 = X_2 = X^*$ and $\bar{P}_{X_1} = \bar{P}_{X_2} = \bar{P}_{X^*}$. Hence,

$$X_i'\bar{P}_{X_j} = X^{*\prime}\bar{P}_{X^*} = 0 \qquad \text{for } i, j = 1, 2.$$

This means that $\sigma^{ij} X_i'\bar{P}_{X_j} = 0$ from part (b) is satisfied for $i, j = 1, 2$.

d. If $X_i = X_j C'$ where C is a non-singular matrix, then $X_i'\bar{P}_{X_j} = C X_j'\bar{P}_{X_j} = 0$. Alternatively, $X_i'X_i = C X_j'X_j C'$ and $\left(X_i'X_i\right)^{-1} = C'^{-1}\left(X_j'X_j\right)^{-1} C^{-1}$. Therefore,

$$P_{X_i} = P_{X_j} \quad \text{and} \quad \bar{P}_{X_i} = \bar{P}_{X_j}.$$

Hence, $X_i'\bar{P}_{X_j} = X_i'\bar{P}_{X_i} = 0$. Note that when $X_i = X_j$ then $C = I_K$ where K is the number of regressors.

10.3 *What Happens to Zellner's SUR Estimator when the Set of Regressors in One Equation Are Orthogonal to Those in the Second Equation?*

a. If X_1 and X_2 are orthogonal, then $X_1'X_2 = 0$. From (10.6) we get

$$\hat{\beta}_{GLS} = \begin{bmatrix} \sigma^{11}X_1'X_1 & 0 \\ 0 & \sigma^{22}X_2'X_2 \end{bmatrix}^{-1} \begin{bmatrix} \sigma^{11}X_1'y_1 + \sigma^{12}X_1'y_2 \\ \sigma^{21}X_2'y_1 + \sigma^{22}X_2'y_2 \end{bmatrix}$$

$$= \begin{bmatrix} \left(X_1'X_1\right)^{-1}/\sigma^{11} & 0 \\ 0 & \left(X_2'X_2\right)^{-1}/\sigma^{22} \end{bmatrix} \begin{bmatrix} \sigma^{11}X_1'y_1 + \sigma^{12}X_1'y_2 \\ \sigma^{21}X_2'y_1 + \sigma^{22}X_2'y_2 \end{bmatrix}$$

$$= \begin{bmatrix} \left(X_1'X_1\right)^{-1}X_1'y_1 + \sigma^{12}\left(X_1'X_1\right)^{-1}X_1'y_2/\sigma^{11} \\ \sigma^{21}\left(X_2'X_2\right)^{-1}X_2'y_1/\sigma^{22} + \left(X_2'X_2\right)^{-1}X_2'y_2 \end{bmatrix}$$

$$= \begin{bmatrix} \hat{\beta}_{1,ols} + (\sigma^{12}/\sigma^{11})\left(X_1'X_1\right)^{-1}X_1'y_2 \\ \hat{\beta}_{2,ols} + (\sigma^{21}/\sigma^{22})\left(X_2'X_2\right)^{-1}X_2'y_1 \end{bmatrix}$$

as required.

b. $\text{var}\left(\hat{\beta}_{GLS}\right) = (X'\Omega^{-1}X)^{-1} = \begin{bmatrix} \left(X_1'X_1\right)^{-1}/\sigma^{11} & 0 \\ 0 & \left(X_2'X_2\right)^{-1}/\sigma^{22} \end{bmatrix}$

If you have doubts, note that

$$\hat{\beta}_{1,GLS} = \beta_1 + \left(X_1'X_1\right)^{-1}X_1'u_1 + (\sigma^{12}/\sigma^{11})\left(X_1'X_1\right)^{-1}X_1'u_2$$

using $X_1'X_2 = 0$. Hence, $E(\hat{\beta}_{1,GLS}) = \beta_1$ and

$$\begin{aligned} \text{var}\left(\hat{\beta}_{1,GLS}\right) &= E\left(\hat{\beta}_{1,GLS} - \beta_1\right)\left(\hat{\beta}_{1,GLS} - \beta_1\right)' \\ &= \sigma_{11}\left(X_1'X_1\right)^{-1} + \sigma_{22}(\sigma^{12}/\sigma^{11})^2\left(X_1'X_1\right)^{-1} \\ &\quad + 2\sigma_{12}\sigma^{12}/\sigma^{11}\left(X_1'X_1\right)^{-1}. \end{aligned}$$

But, $\Sigma^{-1} = \begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix} / \left(\sigma_{11}\sigma_{22} - \sigma_{12}^2\right)$.

Hence, $\sigma^{12}/\sigma^{11} = -\sigma_{12}/\sigma_{22}$ substituting that in $\text{var}(\hat{\beta}_{1,GLS})$ we get

$$\begin{aligned} \text{var}\left(\hat{\beta}_{1,GLS}\right) &= \left(X_1'X_1\right)^{-1}\left[\sigma_{11} + \sigma_{22}\left(\sigma_{12}^2/\sigma_{22}^2\right) - 2\sigma_{12}^2/\sigma_{22}\right] \\ &= \left(X_1'X_1\right)^{-1}\left[\sigma_{11}\sigma_{22} + \sigma_{12}^2 - 2\sigma_{12}^2\right]/\sigma_{22} = \left(X_1'X_1\right)^{-1}/\sigma^{11} \end{aligned}$$

since $\sigma^{11} = \sigma_{22}/\left(\sigma_{11}\sigma_{22} - \sigma_{12}^2\right)$. Similarly, one can show that $\text{var}(\hat{\beta}_{2,GLS}) = \left(X_2'X_2\right)^{-1}/\sigma^{22}$.

c. We know that $\text{var}(\hat{\beta}_{1,\text{ols}}) = \sigma_{11}\left(X_1'X_1\right)^{-1}$ and $\text{var}(\hat{\beta}_{2,\text{ols}}) = \sigma_{22}\left(X_2'X_2\right)^{-1}$. If X_1 and X_2 are single regressors, then $\left(X_i'X_i\right)$ are scalars for $i = 1, 2$. Hence,

$$\text{var}\left(\hat{\beta}_{1,\text{GLS}}\right)/\text{var}\left(\hat{\beta}_{1,\text{ols}}\right) = \frac{(1/\sigma^{11})}{\sigma_{11}} = 1/\sigma_{11}\sigma^{11} = \frac{\left(\sigma_{11}\sigma_{22} - \sigma_{22}^2\right)}{\sigma_{11}\sigma_{22}} = 1 - \rho^2$$

where $\rho^2 = \sigma_{12}^2/\sigma_{22}\sigma_{11}$. Similarly,

$$\text{var}\left(\hat{\beta}_{2,\text{GLS}}\right)/\text{var}\left(\hat{\beta}_{2,\text{ols}}\right) = \frac{(1/\sigma^{22})}{\sigma_{22}} = 1/\sigma_{22}\sigma^{22} = \frac{\left(\sigma_{11}\sigma_{22} - \sigma_{12}^2\right)}{\sigma_{22}\sigma_{11}} = 1 - \rho^2.$$

10.4 From (10.13), $\tilde{s}_{ij} = e_i'e_j/[T - K_i - K_j + \text{tr}(B)]$ for $i, j = 1, 2$. where $e_i = y_i - X_i\hat{\beta}_{i,\text{ols}}$ for $i = 1, 2$. This can be rewritten as $e_i = \bar{P}_{X_i}y_i = \bar{P}_{X_i}u_i$ for $i = 1, 2$. Hence,

$$E(e_i'e_j) = E\left(u_i'\bar{P}_{X_i}\bar{P}_{X_j}u_j\right) = E[\text{tr}\left(u_i'\bar{P}_{X_i}\bar{P}_{X_j}u_j\right)] = E[\text{tr}\left(u_ju_i'\bar{P}_{X_i}\bar{P}_{X_j}\right)]$$

$$= \text{tr}[E\left(u_ju_i'\right)\bar{P}_{X_i}\bar{P}_{X_j}] = \sigma_{ji}\text{tr}\left(\bar{P}_{X_i}\bar{P}_{X_j}\right).$$

But $\sigma_{ji} = \sigma_{ij}$ and $\text{tr}\left(\bar{P}_{X_i}\bar{P}_{X_j}\right) = \text{tr}\left(I_T - P_{X_j} - P_{X_i} + P_{X_i}P_{X_j}\right) = T - K_j - K_i + \text{tr}(B)$ where $B = P_{X_i}P_{X_j}$. Hence, $E(\tilde{s}_{ij}) = E\left(e_i'e_j\right)/\text{tr}(\bar{P}_{X_i}\bar{P}_{X_j}) = \sigma_{ij}$. This proves that $\tilde{s}_{ij}$ is unbiased for σ_{ij} for $i, j = 1, 2$.

10.5 *Relative Efficiency of OLS in the Case of Simple Regressions*. This is based on Kmenta (1986, pp. 641–643).

a. Using the results in Chap. 3, we know that for a simple regression $Y_i = \alpha + \beta X_i + u_i$ that $\hat{\beta}_{\text{ols}} = \sum_{i=1}^{n} x_iy_i \Big/ \sum_{i=1}^{n} x_i^2$ and $\text{var}\left(\hat{\beta}_{\text{ols}}\right) = \sigma^2 \Big/ \sum_{i=1}^{n} x_i^2$ where $x_i = X_i - \bar{X}$ and $\text{var}(u_i) = \sigma^2$. Hence, for the first equation in (10.15), we get $\text{var}\left(\hat{\beta}_{12,\text{ols}}\right) = \sigma_{11}/m_{x_1x_1}$ where $m_{x_1x_1} = \sum_{t=1}^{T}\left(X_{1t} - \bar{X}_1\right)^2$ and $\sigma_{11} = \text{var}(u_1)$.

Similarly, for the second equation in (10.15), we get $\text{var}\left(\hat{\beta}_{22,\text{ols}}\right) = \sigma_{22}/m_{x_2x_2}$ where $m_{x_2x_2} = \sum_{t=1}^{T}\left(X_{2t} - \bar{X}_2\right)^2$ and $\sigma_{22} = \text{var}(u_2)$.

b. In order to get rid of the constants in (10.15), one can premultiply (10.15) by $I_2 \otimes (I_T - \bar{J}_T)$ where $\bar{J}_T = J_T/T$ and $J_T = \iota_T\iota_T'$ with ι_T being a vector of ones. Recall, from the FWL Theorem in Chap. 7 that $I_T - \bar{J}_T$ is the orthogonal projection on the constant ι_T, see problem 7.2. This yields

$$\tilde{Y}_1 = \tilde{X}_1\beta_{12} + \tilde{u}_1$$

$$\tilde{Y}_2 = \tilde{X}_2\beta_{22} + \tilde{u}_2$$

where $\tilde{Y}_i = (I_T - \bar{J}_T)Y_i$, $\tilde{X}_i = (I_T - \bar{J}_T)X_i$ and $\tilde{u}_i = (I_T - \bar{J}_T)u_i$ for $i = 1, 2$. Note that $Y_i' = (Y_{i1}, .., Y_{iT}), X_i' = (X_{il}, .., X_{iT})$ and $u_i' = (u_{il}, .., u_{iT})$ for $i = 1, 2$. GLS on this system of two equations yields the same GLS estimators of β_{12} and β_{22} as in (10.15). Note that $\Omega = E\begin{pmatrix}\tilde{u}_1\\ \tilde{u}_2\end{pmatrix}(\tilde{u}_1', \tilde{u}_2') = \Sigma \otimes (I_T - \bar{J}_T)$ where $\Sigma = [\sigma_{ij}]$ for $i, j = 1, 2$. Also, for this transformed system

$$\tilde{X}'\Omega^{-1}\tilde{X} = \begin{bmatrix}\tilde{X}_1' & 0\\ 0 & \tilde{X}_2'\end{bmatrix}(\Sigma^{-1} \otimes (I_T - \bar{J}_T))\begin{bmatrix}\tilde{X}_1 & 0\\ 0 & \tilde{X}_2\end{bmatrix}$$
$$= \begin{bmatrix}\sigma^{11}\tilde{X}_1'\tilde{X}_1 & \sigma^{12}\tilde{X}_1'\tilde{X}_2\\ \sigma^{21}\tilde{X}_2'\tilde{X}_1 & \sigma^{22}\tilde{X}_2'\tilde{X}_2\end{bmatrix}$$

since $\Sigma^{-1} \otimes (I_T - \bar{J}_T)$ is the generalized inverse of $\Sigma \otimes (I_T - \bar{J}_T)$ and $(I_T - \bar{J}_T)\tilde{X}_i = \tilde{X}_i$ for $i = 1, 2$. But, X_i and X_j are Txl vectors, hence, $\tilde{X}_i'\tilde{X}_j = m_{x_ix_j} = \sum_{t=1}^{T}(X_{it} - \bar{X}_i)(X_{jt} - \bar{X}_j)$ for $i, j = 1, 2$. So

$$\tilde{X}'\Omega^{-1}\tilde{X} = \begin{bmatrix}\sigma^{11}m_{x_1x_1} & \sigma^{12}m_{x_1x_2}\\ \sigma^{21}m_{x_2x_1} & \sigma^{22}m_{x_2x_2}\end{bmatrix}.$$

$$\text{But}\quad \Sigma^{-1} = \begin{bmatrix}\sigma_{22} & -\sigma_{12}\\ -\sigma_{12} & \sigma_{11}\end{bmatrix}/(\sigma_{11}\sigma_{22} - \sigma_{12}^2).$$

Hence, $\operatorname{var}\begin{pmatrix}\hat{\beta}_{12,GLS}\\ \hat{\beta}_{22,GLS}\end{pmatrix} = \left(\tilde{X}'\Omega^{-1}\tilde{X}\right)^{-1} = (\sigma_{11}\sigma_{22} - \sigma_{12}^2)$

$$\begin{bmatrix}\sigma_{22}m_{x_1x_1} & -\sigma_{12}m_{x_1x_2}\\ -\sigma_{12}m_{x_2x_1} & \sigma_{11}m_{x_2x_2}\end{bmatrix}^{-1}.$$

Simple inversion of this 2×2 matrix yields

$$\operatorname{var}\left(\hat{\beta}_{12,GLS}\right) = \left(\sigma_{11}\sigma_{22} - \sigma_{12}^2\right)\sigma_{11}m_{x_2x_2}/\left[\sigma_{11}\sigma_{22}m_{x_1x_1}\ m_{x_2x_2} - \sigma_{12}^2m_{x_1x_2}^2\right].$$

where the denominator is the determinant of the matrix to be inverted. Also,

$$\mathrm{var}\left(\hat{\beta}_{22,\mathrm{GLS}}\right) = \left(\sigma_{11}\sigma_{22} - \sigma_{12}^2\right)\sigma_{22}m_{x_1x_1} / \left[\sigma_{11}\sigma_{22}m_{x_1x_1}m_{x_2x_2} - \sigma_{12}^2 m_{x_1x_2}^2\right].$$

c. Defining ρ = correlation $(u_1, u_2) = \sigma_{12}/(\sigma_{11}\sigma_{22})^{1/2}$ and r = sample correlation coefficient between X_1 and $X_2 = m_{x_1x_2}/(m_{x_2x_2}m_{x_1x_1})^{1/2}$, then $\mathrm{var}(\hat{\beta}_{12,\mathrm{GLS}})/\mathrm{var}(\hat{\beta}_{12,\mathrm{ols}})$

$$= \left(\sigma_{11}\sigma_{22} - \sigma_{12}^2\right) m_{x_1x_1}\ m_{x_2x_2} / \left[\sigma_{11}\sigma_{22}m_{x_2x_2}\ m_{x_1x_1}\ - \sigma_{12}^2 m_{x_1x_2}^2\right]$$

$$= \sigma_{11}\sigma_{22}(1-\rho^2)/\sigma_{11}\sigma_{22}(1-\rho^2 r^2) = (1-\rho^2)/(1-\rho^2 r^2)$$

similarly, $\mathrm{var}(\hat{\beta}_{22,\mathrm{GLS}})/\mathrm{var}(\hat{\beta}_{22,\mathrm{ols}})$

$$= \left(\sigma_{11}\sigma_{22} - \sigma_{12}^2\right) m_{x_1x_1}\ m_{x_2x_2} / \left[\sigma_{11}\sigma_{22}m_{x_2x_2}\ m_{x_1x_1}\ - \sigma_{12}^2 m_{x_1x_2}^2\right]$$

$$= \sigma_{11}\sigma_{22}(1-\rho^2)/\sigma_{11}\sigma_{22}(1-\rho^2 r^2) = (1-\rho^2)/(1-\rho^2 r^2).$$

d. Call the relative efficiency ratio $E = (1-\rho^2)/(1-\rho^2 r^2)$. Let $\theta = \rho^2$, then $E = (1-\theta)/(1-\theta r^2)$. So that

$$\partial E/\partial\theta = \frac{-(1-\theta r^2) + r^2(1-\theta)}{(1-\theta r^2)^2} = \frac{-1+\theta r^2 + r^2 - \theta r^2}{(1-\theta r^2)^2} = \frac{-(1-r^2)}{(1-\theta r^2)^2} \leq 0$$

since $r^2 \leq 1$. Hence, the relative efficiency ratio is a non-increasing function of $\theta = \rho^2$. Similarly, let $\lambda = r^2$, then $E = (1-\theta)/(1-\theta\lambda)$ and $\partial E/\partial\lambda = \theta(1-\theta)/(1-\theta\lambda)^2 \geq 0$ since $0 \leq \theta \leq 1$. Hence, the relative efficiency ratio is a non-decreasing function of $\lambda = r^2$. This relative efficiency can be computed for various values of ρ^2 and r^2 between 0 and 1 at 0.1 intervals. See Kmenta (1986, Table 12-1, p. 642).

10.6 *Relative Efficiency of OLS in the Case of Multiple Regressions*. This is based on Binkely and Nelson (1988). Using partitioned inverse formulas, see the solution to problem 7.7(c), we get that the first block of the inverted matrix in (10.17) is

$$\mathrm{var}\left(\hat{\beta}_{1,\mathrm{GLS}}\right) = A_{11} = \left[\sigma^{11}X_1'X_1 - \sigma^{12}X_1'X_2\left(\sigma^{22}X_2'X_2\right)^{-1}\sigma^{21}X_2'X_1\right]^{-1}.$$

But, $\Sigma^{-1} = \begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix} / \left(\sigma_{11}\sigma_{22} - \sigma_{12}^2\right).$

Divide both the matrix on the right hand side and the denominator by $\sigma_{11}\sigma_{22}$, we get

$$\Sigma^{-1} = \left[1/(1-\rho^2)\right] \begin{bmatrix} 1/\sigma_{11} & -\rho^2/\sigma_{12} \\ -\rho^2/\sigma_{21} & 1/\sigma_{22} \end{bmatrix}$$

where $\rho^2 = \sigma_{12}^2/\sigma_{11}\sigma_{22}$. Hence,

$$\text{var}\left(\hat{\beta}_{1,\text{GLS}}\right) = \left[\frac{1}{\sigma_{11}(1-\rho^2)} X_1'X_1 - \sigma_{22}\left(\frac{\rho^2}{\sigma_{12}}\right)^2 \left(X_1'P_{X_2}X_1\right)/(1-\rho^2)\right]^{-1}.$$

But $\rho^2\sigma_{22}/\sigma_{12}^2 = 1/\sigma_{11}$, hence

$$\begin{aligned} \text{var}\left(\hat{\beta}_{1,\text{GLS}}\right) &= \left[\frac{1}{\sigma_{11}(1-\rho^2)} X_1'X_1 - \frac{\rho^2}{\sigma_{11}(1-\rho^2)} X_1'P_{X_2}X_1\right]^{-1} \\ &= \sigma_{11}(1-\rho^2)\left[X_1'X_1 - \rho^2 X_1'P_{X_2}X_1\right]^{-1}. \end{aligned}$$

Add and subtract $\rho^2 X_1'X_1$ from the expression to be inverted, one gets

$$\text{var}\left(\hat{\beta}_{1,\text{GLS}}\right) = \sigma_{11}(1-\rho^2)\left[(1-\rho^2)X_1'X_1 + \rho^2 X_1'\bar{P}_{X_2}X_1\right]^{-1}.$$

Factor out $(1-\rho^2)$ in the matrix to be inverted, one gets

$$\text{var}\left(\hat{\beta}_{1,\text{GLS}}\right) = \sigma_{11}\left\{X_1'X_1 + [\rho^2/(1-\rho^2)]E'E\right\}^{-1}$$

where $E = \bar{P}_{X_2}X_1$ is the matrix whose columns are the OLS residuals of each variable in X_1 regressed on X_2.

10.7 When X_1 and X_2 are orthogonal, then $X_1'X_2 = 0$. R_q^2 is the R^2 of the regression of variable X_q on the other $(K_1 - 1)$ regressors in X_1. R_q^{*2} is the R^2 of the regression of $\begin{bmatrix} X_q \\ \theta e_q \end{bmatrix}$ on the other $(K_1 - 1)$ regressors in $W = \begin{bmatrix} X_1 \\ \theta E \end{bmatrix}$. But $E = \bar{P}_{X_2} X_1 = X_1 - P_{X_2} X_1 = X_1$ since $P_{X_2} X_1 = X_2\left(X_2'X_2\right)^{-1} X_2'X_1 = 0$. So $e_q = X_q$ and $W = \begin{bmatrix} X_1 \\ \theta X_1 \end{bmatrix}$. Regressing $\begin{bmatrix} X_q \\ \theta X_q \end{bmatrix}$ on the other $(K_1 - 1)$ regressors in $\begin{bmatrix} X_1 \\ \theta X_1 \end{bmatrix}$ yields the same R^2 as that of X_q on the other $(K_1 - 1)$

regressors in X_1. Hence, $R_q^2 = R_q^{*2}$ and from (10.22) we get

$$\begin{aligned}
\text{var}\left(\hat{\beta}_{q,SUR}\right) &= \sigma_{11}/\left\{\sum_{t=1}^{T} x_{tq}^2(1-R_q^2) + \theta^2 \sum_{t=1}^{T} e_{tq}^2\left(1-R_q^2\right)\right\} \\
&= \sigma_{11}/\left\{\sum_{t=1}^{T} x_{tq}^2\left(1-R_q^2\right)\right\}(1+\theta^2) \\
&= \sigma_{11}(1-\rho^2)/\sum_{t=1}^{T} x_{tq}^2(1-R_q^2)
\end{aligned}$$

since $1+\theta^2 = 1/(1-\rho^2)$.

10.8 *SUR With Unequal Number of Observations*. This is based on Schmid (1977).

a. Ω given in (10.25) is block-diagonal. Therefore, Ω^{-1} is block-diagonal:

$$\Omega^{-1} = \begin{bmatrix} \sigma^{11}I_T & \sigma^{12}I_T & 0 \\ \sigma^{12}I_T & \sigma^{22}I_T & 0 \\ 0 & 0 & \frac{1}{\sigma_{22}}I_N \end{bmatrix} \text{ with } \Sigma^{-1} = [\sigma^{ij}] \text{ for } i,j = 1,2.$$

$$X = \begin{bmatrix} X_1 & 0 \\ 0 & X_2{}^* \\ 0 & X_2{}^0 \end{bmatrix} \quad \text{and} \quad y = \begin{pmatrix} y_1 \\ y_2{}^* \\ y_2{}^0 \end{pmatrix}$$

where X_1 is TxK_1, X_2^* is TxK_2 and X_2^o is NxK_2. Similarly, y_1 is $Tx1$, y_2^* is $Tx1$ and y_2^o is $Nx1$.

$$\begin{aligned}
X'\Omega^{-1} &= \begin{bmatrix} X_1' & 0 & 0 \\ 0 & X_2{}^{*\prime} & X_2{}^{0\prime} \end{bmatrix} \begin{bmatrix} \sigma^{11}I_T & \sigma^{12}I_T & 0 \\ \sigma^{12}I_T & \sigma^{22}I_T & 0 \\ 0 & 0 & \frac{1}{\sigma_{22}}I_N \end{bmatrix} \\
&= \begin{bmatrix} \sigma^{11}X_1' & \sigma^{12}X_1' & 0 \\ \sigma^{12}X_2^{*\prime} & \sigma^{22}X_2{}^{*\prime} & \frac{1}{\sigma_{22}}X_2{}^{0\prime} \end{bmatrix}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\hat{\beta}_{GLS} &= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y \\
&= \begin{bmatrix} \sigma^{11}X_1'X_1 & \sigma^{12}X_1'X_2{}^* \\ \sigma^{12}X_2^{*\prime}X_1 & \sigma^{22}X_2{}^{*\prime}X_2{}^* + (X_2{}^{0\prime}X_2{}^0/\sigma_{22}) \end{bmatrix}^{-1} \\
&\quad \begin{bmatrix} \sigma^{11}X_1'y_1 + \sigma^{12}X_1'y_2^* \\ \sigma^{12}X_2^{*\prime}y_1 + \sigma^{22}X_2^{*\prime}y_2^* + (X_2^{0\prime}y_2^0/\sigma_{22}) \end{bmatrix}
\end{aligned}$$

b. If $\sigma_{12} = 0$, then from (10.25)

$$\Omega = \begin{bmatrix} \sigma_{11}I_T & 0 & 0 \\ 0 & \sigma_{22}I_T & 0 \\ 0 & 0 & \sigma_{22}I_N \end{bmatrix} \quad \text{and} \quad \Omega^{-1} = \begin{bmatrix} \frac{1}{\sigma_{11}}I_T & 0 & 0 \\ 0 & \frac{1}{\sigma_{22}}I_T & 0 \\ 0 & 0 & \frac{1}{\sigma_{22}}I_N \end{bmatrix}$$

so that $\sigma^{12} = 0$ and $\sigma^{ii} = 1/\sigma_{ii}$ for $i = 1, 2$. From (10.26)

$$\hat{\beta}_{GLS} = \begin{bmatrix} \frac{X_1'X_1}{\sigma_{11}} & 0 \\ 0 & \left[\frac{X_2^{*\prime}X_2{}^*}{\sigma_{22}} + \frac{X_2{}^{0\prime}X_2{}^0}{\sigma_{22}}\right] \end{bmatrix}^{-1} \begin{bmatrix} (X_1'y_1/\sigma_{11}) \\ \left(X_2{}^{*\prime}y_2{}^* + X_2{}^{0\prime}y_2{}^0\right)/\sigma_{22} \end{bmatrix}$$

so that

$$\hat{\beta}_{GLS} = \begin{bmatrix} \left(X_1'X_1\right)^{-1}X_1'y_1 \\ \left[X_2^{*\prime}X_2{}^* + X_2{}^{0\prime}X_2{}^0\right]^{-1}\left[X_2^{*\prime}y_2{}^* + X_2{}^{0\prime}y_2{}^0\right] \end{bmatrix} = \begin{pmatrix} \hat{\beta}_{1,ols} \\ \hat{\beta}_{2,ols} \end{pmatrix}$$

Therefore, SUR with unequal number of observations reduces to OLS on each equation separately if $\sigma_{12} = 0$.

10.9 Grunfeld (1958) investment equation.

a. OLS on each firm.

Firm 1

Autoreg Procedure

Dependent Variable = INVEST

Ordinary Least Squares Estimates

SSE	275298.9	DFE	16
MSE	17206.18	Root MSE	131.1723
SBC	244.7953	AIC	241.962
Reg Rsq	0.8411	Total Rsq	0.8411
Durbin-Watson	1.3985		

Godfrey's Serial Correlation Test

Alternative	LM	Prob>LM
AR(+ 1)	2.6242	0.1052
AR(+ 2)	2.9592	0.2277
AR(+ 3)	3.8468	0.2785

Variable	DF	B Value	Std Error	t Ratio	Approx Prob
Intercept	1	-72.906480	154.0	-0.473	0.6423
VALUE1	1	0.101422	0.0371	2.733	0.0147
CAPITAL1	1	0.465852	0.0623	7.481	0.0001

Covariance of B-Values

	Intercept	VALUE1	CAPITAL1
Intercept	23715.391439	-5.465224899	0.9078247414
VALUE1	-5.465224899	0.0013768903	-0.000726424
CAPITAL1	0.9078247414	-0.000726424	0.0038773611

Firm 2

Autoreg Procedure

Dependent Variable = INVEST

Ordinary Least Squares Estimates

SSE	224688.6	DFE	16
MSE	14043.04	Root MSE	118.5033
SBC	240.9356	AIC	238.1023
Reg Rsq	0.1238	Total Rsq	0.1238
Durbin-Watson	1.1116		

Godfrey's Serial Correlation Test

Alternative	LM	Prob>LM
AR(+ 1)	4.6285	0.0314
AR(+ 2)	10.7095	0.0047
AR(+ 3)	10.8666	0.0125

Variable	DF	B Value	Std Error	t Ratio	Approx Prob
Intercept	1	306.273712	185.2	1.654	0.1177
VALUE1	1	0.015020	0.0913	0.165	0.8713
CAPITAL1	1	0.309876	0.2104	1.473	0.1602

Covariance of B-Values

	Intercept	VALUE1	CAPITAL1
Intercept	34309.129267	-15.87143322	-8.702731269
VALUE1	-15.87143322	0.0083279981	-0.001769902
CAPITAL1	-8.702731269	-0.001769902	0.0442679729

Firm 3

Autoreg Procedure

Dependent Variable = INVEST

Ordinary Least Squares Estimates

SSE	14390.83	DFE	16
MSE	899.4272	Root MSE	29.99045
SBC	188.7212	AIC	185.8879
Reg Rsq	0.6385	Total Rsq	0.6385
Durbin-Watson	1.2413		

Godfrey's Serial Correlation Test

Alternative	LM	Prob>LM
AR(+ 1)	2.7742	0.0958
AR(+ 2)	8.6189	0.0134
AR(+ 3)	11.2541	0.0104

Variable	DF	B Value	Std Error	t Ratio	Approx Prob
Intercept	1	-14.649578	39.6927	-0.369	0.7169
VALUE1	1	0.031508	0.0189	1.665	0.1154
CAPITAL1	1	0.162300	0.0311	5.213	0.0001

Covariance of B-Values

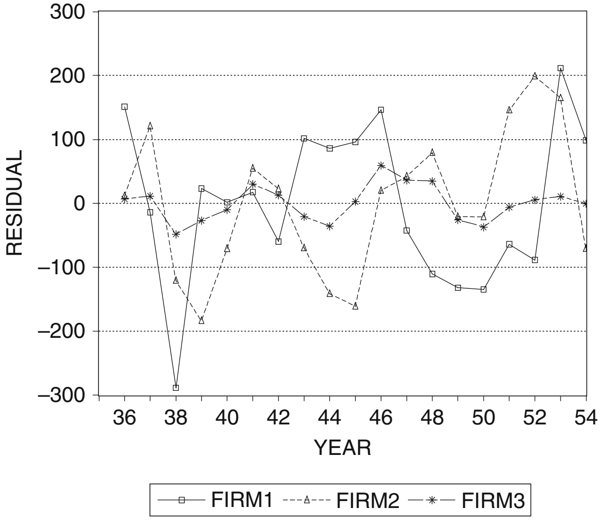
	Intercept	VALUE1	CAPITAL1
Intercept	1575.5078379	-0.706680779	-0.498664487
VALUE1	-0.706680779	0.0003581721	0.00007153
CAPITAL1	-0.498664487	0.00007153	0.0009691442

Plot of RESID*YEAR.

c. SUR model using the first 2 firms

Model: FIRM1

Dependent variable: FIRM1_I



Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter = 0	Prob > |T|
INTERCEP	1	-74.776254	153.974556	-0.486	0.6338
FIRM1_F1	1	0.101594	0.037101	2.738	0.0146
FIRM1_C1	1	0.467861	0.062263	7.514	0.0001

Model: FIRM2

Dependent variable: FIRM2_I

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter = 0	Prob >|T|
INTERCEP	1	309.978987	185.198638	1.674	0.1136
FIRM2_F1	1	0.012508	0.091244	0.137	0.8927
FIRM2_C1	1	0.314339	0.210382	1.494	0.1546

SUR model using the first 2 firms

SYSLIN Procedure

Seemingly Unrelated Regression Estimation

Cross Model Covariance

Sigma	FIRM1	FIRM2
FIRM1	17206.183816	-355.9331435
FIRM2	-355.9331435	14043.039401

Cross Model Correlation

Corr	FIRM1	FIRM2
FIRM1	1	-0.022897897
FIRM2	-0.022897897	1

Cross Model Inverse Correlation

Inv Corr	FIRM1	FIRM2
FIRM1	1.0005245887	0.0229099088
FIRM2	0.0229099088	1.0005245887

Cross Model Inverse Covariance

Inv Sigma	FIRM1	FIRM2
FIRM1	0.0000581491	1.4738406E - 6
FIRM2	1.4738406E - 6	0.000071247

System Weighted MSE: 0.99992 with 32 degrees of freedom. System Weighted R-Square: 0.7338

d. SUR model using the first three firms

Model: FIRM1

Dependent variable: FIRM1_I

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter = 0	Prob > |T|
INTERCEP	1	-27.616240	147.621696	-0.187	0.8540
FIRM1_F1	1	0.088732	0.035366	2.509	0.0233
FIRM1_C1	1	0.481536	0.061546	7.824	0.0001

Model: FIRM2

Dependent variable: FIRM2_I

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter = 0	Prob > |T|
INTERCEP	1	255.523339	167.011628	1.530	0.1456
FIRM2_F1	1	0.034710	0.081767	0.425	0.6769
FIRM2_C1	1	0.353757	0.195529	1.809	0.0892

Model: FIRM3

Dependent variable: FIRM3_I

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter = 0	Prob > |T|
INTERCEP	1	-27.024792	35.444666	-0.762	0.4569
FIRM3_F1	1	0.042147	0.016579	2.542	0.0217
FIRM3_C1	1	0.141415	0.029134	4.854	0.0002

SUR model using the first 3 firms

SYSLIN Procedure

Seemingly Unrelated Regression Estimation

Cross Model Covariance

Sigma	FIRM1	FIRM2	FIRM3
FIRM1	17206.183816	-355.9331435	1432.3838645
FIRM2	-355.9331435	14043.039401	1857.4410783
FIRM3	1432.3838645	1857.4410783	899.4271732

Cross Model Correlation

Corr	FIRM1	FIRM2	FIRM3
FIRM1	1	-0.022897897	0.3641112847
FIRM2	-0.022897897	1	0.5226386059
FIRM3	0.3641112847	0.5226386059	1

Cross Model Inverse Correlation

Inv Corr	FIRM1	FIRM2	FIRM3
FIRM1	1.2424073467	0.3644181288	-0.642833518
FIRM2	0.3644181288	1.4826915084	-0.907600576
FIRM3	-0.642833518	-0.907600576	1.7084100377

Cross Model Inverse Covariance

Inv Sigma	FIRM1	FIRM2	FIRM3
FIRM1	0.000072207	0.0000234438	-0.000163408
FIRM2	0.0000234438	0.000105582	-0.000255377
FIRM3	-0.000163408	-0.000255377	0.0018994423

System Weighted MSE: 0.95532 with 48 degrees of freedom.

System Weighted R-Square: 0.6685

SAS PROGRAM

```
data A; infile 'B:/DATA/grunfeld.dat';
input firm year invest value capital;

data aa; set A; keep firm year invest value capital;
if firm>3 then delete;

data A1; set aa; keep firm year invest value capital;
if firm>1 then delete;
data AA1; set A1;
value1=lag(value);
capital1=lag(capital);

Proc autoreg data=AA1;
    model invest=value1 capital1/ godfrey=3 covb;
    output out=E1 r=resid;
title 'Firm 1';
proc plot data=E1;
    plot resid*year='*';
Title 'Firm 1';
run;

******************************************************;

data A2; set aa; keep firm year invest value capital;
if firm=1 or firm=3 then delete;
```

```
data AA2; set A2;
value1=lag(value);
capital1=lag(capital);

Proc autoreg data=AA2;
   model invest=value1 capital1/ godfrey=3 covb;
   output out=E2 r=resid;
title 'Firm 2';
proc plot data=E2;
     plot resid*year='*';
Title 'Firm 2';
run;

data A3; set aa; keep firm year invest value capital;
if firm<=2 then delete;
data AA3; set A3;
value1=lag(value);
capital1=lag(capital);

Proc autoreg data=AA3;
   model invest=value1 capital1/ godfrey=3 covb;
   output out=E3 r=resid;
title 'Firm 3';
proc plot data=E3;
    plot resid*year='*';
title 'Firm 3';
run;

Proc iml;
use aa;
read all into temp;
sur=temp[1:20,3:5]||temp[21:40,3:5]||temp[41:60,3:5];
c={"F1_i" "F1_f" "F1_c" "F2_i" "F2_f" "F2_c" "F3_i" "F3_f" "F3_c"};
create sur_data from sur [colname=c];
append from sur;
```

```
data surdata; set sur_data;
firm1_i=f1_i; firm1_f1=lag(f1_f); firm1_c1=lag(f1_c);
firm2_i=f2_i; firm2_f1=lag(f2_f); firm2_c1=lag(f2_c);
firm3_i=f3_i; firm3_f1=lag(f3_f); firm3_c1=lag(f3_c);

proc syslin sur data=surdata;
   Firm1: model firm1_i=firm1_f1 firm1_c1;
   Firm2: model firm2_i=firm2_f1 firm2_c1;
title 'SUR model using the first 2 firms';
run;

proc syslin sur data=surdata;
   Firm1: model firm1_i=firm1_f1 firm1_c1;
   Firm2: model firm2_i=firm2_f1 firm2_c1;
   Firm3: model firm3_i=firm3_f1 firm3_c1;
title 'SUR model using the first 3 firms';
run;
```

10.11 Grunfeld (1958) Data-Unequal Observations. The SAS output is given below along with the program.

```
             Ignoring the extra OBS
          BETA                  STD_BETA

        48.910297              116.04249
        0.0834951              0.0257922
        0.2419409              0.0654934

        -58.84498              125.53869
        0.1804117              0.0629202
        0.3852054              0.1213539
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

                 WILKS METHOD
          BETA                  STD_BETA

        47.177574              116.24131
        0.0835601              0.0258332
         0.245064              0.0656327

        -58.93095              135.57919
        0.1803626              0.0679503
        0.3858253              0.1309581
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

```
SRIVASTAVA & ZAATAR METHOD
 BETA              STD_BETA

47.946223          116.04249
0.0835461          0.0257922
0.2435492          0.0654934

-59.66478          135.42233
0.1808292           0.067874
0.3851937          0.1309081
-----------------------------------------

HOCKING & SMITH METHOD
 BETA              STD_BETA

48.769108          116.12029
 0.083533          0.0258131
0.2419114          0.0655068

-60.39359          135.24547
0.1813054          0.0677879
 0.384481          0.1308514
```

SAS PROGRAM

```
data AA; infile 'a:/ch10/grunfeld.dat';
input firm year invest value capital;

data AAA; set AA;
keep firm year invest value capital;
if firm>=3 then delete;

Proc IML;
use AAA; read all into temp;
Y1=temp[1:15,3]; Y2=temp[21:40,3]; Y=Y1//Y2;
X1=temp[1:15,4:5]; X2=temp[21:40,4:5];
N1=15; N2=20; NN=5; K=3;

*---------- SCHMIDT'S FEASIBLE GLS ESTIMATORS ----------*;
        X1=J(N1,1,1)||X1;
        X2=J(N2,1,1)||X2;
        X=(X1//J(N2,K,0))||(J(N1,K,0)//X2);
        BT1=INV(X1`*X1)*X1`*Y1;
        BT2=INV(X2`*X2)*X2`*Y2;
```

```
e1=Y1-X1*BT1; e2=Y2-X2*BT2;
e2_15=e2[1:N1,]; ee=e2[16:N2,];
S11=e1`*e1/N1; S12=e1`*e2_15/N1;
S22_15=e2_15`*e2_15/N1;
S22_4=ee`*ee/NN; S22=e2`*e2/N2;
S_12=S12*SQRT(S22/S22_15);
S_11=S11 - (NN/N2)*((S12/S22_15)**(2))*(S22_15-S22_4);
S1_2=S12*S22/S22_15; ZERO=J(NN,2*N1,0);
OMG1=((S11||S12)//(S12||S22_15))@I(N1);
OMG2=((S11||S12)//(S12||S22))@I(N1);
OMG3=((S11||S_12)//(S_12||S22))@I(N1);
OMG4=((S_11||S1_2)//(S1_2||S22))@I(N1);
OMEGA1=(OMG1//ZERO)||(ZERO //(S22_15@I(NN)));
OMEGA2=(OMG2//ZERO)||(ZERO`//(S22@I(NN)));
OMEGA3=(OMG3//ZERO)||(ZERO`//(S22@I(NN)));
OMEGA4=(OMG4//ZERO)||(ZERO`//(S22@I(NN)));
OMG1_INV=INV(OMEGA1); OMG2_INV=INV(OMEGA2);
OMG3_INV=INV(OMEGA3); OMG4_INV=INV(OMEGA4);

********** Ignoring the extra OBS **********;
BETA1=INV(X`*OMG1_INV*X)*X`*OMG1_INV*Y;
VAR_BT1=INV(X`*OMG1_INV*X);
STD_BT1=SQRT(VECDIAG(VAR_BT1));
OUT1=BETA1||STD_BT1; C={"BETA" "STD_BETA"};
PRINT 'Ignoring the extra OBS',,OUT1(|COLNAME=C|);

********** WILKS METHOD **********;
BETA2=INV(X`*OMG2_INV*X)*X`*OMG2_INV*Y;
VAR_BT2=INV(X`*OMG2_INV*X);
STD_BT2=SQRT(VECDIAG(VAR_BT2));
OUT2=BETA2||STD_BT2;
PRINT 'WILKS METHOD',,OUT2(|COLNAME=C|);

********** SRIVASTAVA & ZAATAR METHOD **********;
BETA3=INV(X`*OMG3_INV*X)*X`*OMG3_INV*Y;
```

```
VAR_BT3=INV(X`*OMG3_INV*X);
STD_BT3=SQRT(VECDIAG(VAR_BT3));
OUT3=BETA3||STD_BT3;
PRINT 'SRIVASTAVA & ZAATAR METHOD',,OUT3(|COLNAME=C|);

********** HOCKING & SMITH METHOD **********;
BETA4=INV(X`*OMG4_INV*X)*X`*OMG4_INV*Y;
VAR_BT4=INV(X`*OMG4_INV*X);
STD_BT4=SQRT(VECDIAG(VAR_BT4));
OUT4=BETA4||STD_BT4;
PRINT 'HOCKING & SMITH METHOD',,OUT4(|COLNAME=C|);
```

10.12 Baltagi and Griffin (1996) Gasoline Data.

a. SUR Model with the first two countries

SYSLIN Procedure

Seemingly Unrelated Regression Estimation

Model: AUSTRIA

Dependent variable: AUS_Y

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter = 0	Prob > |T|
INTERCEP	1	3.713252	0.371877	9.985	0.0001
AUS_X1	1	0.721405	0.208790	3.455	0.0035
AUS_X2	1	-0.753844	0.146377	-5.150	0.0001
AUS_X3	1	-0.496348	0.111424	-4.455	0.0005

Model: BELGIUM

Dependent variable: BEL_Y

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter = 0	Prob > |T|
INTERCEP	1	2.843323	0.445235	6.386	0.0001
BEL_X1	1	0.835168	0.169508	4.927	0.0002
BEL_X2	1	-0.130828	0.153945	-0.850	0.4088
BEL_X3	1	-0.686411	0.092805	-7.396	0.0001

b. SUR Model with the first three countries

SYSLIN Procedure

Seemingly Unrelated Regression Estimation

Model: AUSTRIA

Dependent variable: AUS_Y

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter = 0	Prob > |T|
INTERCEP	1	3.842153	0.369141	10.408	0.0001
AUS_X1	1	0.819302	0.205608	3.985	0.0012
AUS_X2	1	-0.786415	0.145256	-5.414	0.0001
AUS_X3	1	-0.547701	0.109719	-4.992	0.0002

Model: BELGIUM

Dependent variable: BEL_Y

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter = 0	Prob > |T|
INTERCEP	1	2.910755	0.440417	6.609	0.0001
BEL_X1	1	0.887054	0.167952	5.282	0.0001
BEL_X2	1	-0.128480	0.151578	-0.848	0.4100
BEL_X3	1	-0.713870	0.091902	-7.768	0.0001

Model: CANADA

Dependent variable: CAN_Y

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter = 0	Prob > |T|
INTERCEP	1	3.001741	0.272684	11.008	0.0001
CAN_X1	1	0.410169	0.076193	5.383	0.0001
CAN_X2	1	-0.390490	0.086275	-4.526	0.0004
CAN_X3	1	-0.462567	0.070169	-6.592	0.0001

c. Cross Model Correlation

Corr	AUSTRIA	BELGIUM	CANADA
AUSTRIA	1	0.2232164952	-0.211908098
BELGIUM	0.2232164952	1	-0.226527683
CANADA	-0.211908098	-0.226527683	1

Breusch and Pagan (1980) diagonality LM test statistic for the first three countries yields $\lambda_{LM} = T\left(r_{12}^2 + r_{31}^2 + r_{32}^2\right) = 2.77$ which is asympotically distributed as χ_3^2 under the null hypothesis. This does not reject the diagonality of the variance–covariance matrix across the three countries.

SAS PROGRAM

```
Data gasoline;
Input COUNTRY $ YEAR Y X1 X2 X3;
CARDS;

Proc IML;
use GASOLINE;
read all into temp;
sur=temp [1:19,2:5] | | temp[20:38, 2:5] | | temp[39:57,2:5];
c={“AUS_Y” “AUS_X1” “AUS_X2” “AUS_X3” “BEL_Y” “BEL_X1” “BEL_X2”
“BEL_X3” “CAN_Y” “CAN_X1” “CAN_X2” “CAN_X3”};
create SUR_DATA from SUR [colname=c];
append from SUR;
proc syslin sur data=SUR_DATA;
     AUSTRIA: model AUS_Y=AUS_X1 AUS_X2 AUS_X3;
     BELGIUM: model BEL_Y=BEL_X1 BEL_X2 BEL_X3;
title ‘SUR MODEL WITH THE FIRST 2 COUNTRIES’;
proc syslin sur data=SUR_DATA;
     AUSTRIA: model AUS_Y=AUS_X1 AUS_X2 AUS_X3;
     BELGIUM: model BEL_Y=BEL_X1 BEL_X2 BEL_X3;
     CANADA: model CAN_Y=CAN_X1 CAN_X2 CAN_X3;
title ‘SUR MODEL WITH THE FIRST 3 COUNTRIES’;
run;
```

10.13 *Trace Minimization of Singular Systems with Cross-Equation Restrictions.* This is based on Baltagi (1993).

a. Note that $\sum_{i=1}^{3} y_{it} = 1$, implies that $\sum_{i=1}^{3} \bar{y}_i = 1$, where $\bar{y}_i = \sum_{t=1}^{T} y_{it}/T$, for $i = 1, 2, 3$. This means that $\sum_{t=1}^{T} x_t(y_{3t} - \bar{y}_3) = -\sum_{t=1}^{T} x_t(y_{2t} - \bar{y}_2) - \sum_{t=1}^{T} x_t(y_{1t} - \bar{y}_1)$ or $b_3 = -b_2 - b_1$. This shows that the b_i's satisfy the adding up restriction $\beta_1 + \beta_2 + \beta_3 = 0$.

b. Note also that $\beta_1 = \beta_2$ and $\beta_1 + \beta_2 + \beta_3 = 0$ imply $\beta_3 = -2\beta_1$. If we ignore the first equation, and impose $\beta_1 = \beta_2$, we get

$$\begin{bmatrix} y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} \iota & 0 & X \\ 0 & \iota & -2X \end{bmatrix} \begin{bmatrix} \alpha_2 \\ \alpha_3 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_2 \\ \varepsilon_3 \end{bmatrix}$$

which means that the OLS normal equations yield

$$T\hat{\alpha}_2 + \hat{\beta}_1 \sum_{t=1}^{T} x_t = \sum_{t=1}^{T} y_{2t}$$

$$T\hat{\alpha}_3 - 2\hat{\beta}_1 \sum_{t=1}^{T} x_t = \sum_{t=1}^{T} y_{3t}$$

$$\hat{\alpha}_2 \sum_{t=1}^{T} x_t - 2\hat{\alpha}_3 \sum_{t=1}^{T} x_t + 5\hat{\beta}_1 \sum_{t=1}^{T} x_t^2 = \sum_{t=1}^{T} x_t y_{2t} - 2\sum_{t=1}^{T} x_t y_{3t}.$$

Substituting the expressions for $\hat{\alpha}_2$ and $\hat{\alpha}_3$ from the first two equations into the third equation, we get $\hat{\beta}_1 = 0.2b_2 - 0.4b_3$. Using $b_1 + b_2 + b_3 = 0$ from part (a), one gets $\hat{\beta}_1 = 0.4b_1 + 0.6b_2$.

c. Similarly, deleting the second equation and imposing $\beta_1 = \beta_2$ one gets

$$\begin{bmatrix} y_1 \\ y_3 \end{bmatrix} = \begin{bmatrix} \iota & 0 & X \\ 0 & \iota & -2X \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_3 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_3 \end{bmatrix}.$$

Forming the OLS normal equations and solving for $\hat{\beta}_1$, one gets $\hat{\beta}_1 = 0.2b_1 - 0.4b_3$. Using $b_1 + b_2 + b_3 = 0$ gives $\hat{\beta}_1 = 0.6b_1 + 0.4b_2$.

d. Finally, deleting the third equation and imposing $\beta_1 = \beta_2$ one gets

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \iota & 0 & X \\ 0 & \iota & X \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}.$$

Forming the OLS normal equations and solving for $\hat{\beta}_1$ one gets $\hat{\beta}_1 = 0.5b_1 + 0.5b_2$.

Therefore, the estimate of β_1 is not invariant to the deleted equation. Also, this non-invariancy affects Zellner's SUR estimation if the restricted least squares residuals are used rather than the unrestricted least squares residuals in estimating the variance covariance matrix of the disturbances. An alternative solution is given by Im (1994).

10.17 The SUR results are replicated using Stata below:

```
. sureg (Growth: dly = yrt gov m2y inf swo dtot f_pcy d80 d90) (Inequality: gih  = yrt m2y
 civ mlg  mlgldc), corr
```

Seemingly unrelated regression

Equation	Obs	Parms	RMSE	"R-sq"	chi2	P
Growth	119	9	2.313764	0.4047	80.36	0.0000
Inequality	119	5	6.878804	0.4612	102.58	0.0000

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Growth						
yrt	-.0497042	.1546178	-0.32	0.748	-.3527496	.2533412
gov	-.0345058	.0354801	-0.97	0.331	-.1040455	.0350338
m2y	.0084999	.0163819	0.52	0.604	-.023608	.0406078
inf	-.0020648	.0013269	-1.56	0.120	-.0046655	.000536
swo	3.263209	.60405	5.40	0.000	2.079292	4.447125
dtot	17.74543	21.9798	0.81	0.419	-25.33419	60.82505
f_pcy	-1.038173	.4884378	-2.13	0.034	-1.995494	-.0808529
d80	-1.615472	.5090782	-3.17	0.002	-2.613247	-.6176976
d90	-3.339514	.6063639	-5.51	0.000	-4.527965	-2.151063
_cons	10.60415	3.471089	3.05	0.002	3.800944	17.40736

Inequality						
yrt	-1.000843	.3696902	-2.71	0.007	-1.725422	-.2762635
m2y	-.0570365	.0471514	-1.21	0.226	-.1494516	.0353785
civ	.0348434	.5533733	0.06	0.950	-1.049748	1.119435
mlg	.1684692	.0625023	2.70	0.007	.0459669	.2909715
mlgldc	.0344093	.0421904	0.82	0.415	-.0482823	.117101
_cons	33.96115	4.471626	7.59	0.000	25.19693	42.72538

Correlation matrix of residuals:

	Growth	Inequality
Growth	1.0000	
Inequality	0.0872	1.0000

Breusch–Pagan test of independence: chi2(1) = 0.905, Pr = 0.3415.

b. Note that the correlation among the residuals of the two equations is weak (0.0872) and the Breusch–Pagan test for diagonality of the variance–covariance matrix of the disturbances of the two equations is statistically insignificant, not rejecting zero correlation among the two equations.

References

Baltagi, B.H. (1993), "Trace Minimization of Singular Systems With Cross-Equation Restrictions," *Econometric Theory*, Problem 93.2.4, 9: 314–315.

Baltagi, B.H. and J.M. Griffin (1983), "Gasoline Demand in the OECD: An Application of Pooling and Testing Procedures," European Economic Review, 22: 117–137.

Binkley, J.K. and C.H. Nelson (1988), "A Note on the Efficiency of Seemingly Unrelated Regression," The American Statistician, 42: 137–139.

Breusch, T.S. and A.R. Pagan (1980), "The Lagrange Multiplier Test and Its Applications to Model Specification in Econometrics, Review of Economic Studies, 47: 239–253.

Grunfeld, Y. (1958), "The Determinants of Corporate Investment," unpublished Ph.D. dissertation (University of Chicago: Chicago, IL).

Im, Eric Iksoon (1994), "Trace Minimization of Singular Systems With Cross-Equation Restrictions," *Econometric Theory*, Solution 93.2.4, 10: 450.

Kmenta, J. (1986), Elements of Econometrics (Macmillan: New York).

Schmidt, P. (1977), "Estimation of Seemingly Unrelated Regressions With Unequal Numbers of Observations," Journal of Econometrics, 5: 365–377.

CHAPTER 11
Simultaneous Equations Model

11.1 *The Inconsistency of OLS.* The OLS estimator from Eq. (11.14) yields $\hat{\delta}_{ols} = \sum_{t=1}^{T} p_t q_t / \sum_{t=1}^{T} p_t{}^2$ where $p_t = P_t - \bar{P}$ and $q_t = Q_t - \bar{Q}$. Substituting $q_t = \delta p_t + (u_{2t} - \bar{u}_2)$ from (11.14) we get $\hat{\delta}_{ols} = \delta + \sum_{t=1}^{T} p_t(u_{2t} - \bar{u}_2) / \sum_{t=1}^{T} p_t^2$. Using (11.18), we get $\text{plim} \sum_{t=1}^{T} p_t(u_{2t} - \bar{u}_2)/T = (\sigma_{12} - \sigma_{22})/(\delta - \beta)$ where $\sigma_{ij} = \text{cov}(u_{it}, u_{jt})$ for $i, j = 1, 2$ and $t = 1, 2, .., T$. Using (11.20) we get

$$\text{plim}\, \hat{\delta}_{ols} = \delta + [(\sigma_{12} - \sigma_{22})/(\delta - \beta)]/[(\sigma_{11} + \sigma_{22} - 2\sigma_{12})/(\delta - \beta)^2]$$

$$= \delta + (\sigma_{12} - \sigma_{22})(\delta - \beta)/(\sigma_{11} + \sigma_{22} - 2\sigma_{12}).$$

11.2 *When Is the IV Estimator Consistent?*

a. For Eq. (11.30) $y_1 = \alpha_{12}y_2 + \alpha_{13}y_3 + \beta_{11}X_1 + \beta_{12}X_2 + u_1$. When we regress y_2 on X_1, X_2 and X_3 to get $y_2 = \hat{y}_2 + \hat{\nu}_2$, the residuals satisfy $\sum_{t=1}^{T} \hat{y}_{2t}\hat{\nu}_{2t} = 0$ and $\sum_{t=1}^{T} \hat{\nu}_{2t}X_{1t} = \sum_{t=1}^{T} \hat{\nu}_{2t}X_{2t} = \sum_{t=1}^{T} \hat{\nu}_{2t}X_{3t} = 0$.
Similarly, when we regress y_3 on X_1, X_2 and X_4 to get $y_3 = \hat{y}_3 + \hat{\nu}_3$, the residuals satisfy $\sum_{t=1}^{T} \hat{y}_{3t}\hat{\nu}_{3t} = 0$ and $\sum_{t=1}^{T} \hat{\nu}_{3t}X_{1t} = \sum_{t=1}^{T} \hat{\nu}_{3t}X_{2t} = \sum_{t=1}^{T} \hat{\nu}_{3t}X_{4t} = 0$. In the second stage regression $y_1 = \alpha_{12}\hat{y}_2 + \alpha_{13}\hat{y}_3 + \beta_{11}X_1 + \beta_{12}X_2 + \varepsilon_1$ where $\varepsilon_1 = \alpha_{12}(y_2 - \hat{y}_2) + \alpha_{13}(y_3 - \hat{y}_3) + u_1 = \alpha_{12}\hat{\nu}_2 + \alpha_{13}\hat{\nu}_3 + u_1$. For this to yield consistent estimates, we need $\sum_{t=1}^{T} \hat{y}_{2t}\varepsilon_{1t} = \sum_{t=1}^{T} \hat{y}_{2t}u_{1t}$; $\sum_{t=1}^{T} \hat{y}_{3t}\varepsilon_{1t} = \sum_{t=1}^{T} \hat{y}_{3t}u_{1t}$; $\sum_{t=1}^{T} X_{1t}\varepsilon_{1t} = \sum_{t=1}^{T} X_{1t}u_{1t}$ and $\sum_{t=1}^{T} X_{2t}\varepsilon_{1t} = \sum_{t=1}^{T} X_{2t}u_{1t}$. But $\sum_{t=1}^{T} \hat{y}_{2t}\varepsilon_{1t} = \alpha_{12}\sum_{t=1}^{T} \hat{\nu}_{2t}\hat{y}_{2t} + \alpha_{13}\sum_{t=1}^{T} \hat{\nu}_{3t}\hat{y}_{2t} + \sum_{t=1}^{T} \hat{y}_{2t}u_{1t}$ with $\sum_{t=1}^{T} \hat{\nu}_{2t}\hat{y}_{2t} = 0$ because $\hat{\nu}_2$ are the residuals and $\hat{y}_2$ are the predicted values from the same regression. However, $\sum_{t=1}^{T} \hat{\nu}_{3t}\hat{y}_{2t}$ is not necessarily zero

B.H. Baltagi, *Solutions Manual for Econometrics*, Springer Texts in Business and Economics, DOI 10.1007/978-3-642-54548-1_11,

because $\hat{v}_3$ is only orthogonal to X_1, X_2 and X_4, while $\hat{y}_2$ is a perfect linear combination of X_1, X_2 and X_3. Hence, $\sum_{t=1}^{T} \hat{v}_{3t}X_{3t}$ is not necessarily zero, which makes $\sum_{t=1}^{T} \hat{v}_{3t}\hat{y}_{2t}$ not necessarily zero.

Similarly, $\sum_{t=1}^{T} \hat{y}_{3t}\varepsilon_{1t} = \alpha_{12}\sum_{t=1}^{T} \hat{v}_{2t}\hat{y}_{3t} + \alpha_{13}\sum_{t=1}^{T} \hat{v}_{3t}\hat{y}_{3t} + \sum_{t=1}^{T} \hat{y}_{3t}u_{1t}$ with $\sum_{t=1}^{T} \hat{v}_{3t}\hat{y}_{3t} = 0$ and $\sum_{t=1}^{T} \hat{v}_{2t}\hat{y}_{3t}$ not necessarily zero. This is because $\hat{v}_{2t}$ is only orthogonal to X_1, X_2 and X_3, while $\hat{y}_3$ is a perfect linear combination of X_1, X_2 and X_4. Hence, $\sum_{t=1}^{T} \hat{v}_{2t}X_{4t}$ is not necessarily zero, which makes $\sum_{t=1}^{T} \hat{v}_{2t}\hat{y}_{3t}$ not necessarily zero. Since both X_1 and X_2 are included in the first stage regressions, we have $\sum_{t=1}^{T} X_{1t}\varepsilon_{1t} = \sum_{t=1}^{T} X_{1t}u_{1t}$ using the fact that $\sum_{t=1}^{T} X_{1t}\hat{v}_{2t} = \sum_{t=1}^{T} X_{1t}\hat{v}_{3t} = 0$. Also, $\sum_{t=1}^{T} X_{2t}\varepsilon_{1t} = \sum_{t=1}^{T} X_{2t}u_{1t}$ since $\sum_{t=1}^{T} X_{2t}\hat{v}_{2t} = \sum_{t=1}^{T} X_{2t}\hat{v}_{3t} = 0$.

b. The first stage regressions regress y_2 and y_3 on X_2, X_3 and X_4 to get $y_2 = \hat{y}_2 + \hat{v}_2$ and $y_3 = \hat{y}_3 + \hat{v}_3$ with the residuals satisfying $\sum_{t=1}^{T} \hat{v}_{2t}\hat{y}_{2t} = \sum_{t=1}^{T} \hat{v}_{2t}X_{2t} = \sum_{t=1}^{T} \hat{v}_{2t}X_{3t} = \sum_{t=1}^{T} \hat{v}_{2t}X_{4t} = 0$ and $\sum_{t=1}^{T} \hat{v}_{3t}\hat{y}_{3t} = \sum_{t=1}^{T} \hat{v}_{3t}X_{2t} = \sum_{t=1}^{T} \hat{v}_{3t}X_{3t} = \sum_{t=1}^{T} \hat{v}_{3t}X_{4t} = 0$. In the second stage regression $y_1 = \alpha_{12}\hat{y}_2 + \alpha_{13}\hat{y}_3 + \beta_{11}X_1 + \beta_{12}X_2 + \varepsilon_1$ where $\varepsilon_1 = \alpha_{12}(y_2 - \hat{y}_2) + \alpha_{13}(y_3 - \hat{y}_3) + u_1 = \alpha_{12}\hat{v}_2 + \alpha_{13}\hat{v}_3 + u_1$. For this to yield consistent estimates, we need $\sum_{t=1}^{T} \hat{y}_{2t}\varepsilon_{1t} = \sum_{t=1}^{T} \hat{y}_{2t}u_{1t}$; $\sum_{t=1}^{T} \hat{y}_{3t}\varepsilon_{1t} = \sum_{t=1}^{T} \hat{y}_{3t}u_{1t}$; $\sum_{t=1}^{T} X_{1t}\varepsilon_{1t} = \sum_{t=1}^{T} X_{1t}u_{1t}$ and $\sum_{t=1}^{T} X_{2t}\varepsilon_{1t} = \sum_{t=1}^{T} X_{2t}u_{1t}$.

The first two equalities are satisfied because

$$\sum_{t=1}^{T} \hat{y}_{2t}\hat{v}_{2t} = \sum_{t=1}^{T} \hat{y}_{2t}\hat{v}_{3t} = 0 \qquad \text{and} \qquad \sum_{t=1}^{T} \hat{y}_{3t}\hat{v}_{2t} = \sum_{t=1}^{T} \hat{y}_{3t}\hat{v}_{3t} = 0$$

since the *same* set of X's are included in both first stage regressions. $\sum_{t=1}^{T} X_{2t}\varepsilon_{1t} = \sum_{t=1}^{T} X_{2t}u_{1t}$ because $\sum_{t=1}^{T} X_{2t}\hat{v}_{2t} = \sum_{t=1}^{T} X_{2t}\hat{v}_{3t} = 0$ since X_2 is included in both first-stage regressions. However, $\sum_{t=1}^{T} X_{1t}\varepsilon_{1t} \neq \sum_{t=1}^{T} X_{1t}u_{1t}$

since $\sum_{t=1}^{T} X_{1t}\hat{\nu}_{2t} \neq 0$ and $\sum_{t=1}^{T} X_{1t}\hat{\nu}_{3t} \neq 0$

because X_1 was not included in the first stage regressions.

11.3 *Just-Identification and Simple IV.* If Eq. (11.34) is just-identified, then X_2 is of the same dimension as Y_1, i.e., both are Txg_1. Hence, Z_1 is of the same dimension as X, both of dimension Tx $(g_1 + k_1)$. Therefore, $X'Z_1$ is a square non-singular matrix of dimension $(g_1 + k_1)$. Hence, $\left(Z_1'X\right)^{-1}$ exists. But from (11.36), we know that

$$\begin{aligned}\hat{\delta}_{1,2SLS} &= \left(Z_1'P_XZ_1\right)^{-1} Z_1'P_Xy_1 = \left[Z_1'X(X'X)^{-1}X'Z_1\right]^{-1} Z_1'X(X'X)^{-1}X'y_1 \\ &= (X'Z_1)^{-1}(X'X)\left(Z_1'X\right)^{-1}\left(Z_1'X\right)(X'X)^{-1}X'y_1 = (X'Z_1)^{-1}X'y_1\end{aligned}$$

as required. This is exactly the IV estimator with $W = X$ given in (11.41). It is important to emphasize that this is only feasible if $X'Z_1$ is square and non-singular.

11.4 *2SLS Can Be Obtained as GLS.* Premultiplying (11.34) by X' we get $X'y_1 = X'Z_1\delta_1 + X'u_1$ with $X'u_1$ having mean zero and $\text{var}(X'u_1) = \sigma_{11}X'X$ since $\text{var}(u_1) = \sigma_{11}I_T$. Performing GLS on this transformed equation yields

$$\hat{\delta}_{1,GLS} = [Z_1'X(\sigma_{11}X'X)^{-1}X'Z_1]^{-1}Z_1'X(\sigma_{11}X'X)^{-1}X'y_1 = (Z_1'P_XZ_1)^{-1}Z_1'P_Xy_1$$

as required.

11.5 *The Equivalence of 3SLS and 2SLS.*

a. From (11.46), (i) if Σ is diagonal then Σ^{-1} is diagonal and $\hat{\Sigma}^{-1} \otimes P_X$ is block-diagonal with the i-th block consisting of $P_X/\hat{\sigma}_{ii}$. Also, Z is block-diagonal, therefore, $\{Z'[\hat{\Sigma}^{-1} \otimes P_X]Z\}^{-1}$ is block-diagonal with the i-th block consisting of $\hat{\sigma}_{ii}\left(Z_i'P_XZ_i\right)^{-1}$. In other words,

$$Z'\left[\hat{\Sigma}^{-1}\otimes P_X\right]Z = \begin{pmatrix} Z_1' & 0 & .. & 0 \\ 0 & Z_2' & .. & 0 \\ : & : & & : \\ 0 & 0 & .. & Z_G' \end{pmatrix} \begin{bmatrix} P_X/\hat{\sigma}_{11} & 0 & .. & 0 \\ 0 & P_X/\hat{\sigma}_{22} & .. & 0 \\ : & : & & : \\ 0 & 0 & .. & P_X/\hat{\sigma}_{GG} \end{bmatrix}$$

$$\times \begin{pmatrix} Z_1 & 0 & .. & 0 \\ 0 & Z_2 & .. & 0 \\ : & : & & : \\ 0 & 0 & .. & Z_G \end{pmatrix}$$

$$= \begin{pmatrix} Z_1'P_XZ_1/\hat{\sigma}_{11} & 0 & .. & 0 \\ 0 & Z_2'P_XZ_2/\hat{\sigma}_{22} & .. & 0 \\ : & : & & : \\ 0 & 0 & .. & Z_G'P_XZ_G/\hat{\sigma}_{GG} \end{pmatrix}$$

and $Z'\left[\hat{\Sigma}^{-1}\otimes P_X\right]y = \begin{pmatrix} Z_1'P_Xy_1/\hat{\sigma}_{11} \\ Z_2'P_Xy_2/\hat{\sigma}_{22} \\ : \\ Z_G'P_Xy_G/\hat{\sigma}_{GG} \end{pmatrix}$

Hence, from (11.46) $\hat{\delta}_{3SLS} = \{Z'[\hat{\Sigma}^{-1}\otimes P_X]Z\}^{-1}\{Z'[\hat{\Sigma}^{-1}\otimes P_X]y\}$

$$= \begin{pmatrix} \hat{\sigma}_{11}\left[Z_1'\,P_XZ_1\right]^{-1} & 0 & .. & 0 \\ 0 & \hat{\sigma}_{22}\left[Z_2'\,P_XZ_2\right]^{-1} & .. & 0 \\ : & : & .. & : \\ 0 & 0 & .. & \hat{\sigma}_{GG}\left[Z_G'P_XZ_G\right]^{-1} \end{pmatrix}$$

$$\times \begin{pmatrix} Z_1'P_Xy_1/\hat{\sigma}_{11} \\ Z_2'P_Xy_2/\hat{\sigma}_{22} \\ : \\ Z_G'P_Xy_G/\hat{\sigma}_{GG} \end{pmatrix}$$

$$= \begin{pmatrix} \left(Z_1'P_XZ_1\right)^{-1}Z_1'P_Xy_1 \\ \left(Z_2'P_XZ_2\right)^{-1}Z_2'P_Xy_2 \\ : \\ \left(Z_G'P_XZ_G\right)^{-1}Z_G'P_Xy_G \end{pmatrix} = \begin{pmatrix} \hat{\delta}_{1,2SLS} \\ \hat{\delta}_{2,2SLS} \\ : \\ \hat{\delta}_{G,2SLS} \end{pmatrix}$$

(ii) If every equation in the system is just-identified, then $Z_i'X$ is square and non-singular for $i = 1, 2, .., G$. In problem 11.3, we saw that for say the first equation, both Z_1 and X have the same dimension $Tx(g_1 + k_1)$. Hence, $Z_1'X$ is square and of dimension $(g_1 + k_1)$. This holds for every equation of the just-identified system. In fact, problem 11.3 proved that $\hat{\delta}_{i,2SLS} = (X'Z_i)^{-1}X'y_i$ for $i = 1, 2, .., G$. Also, from (11.44), we get

$$\hat{\delta}_{3SLS} = \{\text{diag}\left[Z_i'X\right]\left[\hat{\Sigma}^{-1} \otimes (X'X)^{-1}\right] \text{diag}\left[X'Z_i\right]\}^{-1}$$
$$\{\text{diag}\left[Z_i'X\right]\left[\hat{\Sigma}^{-1} \otimes (X'X)^{-1}\right](I_G \otimes X')y\}.$$

But each $Z_i{}'X$ is square and non-singular, hence

$$\hat{\delta}_{3SLS} = \left[\text{diag}\left(X'Z_i\right)^{-1}\left(\hat{\Sigma} \otimes (X'X)\right) \text{diag}\left(Z_i'X\right)^{-1}\right]$$
$$\{\text{diag}\left(Z_i'X\right)\left[\hat{\Sigma}^{-1} \otimes (X'X)^{-1}\right](I_G \otimes X')y\}.$$

which upon cancellation yields

$$\hat{\delta}_{3SLS} = \text{diag}(X'Z_i)^{-1}\,(I_G \otimes X')y$$
$$= \begin{pmatrix} (X'Z_1)^{-1} & 0 & .. & 0 \\ 0 & (X'Z_2)^{-1} & .. & 0 \\ \vdots & \vdots & :: & \vdots \\ 0 & 0 & .. & (X'Z_G)^{-1} \end{pmatrix}\begin{pmatrix} X'y_1 \\ X'y_2 \\ \vdots \\ X'y_G \end{pmatrix}$$
$$= \begin{pmatrix} (X'Z_1)^{-1}X'y_1 \\ (X'Z_2)^{-1}X'y_2 \\ \vdots \\ (X'Z_G)^{-1}X'y_G \end{pmatrix} = \begin{pmatrix} \hat{\delta}_{1,2SLS} \\ \hat{\delta}_{2,2SLS} \\ \vdots \\ \hat{\delta}_{G,2SLS} \end{pmatrix} \qquad \text{as required.}$$

b. Premultiplying (11.43) by $(I_G \otimes P_X)$ yields $y^* = Z^*\delta + u^*$ where $y^* = (I_G \otimes P_X)y$, $Z^* = (I_G \otimes P_X)Z$ and $u^* = (I_G \otimes P_X)u$. OLS on this transformed model yields

$$\hat{\delta}_{ols} = (Z^{*\prime}Z^*)^{-1}Z^{*\prime}y^* = [Z'(I_G \otimes P_X)Z]^{-1}\,[Z'(I_G \otimes P_X)y]$$
$$= \begin{pmatrix} \left(Z_1'P_XZ_1\right)^{-1} & 0 & .. & 0 \\ 0 & \left(Z_2'P_XZ_2\right)^{-1} & .. & 0 \\ \vdots & \vdots & :: & \vdots \\ 0 & 0 & .. & \left(Z_G'P_XZ_G\right)^{-1} \end{pmatrix}\begin{pmatrix} Z_1'P_Xy_1 \\ Z_2'P_Xy_2 \\ \vdots \\ Z_G'P_Xy_G \end{pmatrix}$$
$$= \begin{pmatrix} \left(Z_1'P_XZ_1\right)^{-1} Z_1'P_Xy_1 \\ \left(Z_2'P_XZ_2\right)^{-1} Z_2'P_Xy_2 \\ \vdots \\ \left(Z_G'P_XZ_G\right)^{-1} Z_G'P_Xy_G \end{pmatrix} = \begin{pmatrix} \hat{\delta}_{1,2SLS} \\ \hat{\delta}_{2,2SLS} \\ \vdots \\ \hat{\delta}_{G,2SLS} \end{pmatrix}$$

which is the 2SLS estimator of each equation in (11.43). Note that u^* has mean zero and $\mathrm{var}(u^*) = \Sigma \otimes P_X$ since $\mathrm{var}(u) = \Sigma \otimes I_T$. The generalized inverse of $\mathrm{var}(u^*)$ is $\Sigma^{-1} \otimes P_X$. Hence, GLS on this transformed model yields

$$\begin{aligned}\hat{\delta}_{GLS} &= [Z^{*\prime}(\Sigma^{-1} \otimes P_X)Z^*]^{-1}Z^{*\prime}(\Sigma^{-1} \otimes P_X)y^* \\ &= [Z'(I_G \otimes P_X)(\Sigma^{-1} \otimes P_X)(I_G \otimes P_X)Z]^{-1}Z'(I_G \otimes P_X)(\Sigma^{-1} \otimes P_X) \\ &\quad \times (I_G \otimes P_X)y \\ &= [Z'(\Sigma^{-1} \otimes P_X)Z]^{-1}Z'[\Sigma^{-1} \otimes P_X]y.\end{aligned}$$

Using the Milliken and Albohali condition for the equivalence of OLS and GLS given in Eq. (9.7) of Chap. 9, we can obtain a necessary and sufficient condition for 2SLS to be equivalent to 3SLS. After all, the last two estimators are respectively OLS and GLS on the transformed * system. This condition gives $Z^{*\prime}(\Sigma^{-1} \otimes P_X)\bar{P}_{Z^*} = 0$ since Z^* is the matrix of regressors and $\Sigma \otimes P_X$ is the corresponding variance-covariance matrix. Note that $Z^* = \mathrm{diag}[P_X Z_i] = \mathrm{diag}[\hat{Z}_i]$ or a block-diagonal matrix with the i-th block being the matrix of regressors of the second-stage regression of 2SLS. Also, $P_{Z^*} = \mathrm{diag}[P_{\hat{Z}_i}]$ and $\bar{P}_{Z^*} = \mathrm{diag}[\bar{P}_{\hat{Z}_i}]$. Hence, the above condition reduces to

$$\sigma^{ij}\hat{Z}_i'\bar{P}_{\hat{Z}_j} = 0 \text{ for } i \neq j \qquad \text{and} \quad i, j = 1, 2, .., G.$$

If Σ is diagonal, this automatically satisfies this condition since $\sigma_{ij} = \sigma^{ij} = 0$ for $i \neq j$. Also, if each equation in the system is just-identified, then $Z_i'X$ and $X'Z_j$ are square and non-singular. Hence,

$$\begin{aligned}\hat{Z}_i'P_{\hat{Z}_j} &= \hat{Z}_i'\hat{Z}_j\left(\hat{Z}_j'\hat{Z}_j\right)^{-1}\hat{Z}_j' = Z_i'X(X'X)^{-1}X'Z_j\left[Z_j'X(X'X)^{-1}X'Z_j\right]^{-1} \\ &\quad Z_j'X(X'X)^{-1}X' = Z_i'X(X'X)^{-1}X' = Z_i'P_X = \hat{Z}_i'\end{aligned}$$

after some cancellations. Hence, $\hat{Z}_i'\bar{P}_{\hat{Z}_j} = \hat{Z}_i' - \hat{Z}_i'P_{\hat{Z}_j} = \hat{Z}_i' - \hat{Z}_i' = 0$, under just-identification of each equation in the system.

11.6 **a.** Writing this system of two equations in the matrix form given by (A.1) we get $B = \begin{bmatrix} 1 & b \\ 1 & -d \end{bmatrix}; \Gamma = \begin{bmatrix} -a & 0 & 0 \\ -c & -e & -f \end{bmatrix}; y_t = \begin{pmatrix} Q_t \\ P_t \end{pmatrix}$

$x_t' = [1, W_t, L_t]$ and $u_t' = (u_{1t}, u_{2t})$.

b. There are two zero restrictions on Γ. These restrictions state that W_t and L_t do not appear in the demand equation. Therefore, the first equation is over-identified by the order condition of identification. There are two excluded exogenous variables and only one right hand side included endogenous variable. However, the supply equation is not identified by the order condition of identification. There are no excluded exogenous variables, but there is one right hand side included endogenous variable.

c. The transformed matrix FB should satisfy the following normalization restrictions:

$$f_{11} + f_{12} = 1 \quad \text{and} \quad f_{21} + f_{22} = 1.$$

These are the same normalization restrictions given in (A.6). Also, $F\Gamma$ must satisfy the following zero restrictions:

$$f_{11}0 - f_{12}e = 0 \quad \text{and} \quad f_{11}0 - f_{12}f = 0.$$

If either e or f are different from zero, so that at least W_t or L_t appear in the supply equation, then $f_{12} = 0$. This means that $f_{11} + 0 = 1$ or $f_{11} = 1$, from the first normalization equation. Hence, the first row of F is indeed the first row of an identity matrix and the demand equation is identified. However, there are no zero restrictions on the second equation and in the absence of any additional restrictions $f_{21} \neq 0$ and the second row of F is not necessarily the second row of an identity matrix. Hence, the supply equation is not identified.

11.7 **a.** In this case,

$$B = \begin{bmatrix} 1 & b \\ 1 & -f \end{bmatrix}; \Gamma = \begin{bmatrix} -a & -c & -d & 0 & 0 \\ -e & 0 & 0 & -g & -h \end{bmatrix}; y_t = \begin{pmatrix} Q_t \\ P_t \end{pmatrix}$$

$x_t' = [1, Y_t, A_t, W_t, L_t]$ and $u_t' = (u_{1t}, u_{2t})$.

b. There are four zero restrictions on Γ. These restrictions state that Y_t and A_t are absent from the supply equation whereas W_t and L_t are absent from the demand equation. Therefore, both equations are over-identified. For each equation, there is one right hand sided included endogenous variable and two excluded exogenous variables.

c. The transformed matrix FB should satisfy the following normalization restrictions:

$$f_{11} + f_{12} = 1 \quad \text{and} \quad f_{21} + f_{22} = 1.$$

Also, $F\Gamma$ must satisfy the following zero restrictions:

$$-f_{21}c + f_{22}0 = 0 \qquad - f_{21}d + f_{22}0 = 0.$$

and

$$f_{11}0 - f_{12}g = 0 \qquad f_{11}0 - f_{12}h = 0.$$

If either c or d are different from zero, so that at least Y_t or A_t appear in the demand equation, then $f_{21} = 0$ and from the second normalization equation we deduce that $f_{22} = 1$. Hence, the second row of F is indeed the second row of an identity matrix and the supply equation is identified. Similarly, if either g or h are different from zero, so that at least W_t or L_t appear in the supply equation, then $f_{12} = 0$ and from the first normalization equation we deduce that $f_{11} = 1$. Hence, the first row of F is indeed the first row of an identity matrix and the demand equation is identified.

11.8 a. From example (A.1) and Eq. (A.5), we get

$$A = [B, \Gamma] = \begin{bmatrix} \beta_{11} & \beta_{12} & \gamma_{11} & \gamma_{12} & \gamma_{13} \\ \beta_{21} & \beta_{22} & \gamma_{21} & \gamma_{22} & \gamma_{23} \end{bmatrix} = \begin{bmatrix} 1 & b & -a & -c & 0 \\ 1 & -e & -d & 0 & -f \end{bmatrix}$$

Therefore, φ for the first equation consists of only one restriction, namely that W_t is not in that equation, or $\gamma_{13} = 0$. This makes $\phi' = (0, 0, 0, 0, 1)$, since

$\alpha_1'\phi = 0$ gives $\gamma_{13} = 0$. Therefore,

$$A\phi = \begin{pmatrix}\gamma_{13}\\ \gamma_{23}\end{pmatrix} = \begin{pmatrix}0\\ -f\end{pmatrix}.$$

This is of rank one as long as $f \neq 0$. Similarly, for ϕ the second equation consists of only one restriction, namely that Y_t is not in that equation, or $\gamma_{22} = 0$. This makes $\phi = (0,0,0,1,0)$, since $\alpha_1'\phi = 0$ gives $\gamma_{22} = 0$. Therefore, $A\phi = \begin{pmatrix}\gamma_{12}\\ \gamma_{22}\end{pmatrix} = \begin{pmatrix}-c\\ 0\end{pmatrix}$. This is of rank one as long as $c \neq 0$.

b. For the supply and demand model given by (A.3) and (A.4), the reduced form is given by

$$\Pi = \begin{bmatrix}\pi_{11} & \pi_{12} & \pi_{13}\\ \pi_{21} & \pi_{22} & \pi_{23}\end{bmatrix} = \begin{bmatrix}ea+bd & ec & bf\\ a-d & c & -f\end{bmatrix}/(e+b).$$

This can be easily verified by solving the two equations in (A.3) and (A.4) for P_t and Q_t in terms of the constant, Y_t and W_t.

From part (a) and Eq. (A.17), we can show that the parameters of the first structural equation can be obtained from $\alpha_1'[W,\phi] = 0$ where $\alpha_1' = (\beta_{11},\beta_{12},\gamma_{11},\gamma_{12},\gamma_{13})$ represents the parameters of the first structural equation. $W' = [\Pi, I_3]$ and $\phi' = (0,0,0,0,1)$ as seen in part (a).

$$\beta_{11}\pi_{11} + \beta_{12}\pi_{21} + \gamma_{11} = 0$$

$$\beta_{11}\pi_{12} + \beta_{12}\pi_{22} + \gamma_{12} = 0$$

$$\beta_{11}\pi_{13} + \beta_{12}\pi_{23} + \gamma_{13} = 0$$

$$\gamma_{13} = 0$$

If we normalize by setting $\beta_{11} = 1$, we get $\beta_{12} = -\pi_{13}/\pi_{23}$ also

$$\gamma_{12} = -\pi_{12} + \frac{\pi_{13}}{\pi_{23}}\pi_{22} = \frac{\pi_{22}\pi_{13} - \pi_{12}\pi_{23}}{\pi_{23}} \text{ and}$$

$$\gamma_{11} = -\pi_{11} + \frac{\pi_{13}}{\pi_{23}}\pi_{21} = \frac{\pi_{21}\pi_{13} - \pi_{11}\pi_{23}}{\pi_{23}}.$$

One can easily verify from the reduced form parameters that $\beta_{12} = -\pi_{13}/\pi_{23} = -bf/-f = b$.

Similarly, $\gamma_{12} = (\pi_{22}\pi_{13} - \pi_{12}\pi_{23})/\pi_{23} = (\text{cbf} + \text{ecf})/-\text{f}(\text{e} + \text{b}) = -\text{c}$.

Finally, $\gamma_{11} = (\pi_{21}\pi_{13} - \pi_{11}\pi_{23})/\pi_{23} = (\text{abf} + \text{aef})/-\text{f}(\text{e} + \text{b}) = -\text{a}$.

Similarly, for the second structural equation, we get,

$$(\beta_{21}, \beta_{22}, \gamma_{21}, \gamma_{22}, \gamma_{23}) \begin{bmatrix} \pi_{11} & \pi_{12} & \pi_{13} & 0 \\ \pi_{21} & \pi_{22} & \pi_{23} & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} = (0, 0, 0, 0, 0)$$

which can be rewritten as

$$\beta_{21}\pi_{11} + \beta_{22}\pi_{21} + \gamma_{21} = 0$$

$$\beta_{21}\pi_{12} + \beta_{22}\pi_{22} + \gamma_{22} = 0$$

$$\beta_{21}\pi_{13} + \beta_{22}\pi_{23} + \gamma_{23} = 0$$

$$\gamma_{22} = 0.$$

If we normalize by setting $\beta_{21} = 1$, we get $\beta_{22} = -\pi_{12}/\pi_{22}$ also

$$\gamma_{23} = -\pi_{13} + \frac{\pi_{12}}{\pi_{22}}\pi_{23} = \frac{\pi_{23}\pi_{12} - \pi_{13}\pi_{22}}{\pi_{22}} \text{ and}$$

$$\gamma_{21} = -\pi_{11} + \frac{\pi_{12}}{\pi_{22}}\pi_{21} = \frac{\pi_{12}\pi_{21} - \pi_{11}\pi_{22}}{\pi_{22}}.$$

One can easily verify from the reduced form parameters that

$\beta_{22} = -\pi_{12}/\pi_{22} = -\text{ec}/\text{c} = -\text{e}$. Similarly,

$\gamma_{23} = (\pi_{23}\pi_{13} - \pi_{13}\pi_{22})/\pi_{22} = (-\text{ecf} - \text{bdf})/\text{c}(\text{e} + \text{b}) = -\text{f}$. Finally,

$\gamma_{21} = (\pi_{12}\pi_{21} - \pi_{11}\pi_{22})/\pi_{22} = (-\text{dec} - \text{bdc})/\text{c}(\text{e} + \text{b}) = -\text{d}$.

11.9 For the simultaneous model in problem 11.6, the reduced form parameters are given

by $\pi = \begin{bmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{21} & \pi_{22} & \pi_{23} \end{bmatrix} = \begin{bmatrix} \text{ab} + \text{bc} & \text{be} & \text{bf} \\ \text{a} - \text{c} & -\text{e} & -\text{f} \end{bmatrix} / (\text{b} + \text{d})$

This can be easily verified by solving for P_t and Q_t in terms of the constant, W_t and L_t. From the solution to 11.6, we get

$$A = [B, \Gamma] = \begin{bmatrix} \beta_{11} & \beta_{12} & \gamma_{11} & \gamma_{12} & \gamma_{13} \\ \beta_{21} & \beta_{22} & \gamma_{21} & \gamma_{22} & \gamma_{23} \end{bmatrix} = \begin{bmatrix} 1 & b & -a & 0 & 0 \\ 1 & -d & -c & -e & -f \end{bmatrix}$$

For the first structural equation,

$$\phi = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

with $\alpha_1'\phi = 0$ yielding $\gamma_{12} = \gamma_{13} = 0$. Therefore, (A.17) reduces to

$$(\beta_{11}, \beta_{12}, \gamma_{11}, \gamma_{12}, \gamma_{13}) \begin{bmatrix} \pi_{11} & \pi_{12} & \pi_{13} & 0 & 0 \\ \pi_{21} & \pi_{22} & \pi_{23} & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix} = (0, 0, 0, 0, 0)$$

This can be rewritten as

$$\beta_{11}\pi_{11} + \beta_{12}\pi_{21} + \gamma_{11} = 0$$

$$\beta_{11}\pi_{12} + \beta_{12}\pi_{22} + \gamma_{12} = 0$$

$$\beta_{11}\pi_{13} + \beta_{12}\pi_{23} + \gamma_{13} = 0$$

$$\gamma_{12} = 0$$

$$\gamma_{13} = 0$$

If we normalize by setting $\beta_{11} = 1$, we get $\beta_{12} = -\pi_{13}/\pi_{23}$; $\beta_{12} = -\pi_{12}/\pi_{22}$. Also, $\gamma_{11} = -\pi_{11} + (\pi_{12}/\pi_{22})\pi_{21} = \dfrac{\pi_{12}\pi_{21} - \pi_{11}\pi_{22}}{\pi_{22}}$ and $\gamma_{11} = -\pi_{11} + (\pi_{13}/\pi_{23})\pi_{21} = \dfrac{\pi_{21}\pi_{13} - \pi_{11}\pi_{23}}{\pi_{23}}$ one can easily verify from the reduced form parameters that

$\beta_{12} = -\pi_{13}/\pi_{23} = -bf/-f = b$. Also

$\beta_{12} = -\pi_{12}/\pi_{22} = -be/-e = b$. Similarly,

$$\gamma_{11} = \frac{\pi_{12}\pi_{21} - \pi_{11}\pi_{22}}{\pi_{22}} = \frac{abe - bce + ead + ebc}{-e(b+d)} = \frac{a(be+ed)}{-(be+ed)} = -a$$

Also,

$$\gamma_{11} = \frac{\pi_{21}\pi_{13} - \pi_{11}\pi_{23}}{\pi_{23}} = \frac{abf - bcf + adf + bcf}{-f(b+d)} = \frac{a(bf+df)}{-(bf+df)} = -a$$

For the second structural equation there are no zero restrictions. Therefore, (A.15) reduces to

$$(\beta_{21}, \beta_{22}, \gamma_{21}, \gamma_{22}, \gamma_{23}) \begin{bmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{21} & \pi_{22} & \pi_{23} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = (0,0,0,0,0)$$

This can be rewritten as

$$\beta_{21}\pi_{11} + \beta_{22}\pi_{21} + \gamma_{21} = 0$$

$$\beta_{21}\pi_{12} + \beta_{22}\pi_{22} + \gamma_{22} = 0$$

$$\beta_{21}\pi_{13} + \beta_{22}\pi_{23} + \gamma_{23} = 0$$

Three equations in five unknowns. When we normalize by setting $\beta_{21} = 1$, we still get three equations in four unknowns for which there is no unique solution.

11.10 *Just-Identified Model.*

a. The *generalized instrumental variable* estimator for δ_1 based on W is given below (11.41)

$$\begin{aligned}
\hat{\delta}_{1,IV} &= \left(Z_1' P_W Z_1\right)^{-1} Z_1' P_W y_1 \\
&= \left[Z_1' W \left(W'W\right)^{-1} W' Z_1\right]^{-1} Z_1' W (W'W)^{-1} W' y_1 \\
&= (W'Z_1)^{-1} (W'W) \left(Z_1' W\right)^{-1} \left(Z_1' W\right) (W'W)^{-1} W^{-1} y_1 \\
&= (W'Z_1)^{-1} W' y_1
\end{aligned}$$

since $W'Z_1$ is square and non-singular under just-identification. This is exactly the expression for the *simple instrumental variable* estimator for δ_1 given in (11.38).

b. Let the minimized value of the criterion function be

$Q = (y_1 - Z_1\hat{\delta}_{1,IV})'P_W(y_1 - Z_1\hat{\delta}_{1,IV})$.

Substituting the expression for $\hat{\delta}_{1,IV}$ and expanding terms, we get

$$Q = y_1'P_W y_1 - y_1'W\left(Z_1'W\right)^{-1}Z_1'P_W y_1 - y_1'P_W Z_1(W'Z_1)^{-1}W'y_1$$
$$+ y_1'W\left(Z_1'W\right)^{-1}Z_1'P_W Z_1(W'Z_1)^{-1}W'y_1$$
$$Q = y_1'P_W y_1 - y_1'W\left(Z_1'W\right)^{-1}Z_1'W(W'W)^{-1}W'y_1$$
$$- y_1'W(W'W)^{-1}W'Z_1\left(W_1'Z_1\right)^{-1}W'y_1$$
$$+ y_1'W\left(Z_1'W\right)^{-1}Z_1'W(W'W)^{-1}W'Z_1(W'Z_1)^{-1}W'y_1$$
$$Q = y_1'P_W y_1 - y_1'P_W y_1 - y_1'P_W y_1 + y_1'P_W y_1 = 0$$

as required.

c. Let $\hat{Z}_1 = P_W Z_1$ be the set of second stage regressors obtained from regressing each variable in Z_1 on the set of instrumental variables W. For the just-identified case

$$P_{\hat{Z}_1} = P_{P_W Z_1} = \hat{Z}_1\left(\hat{Z}_1'\hat{Z}_1\right)^{-1}\hat{Z}_1' = P_W Z_1\left(Z_1'P_W Z_1\right)^{-1}Z_1'P_W$$
$$= W(W'W)^{-1}W'Z_1\left[Z_1'W(W'W)^{-1}W'Z_1\right]^{-1}Z_1^{-1}W(W'W)^{-1}W'$$
$$= W(W'W)^{-1}W'Z_1(W'Z_1)^{-1}(W'W)\left(Z_1'W\right)^{-1}Z_1'W(W'W)^{-1}W'$$
$$= W(W'W)^{-1}W' = P_W.$$

Hence, the residual sum of squares of the second stage regression for the just-identified model, yields $y_1'\bar{P}_{\hat{Z}_1}y_1 = y_1'\bar{P}_W y_1$ since $P_{\hat{Z}_1} = P_W$. This is exactly the residual sum of squares from regression y_1 on the matrix of instruments W.

11.11 *The More Valid Instruments the Better.* If the set of instruments W_1 is spanned by the space of the set of instruments W_2, then $P_{W_2}W_1 = W_1$. The corresponding two instrumental variables estimators are

$$\hat{\delta}_{1,W_i} = \left(Z_1'P_{W_i}Z_1\right)^{-1}Z_1'P_{W_i}y_1 \qquad \text{for } i = 1, 2$$

with asymptotic covariance matrices

σ_{11} plim $\left(Z_1' P_{W_i} Z_1/T\right)^{-1}$ for i = 1, 2.

The difference between the asymptotic covariance matrix of $\hat{\delta}_{1,W_1}$ and that of $\hat{\delta}_{1,W_2}$ is positive semi-definite if

$$\sigma_{11}\left[\text{plim}\frac{Z_1'P_{W_1}Z_1}{T}\right]^{-1} - \sigma_{11}\left[\text{plim}\frac{Z_1'P_{W_2}Z_1}{T}\right]^{-1}$$

is positive semi-definite. This holds, if $Z_1'P_{W_2}Z_1 - Z_1'P_{W_1}Z_1$ is positive semi-definite. To show this, we prove that $P_{W_2} - P_{W_1}$ is idempotent.

$$\begin{aligned}(P_{W_2} - P_{W_1})(P_{W_2} - P_{W_1}) &= P_{W_2} - P_{W_2}P_{W_1} - P_{W_1}P_{W_2} + P_{W_1}\\ &= P_{W_2} - P_{W_1} - P_{W_1} + P_{W_1} = P_{W_2} - P_{W_1}.\end{aligned}$$

This uses the fact that $P_{W_2}P_{W_1} = P_{W_2}W_1\left(W_1'W_1\right)^{-1}W_1' = W_1\left(W_1'W_1\right)^{-1} W_1 = P_{W_1}$, since $P_{W_2}W_1 = W_1$.

11.12 *Testing for Over-Identification*. The main point here is that W spans the same space as $[\hat{Z}_1, W^*]$ and that both are of full rank ℓ. In fact, W^* is a subset of instruments W, of dimension $(\ell - k_1 - g_1)$, that are linearly independent of $\hat{Z}_1 = P_W Z_1$.

a. Using instruments W, the second stage regression on $y_1 = Z_1\delta_1 + W^*\gamma + u_1$ regresses y_1 on $P_W[Z_1, W^*] = [\hat{Z}_1, W^*]$ since $\hat{Z}_1 = P_W Z_1$ and $W^* = P_W W^*$. But the matrix of instruments W is of the same dimension as the matrix of regressors $[\hat{Z}_1, W^*]$. Hence, this equation is just-identified, and from problem 11.10, we deduce that the residual sum of squares of this second stage regression is exactly the residual sum of squares obtained by regressing y_1 on the matrix of instruments W, i.e.,

$$\text{URSS}^* = y_1'\bar{P}_W y_1 = y_1'y_1 - y_1'P_W y_1.$$

b. Using instruments W, the second stage regression on $y_1 = Z_1\delta_1 + u_1$ regresses y_1 on $\hat{Z}_1 = P_W Z_1$. Hence, the $\text{RRSS}^* = y_1'\bar{P}_{\hat{Z}_1}y_1 = y_1'y_1 - y_1'P_{\hat{Z}_1}y_1$ where $\bar{P}_{\hat{Z}_1} = I - P_{\hat{Z}_1}$ and $P_{\hat{Z}_1} = \hat{Z}_1\left(\hat{Z}_1'\hat{Z}_1\right)^{-1}\hat{Z}_1' =$

$P_W Z_1 \left(Z_1' P_W Z_1\right)^{-1} Z_1' P_W$. Therefore, using the results in part (a) we get

$$RRSS^* - URSS^* = y_1' P_W y_1 - y_1' P_{\hat{Z}_1} y_1$$

as given in (11.49).

c. Hausman (1983) proposed regressing the 2SLS residuals $(y_1 - Z_1\tilde{\delta}_{1,2SLS})$ on the set of instruments W and obtaining nR_u^2 where R_u^2 is the uncentered R^2 of this regression. Note that the regression sum of squares yields

$$\begin{aligned}
&(y_1 - Z_1\tilde{\delta}_{1,2SLS})' P_W (y_1 - Z_1\tilde{\delta}_{1,2SLS}) \\
&= y_1' P_W y_1 - y_1' P_W Z_1 (Z_1' P_W Z_1)^{-1} Z_1' P_W y_1 - y_1' P_W Z_1 (Z_1' P_W Z_1)^{-1} Z_1' P_W y_1 \\
&\quad + y_1' P_W Z_1 (Z_1' P_W Z_1)^{-1} Z_1' P_W Z_1 \left(Z_1' P_W Z_1\right)^{-1} Z_1' P_W y_1 \\
&= y_1' P_W y_1 - y_1' P_W Z_1 \left(Z_1' P_W Z_1\right)^{-1} Z_1' P_W y_1 \\
&= y_1' P_W y_1 - y_1' P_{\hat{Z}_1} y_1 = RRSS^* - URSS^*
\end{aligned}$$

as given in part (b). The total sum of squares of this regression, uncentered, is the 2SLS residuals sum of squares given by

$$(y_1 - Z_1\tilde{\delta}_{1,2SLS})'(y_1 - Z_1\tilde{\delta}_{1,2SLS}) = T\tilde{\sigma}_{11}$$

where $\tilde{\sigma}_{11}$ is given in (11.50). Hence, the test statistic

$$\begin{aligned}
\frac{RRSS^* - URSS^*}{\tilde{\sigma}_{11}} &= \frac{\text{Regression Sum of Squares}}{\text{Uncentered Total Sum of Squares / T}} \\
&= T(\text{uncentered } R^2) = T\, R_u^2 \text{ as required.}
\end{aligned}$$

This is asymptotically distributed as χ^2 with $\ell - (g_1 + k_1)$ degrees of freedom. Large values of this test statistic reject the null hypothesis.

d. The GNR for the unrestricted model given in (11.47) computes the residuals from the restricted model under H_o; $\gamma = 0$. These are the 2SLS residuals $(y_1 - Z_1\tilde{\delta}_{1,2SLS})$ based on the set of instruments W. Next, one differentiates the model with respect to its parameters and evaluates these derivatives at the restricted estimates. For instrumental variables W one

has to premultiply the right hand side of the GNR by P_W, this yields $(y_1 - Z_1\tilde{\delta}_{1,2SLS}) = P_W Z_1 b_1 + P_W W^* b_2 +$ residuals.
But $\hat{Z}_1 = P_W Z_1$ and $P_W W^* = W^*$ since W^* is a subset of W. Hence, the GNR becomes $(y_1 - Z_1\tilde{\delta}_{1,2SLS}) = \hat{Z}_1 b_1 + W^* b_2 +$ residuals. However, $[\hat{Z}_1, W^*]$ spans the same space as W, see part (a). Hence, this GNR regresses 2SLS residuals on the matrix of instruments W and computes TR_u^2 where R_u^2 is the uncentered R^2 of this regression. This is Hausman's (1983) test statistic derived in part (c).

11.15 a. The artificial regression in (11.55) is given by

$$y_1 = Z_1\delta_1 + (Y_1 - \hat{Y}_1)\eta + \text{residuals}$$

This can be rewritten as $y_1 = Z_1\delta_1 + \bar{P}_W Y_1 \eta +$ residuals
where we made use of the fact that $\hat{Y}_1 = P_W Y_1$ with $Y_1 - \hat{Y}_1 = Y_1 - P_W Y_1 = \bar{P}_W Y_1$. Note that to residual out the matrix of regressors $\bar{P}_W Y_1$, we need

$$\bar{P}_{\bar{P}_W Y_1} = I - \bar{P}_W Y_1 (Y_1' \bar{P}_W Y_1)^{-1} Y_1' \bar{P}_W$$

so that, by the FWL Theorem, the OLS estimate of δ_1 can also be obtained from the following regression

$$\bar{P}_{\bar{P}_W Y_1} y_1 = \bar{P}_{\bar{P}_W Y_1} Z_1 \delta_1 + \text{residuals}.$$

Note that

$$\bar{P}_{\bar{P}_W Y_1} Z_1 = Z_1 - \bar{P}_W Y_1 (Y_1' \bar{P}_W Y_1)^{-1} Y_1' \bar{P}_W Z_1$$

with $\bar{P}_W Z_1 = [\bar{P}_W Y_1, 0]$ and $\bar{P}_W X_1 = 0$ since X_1 is part of the instruments in W. Hence, $\bar{P}_{\bar{P}_W Y_1} Z_1 = [Y_1, X_1] - [\bar{P}_W Y_1, 0] = [P_W Y_1, X_1] = [\hat{Y}_1, X_1] = P_W[Y_1, X_1] = P_W Z_1 = \hat{Z}_1$
In this case,

$$\hat{\delta}_{1,ols} = \left(Z_1' \bar{P}_{\bar{P}_W Y_1} Z_1\right)^{-1} Z_1' \bar{P}_{\bar{P}_W Y_1} y_1 = \left(\hat{Z}_1' \hat{Z}_1\right)^{-1} \hat{Z}_1' y_1$$
$$= \left(Z_1' P_W Z_1\right)^{-1} Z_1' P_W y_1 = \hat{\delta}_{1,IV}$$

as required.

b. The FWL Theorem also states that the residuals from the regressions in part (a) are the same. Hence, their residuals sum of squares are the same. The last regression computes an estimate of the var$(\hat{\delta}_{1,ols})$ as $\tilde{s}^2\left(Z_1'P_W Z_1\right)^{-1}$ where $\tilde{s}^2$ is the residual sum of squares divided by the degrees of freedom of the regression. In (11.55), this is the MSE of that regression, since it has the same residuals sum of squares. Note that when $\eta \neq 0$ in (11.55), IV estimation is necessary and $\tilde{s}_{11}$ underestimates σ_{11} and will have to be replaced by

$$(y_1 - Z_1\hat{\delta}_{1,IV})'(y_1 - Z_1\hat{\delta}_{1,IV})/T.$$

11.16 *Recursive Systems.*

a. The order condition of identification yields one excluded exogenous variable (x_3) from the first equation and no right hand side endogenous variable i.e., $k_2 = 1$ and $g_1 = 0$. Therefore, the first equation is over-identified with degree of over-identification equal to one. The second equation has two excluded exogenous variables (x_1 and x_2) and one right hand side endogenous variable (y_1) so that $k_2 = 2$ and $g_1 = 1$. Hence, the second equation is over-identified of degree one. The rank condition of identification can be based on

$$A = [B, \Gamma] = \begin{bmatrix} 1 & 0 & \gamma_{11} & \gamma_{12} & 0 \\ \beta_{21} & 1 & 0 & 0 & \gamma_{23} \end{bmatrix}$$

with $y_t' = (y_{1t}, y_{2t})$ and $x_t' = (x_{1t}, x_{2t}, x_{3t})$. For the first equation, the set of zero restrictions yield

$$\phi_1 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{so that } \alpha_1'\phi_1 = (\beta_{12}, \gamma_{13}) = 0$$

where α_1' is the first row of A. For the second equation, the set of zero restrictions yield

$$\phi_2 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \text{ so that } \alpha_2'\phi_2 = (\gamma_{21}, \gamma_{22}) = 0$$

where α_2' is the second row of A. Hence, for the first equation

$$A\phi_1 = \begin{bmatrix} 0 & 0 \\ 1 & \gamma_{23} \end{bmatrix}$$

which is of rank 1 in general. Hence, the first equation is identified by the rank condition of identification. Similarly, for the second equation

$$A\phi_2 = \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ 0 & 0 \end{bmatrix}$$

which is of rank 1 as long as either γ_{11} or γ_{12} are different from zero. This ensures that either x_1 or x_2 appear in the first equation to identify the second equation.

b. The first equation is already in reduced form format

$$y_{1t} = -\gamma_{11}x_{1t} - \gamma_{12}x_{2t} + u_{1t}$$

y_{1t} is a function of x_{1t}, x_{2t} and only u_{1t}, not u_{2t}. For the second equation, one can substitute for y_{1t} to get

$$\begin{aligned} y_{2t} &= -\gamma_{23}x_{3t} - \beta_{21}(-\gamma_{11}x_{1t} - \gamma_{12}x_{2t} + u_{1t}) + u_{2t} \\ &= \beta_{21}\gamma_{11}x_{1t} + \beta_{21}\gamma_{12}x_{2t} - \gamma_{23}x_{3t} + u_{2t} - \beta_{21}u_{1t} \end{aligned}$$

so that y_{2t} is a function of x_{1t}, x_{2t}, x_{3t} and the error is a linear combination of u_{1t} and u_{2t}.

c. OLS on the first structural equation is also OLS on the first reduced form equation with only exogenous variables on the right hand side. Since x_1 and x_2 are not correlated with u_{1t}, this yields consistent estimates for γ_{11}

and γ_{12}. OLS on the second equation regresses y_{2t} on y_{1t} and x_{3t}. Since x_{3t} is not correlated with u_{2t}, the only possible correlation is from y_{1t} and u_{2t}. However, from the first reduced form equation y_{1t} is only a function of exogenous variables and u_{1t}. Since u_{1t} and u_{2t} are not correlated because Σ is diagonal for a recursive system, OLS estimates of β_{21} and γ_{23} are consistent.

d. Assuming that $u_t \sim N(0, \Sigma)$ where $u_t' = (u_{1t}, u_{2t})$, this exercise shows that OLS is MLE for this recursive system. Conditional on the set of x's, the likelihood function is given by

$$L(B, \Gamma, \Sigma) = (2\pi)^{-T/2}|B|^T|\Sigma|^{-T/2} \exp\left(-\frac{1}{2}\sum_{t=1}^{T} u_t'\Sigma^{-1}u_t\right)$$

so that

$$\log L = -(T/2)\log 2\pi + T\log|B| - (T/2)\log|\Sigma| - \frac{1}{2}\sum_{t=1}^{T} u_t'\Sigma^{-1}u_t.$$

Since B is triangular, $|B| = 1$ so that $\log|B| = 0$. Therefore, the only place in logL where B and Γ parameters appear is in $\sum_{t=1}^{T} u_t'\Sigma^{-1}u_t$. Maximizing logL with respect to B and Γ is equivalent to minimizing $\sum_{t=1}^{T} u_t'\Sigma^{-1}u_t$ with respect to B and Γ. But Σ is diagonal, so Σ^{-1} is diagonal and

$$\sum_{t=1}^{T} u_t'\Sigma^{-1}u_t = \sum_{t=1}^{T} u_{1t}^2/\sigma_{11} + \sum_{t=1}^{T} u_{2t}^2/\sigma_{22}.$$

The partial derivatives of logL with respect to the coefficients of the first structural equation are simply the partial derivatives of $\sum_{t=1}^{T} u_{1t}^2/\sigma_{11}$ with respect to those coefficients. Setting these partial derivatives equal to zero yields the OLS estimates of the first structural equation. Similarly, the partial derivatives of logL with respects to the coefficients of the second structural equation are simply the derivatives of $\sum_{t=1}^{T} u_{2t}^2/\sigma_{22}$ with respect to those coefficients. Setting these partial derivatives equal to zero yields the OLS estimates of the second structural equation. Hence, MLE is equivalent to OLS for a recursive system.

11.17 *Hausman's Specification Test: 2SLS Versus 3SLS*. This is based on Baltagi (1989).

a. The two equations model considered is

$$y_1 = Z_1\delta_1 + u_1 \qquad \text{and} \qquad y_2 = Z_2\delta_2 + u_2 \tag{1}$$

where y_1 and y_2 are endogenous; $Z_1 = [y_2, x_1, x_2]$, $Z_2 = [y_1, x_3]$; and x_1, x_2 and x_3 are exogenous (the y's and x's are Tx1 vectors). As noted in the problem, $\tilde{\delta} = $ 2SLS, $\tilde{\tilde{\delta}} = $ 3SLS, and the corresponding residuals are denoted by $\tilde{u}$ and $\tilde{\tilde{u}}$. $\delta_1' = (\alpha, \beta_1, \beta_2)$ and $\delta_2' = (\gamma, \beta_3)$ with $\alpha\gamma \neq 1$, so that the system is complete and we can solve for the reduced form.

Let $X = [x_1, x_2, x_3]$ and $P_X = X(X'X)^{-1}X'$, the 3SLS equations for model (1) are given by

$$Z'\left(\tilde{\Sigma}^{-1} \otimes P_X\right) Z\tilde{\tilde{\delta}} = Z'\left(\tilde{\Sigma}^{-1} \otimes P_X\right) y \tag{2}$$

where $Z = \text{diag}[Z_i]$, $y' = \left(y_1', y_2'\right)$, and $\tilde{\Sigma}^{-1} = [\tilde{\sigma}^{ij}]$ is the inverse of the estimated variance-covariance matrix obtained from 2SLS residuals. Equation (2) can also be written as

$$Z'\left(\tilde{\Sigma}^{-1} \otimes P_X\right) \tilde{\tilde{u}} = 0, \tag{3}$$

or more explicitly as

$$\tilde{\sigma}^{11}Z_1'P_X\tilde{\tilde{u}}_1 + \tilde{\sigma}^{12}Z_1'P_X\tilde{\tilde{u}}_2 = 0 \tag{4}$$

$$\tilde{\sigma}^{12}Z_2'P_X\tilde{\tilde{u}}_1 + \tilde{\sigma}^{22}Z_2'P_X\tilde{\tilde{u}}_2 = 0 \tag{5}$$

Using the fact that $Z_1'X$ is a square non-singular matrix, one can premultiply (4) by $(X'X)\left(Z_1'X\right)^{-1}$ to get

$$\tilde{\sigma}^{11}X'\tilde{\tilde{u}}_1 + \tilde{\sigma}^{12}X'\tilde{\tilde{u}}_2 = 0. \tag{6}$$

b. Writing Eq. (5) and (6) in terms of $\tilde{\tilde{\delta}}_1$ and $\tilde{\tilde{\delta}}_2$, one gets

$$\tilde{\sigma}^{12}\left(Z_2'P_XZ_1\right)\tilde{\tilde{\delta}}_1 + \tilde{\sigma}^{22}\left(Z_2'P_XZ_2\right)\tilde{\tilde{\delta}}_2 = \tilde{\sigma}^{12}\left(Z_2'P_Xy_1\right) + \tilde{\sigma}^{22}\left(Z_2'P_Xy_2\right) \tag{7}$$

and

$$\tilde{\sigma}^{11}X'Z_1\tilde{\tilde{\delta}}_1 + \tilde{\sigma}^{12}X'Z_2\tilde{\tilde{\delta}}_2 = \tilde{\sigma}^{11}X'y_1 + \tilde{\sigma}^{12}X'y_2. \tag{8}$$

Premultiply (8) by $\tilde{\sigma}^{12}Z_2'X(X'X)^{-1}$ and subtract it from $\tilde{\sigma}^{11}$ times (7). This eliminates $\tilde{\tilde{\delta}}_1$, and solves for $\tilde{\tilde{\delta}}_2$:

$$\tilde{\tilde{\delta}}_2 = \left(Z_2'P_XZ_2\right)^{-1} Z_2'P_Xy_2 = \tilde{\delta}_2. \tag{9}$$

Therefore, the 3SLS estimator of the over-identified equation is equal to its 2SLS counterpart.

Substituting (9) into (8), and rearranging terms, one gets

$$\tilde{\sigma}^{11}X'Z_1\tilde{\tilde{\delta}}_1 = \tilde{\sigma}^{11}X'y_1 + \tilde{\sigma}^{12}X'\tilde{u}_2.$$

Premultiplying by $Z_1'X(X'X)^{-1}/\tilde{\sigma}^{11}$, and solving for $\tilde{\tilde{\delta}}_1$, one gets

$$\tilde{\tilde{\delta}}_1 = \left(Z_1'P_XZ_1\right)^{-1} Z_1'P_Xy_1 + \left(\tilde{\sigma}^{12}/\tilde{\sigma}^{11}\right)\left(Z_1'P_XZ_1\right)^{-1} Z_1'P_X\tilde{u}_2. \tag{10}$$

Using the fact that $\tilde{\sigma}^{12} = -\tilde{\sigma}_{12}/|\tilde{\Sigma}|$, and $\tilde{\sigma}^{11} = \tilde{\sigma}_{22}/|\tilde{\Sigma}|$, (10) becomes

$$\tilde{\tilde{\delta}}_1 = \tilde{\delta}_1 - (\tilde{\sigma}_{12}/\tilde{\sigma}_{22})\left(Z_1'P_XZ_1\right)^{-1} Z_1'P_X\tilde{u}_2. \tag{11}$$

Therefore, the 3SLS estimator of the just-identified equation differs from its 2SLS (or indirect least squares) counterpart by a linear combination of the 2SLS (or 3SLS) residuals of the over-identified equation; see Theil (1971).

c. A Hausman-type test based on $\tilde{\tilde{\delta}}_1$ and $\tilde{\delta}_1$ is given by

$$m = (\tilde{\delta}_1 - \tilde{\tilde{\delta}}_1)'\left[V\left(\tilde{\delta}_1\right) - V\left(\tilde{\tilde{\delta}}_1\right)\right]^{-1}\left(\tilde{\delta}_1 - \tilde{\tilde{\delta}}_1\right), \tag{12}$$

where $V(\tilde{\delta}_1) = \tilde{\sigma}_{11}\left(Z_1'P_XZ_1\right)^{-1}$

and $V(\tilde{\tilde{\delta}}_1) = (1/\tilde{\sigma}^{11})\left(Z_1'P_XZ_1\right)^{-1} + \left(\tilde{\sigma}_{12}^2/\tilde{\sigma}_{22}\right)\left(Z_1'P_XZ_1\right)^{-1}$
$\left(Z_1'P_XZ_2\right)\left(Z_2'P_XZ_2\right)^{-1}\left(Z_2'P_XZ_1\right)\left(Z_1'P_XZ_1\right)^{-1}.$

The latter term is obtained by using the partitioned inverse of $Z'(\tilde{\Sigma}^{-1} \otimes P_X)Z$. Using (11), the Hausman test statistic becomes

$$m = \left(\tilde{\sigma}_{12}^2/\tilde{\sigma}_{22}^2\right)\left(\tilde{u}_2'P_XZ_1\right)\left[\left(\tilde{\sigma}_{11} - 1/\tilde{\sigma}^{11}\right)\left(Z_1'P_XZ_1\right)\right.$$
$$\left. - \left(\tilde{\sigma}_{12}^2/\tilde{\sigma}_{22}\right)\left(Z_1'P_XZ_2\right)\left(Z_2'P_XZ_2\right)^{-1}\left(Z_2'P_XZ_1\right)\right]^{-1}\left(Z_1'P_X\tilde{u}_2\right).$$

However, $(\tilde{\sigma}_{11} - 1/\tilde{\sigma}^{11}) = \left(\tilde{\sigma}_{12}^2/\tilde{\sigma}_{22}\right)$; this enables us to write

$$m = \tilde{u}_2'\hat{Z}_1\left[\hat{Z}_1'\hat{Z}_1 - \hat{Z}_1'\hat{Z}_2\left(\hat{Z}_2'\hat{Z}_2\right)^{-1}\hat{Z}_2'\hat{Z}_1\right]^{-1}\hat{Z}_1'\tilde{u}_2/\tilde{\sigma}_{22}, \tag{13}$$

where $\hat{Z}_i = P_X Z_i$ is the matrix of second stage regressors of 2SLS, for $i = 1, 2$. This statistic is asymptotically distributed as χ_3^2, and can be given the following interpretation:

Claim: $m = TR^2$, where R^2 is the R-squared of the regression of $\tilde{u}_2$ on the set of second stage regressors of both equations, $\hat{Z}_1$ and $\hat{Z}_2$.

Proof. This result follows immediately using the fact that $\tilde{\sigma}_{22} = \tilde{u}_2'\tilde{u}_2/T$ = total sums of squares/T, and the fact that the regression sum of squares is given by

$$RSS = \tilde{u}_2'\left[\hat{Z}_1\ \hat{Z}_2\right]\begin{bmatrix}\hat{Z}_1'\hat{Z}_1 & \hat{Z}_1'\hat{Z}_2 \\ \hat{Z}_2'\hat{Z}_1 & \hat{Z}_2'\hat{Z}_2\end{bmatrix}^{-1}\begin{bmatrix}\hat{Z}_1' \\ \hat{Z}_2'\end{bmatrix}\tilde{u}_2, \tag{14}$$

$$= \tilde{u}_2'\hat{Z}_1\left[\hat{Z}_1'\hat{Z}_1 - \hat{Z}_1'\hat{Z}_2\left(\hat{Z}_2'\hat{Z}_2\right)^{-1}\hat{Z}_2'\hat{Z}_1\right]^{-1}\hat{Z}_1'\tilde{u}_2,$$

where the last equality follows from partitioned inverse and the fact that $\hat{Z}_2'\tilde{u}_2 = 0$.

11.18 a. Using the order condition of identification, the first equation has two excluded exogenous variables x_2 and x_3 and only one right hand sided included endogenous variable y_2. Hence,

$$k_2 = 2 > g_1 = 1$$

and the first equation is over-identified with degree of over-identification equal to one. Similarly, the second equation has one excluded exogenous variable x_1 and only one right hand side included endogenous variable y_1. Hence, $k_2 = 1 = g_1$ and the second equation is just-identified. Using the rank condition of identification

$$A = [B\ \Gamma] = \begin{bmatrix} 1 & -\beta_{12} & \gamma_{11} & 0 & 0 \\ -\beta_{21} & 1 & 0 & \gamma_{22} & \gamma_{23}\end{bmatrix}$$

with $y'_t = (y_{1t}, y_{2t})$ and $x'_t = (x_{1t}, x_{2t}, x_{3t})$. The first structural equation has the following zero restrictions:

$$\phi_1 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{so that } \alpha'_1\phi_1 = (\gamma_{12}, \gamma_{13}) = 0$$

where α'_1 is the first row of A. Similarly, the second structural equation has the following zero restriction:

$$\phi_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \quad \text{so that } \alpha'_2\phi_2 = \gamma_{21} = 0$$

where α'_2 is the second row of A. Hence, $A\phi_1 = \begin{bmatrix} 0 & 0 \\ \gamma_{22} & \gamma_{23} \end{bmatrix}$ which has rank one provided either γ_{22} or γ_{23} is different from zero. Similarly, $A\phi_2 = \begin{pmatrix} \gamma_{11} \\ 0 \end{pmatrix}$ which has rank one provided $\gamma_{11} \neq 0$. Hence, both equations are identified by the rank condition of identification.

b. For the first equation, the OLS normal equations are given by

$$(Z'_1Z_1)\begin{pmatrix} \hat{\beta}_{12} \\ \hat{\gamma}_{11} \end{pmatrix}_{ols} = Z'_1y_1$$

where $Z_1 = [y_2, x_1]$. Hence,

$$Z'_1Z_1 = \begin{bmatrix} y'_2y_2 & y'_2x_1 \\ x'_1y_2 & x'_1x_1 \end{bmatrix} = \begin{bmatrix} 8 & 10 \\ 10 & 20 \end{bmatrix}$$

$$Z'_1y_1 = \begin{pmatrix} y'_2y_1 \\ x'_1y_1 \end{pmatrix} = \begin{pmatrix} 4 \\ 5 \end{pmatrix}$$

obtained from the matrices of cross-products provided by problem 11.18. Hence, the OLS normal equations are

$$\begin{bmatrix} 8 & 10 \\ 10 & 20 \end{bmatrix}\begin{pmatrix} \hat{\beta}_{12} \\ \hat{\gamma}_{11} \end{pmatrix}_{ols} = \begin{pmatrix} 4 \\ 5 \end{pmatrix}$$

solving for the OLS estimators, we get

$$\begin{pmatrix}\hat{\beta}_{12}\\ \hat{\gamma}_{11}\end{pmatrix}_{\text{ols}} = \frac{1}{60}\begin{bmatrix}20 & -10\\ -10 & 8\end{bmatrix}\begin{pmatrix}4\\ 5\end{pmatrix} = \begin{pmatrix}0.5\\ 0\end{pmatrix}$$

OLS on the second structural equation can be deduced in a similar fashion.

c. For the first equation, the 2SLS normal equations are given by

$$(Z_1'P_XZ_1)\begin{pmatrix}\hat{\beta}_{12}\\ \hat{\gamma}_{11}\end{pmatrix}_{\text{2SLS}} = Z_1'P_Xy_1$$

where $Z_1 = [y_2, x_1]$ and $X = [x_1, x_2, x_3]$. Hence,

$$Z_1'X = \begin{bmatrix}y_2'x_1 & y_2'x_2 & y_2'x_3\\ x_1'x_1 & x_1'x_2 & x_1'x_3\end{bmatrix} = \begin{bmatrix}10 & 20 & 30\\ 20 & 0 & 0\end{bmatrix}$$

$$(X'X)^{-1} = \begin{bmatrix}1/20 & 0 & 0\\ 0 & 1/20 & 0\\ 0 & 0 & 1/10\end{bmatrix} \quad \text{and } X'y_1 = \begin{bmatrix}x_1'y_1\\ x_2'y_1\\ x_3'y_1\end{bmatrix} = \begin{pmatrix}5\\ 40\\ 20\end{pmatrix}$$

$$Z_1'X(X'X)^{-1} = \begin{bmatrix}0.5 & 1 & 3\\ 1 & 0 & 0\end{bmatrix} \quad \text{and}$$

$$Z_1'P_XZ_1 = \begin{bmatrix}0.5 & 1 & 3\\ 1 & 0 & 0\end{bmatrix}\begin{bmatrix}10 & 20\\ 20 & 0\\ 30 & 0\end{bmatrix} = \begin{bmatrix}115 & 10\\ 10 & 20\end{bmatrix}$$

with

$$Z_1'P_Xy_1 = \begin{bmatrix}0.5 & 1 & 3\\ 1 & 0 & 0\end{bmatrix}\begin{pmatrix}5\\ 40\\ 20\end{pmatrix} = \begin{pmatrix}102.5\\ 5\end{pmatrix}.$$

Therefore, the 2SLS normal equations are

$$\begin{bmatrix}115 & 10\\ 10 & 20\end{bmatrix}\begin{pmatrix}\hat{\beta}_{12}\\ \hat{\gamma}_{11}\end{pmatrix}_{\text{2SLS}} = \begin{pmatrix}102.5\\ 5\end{pmatrix}$$

solving for the 2SLS estimates, we get

$$\begin{pmatrix}\hat{\beta}_{12}\\ \hat{\gamma}_{11}\end{pmatrix}_{\text{2SLS}} = \frac{1}{2200}\begin{bmatrix}20 & -10\\ -10 & 115\end{bmatrix}\begin{pmatrix}102.5\\ 5\end{pmatrix} = \frac{1}{2200}\begin{pmatrix}2000\\ -250\end{pmatrix} = \begin{pmatrix}10/11\\ -9/44\end{pmatrix}$$

2SLS on the second structural equation can be deduced in a similar fashion.

d. The first equation is over-identified and the reduced form parameter estimates will give more than one solution to the structural form parameters.

However, the second equation is just-identified. Hence, 2SLS reduces to indirect least squares which in this case solves uniquely for the structural parameters from the reduced form parameters. ILS on the second equation yields

$$\begin{pmatrix}\hat{\beta}_{21}\\ \hat{\gamma}_{22}\\ \hat{\gamma}_{23}\end{pmatrix}_{ILS} = (X'Z_1)^{-1}X'y_2$$

where $Z_1 = [y_1, x_2, x_3]$ and $X = [x_1, x_2, x_3]$. Therefore,

$$X'Z^1 = \begin{bmatrix} x_1'y_1 & x_1'x_2 & x_1'x_3 \\ x_2'y_1 & x_2'x_2 & x_2'x_3 \\ x_3'y_1 & x_3'x_2 & x_3'x_3 \end{bmatrix} = \begin{bmatrix} 5 & 0 & 0 \\ 40 & 20 & 0 \\ 20 & 0 & 10 \end{bmatrix} \quad \text{and}$$

$$X'y_2 = \begin{pmatrix} x_1'y_2 \\ x_2'y_2 \\ x_3'y_2 \end{pmatrix} = \begin{pmatrix} 10 \\ 20 \\ 30 \end{pmatrix}$$

so that

$$\begin{pmatrix}\hat{\beta}_{21}\\ \hat{\gamma}_{22}\\ \hat{\gamma}_{23}\end{pmatrix}_{ILS} = \begin{bmatrix} 5 & 0 & 0 \\ 40 & 20 & 0 \\ 20 & 0 & 10 \end{bmatrix}^{-1} \begin{pmatrix} 10 \\ 20 \\ 30 \end{pmatrix}$$

Note that $X'Z_1$ for the first equation is of dimension 3×2 and is not even square.

11.19 Supply and Demand Equations for Traded Money. This is based on Laffer (1970).

b. OLS Estimation

SYSLIN Procedure

Ordinary Least Squares Estimation

Model: SUPPLY

Dependent variable: LNTM_P

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	0.28682	0.14341	152.350	0.0001
Error	18	0.01694	0.00094		
C Total	20	0.30377			

Root MSE	0.03068	R-Square	0.9442
Dep Mean	5.48051	Adj R-SQ	0.9380
C.V.	0.55982		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob > |T|
INTERCEP	1	3.661183	0.272547	13.433	0.0001
LNRM_P	1	0.553762	0.089883	6.161	0.0001
LNI	1	0.165917	0.012546	13.225	0.0001

OLS ESTIMATION OF MONEY DEMAND FUNCTION

SYSLIN Procedure

Ordinary Least Squares Estimation

Model: DEMAND

Dependent variable: LNTM_P

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	4	0.29889	0.07472	245.207	0.0001
Error	16	0.00488	0.00030		
C Total	20	0.30377			

Root MSE	0.01746	R-Square	0.9839
Dep Mean	5.48051	Adj R-SQ	0.9799
C.V.	0.31852		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob > |T|
INTERCEP	1	3.055597	0.647464	4.719	0.0002
LNY_P	1	0.618770	0.055422	11.165	0.0001
LNI	1	-0.015724	0.021951	-0.716	0.4841
LNS1	1	-0.305535	0.101837	-3.000	0.0085
LNS2	1	0.147360	0.202409	0.728	0.4771

c. 2SLS Estimation

SYSLIN Procedure

Two-Stage Least Squares Estimation

Model: SUPPLY

Dependent variable: LNTM_P

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	0.29742	0.14871	150.530	0.0001
Error	18	0.01778	0.00099		
C Total	20	0.30377			

Root MSE	0.03143	R-Square	0.9436
Dep Mean	5.48051	Adj R-SQ	0.9373
C.V.	0.57350		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob > |T|
INTERCEP	1	3.741721	0.280279	13.350	0.0001
LNRM_P	1	0.524645	0.092504	5.672	0.0001
LNI	1	0.177757	0.013347	13.318	0.0001

Test for Overidentifying Restrictions

Numerator:	0.001342	DF:	2	F Value:	1.4224
Denominator:	0.000944	DF:	16	Prob>F:	0.2700

2SLS ESTIMATION OF MONEY DEMAND FUNCTION

SYSLIN Procedure

Two-Stage Least Squares Estimation

Model: DEMAND

Dependent variable: LNTM_P

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	4	0.30010	0.07503	154.837	0.0001
Error	16	0.00775	0.00048		
C Total	20	0.30377			

Root MSE	0.02201	R-Square	0.9748
Dep Mean	5.48051	Adj R-SQ	0.9685
C.V.	0.40165		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob > |T|
INTERCEP	1	1.559954	1.222958	1.276	0.2203
LNY_P	1	0.761774	0.111639	6.824	0.0001
LNI	1	-0.083170	0.049518	-1.680	0.1125
LNS1	1	-0.187052	0.147285	-1.270	0.2222
LNS2	1	0.228561	0.259976	0.879	0.3923

The total number of instruments equals the number of parameters in the equation. The equation is just identified, and the test for over identification is not computed.

d. 3SLS Estimation

SYSLIN Procedure
Three-Stage Least Squares Estimation

Cross Model Covariance

Sigma	SUPPLY	DEMAND
SUPPLY	0.0009879044	-0.000199017
DEMAND	-0.000199017	0.000484546

Cross Model Covariance

Corr	SUPPLY	DEMAND
SUPPLY	1	-0.287650003
DEMAND	-0.287650003	1

Cross Model Inverse Correlation

Inv Corr	SUPPLY	DEMAND
SUPPLY	1.0902064321	0.3135978834
DEMAND	0.3135978834	1.0902064321

Cross Model Inverse Covariance

Inv Sigma	SUPPLY	DEMAND
SUPPLY	1103.5545963	453.26080422
DEMAND	453.26080422	2249.9545257

System Weighted MSE: 0.53916 with 34 degrees of freedom.

System Weighted R-Square: 0.9858

Model: SUPPLY

Dependent variable: LNTM_P

3SLS ESTIMATION OF MODEY SUPPLY AND DEMAND FUNCTION

SYSLIN Procedure
Three-Stage Least Squares Estimation

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob > |T|
INTERCEP	1	3.741721	0.280279	13.350	0.0001
LNRM_P	1	0.524645	0.092504	5.672	0.0001
LNI	1	0.177757	0.013347	13.318	0.0001

Model: DEMAND

Dependent variable: LNTM_P

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob > |T|
INTERCEP	1	1.887370	1.186847	1.590	0.1313
LNY_P	1	0.764657	0.110168	6.941	0.0001
LNI	1	-0.077564	0.048797	-1.590	0.1315
LNS1	1	-0.234061	0.141070	-1.659	0.1165
LNS2	1	0.126953	0.250346	0.507	0.6190

e. HAUSMAN TEST

Hausman Test
2SLS vs OLS for SUPPLY Equation

HAUSMAN
6.7586205

2SLS vs OLS for DEMAND Equation

HAUSMAN
2.3088302

SAS PROGRAM

```
Data Laffer1; Input TM RM Y S2 i S1 P;
Cards;

Data Laffer; Set Laffer1;

LNTM_P=LOG(TM/P);
```

```
LNRM_P=LOG(RM/P);
LNi=LOG(i);
LNY_P=LOG(Y/P);
LNS1=LOG(S1);
LNS2=LOG(S2);
T=21;

Proc syslin data=Laffer outest=ols_s outcov;
 SUPPLY: MODEL LNTM_P=LNRM_P LNi;
TITLE 'OLS ESTIMATION OF MONEY SUPPLY FUNCTION';

Proc syslin data=Laffer outest=ols_s outcov;
 DEMAND: MODEL LNTM_P=LNY_P LNi LNS1 LNS2;
TITLE 'OLS ESTIMATION OF MONEY DEMAND FUNCTION';

Proc syslin 2SLS data=Laffer outest=TSLS_s outcov;
 ENDO LNTM_P LNi;
 INST LNRM_P LNY_P LNS1 LNS2;
 SUPPLY: MODEL LNTM_P=LNRM_P LNi/overid;
TITLE '2SLS ESTIMATION OF MONEY SUPPLY FUNCTION';

Proc syslin 2SLS data=Laffer outest=TSLS_d outcov;
 ENDO LNTM_P LNi;
 INST LNRM_P LNY_P LNS1 LNS2;
 DEMAND: MODEL LNTM_P=LNY_P LNi LNS1 LNS2/overid;
TITLE '2SLS ESTIMATION OF MONEY DEMAND FUNCTION';

Proc SYSLIN 3SLS data=Laffer outest=S_3SLS outcov;
 ENDO LNTM_P LNi;
 INST LNRM_P LNY_P LNS1 LNS2;
 SUPPLY: MODEL LNTM_P=LNRM_P LNi;
 DEMAND: MODEL LNTM_P=LNY_P LNi LNS1 LNS2;
TITLE '3SLS ESTIMATION';
RUN;
```

```
PROC IML;
TITLE 'HAUSMAN TEST';

use ols_s; read all var {intercep lnrm_p lni} into ols1;
olsbt_s=ols1[1,]; ols_v_s=ols1[2:4,];
use ols_d; read all var {intercep lny_p lni lns1 lns2} into
ols2;
olsbt_d=ols2[1,]; ols_v_d=ols2[2:6,];

use tsls_s; read all var {intercep lnrm_p lni} into tsls1;
tslsbt_s=tsls1[13,]; tsls_v_s=tsls1[14:16,];
use tsls_d; read all var {intercep lny_p lni lns1 lns2} into
tsls2;
tslsbt_d=tsls2[13,]; tsls_v_d=tsls2[14:18,];

d=tslsbt_s '-olsbt_s'; varcov=tsls_v_s-ols_v_s;
Hausman=d '*inv(varcov)*d;
print 'Hauman Test',, '2SLS vs OLS for SUPPLY Equation',,
Hausman;

d=tslsbt_d '-olsbt_d'; varcov=tsls_v_d-ols_v_d;
Hausman=d '*inv(varcov)*d;
print '2SLS vs OLS for DEMAND Equation',, Hausman;
```

11.20 a. Writing this system of two equations in matrix form, as described in (A.1) we get

$$B = \begin{bmatrix} 1 & \alpha_1 \\ 1 & -\beta_1 \end{bmatrix}; \Gamma = \begin{bmatrix} -\alpha_o & -\alpha_2 \\ -\beta_o & 0 \end{bmatrix}; y_t = \begin{pmatrix} Q_t \\ P_t \end{pmatrix}$$

$x_t' = [1, X_t]$ and $u_t' = (u_{1t}, u_{2t})$.

There are no zero restrictions for the demand equation and only one zero restriction for the supply equation (X_t does not appear in the supply equation). Therefore, the demand equation is not identified by the order condition since the number of excluded exogenous variables $k_2 = 0$ is less than the number of right hand side included endogenous variables $g_1 = 1$.

Similarly, the supply equation is just-identified by the order condition since $k_2 = g_1 = 1$.

By the rank condition, only the supply equation need be checked for identification. In this case, $\phi' = (0, 1)$ since $\gamma_{22} = 0$. Therefore,

$$A\phi = [B, \Gamma]\phi = \begin{pmatrix} \gamma_{12} \\ \gamma_{22} \end{pmatrix} = \begin{pmatrix} -\alpha_2 \\ 0 \end{pmatrix}.$$

This is of rank one as long as $\alpha_2 \neq 0$, i.e., as long as X_t is present in the demand equation. One can also premultiply this system of two equations by

$$F = \begin{bmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{bmatrix},$$

the transformed matrix FB should satisfy the following normalization restrictions:

$$f_{11} + f_{12} = 1 \quad \text{and } f_{21} + f_{22} = 1.$$

Also, $F\Gamma$ must satisfy the following zero restriction:

$$-\alpha_2 f_{21} + 0\, f_{22} = 0.$$

If $\alpha_2 \neq 0$, then $f_{21} = 0$ which also gives from the second normalization equation that $f_{22} = 1$. Hence, the second row of F is indeed the second row of an identity matrix and the supply equation is identified. However, for the demand equation, the only restriction is the normalization restriction $f_{11} + f_{12} = 1$ which is not enough for identification.

b. The OLS estimator of the supply equation yields $\hat{\beta}_{1,ols} = \sum\limits_{t=1}^{T} p_t q_t / \sum\limits_{t=1}^{T} p_t^2$ where $p_t = P_t - \bar{P}$ and $q_t = Q_t - \bar{Q}$. Substituting $q_t = \beta_1 p_t + (u_{2t} - \bar{u}_2)$ we get $\hat{\beta}_{1,ols} = \beta_1 + \sum\limits_{t=1}^{T} p_t(u_{2t} - \bar{u}_2) / \sum\limits_{t=1}^{T} p_t^2$.

The reduced form equation for P_t yields $P_t = \dfrac{\alpha_o - \beta_o}{\alpha_1 + \beta_1} + \dfrac{\alpha_2}{\alpha_1 + \beta_1} X_t + \dfrac{u_{1t} - u_{2t}}{\alpha_1 + \beta_1}$

Hence, $p_t = \frac{\alpha_2}{\alpha_1+\beta_1} x_t + \frac{(u_{1t}-\bar{u}_1)-(u_{2t}-\bar{u}_2)}{\alpha_1+\beta_1}$ where $x_t = X_t - \bar{X}$.

Defining $m_{xx} = \underset{T\to\infty}{\text{plim}} \sum_{t=1}^{T} x_t^2/T$, we get

$$m_{pp} = \underset{T\to\infty}{\text{plim}} \sum_{t=1}^{T} p_t^2/T = \left(\frac{\alpha_2}{\alpha_1+\beta_1}\right)^2 m_{xx} + \frac{m_{u_1u_1}+m_{u_2u_2}-2m_{u_1u_2}}{(\alpha_1+\beta_1)^2}$$

using the fact that $m_{xu_1} = m_{xu_2} = 0$. Also,

$$m_{pu_2} = \text{plim} \sum_{t=1}^{T} p_t u_{2t}/T = \frac{m_{u_1u_2}-m_{u_2u_2}}{(\alpha_1+\beta_1)}$$

using the fact that $m_{xu_2} = 0$. Hence,

$$\underset{T\to\infty}{\text{plim}}\, \hat{\beta}_{1,ols} = \beta_1 + \frac{m_{pu_2}}{m_{pp}} = \beta_1 + \frac{(\sigma_{12}-\sigma_{22})(\alpha_1+\beta_1)}{[\sigma_{11}+\sigma_{22}-2\sigma_{12}+\alpha_2^2 m_{xx}]}$$

where $m_{u_1u_2} = \sigma_{12}$, $m_{u_1u_1} = \sigma_{11}$ and $m_{u_2u_2} = \sigma_{22}$. The second term gives the simultaneous equation bias of $\hat{\beta}_{1,ols}$.

c. If $\sigma_{12} = 0$, then

$$\underset{T\to\infty}{\text{plim}}\left(\hat{\beta}_{1,ols}-\beta_1\right) = -\frac{\sigma_{22}(\alpha_1+\beta_1)}{[\sigma_{11}+\sigma_{22}+\alpha_2^2 m_{xx}]}.$$

The denominator is positive and α_1 and β_1 are positive, hence this bias is negative.

11.21 a. X_2, X_3, X_4 and X_5 are excluded from the first equation. y_2 is the only included right hand side endogenous variable. Therefore, $k_2 = 4$ and $g_1 = 1$, so the first equation is over-identified with degree of over-identification equal to $k_2 - g_1 = 3$. X_1, X_3, X_4 and X_5 are excluded from the second equation, while y_1 and y_3 are the included right hand side endogenous variables. Therefore, $k_2 = 4$ and $g_1 = 2$, so the second equation is over-identified with degree of over-identification equal to $k_2 - g_1 = 2$. X_1 and X_2 are excluded from the third equation and there are no right hand side included endogenous variables. Therefore, $k_2 = 2$ and $g_1 = 0$ and the third equation is over-identified with degree of over-identification equal to $k_2 - g_1 = 2$.

b. Regress y_1 and y_3 on all the X's including the constant, i.e., $[1, X_1, X_2, .., X_5]$. Get $\hat{y}_1$ and $\hat{y}_3$ and regress y_2 on a constant, $\hat{y}_1$, $\hat{y}_3$ and X_2 in the second-step regression.

c. For the first equation, regressing y_2 on X_2 and X_3 yields

$$y_2 = \hat{\pi}_{21} + \hat{\pi}_{22}X_2 + \hat{\pi}_{23}X_3 + \hat{v}_2 = \hat{y}_2 + \hat{v}_2$$

with $\sum_{t=1}^{T} \hat{v}_{2t} = \sum_{t=1}^{T} \hat{v}_{2t}X_{2t} = \sum_{t=1}^{T} \hat{v}_{2t}X_{3t} = 0$ by the property of least squares. Replacing y_2 by $\hat{y}_2$ in Eq. (1) yields

$$y_1 = \alpha_1 + \beta_2(\hat{y}_2 + \hat{v}_2) + \gamma_1 X_1 + u_1 = \alpha_1 + \beta_2\hat{y}_2 + \gamma_1 X_1 + (\beta_2\hat{v}_2 + u_1).$$

In this case, $\sum_{t=1}^{T} \hat{y}_{2t}\hat{v}_{2t} = 0$ because the predictors and residuals of the same regression are uncorrelated. Also, $\sum_{t=1}^{T} \hat{y}_{2t}u_{1t} = 0$ since $\hat{y}_{2t}$ is a linear combination of X_2 and X_3 and the latter are exogenous. However, $\sum_{t=1}^{T} X_{1t}\hat{v}_{2t}$ is not necessarily zero since X_1 was not included in the first stage regression. Hence, this estimation method does not necessarily lead to consistent estimates.

d. The test for over-identification for Eq. (1) can be obtained from the F-test given in Eq. (11.48) with $RRSS^*$ obtained from the second stage regression residual sum of squares of 2SLS run on Eq. (1), i.e., the regression of y_1 on $[1, \hat{y}_2, X_1]$ where $\hat{y}_2$ is obtained from the regression of y_2 on $[1, X_1, X_2.., X_5]$. The $URSS^*$ is obtained from the residual sum of squares of the regression of y_1 on all the set of instruments, i.e., $[1, X_1, .., X_5]$. The URSS is the 2SLS residual sum of squares of Eq. (1) as reported from any 2SLS package. This is obtained as

$$URSS = \sum_{t=1}^{T} \left(y_{1t} - \hat{\alpha}_{1,2SLS} - \hat{\beta}_{2,2SLS}y_{2t} - \hat{\gamma}_{1,2SLS}X_{1t}\right)^2.$$

Note that this differs from $RRSS^*$ in that y_{2t} and not $\hat{y}_{2t}$ is used in the computation of the residuals. ℓ, the number of instruments is 6 and $k_1 = 2$ while $g_1 = 1$. Hence, the numerator degrees of freedom of the F-test is

$\ell - (g_1 + k_1) = 3$ whereas the denominator degrees of freedom of the F-test is $T - \ell = T - 6$. This is equivalent to running 2SLS residuals on the matrix of all predetermined variables $[1, X_1, .., X_5]$ and computing T times the uncentered R^2 of this regression. This is asymptotically distributed as χ^2 with three degrees of freedom. Large values of this statistic reject the over-identification restrictions.

11.23 *Identification and Estimation of a Simple Two-Equation Model.* This solution is based upon Singh and Bhat (1988).

a. *Identification with no further information available.* The structural model being studied is $\begin{bmatrix} 1 & -\beta \\ -1 & 1 \end{bmatrix}\begin{bmatrix} y_{t1} \\ y_{t2} \end{bmatrix} = \begin{bmatrix} \alpha \\ \gamma \end{bmatrix} + \begin{bmatrix} u_{t1} \\ u_{t2} \end{bmatrix}$, with $\Sigma = E\begin{bmatrix} u_{t1}^2 & u_{t1}u_{t2} \\ u_{t2}u_{t1} & u_{t2}^2 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$. The reduced-form equation is given by $\begin{bmatrix} y_{t1} \\ y_{t2} \end{bmatrix} = \begin{bmatrix} \pi_{11} \\ \pi_{12} \end{bmatrix} + \begin{bmatrix} \nu_{t1} \\ \nu_{t2} \end{bmatrix}$, with $\Omega = E\begin{bmatrix} \nu_{t1}^2 & \nu_{t1}\nu_{t2} \\ \nu_{t2}\nu_{t1} & \nu_{t2}^2 \end{bmatrix} = \begin{bmatrix} \omega_1^2 & \omega_{12} \\ \omega_{12} & \omega_2^2 \end{bmatrix}$.

The structural and reduced form parameters are related as follows:

$$\begin{bmatrix} 1 & -\beta \\ -1 & 1 \end{bmatrix}\begin{bmatrix} \pi_{11} \\ \pi_{12} \end{bmatrix} = \begin{bmatrix} \alpha \\ \gamma \end{bmatrix}$$

or

$$\alpha = \pi_{11} - \beta\pi_{12}$$

$$\gamma = \pi_{12} - \pi_{11}$$

We assume that the reduced-form parameters π_{11}, π_{12}, ω_1^2, ω_{12} and ω_2^2 have been estimated. The identification problem is one of recovering α, γ, β, σ_1^2 and σ_{12} and σ_2^2 from the reduced-form parameters.

From the second equation, it is clear that γ is identified. We now examine the relation between Σ and Ω given by $\Sigma = B\Omega B'$ where $B = \begin{bmatrix} 1 & -\beta \\ -1 & 1 \end{bmatrix}$. Thus, we have

$$\sigma_1^2 = \omega_1^2 - 2\beta\omega_{12} + \beta^2\omega_2^2,$$

$$\sigma_{12} = -\omega_1^2 + (1+\beta)\omega_{12} - \beta\omega_2^2,$$

$$\sigma_2^2 = \omega_1^2 - 2\omega_{12} + \omega_2^2.$$

Knowledge of ω_1^2, ω_{12}, and ω_2^2 is enough to recover σ_2^2. Hence, σ_2^2 is identified. Note, however, that σ_1^2 and σ_{12} depend on β. Also, α is dependent on β. Hence, given any $\beta = \beta^*$, we can solve for α, σ_1^2 and σ_{12}. Thus, in principle, we have an infinity of solutions as the choice of β varies. Identification could have been studied using rank and order conditions. However, in a simple framework as this, it is easier to check identification directly. Identification can be reached by imposing an extra restriction $f\left(\beta, \alpha, \sigma_1^2, \sigma_{12}\right) = 0$, which is independent of the three equations we have for σ_1^2, σ_{12} and α. For example, $\beta = 0$ gives a recursive structure. We now study the particular case where $f\left(\beta, \alpha, \sigma_1^2, \sigma_{12}\right) = \sigma_{12} = 0$.

b. *Identification with $\sigma_{12} = 0$*: When $\sigma_{12} = 0$, one can solve immediately for β from the σ_{12} equation:

$$\beta = \frac{\omega_{12} - \omega_1^2}{\omega_2^2 - \omega_{12}}.$$

Therefore, β can be recovered from ω_1^2, ω_{12}, and ω_2^2. Given β, all the other parameters $\left(\sigma_1^2, \alpha\right)$ can be recovered as discussed above.

c. *OLS estimation of β when $\sigma_{12} = 0$*: The OLS estimator of β is

$$\hat{\beta}_{ols} = \frac{\sum_{t=1}^{T}(y_{t1} - \bar{y}_1)(y_{t2} - \bar{y}_2)}{\sum_{t=1}^{T}(y_{t2} - \bar{y}_2)^2} = \beta + \frac{\sum_{t=1}^{T}(u_{t1} - \bar{u}_1)(y_{t1} - \bar{y}_1)}{\sum_{t=1}^{T}(y_{t2} - \bar{y}_2)^2}$$

$$+ \frac{\sum_{t=1}^{T}(u_{t1} - \bar{u}_1)(u_{t2} - \bar{u}_2)}{\sum_{t=1}^{T}(y_{t2} - \bar{y}_2)^2}.$$

The second term is non-zero in probability limit as y_{t1} and u_{t1} are correlated from the first structural equation. Hence $\hat{\beta}_{ols}$ is not consistent.

d. From the second structural equation, we have $y_{t2} - y_{t1} = \gamma + u_{t2}$, which means that $(y_{t2} - \bar{y}_2) - (y_{t1} - \bar{y}_1) = u_{t2} - \bar{u}_2$, or $z_t = u_{t2} - \bar{u}_2$. When $\sigma_{12} = 0$, this z_t is clearly uncorrelated with u_{t1} and by definition correlated with y_{t2}. Therefore, $z_t = (y_{t2} - \bar{y}_2) - (y_{t1} - \bar{y}_1)$ may be taken as instruments, and thus

$$\tilde{\beta}_{IV} = \frac{\sum_{t=1}^{T} z_t(y_{t1} - \bar{y}_1)}{\sum_{t=1}^{T} z_t(y_{t2} - \bar{y}_2)} = \frac{\sum_{t=1}^{T} (y_{t1} - \bar{y}_1)[(y_{t2} - \bar{y}_2) - (y_{t1} - \bar{y}_1)]}{\sum_{t=1}^{T} (y_{t2} - \bar{y}_2)[(y_{t2} - \bar{y}_2) - (y_{t1} - \bar{y}_1)]}$$

is a consistent estimator of β.

e. From the identification solution of β as a function of reduced-form parameters in part (b), one can obtain an alternate estimator of β by replacing the reduced-form parameters by their sample moments. This yields the indirect least squares estimator of β:

$$\tilde{\beta}_{ILS} = \frac{\sum_{t=1}^{T} (y_{t1} - \bar{y}_1)(y_{t2} - \bar{y}_2) - \sum_{t=1}^{T} (y_{t1} - \bar{y}_1)^2}{\sum_{t=1}^{T} (y_{t2} - \bar{y}_2)^2 - \sum_{t=1}^{T} (y_{t1} - \bar{y}_1)(y_{t2} - \bar{y}_2)} = \hat{\beta}_{IV}$$

obtained in part (d). The OLS estimators of the reduced-form parameters are consistent. Therefore, $\tilde{\beta}_{ILS}$ which is a continuous tranform of those estimators is also consistent. It is however, useful to show this consistency directly since it brings out the role of the identifying restriction.

From the structural equations, we have

$$\begin{bmatrix} 1 & -\beta \\ -1 & 1 \end{bmatrix} \begin{bmatrix} y_{t1} - \bar{y}_1 \\ y_{t2} - \bar{y}_2 \end{bmatrix} = \begin{bmatrix} u_{t1} - \bar{u}_1 \\ u_{t2} - \bar{u}_2 \end{bmatrix}.$$

Using the above relation and with some algebra, it can be shown that for $\beta \neq 1$:

$$\tilde{\beta}_{ILS} = \frac{\sum_{t=1}^{T}(u_{t1}-\bar{u}_1)(u_{t2}-\bar{u}_1) + \beta\sum_{t=1}^{T}(u_{t2}-\bar{u}_2)^2}{\sum_{t=1}^{T}(u_{t1}-\bar{u}_1)(u_{t2}-\bar{u}_2) + \sum_{t=1}^{T}(u_{t2}-\bar{u}_2)^2} \quad \text{since,}$$

$\text{plim}\frac{1}{T}\sum_{t=1}^{T}(u_{t1}-\bar{u}_1)(u_{t2}-\bar{u}_2) = \sigma_{12}$, and $\text{plim}\frac{1}{T}\sum_{t=1}^{T}(u_{t2}-\bar{u}_2)^2 = \sigma_2^2$. Therefore, $\text{plim}\,\tilde{\beta}_{ILS} = \left(\sigma_{12}+b\sigma_2^2\right)/\left(\sigma_{12}+\sigma_2^2\right)$. The restriction $\sigma_{12} = 0$ implies $\text{plim}\,\tilde{\beta}_{ILS} = \beta$ which proves consistency. It is clear that $\tilde{\beta}_{ILS}$ can be interpreted as a method-of-moments estimator.

11.24 *Errors in Measurement and the Wald (1940) Estimator*. This is based on Farebrother (1987).

a. The simple IV estimator of β with instrument z is given by $\hat{\beta}_{IV} = \sum_{i=1}^{n} z_i y_i \Big/ \sum_{i=1}^{n} z_i x_i$. But $\sum_{i=1}^{n} z_i y_i$ = sum of the y_i's from the second sample minus the sum of the y_i's from the first sample. If n is even, then $\sum_{i=1}^{n} z_i y_i = \frac{n}{2}(\bar{y}_2 - \bar{y}_1)$. Similarly, $\sum_{i=1}^{n} z_i x_i = \frac{n}{2}(\bar{x}_2 - \bar{x}_1)$, so that $\hat{\beta}_{IV} = (\bar{y}_2 - \bar{y}_1)/(\bar{x}_2 - \bar{x}_1) = \hat{\beta}_W$.

b. Let $\rho^2 = \sigma_u^2/\left(\sigma_u^2+\sigma_*^2\right)$ and $\tau^2 = \sigma_*^2/\left(\sigma_u^2+\sigma_*^2\right)$ then $w_i = \rho^2 x_i^* - \tau^2 u_i$ has mean $E(w_i) = 0$ since $E(x_i^*) = E(u_i) = 0$. Also,

$$\text{var}(w_i) = (\rho^2)^2\sigma_*^2 + (\tau^2)^2\sigma_u^2 = \sigma_*^2\sigma_u^2/\left(\sigma_*^2+\sigma_u^2\right).$$

Since w_i is a linear combination of Normal random variables, it is also Normal.

$$E(x_i w_i) = \rho^2 E\left(x_i x_i^*\right) - \tau^2 E(x_i u_i) = \rho^2\sigma_*^2 - \tau^2\sigma_u^2 = 0.$$

c. Now $\tau^2 x_i + w_i = \tau^2(x_i^* + u_i) + \rho^2 x_i^* - \tau^2 u_i = (\tau^2+\rho^2)x_i^* = x_i^*$ since $\tau^2 + \rho^2 = 1$.

Using this result and the fact that $y_i = \beta x_i^* + \varepsilon_i$, we get

$$\hat{\beta}_W = \sum_{i=1}^{n} z_i y_i \Big/ \sum_{i=1}^{n} z_i x_i$$

$$= \beta\tau^2 + \beta\left(\sum_{i=1}^{n} z_i w_i \Big/ \sum_{i=1}^{n} z_i x_i\right) + \left(\sum_{i=1}^{n} z_i \varepsilon_i \Big/ \sum_{i=1}^{n} z_i x_i\right)$$

so that $E(\hat{\beta}_W/x_1,..,x_n) = \beta\tau^2$. This follows because w_i and ε_i are independent of x_i and therefore independent of z_i which is a function of x_i.

Similarly, $\hat{\beta}_{ols} = \sum_{i=1}^{n} x_i y_i \Big/ \sum_{i=1}^{n} x_i^2 = \beta\tau^2 + \beta \sum_{i=1}^{n} w_i x_i \Big/ \sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n} x_i \varepsilon_i \Big/ \sum_{i=1}^{n} x_i^2$ so that $E(\hat{\beta}_{ols}/x_1,..,x_n) = \beta\tau^2$ since w_i and ε_i are independent of x_i. Therefore, the bias of $\hat{\beta}_{ols}$ = bias of $\hat{\beta}_W = (\beta\tau^2 - \beta) = -\beta\sigma_u^2/\left(\sigma_*^2 + \sigma_u^2\right)$ for all choices of z_i which are a function of x_i.

11.25 *Comparison of t-ratios*. This is based on Farebrother (1991). Let $Z_1 = [y_2, X]$, then by the FWL-Theorem, OLS on the first equation is equivalent to that on $\bar{P}_X y_1 = \bar{P}_X y_2 \alpha + \bar{P}_X u_1$. This yields $\hat{\alpha}_{ols} = \left(y_2'\bar{P}_X y_2\right)^{-1} y_2'\bar{P}_X y_1$ and the residual sum of squares are equivalent

$$y_1'\bar{P}_Z y_1 = y_1'\bar{P}_X y_1 - y_1'\bar{P}_X y_2 \left(y_2'\bar{P}_X y_2\right)^{-1} y_2'\bar{P}_X y_1$$

so that $\text{var}(\hat{\alpha}_{ols}) = s_1^2\left(y_2'\bar{P}_X y_2\right)^{-1}$ where $s_1^2 = y_1'\bar{P}_Z y_1/(T-K)$. The t-ratio for $H_o^a; \alpha = 0$ is

$$t = \frac{y_2'\bar{P}_X y_1}{y_2'\bar{P}_X y_2}\left(\frac{(T-K)y_2'\bar{P}_X y_2}{y_1'\bar{P}_Z y_1}\right)^{1/2}$$

$$= (T-K)^{1/2}(y_2'\bar{P}_X y_1)\left[(y_1'\bar{P}_X y_1)(y_2'\bar{P}_X y_2) - (y_1'\bar{P}_X y_2)^2\right]^{1/2}.$$

This expression is unchanged if the roles of y_1 and y_2 are reversed so that the t-ratio for $H_o^b; \gamma = 0$ in the second equation is the same as that for $H_o^a; \alpha = 0$ in the first equation.

For y_1 and y_2 jointly Normal with correlation coefficient ρ, Farebrother (1991) shows that the above two tests correspond to a test for $\rho = 0$ in the bivariate Normal model so that it makes sense that the two statistics are identical.

11.28 This gives the backup runs for Crime Data for 1987 for Tables 11.2 to 11.6

. ivreg lcrmrte (lprbarr lpolpc= ltaxpc lmix) lprbconv lprbpris lavgsen ldensity lwcon lwtuc lwtrd lwfir lwser lwmfg lwfed lwsta lwloc lpctymle lpctmin west central urban if year==87

Instrumental variables (2SLS) regression

Source	SS	df	MS
Model	22.6350465	20	1.13175232
Residual	4.16465515	69	.060357321
Total	26.7997016	89	.301120243

Number of obs = 90
F(20,69) = 17.35
Prob > F = 0.0000
R-squared = 0.8446
Adj R-squared = 0.7996
Root MSE = .24568

lcrmrte	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lprbarr	-.4393081	.2267579	-1.94	0.057	-.8916777	.0130615
lpolpc	.5136133	.1976888	2.60	0.011	.1192349	.9079918
lprbconv	-.2713278	.0847024	-3.20	0.002	-.4403044	-.1023512
lprbpris	-.0278416	.1283276	-0.22	0.829	-.2838482	.2281651
lavgsen	-.280122	.1387228	-2.02	0.047	-.5568663	-.0033776
ldensity	.3273521	.0893292	3.66	0.000	.1491452	.505559
lwcon	.3456183	.2419206	1.43	0.158	-.137	.8282366
lwtuc	.1773533	.1718849	1.03	0.306	-.1655477	.5202542
lwtrd	.212578	.3239984	0.66	0.514	-.433781	.8589371
lwfir	-.3540903	.2612516	-1.36	0.180	-.8752731	.1670925
lwser	-.2911556	.1122454	-2.59	0.012	-.5150789	-.0672322
lwmfg	.0642196	.1644108	0.39	0.697	-.263771	.3922102
lwfed	.2974661	.3425026	0.87	0.388	-.3858079	.9807402
lwsta	.0037846	.3102383	0.01	0.990	-.615124	.6226931
lwloc	-.4336541	.5166733	-0.84	0.404	-1.464389	.597081
lpctymle	.0095115	.1869867	0.05	0.960	-.3635166	.3825397
lpctmin	.2285766	.0543079	4.21	0.000	.1202354	.3369179
west	-.0952899	.1301449	-0.73	0.467	-.3549219	.1643422
central	-.1792662	.0762815	-2.35	0.022	-.3314437	-.0270888
urban	-.1139416	.143354	-0.79	0.429	-.3999251	.1720419
_cons	-1.159015	3.898202	-0.30	0.767	-8.935716	6.617686

Instrumented: lprbarr lpolpc
Instruments: lprbconv lprbpris lavgsen ldensity lwcon lwtuc lwtrd lwfir lwser lwmfg lwfed lwsta lwloc lpctymle lpctmin west central urban ltaxpc lmix

. estimates store b2sls

. reg lcrmrte lprbarr lprbconv lprbpris lavgsen lpolpc ldensity lwcon lwtuc lwtrd lwfir lwser lwmfg lwfed lwsta lwloc lpctymle lpctmin west central urban if year==87

Source	SS	df	MS
Model	22.8072483	20	1.14036241
Residual	3.99245334	69	.057861643
Total	26.7997016	89	.301120243

Number of obs	=	90
F(20,69)	=	19.71
Prob > F	=	0.0000
R-squared	=	0.8510
Adj R-squared	=	0.8078
Root MSE	=	.24054

lcrmrte	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lprbarr	-.4522907	.0816261	-5.54	0.000	-.6151303	-.2894511
lprbconv	-.3003044	.0600259	-5.00	0.000	-.4200527	-.180556
lprbpris	-.0340435	.1251096	-0.27	0.786	-.2836303	.2155433
lavgsen	-.2134467	.1167513	-1.83	0.072	-.4463592	.0194659
lpolpc	.3610463	.0909534	3.97	0.000	.1795993	.5424934
ldensity	.3149706	.0698265	4.51	0.000	.1756705	.4542707
lwcon	.2727634	.2198714	1.24	0.219	-.165868	.7113949
lwtuc	.1603777	.1666014	0.96	0.339	-.171983	.4927385
lwtrd	.1325719	.3005086	0.44	0.660	-.4669263	.7320702
lwfir	-.3205858	.251185	-1.28	0.206	-.8216861	.1805146
lwser	-.2694193	.1039842	-2.59	0.012	-.4768622	-.0619765
lwmfg	.1029571	.1524804	0.68	0.502	-.2012331	.4071472
lwfed	.3856593	.3215442	1.20	0.234	-.2558039	1.027123
lwsta	-.078239	.2701264	-0.29	0.773	-.6171264	.4606485
lwloc	-.1774064	.4251793	-0.42	0.678	-1.025616	.670803
lpctymle	.0326912	.1580377	0.21	0.837	-.2825855	.3479678
lpctmin	.2245975	.0519005	4.33	0.000	.1210589	.3281361
west	-.087998	.1243235	-0.71	0.481	-.3360167	.1600207
central	-.1771378	.0739535	-2.40	0.019	-.3246709	-.0296046
urban	-.0896129	.1375084	-0.65	0.517	-.3639347	.184709
_cons	-3.395919	3.020674	-1.12	0.265	-9.421998	2.630159

```
. estimates store bols
. hausman b2sls bols
```

	---- Coefficients ----			
	(b) b2sls	(B) bols	(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
lprbarr	-.4393081	-.4522907	.0129826	.2115569
lpolpc	.5136133	.3610463	.152567	.1755231
lprbconv	-.2713278	-.3003044	.0289765	.0597611
lprbpris	-.0278416	-.0340435	.0062019	.0285582
lavgsen	-.280122	-.2134467	-.0666753	.0749208
ldensity	.3273521	.3149706	.0123815	.0557132
lwcon	.3456183	.2727634	.0728548	.1009065
lwtuc	.1773533	.1603777	.0169755	.0422893
lwtrd	.212578	.1325719	.0800061	.1211178
lwfir	-.3540903	-.3205858	-.0335045	.0718228
lwser	-.2911556	-.2694193	-.0217362	.0422646
lwmfg	.0642196	.1029571	-.0387375	.0614869
lwfed	.2974661	.3856593	-.0881932	.1179718

lwsta	.0037846	-.078239	.0820236	.1525764
lwloc	-.4336541	-.1774064	-.2562477	.293554
lpctymle	.0095115	.0326912	-.0231796	.0999404
lpctmin	.2285766	.2245975	.0039792	.0159902
west	-.0952899	-.087998	-.0072919	.0384885
central	-.1792662	-.1771378	-.0021284	.0187016
urban	-.1139416	-.0896129	-.0243287	.0405192

b = consistent under Ho and Ha; obtained from ivreg
B = inconsistent under Ha, efficient under Ho; obtained from regress

```
Test: Ho: difference in coefficients not systematic
          chi2(20) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                   =        0.87
        Prob > chi2 =    1.0000
```

```
. ivreg lcrmrte (lprbarr lpolpc= ltaxpc lmix) lprbconv lprbpris lavgsen ldensity lwcon lwtuc lwtrd
lwfir lwser lwmfg lwfed lwsta lwloc lpctymle lpctmin west central urban if year==87, first
```

First-stage regressions

Source	SS	df	MS
Model	6.84874028	20	.342437014
Residual	7.59345096	69	.110050014
Total	14.4421912	89	.162271812

Number of obs	=	90
F(20,69)	=	3.11
Prob > F	=	0.0002
R-squared	=	0.4742
Adj R-squared	=	0.3218
Root MSE	=	.33174

lprbarr	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lprbconv	-.1946392	.0877581	-2.22	0.030	-.3697119	-.0195665
lprbpris	-.0240173	.1732583	-0.14	0.890	-.3696581	.3216236
lavgsen	.1565061	.1527134	1.02	0.309	-.1481488	.4611611
ldensity	-.2211654	.0941026	-2.35	0.022	-.408895	-.0334357
lwcon	-.2024569	.3020226	-0.67	0.505	-.8049755	.4000616
lwtuc	-.0461931	.230479	-0.20	0.842	-.5059861	.4135999
lwtrd	.0494793	.4105612	0.12	0.904	-.769568	.8685266
lwfir	.050559	.3507405	0.14	0.886	-.6491492	.7502671
lwser	.0551851	.1500094	0.37	0.714	-.2440754	.3544456
lwmfg	.0550689	.2138375	0.26	0.798	-.3715252	.481663
lwfed	.2622408	.4454479	0.59	0.558	-.6264035	1.150885
lwsta	-.4843599	.3749414	-1.29	0.201	-1.232347	.2636277
lwloc	.7739819	.5511607	1.40	0.165	-.3255536	1.873517
lpctymle	-.3373594	.2203286	-1.53	0.130	-.776903	.1021842
lpctmin	-.0096724	.0729716	-0.13	0.895	-.1552467	.1359019
west	.0701236	.1756211	0.40	0.691	-.280231	.4204782
central	.0112086	.1034557	0.11	0.914	-.1951798	.217597

urban	-.0150372	.2026425	-0.07	0.941	-.4192979	.3892234
ltaxpc	-.1938134	.1755345	-1.10	0.273	-.5439952	.1563684
lmix	.2682143	.0864373	3.10	0.003	.0957766	.4406519
_cons	-4.319234	3.797113	-1.14	0.259	-11.89427	3.255799

Source	SS	df	MS
Model	6.99830344	20	.349915172
Residual	5.46683312	69	.079229465
Total	12.4651366	89	.140057714

Number of obs	=	90
F(20,69)	=	4.42
Prob > F	=	0.0000
R-squared	=	0.5614
Adj R-squared	=	0.4343
Root MSE	=	.28148

lpolpc	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lprbconv	.0037744	.0744622	0.05	0.960	-.1447736	.1523223
lprbpris	-.0487064	.1470085	-0.33	0.741	-.3419802	.2445675
lavgsen	.3958972	.1295763	3.06	0.003	.1373996	.6543948
ldensity	.0201292	.0798454	0.25	0.802	-.1391581	.1794165
lwcon	-.5368469	.2562641	-2.09	0.040	-1.04808	-.025614
lwtuc	-.0216638	.1955598	-0.11	0.912	-.411795	.3684674
lwtrd	-.4207274	.3483584	-1.21	0.231	-1.115683	.2742286
lwfir	.0001257	.2976009	0.00	1.000	-.5935718	.5938232
lwser	.0973089	.1272819	0.76	0.447	-.1566116	.3512293
lwmfg	.1710295	.1814396	0.94	0.349	-.1909327	.5329916
lwfed	.8555422	.3779595	2.26	0.027	.1015338	1.609551
lwsta	-.1118764	.3181352	-0.35	0.726	-.7465387	.5227859
lwloc	1.375102	.4676561	2.94	0.004	.4421535	2.30805
lpctymle	.4186939	.1869473	2.24	0.028	.0457442	.7916436
lpctmin	-.0517966	.0619159	-0.84	0.406	-.1753154	.0717222
west	.1458865	.1490133	0.98	0.331	-.151387	.4431599
central	.0477227	.0877814	0.54	0.588	-.1273964	.2228419
urban	-.1192027	.1719407	-0.69	0.490	-.4622151	.2238097
ltaxpc	.5601989	.1489398	3.76	0.000	.2630721	.8573258
lmix	.2177256	.0733414	2.97	0.004	.0714135	.3640378
_cons	-16.33148	3.221824	-5.07	0.000	-22.75884	-9.904113

Instrumental variables (2SLS) regression

Source	SS	df	MS
Model	22.6350465	20	1.13175232
Residual	4.16465515	69	.060357321
Total	26.7997016	89	.301120243

Number of obs	=	90
F(20,69)	=	17.35
Prob > F	=	0.0000
R-squared	=	0.8446
Adj R-squared	=	0.7996
Root MSE	=	.24568

lcrmrte	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
lprbarr	-.4393081	.2267579	-1.94	0.057	-.8916777	.0130615
lpolpc	.5136133	.1976888	2.60	0.011	.1192349	.9079918
lprbconv	-.2713278	.0847024	-3.20	0.002	-.4403044	-.1023512
lprbpris	-.0278416	.1283276	-0.22	0.829	-.2838482	.2281651
lavgsen	-.280122	.1387228	-2.02	0.047	-.5568663	-.0033776
ldensity	.3273521	.0893292	3.66	0.000	.1491452	.505559
lwcon	.3456183	.2419206	1.43	0.158	-.137	.8282366

if year==87

Source	SS	df	MS
Model	6.99830344	20	.349915172
Residual	5.46683312	69	.079229465
Total	12.4651366	89	.140057714

Number of obs	=	90
F(20,69)	=	4.42
Prob > F	=	0.0000
R-squared	=	0.5614
Adj R-squared	=	0.4343
Root MSE	=	.28148

lpolpc	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
lmix	.2177256	.0733414	2.97	0.004	.0714135	.3640378
ltaxpc	.5601989	.1489398	3.76	0.000	.2630721	.8573258
lprbconv	.0037744	.0744622	0.05	0.960	-.1447736	.1523223
lprbpris	-.0487064	.1470085	-0.33	0.741	-.3419802	.2445675
lavgsen	.3958972	.1295763	3.06	0.003	.1373996	.6543948
ldensity	.0201292	.0798454	0.25	0.802	-.1391581	.1794165
lwcon	-.5368469	.2562641	-2.09	0.040	-1.04808	-.025614
lwtuc	-.0216638	.1955598	-0.11	0.912	-.411795	.3684674
lwtrd	-.4207274	.3483584	-1.21	0.231	-1.115683	.2742286
lwfir	.0001257	.2976009	0.00	1.000	-.5935718	.5938232
lwser	.0973089	.1272819	0.76	0.447	-.1566116	.3512293
lwmfg	.1710295	.1814396	0.94	0.349	-.1909327	.5329916
lwfed	.8555422	.3779595	2.26	0.027	.1015338	1.609551
lwsta	-.1118764	.3181352	-0.35	0.726	-.7465387	.5227859
lwloc	1.375102	.4676561	2.94	0.004	.4421535	2.30805
lpctymle	.4186939	.1869473	2.24	0.028	.0457442	.7916436
lpctmin	-.0517966	.0619159	-0.84	0.406	-.1753154	.0717222
west	.1458865	.1490133	0.98	0.331	-.151387	.4431599
central	.0477227	.0877814	0.54	0.588	-.1273964	.2228419
urban	-.1192027	.1719407	-0.69	0.490	-.4622151	.2238097
_cons	-16.33148	3.221824	-5.07	0.000	-22.75884	-9.904113

The test that the additional instruments are jointly significant in the first stage regression for lpolpc is performed as follows:

. test lmix=ltaxpc=0

(1) lmix − ltaxpc = 0
(2) lmix = 0

F(2,69) = 10.56
Prob > F = 0.0001

```
. ivregress 2sls lcrmrte (lprbarr lpolpc= ltaxpc lmix) lprbconv lprbpris lavgsen ldensity lwcon
lwtuc lwtrd lwfir lwser lwmfg lwfed lwsta lwloc lpctymle lpctmin west central urban if year==87
```

Instrumental variables (2SLS) regression

Number of obs	=	90
Wald chi2(20)	=	452.49
Prob > chi2	=	0.0000
R-squared	=	0.8446
Root MSE	=	.21511

lcrmrte	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
lprbarr	-.4393081	.1985481	-2.21	0.027	-.8284552	-.050161
lpolpc	.5136133	.1730954	2.97	0.003	.1743526	.852874
lprbconv	-.2713278	.074165	-3.66	0.000	-.4166885	-.1259672
lprbpris	-.0278416	.112363	-0.25	0.804	-.2480691	.1923859
lavgsen	-.280122	.121465	-2.31	0.021	-.5181889	-.042055
ldensity	.3273521	.0782162	4.19	0.000	.1740512	.4806531
lwcon	.3456183	.2118244	1.63	0.103	-.06955	.7607865
lwtuc	.1773533	.1505016	1.18	0.239	-.1176244	.4723309
lwtrd	.212578	.2836914	0.75	0.454	-.3434468	.7686029
lwfir	-.3540903	.2287506	-1.55	0.122	-.8024333	.0942527
lwser	-.2911556	.0982815	-2.96	0.003	-.4837837	-.0985274
lwmfg	.0642196	.1439573	0.45	0.656	-.2179316	.3463707
lwfed	.2974661	.2998936	0.99	0.321	-.2903145	.8852468
lwsta	.0037846	.2716432	0.01	0.989	-.5286262	.5361954
lwtuc	.1773533	.1718849	1.03	0.306	-.1655477	.5202542
lwtrd	.212578	.3239984	0.66	0.514	-.433781	.8589371
lwfir	-.3540903	.2612516	-1.36	0.180	-.8752731	.1670925
lwser	-.2911556	.1122454	-2.59	0.012	-.5150789	-.0672322
lwmfg	.0642196	.1644108	0.39	0.697	-.263771	.3922102
lwfed	.2974661	.3425026	0.87	0.388	-.3858079	.9807402
lwsta	.0037846	.3102383	0.01	0.990	-.615124	.6226931
lwloc	-.4336541	.5166733	-0.84	0.404	-1.464389	.597081
lpctymle	.0095115	.1869867	0.05	0.960	-.3635166	.3825397
lpctmin	.2285766	.0543079	4.21	0.000	.1202354	.3369179
west	-.0952899	.1301449	-0.73	0.467	-.3549219	.1643422
central	-.1792662	.0762815	-2.35	0.022	-.3314437	-.0270888
urban	-.1139416	.143354	-0.79	0.429	-.3999251	.1720419
_cons	-1.159015	3.898202	-0.30	0.767	-8.935716	6.617686

Instrumented: lprbarr lpolpc
Instruments: lprbconv lprbpris lavgsen ldensity lwcon lwtuc lwtrd lwfir lwser lwmfg lwfed lwsta lwloc lpctymle lpctmin west central urban ltaxpc lmix

```
. reg lprbarr lmix ltaxpc lprbconv lprbpris lavgsen ldensity lwcon lwtuc lwtrd lwfir lwser lwmfg
lwfed lwsta lwloc lpctymle lpctmin west central urban if year==87
```

Source	SS	df	MS			
Model	6.84874028	20	.342437014	Number of obs	=	90
Residual	7.59345096	69	.110050014	F(20, 69)	=	3.11
				Prob > F	=	0.0002
				R-squared	=	0.4742
				Adj R-squared	=	0.3218
Total	14.4421912	89	.162271812	Root MSE	=	.33174

lprbarr	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lmix	.2682143	.0864373	3.10	0.003	.0957766	.4406519
ltaxpc	-.1938134	.1755345	-1.10	0.273	-.5439952	.1563684
lprbconv	-.1946392	.0877581	-2.22	0.030	-.3697119	-.0195665
lprbpris	-.0240173	.1732583	-0.14	0.890	-.3696581	.3216236
lavgsen	.1565061	.1527134	1.02	0.309	-.1481488	.4611611
ldensity	-.2211654	.0941026	-2.35	0.022	-.408895	-.0334357
lwcon	-.2024569	.3020226	-0.67	0.505	-.8049755	.4000616
lwtuc	-.0461931	.230479	-0.20	0.842	-.5059861	.4135999
lwtrd	.0494793	.4105612	0.12	0.904	-.769568	.8685266
lwfir	.050559	.3507405	0.14	0.886	-.6491492	.7502671
lwser	.0551851	.1500094	0.37	0.714	-.2440754	.3544456
lwmfg	.0550689	.2138375	0.26	0.798	-.3715252	.481663
lwfed	.2622408	.4454479	0.59	0.558	-.6264035	1.150885
lwsta	-.4843599	.3749414	-1.29	0.201	-1.232347	.2636277
lwloc	.7739819	.5511607	1.40	0.165	-.3255536	1.873517
lpctymle	-.3373594	.2203286	-1.53	0.130	-.776903	.1021842
lpctmin	-.0096724	.0729716	-0.13	0.895	-.1552467	.1359019
west	.0701236	.1756211	0.40	0.691	-.280231	.4204782
central	.0112086	.1034557	0.11	0.914	-.1951798	.217597
urban	-.0150372	.2026425	-0.07	0.941	-.4192979	.3892234
_cons	-4.319234	3.797113	-1.14	0.259	-11.89427	3.255799

Then we can test that the additional instruments are jointly significant in the first stage regression for lprbarr as follows:

```
. test lmix=ltaxpc=0
(1) lmix - ltaxpc = 0
(2) lmix = 0

          F(2, 69) = 5.78
        Prob > F = 0.0048
```

For lpolpc, the first stage regression is given by

```
. reg lpolpc lmix ltaxpc lprbconv lprbpris lavgsen ldensity lwcon lwtuc lwtrd lwfir lwser lwmfg
lwfed lwsta lwloc lpctymle lpctmin west central urban
```

lwloc	-.4336541	.4523966	-0.96	0.338	-1.320335	.453027
lpctymle	.0095115	.1637246	0.06	0.954	-.3113827	.3304058
lpctmin	.2285766	.0475517	4.81	0.000	.135377	.3217763
west	-.0952899	.1139543	-0.84	0.403	-.3186361	.1280564
central	-.1792662	.0667917	-2.68	0.007	-.3101756	-.0483569

urban	-.1139416	.1255201	-0.91	0.364	-.3599564	.1320732
_cons	-1.159015	3.413247	-0.34	0.734	-7.848855	5.530825

Instrumented: lprbarr lpolpc
Instruments: lprbconv lprbpris lavgsen ldensity lwcon lwtuc lwtrd lwfir lwser lwmfg lwfed lwsta lwloc lpctymle lpctmin west central urban ltaxpc lmix

. estat firststage

Shea's partial R-squared

Variable	Shea's Partial R-sq.	Shea's Adj. Partial R-sq.
lprbarr	0.1352	-0.0996
lpolpc	0.2208	0.0093

Minimum eigenvalue statistic = 5.31166

Critical Values # of endogenous regressors: 2
Ho: Instruments are weak # of excluded instruments: 2

	5%	10%	20%	30%
2SLS relative bias		(not available)		
	10%	15%	20%	25%
2SLS Size of nominal 5% Wald test	7.03	4.58	3.95	3.63
LIML Size of nominal 5% Wald test	7.03	4.58	3.95	3.63

One can run the same IV regressions for other years, but this is not done here to save space.

11.30 *Growth and Inequality Reconsidered*

a. The 3SLS results are replicated using Stata below:

reg3 (Growth: dly = yrt gov m2y inf swo dtot f_pcy d80 d90) (Inequality: gih = yrt m2y civ mlg mlgldc), exog(commod f_civ f_dem f_dtot f_flit f_gov f_inf f_m2y f_swo f_yrt pop urb lex lfr marea oil legor_fr legor_ge legor_mx legor_ sc legor_uk) endog (yrt gov m2y inf swo civ mlg mlgldc)

Three-stage least-squares regression

Equation	Obs	Parms	RMSE	"R-sq"	chi2	P
Growth	119	9	2.34138	0.3905	65.55	0.0000
Inequality	119	5	7.032975	0.4368	94.28	0.0000

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Growth						
yrt	-.0280625	.1827206	-0.15	0.878	-.3861882	.3300632
gov	-.0533221	.0447711	-1.19	0.234	-.1410718	.0344276
m2y	.0085368	.0199759	0.43	0.669	-.0306152	.0476889
inf	-.0008174	.0025729	-0.32	0.751	-.0058602	.0042254
swo	4.162776	.9499015	4.38	0.000	2.301003	6.024548
dtot	26.03736	23.05123	1.13	0.259	-19.14221	71.21694
f_pcy	-1.38017	.5488437	-2.51	0.012	-2.455884	-.3044564
d80	-1.560392	.545112	-2.86	0.004	-2.628792	-.4919922
d90	-3.413661	.6539689	-5.22	0.000	-4.695417	-2.131906
_cons	13.00837	3.968276	3.28	0.001	5.230693	20.78605
Inequality						
yrt	-1.244464	.4153602	-3.00	0.003	-2.058555	-.4303731
m2y	-.120124	.0581515	-2.07	0.039	-.2340989	-.0061492
civ	.2531189	.7277433	0.35	0.728	-1.173232	1.67947
mlg	.292672	.0873336	3.35	0.001	.1215012	.4638428
mlgldc	-.0547843	.0576727	-0.95	0.342	-.1678207	.0582522
_cons	33.13231	5.517136	6.01	0.000	22.31893	43.9457

Endogenous variables: dly gih yrt gov m2y inf swo civ mlg mlgldc

Exogenous variables: dtot f_pcy d80 d90 commod f_civ f_dem f_dtot f_flit f_gov f_inf f_m2y f_swo f_yrt pop urb lex lfr marea oil legor_fr legor_ge legor_mx legor_sc legor_uk

b. The 3sls results of the respecified equations are replicated using Stata below:

```
. reg3 (Growth: dly = gih yrt gov m2y inf swo dtot f_pcy d80 d90) (Inequality: gih = dly yrt m2y civ mlg mlgldc), exog(f_civ f_dem f_dtot f_flit f_gov f_inf f_m2y f_swo f_yrt pop urb lex lfr marea commod oil legor_fr legor_ge legor_mx legor_sc legor_uk) endog(yrt gov m2y inf swo civ mlg mlgldc)
```

Three-stage least-squares regression

Equation	Obs	Parms	RMSE	"R-sq"	chi2	P
Growth	119	10	2.366525	0.3773	65.26	0.0000
Inequality	119	6	6.969627	0.4469	96.06	0.0000

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Growth						
gih	-.0181063	.0452471	-0.40	0.689	-.106789	.0705763
yrt	-.0642514	.1971126	-0.33	0.744	-.4505851	.3220822
gov	-.0559828	.0457033	-1.22	0.221	-.1455597	.0335941
m2y	.0084914	.0201616	0.42	0.674	-.0310246	.0480074
inf	.0002947	.0031579	0.09	0.926	-.0058948	.0064841
swo	4.148469	.9771934	4.25	0.000	2.233205	6.063733
dtot	28.56603	23.60078	1.21	0.226	-17.69064	74.8227
f_pcy	-1.35157	.5565555	-2.43	0.015	-2.442399	-.2607418
d80	-1.581281	.5498957	-2.88	0.004	-2.659057	-.503505
d90	-3.392124	.6624323	-5.12	0.000	-4.690467	-2.09378
_cons	13.67011	4.382862	3.12	0.002	5.079854	22.26036
Inequality						
dly	.2642484	.3586634	0.74	0.461	-.438719	.9672158
yrt	-1.221263	.4122647	-2.96	0.003	-2.029286	-.4132386
m2y	-.115859	.0577899	-2.00	0.045	-.2291251	-.0025929
civ	.1510064	.7449721	0.20	0.839	-1.309112	1.611125
mlg	.2951145	.0870164	3.39	0.001	.1245654	.4656635
mlgldc	-.0451182	.0581734	-0.78	0.438	-.159136	.0688996
_cons	32.06252	5.611046	5.71	0.000	21.06507	43.05997

Endogenous variables: dly gih yrt gov m2y inf swo civ mlg mlgldc

Exogenous variables: dtot f_pcy d80 d90 f_civ f_dem f_dtot f_flit f_gov f_inf f_m2y f_swo f_yrt pop urb lex lfr marea commod oil legor_fr legor_ge legor_mx legor_sc legor_uk

11.31 *Married Women Labor Supply*

b. The following Stata output replicates Equation 2 of Table IV of Mroz (1987, p. 770) which runs 2SLS using the following instrumental variables:

```
. local B unem city motheduc fatheduc
. local E exper expersq
. local control age educ
. ivregress 2sls hours (lwage= ‘B’ ‘E’) nwifeinc kidslt6 kidsge6 ‘control’, vce(robust)
```

Instrumental variables (2SLS) regression

Number of obs	=	428
Wald chi2(6)	=	18.67
Prob > chi2	=	0.0048
R-squared	=	.
Root MSE	=	1141.3

hours	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf. Interval]	
lwage	1261.574	460.7772	2.74	0.006	358.4673	2164.681
nwifeinc	-8.341979	4.586405	-1.82	0.069	-17.33117	.6472101
kidslt6	-234.7069	182.201	-1.29	0.198	-591.8143	122.4005
kidsge6	-59.78876	49.04792	-1.22	0.223	-155.9209	36.34341
age	-10.23035	9.264046	-1.10	0.269	-28.38755	7.926846
educ	-147.895	52.71884	-2.81	0.005	-251.222	-44.56794
_cons	2374.632	531.9775	4.46	0.000	1331.976	3417.289

Instrumented: lwage

Instruments: nwifeinc kidslt6 kidsge6 age educ unem city motheduc fatheduc exper expersq

. estat overid

Test of overidentifying restrictions:

Score chi2(5) = 5.9522 (p = 0.3109)

. estat firststage, forcenonrobust all

First-stage regression summary statistics

Variable	R-sq.	Adjusted R-sq.	Partial R-sq.	Robust F(6,416)	Prob > F
lwage	0.1719	0.1500	0.0468	2.58071	0.0182

Shea's partial R-squared

Variable	Shea's Partial R-sq.	Shea's Adj. Partial R-sq.
lwage	0.0468	0.0240

Minimum eigenvalue statistic = 3.40543

Critical Values # of endogenous regressors: 1
Ho: Instruments are weak # of excluded instruments: 6

	5%	10%	20%	30%
2SLS relative bias	19.28	11.12	6.76	5.15

	10%	15%	20%	25%
2SLS Size of nominal 5% Wald test	29.18	16.23	11.72	9.38
LIML Size of nominal 5% Wald test	4.45	3.34	2.87	2.61

Similarly, the following Stata output replicates Equation 3 of Table IV of Mroz(1987, p. 770) which runs 2SLS using the following instrumental variables:

```
. local F2 age2 educ2 age_educ

. ivregress 2sls hours (lwage= 'B' 'E' 'F2') nwifeinc kidslt6 kidsge6
'control', vce(robust)
```

Instrumental variables (2SLS) regression

Number of obs	=	428
Wald chi2(6)	=	24.56
Prob > chi2	=	0.0004
R-squared	=	.
Root MSE	=	942.03

hours	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
lwage	831.2505	312.1889	2.66	0.008	219.3715	1443.129
nwifeinc	-6.963789	3.814549	-1.83	0.068	-14.44017	.5125885
kidslt6	-270.9764	154.5197	-1.75	0.079	-573.8294	31.87667
kidsge6	-78.37192	39.257	-2.00	0.046	-155.3142	-1.429619
age	-9.38908	7.589533	-1.24	0.216	-24.26429	5.486132
educ	-102.9946	36.35213	-2.83	0.005	-174.2435	-31.74576
_cons	2287.175	432.9381	5.28	0.000	1438.632	3135.718

```
Instrumented: lwage
Instruments: nwifeinc kidslt6 kidsge6 age educ unem city motheduc fatheduc
             exper expersq age2 educ2 age_educ

. estat overid

Test of overidentifying restrictions:
Score chi2(8) = 19.8895 (p = 0.0108)
```

. estat firststage, forcenonrobust all

First-stage regression summary statistics

Variable	R-sq.	Adjusted R-sq.	Partial R-sq.	Robust F(6,416)	Prob > F
lwage	0.1846	0.1570	0.0614	2.3555	0.0133

Shea's partial R-squared

Variable	Shea's Partial R-sq.	Shea's Adj. PartialR-sq.
lwage	0.0614	0.0320

Minimum eigenvalue statistic = 3.0035

Critical Values # of endogenous regressors: 1
Ho: Instruments are weak # of excluded instruments: 9

	5%	10%	20%	30%
2SLS relative bias	20.53	11.46	6.65	4.92

	10%	15%	20%	25%
2SLS Size of nominal 5% Wald test	36.19	19.71	14.01	11.07
LIML Size of nominal 5% Wald test	3.81	2.93	2.54	2.32

Similarly, one can generate the rest of the regressions in Table IV and their related diagnostics.

References

Baltagi, B.H. (1989), "A Hausman Specification Test in a Simultaneous Equations Model," *Econometric Theory*, Solution 88.3.5, 5: 453–467.

Farebrother, R.W. (1987), "The Exact Bias of Wald's Estimation," *Econometric Theory*, Solution 85.3.1, 3: 162.

Farebrother, R.W. (1991), "Comparison of t-Ratios," *Econometric Theory*, Solution 90.1.4, 7: 145–146.

Hausman, J.A. (1983), "Specification and Estimation of Simultaneous Equation Models," Chapter 7 in Griliches, Z. and Intriligator, M.D. (eds.) Handbook of Econometrics, Vol. I (North Holland: Amsterdam).

Laffer, A.B., (1970), "Trade Credit and the Money Market," Journal of Political Economy, 78: 239–267.

Mroz, T.A. (1987), "The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions," Econometrica, 55: 765–799.

Singh, N. and A N. Bhat (1988), "Identification and Estimation of a Simple Two-Equation Model," *Econometric Theory*, Solution 87.3.3, 4: 542–545.

Theil, H. (1971), Principles of Econometrics (Wiley: New York).

Wald, A. (1940), "Fitting of Straight Lines if Both Variables are Subject to Error," Annals of Mathematical Statistics, 11: 284–300.

CHAPTER 12
Pooling Time-Series of Cross-Section Data

12.1 *Fixed Effects and the Within Transformation.*

a. Premultiplying (12.11) by Q one gets

$$Qy = \alpha Q\iota_{NT} + QX\beta + QZ_{\mu}\mu + Q\nu$$

But $PZ_{\mu} = Z_{\mu}$ and $QZ_{\mu} = 0$. Also, $P\iota_{NT} = \iota_{NT}$ and $Q\iota_{NT} = 0$. Hence, this transformed equation reduces to (12.12)

$$Qy = QX\beta + Q\nu$$

Now $E(Q\nu) = QE(\nu) = 0$ and $\text{var}(Q\nu) = Q\,\text{var}(\nu)Q' = \sigma_{\nu}^2 Q$, since

$$\text{var}(\nu) = \sigma_{\nu}^2 I_{NT}$$

and Q is symmetric and idempotent.

b. For the general linear model $y = X\beta + u$ with $E(uu') = \Omega$, a necessary and sufficient condition for OLS to be equivalent to GLS is given by $X'\Omega^{-1}\bar{P}_X$ where $\bar{P}_X = I - P_X$ and $P_X = X(X'X)^{-1}X'$, see Eq. (9.7) of Chap. 9. For Eq. (12.12), this condition can be written as

$$(X'Q)(Q/\sigma_{\nu}^2)\bar{P}_{QX} = 0$$

using the fact that Q is idempotent, the left hand side can be written as

$$(X'Q)\bar{P}_{QX}/\sigma_{\nu}^2$$

which is clearly 0, since $\bar{P}_{QX}$ is the orthogonal projection of QX.

One can also use Zyskind's condition $P_X\Omega = \Omega P_X$ given in Eq. (9.8) of Chap. 9. For Eq. (12.12), this condition can be written as

$$P_{QX}(\sigma_{\nu}^2 Q) = (\sigma_{\nu}^2 Q)P_{QX}$$

B.H. Baltagi, *Solutions Manual for Econometrics*, Springer Texts in Business and Economics, DOI 10.1007/978-3-642-54548-1_12,

But, $P_{QX} = QX(X'QX)^{-1}X'Q$. Hence, $P_{QX}Q = P_{QX}$ and $QP_{QX} = P_{QX}$ and the condition is met. Alternatively, we can verify that OLS and GLS yield the same estimates. Note that $Q = I_{NT} - P$ where $P = I_N \otimes \bar{J}_T$ is idempotent and is therefore its own generalized inverse. The variance–covariance matrix of the disturbance $\tilde{\nu} = Q\nu$ in (12.12) is $E(\tilde{\nu}\tilde{\nu}') = E(Q\nu\nu'Q) = \sigma_\nu^2 Q$ with generalized inverse Q/σ_ν^2. OLS on (12.12) yields

$$\hat{\beta} = (X'QQX)^{-1}X'QQy = (X'QX)^{-1}X'Qy$$

which is $\tilde{\beta}$ given by (12.13). Also, GLS on (12.12) using generalized inverse yields

$$\hat{\beta} = (X'QQQX)^{-1}X'QQQy = (X'QX)^{-1}X'Qy = \tilde{\beta}.$$

c. The Frisch–Waugh–Lovell (FWL) theorem of Davidson and MacKinnon (1993, p. 19) states that for

$$y = X_1\beta_1 + X_2\beta_2 + u \tag{12.1}$$

If we premultiply by the orthogonal projection of X_2 given by

$$M_2 = I - X_2(X_2'X_2)^{-1}X_2',$$

then $M_2X_2 = 0$ and (1) becomes

$$M_2y = M_2X_1\beta_1 + M_2u \tag{12.2}$$

The OLS estimate of β_1 from (2) is the same as that from (1) *and* the residuals from (1) are the same as the residuals from (2). This was proved in Sect. 7.3. Here we will just use this result. For the model in (12.11)

$$y = Z\delta + Z_\mu\mu + \nu$$

Let $Z = X_1$ and $Z_\mu = X_2$. In this case, $M_2 = I - Z_\mu\left(Z_\mu'Z_\mu\right)^{-1}Z_\mu' = I - P = Q$. In this case, premultiplying by M_2 is equivalent to premultiplying by Q and Eq. (2) above becomes Eq. (12.12) in the text. By the FWL theorem, OLS on (12.12) which yields (12.13) is the same as OLS on

(12.11). Note that

$Z = [\iota_{NT}, X]$ and $QZ = [0, QX]$ since $Q\iota_{NT} = 0$.

12.2 *Variance–Covariance Matrix of Random Effects.*

a. From (12.17) we get

$$\Omega = \sigma_\mu^2 (I_N \otimes J_T) + \sigma_\nu^2 (I_N \otimes I_T)$$

Replacing J_T by $T\bar{J}_T$, and I_T by $(E_T + \bar{J}_T)$ where E_T is by definition $(I_T - \bar{J}_T)$, one gets

$$\Omega = T\sigma_\mu^2 (I_N \otimes \bar{J}_T) + \sigma_\nu^2 (I_N \otimes E_T) + \sigma_\nu^2 (I_N \otimes \bar{J}_T)$$

collecting terms with the same matrices, we get

$$\Omega = (T\sigma_\mu^2 + \sigma_\nu^2)(I_N \otimes \bar{J}_T) + \sigma_\nu^2 (I_N \otimes E_T) = \sigma_1^2 P + \sigma_\nu^2 Q$$

where $\sigma_1^2 = T\sigma_\mu^2 + \sigma_\nu^2$.

b. $P = Z_\mu \left(Z_\mu' Z_\mu\right)^{-1} Z_\mu' = I_N \otimes \bar{J}_T$ is a projection matrix of Z_μ. Hence, it is by definition symmetric and idempotent. Similarly, $Q = I_{NT} - P$ is the orthogonal projection matrix of Z_μ. Hence, Q is also symmetric and idempotent. By definition, $P + Q = I_{NT}$. Also, $PQ = P(I_{NT} - P) = P - P^2 = P - P = 0$.

c. From (12.18) and (12.19) one gets

$$\Omega\Omega^{-1} = (\sigma_1^2 P + \sigma_\nu^2 Q)\left(\frac{1}{\sigma_1^2} P + \frac{1}{\sigma_\nu^2} Q\right) = P + Q = I_{NT}$$

since $P^2 = P$, $Q^2 = Q$ and $PQ = 0$ as verified in part (b). Similarly, $\Omega^{-1}\Omega = I_{NT}$.

d. From (12.20) one gets

$$\Omega^{-1/2}\Omega^{-1/2} = \left(\frac{1}{\sigma_1} P + \frac{1}{\sigma_\nu} Q\right)\left(\frac{1}{\sigma_1} P + \frac{1}{\sigma_\nu} Q\right) = \frac{1}{\sigma_1^2} P^2 + \frac{1}{\sigma_\nu^2} Q^2$$
$$= \frac{1}{\sigma_1^2} P + \frac{1}{\sigma_\nu^2} Q = \Omega^{-1}$$

using the fact that $P^2 = P$, $Q^2 = Q$ and $PQ = 0$.

12.3 *Fuller and Battese (1974) Transformation.*

From (12.20) one gets $\sigma_\nu\Omega^{-1/2} = Q + (\sigma_\nu/\sigma_1)P$

Therefore,

$$y^* = \sigma_\nu\Omega^{-1/2}y = Qy + (\sigma_\nu/\sigma_1)Py = y - Py + (\sigma_\nu/\sigma_1)Py$$

$$= y - (1 - (\sigma_\nu/\sigma_1))Py = y - \theta Py$$

where $\theta = 1 - (\sigma_\nu/\sigma_1)$. Recall that the typical element of Py is $\bar{y}_{i.}$, therefore, the typical element of y^* is $y^*_{it} = y_{it} - \theta\bar{y}_{i.}$.

12.4 *Unbiased Estimates of the Variance-Components.* $E(u'Pu) = E(tr(uu'P)) = tr(E(uu')P) = tr(\Omega P)$. From (12.21), $\Omega P = \sigma_1^2 P$ since $PQ = 0$. Hence, from (12.18),

$$E\left(\hat{\sigma}_1^2\right) = \frac{E(u'Pu)}{tr(P)} = \frac{tr\left(\sigma_1^2 P\right)}{tr(P)} = \sigma_1^2.$$

Similarly, $E(u'Qu) = tr(\Omega Q) = tr\left(\sigma_\nu^2 Q\right)$ where the last equality follows from (12.18) and the fact that $\Omega Q = \sigma_\nu^2 Q$ since $PQ = 0$. Hence, from (12.22),

$$E\left(\hat{\sigma}_\nu^2\right) = \frac{E(u'Qu)}{tr(Q)} = \frac{tr\left(\sigma_\nu^2 Q\right)}{tr(Q)} = \sigma_\nu^2.$$

12.5 *Swamy and Arora (1972) Estimates of the Variance-Components.*

a. $\hat{\hat{\sigma}}_\nu^2$ given by (12.23) is the s^2 from the regression given by (12.11). In fact

$$\hat{\hat{\sigma}}_\nu^2 = y'Q[I_{NT} - P_{QX}]Qy/[N(T-1) - K]$$

where $P_{QX} = QX(X'QX)^{-1}X'Q$. Substituting Qy from (12.12) into $\hat{\hat{\sigma}}_\nu^2$, one gets

$$\hat{\hat{\sigma}}_\nu^2 = \nu'Q[I_{NT} - P_{QX}]Q\nu/[N(T-1) - K]$$

with $Q\nu \sim \left(0, \sigma_\nu^2 Q\right)$. Therefore,

$$E[\nu'Q[I_{NT} - P_{QX}]Q\nu] = E[tr\{Q\nu\nu'Q[I_{NT} - P_{QX}]\}]$$

$$= \sigma_\nu^2 tr\{Q - QP_{QX}\} = \sigma_\nu^2\{N(T-1) - tr(P_Q X)\}$$

where the last equality follows from the fact that $QP_{QX} = P_{QX}$. Also,

$$\mathrm{tr}(P_{QX}) = \mathrm{tr}(X'QX)(X'QX)^{-1} = \mathrm{tr}(I_K) = K.$$

Hence, $E\left(\hat{\hat{\sigma}}_{\nu}^{2}\right) = \sigma_{\nu}^{2}$.

b. Similarly, $\hat{\hat{\sigma}}_1^2$ given by (12.26) is the s^2 from the regression given by $Py = PZ\delta + Pu$

In fact,

$$\hat{\hat{\sigma}}_1^2 = y'P[I_{NT} - P_{PZ}]Py/(N - K - 1)$$

where $P_{PZ} = PZ(Z'PZ)^{-1}Z'P$. Substituting Py from (3) into $\hat{\hat{\sigma}}_1^2$ one gets

$$\hat{\hat{\sigma}}_1^2 = u'P[I_{NT} - P_{PZ}]Pu/(N - K - 1) = u'P[I_{NT} - P_{PZ}]Pu/(N - K - 1)$$

with $Pu \sim \left(0, \sigma_1^2 P\right)$ as can be easily verified from (12.18). Therefore,

$$E(u'P[I_{NT} - P_{PZ}]Pu) = E[\mathrm{tr}\{Puu'P(I_{NT} - P_{PZ})\} = \sigma_1^2 \mathrm{tr}\{P - PP_{PZ}\}$$
$$= \sigma_1^2\{N - \mathrm{tr}(P_{PZ})\}$$

where the last equality follows from the fact that $PP_{PZ} = P_{PZ}$. Also,

$$\mathrm{tr}(P_{PZ}) = \mathrm{tr}(Z'PZ)(Z'PZ)^{-1} = \mathrm{tr}(I_{K'}) = K'.$$

Hence, $E\left(\hat{\hat{\sigma}}_1^2\right) = \sigma_1^2$.

12.6 *System Estimation.*

a. OLS on (12.27) yields

$$\hat{\delta}_{ols} = \left[(Z'Q, Z'P)\begin{pmatrix} QZ \\ PZ \end{pmatrix}\right]^{-1}(Z'Q, Z'P)\begin{pmatrix} QY \\ Py \end{pmatrix}$$
$$= (Z'QZ + Z'PZ)^{-1}(Z'Qy + Z'Py)$$
$$= (Z'(Q + P)Z)^{-1}Z'(Q + P)y = (Z'Z)^{-1}Z'y$$

since $Q + P = I_{NT}$.

b. GLS on (12.27) yields

$$\hat{\delta}_{gls} = \left((Z'Q, Z'P)\begin{bmatrix}\sigma_\nu^2 Q & 0\\ 0 & \sigma_1^2 P\end{bmatrix}^{-1}\begin{pmatrix}QZ\\ PZ\end{pmatrix}\right)^{-1}$$

$$(Z'Q, Z'P)\begin{bmatrix}\sigma_\nu^2 Q & 0\\ 0 & \sigma_1^2 P\end{bmatrix}^{-1}\begin{pmatrix}Qy\\ Py\end{pmatrix}$$

Using generalized inverse

$$\begin{bmatrix}\sigma_\nu^2 Q & 0\\ 0 & \sigma_1^2 P\end{bmatrix}^{-1} = \begin{bmatrix}Q/\sigma_\nu^2 & 0\\ 0 & P/\sigma_1^2\end{bmatrix}$$

one gets

$$\hat{\delta}_{GLS} = \left[(Z'QZ)/\sigma_\nu^2 + (Z'PZ)/\sigma_1^2\right]^{-1}\left[(Z'Qy)/\sigma_\nu^2 + (Z'Py)/\sigma_1^2\right]$$
$$= (Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}y$$

where Ω^{-1} is given by (12.19).

12.7 *Random Effects Is More Efficient than Fixed Effects.* From (12.30) we have

$$\mathrm{var}(\hat{\beta}_{GLS}) = \sigma_\nu^2[W_{XX} + \phi^2 B_{XX}]^{-1}$$

where $W_{XX} = X'QX$, $B_{XX} = X'(P - \bar{J}_{NT})X$ and $\phi^2 = \sigma_\nu^2/\sigma_1^2$. The Within estimator is given by (12.13) with $\mathrm{var}(\tilde{\beta}_{Within}) = \sigma_\nu^2 W_{XX}^{-1}$ Hence,

$$(\mathrm{var}(\hat{\beta}_{GLS}))^{-1} - (\mathrm{var}(\tilde{\beta}_{Within}))^{-1} = \frac{1}{\sigma_\nu^2}[W_{xx} + \phi^2 B_{XX}] - \frac{1}{\sigma_\nu^2}W_{XX} = \phi^2 B_{XX}/\sigma_\nu^2$$

which is positive semi-definite. Hence, $\mathrm{var}(\tilde{\beta}_{Within}) - \mathrm{var}(\tilde{\beta}_{GLS})$ is positive semi-definite. This last result uses the well known fact that if $A^{-1} - B^{-1}$ is positive semi-definite, then $B - A$ is positive semi-definite.

12.8 *Maximum Likelihood Estimation of the Random Effects Model.*

a. Differentiating (12.34) with respect to ϕ^2 yields

$$\frac{\partial L_c}{\partial \phi^2} = -\frac{NT}{2}\cdot\frac{d'\left(P - \bar{J}_{NT}\right)d}{d'\left[Q + \phi^2(P - \bar{J}_{NT})\right]d} + \frac{N}{2}\cdot\frac{1}{\phi^2}$$

setting $\partial L_c/\partial\phi^2 = 0$, we get $Td'(P - \bar{J}_{NT})d\phi^2 = d'Qd + \phi^2 d'(P - \bar{J}_{NT})d$ solving for ϕ^2, we get $\hat{\phi}^2(T-1)d'\left(P - \bar{J}_{NT}\right)d = d'Qd$ which yields (12.35).

b. Differentiating (12.34) with respect to β yields

$$\frac{\partial L_c}{\partial \beta} = -\frac{NT}{2} \cdot \frac{1}{d'\left[Q + \phi^2(P - \bar{J}_{NT})\right] d} \cdot \frac{\partial}{\partial \beta} d'[Q + \phi^2(P - \bar{J}_{NT})]d$$

setting $\frac{\partial L_c}{\partial \beta} = 0$ is equivalent to solving $\frac{\partial d'[Q + \phi^2(P - \bar{J}_{NT})]d}{\partial \beta} = 0$

Using the fact that $d = y - X\beta$, this yields

$$-2X'\left[Q + \phi^2\left(P - \bar{J}_{NT}\right)\right](y - X\hat{\beta}) = 0$$

Solving for $\hat{\beta}$, we get $X'[Q+\phi^2(P-\bar{J}_{NT})]X\hat{\beta} = X'[Q+\phi^2(P-\bar{J}_{NT})]y$ which yields (12.36).

12.9 *Prediction in the Random Effects Model.*

a. From (12.2) and (12.38), $E(u_{i,T+S}u_{jt}) = \sigma_\mu{}^2$, for $i = j$ and zero otherwise. The only correlation over time occurs because of the presence of the same individual across the panel. The ν_{it}'s are not correlated for different time periods. In vector form

$$w = E(u_{i,T+S}\ u) = \sigma_\mu{}^2(0, .., 0, .., 1.., 1, .., 0, .., 0)'$$

where there are T ones for the i-th individual. This can be rewritten as $w = \sigma_\mu{}^2(\ell_i \otimes \iota_T)$ where ℓ_i is the i-th column of I_N, i.e., ℓ_i is a vector that has 1 in the i-th position and zero elsewhere. ι_T is a vector of ones of dimension T.

b. $\left(\ell_i{}' \otimes \iota_T{}'\right) P = \left(\ell_i{}' \otimes \iota_T{}'\right)\left(I_N \otimes \frac{\iota_T \iota_T{}'}{T}\right) = \left(\ell_i{}' \otimes \iota_T{}'\right).$

Therefore, in (12.39)

$$w'\Omega^{-1} = \sigma_\mu^2\left(\ell_i{}' \otimes \iota_T{}'\right)\left[\frac{1}{\sigma_1^2} P + \frac{1}{\sigma_\nu^2} Q\right] = \frac{\sigma_\mu^2}{\sigma_1^2}\left(\ell_i{}' \otimes \iota_T{}'\right)$$

since $\left(\ell_i' \otimes \iota_T{}'\right) Q = \left(\ell_i{}' \otimes \iota_T{}'\right)(I_{NT} - P) = \left(\ell_i{}' \otimes \iota_T{}'\right) - \left(\ell_i{}' \otimes \iota_T{}'\right) = 0.$

12.10 Using the motor gasoline data on diskette, the following SAS output and program replicates the results in Tables 12.1–12.7. The program uses the IML procedure (SAS matrix language). This can be easily changed into GAUSS or any other matrix language.

RESULTS OF OLS

LOOK1	PARAMETER	STANDARD ERROR	T-STATISTICS
INTERCEPT	2.39133	0.11693	20.45017
INCOME	0.88996	0.03581	24.85523
PRICE	-0.89180	0.03031	-29.4180
CAR	-0.76337	0.01861	-41.0232

RESULTS OF BETWEEN

LOOK1	PARAMETER	STANDARD ERROR	T-STATISTICS
INTERCEPT	2.54163	0.52678	4.82480
INCOME	0.96758	0.15567	6.21571
PRICE	-0.96355	0.13292	-7.24902
CAR	-0.79530	0.08247	-9.64300

RESULTS OF WITHIN

LOOK1	PARAMETER	STANDARD ERROR	T-STATISTICS
INCOME	0.66225	0.07339	9.02419
PRICE	-0.32170	0.04410	-7.29496
CAR	-0.64048	0.02968	-21.5804

RESULTS OF WALLACE-HUSSAIN

LOOK1	PARAMETER	STANDARD ERROR	T-STATISTICS
INTERCEPT	1.90580	0.19403	9.82195
INCOME	0.54346	0.06353	8.55377
PRICE	-0.47111	0.04550	-10.3546
CAR	-0.60613	0.02840	-21.3425

RESULTS OF AMEMIYA

LOOK1	PARAMETER	STANDARD ERROR	T-STATISTICS
INTERCEPT	2.18445	0.21453	10.18228
INCOME	0.60093	0.06542	9.18559
PRICE	-0.36639	0.04138	-8.85497
CAR	-0.62039	0.02718	-22.8227

RESULTS OF SWAMY-ARORA

LOOK1	PARAMETER	STANDARD ERROR	T-STATISTICS
INTERCEPT	1.99670	0.17824	11.20260
INCOME	0.55499	0.05717	9.70689
PRICE	-0.42039	0.03866	-10.8748
CAR	-0.60684	0.02467	-24.5964

RESULTS OF NERLOVE

LOOK1	PARAMETER	STANDARD ERROR	T-STATISTICS
INTERCEPT	2.40173	1.41302	1.69972
INCOME	0.66198	0.07107	9.31471
PRICE	-0.32188	0.04271	-7.53561
CAR	-0.64039	0.02874	-22.2791

RESULTS OF MAXIMUM LIKELIHOOD

LOOK1	PARAMETER	STANDARD ERROR	T-STATISTICS
INTERCEPT	2.10623	0.20096	10.48088
INCOME	0.58044	0.06286	9.23388
PRICE	-0.38582	0.04072	-9.47515
CAR	-0.61401	0.02647	-23.2001

ONE-WAY ERROR COMPONENT MODEL WITH GASOLINE DATA:
BETA, VARIANCES OF BETA, AND THETA

ESTIMATORS	BETA1	BETA2	BETA3	STD_BETA1	STD_BETA2	STD_BETA3	THETA
OLS	0.88996	-0.89180	-0.76337	0.03581	0.03031	0.01861	0.00000
BETWEEN	0.96758	-0.96355	-0.79530	0.15567	0.13292	0.08247	.
WITHIN	0.66225	-0.32170	-0.64048	0.07339	0.04410	0.02968	1.00000
WALLACE & HUSSAIN	0.54346	-0.47111	-0.60613	0.06353	0.04550	0.02840	0.84802
AMEMIYA	0.60093	-0.36639	-0.62039	0.06542	0.04138	0.02718	0.93773
SWAMY & ARORA	0.55499	-0.42039	-0.60684	0.05717	0.03866	0.02467	0.89231
NERLOVE	0.66198	-0.32188	-0.64039	0.07107	0.04271	0.02874	0.99654
IMLE	0.58044	-0.38582	-0.61401	0.06286	0.04072	0.02647	0.92126

NEGATIVE VAR_MHU

NEGA_VAR	
OLS ESTIMATOR	.
BETWEEN ESTIMATOR	.
WITHIN ESTIMATOR	.
WALLACE & HUSSAIN ESTIMATOR	0
AMEMIYA ESTIMATOR	0
SWAMY & ARORA ESTIMATOR	0
NERLOVE ESTIMATOR	.
IMLE	.

SAS PROGRAM

```
Options linesize=162;
   Data Gasoline;
```

```
Infile 'b:/gasoline.dat' firstobs=2;
Input @1 Country $ @10 Year @15 Lgaspcar @29 Lincomep
      @44 Lprpmg @61 Lcarpcap ;

Proc IML;
      Use Gasoline; Read all into Temp;
      N=18; T=19; NT=N*T;
      One=Repeat(1,NT,1);
      X=Temp[,3:5]; Y=Temp[,2]; Z=One||X; K=NCOL(X);
      l_t=J(T,1,1); JT=(l_t*l_t`); Z_U=I(N)@l_t;
      P=Z_U*INV(Z_U`*Z_U)*Z_U`;     Q=I(NT)-P;
      JNT=Repeat(JT,N,N);       JNT_BAR=JNT/NT;

* --------- OLS ESTIMATORS ------------ *;

      OLS_BETA=INV(Z`*Z)*Z`*Y;
      OLS_RES=Y-Z*OLS_BETA;
      VAR_REG=SSQ(OLS_RES)/(NT-NCOL(Z));
      VAR_COV=VAR_REG*INV(Z`*Z);
      STD_OLS=SQRT(VECDIAG(VAR_COV));
      T_OLS=OLS_BETA/STD_OLS;

   LOOK1=OLS_BETA||STD_OLS||T_OLS;
   CTITLE={'PARAMETER' 'STANDARD ERROR' 'T-STATISTICS'};
   RTITLE={'INTERCEPT' 'INCOME' 'PRICE' 'CAR'};
   PRINT 'RESULTS OF OLS',,
   LOOK1(|COLNAME=CTITLE ROWNAME=RTITLE FORMAT=8.5|);

* --------- BETWEEN ESTIMATOR ------------------ *;

      BW_BETA=INV(Z`*P*Z)*Z`*P*Y;
      BW_RES=P*Y-P*Z*BW_BETA;
      VAR_BW=SSQ(BW_RES)/(N-NCOL(Z));
      V_C_BW=VAR_BW*INV (Z`*P*Z);
```

```
    STD_BW=SQRT(VECDIAG(V_C_BW));
    T_BW=BW_BETA/STD_BW;

  LOOK1=BW_BETA||STD_BW||T_BW;
  CTITLE={'PARAMETER' 'STANDARD ERROR' 'T-STATISTICS'};
  RTITLE={'INTERCEPT' 'INCOME' 'PRICE' 'CAR'};
  PRINT 'RESULTS OF BETWEEN',,
  LOOK1(|COLNAME=CTITLE ROWNAME=RTITLE FORMAT=8.5|);

* --------- WITHIN ESTIMATORS ------------ *;

    WT_BETA=INV(X`*Q*X)*X`*Q*Y;
    WT_RES=Q*Y-Q*X*WT_BETA;
    VAR_WT=SSQ(WT_RES)/(NT-N-NCOL(X));
    V_C_WT=VAR_WT*INV(X`*Q*X);
    STD_WT=SQRT(VECDIAG(V_C_WT));
    T_WT=WT_BETA/STD_WT;
  LOOK1=WT_BETA||STD_WT||T_WT;
  CTITLE={'PARAMETER' 'STANDARD ERROR' 'T-STATISTICS'};
  RTITLE={'INCOME' 'PRICE' 'CAR'};
  PRINT 'RESULTS OF WITHIN',,
  LOOK1(|COLNAME=CTITLE ROWNAME=RTITLE FORMAT=8.5|);

* -- WALLACE & HUSSAIN ESTIMATOR OF VARIANCE COMPONENTS --- *;

    WH_V_V=(OLS_RES`*Q*OLS_RES)/(NT-N);
    WH_V_1=(OLS_RES`*P*OLS_RES)/N;

    ******* Checking for negative VAR_MHU *******;

    WH_V_MHU=(WH_V_1-WH_V_V)/T;
    IF WH_V_MHU<0 THEN NEGA_WH=1; ELSE NEGA_WH=0;
    WH_V_MHU=WH_V_MHU # (WH_V_MHU>0);
    WH_V_1=(T*WH_V_MHU)+WH_V_V;
```

```
    OMEGA_WH=(Q/WH_V_V)+(P/WH_V_1);
    WH_BETA=INV(Z`*OMEGA_WH*Z)*Z`*OMEGA_WH*Y;
    THETA_WH=1-(SQRT(WH_V_V)/SQRT(WH_V_1));
    OMEGAWH=(Q/SQRT(WH_V_V))+(P/SQRT(WH_V_1));
    WH_RES=(OMEGAWH*Y)-(OMEGAWH*Z*WH_BETA);
    VAR_WH=SSQ(WH_RES)/(NT-NCOL(Z));
    V_C_WH=INV(Z`*OMEGA_WH*Z);
    STD_WH=SQRT(VECDIAG(V_C_WH));
    T_WH=WH_BETA/STD_WH;

  LOOK1=WH_BETA||STD_WH||T_WH;
  CTITLE={'PARAMETER' 'STANDARD ERROR' 'T-STATISTICS'};
  RTITLE={'INTERCEPT' 'INCOME' 'PRICE' 'CAR'};
  PRINT 'RESULTS OF WALLACE-HUSSAIN',,
  LOOK1(|COLNAME=CTITLE ROWNAME=RTITLE FORMAT=8.5|);
  FREE OMEGA_WH OMEGAWH WH_RES;

* -- AMEMIYA ESTIMATOR OF VARIANCE COMPONENTS --- *;

    Y_BAR=Y[:];     X_BAR=X[:,];
    ALPHA_WT=Y_BAR-X_BAR*WT_BETA;
    LSDV_RES=Y-ALPHA_WT*ONE-X*WT_BETA;
    AM_V_V=(LSDV_RES`*Q*LSDV_RES)/(NT-N);

    AM_V_1=(LSDV_RES`*P*LSDV_RES)/N;

    ***** Checking for negative VAR_MHU *********;

    AM_V_MHU=(AM_V_1-AM_V_V)/T;
    IF AM_V_MHU<0 THEN NEGA_AM=1; ELSE NEGA_AM=0;
    AM_V_MHU=AM_V_MHU # (AM_V_MHU>0);
    AM_V_1=(T*AM_V_MHU)+AM_V_V;

    OMEGA_AM=(Q/AM_V_V)+(P/AM_V_1);
    AM_BETA=INV(Z`*OMEGA_AM*Z)*Z`*OMEGA_AM*Y;
```

```
        THETA_AM=1-(SQRT(AM_V_V)/SQRT(AM_V_1));
        OMEGAAM=(Q/SQRT(AM_V_V))+(P/SQRT(AM_V_1));
        AM_RES=(OMEGAAM*Y)-(OMEGAAM*Z*AM_BETA);
        VAR_AM=SSQ(AM_RES)/(NT-NCOL(Z));
        V_C_AM=INV(Z`*OMEGA_AM*Z);
        STD_AM=SQRT(VECDIAG(V_C_AM));
        T_AM=AM_BETA/STD_AM;

    LOOK1=AM_BETA||STD_AM||T_AM;
    CTITLE={'PARAMETER' 'STANDARD ERROR' 'T-STATISTICS'};
    RTITLE={'INTERCEPT' 'INCOME' 'PRICE' 'CAR'};
    PRINT 'RESULTS OF AMEMIYA',,
    LOOK1(|COLNAME=CTITLE ROWNAME=RTITLE FORMAT=8.5|);
    FREE OMEGA_AM OMEGAAM AM_RES;

* --- SWAMY & ARORA ESTIMATOR OF VARIANCE COMPONENTS ---- *;

        SA_V_V=(Y`*Q*Y-Y`*Q*X*INV(X`*Q*X)*X`*Q*Y)/(NT-N-K);
        SA_V_1=(Y`*P*Y-Y`*P*Z*INV(Z`*P*Z)*Z`*P*Y)/(N-K-1);

        ****** Checking for negative VAR_MHU ********;

        SA_V_MHU=(SA_V_1-SA_V_V)/T;
        IF SA_V_MHU<0 THEN NEGA_SA=1; ELSE NEGA_SA=0;
        SA_V_MHU=SA_V_MHU # (SA_V_MHU>0);
        SA_V_1=(T*SA_V_MHU)+SA_V_V;

        OMEGA_SA=(Q/SA_V_V)+(P/SA_V_1);
        SA_BETA=INV(Z`*OMEGA_SA*Z)*Z`*OMEGA_SA*Y;
        THETA_SA=1-(SQRT(SA_V_V)/SQRT(SA_V_1));
        OMEGASA=(Q/SQRT(SA_V_V))+(P/SQRT(SA_V_1));
        SA_RES=(OMEGASA*Y)-(OMEGASA*Z*SA_BETA);
        VAR_SA=SSQ(SA_RES)/(NT-NCOL(Z));
        V_C_SA=INV(Z`*OMEGA_SA*Z);
```

```
    STD_SA=SQRT(VECDIAG(V_C_SA));
    T_SA=SA_BETA/STD_SA;

  LOOK1=SA_BETA||STD_SA||T_SA;
  CTITLE={'PARAMETER' 'STANDARD ERROR' 'T-STATISTICS'};
  RTITLE={'INTERCEPT' 'INCOME' 'PRICE' 'CAR'};
  PRINT 'RESULTS OF SWAMY-ARORA',,
  LOOK1(|COLNAME=CTITLE ROWNAME=RTITLE FORMAT=8.5|);
  FREE OMEGA_SA OMEGASA SA_RES;

*-- NERLOVE ESTIMATOR OF VARIANCE COMPONENTS AND BETA --*;

    MHU=P*LSDV_RES;
    MEAN_MHU=MHU[:];
    DEV_MHU=MHU-(ONE*MEAN_MHU);
    VAR_MHU=SSQ(DEV_MHU)/T*(N-1);
    NL_V_V=SSQ(WT_RES)/NT;
    NL_V_1=T*VAR_MHU+NL_V_V;
    OMEGA_NL=(Q/NL_V_V)+(P/NL_V_1);
    NL_BETA=INV(Z`*OMEGA_NL*Z)*Z`*OMEGA_NL*Y;
    THETA_NL=1-(SQRT(NL_V_V)/SQRT(NL_V_1));
    OMEGANL=(Q/SQRT(NL_V_V))+(P/SQRT(NL_V_1));
    NL_RES=(OMEGANL*Y)-(OMEGANL*Z*NL_BETA);
    VAR_NL=SSQ(NL_RES)/(NT-NCOL(Z));
    V_C_NL=INV(Z`*OMEGA_NL*Z);
    STD_NL=SQRT(VECDIAG(V_C_NL));
    T_NL=NL_BETA/STD_NL;

  LOOK1=NL_BETA| |STD_NL| |T_NL;
  CTITLE={'PARAMETER' 'STANDARD ERROR' 'T-STATISTICS'};
  RTITLE={'INTERCEPT' 'INCOME' 'PRICE' 'CAR'};
  PRINT 'RESULTS OF NERLOVE',,
  LOOK1(|COLNAME=CTITLE ROWNAME=RTITLE FORMAT=8.5|);
  FREE OMEGA_NL OMEGANL NL_RES;
```

```
        *--- MAXIMUM LIKELIHOOD ESTIMATION ----*;

    /* START WITH WITHIN AND BETWEEN BETA SUGGESTED BY
          BREUSCH(1987) */;


          CRITICAL=1;
          BETA_W=WT_BETA;
          BETA_B=BW_BETA[2:K+1,];
          BETA_MLE=WT_BETA;

          DO WHILE (CRITICAL>0.0001);
             WT_RES=Y - X*BETA_W;
             BW_RES=Y - X*BETA_B;

PHISQ_W=(WT_RES`*Q*WT_RES)/((T-1)*(WT_RES`*(P-JNT_BAR)*WT_RES));
PHISQ_B=(BW_RES`*Q*BW_RES)/((T-1)*(BW_RES`*(P-JNT_BAR)*BW_RES));
CRITICAL=PHISQ_W-PHISQ_B;
BETA_W=INV(X`*(Q+PHISQ_W*(P-JNT_BAR))*X)*X`*(Q+PHISQ_W*(P-JNT_
                    BAR))*Y;

BETA_B=INV(X`* (Q+PHISQ_B*(P-JNT_BAR))*X)*X`*(Q+PHISQ_B*(P-JNT_BAR))*Y;

BETA_MLE=(BETA_W+BETA_B)/2;
END;

D_MLE=Y-X*BETA_MLE;

PHISQ_ML=(D_MLE` *Q*D_MLE)/((T-1)*D_MLE` *(P-JNT_BAR)*D_MLE);
 THETA_ML=1-SQRT(PHISQ_ML);
VAR_V_ML=D_MLE` *(Q+PHISQ_ML*(P-JNT_BAR))*D_MLE/NT;
VAR_1_ML=VAR_V_ML/PHISQ_ML;
OMEGA_ML=(Q/VAR_V_ML)+(P/VAR_1_ML);
ML_BETA=INV(Z`*OMEGA_ML*Z)*Z` *OMEGA_ML*Y;
OMEGAML=(Q/SQRT(VAR_V_ML))+(P/SQRT(VAR_1_ML));
```

```
ML_RES=(OMEGAML*Y)-(OMEGAML*Z*ML_BETA);
VAR_ML=SSQ(ML_RES) / (NT-NCOL(Z));
V_C_ML=INV(Z` *OMEGA_ML*Z);
STD_ML=SQRT (VECDIAG (V_C_ML));
T_ML=ML_BETA/STD_ML;

LOOK1=ML_BETA| |STD_ML| |T_ML;
CTITLE={'PARAMETER' 'STANDARD ERROR' 'T-STATISTICS'};
RTITLE={'INTERCEPT' 'INCOME' 'PRICE' 'CAR'};
PRINT 'RESULTS OF MAXIMUM LIKELIHOOD',,
LOOK1(|COLNAME=CTITLE ROWNAME=RTITLE FORMAT=8.5|);
FREE OMEGA_ML;

*--------- PRINT AND OUTPUT INFORMATION ---------------*;

BETA=OLS_BETA` [,2:K+1]//BW_BETA` [,2:K+1]//WT_BETA`//WH_BETA` [,2: K+1]//
AM_BETA` [,2:K+1]//SA_BETA`[,2:K+1]//NL_BETA`[,2:K+1]//ML_BETA`[,2:K+1];
STD_ERR=STD_OLS`[,2:K+1]//STD_BW`[,2:K+1]//STD_WT`//STD_WH`[,2:K+1]//
STD_AM`[,2:K+1]//STD_SA`[,2:K+1]//STD_NL`[,2:K+1]//STD_ML`[,2:K+1];
THETAS={0,.,1}//THETA_WH//THETA_AM//THETA_SA//THETA_NL//THETA_ML;

     NEGA_VAR={.,.,.}//NEGA_WH//NEGA_AM//NEGA_SA//{.,.};
     OUTPUT=BETA| |STD_ERR| |THETAS| |NAGA_VAR;
     C2={"BETA1" "BETA2" "BETA3" "STD_BETA1" "STD_BETA2"
          "STD_BETA3" "THETA"};
     R={"OLS ESTIMATOR" "BETWEEN ESTIMATOR" "WITHIN
        ESTIMATOR" "WALLACE & HUSSAIN ESTIMATOR"
        "AMEMIYA ESTIMATOR" "SWAMY & ARORA ESTIMATOR"
        "NERLOVE ESTIMATOR" "IMLE"};

   PRINT 'ONE-WAY ERROR COMPONENT MODEL WITH GASOLINE DATA:
          BETA, VARIANCES OF BETA, AND THETA'
          ,,OUTPUT (|ROWNAME=R COLNAME=C2 FORMAT=8.5|);

   PRINT 'NEGATIVE VAR_MHU',,NEGA_VAR (|ROWNAME=R|);
```

12.11 *Bounds on s^2 in the Random Effects Model.*

a. This solution is based on Baltagi and Krämer (1994). From (12.3), one gets $\hat{\delta}_{ols} = (Z'Z)^{-1}Z'y$ and $\hat{u}_{ols} = y - Z\hat{\delta}_{ols} = \bar{P}_Z u$ where $\bar{P}_Z = I_{NT} - P_Z$ with $P_Z = Z(Z'Z)^{-1}Z'$. Also,

$$E(s^2) = E[\hat{u}'\hat{u}/(NT-K')] = E\left[u'\bar{P}_Z u/(NT-K')\right] = tr(\Omega\bar{P}_Z)/(NT-K')$$

which from (12.17) reduces to

$$E(s^2) = \sigma_\nu^2 + \sigma_\mu^2(NT - tr(I_N \otimes J_T)P_Z)/(NT-K')$$

since $tr(I_{NT}) = tr(I_N \otimes J_T) = NT$

and $tr(P_Z) = K'$. By adding and subtracting σ_μ^2, one gets

$$E(s^2) = \sigma^2 + \sigma_\mu^2[K' - tr(I_N \otimes J_T)P_Z]/(NT-K')$$

where $\sigma^2 = E\left(u_{it}^2\right) = \sigma_\mu^2 + \sigma_\nu^2$ for all i and t.

b. Nerlove (1971) derived the characteristic roots and vectors of Ω given in (12.17). These characteristic roots turn out to be σ_ν^2 with multiplicity N(T-1) and $\left(T\sigma_\mu^2 + \sigma_\nu^2\right)$ with multiplicity N. Therefore, the smallest $(n-K')$ characteristic roots are made up of the (n-N) σ_ν^2's and $(N-K')$ of the $\left(T\sigma_\mu^2 + \sigma_\nu^2\right)$'s. This implies that the mean of the $(n-K')$ smallest characteristic roots of $\Omega = \left[(n-N)\sigma_\nu^2 + (N-K')\left(T\sigma_\mu^2 + \sigma_\nu^2\right)\right]/(n-K')$. Similarly, the largest $(n-K')$ characteristic roots are made up of the $N\left(T\sigma_\mu^2 + \sigma_\nu^2\right)$'s and $(n-N-K')$ of the σ_ν^2's. This implies that the mean of the $(n-K')$ largest characteristic roots of $\Omega = \left[N\left(T\sigma_\mu^2 + \sigma_\nu^2\right) + (n-N-K')\sigma_\nu^2\right]/(n-K')$. Using the Kiviet and Krämer (1992) inequalities, one gets

$$0 \le \sigma_\nu^2 + (n-TK')\sigma_\mu^2/(n-K') \le E(s^2) \le \sigma_\nu^2 + n\sigma_\mu^2/(n-K') \le n\sigma^2/(n-K').$$

As $n \to \infty$, both bounds tend to σ^2, and s^2 is asymptotically unbiased, irrespective of the particular evolution of X.

12.12 $M = I_{NT} - Z(Z'Z)^{-1}Z'$ and $M^* = I_{NT} - Z^*(Z^{*\prime}Z^*)^{-1}Z^{*\prime}$ are both symmetric and idempotent. From (12.43), it is clear that $Z = Z^*I^*$ with $I^* = (\iota_N \otimes I_{K'})$, ι_N being a vector of ones of dimension N and $K' = K + 1$.

$$MM^* = I_{NT} - Z(Z'Z)^{-1}Z' - Z^*(Z^{*\prime}Z^*)^{-1}Z^{*\prime} + Z(Z'Z)^{-1}Z'Z^*(Z^{*\prime}Z^*)^{-1}Z^{*\prime}$$

Substituting $Z = Z^*I^*$, the last term reduces to

$$Z(Z'Z)^{-1}Z'Z^*(Z^{*\prime}Z^*)^{-1}Z^{*\prime} = Z(Z'Z)^{-1}I^{*\prime}Z^{*\prime}Z^*(Z^{*\prime}Z^*)^{-1}Z^{*\prime} = Z(Z'Z)^{-1}Z'$$

Hence, $MM^* = I_{NT} - Z^*(Z^{*\prime}Z^*)^{-1}Z^{*\prime} = M^*$.

12.13 This problem differs from problem 12.12 in that $\dot{Z} = \Sigma^{-1/2} Z$ and $\dot{Z}^* = \Sigma^{-1/2}Z^*$. Since $Z = Z^*I^*$, premultiplying both sides by $\Sigma^{-1/2}$ one gets $\dot{Z} = \dot{Z}^*I^*$. Define

$$\dot{M} = I_{NT} - \dot{Z}(\dot{Z}'\dot{Z})^{-1}\dot{Z}'$$

and

$$\dot{M}^* = I_{NT} - \dot{Z}^*(\dot{Z}^{*\prime}\dot{Z}^*)^{-1}\dot{Z}^{*\prime}$$

Both are projection matrices that are symmetric and idempotent. The proof of $\dot{M}\dot{M}^* = \dot{M}^*$ is the same as that of $MM^* = M^*$ given in problem 12.12 with $\dot{Z}$ replacing Z and $\dot{Z}^*$ replacing Z^*.

12.16 **a.** $\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y = \beta + (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}u$ with $E(\hat{\beta}_{GLS}) = \beta$

$\tilde{\beta}_{Within} = (X'QX)^{-1}X'Qy = \beta + (X'QX)^{-1}X'Qu$ with $E(\tilde{\beta}_{Within}) = \beta$

Therefore, $\hat{q} = \hat{\beta}_{GLS} - \tilde{\beta}_{Within}$ has $E(\hat{q}) = 0$.

$$\begin{aligned}
\text{cov}(\hat{\beta}_{GLS}, \hat{q}) &= E(\hat{\beta}_{GLS} - \beta)(\hat{q}') \\
&= E(\hat{\beta}_{GLS} - \beta)[(\hat{\beta}_{GLS} - \beta)' - (\tilde{\beta}_{Within} - \beta)'] \\
&= \text{var}(\hat{\beta}_{GLS}) - \text{cov}(\hat{\beta}_{GLS}, \tilde{\beta}_{Within}) \\
&= E[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}uu'\Omega^{-1}X(X'\Omega^{-1}X)^{-1}] \\
&\quad - E[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}uu'QX(X'QX)^{-1}]
\end{aligned}$$

$$= (X'\Omega^{-1}X)^{-1} - (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\Omega QX(X'QX)^{-1}$$

$$= (X'\Omega^{-1}X)^{-1} - (X'\Omega^{-1}X)^{-1} = 0.$$

b. Using the fact that $\tilde{\beta}_{\text{Within}} = -\hat{q} + \hat{\beta}_{\text{GLS}}$, one gets

$\text{var}(\tilde{\beta}_{\text{Within}}) = \text{var}(-\hat{q} + \hat{\beta}_{\text{GLS}}) = \text{var}(\hat{q}) + \text{var}(\hat{\beta}_{\text{GLS}})$,

since $\text{cov}(\hat{\beta}_{\text{GLS}}, \hat{q}) = 0$. Therefore,

$\text{var}(\hat{q}) = \text{var}(\tilde{\beta}_{\text{Within}}) - \text{var}(\hat{\beta}_{\text{GLS}}) = \sigma_\nu^2(X'QX)^{-1} - (X'\Omega^{-1}X)^{-1}$.

Detailed solutions for problems 12.17, 12.18 and 12.21 are given in Baltagi (2009).

References

Baltagi, B.H. (2009), *A Companion to Econometric Analysis of Panel Data* (Wiley: Chichester).

Baltagi, B.H. and W. Krämer (1994), "Consistency, Asymptotic Unbiasedness and Bounds on the Bias of s^2 in the Linear Regression Model with Error Components Disturbances," Statistical Papers, 35: 323–328.

Davidson, R. and J.G. MacKinnon, 1993, *Estimation and Inference in Econometrics* (Oxford University Press, New York).

Fuller, W.A. and G.E. Battese (1974), "Estimation of Linear Models with Cross-Error Structure," Journal of Econometrics, 2: 67–78.

Nerlove, M. (1971), "A Note on Error Components Models," Econometrica, 39: 383–396.

Swamy, P.A.V.B. and S.S. Arora (1972), "The Exact Finite Sample Properties of the Estimators of Coefficients in the Error Components Regression Models," Econometrica, 40: 261–275.

CHAPTER 13
Limited Dependent Variables

13.1 *The Linear Probability Model*

y_i	u_i	Prob.
1	$1 - x_i'\beta$	π_i
0	$-x_i'\beta$	$1 - \pi_i$

a. Let $\pi_i = \Pr[y_i = 1]$, then $y_i = 1$ when $u_i = 1 - x_i'\beta$ with probability π_i as shown in the table above. Similarly, $y_i = 0$ when $u_i = -x_i'\beta$ with probability $1 - \pi_i$. Hence, $E(u_i) = \pi_i\left(1 - x_i'\beta\right) + (1 - \pi_i)\left(-x_i'\beta\right)$.

For this to equal zero, we get, $\pi_i - \pi_i x_i'\beta + \pi_i x_i'\beta - x_i'\beta = 0$ which gives $\pi_i = x_i'\beta$ as required.

b.
$$\begin{aligned}\mathrm{var}(u_i) = E(u_i^2) &= \left(1 - x_i'\beta\right)^2\pi_i + \left(-x_i'\beta\right)^2(1 - \pi_i)\\ &= \left[1 - 2x_i'\beta + \left(x_i'\beta\right)^2\right]\pi_i + \left(x_i'\beta\right)^2(1 - \pi_i)\\ &= \pi_i - 2x_i'\beta\pi_i + (x_i'\beta)^2\\ &= \pi_i - \pi_i^2 = \pi_i(1 - \pi_i) = x_i'\beta\left(1 - x_i'\beta\right)\end{aligned}$$

using the fact that $\pi_i = x_i'\beta$.

13.2 a. Since there are no slopes and only a constant, $x_i'\beta = \alpha$ and (13.16) becomes $\log\ell = \sum_{i=1}^{n}\{y_i \log F(\alpha) + (1 - y_i)\log[1 - F(\alpha)]\}$ differentiating with respect to α we get

$$\frac{\partial \log \ell}{\partial \alpha} = \sum_{i=1}^{n}\frac{y_i}{F(\alpha)}\cdot f(\alpha) + \sum_{i=1}^{n}\frac{(1 - y_i)}{1 - F(\alpha)}(-f(\alpha)).$$

Setting this equal to zero yields $\sum_{i=1}^{n}(y_i - F(\alpha))f(\alpha) = 0$.

Therefore, $\hat{F}(\alpha) = \sum_{i=1}^{n} y_i/n = \bar{y}$. This is the proportion of the sample with $y_i = 1$.

B.H. Baltagi, *Solutions Manual for Econometrics*, Springer Texts in Business and Economics, DOI 10.1007/978-3-642-54548-1_13,

b. Using $\hat{F}(\alpha) = \bar{y}$, the value of the maximized likelihood, from (13.16), is

$$\log \ell_r = \sum_{i=1}^{n} \{y_i \log \bar{y} + (1-y_i)\log(1-\bar{y})\} = n\bar{y}\log\bar{y} + (n-n\bar{y})\log(1-\bar{y})$$

$$= n[\bar{y}\log\bar{y} + (1-\bar{y})\log(1-\bar{y})] \qquad \text{as required.}$$

c. For the empirical example in Sect. 13.9, we know that $\bar{y} = 218/595 = 0.366$. Substituting in (13.33) we get,

$$\log \ell_r = n[0.366\log 0.366 + (1-0.366)\log(1-0.366)] = -390.918.$$

13.3 Union participation example. See Tables 13.3–13.5. These were run using EViews.

a. OLS ESTIMATION

LS // Dependent Variable is UNION

Sample: 1 595

Included observations: 595

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.195872	0.227010	5.267922	0.0000
EX	-0.001974	0.001726	-1.143270	0.2534
WKS	-0.017809	0.003419	-5.209226	0.0000
OCC	0.318118	0.046425	6.852287	0.0000
IND	0.030048	0.038072	0.789229	0.4303
SOUTH	-0.170130	0.039801	-4.274471	0.0000
SMSA	0.084522	0.038464	2.197419	0.0284
MS	0.098953	0.063781	1.551453	0.1213
FEM	-0.108706	0.079266	-1.371398	0.1708
ED	-0.016187	0.008592	-1.883924	0.0601
BLK	0.050197	0.071130	0.705708	0.4807

R-squared	0.233548	Mean dependent var	0.366387
Adjusted R-squared	0.220424	S.D. dependent var	0.482222
S.E. of regression	0.425771	Akaike info criterion	-1.689391
Sum squared resid	105.8682	Schwarz criterion	-1.608258
Log likelihood	-330.6745	F-statistic	17.79528
Durbin-Watson stat	1.900963	Prob(F-statistic)	0.000000

LOGIT ESTIMATION

LOGIT // Dependent Variable is UNION

Sample: 1 595

Included observations: 595

Convergence achieved after 4 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.380828	1.338629	3.272624	0.0011
EX	-0.011143	0.009691	-1.149750	0.2507
WKS	-0.108126	0.021428	-5.046037	0.0000
OCC	1.658222	0.264456	6.270325	0.0000
IND	0.181818	0.205470	0.884888	0.3766
SOUTH	-1.044332	0.241107	-4.331411	0.0000
SMSA	0.448389	0.218289	2.054110	0.0404
MS	0.604999	0.365043	1.657336	0.0980
FEM	-0.772222	0.489665	-1.577040	0.1153
ED	-0.090799	0.049227	-1.844501	0.0656
BLK	0.355706	0.394794	0.900992	0.3680

Log likelihood	-312.3367
Obs with Dep=1	218
Obs with Dep=0	377

Variable	Mean All	Mean D=1	Mean D=0
C	1.000000	1.000000	1.000000
EX	22.85378	23.83028	22.28912
WKS	46.45210	45.27982	47.12997
OCC	0.512605	0.766055	0.366048
IND	0.405042	0.513761	0.342175
SOUTH	0.292437	0.197248	0.347480
SMSA	0.642017	0.646789	0.639257
MS	0.805042	0.866972	0.769231
FEM	0.112605	0.059633	0.143236
ED	12.84538	11.84862	13.42175
BLK	0.072269	0.082569	0.066313

PROBIT ESTIMATION

PROBIT // Dependent Variable is UNION

Sample: 1 595

Included observations: 595

Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.516784	0.762606	3.300242	0.0010
EX	-0.006932	0.005745	-1.206501	0.2281
WKS	-0.060829	0.011785	-5.161707	0.0000
OCC	0.955490	0.152136	6.280522	0.0000
IND	0.092827	0.122773	0.756089	0.4499
SOUTH	-0.592739	0.139100	-4.261243	0.0000
SMSA	0.260701	0.128629	2.026756	0.0431
MS	0.350520	0.216282	1.620664	0.1056
FEM	-0.407026	0.277034	-1.469226	0.1423
ED	-0.057382	0.028842	-1.989533	0.0471
BLK	0.226482	0.228843	0.989683	0.3227

Log likelihood	-313.3795
Obs with Dep=1	218
Obs with Dep=0	377

d. Dropping the industry variable (IND).

OLS ESTIMATION

LS // Dependent Variable is UNION

Sample: 1 595

Included observations: 595

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.216753	0.225390	5.398425	0.0000
EX	-0.001848	0.001718	-1.075209	0.2827
WKS	-0.017874	0.003417	-5.231558	0.0000
OCC	0.322215	0.046119	6.986568	0.0000
SOUTH	-0.173339	0.039580	-4.379418	0.0000
SMSA	0.085043	0.038446	2.212014	0.0274
MS	0.100697	0.063722	1.580267	0.1146
FEM	-0.114088	0.078947	-1.445122	0.1490
ED	-0.017021	0.008524	-1.996684	0.0463
BLK	0.048167	0.071061	0.677822	0.4982

R-squared	0.232731	Mean dependent var	0.366387
Adjusted R-squared	0.220927	S.D. dependent var	0.482222
S.E. of regression	0.425634	Akaike info criterion	-1.691687
Sum squared resid	105.9811	Schwarz criterion	-1.617929
Log likelihood	-330.9916	F-statistic	19.71604
Durbin-Watson stat	1.907714	Prob(F-statistic)	0.000000

LOGIT ESTIMATION

LOGIT // Dependent Variable is UNION

Sample: 1 595

Included observations: 595

Convergence achieved after 4 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.492957	1.333992	3.368053	0.0008
EX	-0.010454	0.009649	-1.083430	0.2791
WKS	-0.107912	0.021380	-5.047345	0.0000
OCC	1.675169	0.263654	6.353652	0.0000
SOUTH	-1.058953	0.240224	-4.408193	0.0000
SMSA	0.449003	0.217955	2.060074	0.0398
MS	0.618511	0.365637	1.691599	0.0913
FEM	-0.795607	0.489820	-1.624285	0.1049
ED	-0.096695	0.048806	-1.981194	0.0480
BLK	0.339984	0.394027	0.862845	0.3886

Log likelihood	-312.7267
Obs with Dep=1	218
Obs with Dep=0	377

Variable	Mean All	Mean D=1	Mean D=0
C	1.000000	1.000000	1.000000
EX	22.85378	23.83028	22.28912
WKS	46.45210	45.27982	47.12997
OCC	0.512605	0.766055	0.366048
SOUTH	0.292437	0.197248	0.347480
SMSA	0.642017	0.646789	0.639257
MS	0.805042	0.866972	0.769231
FEM	0.112605	0.059633	0.143236
ED	12.84538	11.84862	13.42175
BLK	0.072269	0.082569	0.066313

PROBIT ESTIMATION

PROBIT // Dependent Variable is UNION

Sample: 1 595

Included observations: 595

Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.570491	0.759181	3.385875	0.0008
EX	-0.006590	0.005723	-1.151333	0.2501
WKS	-0.060795	0.011777	-5.162354	0.0000
OCC	0.967972	0.151305	6.397481	0.0000
SOUTH	-0.601050	0.138528	-4.338836	0.0000
SMSA	0.261381	0.128465	2.034640	0.0423
MS	0.357808	0.216057	1.656085	0.0982
FEM	-0.417974	0.276501	-1.511657	0.1312
ED	-0.060082	0.028625	-2.098957	0.0362
BLK	0.220695	0.228363	0.966423	0.3342

Log likelihood -313.6647
Obs with Dep = 1 218
Obs with Dep = 0 377

f. The restricted regressions omitting IND, FEM and BLK are given below:

LS // Dependent Variable is UNION
Sample: 1 595
Included observations: 595

Variable	Coefficient	Std.Error	t-Statistic	Prob.
C	1.153900	0.218771	5.274452	0.0000
EX	-0.001840	0.001717	-1.071655	0.2843
WKS	-0.017744	0.003412	-5.200421	0.0000
OCC	0.326411	0.046051	7.088110	0.0000
SOUTH	-0.171713	0.039295	-4.369868	0.0000
SMSA	0.086076	0.038013	2.264382	0.0239
MS	0.158303	0.045433	3.484351	0.0005
ED	-0.017204	0.008507	-2.022449	0.0436

R-squared	0.229543	Mean dependent var	0.366387
Adjusted R-squared	0.220355	S.D. dependent var	0.482222
S.E. of regression	0.425790	Akaike info criterion	-1.694263
Sum squared resid	106.4215	Schwarz criterion	-1.635257
Log likelihood	-332.2252	F-statistic	24.98361
Durbin-Watson stat	1.912059	Prob(F-statistic)	0.000000

LOGIT // Dependent Variable is UNION
Sample: 1 595
Included observations: 595
Convergence achieved after 4 iterations

Variable	Coefficient	Std.Error	t-Statistic	Prob.
C	4.152595	1.288390	3.223088	0.0013
EX	-0.011018	0.009641	-1.142863	0.2536
WKS	-0.107116	0.021215	-5.049031	0.0000
OCC	1.684082	0.262193	6.423059	0.0000
SOUTH	-1.043629	0.237769	-4.389255	0.0000
SMSA	0.459707	0.215149	2.136687	0.0330
MS	0.975711	0.272560	3.579800	0.0004
ED	-0.100033	0.048507	-2.062229	0.0396

Log likelihood	-314.2744
Obs with Dep=1	218
Obs with Dep=0	377

Variable	Mean All	Mean D=1	Mean D=0
C	1.000000	1.000000	1.000000
EX	22.85378	23.83028	22.28912
WKS	46.45210	45.27982	47.12997
OCC	0.512605	0.766055	0.366048
SOUTH	0.292437	0.197248	0.347480
SMSA	0.642017	0.646789	0.639257
MS	0.805042	0.866972	0.769231
ED	12.84538	11.84862	13.42175

PROBIT // Dependent Variable is UNION
Sample: 1 595
Included observations: 595
Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.411706	0.741327	3.253228	0.0012
EX	-0.006986	0.005715	-1.222444	0.2220
WKS	-0.060491	0.011788	-5.131568	0.0000
OCC	0.971984	0.150538	6.456745	0.0000
SOUTH	-0.580959	0.136344	-4.260988	0.0000
SMSA	0.273201	0.126988	2.151388	0.0319
MS	0.545824	0.155812	3.503105	0.0005
ED	-0.063196	0.028464	-2.220210	0.0268

Log likelihood	-315.1770
Obs with Dep=1	218
Obs with Dep=0	377

13.4 Occupation regression.

a. OLS Estimation

LS // Dependent Variable is OCC

Sample: 1 595

Included observations: 595

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.111943	0.182340	11.58245	0.0000
ED	-0.111499	0.006108	-18.25569	0.0000
WKS	-0.001510	0.003044	-0.496158	0.6200
EX	-0.002870	0.001533	-1.872517	0.0616
SOUTH	-0.068631	0.035332	-1.942452	0.0526
SMSA	-0.079735	0.034096	-2.338528	0.0197
IND	0.091688	0.033693	2.721240	0.0067
MS	0.006271	0.056801	0.110402	0.9121
FEM	-0.064045	0.070543	-0.907893	0.3643
BLK	0.068514	0.063283	1.082647	0.2794

R-squared	0.434196	Mean dependent var	0.512605
Adjusted R-squared	0.425491	S.D. dependent var	0.500262
S.E. of regression	0.379180	Akaike info criterion	-1.922824
Sum squared resid	84.10987	Schwarz criterion	-1.849067
Log likelihood	-262.2283	F-statistic	49.88075
Durbin-Watson stat	1.876105	Prob(F-statistic)	0.000000

LOGIT ESTIMATION

LOGIT // Dependent Variable is OCC

Sample: 1 595

Included observations: 595

Convergence achieved after 5 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	11.62962	1.581601	7.353069	0.0000
ED	-0.806320	0.070068	-11.50773	0.0000
WKS	-0.008424	0.023511	-0.358297	0.7203
EX	-0.017610	0.011161	-1.577893	0.1151
SOUTH	-0.349960	0.260761	-1.342073	0.1801
SMSA	-0.601945	0.247206	-2.434995	0.0152
IND	0.689620	0.241028	2.861157	0.0044

MS	-0.178865	0.417192	-0.428735	0.6683
FEM	-0.672117	0.503002	-1.336212	0.1820
BLK	0.333307	0.441064	0.755687	0.4501

Log likelihood	-244.2390
Obs with Dep=1	305
Obs with Dep=0	290

Variable	Mean All	Mean D=1	Mean D=0
C	1.000000	1.000000	1.000000
ED	12.84538	11.10164	14.67931
WKS	46.45210	46.40984	46.49655
EX	22.85378	23.88525	21.76897
SOUTH	0.292437	0.298361	0.286207
SMSA	0.642017	0.554098	0.734483
IND	0.405042	0.524590	0.279310
MS	0.805042	0.816393	0.793103
FEM	0.112605	0.095082	0.131034
BLK	0.072269	0.091803	0.051724

PROBIT ESTIMATION

PROBIT // Dependent Variable is OCC

Sample: 1 595

Included observations: 595

Convergence achieved after 4 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.416131	0.847427	7.571312	0.0000
ED	-0.446740	0.034458	-12.96473	0.0000
WKS	-0.003574	0.013258	-0.269581	0.7876
EX	-0.010891	0.006336	-1.718878	0.0862
SOUTH	-0.240756	0.147920	-1.627608	0.1041
SMSA	-0.327948	0.139849	-2.345016	0.0194
IND	0.371434	0.135825	2.734658	0.0064
MS	-0.097665	0.245069	-0.398522	0.6904
FEM	-0.358948	0.296971	-1.208697	0.2273
BLK	0.215257	0.252219	0.853453	0.3938

Log likelihood	-246.6581
Obs with Dep=1	305
Obs with Dep=0	290

13.5 *Truncated Uniform Density.*

$$\Pr\left[x > -\frac{1}{2}\right] = \int_{-1/2}^{1} \frac{1}{2}dx = \frac{1}{2}\left[\frac{3}{2}\right] = \frac{3}{4}. \text{ So that}$$

$$f\left(x/x > -\frac{1}{2}\right) = \frac{f(x)}{\Pr\left[x > -\frac{1}{2}\right]} = \frac{1/2}{3/4} = \frac{2}{3} \text{ for } -\frac{1}{2} < x < 1.$$

b. $$E\left[x/x > -\frac{1}{2}\right] = \int_{-1/2}^{1} xf\left(x/x > -\frac{1}{2}\right)dx = \int_{-1/2}^{1} x\cdot\frac{2}{3}dx = \frac{2}{3}\cdot\frac{1}{2}[x^2]_{-1/2}^{1}$$

$$= \frac{1}{3}\left[1 - \frac{1}{4}\right] = \frac{1}{4}$$

$$E(x) = \int_{-1}^{1} x\cdot\frac{1}{2}dx = \frac{1}{2}\cdot\frac{1}{2}\cdot[1-(-1)^2] = 0.$$

Therefore, $\frac{1}{4} = E\left[x/x > -\frac{1}{2}\right] > E(x) = 0$. Because the density is truncated from below, the new mean shifts to the right.

c. $$E(x^2) = \int_{-1}^{1} x^2\cdot\frac{1}{2}dx = \frac{1}{2}\cdot\frac{1}{3}\cdot[x^3]_{-1}^{1} = \frac{1}{3}$$

$$\text{var}(x) = E(x^2) - (E(x))^2 = E(x^2) = \frac{1}{3}$$

$$E\left(x^2/x > -\frac{1}{2}\right) = \int_{-1/2}^{1} x^2\cdot\frac{2}{3}\cdot dx = \frac{2}{3}\cdot\frac{1}{3}[x^3]_{-1/2}^{1} = \frac{2}{9}\left[1+\frac{1}{8}\right] = \frac{1}{4}$$

$$\text{var}\left(x/x > -\frac{1}{2}\right) = E\left(x^2/x > -\frac{1}{2}\right) - \left(E\left[x/x > -\frac{1}{2}\right]\right)^2 = \frac{1}{4} - \left(\frac{1}{4}\right)^2$$

$$= \frac{1}{4} - \frac{1}{16} = \frac{3}{16} < \text{var}(x) = \frac{1}{3}.$$

Therefore, as expected, truncation reduces the variance.

13.6 *Truncated Normal Density.*

a. From the Appendix, Eq. (A.1), using $c = 1,\ \mu = 1,\ \sigma^2 = 1$ and

$\Phi(0) = \frac{1}{2}$, we get, $f(x/x > 1) = \dfrac{\phi(x-1)}{1-\Phi(0)} = 2\phi(x-1)$ for $x > 1$

Similarly, using Eq. (A.2), for $c = 1,\ \mu = 1$ and $\sigma^2 = 1$ with $\Phi(0) = \frac{1}{2}$,

we get $f(x/x < 1) = \dfrac{\phi(x-1)}{\Phi(0)} = 2\phi(x-1)$ for $x < 1$

b. The conditional mean is given in (A.3) and for this example we get

$$E(x/x > 1) = 1 + 1 \cdot \frac{\phi(c^*)}{1 - \Phi(c^*)} = 1 + \frac{\phi(0)}{1 - \Phi(0)} = 1 + 2\phi(0)$$
$$= 1 + \frac{2}{\sqrt{2\pi}} \approx 1.8 > 1 = E(x)$$

with $c^* = \frac{c - \mu}{\sigma} = \frac{1-1}{1} = 0$. Similarly, using (A.4) we get,

$$E(x/x < 1) = 1 - 1 \cdot \frac{\phi(c^*)}{\Phi(c^*)} = 1 - \frac{\phi(0)}{1/2} = 1 - 2\phi(0) = 1 - \frac{2}{\sqrt{2\pi}}$$
$$\approx 1 - 0.8 = 0.2 < 1 = E(x).$$

c. From (A.5) we get, $\text{var}(x/x > 1) = 1(1 - \delta(c^*)) = 1 - \delta(0)$ where

$$\delta(0) = \frac{\phi(0)}{1 - \Phi(0)} \left[\frac{\phi(0)}{1 - \Phi(0)} - 0 \right] \qquad \text{for} \qquad x > 1$$
$$= 2\phi(0)[2\phi(0)] = 4\phi^2(0) = \frac{4}{2\pi} = \frac{2}{\pi} = 0.64 \qquad \text{for} \qquad x > 1$$

From (A.6), we get $\text{var}(x/x > 1) = 1 - \delta(0)$ where

$$\delta(0) = \frac{-\phi(0)}{\Phi(0)} \left[\frac{-\phi(0)}{\Phi(0)} - 0 \right] = 4\phi^2(0) = \frac{4}{2\pi} = 0.64 \qquad \text{for} \qquad x < 1.$$

Both conditional truncated variances are less than the unconditional $\text{var}(x) = 1$.

13.7 *Censored Normal Distribution*.

a. From the Appendix we get,

$$E(y) = \Pr[y = c]\, E(y/y = c) + \Pr[y > c]\, E(y/y > c)$$
$$= c\Phi(c^*) + (1 - \Phi(c^*))E(y^*/y^* > c)$$
$$= c\Phi(c^*) + (1 - \Phi(c^*)) \left[\mu + \sigma \frac{\phi(c^*)}{1 - \Phi(c^*)} \right]$$

where $E(y^*/y^* > c)$ is obtained from the mean of a truncated normal density, see (A.3).

b. Using the result on conditional variance given in Chap. 2 we get, $\text{var}(y) =$ E(conditional variance) + var(conditional mean). But

$$E(\text{conditional variance}) = P[y = c]\, \text{var}(y/y = c) + P[y > c]\, \text{var}(y/y > c)$$

$$= \Phi(c^*) \cdot 0 + (1 - \Phi(c^*))\sigma^2(1 - \delta(c^*))$$

where $\mathrm{var}(y/y > c)$ is given by (A.5).

$$\begin{aligned}\text{var(conditional mean)} &= P[y = c] \cdot (c - E(y))^2 + \Pr(y > c)[E(y/y{>}c) - E(y)]^2 \\ &= \Phi(c^*)(c - E(y))^2 + [1 - \Phi(c^*)][E(y/y > c) - E(y)]^2\end{aligned}$$

where $E(y)$ is given by (A.7) and $E(y/y > c)$ is given by (A.3). This gives

$$\begin{aligned}\text{var(conditional mean)} = {} & \Phi(c^*)\left\{c - c\Phi(c^*) - (1 - \Phi(c^*))\right. \\ & \left.\left[\mu + \sigma\frac{\phi(c^*)}{1 - \Phi(c^*)}\right]\right\}^2 \\ & + [1 - \Phi(c^*)]\left\{\mu + \sigma\frac{\phi(c^*)}{1 - \Phi(c^*)} - c\Phi(c^*)\right. \\ & \left. - (1 - \Phi(c^*))\left[\mu + \frac{\sigma\phi(c^*)}{1 - \Phi(c^*)}\right]\right\}^2 \\ = {} & \Phi(c^*)\left\{(1 - \Phi(c^*))\left(c - \mu - \sigma\frac{\phi(c^*)}{1 - \Phi(c^*)}\right)\right\}^2 \\ & + (1 - \Phi(c^*))\left\{\Phi(c^*)\left(c - \mu - \sigma\frac{\phi(c^*)}{1 - \Phi(c^*)}\right)\right\}^2 \\ = {} & \{\Phi(c^*)[1 - \Phi(c^*)]^2 \\ & + (1 - \Phi(c^*))\Phi^2(c^*)\}\sigma^2\left[c^* - \frac{\phi(c^*)}{1 - \Phi(c^*)}\right]^2 \\ = {} & \Phi(c^*)[1 - \Phi(c^*)]\sigma^2\left[c^* - \frac{\phi(c^*)}{1 - \Phi(c^*)}\right]^2.\end{aligned}$$

Summing the two terms, we get the expression in (A.8)

$$\mathrm{var}(y) = \sigma^2(1-\Phi(c^*))\left\{1-\delta(c^*)+\left[c^*-\frac{\phi(c^*)}{1-\Phi(c^*)}\right]^2\Phi(c^*)\right\}.$$

c. For the special case where $c = 0$ and $c^* = -\mu/\sigma$, the mean in (A.7) simplifies to

$$E(y) = 0.\Phi(c^*) + \left(1-\Phi\left(-\frac{\mu}{\sigma}\right)\right)\left[\mu+\frac{\sigma\phi(-\mu/\sigma)}{1-\Phi(-\mu/\sigma)}\right]$$

$$= \Phi\left(\frac{\mu}{\sigma}\right)\left[\mu+\sigma\frac{\phi(\mu/\sigma)}{\Phi(\mu/\sigma)}\right]$$

as required. Similarly, from part (b), using $c^* = -\mu/\sigma$ and $\Phi(-\mu/\sigma) = \Phi(\mu/\sigma)$, we get $\mathrm{var}(y) = \sigma^2\Phi\left(\frac{\mu}{\sigma}\right)\left[1-\delta\left(-\frac{\mu}{\sigma}\right)+\left(-\frac{\mu}{\sigma}-\frac{\phi(-\mu/\sigma)}{1-\Phi(-\mu/\sigma)}\right)^2\Phi\left(-\frac{\mu}{\sigma}\right)\right]$ where $\delta\left(-\frac{\mu}{\sigma}\right) = \frac{\phi(-\mu/\sigma)}{1-\Phi(-\mu/\sigma)}$

$$\left[\frac{\phi(-\mu/\sigma)}{1-\Phi(-\mu/\sigma)}+\frac{\mu}{\sigma}\right] = \frac{\phi(\mu/\sigma)}{\Phi(\mu/\sigma)}\left[\frac{\phi(-\mu/\sigma)}{\Phi(\mu/\sigma)}+\frac{\mu}{\sigma}\right].$$

13.8 Fixed vs. adjustable mortgage rates. This is based on Dhillon et al. (1987).

a. The OLS regression of Y on all variables in the data set is given below. This was done using EViews. The $R^2 = 0.434$ and the F-statistic for the significance of all slopes is equal to 3.169. This is distributed as F(15,62) under the null hypothesis. This has a p-value of 0.0007. Therefore, we reject H_o and we conclude that this is a significant regression. As explained in Sect. 13.6, using BRMR this also rejects the insignificance of all slopes in the logit specification.

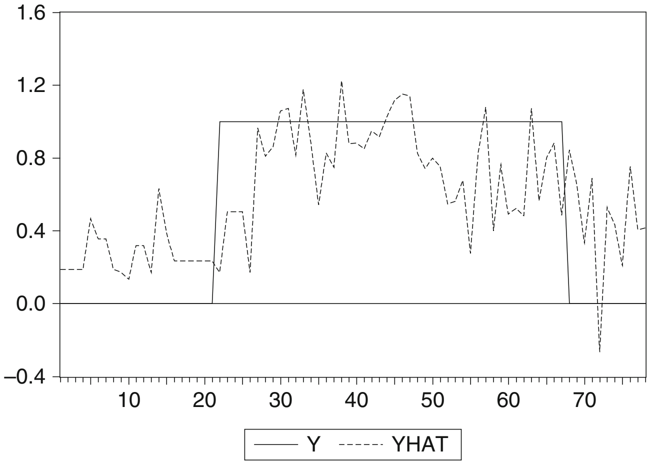
Unrestricted Least Squares

LS // Dependent Variable is Y
Sample: 1 78
Included observations: 78

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.272832	1.411806	0.901563	0.3708
BA	0.000398	0.007307	0.054431	0.9568
BS	0.017084	0.020365	0.838887	0.4048
NW	-0.036932	0.025320	-1.458609	0.1497
FI	-0.221726	0.092813	-2.388949	0.0200
PTS	0.178963	0.091050	1.965544	0.0538
MAT	0.214264	0.202497	1.058108	0.2941
MOB	0.020963	0.009194	2.279984	0.0261
MC	0.189973	0.150816	1.259635	0.2125
FTB	-0.013857	0.136127	-0.101797	0.9192
SE	0.188284	0.360196	0.522728	0.6030
YLD	0.656227	0.366117	1.792399	0.0779
MARG	0.129127	0.054840	2.354621	0.0217
CB	0.172202	0.137827	1.249403	0.2162
STL	-0.001599	0.005994	-0.266823	0.7905
LA	-0.001761	0.007801	-0.225725	0.8222

R-squared	0.433996	Mean dependent var	0.589744
Adjusted R-squared	0.297059	S.D. dependent var	0.495064
S.E. of regression	0.415069	Akaike info criter	-1.577938
Sum squared resid	10.68152	Schwarz criterion	-1.094510
Log likelihood	-33.13764	F-statistic	3.169321
Durbin-Watson stat	0.905968	Prob(F-statistic)	0.000702

Plot of Y and YHAT



b. The URSS from part (a) is 10.6815 while the RRSS by including only the cost variables is 14.0180 as shown in the enclosed output from EViews. The Chow-F statistic for insignificance of 10 personal characteristics variables is

$$F = \frac{(14.0180 - 10.6815)/10}{10.6815/62} = 1.9366$$

which is distributed as F(10,62) under the null hypothesis. This has a 5% critical value of 1.99. Hence, we cannot reject H_o. The principal agent theory suggests that personal characteristics are important in making this mortgage choice. Briefly, this theory suggests that information is asymmetric and the borrower knows things about himself or herself that the lending institution does not. Not rejecting H_o does not provide support for the principal agent theory.

TESTING THE EFFICIENT MARKET HYPOTHESIS WITH THE LINEAR PROBABILITY MODEL

Restricted Least Squares

LS // Dependent Variable is Y
Sample: 1 78
Included observations: 78

Variable	Coefficient	Std. Error	t-Statistic	Prob.
FI	-0.237228	0.078592	-3.018479	0.0035
MARG	0.127029	0.051496	2.466784	0.0160
YLD	0.889908	0.332037	2.680151	0.0091
PTS	0.054879	0.072165	0.760465	0.4495
MAT	0.069466	0.196727	0.353108	0.7250
C	1.856435	1.289797	1.439324	0.1544

R-squared	0.257199	Mean dependent var	0.589744
Adjusted R-squared	0.205616	S.D. dependent var	0.495064
S.E. of regression	0.441242	Akaike info criter	-1.562522
Sum squared resid	14.01798	Schwarz criterion	-1.381236
Log likelihood	-43.73886	F-statistic	4.986087
Durbin-Watson stat	0.509361	Prob(F-statistic)	0.000562

c. The logit specification output using EViews is given below. The unrestricted log-likelihood is equal to -30.8963. The restricted specification output is also given showing a restricted log-likelihood of -41.4729. Therefore, the LR test statistic is given by $LR = 2(41.4729 - 30.8963) = 21.1532$ which is distributed as χ^2_{10} under the null hypothesis. This is significant given that the 5% critical value of χ^2_{10} is 18.31. This means that the logit specification does not reject the principal agent theory as personal characteristics are not jointly insignificant.

TESTING THE EFFICIENT MARKET HYPOTHESIS WITH THE LOGIT MODEL

Unrestricted Logit Model

LOGIT // Dependent Variable is Y
Sample: 1 78
Included observations: 78
Convergence achieved after 5 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.238872	10.47875	0.404521	0.6872
BA	0.010478	0.075692	0.138425	0.8904
BS	0.198251	0.172444	1.149658	0.2547
NW	-0.244064	0.185027	-1.319072	0.1920
FI	-1.717497	0.727707	-2.360149	0.0214
PTS	1.499799	0.719917	2.083294	0.0414
MAT	2.057067	1.631100	1.261153	0.2120
MOB	0.153078	0.097000	1.578129	0.1196
MC	1.922943	1.182932	1.625575	0.1091
FTB	-0.110924	0.983688	-0.112763	0.9106
SE	2.208505	2.800907	0.788496	0.4334
YLD	4.626702	2.919634	1.584686	0.1181
MARG	1.189518	0.485433	2.450426	0.0171
CB	1.759744	1.242104	1.416744	0.1616
STL	-0.031563	0.051720	-0.610265	0.5439
LA	-0.022067	0.061013	-0.361675	0.7188

Log likelihood	-30.89597
Obs with Dep=1	46
Obs with Dep=0	32

Variable	Mean All	Mean D=1	Mean D=0
C	1.000000	1.000000	1.000000
BA	36.03846	35.52174	36.78125
BS	16.44872	15.58696	17.68750
NW	3.504013	2.075261	5.557844
FI	13.24936	13.02348	13.57406
PTS	1.497949	1.505217	1.487500
MAT	1.058333	1.027609	1.102500
MOB	4.205128	4.913043	3.187500
MC	0.602564	0.695652	0.468750
FTB	0.615385	0.521739	0.750000
SE	0.102564	0.043478	0.187500
YLD	1.606410	1.633261	1.567813
MARG	2.291923	2.526304	1.955000
CB	0.358974	0.478261	0.187500
STL	13.42218	11.72304	15.86469
LA	5.682692	4.792174	6.962812

Restricted Logit Model

LOGIT // Dependent Variable is Y
Sample: 1 78
Included observations: 78
Convergence achieved after 4 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
FI	-1.264608	0.454050	-2.785172	0.0068
MARG	0.717847	0.313845	2.287265	0.0251
YLD	4.827537	1.958833	2.464497	0.0161
PTS	0.359033	0.423378	0.848019	0.3992
MAT	0.550320	1.036613	0.530883	0.5971
C	6.731755	7.059485	0.953576	0.3435

Log likelihood	-41.47292
Obs with Dep=1	46
Obs with Dep=0	32

Variable	Mean All	Mean D=1	Mean D=0
FI	13.24936	13.02348	13.57406
MARG	2.291923	2.526304	1.955000
YLD	1.606410	1.633261	1.567813
PTS	1.497949	1.505217	1.487500
MAT	1.058333	1.027609	1.102500
C	1.000000	1.000000	1.000000

d. Similarly, the probit specification output using EViews is given below. The unrestricted log-likelihood is equal to -30.7294. The restricted log-likelihood is -41.7649. Therefore, the LR test statistic is given by LR $= 2(41.7649 - 30.7294) = 22.0710$ which is distributed as χ^2_{10} under the null hypothesis. This is significant given that the 5% critical value of χ^2_{10} is 18.31. This means that the probit specification does not reject the principal agent theory as personal characteristics are not jointly insignificant.

TESTING THE EFFICIENT MARKET HYPOTHESIS WITH THE PROBIT MODEL

Unrestricted Probit Model

PROBIT // Dependent Variable is Y
Sample: 1 78
Included observations: 78
Convergence achieved after 5 iterations

Variable	Coefficien	Std. Error	t-Statistic	Prob.
C	3.107820	5.954673	0.521913	0.6036
BA	0.003978	0.044546	0.089293	0.9291
BS	0.108267	0.099172	1.091704	0.2792
NW	-0.128775	0.103438	-1.244943	0.2178
FI	-1.008080	0.418160	-2.410750	0.0189
PTS	0.830273	0.379895	2.185533	0.0326
MAT	1.164384	0.924018	1.260131	0.2123
MOB	0.093034	0.056047	1.659924	0.1020
MC	1.058577	0.653234	1.620518	0.1102
FTB	-0.143447	0.550471	-0.260589	0.7953
SE	1.127523	1.565488	0.720237	0.4741
YLD	2.525122	1.590796	1.587332	0.1175
MARG	0.705238	0.276340	2.552069	0.0132
CB	1.066589	0.721403	1.478493	0.1443
STL	-0.016130	0.029303	-0.550446	0.5840
LA	-0.014615	0.035920	-0.406871	0.6855

Log likelihood	-30.72937
Obs with Dep=1	46
Obs with Dep=0	32

Restricted Probit Model

```
=========================================================
PROBIT // Dependent Variable is Y
Sample: 1 78
Included observations: 78
Convergence achieved after 3 iterations
=========================================================
Variable      Coefficient    Std. Error    t-Statistic    Prob.
=========================================================
FI            -0.693584      0.244631      -2.835225      0.0059
MARG          0.419997       0.175012      2.399811       0.0190
YLD           2.730187       1.099487      2.483146       0.0154
PTS           0.235534       0.247390      0.952076       0.3442
MAT           0.221568       0.610572      0.362886       0.7178
C             3.536657       4.030251      0.877528       0.3831
=========================================================
Log likelihood    -41.76443
Obs with Dep=1        46
Obs with Dep=0        32
```

13.13 *Problem Drinking and Employment.* The following Stata output replicates the OLS results given in Table 5 of Mullahy and Sindelar (1996, p. 428) for males. The first regression is for employment, given in column 1 of Table 5 of the paper, and the second regression is for unemployment, given in column 3 of Table 5 of the paper. Robust standard errors are reported.

```
. reg emp hvdrnk90 ue88 age agesq educ married famsize white hlstat1 hlstat2
hlstat3 hlstat4 region1 region2 region3 msa1 msa2 q1 q2 q3, robust

Regression with robust standard errors            Number of obs =    9822
                                                  F (20, 9801)  =   46.15
                                                  Prob > F      =  0.0000
                                                  R-squared     =  0.1563
                                                  Root MSE      =  .27807

------------------------------------------------------------------------------
         |              Robust
     emp |    Coef.    Std. Err.     t     P>|t|     [95% Conf. Interval]
------------------------------------------------------------------------------
hvdrnk90 | -.0155071   .0101891   -1.52   0.128    -.0354798    .0044657
    ue88 | -.0090938   .0022494   -4.04   0.000     -.013503   -.0046846
     age |  .0162668   .0029248    5.56   0.000     .0105336    .0220001
   agesq | -.0002164   .0000362   -5.98   0.000    -.0002873   -.0001455
```

educ	.0078258	.0011271	6.94	0.000	.0056166	.0100351
married	.0505682	.0098396	5.14	0.000	.0312805	.0698558
famsize	.0020612	.0021796	0.95	0.344	-.0022113	.0063336
white	.0773332	.0104289	7.42	0.000	.0568905	.097776
hlstat1	.5751898	.0306635	18.76	0.000	.5150831	.6352965
hlstat2	.5728	.0306427	18.69	0.000	.512734	.632866
hlstat3	.537617	.0308845	17.41	0.000	.4770769	.598157
hlstat4	.3947391	.0354291	11.14	0.000	.3252908	.4641874
region1	-.0013608	.0094193	-0.14	0.885	-.0198247	.017103
region2	.0050446	.0084215	0.60	0.549	-.0114633	.0215526
region3	.0254332	.0081999	3.10	0.002	.0093596	.0415067
msa1	-.0159492	.0083578	-1.91	0.056	-.0323322	.0004337
msa2	.0073081	.0072395	1.01	0.313	-.0068827	.0214989
q1	-.0155891	.0079415	-1.96	0.050	-.0311561	-.000022
q2	-.0068915	.0077786	-0.89	0.376	-.0221392	.0083561
q3	-.0035867	.0078474	-0.46	0.648	-.0189692	.0117957
_cons	-.0957667	.0623045	-1.54	0.124	-.2178964	.0263631

```
. reg unemp hvdrnk90 ue88 age agesq educ married famsize white hlstat1 hlstat2
hlstat3 hlstat4 region1 region2 region3 msa1 msa2 q1 q2 q3, robust
```

Regression with robust standard errors

Number of obs	=	9822
F(20, 9801)	=	3.37
Prob > F	=	0.0000
R-squared	=	0.0099
Root MSE	=	.17577

emp	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
hvdrnk90	.0100022	.0066807	1.50	0.134	-.0030934	.0230977
ue88	.0045029	.0014666	3.07	0.002	.0016281	.0073776
age	-.0014753	.0017288	-0.85	0.393	-.0048641	.0019134
agesq	.0000123	.0000206	0.60	0.551	-.0000281	.0000527
educ	-.0028141	.0006307	-4.46	0.000	-.0040504	-.0015777
married	-.0092854	.0060161	-1.54	0.123	-.0210782	.0025073
famsize	.0003859	.0013719	0.28	0.778	-.0023033	.0030751
white	-.0246801	.0063618	-3.88	0.000	-.0371506	-.0122096
hlstat1	.0150194	.0113968	1.32	0.188	-.0073206	.0373594
hlstat2	.0178594	.0114626	1.56	0.119	-.0046097	.0403285
hlstat3	.0225153	.0116518	1.93	0.053	-.0003245	.0453552
hlstat4	.0178865	.0136228	1.31	0.189	-.0088171	.0445901
region1	.0007911	.005861	0.13	0.893	-.0106977	.01228
region2	-.0029056	.0053543	-0.54	0.587	-.0134011	.0075898
region3	-.0065005	.005095	-1.28	0.202	-.0164877	.0034868
msa1	-.0008801	.0052004	-0.17	0.866	-.011074	.0093139

msa2	-.0055184	.0047189	-1.17	0.242	-.0147685	.0037317
q1	.0145704	.0051986	2.80	0.005	.00438	.0247607
q2	.0022831	.0047579	0.48	0.631	-.0070434	.0116096
q3	.000043	.0047504	0.01	0.993	-.0092687	.0093547
_cons	.0927746	.0364578	2.54	0.011	.0213098	.1642394

The following Stata output replicates the OLS results given in Table 6 of Mullahy and Sindelar (1996, p. 429) for females. The first regression is for employment, given in column 1 of Table 6 of the paper, and the second regression is for unemployment, given in column 3 of Table 6 of the paper. Robust standard errors are reported.

```
. reg emp hvdrnk90 ue88 age agesq educ married famsize white hlstat1 hlstat2
hlstat3 hlstat4 region1 region2 region3 msa1 msa2 q1 q2 q3, robust
```

```
Regression with robust standard errors          Number of obs  =  12534
                                                F(20, 12513)   = 117.99
                                                Prob > F       = 0.0000
                                                R-squared      = 0.1358
                                                Root MSE       = .42932
```

emp	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
hvdrnk90	.0059878	.0120102	0.50	0.618	-.017554	.0295296
ue88	-.0168969	.002911	-5.80	0.000	-.0226028	-.011191
age	.04635	.0036794	12.60	0.000	.0391378	.0535622
agesq	-.0005898	.0000449	-13.13	0.000	-.0006778	-.0005018
educ	.0227162	.0015509	14.65	0.000	.0196762	.0257563
married	.0105416	.0111463	0.95	0.344	-.0113068	.0323901
famsize	-.0662794	.0030445	-21.77	0.000	-.072247	-.0603118
white	-.0077594	.0104111	-0.75	0.456	-.0281668	.012648
hvdrnk90	.0059878	.0120102	0.50	0.618	-.017554	.0295296
ue88	-.0168969	.002911	-5.80	0.000	-.0226028	-.011191
age	.04635	.0036794	12.60	0.000	.0391378	.0535622
agesq	-.0005898	.0000449	-13.13	0.000	-.0006778	-.0005018
educ	.0227162	.0015509	14.65	0.000	.0196762	.0257563
married	.0105416	.0111463	0.95	0.344	-.0113068	.0323901
famsize	-.0662794	.0030445	-21.77	0.000	-.072247	-.0603118
white	-.0077594	.0104111	-0.75	0.456	-.0281668	.012648
hlstat1	.4601695	.0253797	18.13	0.000	.4104214	.5099177
hlstat2	.4583823	.0252973	18.12	0.000	.4087957	.5079689

```
 hlstat3 |   .4096243   .0251983    16.26   0.000    .3602317    .4590169
 hlstat4 |   .2494427    .027846     8.96   0.000    .1948602    .3040251
 region1 |  -.0180596   .0129489    -1.39   0.163   -.0434415    .0073223
 region2 |   .0095951   .0114397     0.84   0.402   -.0128285    .0320186
 region3 |   .0465464   .0108841     4.28   0.000    .0252119     .067881
    msa1 |  -.0256183   .0109856    -2.33   0.020   -.0471518   -.0040848
    msa2 |   .0051885   .0103385     0.50   0.616   -.0150765    .0254534
      q1 |  -.0058134   .0107234    -0.54   0.588   -.0268329    .0152061
      q2 |  -.0061301   .0109033    -0.56   0.574   -.0275022    .0152421
      q3 |  -.0168673   .0109023    -1.55   0.122   -.0382376    .0045029
   _cons |  -.5882924   .0782545    -7.52   0.000   -.7416831   -.4349017
-------------------------------------------------------------------------------
```

```
. reg unemp hvdrnk90 ue88 age agesq educ married famsize white hlstat1 hlstat2
hlstat3 hlstat4 region1 region2 region3 msa1 msa2 q1 q2 q3, robust

  Regression with robust standard errors               Number of obs  =   12534
                                                       F(20, 12513)   =    5.99
                                                       Prob > F       =  0.0000
                                                       R-squared      =  0.0141
                                                       Root MSE       =  .18409

-------------------------------------------------------------------------------
          |                Robust
      emp |      Coef.    Std. Err.       t     P>|t|     [95% Conf. Interval]
-------------------------------------------------------------------------------
 hvdrnk90 |    .0149286   .0059782     2.50   0.013    .0032104    .0266468
     ue88 |    .0038119   .0013782     2.77   0.006    .0011105    .0065133
      age |   −.0013974   .0015439    -0.91   0.365   -.0044237    .0016289
    agesq |    4.43e-06   .0000181     0.24   0.807   -.0000311      .00004
     educ |   −.0011631   .0006751    -1.72   0.085   -.0024865    .0001602
  married |   −.0066296   .0058847    -1.13   0.260   -.0181645    .0049053
  famsize |    .0013304   .0013075     1.02   0.309   -.0012325    .0038933
    white |   −.0308826   .0051866    -5.95   0.000   -.0410493    -.020716
  hlstat1 |     .008861   .0092209     0.96   0.337   -.0092135    .0269354
  hlstat2 |    .0079536   .0091305     0.87   0.384   -.0099435    .0258507
  hlstat3 |    .0224927   .0093356     2.41   0.016    .0041934    .0407919
  hlstat4 |    .0193116   .0106953     1.81   0.071   -.0016528     .040276
  region1 |    .0020325   .0055618     0.37   0.715   -.0088694    .0129344
  region2 |   −.0005405   .0049211    -0.11   0.913   -.0101866    .0091057
  region3 |   −.0079708   .0046818    -1.70   0.089   -.0171479    .0012063
     msa1 |    −.002055   .0049721    -0.41   0.679   -.0118011     .007691
     msa2 |   −.0130041   .0041938    -3.10   0.002   -.0212246   -.0047835
       q1 |    .0025441   .0043698     0.58   0.560   -.0060214    .0111095
       q2 |    .0080984   .0046198     1.75   0.080   -.0009571    .0171539
       q3 |    .0102601   .0046839     2.19   0.029     .001079    .0194413
    _cons |    .0922081   .0350856     2.63   0.009     .023435    .1609813
-------------------------------------------------------------------------------
```

The corresponding probit equation for employment for males is given by the following stata output (this replicates Table 13.6 in the text):

```
. probit emp hvdrnk90 ue88 age agesq educ married famsize white hlstat1
hlstat2 hlstat3 hlstat4 region1 region2 region3 msa1 msa2 q1 q2 q3, robust

    Probit regression                                  Number of obs   =    9822
                                                       Wald chi2(20)   =  928.34
                                                       Prob > chi2     =  0.0000
    Log pseudolikelihood = -2698.1797                  Pseudo R2       =  0.1651
```

emp	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
hvdrnk90	-.1049465	.0589878	-1.78	0.075	-.2205606	.0106675
ue88	-.0532774	.0142024	-3.75	0.000	-.0811135	-.0254413
age	.0996338	.0171184	5.82	0.000	.0660824	.1331853
agesq	-.0013043	.0002051	-6.36	0.000	-.0017062	-.0009023
educ	.0471834	.0066738	7.07	0.000	.034103	.0602638
married	.2952921	.0540855	5.46	0.000	.1892866	.4012976
famsize	.0188906	.0140462	1.34	0.179	-.0086395	.0464206
white	.3945226	.0483378	8.16	0.000	.2997822	.489263
hlstat1	1.816306	.0983443	18.47	0.000	1.623554	2.009057
hlstat2	1.778434	.0991528	17.94	0.000	1.584098	1.97277
hlstat3	1.547836	.0982635	15.75	0.000	1.355244	1.740429
hlstat4	1.043363	.1077276	9.69	0.000	.8322209	1.254505
region1	.0343123	.0620016	0.55	0.580	-.0872085	.1558331
region2	.0604907	.0537881	1.12	0.261	-.044932	.1659135
region3	.1821206	.0542342	3.36	0.001	.0758236	.2884176
msa1	-.0730529	.0518715	-1.41	0.159	-.1747192	.0286134
msa2	.0759533	.0513087	1.48	0.139	-.02461	.1765166
q1	-.1054844	.0527723	-2.00	0.046	-.2089162	-.0020525
q2	-.0513229	.052818	-0.97	0.331	-.1548444	.0521985
q3	-.0293419	.0543746	-0.54	0.589	-.1359142	.0772303
_cons	-3.017454	.3592294	-8.40	0.000	-3.72153	-2.313377

We can see how the probit model fits by looking at its predictions.

```
. estat classification
Probit model for emp
```

	True		
Classified	D	∼D	Total
+	8743	826	9569
−	79	174	253
Total	8822	1000	9822

Classified + if predicted Pr(D) >= .5

True D defined as emp != 0

Sensitivity	Pr (+| D)	99.10%
Specificity	Pr (−| ∼D)	17.40%
Positive predictive value	Pr (D| +)	91.37%
Negative predictive value	Pr (∼ D| −)	68.77%
False + rate for true ∼D	Pr (+| ∼ D)	82.60%
False − rate for true D	Pr(−| D)	0.90%
False + rate for classified +	Pr(∼D| +)	8.63%
False − rate for classified −	Pr (D| −)	31.23%
Correctly classified		90.79%

We could have alternatively run a logit regression on employment for males

.logit emp hvdrnk90 ue88 age agesq educ married famsize white hlstat1 hlstat2 hlstat3 hlstat4 region1 region2 region3 msa1 msa2 q1 q2 q3, robust

Logistic regression	Number of obs	=	9822
	Wald chi2(20)	=	900.15
	Prob > chi2	=	0.0000
Log pseudolikelihood = -2700.0567	Pseudo R2	=	0.1646

emp	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
hvdrnk90	-.1960754	.1114946	-1.76	0.079	-.4146008	.02245
ue88	-.1131074	.0273316	-4.14	0.000	-.1666764	-.0595384
age	.1884486	.0332284	5.67	0.000	.123322	.2535751
agesq	-.0024584	.0003965	-6.20	0.000	-.0032356	-.0016813
educ	.0913569	.0127978	7.14	0.000	.0662738	.1164401
married	.5534291	.1057963	5.23	0.000	.3460721	.760786
famsize	.0365059	.0276468	1.32	0.187	-.0176808	.0906927

white	.7224036	.0912559	7.92	0.000	.5435454	.9012619
hlstat1	3.145481	.1721925	18.27	0.000	2.80799	3.482972
hlstat2	3.067279	.1741295	17.61	0.000	2.725992	3.408567
hlstat3	2.613691	.1707421	15.31	0.000	2.279042	2.948339
hlstat4	1.725571	.1844904	9.35	0.000	1.363976	2.087166
region1	.0493715	.1220065	0.40	0.686	-.1897568	.2884999
region2	.1146108	.105504	1.09	0.277	-.0921733	.3213948
region3	.3738274	.1066491	3.51	0.000	.1647991	.5828558
msa1	-.1690904	.1016459	-1.66	0.096	-.3683127	.0301319
msa2	.1345974	.1021625	1.32	0.188	-.0656374	.3348323
q1	-.1954528	.1034703	-1.89	0.059	-.3982508	.0073453
q2	-.1052494	.1033014	-1.02	0.308	-.3077163	.0972176
q3	-.0418287	.1074896	-0.39	0.697	-.2525045	.168847
_cons	-5.538271	.6935383	-7.99	0.000	-6.897581	-4.178961

And the corresponding predictions for the logit model are given by

```
. estat classification
Logistic model for emp
```

Classified	True D	True ~D	Total
+	8740	822	9562
−	82	178	260
Total	8822	1000	9822

Classified + if predicted Pr(D) >= .5

True D defined as emp !=0

Sensitivity	Pr(+— D)	99.07%
Specificity	Pr(−—~D)	17.80%
Positive predictive value	Pr(D— +)	91.40%
Negative predictive value	Pr(~D— −)	68.46%
False + rate for true ~D	Pr(+| ~D)	82.20%
False − rate for true D	Pr(−— D)	0.93%
False + rate for classified +	Pr(~D— +)	8.60%
False − rate for classified −	Pr(D— −)	31.54%
Correctly classified		90.80%

The marginal effects for the probit model can be obtained as follows:

```
.dprobit emp hvdrnk90 ue88 age agesq educ married famsize white hlstat1 hlstat2 hlstat3
hlstat4 region1 region2 region3 msa1 msa2 q1 q2 q3, robust

Iteration 0:   log pseudolikelihood =-3231.8973
Iteration 1:   log pseudolikelihood =-2707.0435
Iteration 2:   log pseudolikelihood =-2698.2015
Iteration 3:   log pseudolikelihood =-2698.1797

Probit regression, reporting marginal effects        Number of obs  =    9822
                                                     Wald chi2(20)  =  928.34
                                                     Prob > chi2    =  0.0000
Log pseudolikelihood = -2698.1797                    Pseudo R2      =  0.1651
```

emp	dF/dx	Robust Std. Err.	z	P>|z|	x-bar	[95% Conf.	Interval]
hvdrnk90*	-.0161704	.0096242	-1.78	0.075	.099165	-.035034	.002693
ue88	-.0077362	.0020463	-3.75	0.000	5.56921	-.011747	-.003725
age	.0144674	.0024796	5.82	0.000	39.1757	.009607	.019327
agesq	-.0001894	.0000297	-6.36	0.000	1627.61	-.000248	-.000131
educ	.0068513	.0009621	7.07	0.000	13.3096	.004966	.008737
married*	.0488911	.010088	5.46	0.000	.816432	.029119	.068663
famsize	.002743	.002039	1.34	0.179	2.7415	-.001253	.006739
white*	.069445	.0100697	8.16	0.000	.853085	.049709	.089181
hlstat1*	.2460794	.0148411	18.47	0.000	.415903	.216991	.275167
hlstat2*	.1842432	.0099207	17.94	0.000	.301873	.164799	.203687
hlstat3*	.130786	.0066051	15.75	0.000	.205254	.11784	.143732
hlstat4*	.0779836	.0041542	9.69	0.000	.053451	.069841	.086126
region1*	.0049107	.0087468	0.55	0.580	.203014	-.012233	.022054
region2*	.0086088	.0075003	1.12	0.261	.265628	-.006092	.023309
region3*	.0252543	.0071469	3.36	0.001	.318265	.011247	.039262
msa1*	-.0107946	.0077889	-1.41	0.159	.333232	-.026061	.004471
msa2*	.0109542	.0073524	1.48	0.139	.434942	-.003456	.025365
q1*	-.0158927	.0082451	-2.00	0.046	.254632	-.032053	.000267
q2*	-.0075883	.0079484	-0.97	0.331	.252698	-.023167	.00799
q3*	-.0043066	.0080689	-0.54	0.589	.242822	-.020121	.011508
obs. P	.8981877						
pred. P	.9224487	(at x-bar)					

(*) dF/dx is for discrete change of dummy variable from 0 to 1
z and P>|z| correspond to the test of the underlying coefficient being 0

One can also run logit and probit for the unemployment variable and repeat this for females. This is not done here to save space.

13.15 *Fertility and Female Labor Supply*

a. Carrasco (2001, p. 391) Table 4, column 1, ran a fertility probit equation, which we replicate below using Stata:

```
probit f dsex ags26l educ_2 educ_3 age drace inc
```

Probit regression	Number of obs	=	5768
	LR chi2 (7)	=	964.31
	Prob > chi2	=	0.0000
Log likelihood = −1561.1312	Pseudo R2	=	0.2360

f	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
dsex	.3250503	.0602214	5.40	0.000	.2070184	.4430822
ags26l	-2.135365	.1614783	-13.22	0.000	-2.451857	-1.818873
educ_2	.0278467	.1145118	0.24	0.808	-.1965922	.2522856
educ_3	.3071582	.1255317	2.45	0.014	.0611207	.5531958
age	-.0808522	.0048563	-16.65	0.000	-.0903703	-.071334
drace	-.0916409	.0629859	-1.45	0.146	-.215091	.0318093
inc	.003161	.0029803	1.06	0.289	-.0026803	.0090022
_cons	1.526893	.1856654	8.22	0.000	1.162996	1.890791

For part (b) the predicted probabilities are obtained as follows:

. lstat

Probit model for f

Classified	True D	True ~D	Total
+	2	3	5
-	654	5109	5763
Total	656	5112	5768

Classified + if predicted Pr(D) >= .5
True D defined as f != 0

Sensitivity	Pr(+| D)	0.30%
Specificity	Pr(-| ~D)	99.94%
Positive predictive value	Pr(D| +)	40.00%
Negative predictive value	Pr(~D| -)	88.65%
False + rate for true ~D	Pr(+| ~D)	0.06%
False - rate for true D	Pr(-| D)	99.70%
False + rate for classified +	Pr(~D| +)	60.00%
False - rate for classified -	Pr(D| -)	11.35%
Correctly classified		88.61%

The estimates reveal that having children of the same sex has a significant and positive effect on the probability of having an additional child. The marginal effects are given by dprobit in Stata

. dprobit f dsex ags26l educ_2 educ_3 age drace inc

Probit regression, reporting marginal effects

Number of obs = 5768
LR chi2 (7) = 964.31
Prob > chi2 = 0.0000
Pseudo R2 = 0.2360

Log likelihood = −1561.1312

f	dF/dx	Std. Err.	z	P>|z|	x-bar	[95% C.I.]	
dsex*	.0302835	.0069532	5.40	0.000	.256415	.016655	.043912
ags26l*	-.1618148	.0066629	-13.22	0.000	.377601	-.174874	-.148756
educ_2*	.0022157	.0090239	0.24	0.808	.717753	-.015471	.019902
educ_3*	.0288636	.0140083	2.45	0.014	.223994	.001408	.056319
age	-.0065031	.0007644	-16.65	0.000	32.8024	-.008001	-.005005
drace*	-.0077119	.0055649	-1.45	0.146	.773232	-.018619	.003195
inc	.0002542	.000241	1.06	0.289	12.8582	-.000218	.000727
obs. P	.1137309						
pred. P	.0367557	(at x-bar)					

(*) dF/dx is for discrete change of dummy variable from 0 to 1 z and P> |z| correspond to the test of the underlying coefficient being 0

If we replace same sex by its components: same sex female and same sex male variables, the results do not change indicating that having both boys or girls does not matter, see Carrasco (2001,p.391) Table 4, column 2.

. probit f dsexm dsexf ags26l educ_2 educ_3 age drace inc

Probit regression

Number of obs = 5768
LR chi2 (8) = 964.32
Prob > chi2 = 0.0000
Pseudo R2 = 0.2360

Log likelihood = −1561.1284

f	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
dsexm	.328542	.0764336	4.30	0.000	.1787349	.4783491
dsexf	.3209239	.0820417	3.91	0.000	.1601252	.4817226
ags26l	-2.135421	.1614518	-13.23	0.000	-2.451861	-1.818981
educ_2	.027657	.1145384	0.24	0.809	-.1968342	.2521482
educ_3	.3068706	.1255904	2.44	0.015	.0607179	.5530233
age	-.0808669	.0048605	-16.64	0.000	-.0903934	-.0713404
drace	-.0918074	.0630233	-1.46	0.145	-.2153308	.031716
inc	.0031709	.0029829	1.06	0.288	-.0026754	.0090173
_cons	1.527551	.1858818	8.22	0.000	1.163229	1.891872

. lstat

Probit model for f

Classified	True D	True ~D	Total
+	2	3	5
-	654	5109	5763
Total	656	5112	5768

Classified + if predicted Pr(D) >= .5
True D defined as f != 0

Sensitivity	Pr(+| D)	0.30%
Specificity	Pr(-| ~D)	99.94%
Positive predictive value	Pr(D| +)	40.00%
Negative predictive value	Pr(~D| -)	88.65%
False + rate for true ~D	Pr(+| ~D)	0.06%
False - rate for true D	Pr(-| D)	99.70%
False + rate for classified +	Pr(~D| +)	60.00%
False - rate for classified -	Pr(D| -)	11.35%
Correctly classified		88.61%

. dprobit f dsexm dsexf ags26l educ_2 educ_3 age drace inc

Probit regression, reporting marginal effects

Number of obs = 5768
LR chi2 (7) = 964.32
Prob > chi2 = 0.0000
Log likelihood = −1561.1284
Pseudo R2 = 0.2360

f	dF/dx	Std. Err.	z	P>|z|	x-bar	[95% C.I.]	
dsexm*	.0325965	.0095475	4.30	0.000	.145111	.013884	.051309
dsexf*	.032261	.0103983	3.91	0.000	.111304	.011881	.052641
ags26l*	-.16182	.0066634	-13.23	0.000	.377601	-.17488	-.14876
educ_2*	.0022008	.0090273	0.24	0.809	.717753	-.015492	.019894
educ_3*	.0288323	.01401	2.44	0.015	.223994	.001373	.056291
age	-.0065042	.0007645	-16.64	0.000	32.8024	-.008003	-.005006
drace*	-.0077266	.0055692	-1.46	0.145	.773232	-.018642	.003189
inc	.000255	.0002412	1.06	0.288	12.8582	-.000218	.000728
obs. P	.1137309						
pred. P	.0367556	(at x-bar)					

(*) dF/dx is for discrete change of dummy variable from 0 to 1
z and P > |z| correspond to the test of the underlying coefficient being 0

c. Carrasco (2001, p. 392) Table 5, column 4, ran a female labor force participation OLS equation, which we replicate below using Stata 10:

```
. reg dhw f ags26l fxag26l educ_2 educ_3 age drace inc dhwl
```

Source	SS	df	MS			
				Number of obs	=	5768
				F(9, 5758)	=	445.42
Model	419.800476	9	46.6444973	Prob > F	=	0.0000
Residual	602.985758	5758	.104721389	R-squared	=	0.4104
				Adj R-squared	=	0.4095
Total	1022.78623	5767	.177351523	Root MSE	=	.32361

dhw	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
f	-.0888995	.0144912	-6.13	0.000	-.1173077	-.0604912
ags26l	-.0194454	.0093334	-2.08	0.037	-.0377424	-.0011484
fxag26l	-.0581458	.1629414	-0.36	0.721	-.3775723	.2612806
educ_2	.0491989	.0186018	2.64	0.008	.0127323	.0856655
educ_3	.0725501	.0207404	3.50	0.000	.0318912	.1132091
age	.0014193	.0007854	1.81	0.071	-.0001203	.002959
drace	-.0098333	.010379	-0.95	0.343	-.03018	.0105134
inc	-.0018149	.0004887	-3.71	0.000	-.002773	-.0008568
dhwl	.6253973	.0103188	60.61	0.000	.6051686	.645626
_cons	.2373022	.032744	7.25	0.000	.1731117	.3014927

Carrasco (2001, p. 392) Table 5, column 1, ran a female labor force participation probit equation, which we replicate below using Stata:

```
. probit dhw f ags26l fxag26l educ_2 educ_3 age drace inc dhwl
```

Probit regression

	Number of obs	=	5768
	LR chi2 (7)	=	2153.17
	Prob > chi2	=	0.0000
Log likelihood = −2036.8086	Pseudo R2	=	0.3458

dhw	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
f	-.4103849	.0690538	-5.94	0.000	-.5457279	-.2750419
ags26l	-.1064159	.0480907	-2.21	0.027	-.200672	-.0121598
fxag26l	-.1886427	.7087803	-0.27	0.790	-1.577827	1.200541
educ_2	.2338264	.0858408	2.72	0.006	.0655816	.4020713
educ_3	.3773278	.1001949	3.77	0.000	.1809494	.5737062
age	.0091203	.0041132	2.22	0.027	.0010586	.017182
drace	-.0577508	.0542972	-1.06	0.288	-.1641714	.0486699
inc	-.0088483	.0024217	-3.65	0.000	-.0135948	-.0041019
dhwl	1.932025	.0462191	41.80	0.000	1.841438	2.022613
_cons	-.8540838	.1638299	-5.21	0.000	-1.175184	-.5329831

```
. lstat
```

Probit model for dhw

Classified	------ True ------ D	~D	Total
+	4073	378	4451
-	366	951	1317
Total	4439	1329	5768

Classified + if predicted Pr(D) >= .5
True D defined as dhw != 0

Sensitivity	Pr(+| D)	91.75%
Specificity	Pr(-| ~D)	71.56%
Positive predictive value	Pr(D| +)	91.51%
Negative predictive value	Pr(~D| -)	72.21%
False + rate for true ~D	Pr(+| ~D)	28.44%
False - rate for true D	Pr(-| D)	8.25%
False + rate for classified +	Pr(~D| +)	8.49%
False - rate for classified -	Pr(D| -)	27.79%
Correctly classified		87.10%

The marginal effects are given by dprobit in Stata:

. dprobit dhw f ags26l fxag26l educ_2 educ_3 age drace inc dhwl

Probit regression, reporting marginal effects

Number of obs = 5768
LR chi2 (9) = 2153.17
Prob > chi2 = 0.0000
Log likelihood = −2036.8086
Pseudo R2 = 0.3458

dhw	dF/dx	Std. Err.	z	P>|z|	x-bar	[95% C.I.]	
f*	-.1200392	.0224936	-5.94	0.000	.113731	-.164126	-.075953
ags26l*	-.0275503	.0125892	-2.21	0.027	.377601	-.052225	-.002876
fxag26l*	-.0524753	.2127127	-0.27	0.790	.000693	-.469385	.364434
educ_2*	.0626367	.0239923	2.72	0.006	.717753	.015613	.109661
educ_3*	.0870573	.0206089	3.77	0.000	.223994	.046665	.12745
age	.0023327	.0010504	2.22	0.027	32.8024	.000274	.004391
drace*	-.0145508	.0134701	-1.06	0.288	.773232	-.040952	.01185
inc	-.0022631	.0006189	-3.65	0.000	12.8582	-.003476	-.00105
dhwl*	.6249756	.0134883	41.80	0.000	.771671	.598539	.651412
obs. P	.7695908						
pred. P	.8271351	(at x-bar)					

(*) dF/dx is for discrete change of dummy variable from 0 to 1 z and P> |z| correspond to the test of the underlying coefficient being 0

d. The 2sls estimates in Table 5, column 5, of Carrasco (2001, p. 392) using as instruments the same sex variables and their interactions with ags26l is given below, along with the over-identification test and the first stage diagnostics:

```
. ivregress 2sls dhw (f fxag26l =dsexm dsexf sexm_26l sexf_26l) ags26l educ_2 e
> duc_3 age drace inc dhwl

Instrumental variables (2SLS) regression          Number of obs   =      5768
                                                  Wald chi2(9)    =   3645.96
                                                  Prob > chi2     =    0.0000
                                                  R-squared       =    0.3565
                                                  Root MSE        =     .3378
```

dhw	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
f	-.2164685	.2246665	-0.96	0.335	-.6568067	.2238697
fxag26l	-3.366305	3.512783	-0.96	0.338	-10.25123	3.518623
ags26l	-.0385731	.0467522	-0.83	0.409	-.1302058	.0530596
educ_2	.0331807	.0288653	1.15	0.250	-.0233943	.0897557
educ_3	.064607	.0348694	1.85	0.064	-.0037357	.1329497
age	-.0001934	.0030344	-0.06	0.949	-.0061407	.0057539
drace	-.0163251	.012366	-1.32	0.187	-.0405621	.0079118
inc	-.0017194	.0005162	-3.33	0.001	-.0027312	-.0007076
dhwl	.6230639	.017256	36.11	0.000	.5892427	.6568851
_cons	.3330965	.141537	2.35	0.019	.0556891	.610504

```
Instrumented: f fxag26l
Instruments:  ags26l educ_2 educ_3 age drace inc dhwl dsexm dsexf sexm_26l
              sexf_26l

. estat overid

Tests of overidentifying restrictions:

Sargan (score) chi2(2)  = .332468 (p = 0.8468)
Basmann chi2(2)         = .331796 (p = 0.8471)

. estat firststage

Shea's partial R-squared
```

Variable	Shea's Partial R-sq.	Shea's Adj. Partial R-sq.
f	0.0045	0.0028
fxag26l	0.0023	0.0006

Minimum eigenvalue statistic = 3.36217

Critical Values Ho: Instruments are weak	# of endogenous regressors: 2 # of excluded instruments: 4			
	5%	10%	20%	30%
2SLS relative bias	11.04	7.56	5.57	4.73
	10%	15%	20%	25%
2SLS Size of nominal 5% Wald test	16.87	9.93	7.54	6.28
LIML Size of nominal 5% Wald test	4.72	3.39	2.99	2.79

e. So far, heterogeneity across the individuals is not taken into account. Carrasco (2001, p. 393) Table 7, column 4, ran a female labor force participation fixed effects equation with robust standard errors, which we replicate below using Stata:

```
. xtreg dhw f ags26l fxag26l dhwl, fe r
```

Fixed-effects (within) regression		Number of obs =	5768
Group variable: ident		Number of groups =	1442
R-sq: within = 0.0059		Obs per group: min =	4
between = 0.6185		avg =	4.0
overall = 0.2046		max =	4
		F(4,4322) =	4.64
corr(u_i, Xb) = 0.4991		Prob > F =	0.0010

(Std. Err. adjusted for clustering on ident)

dhw	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
f	-.0547777	.0155326	-3.53	0.000	-.0852296	-.0243257
ags26l	.0012836	.0126213	0.10	0.919	-.0234607	.0260279
fxag26l	-.2204885	.2013721	-1.09	0.274	-.615281	.1743041
dhwl	.0356233	.0236582	1.51	0.132	-.0107588	.0820055
_cons	.7479995	.0193259	38.70	0.000	.7101108	.7858881
sigma_u	.33260036					
sigma_e	.27830212					
rho	.58818535	(fraction of variance due to u_i)				

Note that only fertility is significant in this equation.

Fixed effects 2sls using as instruments the same sex variables and their interactions with ags26l is given below:

```
. xtivreg dhw (f fxag26l =dsexm dsexf sexm_26l sexf_26l)ags26l age inc dhwl, fe

Fixed-effects (within) IV regression          Number of obs        =      5768
Group variable: ident                         Number of groups     =      1442

R-sq: within = .                              Obs per group: min   =         4
between = 0.1125                              avg                  =       4.0
overall = 0.0332                              max                  =         4

                                              Wald chi2(6)         =  39710.29
corr(u_i, Xb) = 0.0882                        Prob > chi2          =    0.0000

------------------------------------------------------------------------------
         dhw |     Coef.    Std. Err.      z     P>|z|     [95% Conf. Interval]
------------------------------------------------------------------------------
           f | -.2970225    .156909    -1.89    0.058    -.6045584    .0105134
     fxag26l |   -2.1887   2.433852    -0.90    0.369    -6.958963    2.581562
      ags26l | -.0584866   .0467667    -1.25    0.211    -.1501476    .0331744
         age |   .000651   .0043265     0.15    0.880    -.0078287    .0091307
         inc | -.0011213   .0010482    -1.07    0.285    -.0031758    .0009331
        dhwl |  .0362943   .0160524     2.26    0.024     .0048322    .0677565
       _cons |  .7920305   .1623108     4.88    0.000     .4739071    1.110154
------------------------------------------------------------------------------
     sigma_u |  .3293446
     sigma_e | .29336161
         rho | .55759255          (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:          F(1441,4320) = 2.16          Prob > F = 0.0000
------------------------------------------------------------------------------
Instrumented:      f fxag26l
Instruments:     ags26l age inc dhwl dsexm dsexf sexm_26l sexf_26l
------------------------------------------------------------------------------
```

13.16 *multinomial logit model*

a. Table II of Terza (2002, p. 399) columns 3,4, 9 and 10 are replicated below for the male data using Stata:

. mlogit y alc90th ue88 age agesq schooling married famsize white excellent verygood good fair northeast midwest south centercity othermsa q1 q2 q3, baseoutcome(1)

Multinomial logistic regression	Number of obs	=	9822
	LR chi2 (40)	=	1276.47
	Prob > chi2	=	0.0000
Log likelihood = −3217.481	Pseudo R2	=	0.1655

y	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
2						
alc90th	.1270931	.21395	0.59	0.552	-.2922412	.5464274
ue88	.0458099	.051355	0.89	0.372	-.0548441	.1464639
age	.1617634	.0663205	2.44	0.015	.0317776	.2917492
agesq	-.0024377	.0007991	-3.05	0.002	-.004004	-.0008714
schooling	-.0092135	.0245172	-0.38	0.707	-.0572664	.0388393
married	.4004928	.1927458	2.08	0.038	.022718	.7782677
famsize	.0622453	.0503686	1.24	0.217	-.0364753	.1609659
white	.0391309	.1705625	0.23	0.819	-.2951653	.3734272
excellent	2.91833	.4486757	6.50	0.000	2.038942	3.797719
verygood	2.978336	.4505932	6.61	0.000	2.09519	3.861483
good	2.493939	.4446815	5.61	0.000	1.622379	3.365499
fair	1.460263	.4817231	3.03	0.002	.5161027	2.404422
northeast	.0849125	.2374365	0.36	0.721	-.3804545	.5502796
midwest	.0158816	.2037486	0.08	0.938	-.3834583	.4152215
south	.1750244	.2027444	0.86	0.388	-.2223474	.5723962
centercity	-.2717445	.1911074	-1.42	0.155	-.6463081	.1028192
othermsa	-.0921566	.1929076	-0.48	0.633	-.4702486	.2859354
q1	.422405	.1978767	2.13	0.033	.0345738	.8102362
q2	-.0219499	.2056751	-0.11	0.915	-.4250657	.3811659
q3	-.0365295	.2109049	-0.17	0.862	-.4498954	.3768364
_cons	-6.113244	1.427325	-4.28	0.000	-8.910749	-3.315739
3						
alc90th	-.1534987	.1395003	-1.10	0.271	-.4269144	.1199169
ue88	-.0954848	.033631	-2.84	0.005	-.1614004	-.0295693
age	.227164	.0409884	5.54	0.000	.1468282	.3074999
agesq	-.0030796	.0004813	-6.40	0.000	-.0040228	-.0021363
schooling	.0890537	.0152314	5.85	0.000	.0592008	.1189067
married	.7085708	.1219565	5.81	0.000	.4695405	.9476012
famsize	.0622447	.0332365	1.87	0.061	-.0028975	.127387
white	.7380044	.1083131	6.81	0.000	.5257147	.9502941
excellent	3.702792	.1852415	19.99	0.000	3.339725	4.065858
verygood	3.653313	.1894137	19.29	0.000	3.282069	4.024557
good	2.99946	.1786747	16.79	0.000	2.649264	3.349656
fair	1.876172	.1885159	9.95	0.000	1.506688	2.245657
northeast	.088966	.1491191	0.60	0.551	-.203302	.3812341
midwest	.1230169	.1294376	0.95	0.342	-.130676	.3767099
south	.4393047	.1298054	3.38	0.001	.1848908	.6937185
centercity	-.2689532	.1231083	-2.18	0.029	-.510241	-.0276654
othermsa	.0978701	.1257623	0.78	0.436	-.1486195	.3443598
q1	-.0274086	.1286695	-0.21	0.831	-.2795961	.224779
q2	-.110751	.126176	-0.88	0.380	-.3580514	.1365494
q3	-.0530835	.1296053	-0.41	0.682	-.3071052	.2009382
_cons	-6.237275	.8886698	-7.02	0.000	-7.979036	-4.495515

(y==1 is the base outcome)

```
**using bootstrap for the var-cov matrix

. mlogit y alc90th ue88 age agesq schooling married famsize white excellent ver >
ygood good fair northeast midwest south centercity othermsa q1 q2 q3, baseout
> come(1) vce(bootstrap)

(running mlogit on estimation sample)

  Bootstrap replications (50)

----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
.................................................. 50

Multinomial logistic regression                  Number of obs   =     9822
                                                 Replications    =       50
                                                 Wald chi2 (40)  =  7442.69
                                                 Prob > chi2     =   0.0000
Log likelihood = -3217.481                       Pseudo R2       =   0.1655
```

y	Observed Coef.	Bookstrap Std. Err.	z	P>|z|	Normal-based [95% Conf. Interval]	
2						
alc90th	.1270931	.1933016	0.66	0.511	-.251771	.5059573
ue88	.0458099	.0566344	0.81	0.419	-.0651914	.1568112
age	.1617634	.0615543	2.63	0.009	.0411192	.2824076
agesq	-.0024377	.0007266	-3.35	0.001	-.0038619	-.0010135
schooling	-.0092135	.0249799	-0.37	0.712	-.0581732	.0397462
married	.4004928	.2069088	1.94	0.053	-.0050409	.8060266
famsize	.0622453	.0534164	1.17	0.244	-.0424489	.1669395
white	.0391309	.1817052	0.22	0.829	-.3170046	.3952665
excellent	2.91833	.5134264	5.68	0.000	1.912033	3.924628
verygood	2.978336	.5473854	5.44	0.000	1.905481	4.051192
good	2.493939	.4904972	5.08	0.000	1.532582	3.455296
fair	1.460263	.5156181	2.83	0.005	.4496697	2.470855
northeast	.0849125	.2058457	0.41	0.680	-.3185377	.4883627
midwest	.0158816	.197601	0.08	0.936	-.3714093	.4031725
south	.1750244	.2211406	0.79	0.429	-.2584032	.608452
centercity	-.2717445	.1708023	-1.59	0.112	-.6065108	.0630218
othermsa	-.0921566	.191577	-0.48	0.630	-.4676406	.2833275
q1	.422405	.2392306	1.77	0.077	-.0464783	.8912883
q2	-.0219499	.2404712	-0.09	0.927	-.4932649	.4493651
q3	-.0365295	.2500046	-0.15	0.884	-.5265295	.4534704
_cons	-6.113244	1.259449	-4.85	0.000	-8.581719	-3.644769

3						
alc90th	-.1534987	.1129983	-1.36	0.174	-.3749714	.0679739
ue88	-.0954848	.0349536	-2.73	0.006	-.1639927	-.026977
age	.227164	.0431	5.27	0.000	.1426896	.3116385
agesq	-.0030796	.0005224	-5.90	0.000	-.0041034	-.0020558
schooling	.0890537	.0173814	5.12	0.000	.0549868	.1231207
married	.7085708	.1286085	5.51	0.000	.4565028	.9606389
famsize	.0622447	.0361903	1.72	0.085	-.0086869	.1331764
white	.7380044	.1320206	5.59	0.000	.4792488	.9967599
excellent	3.702792	.2019607	18.33	0.000	3.306956	4.098627
verygood	3.653313	.2090086	17.48	0.000	3.243663	4.062962
good	2.99946	.2053791	14.60	0.000	2.596925	3.401996
fair	1.876172	.2063004	9.09	0.000	1.471831	2.280514
northeast	.088966	.1624429	0.55	0.584	-.2294162	.4073482
midwest	.1230169	.1410455	0.87	0.383	-.1534272	.399461
south	.4393047	.1340076	3.28	0.001	.1766547	.7019547
centercity	-.2689532	.098325	-2.74	0.006	-.4616666	-.0762398
othermsa	.0978701	.1067784	0.92	0.359	-.1114117	.307152
q1	-.0274086	.1206965	-0.23	0.820	-.2639694	.2091523
q2	-.110751	.1303469	-0.85	0.396	-.3662263	.1447243
q3	-.0530835	.1329726	-0.40	0.690	-.313705	.2075381
_cons	-6.237275	.8224026	-7.58	0.000	-7.849155	-4.625396

(y==1 is the base outcome)

```
**using robust for the var-cov matrix

. mlogit y alc90th ue88 age agesq schooling married famsize white excellent ver >
ygood good fair northeast midwest south centercity othermsa q1 q2 q3, baseout
> come(1) vce(robust)

Iteration 0: log pseudolikelihood = -3855.7148
Iteration 1: log pseudolikelihood = -3692.5753
Iteration 2: log pseudolikelihood = -3526.5092
Iteration 3: log pseudolikelihood = -3236.3918
Iteration 4: log pseudolikelihood = -3219.1826
Iteration 5: log pseudolikelihood = -3217.5569
Iteration 6: log pseudolikelihood = -3217.4813
Iteration 7: log pseudolikelihood = -3217.481
```

Multinomial logistic regression	Number of obs	=	9822
	Wald chi2 (40)	=	1075.69
	Prob > chi2	=	0.0000
Log pseudolikelihood = −3217.481	Pseudo R2	=	0.1655

y	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
2						
alc90th	.1270931	.2152878	0.59	0.555	-.2948632	.5490494
ue88	.0458099	.0500181	0.92	0.360	-.0522238	.1438436
age	.1617634	.0668732	2.42	0.016	.0306944	.2928324
agesq	-.0024377	.0008087	-3.01	0.003	-.0040227	-.0008527
schooling	-.0092135	.0234188	-0.39	0.694	-.0551135	.0366864
married	.4004928	.204195	1.96	0.050	.0002779	.8007078
famsize	.0622453	.0517847	1.20	0.229	-.0392509	.1637416
white	.0391309	.1711588	0.23	0.819	-.2963342	.3745961
excellent	2.91833	.4548999	6.42	0.000	2.026743	3.809918
verygood	2.978336	.4566665	6.52	0.000	2.083286	3.873386
good	2.493939	.4507366	5.53	0.000	1.610511	3.377366
fair	1.460263	.48807	2.99	0.003	.5036629	2.416862
northeast	.0849125	.23845	0.36	0.722	-.382441	.552266
midwest	.0158816	.2044175	0.08	0.938	-.3847694	.4165326
south	.1750244	.2022599	0.87	0.387	-.2213977	.5714466
centercity	-.2717445	.1911311	-1.42	0.155	-.6463546	.1028656
othermsa	-.0921566	.1955115	-0.47	0.637	-.475352	.2910389
q1	.422405	.1970871	2.14	0.032	.0361213	.8086887
q2	-.0219499	.2049964	-0.11	0.915	-.4237355	.3798357
q3	-.0365295	.2109886	-0.17	0.863	-.4500595	.3770005
_cons	-6.113244	1.412512	-4.33	0.000	-8.881717	-3.344771
3						
alc90th	-.1534987	.1392906	-1.10	0.270	-.4265033	.1195059
ue88	-.0954848	.0335442	-2.85	0.004	-.1612303	-.0297394
age	.227164	.0411389	5.52	0.000	.1465333	.3077948
agesq	-.0030796	.000487	-6.32	0.000	-.004034	-.0021251
schooling	.0890537	.0160584	5.55	0.000	.0575798	.1205276
married	.7085708	.1315325	5.39	0.000	.4507719	.9663698
famsize	.0622447	.035511	1.75	0.080	-.0073556	.1318451
white	.7380044	.1139831	6.47	0.000	.5146017	.9614071
excellent	3.702792	.190178	19.47	0.000	3.33005	4.075534
verygood	3.653313	.1929514	18.93	0.000	3.275135	4.03149
good	2.99946	.1849776	16.22	0.000	2.636911	3.36201
fair	1.876172	.1956878	9.59	0.000	1.492631	2.259713
northeast	.088966	.1505301	0.59	0.555	-.2060675	.3839996
midwest	.1230169	.1302651	0.94	0.345	-.1322981	.3783319
south	.4393047	.1341061	3.28	0.001	.1764616	.7021478
centercity	-.2689532	.1266976	-2.12	0.034	-.5172758	-.0206306
othermsa	.0978701	.1275274	0.77	0.443	-.152079	.3478193
q1	-.0274086	.1288453	-0.21	0.832	-.2799406	.2251235
q2	-.110751	.12602	-0.88	0.379	-.3577457	.1362437
q3	-.0530835	.1321321	-0.40	0.688	-.3120576	.2058907
_cons	-6.237275	.8601993	-7.25	0.000	-7.923235	-4.551316

(y==1 is the base outcome)

b. For the female data, the multinomial logit estimates yield:

```
. mlogit y alc90th ue88 age agesq schooling married famsize white excellent verygood good fair northeast midwest south centercity othermsa q1 q2 q3, baseoutcome(1)
```

Multinomial logistic regression — Number of obs = 12534; LR chi2 (40) = 6680.38; Prob > chi2 = 0.0000; Pseudo R2 = 0.3523

Log likelihood = −6140.5419

y	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
2						
alc90th	-.1241993	.2365754	-0.52	0.600	-.5878785	.3394799
ue88	-.001862	.0514214	-0.04	0.971	-.1026462	.0989221
age	-.0392239	.0612728	-0.64	0.522	-.1593164	.0808687
agesq	.0004834	.0007411	0.65	0.514	-.0009691	.0019359
schooling	-.0121174	.0254645	-0.48	0.634	-.0620269	.037792
married	.0117958	.2220045	0.05	0.958	-.423325	.4469167
famsize	.0092434	.0495871	0.19	0.852	-.0879456	.1064324
white	.2817941	.1935931	1.46	0.146	-.0976414	.6612296
excellent	.0420423	.4579618	0.09	0.927	-.8555463	.939631
verygood	.0449091	.4574373	0.10	0.922	-.8516516	.9414698
good	.0182444	.4544742	0.04	0.968	-.8725086	.9089974
fair	.2925131	.4839658	0.60	0.546	-.6560424	1.241069
northeast	-.1721726	.2163151	-0.80	0.426	-.5961425	.2517973
midwest	-.2643294	.1944624	-1.36	0.174	-.6454687	.11681
south	-.0161982	.1814209	-0.09	0.929	-.3717766	.3393803
centercity	-.0812978	.1869101	-0.43	0.664	-.447635	.2850393
othermsa	.044578	.1738872	0.26	0.798	-.2962347	.3853908
q1	22.30515	1.328553	16.79	0.000	19.70123	24.90906
q2	22.24068	1.32893	16.74	0.000	19.63603	24.84534
q3	18.65596	1.360765	13.71	0.000	15.98891	21.32301
_cons	-23.50938	.	.	.	.	.
3						
alc90th	.1288509	.0812475	1.59	0.113	-.0303912	.2880931
ue88	.0148758	.0188237	0.79	0.429	-.022018	.0517696
age	.0175243	.0230613	0.76	0.447	-.0276751	.0627236
agesq	-.0002381	.00028	-0.85	0.395	-.0007868	.0003106
schooling	.0035127	.0095824	0.37	0.714	-.0152685	.0222939
married	-.0997914	.078327	-1.27	0.203	-.2533095	.0537266
famsize	.0027002	.0184619	0.15	0.884	-.0334844	.0388849
white	-.0277798	.066196	-0.42	0.675	-.1575217	.1019621
excellent	-.1178398	.1636194	-0.72	0.471	-.4385278	.2028483
verygood	-.1170045	.1633395	-0.72	0.474	-.437144	.203135
good	-.1144024	.1622966	-0.70	0.481	-.4324979	.2036931
fair	-.0344312	.1775054	-0.19	0.846	-.3823353	.3134729
northeast	-.0548967	.0819514	-0.67	0.503	-.2155184	.105725
midwest	.0572296	.0720545	0.79	0.427	-.0839946	.1984538

south	.1500468	.0698025	2.15	0.032	.0132365	.2868571
centercity	.0366537	.06934	0.53	0.597	-.0992503	.1725576
othermsa	.0638026	.0655869	0.97	0.331	-.0647453	.1923505
q1	22.22019	.4972432	44.69	0.000	21.24561	23.19477
q2	22.22849	.4972341	44.70	0.000	21.25392	23.20305
q3	19.74301	.4973915	39.69	0.000	18.76815	20.71788
_cons	-21.64045	.	.	.	.	.

(y==1 is the base outcome)

13.17 *Tobit estimation of Married Women Labor Supply*

a. A detailed summary of the hours of work show that mean hours of work is 741, the median is 288, the minimum is zero and the maximum is 4950.

```
. sum hours, detail
```

hours worked, 1975

	Percentiles	Smallest		
1%	0	0		
5%	0	0		
10%	0	0	Obs	753
25%	0	0	Sum of Wgt.	753
50%	288		Mean	740.5764
		Largest	Std. Dev.	871.3142
75%	1516	3640		
90%	1984	3686	Variance	759188.5
95%	2100	4210	Skewness	.9225315
99%	3087	4950	Kurtosis	3.193949

b. Using the notation of solution 11.31, OLS on this model yields

```
. reg hours nwifeinc kidslt6 kidsge6 'control' 'E', r
```

Linear regression

Number of obs	=	753
F(7, 745)	=	45.81
Prob > F	=	0.0000
R-squared	=	0.2656
Root MSE	=	750.18

hours	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf. Interval]	
nwifeinc	-3.446636	2.240662	-1.54	0.124	-7.845398	.9521268
kidslt6	-442.0899	57.46384	-7.69	0.000	-554.9002	-329.2796
kidsge6	-32.77923	22.80238	-1.44	0.151	-77.5438	11.98535
age	-30.51163	4.244791	-7.19	0.000	-38.84481	-22.17846
educ	28.76112	13.03905	2.21	0.028	3.163468	54.35878
exper	65.67251	10.79419	6.08	0.000	44.48186	86.86316
expersq	-.7004939	.3720129	-1.88	0.060	-1.430812	.0298245
_cons	1330.482	274.8776	4.84	0.000	790.8556	1870.109

Tobit estimation with left censoring at zero is represented by the option ll(0)

```
. tobit hours nwifeinc kidslt6 kidsge6 'control' 'E', ll(0)
```

Tobit regression

Number of obs	=	753
LR chi2 (7)	=	271.59
Prob > chi2	=	0.0000
Pseudo R2	=	0.0343

Log likelihood = −3819.0946

hours	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
nwifeinc	-8.814243	4.459096	-1.98	0.048	-17.56811	-.0603724
kidslt6	-894.0217	111.8779	-7.99	0.000	-1113.655	-674.3887
kidsge6	-16.218	38.64136	-0.42	0.675	-92.07675	59.64075
age	-54.40501	7.418496	-7.33	0.000	-68.96862	-39.8414
educ	80.64561	21.58322	3.74	0.000	38.27453	123.0167
exper	131.5643	17.27938	7.61	0.000	97.64231	165.4863
expersq	-1.864158	.5376615	-3.47	0.001	-2.919667	-.8086479
_cons	965.3053	446.4358	2.16	0.031	88.88528	1841.725
/sigma	1122.022	41.57903			1040.396	1203.647

Obs. summary: 325 left-censored observations at hours<=0
428 uncensored observations
0 right-censored observations

c. This replicates Table 17.1 of Wooldridge (2009, p. 585) using Stata

```
. reg inlf nwifeinc kidslt6 kidsge6 'control' 'E' , r
```

Linear regression

Number of obs	=	753
F(7, 745)	=	62.48
Prob > F	=	0.0000
R-squared	=	0.2642
Root MSE	=	.42713

inlf	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
nwifeinc	-.0034052	.0015249	-2.23	0.026	-.0063988	-.0004115
kidslt6	-.2618105	.0317832	-8.24	0.000	-.3242058	-.1994152
kidsge6	.0130122	.0135329	0.96	0.337	-.013555	.0395795
age	-.0160908	.002399	-6.71	0.000	-.0208004	-.0113812
educ	.0379953	.007266	5.23	0.000	.023731	.0522596
exper	.0394924	.00581	6.80	0.000	.0280864	.0508983
expersq	-.0005963	.00019	-3.14	0.002	-.0009693	-.0002233
_cons	.5855192	.1522599	3.85	0.000	.2866098	.8844287

The Logit estimates yield:

```
. logit inlf nwifeinc kidslt6 kidsge6 'control' 'E' , r
Iteration 0: log pseudolikelihood = -514.8732
Iteration 1: log pseudolikelihood = -402.38502
Iteration 2: log pseudolikelihood = -401.76569
Iteration 3: log pseudolikelihood = -401.76515
Iteration 4: log pseudolikelihood = -401.76515
```

```
Logistic regression                              Number of obs  =     753
                                                 Wald chi2 (7)  =  158.48
                                                 Prob > chi2    =  0.0000
Log pseudolikelihood = -401.76515                Pseudo R2      =  0.2197
```

inlf	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
nwifeinc	-.0213452	.0090782	-2.35	0.019	-.039138	-.0035523
kidslt6	-1.443354	.2031615	-7.10	0.000	-1.841543	-1.045165
kidsge6	.0601122	.0798825	0.75	0.452	-.0964546	.2166791
age	-.0880244	.0144393	-6.10	0.000	-.1163248	-.0597239
educ	.2211704	.0444509	4.98	0.000	.1340482	.3082925
exper	.2058695	.0322914	6.38	0.000	.1425796	.2691594
expersq	-.0031541	.0010124	-3.12	0.002	-.0051384	-.0011698
_cons	.4254524	.8597308	0.49	0.621	-1.259589	2.110494

```
. estat classification
```

Logistic model for inlf

Classified	True D	True ~D	Total
+	347	118	465
-	81	207	288
Total	428	325	753

Classified + if predicted Pr(D) >= .5
True D defined as inlf != 0

Sensitivity	Pr(+| D)	81.07%
Specificity	Pr(-| ~D)	63.69%
Positive predictive value	Pr(D| +)	74.62%
Negative predictive value	Pr(~D| -)	71.88%
False + rate for true ~D	Pr(+| ~D)	36.31%
False - rate for true D	Pr(-| D)	18.93%
False + rate for classified +	Pr(~D| +)	25.38%
False - rate for classified -	Pr(D| -)	28.13%
Correctly classified		73.57%

. mfx

Marginal effects after logit
y = Pr(inlf) (predict)
= .58277201

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
nwifeinc	-.0051901	.00221	-2.35	0.019	-.009523	-.000857	20.129
kidslt6	-.3509498	.04988	-7.04	0.000	-.448718	-.253182	.237716
kidsge6	.0146162	.01941	0.75	0.451	-.023428	.05266	1.35325
age	-.021403	.00353	-6.07	0.000	-.028317	-.014489	42.5378
educ	.0537773	.01086	4.95	0.000	.032498	.075057	12.2869
exper	.0500569	.00788	6.35	0.000	.034604	.06551	10.6308
expersq	-.0007669	.00025	-3.11	0.002	-.001251	-.000283	178.039

. margeff

Average partial effects after logit

y = Pr(inlf)

variable	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
nwifeinc	-.0038118	.0015923	-2.39	0.017	-.0069327	-.0006909
kidslt6	-.240805	.0262576	-9.17	0.000	-.292269	-.189341
kidsge6	.0107335	.0142337	0.75	0.451	-.017164	.038631
age	-.0157153	.0023842	-6.59	0.000	-.0203883	-.0110423
educ	.0394323	.0074566	5.29	0.000	.0248176	.0540471
exper	.0367123	.0051935	7.07	0.000	.0265332	.0468914
expersq	-.0005633	.0001767	-3.19	0.001	-.0009096	-.0002169

```
. probit inlf nwifeinc kidslt6 kidsge6 'control' 'E' , r

Iteration 0: log pseudolikelihood = -514.8732
Iteration 1: log pseudolikelihood = -402.06651
Iteration 2: log pseudolikelihood = -401.30273
Iteration 3: log pseudolikelihood = -401.30219
Iteration 4: log pseudolikelihood = -401.30219

Probit regression                                  Number of obs   =     753
                                                   Wald chi2 (7)   =  185.10
                                                   Prob > chi2     =  0.0000
Log pseudolikelihood = -401.30219                  Pseudo R2       =  0.2206
```

inlf	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf. Interval]	
nwifeinc	-.0120237	.0053106	-2.26	0.024	-.0224323	-.0016152
kidslt6	-.8683285	.1162037	-7.47	0.000	-1.096084	-.6405735
kidsge6	.036005	.0452958	0.79	0.427	-.0527731	.124783
age	-.0528527	.0083532	-6.33	0.000	-.0692246	-.0364807
educ	.1309047	.0258192	5.07	0.000	.0803	.1815095
exper	.1233476	.0188537	6.54	0.000	.086395	.1603002
expersq	-.0018871	.0006007	-3.14	0.002	-.0030645	-.0007097
_cons	.2700768	.505175	0.53	0.593	-.7200481	1.260202

```
. mfx

Marginal effects after probit
y = Pr(inlf) (predict)
  = .58154201
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
nwifeinc	-.0046962	.00208	-2.26	0.024	-.008766	-.000626	20.129
kidslt6	-.3391514	.04565	-7.43	0.000	-.428628	-.249675	.237716
kidsge6	.0140628	.01769	0.80	0.427	-.020603	.048729	1.35325
age	-.0206432	.00327	-6.31	0.000	-.027056	-.014231	42.5378
educ	.0511287	.01011	5.06	0.000	.031308	.07095	12.2869
exper	.0481771	.00739	6.52	0.000	.033694	.06266	10.6308
expersq	-.0007371	.00024	-3.14	0.002	-.001198	-.000276	178.039

```
. margeff
```

Average partial effects after probit

y = Pr(inlf)

Variable	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
nwifeinc	-.0036162	.0015759	-2.29	0.022	-.0067049	-.0005275
kidslt6	-.2441788	.0257356	-9.49	0.000	-.2946198	-.1937379
kidsge6	.0108274	.0135967	0.80	0.426	-.0158217	.0374765
age	-.0158917	.0023447	-6.78	0.000	-.0204873	-.011296
educ	.0393088	.0073669	5.34	0.000	.02487	.0537476
exper	.037046	.0051959	7.13	0.000	.0268621	.0472299
expersq	-.0005675	.0001775	-3.20	0.001	-.0009154	-.0002197

. dprobit inlf nwifeinc kidslt6 kidsge6 ‘control’ ‘E’ , r

Iteration 0: log pseudolikelihood = -514.8732

Iteration 1: log pseudolikelihood = -405.78215

Iteration 2: log pseudolikelihood = -401.32924

Iteration 3: log pseudolikelihood = -401.30219

Iteration 4: log pseudolikelihood = -401.30219

Probit regression, reporting marginal effects

Log pseudolikelihood = −401.30219

Number of obs = 753
Wald chi2 (7) = 185.10
Prob > chi2 = 0.0000
Pseudo R2 = 0.2206

inlf	dF/dx	Robust Std. Err.	z	P>|z|	x-bar	[95% C.I.]	
nwifeinc	-.0046962	.0020767	-2.26	0.024	20.129	-.008766	-.000626
kidslt6	-.3391514	.045652	-7.47	0.000	.237716	-.428628	-.249675
kidsge6	.0140628	.0176869	0.79	0.427	1.35325	-.020603	.048729
age	-.0206432	.0032717	-6.33	0.000	42.5378	-.027056	-.014231
educ	.0511287	.010113	5.07	0.000	12.2869	.031308	.07095
exper	.0481771	.0073896	6.54	0.000	10.6308	.033694	.06266
expersq	-.0007371	.000235	-3.14	0.002	178.039	-.001198	-.000276
obs. P	.5683931						
pred. P	.581542	(at x-bar)					

z and P> |z| correspond to the test of the underlying coefficient being 0

```
. estat classification

Probit model for inlf

                 ------ True------
Classified  |       D          ~D   |   Total
------------------------------------------------
+           |     348         120   |     468
-           |      80         205   |     285
------------------------------------------------
Total       |     428         325   |     753

Classified + if predicted Pr(D) >= .5
True D defined as inlf != 0
--------------------------------------------------
Sensitivity                  Pr( +| D)      81.31%
Specificity                  Pr( -| ~D)     63.08%
Positive predictive value    Pr( D| +)      74.36%
Negative predictive value    Pr(~D| -)      71.93%
--------------------------------------------------
False + rate for true ~D     Pr( +| ~D)     36.92%
False - rate for true D      Pr( -| D)      18.69%
False + rate for classified + Pr(~D| +)     25.64%
False - rate for classified - Pr( D| -)     28.07%
--------------------------------------------------
Correctly classified                        73.44%
--------------------------------------------------
```

d. Wooldridge (2009, Chapter 17) recommends one obtain the estimates of (β/σ^2) from a probit using an indicator of labor force participation. Then comparing those with the Tobit estimates generated by dividing β by σ^2. If these estimates are different or have different signs, then the Tobit estimation may not be appropriate. Part (c) gave such probit estimates. For (kidslt6) this was estimated at -0.868. From part (b) the tobit estimation gave a β estimate for (kidslt6) of -894 and an estimate of σ^2 of 1122. The resulting estimate of (β/σ^2) is -0.797. These have the same sign but with different magnitudes.

13.18 *Heckit Estimation of Married Women's Earnings*

a. OLS on this model yields

```
. reg lwage educ exper expersq

Source   |       SS          df        MS           Number of obs  =     428
-------------------------------------------         F( 3, 424)     =   26.29
Model    |   35.0222967       3    11.6740989       Prob > F       =  0.0000
Residual |   188.305144     424    .444115906       R-squared      =  0.1568
-------------------------------------------         Adj R-squared  =  0.1509
Total    |   223.327441     427    .523015084       Root MSE       =  .66642
------------------------------------------------------------------------------
   lwage |     Coef.     Std. Err.      t     P>|t|     [95% Conf. Interval]
------------------------------------------------------------------------------
    educ |   .1074896    .0141465     7.60    0.000     .0796837    .1352956
   exper |   .0415665    .0131752     3.15    0.002     .0156697    .0674633
 expersq |  -.0008112    .0003932    -2.06    0.040    -.0015841   -.0000382
   _cons |  -.5220406    .1986321    -2.63    0.009    -.9124667   -.1316144
------------------------------------------------------------------------------
Heckman two-step estimates

. heckman lwage educ exper expersq, select ( educ exper expersq age kidslt6
kidsge6 nwifeinc) twostep

Heckman selection model – two-step estimates       Number of obs    =     753
(regression model with sample selection)           Censored obs     =     325
                                                   Uncensored obs   =     428

                                                   Wald chi2(3)     =   51.53
                                                   Prob > chi2      =  0.0000
------------------------------------------------------------------------------
   lwage |     Coef.     Std. Err.      z     P>|z|     [95% Conf. Interval]
------------------------------------------------------------------------------
lwage
    educ |   .1090655     .015523     7.03    0.000     .0786411      .13949
   exper |   .0438873    .0162611     2.70    0.007     .0120163    .0757584
 expersq |  -.0008591    .0004389    -1.96    0.050    -.0017194    1.15e-06
   _cons |  -.5781032    .3050062    -1.90    0.058    -1.175904     .019698
------------------------------------------------------------------------------
select
    educ |   .1309047    .0252542     5.18    0.000     .0814074     .180402
   exper |   .1233476    .0187164     6.59    0.000     .0866641    .1600311
 expersq |  -.0018871       .0006    -3.15    0.002     -.003063   -.0007111
     age |  -.0528527    .0084772    -6.23    0.000    -.0694678   -.0362376
 kidslt6 |  -.8683285    .1185223    -7.33    0.000    -1.100628    -.636029
 kidsge6 |    .036005    .0434768     0.83    0.408     -.049208    .1212179
nwifeinc |  -.0120237    .0048398    -2.48    0.013    -.0215096   -.0025378
   _cons |   .2700768     .508593     0.53    0.595    -.7267473    1.266901
------------------------------------------------------------------------------
mills
  lambda |   .0322619    .1336246     0.24    0.809    -.2296376    .2941613
------------------------------------------------------------------------------
     rho |    0.04861
   sigma |  .66362875
  lambda |  .03226186    .1336246
------------------------------------------------------------------------------
```

b. The inverse mills ratio coefficient lambda is estimated to be .032 with a standard error of 0.134 which is not significant. This does not reject the null hypothesis of no sample selection.

c. The MLE of this Heckman (1976) sample selection model.

```
. heckman lwage educ exper expersq, select ( educ exper expersq age kidslt6
kidsge6 nwifeinc)

Iteration 0: log likelihood = -832.89776
Iteration 1: log likelihood = -832.88509
Iteration 2: log likelihood = -832.88508

Heckman selection model                         Number of obs    =     753
(regression model with sample selection)        Censored obs     =     325
                                                Uncensored obs   =     428

                                                Wald chi2(3)     =   59.67
Log likelihood = −832.8851                      Prob > chi2      =  0.0000
```

lwage	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
lwage						
educ	.1083502	.0148607	7.29	0.000	.0792238	.1374767
exper	.0428369	.0148785	2.88	0.004	.0136755	.0719983
expersq	-.0008374	.0004175	-2.01	0.045	-.0016556	-.0000192
_cons	-.5526973	.2603784	-2.12	0.034	-1.06303	-.0423651
select						
educ	.1313415	.0253823	5.17	0.000	.0815931	.1810899
exper	.1232818	.0187242	6.58	0.000	.0865831	.1599806
expersq	-.0018863	.0006004	-3.14	0.002	-.003063	-.0007095
age	-.0528287	.0084792	-6.23	0.000	-.0694476	-.0362098
kidslt6	-.8673988	.1186509	-7.31	0.000	-1.09995	-.6348472
kidsge6	.0358723	.0434753	0.83	0.409	-.0493377	.1210824
nwifeinc	-.0121321	.0048767	-2.49	0.013	-.0216903	-.002574
_cons	.2664491	.5089578	0.52	0.601	-.7310898	1.263988
/athrho	.026614	.147182	0.18	0.857	-.2618573	.3150854
/lnsigma	-.4103809	.0342291	-11.99	0.000	-.4774687	-.3432931
rho	.0266078	.1470778			-.2560319	.3050564
sigma	.6633975	.0227075			.6203517	.7094303
lambda	.0176515	.0976057			-.1736521	.2089552

LR test of indep. eqns. (rho = 0): chi2(1) = 0.03 Prob > chi2 = 0.8577

This yields the same results as the two-step Heckman procedure and the LR test for (rho $=$ 0) is not significant.

References

Carrasco, R. (2001), "Binary Choice with Binary Endogenous Regressors in Panel Data: Estimating the Effect Fertility on Female Labor Participation," Journal of Business & Economic Statistics, 19: 385–394.

Dhillon, U.S., J.D. Shilling and C.F. Sirmans (1987), "Choosing Between Fixed and Adjustable Rate Mortgages," Journal of Money, Credit and Banking, 19: 260–267.

Heckman, J. (1976), "The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables and a Simple Estimator for Such Models," Annals of Economic and Social Measurement, 5: 475–492.

Mullahy, J. and J. Sindelar (1996), "Employment, Unemployment, and Problem Drinking," Journal of Health Economics, 15: 409–434.

Terza, J. (2002), "Alcohol Abuse and Employment: A Second Look," Journal of Applied Econometrics, 17: 393–404.

Wooldridge, J.M. (2009), *Introductory Econometrics: A Modern Approach* (South-Western: Ohio).

CHAPTER 14
Time-Series Analysis

14.1 The AR(1) Model. $y_t = \rho y_{t-1} + \varepsilon_t$ with $|\rho| < 1$ and $\varepsilon_t \sim \text{IIN}\left(0, \sigma_\varepsilon^2\right)$. Also, $y_o \sim \text{N}\left(0, \sigma_\varepsilon^2/1 - \rho^2\right)$.

a. By successive substitution

$$y_t = \rho y_{t-1} + \varepsilon_t = \rho(\rho y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t = \rho^2 y_{t-2} + \rho\varepsilon_{t-1} + \varepsilon_t$$
$$= \rho^2(\rho y_{t-3} + \varepsilon_{t-2}) + \rho\varepsilon_{t-1} + \varepsilon_t = \rho^3 y_{t-3} + \rho^2\varepsilon_{t-2} + \rho\varepsilon_{t-1} + \varepsilon_t$$
$$= \cdots = \rho^t y_o + \rho^{t-1}\varepsilon_1 + \rho^{t-2}\varepsilon_2 + \cdots + \varepsilon_t$$

Then, $E(y_t) = \rho^t E(y_o) = 0$ for every t, since $E(y_o) = E(\varepsilon_t) = 0$.

$$\begin{aligned}\text{var}(y_t) &= \rho^{2t}\text{var}(y_o) + \rho^{2(t-1)}\text{var}(\varepsilon_1) + \rho^{2(t-2)}\text{var}(\varepsilon_2) + \cdots + \text{var}(\varepsilon_t)\\ &= \rho^{2t}\left(\frac{\sigma_\varepsilon^2}{1-\rho^2}\right) + (\rho^{2(t-1)} + \rho^{2(t-2)} + \cdots + 1)\sigma_\varepsilon^2\\ &= \frac{\rho^{2t}}{1-\rho^2}\cdot\sigma_\varepsilon^2 + \frac{1-\rho^{2t}}{1-\rho^2}\cdot\sigma_\varepsilon^2 = \frac{\sigma_\varepsilon^2}{1-\rho^2} = \text{var}(y_o) \quad \text{for every t.}\end{aligned}$$

If $\rho = 1$, then $\text{var}(y_t) = \sigma_\varepsilon^2/0 \to \infty$. Also, if $|\rho| > 1$, then $1 - \rho^2 < 0$ and $\text{var}(y_t) < 0$.

b. The AR(1) series y_t has zero mean and constant variance $\sigma^2 = \text{var}(y_t)$, for $t = 0, 1, 2, \ldots$ In part (a) we could have stopped the successive substitution at y_{t-s}, this yields

$$y_t = \rho^s y_{t-s} + \rho^{s-1}\varepsilon_{t-s+1} + \cdots + \varepsilon_t$$

Therefore, $\text{cov}(y_t, y_{t-s}) = \text{cov}(\rho^s y_{t-s} + \rho^{s-1}\varepsilon_{t-s+1} + \cdots + \varepsilon_t, y_{t-s}) = \rho^s\text{var}(y_{t-s}) = \rho^s\sigma^2$ which only depends on s the distance between t and t-s. Therefore, the AR(1) series y_t is said to be covariance-stationary or weakly stationary.

B.H. Baltagi, *Solutions Manual for Econometrics*, Springer Texts in Business and Economics, DOI 10.1007/978-3-642-54548-1_14,

c. First one generates $y_o = 0.5\,N(0,1)/(1-\rho^2)^{1/2}$ for various values of ρ. Then $y_t = \rho y_{t-1} + \varepsilon_t$ where $\varepsilon_t \sim \text{IIN}(0, 0.25)$ for $t = 1, 2, .., T$.

14.2 The MA(1) Model. $y_t = \varepsilon_t + \theta\varepsilon_{t-1}$ with $\varepsilon_t \sim \text{IIN}\left(0, \sigma_\varepsilon^2\right)$

a. $E(y_t) = E(\varepsilon_t) + \theta E(\varepsilon_{t-1}) = 0$ for all t. Also,

$\text{var}(y_t) = \text{var}(\varepsilon_t) + \theta^2\text{var}(\varepsilon_{t-1}) = (1+\theta^2)\sigma_\varepsilon^2$ for all t.

Therefore, the mean and variance are independent of t.

b. $\text{cov}(y_t, y_{t-1}) = \text{cov}(\varepsilon_t + \theta\varepsilon_{t-1}, \varepsilon_{t-1} + \theta\varepsilon_{t-2}) = \theta\text{var}(\varepsilon_{t-1}) = \theta\sigma_\varepsilon^2$

$$\text{cov}(y_t, y_{t-s}) = \text{cov}(\varepsilon_t + \theta\varepsilon_{t-1}, \varepsilon_{t-s} + \theta\varepsilon_{t-s-1}) = \begin{cases} \theta\sigma_\varepsilon^2 & \text{when } s = 1 \\ 0 & \text{when } s > 1 \end{cases}$$

Since $E(y_t) = 0$ for all t and $\text{cov}(y_t, y_{t-s}) = \begin{cases} \theta\sigma_\varepsilon^2 & \text{when } s = 1 \\ 0 & \text{when } s > 1 \end{cases}$
for all t and s, then the MA(1) process is covariance stationary.

c. First one generates $\varepsilon_t \sim \text{IIN}(0, 0.25)$. Then, for various values of θ, one generates $y_t = \varepsilon_t + \theta\varepsilon_{t-1}$.

14.3

a. The sample autocorrelation function for Income using EViews is as follows:

Correlogram of Income
Sample: 1959 2007
Included observations: 49

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob.
. |*******	. |*******	1	0.935	0.935	45.538	0
. |*******	. | .	2	0.869	-0.043	85.714	0
. |******	. | .	3	0.804	-0.031	120.81	0
. |******	. | .	4	0.737	-0.05	150.95	0
. |*****	. | .	5	0.671	-0.026	176.52	0
. |*****	. | .	6	0.607	-0.026	197.93	0
. |****	. | .	7	0.545	-0.022	215.62	0
. |****	. | .	8	0.484	-0.034	229.91	0
. |***	. | .	9	0.428	-0.006	241.33	0
. |***	. | .	10	0.373	-0.023	250.25	0
. |**	. | .	11	0.324	0.005	257.17	0
. |**	. | .	12	0.278	-0.022	262.37	0
. |**	. | .	13	0.232	-0.025	266.12	0
. |*.	. | .	14	0.189	-0.022	268.68	0
. |*.	. | .	15	0.15	-0.009	270.33	0

. |*.	.*| .	16	0.105	-0.077	271.17	0
. | .	. | .	17	0.061	-0.033	271.46	0
. | .	. | .	18	0.016	-0.056	271.48	0
. | .	. | .	19	-0.029	-0.038	271.56	0
.*| .	. | .	20	-0.072	-0.024	272	0

The sample autocorrelation function for differenced Income is as follows:

Correlogram of Differenced Income
Sample: 1959 2007
Included observations: 48

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob.
. | .	. | .	1	0.021	0.021	0.0234	0.879
. |*.	. |*.	2	0.145	0.145	1.1253	0.57
.*| .	.*| .	3	-0.141	-0.15	2.1907	0.534
. | .	. | .	4	-0.008	-0.022	2.1944	0.7
.*| .	.*| .	5	-0.145	-0.106	3.3743	0.642
. |*.	. |*.	6	0.112	0.11	4.0884	0.665
**| .	**| .	7	-0.309	-0.306	9.6689	0.208
. | .	. | .	8	0.05	0.026	9.8171	0.278
**| .	.*| .	9	-0.232	-0.166	13.135	0.157
.*| .	**| .	10	-0.129	-0.228	14.18	0.165
. |*.	. |*.	11	0.077	0.182	14.561	0.203
. |*.	. | .	12	0.137	0.024	15.804	0.2
. | .	. | .	13	-0.019	-0.056	15.83	0.258
. |**	. |*.	14	0.209	0.11	18.905	0.169
. | .	. | .	15	-0.057	-0.022	19.145	0.207
. |*.	. | .	16	0.113	0.043	20.371	0.256
.*| .	.*| .	17	-0.06	-0.149	20.094	0.216
.*| .	. | .	18	-0.074	-0.054	20.813	0.289
. |*.	. |*.	19	0.083	0.185	21.386	0.316
. |*.	. | .	20	0.089	0	22.069	0.337

b. The Augmented Dickey–Fuller test statistic for Income using a constant and linear trend is as follows:

Null Hypothesis: Y has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic based on SIC, MAXLAG=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.843738	0.6677
Test critical values:	1% level	-4.161144	
	5% level	-3.506374	
	10% level	-3.183002	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(Y)
Method: Least Squares

Sample (adjusted): 1960 2007
Included observations: 48 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
Y(-1)	-0.165599	0.089817	-1.843738	0.0718
C	1771.254	798.2947	2.218797	0.0316
@TREND(1959)	67.25568	34.64899	1.941058	0.0585

R-squared	0.100578	Mean dependent var	394.3542
Adjusted R-squared	0.060604	S.D. dependent var	277.5483
S.E. of regression	269.0066	Akaike info criterion	14.08781
Sum squared resid	3256405.	Schwarz criterion	14.20476
Log likelihood	-335.1074	Hannan-Quinn criter.	14.13201
F-statistic	2.516070	Durbin-Watson stat	1.806165
Prob(F-statistic)	0.092083		

This does not reject the null hypothesis of unit root for Income.

c. The Augmented Dickey–Fuller test statistic for differenced Income using a constant and linear trend is as follows:

Null Hypothesis: D(Y) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic based on SIC, MAXLAG=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-6.717325	0.0000
Test critical values:	1% level	-4.165756	
	5% level	-3.508508	
	10% level	-3.184230	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(Y,2)
Method: Least Squares

Sample (adjusted): 1961 2007
Included observations: 47 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
D(Y(-1))	-1.002187	0.149194	-6.717325	0.0000
C	330.8657	96.73746	3.420244	0.0014
@TREND(1959)	2.866879	3.046308	0.941100	0.3518

R-squared	0.506532	Mean dependent var	9.914894
Adjusted R-squared	0.484102	S.D. dependent var	388.6497
S.E. of regression	279.1518	Akaike info criterion	14.16309
Sum squared resid	3428732.	Schwarz criterion	14.28118
Log likelihood	-329.8326	Hannan-Quinn criter.	14.20753

F-statistic	22.58242	Durbin-Watson stat	2.018360
Prob(F-statistic)	0.000000		

This rejects the null hypothesis of unit root for differenced Income. We conclude that Income is I(1).

d. Let R1 denote the ols residuals from the regression of Consumption on Income and a constant. This tests R1 for unit roots: This ADF includes a constant

Null Hypothesis: R1 has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic based on SIC, MAXLAG=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.502798	0.5237
Test critical values:	1% level	-3.574446	
	5% level	-2.923780	
	10% level	-2.599925	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(R1)
Method: Least Squares

Sample (adjusted): 1960 2007
Included observations: 48 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
R1(-1)	-0.094140	0.062643	-1.502798	0.1397
C	-1.110596	26.49763	-0.041913	0.9667

R-squared	0.046798	Mean dependent var	0.149870
Adjusted R-squared	0.026076	S.D. dependent var	185.9291
S.E. of regression	183.4889	Akaike info criterion	13.30296
Sum squared resid	1548737.	Schwarz criterion	13.38093
Log likelihood	-317.2710	Hannan-Quinn criter.	13.33242
F-statistic	2.258401	Durbin-Watson stat	2.408726
Prob(F-statistic)	0.139726		

This ADF includes a constant and a linear trend.

Null Hypothesis: R1 has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 1 (Automatic based on SIC, MAXLAG=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.000195	0.9342
Test critical values:	1% level	-4.165756	
	5% level	-3.508508	
	10% level	-3.184230	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(R1)
Method: Least Squares

Sample (adjusted): 1961 2007
Included observations: 47 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
R1(-1)	-0.062470	0.062458	-1.000195	0.3228
D(R1(-1))	-0.324625	0.145177	-2.236060	0.0306
C	-121.4508	54.46854	-2.229742	0.0310
@TREND(1959)	4.793820	1.937792	2.473856	0.0174

R-squared	0.209855	Mean dependent var	-0.103081
Adjusted R-squared	0.154728	S.D. dependent var	187.9309
S.E. of regression	172.7812	Akaike info criterion	13.22319
Sum squared resid	1283693.	Schwarz criterion	13.38065
Log likelihood	-306.7451	Hannan-Quinn criter.	13.28245
F-statistic	3.806787	Durbin-Watson stat	2.032797
Prob(F-statistic)	0.016614		

Both ADF tests do not reject the null hypothesis of unit roots in the ols residuals.

f. Correlogram of log(consumption)

Correlogram of log(consumption)
Sample: 1959 2007
Included Observations: 49

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
. |*******	. |*******	1	0.937	0.937	45.664	0
. |*******	.*| .	2	0.87	-0.059	85.895	0
. |******	. | .	3	0.801	-0.053	120.74	0
. |******	. | .	4	0.733	-0.027	150.61	0
. |*****	. | .	5	0.667	-0.03	175.86	0
. |*****	. | .	6	0.603	-0.021	196.97	0
. |****	. | .	7	0.541	-0.017	214.41	0
. |****	. | .	8	0.482	-0.024	228.6	0
. |***	. | .	9	0.425	-0.027	239.89	0
. |***	. | .	10	0.373	0.003	248.81	0
. |**	. | .	11	0.324	-0.016	255.73	0
. |**	. | .	12	0.275	-0.041	260.85	0
. |**	. | .	13	0.228	-0.027	264.44	0
. |*.	. | .	14	0.183	-0.011	266.84	0
. |*.	. | .	15	0.142	-0.01	268.33	0
. |*.	.*| .	16	0.099	-0.063	269.07	0
. | .	. | .	17	0.055	-0.038	269.3	0
. | .	. | .	18	0.012	-0.036	269.32	0
. | .	. | .	19	-0.029	-0.028	269.39	0
.*| .	. | .	20	-0.068	-0.022	269.78	0

For log(consumption), the ADF with a Constant and Linear Trend yields:

Null Hypothesis: LOGC has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 1 (Automatic based on SIC, MAXLAG=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-3.201729	0.0965
Test critical values:	1% level	-4.165756	
	5% level	-3.508508	
	10% level	-3.184230	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(LOGC)
Method: Least Squares
Sample (adjusted): 1961 2007
Included observations: 47 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
LOGC(-1)	-0.216769	0.067704	-3.201729	0.0026
D(LOGC(-1))	0.447493	0.129784	3.447996	0.0013
C	1.987148	0.614623	3.233118	0.0024
@TREND(1959)	0.004925	0.001590	3.097180	0.0034

R-squared	0.311464	Mean dependent var	0.024014
Adjusted R-squared	0.263427	S.D. dependent var	0.015980
S.E. of regression	0.013715	Akaike info criterion	-5.659384
Sum squared resid	0.008088	Schwarz criterion	-5.501924
Log likelihood	136.9955	Hannan-Quinn criter.	-5.600131
F-statistic	6.483785	Durbin-Watson stat	1.978461
Prob(F-statistic)	0.001019		

We do not reject the null hypothesis that log(consumption) has unit root at the 5% level, but we do so at the 10% level.

For differenced log(consumption), the ADF with a Constant and Linear Trend yields:

Null Hypothesis: D(LOGC) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 1 (Automatic based on SIC, MAXLAG=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-5.143889	0.0006
Test critical values:	1% level	-4.170583	
	5% level	-3.510740	
	10% level	-3.185512	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(LOGC,2)
Method: Least Squares
Sample (adjusted): 1962 2007
Included observations: 46 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
D(LOGC(-1))	-0.848482	0.164950	-5.143889	0.0000
D(LOGC(-1),2)	0.253834	0.143686	1.766584	0.0846
C	0.026204	0.006566	3.990611	0.0003
@TREND(1959)	-0.000215	0.000165	-1.307036	0.1983

R-squared	0.403764	Mean dependent var	0.000300
Adjusted R-squared	0.361175	S.D. dependent var	0.018320
S.E. of regression	0.014642	Akaike info criterion	-5.526861
Sum squared resid	0.009005	Schwarz criterion	-5.367849
Log likelihood	131.1178	Hannan-Quinn criter.	-5.467294
F-statistic	9.480625	Durbin-Watson stat	2.028804
Prob(F-statistic)	0.000066		

We do reject the null hypothesis that differenced log(consumption) has unit root at the 5% level.

Correlogram of log Income
Sample: 1959 2007
Included Observations: 49

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
. |*******	. |*******	1	0.935	0.935	45.507	0
. |*******	. | .	2	0.867	-0.057	85.463	0
. |******	. | .	3	0.798	-0.042	120.07	0
. |******	. | .	4	0.729	-0.041	149.58	0
. |*****	. | .	5	0.661	-0.035	174.36	0
. |*****	. | .	6	0.596	-0.012	194.98	0
. |****	. | .	7	0.534	-0.016	211.96	0
. |****	. | .	8	0.475	-0.026	225.69	0

. |***	. | .	9	0.418	-0.015	236.64	0
. |***	. | .	10	0.365	-0.021	245.17	0
. |**	. | .	11	0.315	-0.009	251.7	0
. |**	. | .	12	0.268	-0.023	256.54	0
. |**	. | .	13	0.222	-0.023	259.97	0
. |*.	. | .	14	0.179	-0.017	262.26	0
. |*.	. | .	15	0.141	0.002	263.72	0
. |*.	.*| .	16	0.098	-0.079	264.44	0
. | .	. | .	17	0.055	-0.035	264.68	0
. | .	. | .	18	0.011	-0.043	264.69	0
. | .	. | .	19	-0.031	-0.034	264.77	0
.*| .	. | .	20	-0.071	-0.018	265.2	0

For log(Income), the ADF with a Constant and Linear Trend yields:

Null Hypothesis: LOGY has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic based on SIC, MAXLAG=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.567210	0.7912
Test critical values:	1% level	-4.161144	
	5% level	-3.506374	
	10% level	-3.183002	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation

Dependent Variable: D(LOGY)

Method: Least Squares

Sample (adjusted): 1960 2007

Included observations: 48 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
LOGY(-1)	-0.079684	0.050844	-1.567210	0.1241
C	0.765061	0.469055	1.631070	0.1099
@TREND(1959)	0.001460	0.001136	1.285021	0.2054

R-squared	0.118240	Mean dependent var	0.022569
Adjusted R-squared	0.079051	S.D. dependent var	0.016010
S.E. of regression	0.015364	Akaike info criterion	-5.453080
Sum squared resid	0.010623	Schwarz criterion	-5.336130
Log likelihood	133.8739	Hannan-Quinn criter.	-5.408884
F-statistic	3.017149	Durbin-Watson stat	1.746698
Prob(F-statistic)	0.058936		

We do not reject the null hypothesis that log(income) has unit root at the 5% level.

For differenced log(income), the ADF with a Constant and Linear Trend yields:

Null Hypothesis: D(LOGY) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic based on SIC, MAXLAG=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-6.336919	0.0000
Test critical values :	1% level	-4.165756	
	5% level	-3.508508	
	10% level	-3.184230	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation

Dependent Variable: D(LOGY,2)

Method: Least Squares

Sample (adjusted): 1961 2007

Included observations: 47 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
D(LOGY(-1))	-0.924513	0.145893	-6.336919	0.0000
C	0.029916	0.006480	4.616664	0.0000
@TREND(1959)	-0.000347	0.000172	-2.019494	0.0496

R-squared	0.477988	Mean dependent var	0.000278
Adjusted R-squared	0.454260	S.D. dependent var	0.020895
S.E. of regression	0.015436	Akaike info criterion	-5.442528
Sum squared resid	0.010484	Schwarz criterion	-5.324434
Log likelihood	130.8994	Hannan-Quinn criter.	-5.398088
F-statistic	20.14464	Durbin-Watson stat	2.057278
Prob(F-statistic)	0.000001		

We do reject the null hypothesis that differenced log(income) has unit root at the 5% level.

The OLS regression of log(consumption) on log(income) and a constant yields:

Dependent Variable: LOGC

Method: Least Squares

Sample: 1959 2007

Included observations: 49

	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.625988	0.107332	-5.832253	0.0000
LOGY	1.053340	0.010972	95.99861	0.0000

R-squared	0.994926	Mean dependent var	9.672434
Adjusted R-squared	0.994818	S.D. dependent var	0.335283
S.E. of regression	0.024136	Akaike info criterion	-4.570278
Sum squared resid	0.027379	Schwarz criterion	-4.493061
Log likelihood	113.9718	Hannan-Quinn criter.	-4.540982
F-statistic	9215.733	Durbin-Watson stat	0.205803
Prob(F-statistic)	0.000000		

Let R2 denote the ols residuals from the regression of Log(Consumption) on log(Income) and a constant.

This tests R2 for unit roots:

Null Hypothesis: R2 has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic based on SIC, MAXLAG=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.898473	0.6399
Test critical values:	1% level	-4.161144	
	5% level	-3.506374	
	10% level	-3.183002	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation

Dependent Variable: D(R2)

Method: Least Squares

Sample (adjusted): 1960 2007

Included observations: 48 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
R2(-1)	-0.122648	0.064603	-1.898473	0.0641
C	-0.005115	0.003074	-1.663739	0.1031
@TREND(1959)	0.000201	0.000109	1.838000	0.0727

R-squared	0.126883	Mean dependent var	-0.000115
Adjusted R-squared	0.088078	S.D. dependent var	0.010949
S.E. of regression	0.010455	Akaike info criterion	-6.222927
Sum squared resid	0.004919	Schwarz criterion	-6.105977
Log likelihood	152.3503	Hannan-Quinn criter.	-6.178732
F-statistic	3.269755	Durbin-Watson stat	2.497289
Prob(F-statistic)	0.047220		

We do not reject the null hypothesis of unit roots in these ols residuals.

Johansen's Cointegration Tests for Log(consumption) and log(Y)
Sample (adjusted): 1961 2007
Included observations: 47 after adjustments
Trend assumption: Linear deterministic trend
Series: LOGC LOGY
Lags interval (in first differences): 1 to 1

Unrestricted Cointegration Rank Test (Trace)

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None	0.253794	14.26745	15.49471	0.0759
At most 1	0.010751	0.508018	3.841466	0.4760

Trace test indicates no cointegration at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
**MacKinnon-Haug-Michelis (1999) p-values

Unrestricted Cointegration Rank Test (Maximum Eigenvalue)

Hypothesized No. of CE(s)	Eigenvalue	Max-Eigen Statistic	0.05 Critical Value	Prob.**
None	0.253794	13.75944	14.26460	0.0600
At most 1	0.010751	0.508018	3.841466	0.4760

Max-eigenvalue test indicates no cointegration at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
**MacKinnon-Haug-Michelis (1999) p-values

Both tests indicate no cointegration between Log(consumption) and log(Y) at the 5% level

GARCH(1,1) for Log(consumption) and log(Y)

Dependent Variable: LOGC
Method: ML - ARCH (Marquardt) - Normal distribution

Sample: 1959 2007
Included observations: 49
Convergence achieved after 20 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(3) + C(4)*RESID(-1)^2 + C(5)*GARCH(-1)

Coefficient	Std. Error	z-Statistic	Prob.	
C	-0.712055	0.077795	-9.152987	0.0000
LOGY	1.063091	0.007986	133.1172	0.0000
		Variance Equation		
C	0.000127	0.000114	1.112565	0.2659
RESID(-1)^2	1.332738	0.542199	2.458025	0.0140
GARCH(-1)	-0.195201	0.120479	-1.620211	0.1052

R-squared	0.994060	Mean dependent var	9.672434
Adjusted R-squared	0.993521	S.D. dependent var	0.335283
S.E. of regression	0.026989	Akaike info criterion	-4.800472
Sum squared resid	0.032049	Schwarz criterion	-4.607429
Log likelihood	122.6116	Hannan-Quinn criter.	-4.727232
F-statistic	1841.004	Durbin-Watson stat	0.178014
Prob(F-statistic)	0.000000		

14.5 Data Description: This data is obtained from the Citibank data base.

M1: is the seasonally adjusted monetary base. This a monthly average series. We get the quarterly average of M1 by using $(M1_t + M1_{t+1} + M1_{t+2})/3$.

TBILL3: is the T-bill-3 month-rate. This is a monthly series. We calculate a quarterly average of TBILL3 by using $(TBILL3_t + TBILL3_{t+1} + TBILL3_{t+2})/3$. Note that TBILL3 is an annualized rate (per annum).

GNP: This is Quarterly GNP. All series are transformed by taking their natural logarithm.

a. VAR with two lags on each variable

Sample(adjusted): 1959:3 1995:2
Included observations: 144 after adjusting endpoints
Standard errors & t-statistics in parentheses

	LNGNP	LNM1	LNTBILL3
LNGNP(-1)	1.135719	−0.005500	1.437376
	(0.08677)	(0.07370)	(1.10780)
	(13.0886)	(−0.07463)	(1.29751)
LNGNP(-2)	−0.130393	0.037241	−1.131462
	(0.08750)	(0.07431)	(1.11705)
	(−1.49028)	(0.50115)	(−1.01290)

LNM1(-1)	0.160798	1.508925	1.767750
	(0.07628)	(0.06478)	(0.97383)
	(2.10804)	(23.2913)	(1.81525)
LNM1(-2)	−0.163492	−0.520561	−1.892962
	(0.07516)	(0.06383)	(0.95951)
	(−2.17535)	(−8.15515)	(−1.97284)
LNTBILL3(-1)	0.001615	−0.036446	1.250074
	(0.00645)	(0.00547)	(0.08230)
	(0.25047)	(−6.65703)	(15.1901)
LNTBILL3(-2)	−0.008933	0.034629	−0.328626
	(0.00646)	(0.00549)	(0.08248)
	(−1.38286)	(6.31145)	(−3.98453)
C	−0.011276	−0.179754	−1.656048
	(0.07574)	(0.06433)	(0.96696)
	(−0.14888)	(−2.79436)	(−1.71264)

R-squared	0.999256	0.999899	0.946550
Adj. R-squared	0.999223	0.999895	0.944209
Sum sq. resids	0.009049	0.006527	1.474870
S.E. equation	0.008127	0.006902	0.103757
Log likelihood	492.2698	515.7871	125.5229
Akaike AI C	-9.577728	-9.904358	-4.484021
Schwarz SC	-9.433362	-9.759992	-4.339656
Mean dependent	8.144045	5.860579	1.715690
S.D. dependent	0.291582	0.672211	0.439273

Determinant Residual Covariance	2.67E-11
Log Likelihood	1355.989
Akaike Information Criteria	−24.24959
Schwarz Criteria	−24.10523

b. VAR with three lags on each variable

Sample(adjusted): 1959:4 1995:2
Included observations: 143 after adjusting endpoints
Standard errors & t-statistics in parentheses

	LNGNP	LNM1	LNTBILL3
LNGNP(-1)	1.133814	−0.028308	1.660761
	(0.08830)	(0.07328)	(1.11241)
	(12.8398)	(−0.38629)	(1.49295)
LNGNP(-2)	−0.031988	0.103428	0.252378
	(0.13102)	(0.10873)	(1.65053)
	(−0.24414)	(0.95122)	(0.15291)

LNGNP(-3)	−0.105146	−0.045414	−1.527252
	(0.08774)	(0.07281)	(1.10526)
	(−1.19842)	(−0.62372)	(−1.38180)
LNM1(-1)	0.098732	1.375936	1.635398
	(0.10276)	(0.08528)	(1.29449)
	(0.96081)	(16.1349)	(1.26335)
LNM1(-2)	−0.012617	−0.134075	−3.555324
	(0.17109)	(0.14198)	(2.15524)
	(−0.07375)	(−0.94432)	(−1.64962)
LNM1(-3)	−0.085778	−0.253402	1.770995
	(0.09254)	(0.07680)	(1.16577)
	(−0.92693)	(−3.29962)	(1.51917)
LNTBILL3(-1)	0.001412	−0.041461	1.306043
	(0.00679)	(0.00564)	(0.08555)
	(0.20788)	(−7.35638)	(15.2657)
LNTBILL3(-2)	−0.013695	0.039858	−0.579077
	(0.01094)	(0.00908)	(0.13782)
	(−1.25180)	(4.38997)	(−4.20158)
LNTBILL3(-3)	0.006468	0.000144	0.207577
	(0.00761)	(0.00632)	(0.09588)
	(0.84990)	(0.02281)	(2.16504)
C	0.037812	−0.166320	−2.175434
	(0.07842)	(0.06508)	(0.98789)
	(0.48217)	(−2.55566)	(−2.20210)

R-squared	0.999271	0.999907	0.950041
Adj. R-squared	0.999222	0.999901	0.946661
Sum sq. resids	0.008622	0.005938	1.368186
S.E. equation	0.008051	0.006682	0.101425
Log likelihood	491.8106	518.4767	129.5215
Akaike AIC	-9.576480	-9.949432	-4.509499
Schwarz SC	-9.369288	-9.742240	-4.302307
Mean dependent	8.148050	5.866929	1.718861
S.D. dependent	0.288606	0.670225	0.439160

Determinant Residual Covariance	2.18E-11
Log Likelihood	1360.953
Akaike Information Criteria	−24.40808
Schwarz Criteria	−24.20088

d. Pairwise Granger Causality Tests

Sample: 1959:1 1995:2

Lags: 3

Null Hypothesis:	Obs	F-Statistic	Probability
LNTBILL3 does not Granger Cause LNM1	143	20.0752	7.8E-11
LNM1 does not Granger Cause LNTBILL3		1.54595	0.20551

e. Pairwise Granger Causality Tests

Pairwise Granger Causality Tests

Sample: 1959:1 1995:2

Lags: 2

Null Hypothesis:	Obs	F-Statistic	Probability
LNTBILL3 does not Granger Cause LNM1	144	23.0844	2.2E-09
LNM1 does not Granger Cause LNTBILL3		3.99777	0.02051

14.6 *The Simple Deterministic Time Trend Model.* This is based on Hamilton (1994).

$$y_t = \alpha + \beta t + u_t \qquad t = 1, \ldots, T \qquad \text{where } u_t \sim \text{IIN}(0, \sigma^2).$$

In vector form, this can be written as

$$y = X\varphi + u, \quad X = [1, t], \quad \varphi = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}.$$

a. $\hat{\varphi}_{ols} = (X'X)^{-1}X'y$ and $\hat{\varphi}_{ols} - \varphi = (X'X)^{-1}X'u$

Therefore,

$$\hat{\varphi}_{ols} - \varphi = \begin{bmatrix} \hat{\alpha}_{ols} - \alpha \\ \hat{\beta}_{ols} - \beta \end{bmatrix} = \left[\begin{pmatrix} 1' \\ t' \end{pmatrix}(1, t)\right]^{-1} \begin{pmatrix} 1' \\ t' \end{pmatrix} u$$

$$= \begin{bmatrix} 1'1 & 1't \\ t'1 & t't \end{bmatrix}^{-1} \begin{bmatrix} 1'u \\ t'u \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{t=1}^{T} 1 & \sum_{t=1}^{T} t \\ \sum_{t=1}^{T} t & \sum_{t=1}^{T} t^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^{T} u_t \\ \sum_{t=1}^{T} tu_t \end{bmatrix} = \begin{bmatrix} T & \sum_{t=1}^{T} t \\ \sum_{t=1}^{T} t & \sum_{t=1}^{T} t^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^{T} u_t \\ \sum_{t=1}^{T} tu_t \end{bmatrix}$$

b. $$\frac{X'X}{T} = T^{-1}\begin{bmatrix} T & \sum_{t=1}^{T} t \\ \sum_{t=1}^{T} t & \sum_{t=1}^{T} t^2 \end{bmatrix} = T^{-1}\begin{bmatrix} T & T(T+1)/2 \\ T(T+1)/2 & T(T+1)(2T+1)/6 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & (T+1)/2 \\ (T+1)/2 & (T+1)(2T+1)/6 \end{bmatrix}$$

Therefore, $\text{plim}(\frac{X'X}{T})$ diverges as $T \to \infty$ and is not a positive definite matrix.

c. Note that

$$A^{-1}(X'X)A^{-1} = \begin{bmatrix} T^{-\frac{1}{2}} & 0 \\ 0 & T^{-\frac{3}{2}} \end{bmatrix}\begin{bmatrix} \sum_{t=1}^{T} 1 & \sum_{t=1}^{T} t \\ \sum_{t=1}^{T} t & \sum_{t=1}^{T} t^2 \end{bmatrix}\begin{bmatrix} T^{-\frac{1}{2}} & 0 \\ 0 & T^{-\frac{3}{2}} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{T}\sum_{t=1}^{T} 1 & \frac{1}{T^2}\sum_{t=1}^{T} t \\ \frac{1}{T^2}\sum_{t=1}^{T} t & \frac{1}{T^3}\sum_{t=1}^{T} t^2 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & \frac{T(T+1)}{2T^2} \\ \frac{T(T+1)}{2T^2} & \frac{T(T+1)(2T+1)}{6T^3} \end{bmatrix} = \begin{bmatrix} 1 & \frac{1}{2}+\frac{1}{2T} \\ \frac{1}{2}+\frac{1}{2T} & \frac{1}{3}+\frac{1}{2T}+\frac{1}{6T^2} \end{bmatrix}$$

Therefore plim $A^{-1}(X'X)A^{-1} = \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{bmatrix} = Q$ as $T \to \infty$ which is a finite positive definite matrix. Also

$$A^{-1}(X'u) = \begin{bmatrix} \frac{1}{\sqrt{T}} & 0 \\ 0 & \frac{1}{T\sqrt{T}} \end{bmatrix}\begin{bmatrix} 1'u \\ t'u \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{T}} & 0 \\ 0 & \frac{1}{T\sqrt{T}} \end{bmatrix}\begin{bmatrix} \sum_{t=1}^{T} u_t \\ \sum_{t=1}^{T} tu_t \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_t \\ \frac{1}{T\sqrt{T}}\sum_{t=1}^{T} tu_t \end{bmatrix}$$

d. Show that $z_1 = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_t \sim N(0,\sigma^2)$. $u_t \sim N(0,\sigma^2)$, so that $\sum_{t=1}^{T} u_t \sim N(0, T\sigma^2)$.

Therefore, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_t \sim N\left(0, \frac{1}{\sqrt{T}} \cdot T\sigma^2 \cdot \frac{1}{\sqrt{T}}\right) = N(0,\sigma^2)$.

Also, show that $z_2 = \frac{1}{T\sqrt{T}}\sum_{t=1}^{T} tu_t \sim N\left[0, \frac{\sigma^2}{6T^2} \cdot (T+1)(2T+1)\right]$.

Let $\sigma_t^2 = \text{var}\left(\frac{t}{T}u_t\right) = E\left(\frac{t}{T}u_t\right)^2 = \frac{t^2}{T^2}\sigma^2$. Then,

$$\frac{1}{T}\sum_{t=1}^{T}\sigma_t^2 = \frac{\sigma^2}{T^3}\sum_{t=1}^{T}t^2 = \frac{\sigma^2}{T^3}\cdot\frac{T(T+1)(2T+1)}{6} = \frac{\sigma^2(T+1)(2T+1)}{6T^2}$$

Since $\text{var}\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T}tu_t\right) = \frac{1}{T}\text{var}\left(\sum_{t-1}^{T}\frac{t}{T}u_t\right) = \frac{1}{T}\sum_{t=1}^{T}\sigma_t^2 = \frac{\sigma^2(T+1)(2T+1)}{6T^2}$

Therefore $\frac{1}{T\sqrt{T}}\sum_{t=1}^{T}tu_t \sim N\left(0, \frac{\sigma^2(T+1)(2T+1)}{6T^2}\right)$

Now $z_1 = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}u_t \sim N(0,\sigma^2)$ and

$$z_2 = \frac{1}{T\sqrt{T}}\sum_{t=1}^{T}tu_t \sim N\left(0, \frac{(T+1)(2T+1)}{6T^2}\sigma^2\right)$$

with

$$\begin{aligned}\text{cov}(z_1,z_2) = E(z_1z_2) = E\left(\sum_{t=1}^{T}\frac{t}{T^2}u_t^2\right) &= \sigma^2\sum_{t=1}^{T}t/T^2 \\ &= T(T+1)\sigma^2/2T^2 = (T+1)\sigma^2/2T.\end{aligned}$$

Hence,

$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \sim N\left(0, \sigma^2\begin{bmatrix} 1 & \frac{T+1}{2T} \\ \frac{T+1}{2T} & \frac{(T+1)(2^T+1)}{6T^2} \end{bmatrix}\right)$$

Therefore as $T \to \infty$, $\begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$ has an asymptotic distribution that is

$$N\left(0, \sigma^2\begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{bmatrix}\right) \text{ or } N(0,\sigma^2 Q).$$

e. $\begin{bmatrix} \sqrt{T}\,(\hat{\alpha}_{ols} - \alpha) \\ T\sqrt{T}\left(\hat{\beta}_{ols} - \beta\right) \end{bmatrix} = [A^{-1}(X'X)A^{-1}]^{-1}[A^{-1}(X'u)]$

But from part (c), we have plim $A^{-1}(X'X)A^{-1}$ is Q which is finite and positive definite. Therefore plim$[A^{-1}(X'X)A^{-1}]^{-1}$ is Q^{-1}. Also, from part (d)

$$A^{-1}(X'u) = \begin{bmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^{T}u_t \\ \frac{1}{T\sqrt{T}}\sum_{t=1}^{T}tu_t \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$$

has an asymptotic distribution $N(0, \sigma^2 Q)$. Hence, $[A^{-1}(X'X)A^{-1}]^{-1}$ $[A^{-1}(X'u)]$ has an asymptotic distribution $N(0, Q^{-1}\sigma^2 QQ^{-1})$ or $N(0, \sigma^2 Q^{-1})$. Thus,

$$\begin{bmatrix} \sqrt{T}\,(\hat{\alpha}_{ols} - \alpha) \\ T\sqrt{T}\left(\hat{\beta}_{ols} - \beta\right) \end{bmatrix},$$

has an asymptotic distribution $N(0, \sigma^2 Q^{-1})$. Since $\hat{\beta}_{ols}$ has the factor $T\sqrt{T}$ rather than the usual $\sqrt{T}$, it is said to be *superconsistent*. This means that not only does $(\hat{\beta}_{ols} - \beta)$ converge to zero in probability limits, but so does $T(\hat{\beta}_{ols} - \beta)$. Note that the normality assumption is not needed for this result. Using the central limit theorem, all that is needed is that u_t is White noise with finite fourth moments, see Sims et al. (1990) or Hamilton (1994).

14.7 *Test of Hypothesis with a Deterministic Time Trend Model*. This is based on Hamilton (1994).

a. Show that plim $s^2 = \frac{1}{T-2}\sum_{t=1}^{T}(y_t - \hat{\alpha}_{ols} - \hat{\beta}_{ols}t)^2 = \sigma^2$. By the law of large numbers $\frac{1}{T-2}\sum_{t=1}^{T}(y_t - \hat{\alpha}_{ols} - \hat{\beta}_{ols}t)^2 = \frac{1}{T-2}\sum_{t=1}^{T}u_t^2$.

Hence, plim $s^2 = \text{plim}\frac{1}{T-2}\sum_{t=1}^{T}u_t^2 = \text{var}(u_t) = \sigma^2$

b. Show that $t_\alpha = \frac{\hat{\alpha}_{ols}-\alpha_o}{\left[s^2(1,0)(X'X)^{-1}\begin{bmatrix}1\\0\end{bmatrix}\right]^{\frac{1}{2}}}$ has the same asymptotic N(0,1) distribution as $t_\alpha^* = \frac{\sqrt{T}(\hat{\alpha}_{ols}-\alpha_o)}{\sigma\sqrt{q^{11}}}$. Multiplying both the numerator and denominator of t_α by $\sqrt{T}$ gives

$$t_\alpha = \frac{\sqrt{T}\,(\hat{\alpha}_{ols} - \alpha_o)}{\left[s^2(\sqrt{T}, 0)(X'X)^{-1}\begin{bmatrix}\sqrt{T}\\0\end{bmatrix}\right]^{\frac{1}{2}}}$$

From problem 14.6, part (c), we showed that for $A = \begin{bmatrix} \sqrt{T} & 0 \\ 0 & T\sqrt{T} \end{bmatrix}$ the $\text{plim}[A^{-1}(X'X)A^{-1}]^{-1} = Q^{-1}$ where

$$Q = \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{bmatrix}.$$

Now, $[\sqrt{T}, 0]$ in t_α can be rewritten as [1,0]A, because

$$[1,0]A = [1,0]\begin{bmatrix} \sqrt{T} & 0 \\ 0 & T\sqrt{T} \end{bmatrix} = [\sqrt{T}, 0].$$

Therefore, $t_\alpha = \dfrac{\sqrt{T}(\hat{\alpha}_{ols} - \alpha_o)}{\left[s^2[1,0]A(X'X)^{-1}A\begin{pmatrix}1\\0\end{pmatrix}\right]^{\frac{1}{2}}}$

Using plim $s^2 = \sigma^2$ and plim $A(X'X)^{-1}A = Q^{-1}$, we get

$$\text{plim}\left[s^2[1,0]A(X'X)^{-1}A\begin{bmatrix}1\\0\end{bmatrix}\right]^{1/2} = \left[\sigma^2[1,0]Q^{-1}\begin{bmatrix}1\\0\end{bmatrix}\right]^{1/2} = \sigma\sqrt{q^{11}}$$

where q^{11} is the (1,1) element of Q^{-1}. Therefore, t_α has the same asymptotic distribution as

$$\frac{\sqrt{T}\,(\hat{\alpha}_{ols} - \alpha_o)}{\sigma\sqrt{q^{11}}} = t_\alpha^*$$

In problem 14.6, part (e), we showed that $\begin{bmatrix} \sqrt{T}\,(\hat{\alpha}_{ols} - \alpha_o) \\ T\sqrt{T}(\hat{\beta}_{ols} - \beta_o) \end{bmatrix}$ has an asymptotic distribution $N(0, \sigma^2 Q^{-1})$, so that $\sqrt{T}\,(\hat{\alpha}_{ols} - \alpha_o)$ is asymptotically distributed $N(0, \sigma^2 q^{11})$. Thus,

$$t_\alpha^* = \frac{\sqrt{T}\,(\hat{\alpha}_{ols} - \alpha_o)}{\sigma\sqrt{q^{11}}}$$

is asymptotically distributed as N(0,1) under the null hypothesis of $\alpha = \alpha_o$. Therefore, both t_α and t_α^* have the same asymptotic N(0,1) distribution.

c. Similarly, for testing $H_o;\ \beta = \beta_o$, show that

$$t_\beta = (\hat{\beta}_{ols} - \beta_o)/[s^2(0,1)(X'X)^{-1}(0,1)']^{1/2}$$

has the same asymptotic N(0,1) distribution as $t^*_\beta = T\sqrt{T}(\hat{\beta}_{ols} - \beta_o)/\sigma\sqrt{q^{22}}$. Multiplying both numerator and denominator of t_β by $T\sqrt{T}$ we get,

$$t_\beta = T\sqrt{T}(\hat{\beta}_{ols} - \beta_o)/[s^2(0, T\sqrt{T})(X'X)^{-1}(0, T\sqrt{T})']^{1/2}$$

$$= T\sqrt{T}(\hat{\beta}_{ols} - \beta_o)/[s^2(0, 1)A(X'X)^{-1}A(0, 1)']^{1/2}$$

Now $[0, T\sqrt{T}]$ in t_β can be rewritten as [0,1]A because

$$[0, 1]A = [0, 1]\begin{bmatrix} \sqrt{T} & 0 \\ 0 & T\sqrt{T} \end{bmatrix} = [0, T\sqrt{T}].$$

Therefore, $\text{plim}\ t_\beta = T\sqrt{T}\left(\hat{\beta}_{ols} - \beta_o\right)/[\sigma^2(0, 1)Q^{-1}(0, 1)']^{1/2} = T\sqrt{T}\left(\hat{\beta}_{ols} - \beta_o\right)/\sigma\sqrt{q^{22}}$. Using $\text{plim}\ s^2 = \sigma^2$ and $\text{plim}\ A(X'X)^{-1}A = Q^{-1}$, we get that $\text{plim}\ [s^2(0, 1)A(X'X)^{-1}A(0, 1)']^{1/2} = [\sigma^2(0, 1)Q^{-1}(0, 1)']^{1/2} = \sigma\sqrt{q^{22}}$ where q^{22} is the (2,2) element of Q^{-1}. Therefore, t_β has the same asymptotic distribution as

$$T\sqrt{T}\left(\hat{\beta}_{ols} - \beta_o\right)/\sigma\sqrt{q^{22}} = t^*_\beta$$

From problem 14.6, part (e), $T\sqrt{T}\left(\hat{\beta}_{ols} - \beta_o\right)$ has an asymptotic distribution $N(0, \sigma^2 q^{22})$. Therefore, both t_β and t^*_β have the same asymptotic N(0,1) distribution. Also, the usual OLS t-tests for $\alpha = \alpha_o$ and $\beta = \beta_o$ will give asymptotically valid inference.

14.8 *A Random Walk Model*. This is based on Fuller (1976) and Hamilton (1994). $y_t = y_{t-1} + u_t,\ t = 0, 1, .., T$ where $u_t \sim IIN(0, \sigma^2)$ and $y_o = 0$.

a. Show that $y_t = u_1 + \cdot\cdot + u_t$ with $E(y_t) = 0$ and $\text{var}(y_t) = t\sigma^2$. By successive substitution, we get

$$y_t = y_{t-1} + u_t = y_{t-2} + u_{t-1} + u_t = \cdot\cdot = y_o + u_1 + u_2 + \cdot\cdot + u_t$$

substituting $y_o = 0$ we get $y_t = u_1 + \cdot\cdot + u_t$.

Hence, $E(y_t) = E(u_1) + \cdot\cdot + E(u_t) = 0$

$$\text{var}(y_t) = \text{var}(u_1) + \cdot\cdot + \text{var}(u_t) = t\sigma^2$$

and $y_t \sim N(0, t\sigma^2)$.

b. Squaring the random walk equation, we get

$$y_t^2 = (y_{t-1} + u_t)^2 = y_{t-1}^2 + 2y_{t-1}u_t + u_t^2.$$

Solving for $y_{t-1}u_t$ yields $y_{t-1}u_t = \frac{1}{2}\left(y_t^2 - y_{t-1}^2 - u_t^2\right)$

Summing over $t = 1, 2 \ldots, T$, we get

$$\sum_{t=1}^{T}(y_{t-1}u_t) = \frac{1}{2}\sum_{t=1}^{T}\left(y_t^2 - y_{t-1}^2\right) - \frac{1}{2}\sum_{t=1}^{T}u_t^2.$$

But

$$\sum_{t=1}^{2} y_t^2 = y_1^2 + \cdots + y_T^2$$

$$\sum_{t=1}^{T} y_{t-1}^2 = y_o^2 + \cdots + y_{T-1}^2$$

Hence, by subtraction $\sum_{t=1}^{T}\left(y_t^2 - y_{t-1}^2\right) = y_T^2 - y_o^2$ Substituting this result above, we get $\sum_{t=1}^{T} y_{t-1}u_t = \frac{1}{2}\left(y_T^2 - y_o^2\right) - \frac{1}{2}\sum_{t=1}^{T} u_t^2 = \frac{1}{2}y_T^2 - \frac{1}{2}\sum_{t=1}^{T} u_t^2$

Dividing by $T\sigma^2$ we get $\frac{1}{T\sigma^2}\sum_{t=1}^{T} y_{t-1}u_t = \frac{1}{2}\left(\frac{y_T}{\sqrt{T}\sigma}\right)^2 - \frac{1}{2T\sigma^2}\sum_{t=1}^{T} u_t^2$

But, from part (a), $y_T \sim N(0, T\sigma^2)$ and $y_T/\sqrt{T}\sigma \sim N(0,1)$. Therefore, $(y_T/\sqrt{T}\sigma)^2 \sim \chi_1^2$. Also, by the law of large numbers, $\text{plim}\frac{\sum_{t=1}^{T} u_t^2}{T} = \text{var}(u_t) = \sigma^2$. Hence, $\text{plim}\frac{1}{2\sigma^2}\cdot\frac{\sum_{t=1}^{T} u_t^2}{T} = \frac{1}{2}$. Therefore, $\frac{1}{T\sigma^2}\sum_{t=1}^{T} y_{t-1}u_t = \frac{1}{2}\left(\frac{y_T}{\sqrt{T}\sigma}\right)^2 - \frac{1}{2\sigma^2}\cdot\frac{1}{T}\sum_{t=1}^{T} u_t^2$ is asymptotically distributed as $\frac{1}{2}\left(\chi_1^2 - 1\right)$.

c. Show that $E\left(\sum_{t=1}^{T} y_{t-1}^2\right) = \frac{T(T-1)}{2}\sigma^2$. Using the results in part (a), we get

$$y_{t-1} = y_o + u_1 + u_2 + \cdots + u_{t-1}$$

Substituting $y_o = 0$, squaring both sides and taking expected values, we get $E\left(y_{t-1}^2\right) = E\left(u_1^2\right) + \cdots + E\left(u_{t-1}^2\right) = (t-1)\sigma^2$ since the u_t's are independent.

Therefore,

$$E\left(\sum_{t=1}^{T} y_{t-1}^2\right) = \sum_{t=1}^{T} E\left(y_{t-1}^2\right) = \sum_{t=1}^{T}(t-1)\sigma^2 = \frac{T(T-1)}{2}\sigma^2$$

where we used the fact that $\sum_{t=1}^{T} t = T(T+1)/2$ from problem 14.6.

d. For the AR(1) model, $y_t = \rho y_{t-1} + u_t$, show that OLS estimate of ρ satisfies

$$\text{plim}\, T(\hat{\rho} - \rho) = \text{plim}\frac{\sum_{t=1}^{T} y_{t-1}u_t/T\sigma^2}{\sum_{t=1}^{T} y_{t-1}^2/T^2\sigma^2} = 0 \text{ where}$$

$$\hat{\rho} = \frac{\sum_{t=1}^{T} y_{t-1}y_t}{\sum_{t=1}^{T} y_{t-1}^2} = \rho + \frac{\sum_{t=1}^{T} y_{t-1}u_t}{\sum_{t=1}^{T} y_{t-1}^2}.$$

From part (b), $\frac{1}{T\sigma^2}\sum_{t=1}^{T} y_{t-1}u_t$ has an asymptotic distribution $\frac{1}{2}\left(\chi_1^2 - 1\right)$. This implies that $\sum_{t=1}^{T} y_{t-1}u_t/\sigma^2$ converges to an asymptotic distribution of $\frac{1}{2}\left(\chi_1^2 - 1\right)$ at the rate of T. Also, from part (c), $E\left(\sum_{t=1}^{T} y_{t-1}^2\right) = \frac{\sigma^2 T(T-1)}{2}$ implies that $\text{plim}\frac{1}{T^2}\sum_{t=1}^{T} y_{t-1}^2$ is $\sigma^2/2$. This means that $\sum_{t=1}^{T} y_{t-1}^2/\sigma^2$ converges to $\frac{1}{2}$ at the rate of T^2. One can see that the asymptotic distribution of $\hat{\rho}$ when $\rho = 1$ is a ratio of a $\frac{1}{2}(\chi_1^2 - 1)$ random variable to a non-standard distribution in the denominator which is beyond the scope of this book, see Hamilton (1994) or Fuller (1976) for further details. The object of this exercise is to show that if $\rho = 1$, $\sqrt{T}(\hat{\rho} - \rho)$ is no longer Normal as in the standard stationary least squares regression with $|\rho| < 1$. Also, to show that for the non-stationary (random walk) model, $\hat{\rho}$ converges at a faster rate (T) than for the stationary case ($\sqrt{T}$). From part (c) it is clear that one has to divide the denominator of $\hat{\rho}$ by T^2 rather than T to get a convergent distribution.

14.9 Cointegration Example

a. Solving for the *reduced form* from (14.13) and (14.14) we get

$$Y_t = \frac{u_t - v_t}{(\alpha - \beta)} = \frac{1}{(\alpha - \beta)}u_t - \frac{1}{(\alpha - \beta)}\nu_t$$

and

$$C_t = \frac{\beta(u_t - v_t)}{(\alpha - \beta)} + u_t = \frac{\alpha}{(\alpha - \beta)}u_t - \frac{\beta}{(\alpha - \beta)}\nu_t$$

In this case, u_t is I(0) and ν_t is I(1). Therefore both Y_t and C_t are I(1). Note that there are no excluded exogenous variables in (14.13) and (14.14) and only one right hand side endogenous variable in each equation. Hence both equations are unidentified by the order condition of identification. However, a linear combination of the two structural equations will have a mongrel disturbance term that is neither AR(1) nor random walk. Hence, both equations are identified. If $\rho = 1$, then both u_t and ν_t are random walks and the mongrel disturbance is also a random walk. Therefore, the system is unidentified. In such a case, there is no cointegrating relationship between C_t and Y_t. Let $(C_t - \gamma Y_t)$ be another cointegrating relationship, then subtracting it from the first cointegrating relationship, one gets $(\gamma - \beta)Y_t$ which should be I(0). Since Y_t is I(1), this can only happen if $\gamma = \beta$. Differencing both equations in (14.13) and (14.14) we get

$$\begin{aligned}\Delta C_t - \beta\Delta Y_t = \Delta u_t = (\rho - 1)u_{t-1} + \varepsilon_t &= \varepsilon_t + (\rho - 1)(C_{t-1} - \beta Y_{t-1}) \\ &= \varepsilon_t + (\rho - 1)C_{t-1} - \beta(\rho - 1)Y_{t-1}\end{aligned}$$

and $\Delta C_t - \alpha\Delta Y_t = \Delta\nu_t = \eta_t$. Writing them as a VAR, we get (14.17)

$$\begin{bmatrix} 1 & -\beta \\ 1 & -\alpha \end{bmatrix}\begin{bmatrix} \Delta C_t \\ \Delta Y_t \end{bmatrix} = \begin{bmatrix} \varepsilon_t + (\rho - 1)C_{t-1} - \beta(\rho - 1)Y_{t-1} \\ \eta_t \end{bmatrix}$$

Post-multiplying by the inverse of the first matrix, we get

$$\begin{bmatrix} \Delta C_t \\ \Delta Y_t \end{bmatrix} = \left(\frac{1}{\beta - \alpha}\right)\begin{bmatrix} -\alpha & \beta \\ -1 & 1 \end{bmatrix}\begin{bmatrix} \varepsilon_t + (\rho - 1)C_{t-1} - \beta(\rho - 1)Y_{t-1} \\ \eta_t \end{bmatrix}$$

$$= \frac{1}{\beta - \alpha}\begin{bmatrix} -\alpha\varepsilon_t - \alpha(\rho-1)C_{t-1} + \alpha\beta(\rho-1)Y_{t-1} + \beta\eta_t \\ -\varepsilon_t - (\rho-1)C_{t-1} + \beta(\rho-1)Y_{t-1} + \eta_t \end{bmatrix}$$

$$= \frac{1}{(\beta-\alpha)}\begin{bmatrix} -\alpha(\rho-1) & \alpha\beta(\rho-1) \\ -(\rho-1) & \beta(\rho-1) \end{bmatrix}\begin{pmatrix} C_{t-1} \\ Y_{t-1} \end{pmatrix} + \begin{pmatrix} h_t \\ g_t \end{pmatrix}$$

where h_t and g_t are linear combinations of ε_t and η_t. This is Eq. (14.18). This can be rewritten as

$$\Delta C_t = \frac{-\alpha(\rho-1)}{\beta-\alpha}(C_{t-1} - \beta Y_{t-1}) + h_t = -\alpha\delta Z_{t-1} + h_t$$

$$\Delta Y_t = \frac{-(\rho-1)}{\beta-\alpha}(C_{t-1} - \beta Y_{t-1}) + g_t = -\delta Z_{t-1} + g_t$$

where

$$\delta = (\rho-1)/(\beta-\alpha) \quad \text{and} \quad Z_t = C_t - \beta Y_t.$$

These are Eqs. (14.19) and (14.20). This is the *Error-Correction Model* (ECM) representation of the original model. Z_{t-1} is the error correction term. It represents a disequilibrium term showing the departure from long-run equilibrium. Note that if $\rho = 1$, then $\delta = 0$ and Z_{t-1} drops from both ECM equations.

b. $$\hat{\beta}_{ols} = \frac{\sum_{t=1}^{T} C_t Y_t}{\sum_{t=1}^{T} Y_t^2} = \beta + \frac{\sum_{t=1}^{T} Y_t u_t}{\sum_{t=1}^{T} Y_t^2}$$

Since u_t is I(0) if $\rho \neq 1$ and Y_t is I(1), we have plim $\sum_{t=1}^{T} {Y_t}^2/T^2$ is O(1), while plim $\sum_{t=1}^{T} Y_t u_t/T$ is O(1). Hence $T(\hat{\beta}_{ols} - \beta)$ is O(1) or $(\hat{\beta}_{ols} - \beta)$ is O(T).

References

Fuller, W.A. (1976), Introduction to Statistical Time Series (John Wiley and Sons: New York).

Hamilton, J.D. (1994), Time Series Analysis (Princeton University Press: Princeton, New Jersey).

Sims, C.A., J.H. Stock and M.W. Watson (1990), "Inference in Linear Time Series Models with Some Unit Roots," Econometrica, 58: 113–144.