Suggested Answers to Odd-Numbered Questions

A1 **(a)** This Monte Carlo study examines the sampling distribution of the average of 25 numbers drawn from a distribution with mean 2 and variance 9. The average of N numbers drawn randomly from a distribution with mean μ and variance σ^2 has a sampling distribution with mean μ and variance σ^2/N.

(b) *A*bar estimates the mean of this sampling distribution, so should be close to 2.

(c) *A*var estimates its variance, so should be close to 9/25 = 0.36.

A3 **(a)** This Monte Carlo study examines the sampling distribution of the fraction of successes h in a sample of size 50 where the probability of success is 20%. When the true probability of success is p, for a sample of size N then h has a sampling distribution with mean p and variance $p(1–p)/N$.

(b) *h*av should be an estimate of 0.2.

(c) *h*var should be an estimate of $p(1–p)/N$ which here is (0.2)(0.8)/50 = 0.0032

(d) *w*av estimates the mean of the sampling distribution of the estimated variance of h, and so should be approximately 0.0032.

A5 Create 44 observations from a normal distribution with mean 6 and variance 4. Calculate their average A and their median B. Repeat to obtain, say, 1000 *A*s and 1000 *B*s. Find the mean of the 1000 A values and the mean of the 1000 B values and see which is closer to 6. Calculate the variance of the 1000 A values and the variance of the 1000 B values and see which is smaller.

A7 **(i)** Choose values for a and b, say 2 and 6 (be sure not to have zero fall between a and b because then an infinite value of $1/x^2$ would be possible and its distribution would not have a mean). (ii) Get the computer to generate 25 drawings (x values) from $U(2,6)$. (If the computer can only draw from $U(0,1)$, then multiply each of these values by 4 and add 2.) (iii) Use the data to calculate A's estimate A^* and B's estimate B^*. Save them. (iv) Repeat from (ii) 499 times, say, to get 500 A^*s and 500 B^*s. (v) Obtain the mean m of the distribution of $1/x^2$ either algebraically (the integral from a to b of $1/(b–a)x^2$) or by averaging a very large number (10,000, say) of $1/x^2$ values. (vi) Estimate the bias of A^* as the difference between the average of the 500 A^*s and m. Estimate the variance of A^* as the variance of the 500 A^*s. Estimate the MSE of A^* as the average of the 500 values of $(A^* – m)^2$. Compute the estimates for B^* in similar fashion and compare.

A9 The bias of β^* is estimated as the average of the 400 β^*s minus β, the variance of which is the variance of β^* divided by 400. Our estimate of this is 0.01/400. The relevant t statistic is thus 0.04/(0.1/20) = 8 which exceeds the 5% critical value, so the null is rejected.

A11 **(a)** The expected value of $\bar{x}$ is μ, but the expected value of a nonlinear function of $\bar{x}$ is not the nonlinear function of μ, except in large samples.
(b) Use an estimate of the formula for the variance of a nonlinear function.

$$V^*(\theta^*) = \left[3\bar{x}^2\right]^2 V(\bar{x}) = 9\bar{x}^4\left(\frac{6}{27}\right) = 2\bar{x}^4$$

(c) (i) Choose a value for μ, say 2, so that $\theta = 8$. (ii) Generate 27 values of x drawn from $N(2,6)$. (iii) Use the data to calculate θ^* and $V^*(\theta^*)$. Save them. (iv) Repeat from (ii) to get 5000 θ^*s and 5000 $V^*(\theta^*)$s. (v) Estimate the true variance of θ^* by the variance of the 5000 θ^*s and see how close the average of the 5000 $V^*(\theta^*)$s is to this value.

A13 **(a)** This process is estimating the variance of a random variable w by three different formulas differing only by their divisors. Since dividing by N–1 produces an unbiased estimate of this variance, the average of the 4000 a values should be closest to 4, the true variance of w.
(b) Dividing by a larger number shrinks the variance estimate towards zero, making its variance smaller, so the variance of the c values should be smallest.
(c) Step (viii) estimates the MSEs of the three estimators. Since dividing by N+1 produces the estimate with the smallest MSE, C should be the smallest.

A15 Set variables "stick" and "switch" equal to zero. *Set variable "switchguess" equal to four. Draw a number x from a distribution uniform between 0 and 3. Set variable "door" equal to 1 if $0 \leq x < 1$, equal to 2 if $1 \leq x < 2$, and equal to 3 if $2 \leq x \leq 3$. Draw a number y from a distribution uniform between zero and three. Set "doorguess" equal to one if $0 \leq y < 1$, equal to two if $1 \leq y < 2$, and equal to three if $2 \leq y \leq 3$. If door = 1 and doorguess = 2 or 3, set switchguess = 1; if door = 2 and doorguess = 1 or 3, set switchguess = 2; if door = 3 and doorguess = 1 or 2, set switchguess = 3. If doorguess = door, add one to "stick" and if switchguess = door, add one to "switch." Repeat from * until, say, a thousand x values have been drawn. For any original guess, sticking should work one-third of the time, but as is evident from the Monte Carlo structure, switching should work two-thirds of the time.

B1 $k = 3/4$, found by setting the integral of $kx(2–x)$ from 0 to 2 equal to one. $E(x)$ = integral from 0 to 2 of $3x^2(2–x)/4 = 1$. $V(x) = Ex^2 – (Ex)^2 = 1/5$.

B3 Try supply of 100. All units are sold and profit on each unit is \$5, so expected profit is \$500. Try supply of 200. All units are sold with a probability of 0.95, with profit of \$1000. With probability 0.05, only 100 units are sold, with a profit \$500 on the units sold but a loss of \$1000 on the units not sold for a net loss of \$500. Expected profit is thus 0.95(1000) + 0.05(–500) = 925. Try supply of 300. All units are sold with a probability of 0.85, with a profit of \$1500; with probability 0.1 only 200 units are sold, with net profit of zero; and with probability 0.05 only 100 units are sold, with a net loss of \$1,500. Expected profit is thus 0.85(1500) + 0.1(0) + 0.05(−1500) = 1200. For supply of 400 expected profit is 1100. That the supply of 300 is the supply that maximizes expected profit can be verified by calculating expected profit for supplies of 299 and 301. The variance of profit if supply is 300 is calculated as $0.85(1500–1200)^2 + 0.1(–1200)^2 + 0.05(–1500–1200)^2 =$ 585 000.

B5 The probability of finding a minimum price of \$2 is 1/8, namely the probability of checking three stores all of which have a price of \$2. Expected price is thus $2 \times (1/8) + 1 \times (7/8) = 9/8$.

B7 **(a)** True. If α^* is unbiased in small samples it is also unbiased in large samples.
(b) False. Asy Var(α^*) = $(1/N)$ lim$N(4\alpha/N + 16\alpha^2/N^2) = 4\alpha/N$
(c) True, because α^* is asymptotically unbiased and the limit as N goes to infinity of its variance is zero.
(d) Uncertain. Do not know about other estimators, or if α^* is MLE.

B9 **(a)** Both have mean 5%. 1 has variance $30^2 \times 6 = 5400$; 2 has variance $3 \times 10^2 \times 6 = 1800$.
(b) Variance of 2 becomes $10^2 \times 6 + 10^2 (6 – 2 \times 3 + 6) = 1200$.
(c) False. In example of part (a) above, positive correlation will raise the variance of 2 above

1800, but it will not make it as large as 5400. Perfect correlation would make the variance 5400.

(d) False. Although it will have zero variance, it will have an expected return of 5%.

B11 **(a)** The expected value of x is $(\lambda-10)/2$; setting the mean of the data $\bar{x}$ equal to the mean of the distribution we get $\lambda^{mm} = 2x + 10$.

(b) $V(\lambda^{mm}) = 4V(x)/N = (\lambda-10)^2/3N$.

B13 The formula for the variance of a nonlinear function of a random vector produces $V(\alpha^*\beta^*) = \beta^{*2}V(\alpha^*) + \alpha^{*2}V(\beta^*) + 2\alpha^*\beta^*C(\alpha^*, \beta^*)$ which when evaluated at estimated values yields 0.49, so the estimated standard error is 0.7.

B15 $V(y) = [2V(x) + 2C(x_t, x_{t-1})]/4$. $V(x) = \sigma^2/0.36$, calculated by squaring both sides of the expression for x_t, taking expected values and solving for $V(x) = E(x_t^2) = E(x_{t-1}^2)$.

$C(x_t, x_{t-1}) = 0.8V(x)$, calculated by multiplying through the expression for x_t by x_{t-1} and taking expected values. Using these values, $V(y) = 2.5\ \sigma^2$.

C1 **(a)** Unbiased; it is as likely to be too steep as to be too flat, depending on the chosen observation. $\beta_i = y_i/x_i = \beta + \varepsilon_i/x_i$, so $E\beta_i = \beta$, where i denotes the chosen observation.

(b) Variance is $\sigma^2/(x_i)^2$, the variance of ε_i/x_i.

(c) Choose the observation with the largest x value, to minimize variance.

(d) No. As the sample size grows, the variance does not shrink to zero.

C3 **(a)** All three are unbiased. The variance of b^{***} is $(8 + 4)/2^2 = 3.0$, the smallest of the three, so it is preferred.

(b) Any weighted average of the two unbiased estimates is unbiased, so one should choose that weighted average that minimizes variance. Choose the estimator $a\beta^* + (1-a)\ b^{**}$ whose variance, $8a^2 + 4(1-a)^2$, is minimized for $a = 1/3$.

(c) Pool the two data sets and run a single regression.

C5 **(a)** b^{***} is unbiased, so choose a to minimize its variance: $a = Vb^{**}/(Vb^*+Vb^{**})$.

(b) When b^* is a very precise estimator (Vb^* small), more weight is given to b^* in the weighted average, and similarly when b^{**} is very precise more weight is given to b^{**}. For the special case in which an estimator is perfect (zero variance), it is given 100% of the weighting.

C7 **(a)** The average is unbiased with variance $(V_A + V_B)/4$. The first estimate's variance, V_A is smaller than this if V_B is more than three times the size of V_A.

(b) No because in this case the unweighted average is not the optimal way of employing both pieces of information.

D1 When MSE of zero (μ^2) exceeds MSE of the sample mean (σ^2/T).

D3 $\text{MSE}(\beta^{**}) = a^2V\beta^* + (a\beta-\beta)^2$, minimized for $a = \beta^2/(V\beta^*+\beta^2)$. It is not used more often because it involves β which is unknown.

E1 **(a)** The distribution of prices is a horizontal line of height one between one and two. The probability that p is the minimum price is the probability of obtaining price p in the first store and any price greater than p in the second store, plus the probability of obtaining price p in the second store and any price greater than p in the first store. Both these probabilities are $(2-p)$, so the distribution of the minimum price is $2(2-p)$ for $1 \le p \le 2$. Expected price is thus the integral from 1 to 2 of $2p(2-p)$ which is 4/3.

(b) Have the computer produce two values from a $U(1,2)$ distribution. Save the smaller value. Repeat this procedure to obtain a thousand such numbers, say. Average these thousand numbers.

F1 **(a)** Q is the average R^2 from a regression restricting the w slope coefficient to be zero. By allowing the computer to use w it will be more successful in minimizing the sum of squared errors and thus R^2 will be higher, so S will be bigger than Q.

(b) Adjusted R^2 corrects for the addition of an irrelevant explanatory variable and so AQ and AS should be roughly equal.

F3 **(a)** (i) Select values for α, β, δ, and σ^2 the variance of the error such that β is positive and δ is negative because of the context. (ii) Have the computer draw 25 errors from $N(0, \sigma^2)$. (iii) Use these errors, the parameter values and the data on y and r to calculate 25 m observations. (iv) Regress m on a constant, y and r to get β^*. Save it. (v) Regress m on a constant and y

to get β^{**}. Save it. (vi) Repeat from (ii) 599 times, say, to get 600 β^*s and 600 β^{**}s. (vii) Use the 600 β^*s to estimate the bias, variance, and MSE of β^*. Do the same for β^{**} and compare.

(b) Results expected are those associated with an incorrectly omitted (non-orthogonal) explanatory variable. β^* should be unbiased, whereas β^{**} should be biased (in this case upwards) but with a smaller variance than β^*.

(c) Positive correlation between y and r would cause the bias to be negative.

(d) Bias would be zero.

F5 **(a)** Reduction in bias, variance, or MSE.

(b) (i) Choose values for α_0, α_1, α_2, α_3, and σ^2, the variance of ε, making sure that $\alpha_1 + \alpha_2 = 1$. (ii) Choose sample size, say 35 and obtain 35 values of x, q, and w, ensuring that they are not orthogonal. (iii) Get computer to generate 35 errors from $N(0, \sigma^2)$. (iv) Use these errors, in conjunction with the parameter values and the observations on the independent variables, to calculate 35 y values. (v) Regress y on a constant, x, q, and w. Save α_3^*. (vi) Regress $(y-q)$ on a constant, $(x-q)$ and w to get the restricted least squares estimate α_3^{**}. Save it. (vii) Repeat from (iii) 499 times to obtain 500 α_3^* values and 500 α_3^{**} values. (viii) Use the 500 values of α_3^* to estimate its bias, variance, and MSE. Use the 500 values of α_3^{**} to estimate its bias, variance, and MSE. Compare.

(c) Both should be unbiased. The variance and thus the MSE of α_3^{**} should be smaller.

(d) α_3^{**} would be biased, its variance would still be smaller, but it is not possible to say which will have the smaller MSE.

G1 **(a)** Perfect multicollinearity between the constant term and x. Estimation is not possible.

(b) False. The CNLR model requires that the errors be distributed normally, not that the values of the independent variable come from a normal distribution.

(c) True. This would imply that the variance of the error was not the same for all observations, violating one of the assumptions of the CLR model.

(d) Uncertain. It depends on the criterion we adopt. A biased estimator, for example, may have a lower MSE.

G3 No. Its inclusion creates perfect multicollinearity with the constant term, making estimation impossible.

G5 Write the relationship as $E = \alpha + \beta Y + \delta F + \varepsilon$ with $F = \eta + \theta Y + u$ where ε and u are error terms. Substitution yields $E = \alpha + \delta\eta + (\beta{+}\delta\theta)Y + \varepsilon + \delta u$. Regressing E on a constant and Y will produce a slope coefficient estimate which estimates $(\beta{+}\delta\theta)$ rather than β. If $\delta\theta$ were negative this OLS estimate could be negative. This could occur, for example, if δ were positive (expenditure is higher for a bigger family size, which seems reasonable) and θ were negative (family size is smaller when family income is larger, which is possible).

G7 Since the sample x values are all positive, the off-diagonal elements of the $X'X$ matrix are positive and so the off-diagonal elements of the $(X'X)^{-1}$ matrix are negative, implying a negative covariance between α^{OLS} and β^{OLS}. Thus if β^{OLS} is an overestimate, it is more likely that α^{OLS} is an underestimate. Draw an upward-sloping line to represent the true relationship in the NE quadrant. Drawing in an estimating line with a higher slope, crossing the original line at the average of the y values, produces a lower intercept.

G9 **(a)** Consider the data measured as deviations from their means. Then $\delta^{OLS} = \Sigma sy/\Sigma y^2 = \Sigma(y-c)y/\Sigma y^2 = 1-\Sigma cy/\Sigma y^2 = 1 - \beta^{OLS}$.

(b) True. The residuals are the same for both relationships. $s = \delta^{OLS}y + \varepsilon^{OLS}$ implies $c = \beta^{OLS}y + \varepsilon^{OLS}$.

(c) False. SSE is the same, but the variation in S is less than the variation in C (assuming $\beta >$ 0.5), so SST differs.

G11 **(a)** All are unbiased. Variance of β^* is $\sigma^2/\Sigma(x-\bar{x})^2$, of β^{**} is $\sigma^2/\Sigma x^2$, of β^{***} is $N\sigma^2/(\Sigma x)^2$, and of β^{****} is $(\sigma^2/N^2)\,\Sigma(1/x^2)$.

(b) All are linear and unbiased. Thus choose β^{**} because it is the BLUE.

G13 Both are easily seen to be unbiased. $V(\bar{y}) = \sigma^2/N$. and $V(\beta^{OLS}\bar{x}) = (\bar{x}^2)V(\beta^{OLS}) = [(\Sigma x)^2/(N\Sigma x^2)](\sigma^2/N)$

which is less than σ^2/N unless x is a constant, in which case it is equal to σ^2/N. Thus prefer $\beta^{OLS}\bar{x}$.

G15 False. The 0.73 is one drawing from the sampling distribution which is centered at the true value of the slope parameter with a standard error estimated by 0.2.

G17 Work backwards to discover that the original functional form is $y = Ae^{\delta t}K^{\alpha}L^{\beta}$ so that the intercept is an estimate of δ, the growth rate from technical change.

G19 (**a**) R^2 and all parameter estimates are zero.
(**b**) Intercept estimate zero, *y*hat coefficient estimate one, and $R^2 = 0.8$.
(**c**) Intercept estimate is the average *y* value, *e*hat coefficient estimate one, and $R^2 = 0.2$.

H1 (**a**) OLS is unbiased so mean is 3. Variance is $\sigma^2/\Sigma(x-\bar{x})^2$. Calculate numerator as $E\varepsilon^2 = 1/2$, and denominator as $\sum x^2 - N\bar{x}^2 = 20$, so variance is 1/40.
(**b**) Estimator is unbiased because restriction is true, so mean is 3.
Variance is $\sigma^2/\Sigma x^2 = 1/60$.

H3 (**a**) Substitute $R = P - D$ in the estimated equation to get
$D_t = (1 - 0.891)P_t + 0.654\ D_{t-1} + (0.622 - 0.654)P_{t-1}$ with the same residual. The *d* statistic is calculated from the residuals, so it should be the same. R^2 is different because there is a different dependent variable and thus a different SST.
(**b**) The first standard error is the standard error of 1 minus the 0.891 estimate which is just the standard error of the 0.891 estimate, namely 0.027. Similarly, the second standard error is 0.092. The third standard error is the standard error of the 0.622 estimate minus the 0.654 estimate, estimation of which requires knowledge of their estimated covariance.

H5 (**a**) Guess is $2\alpha^{OLS} + 9\beta^{OLS} + E(\text{dice}) = -6 + 18 + 35 = 47$.
(**b**) Expected payoff is $60 - V(W) = 60 - V(2\alpha^{OLS} + 9\beta^{OLS}) - V(\text{dice}) = 60 - 4V(\alpha^{OLS}) - 36C(\alpha^{OLS}, \beta^{OLS}) - 81V(\beta^{OLS}) - 29.2 = 30.8 - 0.29\sigma^2$ where σ^2 is $V(\varepsilon)$.
(**c**) SSE = 10,300 – 500 = 9,800, so $s^2 = 100$, so estimated expected payoff is $1.80.

I1 (**a**) True. On the Ballentine, use one circle for *y*, one for *x* and the third for the time trend. Call the area used by the computer to estimate the coefficient on *x* the blue area. The overlap of the residuals from regressing *y* on time and from regressing *x* on time is the blue area.
(**b**) False. Both numerator and denominator of R^2 have the same area removed, changing its value.

I3 Suppose $\ln y = \alpha + \beta x$ was estimated in 1981. If the relationship is unchanged, then $\ln(y/1.2) = \alpha + \beta x$ should hold for the 1985 data. This implies $\ln y = \alpha + \ln 1.2 + \beta x$ suggesting that if the data are not scaled to 1981 dollars the slope coefficient should be comparable to the 1981 slope estimate, but the intercept estimate will not be comparable to the 1981 intercept estimate.

I5 $w^* = 100w - 1$, so $w = 1 + w^*/100$. The new relationship is $y = 300 + 6w = 300 + 6(1 + w^*/100) = (300 + 6) + (6/100)w^*$. This implies that the new slope coefficient estimate is 6/100 with estimated standard error 0.011, and the new intercept coefficient estimate is 306 with estimated variance $25 + 1.1 + 2 \times 0.5 = 27.1$.

I7 New relationship is $y = \alpha + (\beta/3)(3x) + \varepsilon$ so intercept estimate remains unbiased but slope estimate is now an unbiased estimate of $\beta/3$.

I9 New relationship is $2.2(W/2.2) = -180 + 5.0 \times (2.5H)/2.5$ which becomes $(W/2.2) = -180/2.2 + 5.0 \times (2.5H)/(2.5 \times 2.2)$ so that the new intercept is 81.8 with variance $4 \times (2.2)^{-2} = 0.83$, and the new slope is 0.91 with variance $1 \times (2.5 \times 2.2)^{-2} = 0.033$.

I11 The research assistant must have measured returns differently. In particular, 6% must have been measured as 0.06 rather than 6.0.

I13 New relationship is $\ln w = 1 + 2\text{ed} - 2 \times 12 + 2 \times 12 + 0.5\text{male} + 0.2\text{ed} \times \text{male} - 0.2 \times 12 \times \text{male} + 0.2 \times 12 \times \text{male} + 3 \times \exp - 0.1\exp^2$ which becomes
$\ln w = 25 + 2(\text{ed} - 12) + 2.9\text{male} + 0.2(\text{ed} - 12) \times \text{male} + 3 \times \exp - 0.1\exp^2$.

J1 Uncertain. With the help of the extra regressor minimizing SSE cannot be less successful, so

SSE should remain the same or fall, suggesting that the estimate of σ^2 should remain the same or fall. If σ^2 is estimated with a correction for degrees of freedom, however, it is possible that SSE could change by such a small amount that the reduction in the denominator of the estimator could cause the estimate to increase.

J3 In general, the variance of b^* is larger than that of b^{**}, but in general the estimate of the variance of b^{**} is biased upward, so no firm conclusion can be drawn concerning these relative magnitudes.

J5 Expected pay = $10 - V(\alpha^{OLS}) - 2C(\alpha^{OLS}, \beta^{OLS}) - V(\beta^{OLS}) = 10 - 15 + 12 - 3 = 4$.

J7 Estimate is θ^* the sum of the estimated slope coefficients β^* of experience and γ^* of years with current employer. To calculate a confidence interval we need to estimate the variance of this estimate. This could be done by using the formula for the variance of the sum of two random variables. An easier way would be to substitute $\beta = \theta - \gamma$ into the estimating equation to create a new regression with θ as one of the parameters. OLS estimation will produce θ^* and its estimated variance.

J9 A little algebra shows that $\theta = \bar{x}_B(\hat{\beta} - \hat{\beta}_B) = \bar{X}_B\Omega(\hat{\beta}_W - \hat{\beta}_B)$ which takes the form Ax where is A is a matrix of constants and x is a random variable with variance $V(x)$ the sum of the variances of $\hat{\beta}_W$ and $\hat{\beta}_B$. The required variance is then obtained using the formula $AV(x)A'$.

J11 We want to estimate $\lambda = 400\beta + \theta$. Substitute $\theta = \lambda - 400\beta$ to get a new estimating equation $P = \alpha + \beta\,(\text{sqft} - 400\text{FR}) + \gamma\text{beds} + \eta\text{baths} + \lambda\text{FR} + \varepsilon$.

K1 False. Imposing any constraint, regardless of its truth, inhibits the minimization of SSE and thus lowers R^2.

K3 To incorporate the restrictions we regress $(y-2w)$ on $(x-w)$ without an intercept so that $\beta_x^* = \Sigma(y-2w)(x-w)/\Sigma(x-w)^2 = 90/14$ and $\beta_w^* = 2-\beta_x^* = -68/14$.

K5 **(a)** Regress y on a constant, $(2x+w)$ and z.
(b) Smaller, since minimization of SSE is inhibited.
(c) Yes, since the restriction is true.
(d) Smaller, since incorporating more information into estimation produces more efficient estimates.
(e) Answers to (b) and (d) are unchanged. Answer to (c) is that θ estimate is in general (i.e., when the regressors are correlated) now biased.

K7 **(a)** Substituting the relationship for β_i we get

$$\begin{aligned} y_t &= \alpha + \delta_0 x_t + (\delta_0 + \delta_1 + \delta_2)x_{t-1} + \\ &\quad (\delta_0 + 2\delta_1 + 4\delta_2)x_{t-2} + (\delta_0 + 3\delta_1 + 9\delta_2)x_{t-3} \\ &= \alpha + \delta_0(x_t + x_{t-1} + x_{t-2} + x_{t-3}) + \\ &\quad \delta_1(x_{t-1} + 2x_{t-2} + 3x_{t-3}) + \delta_2(x_{t-1} + 4x_{t-2} + 9x_{t-3}) \end{aligned}$$

(b) $\beta_0^* = 4$, $\beta_1^* = 5$, $\beta_2^* = 4$ and $\beta_3^* = 1$.
(c) $\beta^* = A\delta^*$ where A is a 4×3 matrix with first row (1,0,0), second row (1,1,1), third row (1,2,4), and fourth row (1,3,9). $V(\beta^*) = AVA'$.

L1 X is a column of ones, of length N. $X'X = N$. $(X'X)^{-1} = 1/N$. $X'y = \Sigma y$.
β^{OLS} = ybar. $(X'X)^{-1}X'\varepsilon = \bar{\varepsilon}$. $V(\beta^{OLS}) = \sigma^2/N$. That the OLS estimate of μ is the sample average, and that its variance is σ^2/N are well known and could have been guessed.

L3 **(a)** The restricted OLS estimator is given by
$\beta^* = \beta^{OLS} + (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(r-R\beta^{OLS})$
$E\beta^* = \beta + (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(r-R\beta)$ so bias is $(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(r-R\beta)$.

(b)
$$\begin{aligned} (\beta^* - E\beta^*) &= \beta + (X'X)^{-1}X'\varepsilon + (X'X)^{-1} \\ &\quad R'\left[R(X'X)^{-1}R'\right]^{-1} \\ &\quad \left(r - R\beta - R(X'X)^{-1}X'\varepsilon\right) - E\beta^* \\ &= \left\{I - (X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}R\right\} \\ &\quad (X'X)^{-1}X'\varepsilon \end{aligned}$$

$$\begin{aligned} V(\beta^*) &= E(\beta^* - E\beta^*)(\beta^* - E\beta^*)' \\ &= E\left\{I - (X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}R\right\} \\ &\quad (X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1} \\ &\quad \left\{I - R'\left[R(X'X)^{-1}R'\right]^{-1}R(X'X)^{-1}\right\} \\ &= \sigma^2\left\{I - (X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}R\right\}(X'X)^{-1} \end{aligned}$$

$$\left\{I - R'\left[R(X'X)^{-1}R'\right]^{-1}R(X'X)^{-1}\right\}$$

$$= \sigma^2\left[(X'X)^{-1} - (X'X)^{-1}\right.$$

$$\left.R'\left[R(X'X)^{-1}R'\right]^{-1}R(X'X)^{-1}\right]$$

Thus $V(\beta^{OLS}) - V(\beta^*) = \sigma^2(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R\}(X'X)^{-1}$ which is seen to be nnd.

NOTE: A quicker way to get $V(\beta^*)$ is to write

$\beta^* = (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}r + \{I - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R\}\beta^{OLS} = A + B\beta^{OLS}$

so that $V(\beta^*) = BV(\beta^{OLS})B'$.

L5 **(a)** Let an alternative linear unbiased estimator be $[x_0(X'X)^{-1}X' + d]y$. Unbiasedness implies that $dX = 0$. Then find the variance–covariance matrix of this alternative estimator and show that it must be greater than that of $x_0\beta^{OLS}$ for d nonzero.

(b) $x_0\beta^{OLS} - y_0 = x_0(X'X)^{-1}X'\varepsilon - \varepsilon_0$.

(c) $\sigma^2[1 + x_0(X'X)^{-1}x_0']$

(d) Minimize $x_0(X'X)^{-1}x_0' - 2\lambda\ (Rx_0'-1)$ with respect to x_0', where R is a row vector whose first element is unity and all other elements are zero. First partial is $2(X'X)^{-1}x_0' - 2\lambda R'$ whence the required value of x_0' is $[R(X'X)R']^{-1}X'XR'$. The special nature of R makes $[R(X'X)R']^{-1}$ equal $1/N$ and $X'XR'$ equal the first column of $X'X$ which is the sums of all the regressors. Thus $[R(X'X)R']^{-1}X'XR'$ is the average of all the x values in the data.

L7 **(a)** $E\theta^* = Ea'(X\beta + \varepsilon) = a'X\beta$ so $a'X$ must equal c' for unbiasedness.

(b) $V(\theta^*) = Ea'\varepsilon\varepsilon'a = \sigma^2a'a$

(c) Minimize $\sigma^2a'a - 2\lambda'(X'a-c)$ with respect to a. First partial is $2\sigma^2a - 2X\lambda$. Solving for a we get $X(X'X)^{-1}c$ so that $\theta^* = c'\beta^{OLS}$.

(d) Pick c to be a vector with ith element unity and all other elements zero. This result shows that the ith element of β^{OLS} is the BLUE of the ith element of β. Repeat this for all i to show that β^{OLS} is the BLUE of β.

(e) If c is a vector of values of the independent variables associated with the dependent variable for which we wish a forecast, this result shows that $c'\beta^{OLS}$ is the BLUE forecast.

L9. $E\theta SSE = \theta\sigma^2(N-K)$ so bias $= \sigma^2[\theta\,(N-K)-1]$. $V(\theta SSE) = 2\theta^2\sigma^4(N-K)$. Thus $MSE(\theta SSE) = \sigma^4[\theta\,(N-K)-1]^2 + 2\theta^2\sigma^4(N-K)$.

Minimizing with respect to θ gives $1/(N-K+2)$.

M1 **(a)** Write X as $[X_1'\ X_2']'$. Then $X'X = X_1'X_1 + X_2'X_2$ and $X'y = X_1'y + X_2'y$ so that

$\beta^{OLS} = [X_1'X_1 + X_2'X_2]^{-1}[X_1'y + X_2'y]$

$= [X_1'X_1 + X_2'X_2]^{-1}[X_1'X_1\beta_1^{OLS} + X_2'X_2\beta_2^{OLS}]$

(b) The weight for β_1^{OLS} is the ratio of the sum of the x squares for the first data set to the sum of the x squares for both data sets. This makes sense because the bigger the sum of x squares for a data set, the more precise is its estimate, so the bigger weight it should have.

M3 Use pooled formula from M1(a) above, calculating $(X_1'X_1)^{-1}$, for example, from the variances and covariances of A's regression, divided by A's estimated error variance. This produces pooled estimates of α and β of 126/33 and 96/33, respectively.

N1 It is a solution in the sense that it may lower the MSE of other coefficient estimates. The bias introduced by dropping the variable could be more than offset (in terms of MSE) by the reduction in variance.

N3 False. t statistics may be small, but this smallness reflects accurately the lack of precision with which coefficients are estimated. Inference is not biased. Although variances get bigger, estimates of these variances get bigger as well, keeping variance estimates unbiased.

N5 False. The explanatory variables could together do a good job of explaining variation in the dependent variable but because of collinearity each individually has a low t statistic.

N7 Questionable proposal because although it reduces variance, it introduces bias in the estimate of β_1 by giving x_1 credit for all variation in y that matches joint variation in x_1 and x_2. Note that the β_2 estimate is unaffected. The Ballentine exposits this neatly.

N9 Problem is multicollinearity, probably between the two price indices, and possibly between YD and POP. Address the former by replacing the price indices by their ratio, something that makes better economic sense. Address the latter by dropping POP; there is no good theoretical reason for its inclusion, since the consumption data are in per capita terms.

N11 **(a)** $\Sigma xy/(\Sigma x^2 + k)$.
(b) $\Sigma x^2/(\Sigma x^2 + k) = 1/(1 + k/\Sigma x^2)$.
(c) Optimal shrinking factor is that minimizing MSE = $\beta^2(1-\theta)^2 + \theta^2\sigma^2/\Sigma x^2$. This is minimized for $\theta = \beta^2/(\beta^2 + \sigma^2/\Sigma x^2)$ implying a value for k of σ^2/β^2.
(d) Neither σ^2 nor β^2 are known.

N13 **(a)** $a'\beta^{OLS} = \lambda'X'X(X'X)^{-1}X'y = \lambda'X'y$ which can be calculated.
(b) $E\lambda'X'y = E\lambda'X'(X\beta + \varepsilon) = \lambda'X'X\beta = a'\beta$. $V(\lambda'X'y) = E\lambda'X'\varepsilon\varepsilon'X\lambda = \sigma^2\lambda'X'X\lambda$.

O1 **(a)** The intercepts for the males and the females, 13 and 10, respectively, will not change. Call the new intercept a and the new dummy variable slope b. Then for males the intercept in the new specification is $a + 2b = 13$ and for females is $a + b = 10$. Solving this for a and b we get $a = 7$ and $b = 3$, implying the new equation $y = 7 + 2x + 3D$.
(b) For males we have $a + b = 13$ and for females we have $a - b = 10$. Solving this we get $y = 11.5 + 2x + 1.5D$.

O3 **(a)** To avoid perfect multicollinearity.
(b) No. Need to test the dummy coefficients all equal to one another, not all equal to zero. Further, this should be done using an F test. Restricted SSE from regression with intercept and no dummies. Unrestricted SSE as described in question. Numerator df = number of regions minus one. Denominator df = sample size minus number of regressors minus number of regions.

O5 Nothing. When expressed in raw data instead of logarithms, the dummy enters multiplicatively, ensuring that its influence is commensurate with the size of the economy.

O7 **(a)** No intercept difference between southern males and females.
(b) No intercept difference between northern males and females.
(c) $\beta_2 + \beta_4 = 0$.
(d) The way in which dummy variables are defined affects the way in which hypotheses are tested.

O9 **(a)** $\partial \ln y/\partial K = \beta = (1/y)\partial y/\partial K$
(b) For males $y = e^{\alpha}e^{\beta K}e^{\delta}$ and for females $y = e^{\alpha}e^{\beta K}$ so the percentage change in going from a female to a male is $100(e^{\alpha}e^{\beta K}e^{\delta} - e^{\alpha}e^{\beta K})/e^{\alpha}e^{\beta K} = 100(e^{\delta}-1)$
(c) It is a nonlinear function of δ^{OLS} and so its expected value is not equal to the nonlinear function of the expected value.
(d) If ε is distributed normally, then δ^{OLS} is distributed normally with mean δ and variance $V(\delta^{OLS})$. Thus $\exp(\delta^{OLS})$ is distributed log normally with mean $\exp[\delta + 0.5V(\delta^{OLS})]$. Thus $\exp[\delta^{OLS} - 0.5V(\delta^{OLS})]$ is distributed with mean e^{δ}. This suggests estimating $100(e^{\delta}-1)$ by $100\{\exp[\delta^{OLS} - 0.5V^*(\delta^{OLS})] - 1\}$ where the * indicates an estimate.
(e) Expand $\exp(\delta^{OLS})$ in a Taylor series around $E\delta^{OLS} = \delta$ to get $\exp(\delta^{OLS}) = e^{\delta} + (\delta^{OLS} - \delta)e^{\delta} + (1/2)(\delta^{OLS} - \delta)^2e^{\delta} + \cdots$.
Dropping higher-order terms and taking expectations we get that $E\exp(\delta^{OLS})$ is approximately $e^{\delta}[1 + (1/2)V(\delta^{OLS})]$ suggesting that $100(e^{\delta}-1)$ be estimated by $100\{\exp(\delta^{OLS})\,[1 + (1/2)V^*(\delta^{OLS})]^{-1} - 1\}$.

O11 It avoids the multicollinearity and improves R^2, but is not a good idea. The coefficient on the treatment dummy will measure the difference between the base person in the treatment group and the base person in the control group. This is a case in which you need to write out the specification for each individual to deduce the correct interpretation of the dummy coefficient.

O13 The intercepts are 4, 2.5, 6, and 1, for quarters 1 through 4, respectively. If the new regression has the first quarter as its base, the new regression's intercept has to be 4 to keep the first quarter intercept unchanged (i.e., it needs to increase by 3, the coefficient of D1 which will disappear with the change in base). Because the new regression's intercept is bigger by 3, all the other slope coefficients need to decrease by 3 to keep their respective intercepts unchanged. The D2 coefficient becomes –1.5, the D3 coefficient becomes 2, and the D4 coefficient becomes –3. The D3 coefficient is the old D3 coefficient

(5) less the old D1 coefficient (3), so its variance will be the variance of the difference between these two coefficient estimates, namely the variance of the estimate 5 plus the variance of the estimate 3 minus twice their covariance. A more general way of solving here is to write the new specification as $y = a + 2x + b\text{D2} + c\text{D3} + d\text{D4}$. The intercepts will not change, so quarter 1 intercept will be $a = 4$, quarter 2 intercept will be $a + b = 2.5$, quarter 3 intercept will be a + c = 6, and quarter 4 intercept will be $a + d = 1$. The new D3 coefficient is $c = 6 - a = 1 + 5 - (1 + 3) = 5 - 3$, so its variance is the variance of $(5 - 3)$.

O15 **(a)** Savings of a person in the control group before the experiment.
(b) $(\delta - \gamma) - (\alpha + \beta - \alpha) = \theta = \delta - \gamma - \beta$.
(c)

$$\begin{aligned} y &= \alpha + \beta \text{control after} + \gamma \text{treat before} \\ &\quad + (\theta + \gamma + \beta)\text{treat after} + \varepsilon \\ &= \alpha + \beta(\text{control after} + \text{treat after}) \\ &\quad + \gamma(\text{treat before} + \text{treat after}) \\ &\quad + \theta \text{treat after} + \varepsilon \\ &= \alpha + \beta \text{after} + \gamma \text{treat} + \theta \text{treat} * \text{after} + \varepsilon. \end{aligned}$$

P1 **(a)** Intercept estimate is the estimate of y for the females, and so must be 2, the average of the females. The sum of the intercept and slope estimates is the estimate of y for the males, and so must be 3, the average value of the males; this implies that the slope estimate is 1.
(b) Numerator is $6 + 2 - 3 = 5$. Use $10(X'X)^{-1}$ to get variances and covariances, where $X'X$ has 50 and 20 on the diagonal and 20 on the off-diagonal. Denominator is square root of $9 \times 2/6 + 4 \times 5/6 - 12 \times 2/6 = 14/6$.
(c) As a standard normal; its square is distributed as a chi-square with 1 degree of freedom.

P3 The new regression results are $y = 10 + 3x - 6\text{DM}$, so the numerator of the t statistic is -6. This -6 is obtained as $4 - 10$, so its variance is $V(4) + V(10) - 2C(4,10) = 1 + 4 - 1 = 4$. The t statistic is -3.

P5 Set the thirty-fourth observation on ln$w$ equal to zero; the negative of the slope coefficient on the observation-specific dummy is the forecast.

Q1 Set up $d = \alpha + \beta y + \delta p + \theta_1 \text{D}_{\text{sp}} + \theta_2 \text{D}_{\text{su}} + \theta_3 \text{D}_{\text{fa}}$ where the notation is obvious – the base category is winter. Perform an F test to test $\theta_1 = \theta_2 = \theta_3$. Obtain unrestricted SSE from regressing d on a constant, y, p, and the three dummies. Denominator degrees of freedom $N - K = 48–6 = 42$. Obtain restricted SSE by regressing d on a constant, y, p, and a dummy that is zero in winter and otherwise one (i.e., $\text{D}_{\text{sp}} + \text{D}_{\text{su}} + \text{D}_{\text{fa}}$). Numerator degrees of freedom, the number of restrictions, are two. Slightly easier to set up with one of spring, summer, or fall as the base category.

Q3 Whether the coefficients' differences are significant can be tested by testing whether the observation in question is consistent with the relationship as estimated by using the other observations. Use a Chow test in which this single observation is the second data set. Easiest is to use an observation-specific dummy.

Q5 **(a)** Test λ significantly greater than zero using a one-sided t test.
(b) (i) $-\delta$ (ii) $-\theta$.
(c) Test $\delta = \theta$ using either a t test or an F test. Restricted SSE for F test comes from regressing y on a constant, Ed, IQ, Ex, Sex, and a dummy with value one if French only or if English only, zero if bilingual (DF + DE).
(d) Test $\lambda + \delta = \theta$ using either a t or an F test. Restricted SSE for F test comes from regressing y on a constant, Ed, IQ, Ex, a dummy with value two if a English-only male, one for other males and zero for all others (Sex + DE), and a dummy with value one if French only or if English only, zero if bilingual (DF + DE).
(e) Include a variable ExSex that takes the value zero for females and the value of Ex for males. Test its coefficient against zero with a t test.

Q7 Write A's regression results as $G = \alpha + \beta Y + \delta_1\text{DM} + \delta_2\text{DO} + \delta_3\text{DW}$ and B's as
$G = \phi Y + \lambda_1\text{DM} + \lambda_2\text{DO} + \lambda_3\text{DW} + \lambda_4\text{DQ}$.
(a) (i) δ_2 (ii) $\lambda_4 - \lambda_2$. They should be the same because changing dummy definitions has no impact on estimates of fundamental parameters.
(b) (i) Do a t test for $\delta_2 = 0$. (ii) Do a t test for $\lambda_4 = \lambda_2$ or an F test in which the restricted SSE

is obtained by regressing G on y, DM, DW, and a dummy with value one for observations in either Ontario and Quebec and zero elsewhere (DO + DQ).

(c) Run the regression $G = \alpha + \beta Y + \delta_1\text{DM} + \delta_3\text{DW} + \theta Y\text{DQ}$, where YDQ is a variable taking value Y for all Quebec observations and zero elsewhere.

(d) Get restricted SSE from regression of part (c). Get SSE unrestricted by running four regressions, one for each region, adding together the SSEs. Degrees of freedom for this are $K = 8$. Number of restrictions $J = 3$.

Q9 Restricted SSE from regression given in the question. Unrestricted by replacing CAT with four dummies for categories 2 through 5. Numerator df = difference between number of parameters estimated in unrestricted versus restricted regressions, in this case $8 - 5 = 3$. Denominator d$f = N - 8$.

R1 **(a)** Write the relationship as $y = \alpha + \beta x + \theta\text{D} + \delta\text{DX}$ where the dummy D takes the value one for x values greater than x^* and zero otherwise, and the variable DX takes the value x for x values greater than x^* and zero otherwise. This produces two different lines for x values above and below x^*, but does not guarantee that these lines intersect at $x = x^*$. For this to be true, $\alpha + \beta x^* = \theta + \delta x^*$.

Building this restriction into the estimating equation above we get $y = \theta + \delta x^* - \beta x^* + \beta x + \theta\text{D} + \delta\text{DX} = \theta(1+\text{D}) + \beta(x - x^*) + \delta(x^* + \text{DX})$.

(b) Testing continuity means testing whether or not the two lines intersect at x^* (so there is no jump at that point), or that $\alpha + \beta x^* = \theta + \delta x^*$. Easiest way to do this is to use an F test with the unrestricted regression that given at the beginning of part (a) above, and the restricted regression that given at the end of part (a).

R3 **(a)** Need a period-specific dummy for 1966, for 1967, for 1968, for 1969, and a dummy with value zero before 1970, value one otherwise. Estimating eight parameters.

(b) Five: α, β, and the three slopes of the cubic; the cubic intercept is just α.

(c) Write the intercept during the transition period as $\text{int}(1965+i) = \alpha + \alpha_1 i + \alpha_2 i^2 + \alpha_3 i^3$, where i is the number of years since 1965, running from zero (for 1965 and earlier) to 5 (for 1970 and later). Write out the equation for each time period in the data and substitute this expression into this list of observations. It will be seen that we should regress on an intercept, x and the following three special variables: variable $w1$ with value zero before 1966, value 1 in 1966, value 2 in 1967, value 3 in 1968, value 4 in 1969, and value 5 otherwise; variable $w2$ with values the squares of the $w1$ values; and variable $w3$ with values the cubes of the $w1$ values.

S1 There is some truth in this if the distribution of the errors is not known, but if this distribution is known, building that knowledge into the estimation procedure by using MLE creates better estimates.

S3 Let x be the unknown IQ. The score of 140 has resulted from drawing x from $N(100, 400)$, in conjunction with drawing a testing error $140-x$ from $N(0, 40)$. The likelihood is $(2\pi)^{-1/2}(1/20)\exp[-(1/800)(x-100)^2](2\pi)^{-1/2}(1/\sqrt{40})\exp[-(1/80)(140-x)^2]$. Maximizing yields $x^{\text{MLE}} = 136.4$. It is more likely that the score of 140 was obtained by a person with a lower IQ who had a lucky test, than by someone with an actual IQ of 140.

S5 **(a)** β^{OLS} is BLUE, but α^{OLS} is biased because $E\varepsilon = \lambda/(\lambda-1) \neq 0$.

(b) Yes. If λ is known, $E\varepsilon$ is known, allowing the bias in α^{OLS} to be removed.

(c) Yes. If λ is known, $\sigma^2 = V(\varepsilon)$ is known (equal to $\lambda/[(\lambda-2)(\lambda-1)^2]$), implying that it need not be estimated in the formula $\sigma^2(X'X)^{-1}$.

(d) Use MLE to estimate all three parameters simultaneously.

S7 **(a)** The probability that x comes from group i is $(2\pi)^{-K/2}(\det\Sigma)^{-1/2}\exp[-(1/2)(x-\mu_i)'\Sigma^{-1}(x-\mu_i)]$.

It is more likely to have come from group 1 if $\exp[-(1/2)(x-\mu_1)'\Sigma^{-1}(x-\mu_1)]/\exp[-(1/2)(x-\mu_2)'\Sigma^{-1}(x-\mu_2)] > 1$

manipulation of which produces the desired result.

(b) The probability above for x coming from group i must be multiplied by its prior probability before proceeding. Further, if the cost of misclassifying a group 1 observation is, say, 10 times the cost of misclassifying a group 2 observation, then one would only classify into group 2 if the probability of it being from group 2 was at least 10 times the probability of it coming from group 1, to ensure that the classification decision minimizes expected cost. This implies that the numerator of the final expression in part (a) should be multiplied by 10.

S9 Let $x = g(w)$ where g is an unknown function. Since the density of w is $f(w) = 1$, $f(x) = |dw/dx| \times 1$ which we want to be $3e^{-3x}$. From this, $w = e^{-3x}$ and so values of x can be created by drawing a w from $U(0,1)$ and calculating $x = -(\ln w)/3$.

S11 **(a)** To create an unbiased estimate.
(b) Dividing by $N-1$ creates a new estimate that can be written as $[N/(N-1)]$ times the old estimate. So its variance is $[N/(N-1)]^2$ times the variance of the old estimate.

T1 **(a)** Likelihood is $(2\pi\sigma^2)^{-N/2}\exp[-(1/2\sigma^2)\Sigma(x-\mu)^2$ and log-likelihood is $-(N/2)\ln 2\pi - (N/2)\ln\sigma^2 - (1/2\sigma^2)\Sigma(x-\mu)^2$. First partial with respect to μ is $(1/\sigma^2)\Sigma(x-\mu)$ which when set equal to zero yields $\mu^{MLE} = \bar{x}$.
(b) Second partial wrt μ is $-(1/\sigma^2)N$. Second cross-partial (wrt σ^2) is $-(1/\sigma^4)\Sigma(x-\mu)$, the expected value of which is zero, so the Cramer–Rao lower bound is just the negative of the inverse of the expected value of $-(1/\sigma^2)N$, which is σ^2/N.

T3 Likelihood is $k^N\exp(-k\Sigma x)$ and log-likelihood is $N\ln k - k\Sigma x$. The first partial is $N/k - \Sigma x$ which when set equal to zero yields $k^{MLE} = N/\Sigma x$. Second partial is $-N/k^2$ so the Cramer–Rao lower bound is k^2/N.

T5 **(a)** The MLE formula is given in question T4 which for this case is 1.92.
(b) From question T4 the variance of the MLE is α^2/N the square root of which in this case is estimated as 0.19. The t statistic is 0.08/0.19 = 0.42 so the null is accepted for reasonable significance levels.

T7 **(a)** Likelihood is $\lambda^{\Sigma x}e^{-N\lambda}\Pi(x!)^{-1}$ and the log-likelihood is $\Sigma x\ln\lambda - N\lambda - \Sigma\ln(x!)$. First partial is $\Sigma x/\lambda - N$ which when set equal to zero yields $\lambda^{MLE} = \bar{x}$.
(b) Second partial is $-\Sigma x/\lambda^2$ Expected value of x is λ, so Cramer–Rao lower bound is λ/N.

T9 **(a)** $\mu = \lambda/(\lambda+1)$ and $V(x) = \lambda/(\lambda+2)(\lambda+1)^2$.
(b) $\lambda^{MLE} = -N/\Sigma\ln x$.
(c) $\mu^{MLE} = \lambda^{MLE}/(\lambda^{MLE}+1) = N/(N-\Sigma\ln x)$.
(d) CRLB = λ^2/N.
(e) Asy. Var $\mu^{MLE} = \{\partial[\lambda^{MLE}/(\lambda^{MLE}+1)]/\partial\lambda^{MLE}\}^2V(\lambda^{MLE}) = \lambda^2/N(\lambda+1)^4$.
(f) Sample mean is unbiased so $E\bar{x} = \mu$, and its variance is given by a well-known formula, σ^2/N, which in this case is $\lambda/N(\lambda+2)(\lambda+1)^2$.
(g) μ^{MLE} should have the smaller asy. var. because it is the MLE. This can be verified by showing that the ratio of $V(\bar{x})$ to $V(\mu^{MLE})$ exceeds one.
(h) The sample mean is BLUE and so in small samples is attractive. In sufficiently large samples, however, the properties of the MLE are superior.

T11 **(a)** prob(yes) = prob(blue) × prob(cheat) + prob(green) which can be solved for prob(cheat) as a function of the data which provides prob(yes) and the other two probabilities which are known.
(b) Use maximum likelihood to estimate. Write prob(cheat) as a function of income and a dummy for gender and insert in the expression above for prob(yes). The likelihood function is formed using this probability expression for each yes answer and one minus this expression for each no answer.

T13 $-(N/2)\ln 2\pi - (N/2)\ln\sigma^2 - (2\sigma^2)^{-1}\Sigma(y^\delta - \alpha + \beta x)^2 + N\ln\delta + (\delta-1)\Sigma\ln y$. The last two terms come from the Jacobian.

T15 **(a)** $EP = \int P\theta^P e^{-\theta}/P! = \theta\int\theta^{P-1}e^{-\theta}/(P-1)! = \theta$.
(b) $\Sigma P(\alpha+\beta x) - \Sigma\exp(\alpha+\beta x) - \Sigma\ln(P!)$.

U1 The informed prior will not be centered over the true value of β (except by accident), so β^* will be biased; as the sample size grows, however, the influence of the prior falls, so β^* is consistent. Because it incorporates additional information, the variance of β^* is smaller than that of β^{OLS}.

U3 Expected loss of β^* is the integral from zero to β^* of $4(\beta^*-\beta)\beta$ plus the integral from β^* to one of $2(\beta-\beta^*)\beta$. This turns out to be $\beta^{*3} - \beta^* + 2/3$.

Minimizing this with respect to β^* yields the point estimate $1/\sqrt{3}$.

U5 **(a)** The *t* statistic for testing the null $\beta=1$ is 1.28 which is less than the one-sided 5% critical value 1.65 so the null would be accepted and the venture not undertaken.

(b) A normal distribution with mean 2.28 and variance 1.

(c) Area under posterior to right of 1.0, divided by remaining area = 0.9/0.1 = 9.

(d) The Bayesian will decide not to undertake the project if $1800 \times \text{prob}(\beta \leq 1) > Q \times \text{prob}(\beta \geq 1)$ or if $Q < 1800 \times \text{prob}(\beta \leq 1)/\text{prob}(\beta \geq 1) = 200$.

U7 Find the area under the predictive distribution for y_{T+1} below the value y_T.

U9 Posterior is proportional to $\theta^{0.5}(1-\theta)^{3.5}\theta^{5}(1-\theta)^{95} = \theta^{5.5}(1-\theta)^{98.5}$ which is a beta distribution with mean 6.5/(6.5+99.5) = 0.061. Because of the quadratic loss function, this is the Bayesian point estimate.

U11 The posterior distribution of the forecast is normal with mean 6.4 and standard error 0.1. The Bayesian probability of a fall in the interest rate is the probability that it will be below 6.3, which is the probability of a normal variable being one standard deviation or more below its mean. This is found from the normal tables to be about 16%. The expected payoff to the hedging strategy is $.16 \times 20000 - .84 \times 5000 = -1000$; she should invest in the guaranteed investment certificate.

U13 Expected payoff is the probability of a number being correct times that number. Guessing 52 maximizes this expected payoff ($52 \times 0.0685 = 3.562$).

V1 **(a)** Mean = 0.8η; variance = $0.026\eta^2$.

(b) Write g_2 as proportional to $(\eta/2)^{0.5}[1-(\eta/2)]^{0.5}$. Then $\eta/2$ has mean one-half, implying that η has mean one. Variance of $\eta/2$ is 1/16 so variance of η is 1/4.

(c) The prior on the returns to scale parameter is centered at constant returns to scale. When all inputs are increased by the same percentage amount, what fraction of the output is attributable to the increased labor input? The prior on α is centered such that this fraction is 0.8.

V3 **(a)** Uniform prior for $\ln\sigma$ implies that σ has prior proportional to $d\ln\sigma/d\sigma = 1/\sigma$.

(b) Uniform prior for $\ln[\theta/(1-\theta)]$ implies that θ has prior proportional to $d\ln[\theta/(1-\theta)]/d\theta = [\theta(1-\theta)]^{-1}$.

(c) Uniform prior for $\ln[\rho^2/(1-\rho^2)]$ implies that ρ has prior proportional to absolute value of $d\ln[\rho^2/(1-\rho^2)]/d\rho$, or $[\rho(1-\rho^2)]^{-1}$.

V5 ρ^2 with prior $(\rho^2)^{-0.5}(1-\rho^2)^{-0.5}$ implies that ρ has prior proportional to absolute value of $(\rho^2)^{-0.5}(1-\rho^2)^{-0.5}\, d\rho^2/d\rho$, or $(1-\rho^2)^{-0.5}$. Note that this differs from the answer to 3(c) above.

W1 This Monte Carlo procedure creates 5000 *t* values for testing the true null that $bz = 4$. The 4750th value should cut off the upper 5% of the t distribution with 17 d*f*, so from the *t* tables it should be 1.74.

W3 This Monte Carlo procedure calculates 2000 *F* values with 3 and 3 degrees of freedom. The 20th value should cut off the lower 1%. This can be found as the inverse of the tabulated *F* value that cuts off the highest 1%. This is the inverse of 29.46 = 0.034.

W5 **(a)** (i) choose parameter values, say $\theta = 2$, $\delta = -.2$ and $\sigma^2 = 6$, ensuring that $\beta = 1$; (ii) set ctr = 0; (iii) draw 25 ε values from $N(0,6)$; (iv) create 25 m values as $2 + y - 0.2r + \varepsilon$; (v) regress *m* on *y* and calculate the *t* statistic for testing $\beta = 1$, namely $t = (\text{bols}-1)/\text{sb}$ where sb is the standard error of bols; (vi) add one to ctr if $t > 1.714$; (vii) repeat from (iii) to produce 4000 *t* values; (viii) divide ctr by 4000 to produce an estimate of the type I error.

(b) By omitting a relevant, correlated explanatory variable, bols is biased, in this case upwards since both δ and the correlation are negative. This should cause the *t* values to be biased upward, making the type I error exceed 5%.

W7 **(a)** See if 4 is unusual relative to the 1000 *z* slope estimates, say, in the 2.5% tails.

(b) In the bootstrapping procedure calculate the t statistic for testing the *z* slope equal to 5, to obtain 1000 *t* values. Then see if the *t* value calculated from the original regression is unusual relative to these 1000 *t* values.

W9 **(i)** Specify regression model, say $y = \alpha + \beta x + \delta w + \varepsilon$ and choose values for α, β, and σ^2, the variance of ε. Set $\delta = 0$. (ii) Set sample size = 23, say. Select 23 values for x and w. (iii) Get computer to generate 23 errors from $N(0, \sigma^2)$ and use these to calculate the 23 corresponding y values. (iv) Regress y on a constant, x and w. Save the t statistic on the w coefficient estimate. (v) Repeat from (iii) until you get 10000 t statistic values, say. (vi) Order these t values from smallest to largest. (viii) Check that the values from the t table for degrees of freedom 20 are matched by the appropriate values in your set of 10000. For example, from the t table the value 2.086 is such that it cuts off the upper 5% tail, so the value 9500 in your list of t values should be very close to 2.086.

W11 The 2000 numbers you have calculated should be distributed as a chi-square with 1 degree of freedom. Value 1900 should cut off the upper 5% tail. From the chi-square table this critical value is 3.84.

W13 **(a)** (i) Choose sample size 60, 60 x values, and parameter values $\alpha = 2$ and $\beta = 1$, ensuring thereby that the null is true, required for calculation of the type I error. (ii) Set ctr = 0. (iii) Draw 60 ε values from a distribution uniform between −1 and 1. (iv) Calculate 60 y values as $2 + x + \varepsilon$. (v) Regress y on x, compute the t statistic t^* for testing $\beta = 1$. (vi) Add one to ctr if t^* exceeds the 5% critical value in absolute value. (vii) Repeat from (iii) to obtain 800 t^* values. (viii) Estimate the type I error as d^* = ctr/800.
(b) The type I error estimate d^* is a sample proportion whose standard error under the null hypothesis is se* = square root of $d(1-d)/800$ where in this case $d = 0.05$. To test if the type I error is different from 5%, calculate the absolute value of $(d^*-0.05)$/se* and compare to the 1% critical value from the normal distribution.

X1 Your best guess of the distribution of a sample average is the mean and variance associated with your first sample. Consequently, you should conceptualize the average of the second sample as being drawn out of this distribution. Any average equal to, higher than, or slightly smaller than the mean of this distribution will cause the null to be rejected, so your best guess should be slightly over 50%.

X3 No. Should pool the data. If assume s^2 is same for both samples, then pooled t statistic is 2.8.

X5 With such a large sample it is not surprising that variables are statistically significant (the too-large sample size problem). The low R^2 may result from the fact that the variance of the error term is quite high – it is not uncommon in cross-sectional data for the error term to be playing a dominant role.

X7 Uncertain. If the null hypothesis is true, the t test statistic should continue to take values obtained conceptually from a t distribution. But if the null hypothesis is false, it should become very large in absolute value.

X9 When testing a set of linear restrictions the high value of the F statistic arises from violation in either positive or negative directions and so is a two-sided test, albeit one with only one critical value. When testing equivalence of variances, however, as in the Goldfeld–Quandt test, it can be constructed as either one-sided, to test one variance different from the other in a specific direction, or two-sided, without specifying a specific alternative. In the latter case the two critical values are inverses of one another.

X11 Type I error stays at its chosen level, by tradition often chosen to be 5%; type II error shrinks. So as more and more information becomes available the traditional methodology devotes it entirely to shrinking the type II error.

Y1 False. If σ^2 becomes larger, the variance of β^{OLS} becomes larger, which makes it more difficult to reject the null hypothesis – power falls.

Y3 Size = type I error = .5/5 = 10%; power = 15.5/20 = 77.5%.

Y5 **(a)** The variance of the sample mean here is 256/64, so its standard deviation is 2. When the true mean is 40, the probability of a sample mean greater than 43 is the probability of a z value greater than (43–41)/2 = 1.5 which is 6.7%, the probability of a type I error.
(b) When the true mean is 41, the probability of a sample mean greater than 43 is the probability of a z value greater than (43–41)/2 = 1.0 which is 16%, so power is 16%. When the true mean is 43, 45, and 47, power is 50%, 84%, and 97.7%, respectively.

Z1 Ridiculous; the OLS residuals sum to zero, so the test will always accept the null.

Z3 **(a)** The first equation is a restricted version of the second, where the restriction is that $\beta_3 = -\beta_2$. Consequently, the unrestricted equation, the second equation, will have the higher R^2.

(b) Prefer α_2* since it incorporates the true restriction – it is unbiased and has a smaller variance than the others.

(c) Numerator of the t test is $\beta^*_3 + \beta^*_2$. Denominator is the square root of the sum of their estimated variances plus twice their estimated covariance.

(d) Restricted SSE from first regression, unrestricted SSE from second regression, number of restrictions is one, and other degree of freedom is $N - 4$.

Z5 Elasticity is $\beta_1 + 2\beta_2\ln Q$. For this to be unity, we must have $\beta_1 = 1$ and $\beta_2 = 0$. Use F test with unrestricted SSE from original regression, restricted SSE from regression of $\ln C - \ln Q$ on a constant, numerator degrees of freedom equal 2 (the number of restrictions), and denominator degrees of freedom $N - 3$.

Z7 Estimate $\ln y = \lambda + \alpha\ln L + \beta\ln K + \theta L + \delta K$ and test $\theta = \delta = 0$.

Z9 Unrestricted SSE from original regression. Restricted SSE from regression that results from substituting for each β_i the expression involving the δ_i and grouping on the basis of the δ_i. Since the number of coefficients directly estimated has been reduced by two, there are two restrictions implicit in this structure. The other d*f* is $N - 6$.

Z11 Specify $y = \alpha + \beta x + \theta w + \alpha_1 D_1 + \alpha_2 D_2 + \beta_1 D_1 x + \beta_2 D_2 x + \theta_1 D_1 w + \theta_2 D_2 w + \varepsilon$ where D_1 is a dummy equal to one in the first period, zero otherwise, and D_2 is a dummy equal to zero in the second period, zero otherwise, and then test $\beta_1 = \theta_1 = \beta_2 = \theta_2 = 0$. Numerator F d*f* is 4; denominator d*f* is $N - 9$.

Z13 RESET: Obtain the predicted values from the original regression and form a new variable $c2$ by squaring them and a new variable $c3$ by cubing them. Rerun the regression adding $c2$ and $c3$ as extra explanatory variables and test their coefficients jointly against zero. Rainbow: Create observation-specific dummies for the three observations with the highest y values and for the three observations with the lowest y values. Rerun the regression with these six dummies as extra explanatory variables and test their coefficients jointly against zero.

Z15 Estimate with a Box–Cox transformation on x to get the unrestricted maximized log-likelihood. Run the first regression to get the restricted maximized log-likelihood for the null of equation (1). Use an LR test to test the Box–Cox coefficient λ equal to zero. Run the second regression to get the restricted maximized log-likelihood for the null of equation (2). Use an LR test to test λ equal to unity. Both specifications could be accepted, both rejected, or one accepted and the other rejected.

AA1 SSE for first regression is $0.605 \times (39–5) = 20.54$. For second and third regressions SSE is 25.22 and 57.87, respectively. ΔSSE = 12.11. Number of restrictions is 5. F statistic is $(12.11/5)/(45.75/66) = 3.49$, greater than the 5% critical value of 2.37, so the null that the parameters are the same is rejected.

AA3 **(a)** $F = ((130–100)/2)/(100/(24–4)) = 3$.

(b) $F = ((130–80)/4)/(80/(24–6)) = 2.81$.

(c) (i) Choose values for α, β for the first period, and different values for the second period. Choose a common value for σ^2. (ii) Select 24 observations on x and set counters m and n equal to zero. (iii) Get the computer to generate 24 errors from $N(0, \sigma^2)$ and calculate the corresponding 24 y values. (iv) Perform the first Chow test and if it rejects the null, increase the counter m by one; perform the second Chow test and if it rejects the null, increase the counter n by one. (v) Repeat from (iii) say 500 times. (vi) If $m > n$ the first Chow test is more powerful.

AA5 The numerator of the t statistic is $0.75 + 0.4 - 1 = 0.15$, and the denominator is the square root of $0.015 + 0.015 + 2 \times 0.005 = 0.2$, yielding $t = 0.75$, so the null is accepted.

AA7 Use a Wald statistic to test the nonlinear restriction $\beta_1\beta_2 - 1 = 0$. (This is more reliable than testing $\beta_1 - 1/\beta_2 = 0$; if possible avoid writing restrictions with denominators). The numerator is 0.04, the square of $\beta_1\beta_2-1$ evaluated at the parameter estimates. The first derivative vector of $\beta_1\beta_2-1$ is $(\beta_2, \beta_1)'$ which evaluated at the OLS estimates is $(0.2, 4.0)'$. The estimated variance of $\beta_1\beta_2-1$ using the usual formula for the variance of a nonlinear function is 1.12 and the W statistic is $0.04/1.12 = 0.036$. The critical value at the 5% significance level for a chi-square with one d*f* is 3.84, so the null is not rejected.

AA9 **(a)** $\theta_1\theta_2 + \theta_3 = 0$.
(b) First derivative vector is $(\theta_2, \theta_1, 1)'$. W statistic is given by $0.25/[(0.5,3,1)V^{**}(0.5,3,1)']$ where V^{**} is the block of V^* omitting the intercept ingredients.

BB1 $E[\text{SSE}/\sigma^2] = N-K$, so $E(\text{SSE}) = \sigma^2(N-K)$. Thus $Es^2 = E(\text{SSE})/(N-K) = \sigma^2$. $V(\text{SSE}/\sigma^2) = 2(N-K)$, so $V(\text{SSE}) = 2\sigma^4(N-K)$. Thus $Vs^2 = V(\text{SSE})/(N-K)^2 = 2\sigma^4/(N-K)$.

BB3 **(a)** A t statistic is the ratio of a standard normal to the square root of an independent chi-square divided by its degrees of freedom. The recursive residuals are iid as $N(0, \sigma^2)$ so that their sum is distributed as $N(0, (T-k)\,\sigma^2)$. Thus $(\Sigma e_t)\,\sigma^{-1}(T-k)^{-1/2}$ is distributed as a standard normal, forming the numerator of the t statistic. The expression in curly brackets is a chi-square divided by its degrees of freedom, providing the *es* are divided by σ. This σ cancels out with the σ in the numerator, leaving the t statistic.
(b) Their expected value is zero, their variance is σ^2, and they are independent of one another.
(c) No. The variance of each OLS residual is not σ^2, and they are not independent of one another.

BB5 If the CNLR model applies, under the null hypothesis $(\beta_1^{\text{OLS}} - \beta_2^{\text{OLS}})$ is distributed normally with mean zero and variance V, say, and $(\beta_1^{\text{OLS}} - \beta_2^{\text{OLS}})'V^{-1}(\beta_1^{\text{OLS}} - \beta_2^{\text{OLS}})$ is distributed as a chi-square with degrees of freedom equal to the number of elements in β. Because the two regressions are run separately, β_1^{OLS} and β_2^{OLS} are independent of one another. Thus $V = [\sigma_1^2(X_1'X_1)^{-1} + \sigma_2^2(X_2'X_2)^{-1}]$, which when estimated and employed in the formula above gives rise to the suggested statistic.

BB7 **(a)** Get residuals from regressing y on a constant and x. Use NR^2 from regressing these residuals on a constant, x, w, and z. These are the derivatives of the specification with respect to the parameters.
(b) Dividing by two, the number of restrictions, creates an "asymptotic" F, with 2 and "infinity" degrees of freedom. (This matches producing a chi-square from an F by multiplying the F by the number of restrictions.) NR^2 is a chi-square; dividing by the number of restrictions produces the numerator of the F statistic. The denominator is s^2/σ^2. For an infinite sample size, s^2 becomes σ^2 causing the denominator to become unity.

BB9 **(a)** Numerator is $\beta^{\text{OLS}} - (\delta^{\text{OLS}})^2$ and denominator is the square root of $V^*(\beta^{\text{OLS}}) - 4\delta^{\text{OLS}}C^*(\beta^{\text{OLS}}, \delta^{\text{OLS}}) + 4(\delta^{\text{OLS}})^2V^*(\delta^{\text{OLS}})$ where * denotes estimate of. This comes from the formula for the variance of a nonlinear function of random variables.
(b) Square the asymptotic t to get a W, a chi-square with one d*f*.

BB11 Log-likelihood is $N\ln\theta - \theta\Sigma x$, first partial is $N/\theta - \Sigma x$ and second partial is $-N/\theta^2$, so $\theta^{\text{MLE}} = N/\Sigma x$ and Cramer–Rao lower bound is θ^2/N. W is $(N/\Sigma x - \theta_0)^2(\Sigma x)^2/N$ and LM is $(N/\theta_0 - \Sigma x)^2\theta_0^2/N$, both of which equal $(N - \theta_0\Sigma x)^2/N$.

CC1 **(a)** Likelihood is $(2\pi\sigma^2)^{-N/2}\exp[-(1/2\sigma^2)\Sigma(x-\mu)^2]$, maximized at $\mu = \bar{x}$.
Likelihood ratio λ is
$\exp[-(1/2\sigma^2)\Sigma(x-\mu_0)^2 + (1/2\sigma^2)\Sigma(x-\bar{x})^2]$
$\text{LR} = -2\ln\lambda = (1/\sigma^2)\Sigma(x-\mu_0)^2 - (1/\sigma^2)\Sigma(x-\bar{x})^2 = (\bar{x}-\mu_0)^2/(\sigma^2/N)$, the square root of which is the usual test statistic.
(b) $W = (\bar{x}-\mu_0)'[V(\bar{x}-\mu_0)]^{-1}(\bar{x}-\mu_0) = (\bar{x}-\mu_0)^2/(\sigma^2/N)$.
(c) Partial of log-likelihood wrt μ is $(1/\sigma^2)\Sigma(x-\mu)$, equal to $Q = (N/\sigma^2)(\bar{x}-\mu_0)$ when evaluated at μ_0.
$\text{LM} = Q'[V(Q)]^{-1}Q = (N/\sigma^2)^2(\bar{x}-\mu_0)^2[(N/\sigma^2)^2(\sigma^2/N)]^{-1} = (\bar{x}-\mu_0)^2/(\sigma^2/N)$.

CC3 Write y_1 as $x'Ax$ where $x = (x_1,x_2)'$ and A is a matrix with top row 1/2, −1/2, and bottom row −1/2, 1/2. Write y_2 as $x'Bx$ where $x = (x_1,x_2)'$ and B is a matrix with top row 1/2, 1/2 and bottom row 1/2, 1/2. Confirm that A and B are both idempotent with trace 1 and that $AB = 0$. Use theorem that $x'Ax$ is distributed as a chi-square with d*f* trace A if A is idempotent. And that $x'Ax$ and $x'Bx$ are distributed independently if $AB=0$.

CC5 The relation $1 - [\text{SSE}_R/(N-K)]/v > 1 - [\text{SSE}_U/(N-K+J)]/v$ can be manipulated to become $(N-K-J)\text{SSE}_R < (N-K)\text{SSE}_U$. This same expression can be obtained by manipulating $[(\text{SSE}_R - \text{SSE}_U)/J]/[(\text{SSE}_U/N-K-J)] < 1$.

DD1 **(i)** Specify model as, say, $y = \alpha + \beta x + \delta w + \varepsilon$. Choose values for α, β, and σ^2. (ii) Set sample size $N = 35$, say, and select 35 values for x and for w. (iii) Set $\delta = 0$. (iv) Get computer to generate 35 errors from $N(0, \sigma^2)$ and use to calculate 35 corresponding y values. (v) Run regression and test $\delta = 0$. (vi) If null is accepted, set $\delta^* = 0$. If null is rejected, set $\delta^* = \delta^{\text{ols}}$. Save δ^*. (vii) Repeat from (iv) until, say, 500 values for δ^* have been obtained. (viii) Use the 500 values of δ^* to estimate the MSE of δ^*. This gives the height of the risk function for zero degree of falseness of the restriction $\delta = 0$. (ix) Change the value of δ and repeat from (iv) to graph different points on the risk function.

DD3 Not a good suggestion. Pretest bias continues to exist and should be worse because using a smaller sample for the test means it will have less power.

EE1 Use a non-nested *F* test or a *J* test.

EE3 For the first equation the data allow estimation of diff = β (lnNl − lnNs) and for the second equation diff = $\phi\,(e^{-\delta Ns} - e^{-\delta Nl})$. Use a nonlinear *J* test. First, run nonlinear least squares to estimate the second equation and obtain predicted diff values diff2. Then run OLS on the first equation adding diff2 as an extra regressor. Accept or reject the first equation using a t test of the diff2 coefficient. Second, run OLS on the first equation and obtain predicted diff values diff1. Then run nonlinear least squares on diff = $\phi\,(e^{-\delta Ns} - e^{-\delta Nl}) + \theta$diff1. Accept or reject the second equation using a t test on θ. Alternatively, a *P* test could be employed.

FF1 4(i) Specify $y = \alpha + \beta x + \varepsilon$ and choose values for α and β, say 1 and 2. (ii) Set sample size $N = 35$, say, and select 35 values of x. (iii) Get computer to draw 35 errors from $N(0,1)$, multiply each by the square of its corresponding x value to create heteroskedasticity, and use to calculate the 35 corresponding y values. (iv) Run regression and calculate t statistic for testing $\beta = 2$. Save it. (v) Repeat from (iii) until you have, say, 500 t statistics. (vi) Calculate the percent of these t values that are greater than 1.645. Inference will be biased if this number is not close to 5%.

FF3 **(i)** Specify $y = \alpha + \beta x + \varepsilon$ and choose values for α and β. (ii) Set sample size $N = 40$, say, and select 40 values of x. (iii) Get computer to draw 40 errors from $N(0,1)$, multiply the last 20 of these by 2, and use to calculate the 40 corresponding y values. (iv) Run OLS and save β^{OLS}. (v) Run OLS on first 20 observations to estimate s_1^2, run OLS on last 20 observations to estimate s_2^2, transform the last 20 observations by multiplying them by s_1/s_2. Obtain β^{EGLS} by regressing using first 20 observations and last 20 transformed observations. (v) Repeat from (iii) until you have, say, 500 β^{OLS} and 500 β^{EGLS}. (vi) Use the 500 β^{OLS} to estimate the bias, variance and MSE of β^{OLS}; use the 500 β^{EGLS} to estimate the bias, variance and MSE of β^{EGLS}. Compare.

FF5 Suppose the sample size is 25 and the original error ε is such that $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$ where u is a spherical error with variance σ^2. Then the variance of ε is $\sigma^2/(1-\rho^2)$. Obtain ε_1 by having the computer draw an error from $N[0, \sigma^2/(1-\rho^2)]$. Have the computer draw errors u_2 through u_{25} from $N(0, \sigma^2)$ and use these (and ε_1) to compute ε_2 through ε_{25} using $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$.

FF7 **(i)** Run the regression and obtain the parameter estimates, the residuals and the Goldfeld–Quandt statistic. (ii) Use the residuals to calculate s_1^2 and s_2^2 the estimated error variances for the first and second halves of the data, respectively. (iii) Divide the first half residuals by s_1 and the second half residuals by s_2 to create a set of residuals called the adjusted residuals.

(iv) Use the estimated parameter values and draws with replacement from the adjusted residuals to create a new set of data on y. (v) Calculate the Goldfeld–Quandt statistic. (vi) Repeat from (iv) to obtain, say, 1000 Goldfeld–Quandt statistics. (vii) See if the original Goldfeld–Quandt statistic is in the tails of the distribution of these 1000 values.

GG1 False. OLS always provides the best fit (highest R^2).

GG3 Bad proposal. Both OLS and GLS are unbiased regardless of the presence of nonspherical errors, so under both the null and the alternative their difference has expectation zero.

GG5 No because GLS is BLUE.

HH1 The transformation renders the errors heteroskedastic, but since OLS is unbiased in the presence of nonspherical errors, no bias results.

HH3 Although the R^2 may be affected, this is not of consequence because we are interested in the important properties of bias and efficiency.

HH5 False. Regressing $(y+w)$ on x produces coefficient estimate $\Sigma(y+w)x/\Sigma x^2 = \beta^{OLS} + \alpha^{OLS}$: the estimates are identical.

HH7 Sum the equations for the tth time period to get $C_t = \beta_0 N_t + \beta_1 Y_t + \varepsilon_t^*$ where $\varepsilon_t^* = \Sigma\varepsilon_i$ has mean zero and variance $\sigma^2 N_t$. Use the GLS estimator, found by dividing all data by the square root of N_t before running OLS.

HH9 The two equations must be written as a single equation, the constraints incorporated and a transformation made for heteroskedasticity. Left-hand side vector is a column of ys on top of a column of $(p-w)$s. Columns of observations on explanatory variables are for α_0 a column of ones on top of a column of zeroes; for β_0 a column of zeroes on top of a column of ones; for α_1 a column of xs on top of a column of $-w$s; for α_2 a column of zs on top of a column of zs; and for β_2 a column of zeroes on top of a column of qs. All observations in the bottom half must be multiplied by the square root of two.

HH11 OLS is BLUE (the unconditional variance is constant), but may not be the estimator of choice because a nonlinear estimator that incorporates a correction for the conditional heteroskedasticity may be more efficient.

HH13 Should be multiplying through by s_2/s_1. But more importantly, all that the Goldfeld–Quandt test tells us in this context is that there is heteroskedasticity associated with x; it does not tell us that the form of this heteroskedasticity is such that there are two different variances for the two sets of data. The heteroskedasticity could be that the error variance is connected in some more continuous way to x, such as that the variance is proportional to x or x squared. Further testing should be done to better determine the nature of the heteroskedasticity.

HH15 Correct. The variance of a sample proportion is $p(1-p)/N$ where p is the true proportion and N the sample size. This variance can be estimated for each district. Transforming the data by dividing by the square root of this estimated variance and then running OLS is the standard correction for heteroskedasticity.

HH17 $E(u^*) = 0$ because $E(\hat{\varepsilon}) = 0$.
$V(u^*) = E(u^{*2}) = (1+\sqrt{5})/(2\sqrt{5})[(1-\sqrt{5})/2]^2\ \hat{\varepsilon}^2 + [1-(1+\sqrt{5})/(2\sqrt{5})]\ [(1-(1-\sqrt{5})/2]^2\ \hat{\varepsilon}^2$ which equals $\hat{\varepsilon}^2$.

II1 **(a)** For a 2-week period data, we have $y^* = 2\alpha + \beta x^* + \varepsilon^*$, where * denotes the sum of the 2 weeks data. The error ε^* is spherical, with variance $2\sigma^2$ (where $V(\varepsilon) = \sigma^2$), so OLS is applicable.
(b) For the moving average data, we have $y_t^* = \alpha + \beta_t x_t^* + \varepsilon_t^*$, where $y_t^* = (y_{t-1}+y_t+y_{t-2})/3$, $x_t^* = (x_{t-1}+x_t+x_{t-2})/3$, and $\varepsilon_t^* = (\varepsilon_{t-1}+\varepsilon_t+\varepsilon_{t-2})/3$. Although ε_t^* is homoskedastic, with variance $\sigma^2/3$, there is autocorrelation. Adjacent ε_t^* have covariance $2\sigma^2/9$, ε_t^* two periods apart have covariance $\sigma^2/9$ and all other covariances are zero. Thus GLS should be used, with the error variance–covariance matrix having 3s down the diagonal, 2s beside the diagonal, 1s beside these 2s and all other elements zero.

II3 **(a)** Use a Wald test. See answer AA9.
(b) Incorporating the restriction, the equation can be rewritten as $y_t - \alpha_2 y_{t-1} = \alpha_1 + \alpha_3$

$(x_t - \alpha_2 x_{t-1}) + \varepsilon_t$. Estimate by doing an iterative search over α_2. Select a value for α_2 and regress $y_t - \alpha_2 y_{t-1}$ on a constant and $x_t - \alpha_2 x_{t-1}$ Change the value of α_2 and repeat, continuing until SSE is minimized.

(c) The relationship can be written as $y_t - \alpha_2 y_{t-1} = \alpha_1 + \alpha_3(x_t - \alpha_2 x_{t-1}) + \varepsilon_t$ or $(1 - \alpha_2 L)y_t = \alpha_1 + \alpha_3(1 - \alpha_2 L)x_t + \varepsilon_t$. This implies $y_t = \alpha_1/(1 - \alpha_2 L) + \alpha_3 x_t + \varepsilon_t/(1 - \alpha_2 L)$ or $y_t = \alpha_1/(1 - \alpha_2) + \alpha_3 x_t + u_t$ where $u_t = \alpha_2 u_{t-1} + \varepsilon_t$, a first-order autocorrelated error.

(d) That finding a first-order autocorrelated error may imply a dynamic misspecification rather than a genuine autocorrelated error.

II5 The DW statistic is not relevant here because the data are cross-sectional. An exception is if the data are ordered according to the value of an independent variable, in which case the DW statistic may indicate a nonlinearity.

JJ1 Use a Goldfeld–Quandt test. Run the regression using the N_m male observations to get s^2_m and using the N_f female observations to get s^2_m. Then s^2_m/s^2_f is distributed as an F with N_m and N_f degrees of freedom.

JJ3 F = (45/(18–3))/(14/(10–3)) = 1.5 which is less than the 5% critical value of 3.51 for degrees of freedom 15 and 7 so the null is accepted.

JJ5 The low DW in a cross-sectional context may mean that the data have been ordered according to the value of x and an inappropriate functional form is being used. The high Breusch–Pagan statistic may mean that z should appear as an explanatory variable.

KK1 **(a)** $\beta^{OLS} = 3$; $\beta^{GLS} = 4$.

(b) $V(\beta^{OLS}) = (X'X)^{-1}X'VX(X'X)^{-1} = 1.7/9$; $V(\beta^{GLS}) = (X'V^{-1}X)^{-1} = 1/8$.

(c) $(\varepsilon^{OLS'}\varepsilon^{OLS}/2)(X'X)^{-1} = 4/3$.

(d) $(\varepsilon^{GLS'}V^{-1}\varepsilon^{GLS}/2)(X'V^{-1}X)^{-1} = 1.0$.

KK3 **(a)** Divide the data by the square root of x to deal with the heteroskedasticity. This produces transformed observations 3, 5, 5 on y and 1, 2, 3 on x. Running OLS on this gives $\beta^{GLS} = 2$. Estimated variance is $(\varepsilon^{GLS'}\varepsilon^{GLS}/2)(X'X)^{-1} = (3/2)/14 = 3/28$.

(b) $(X'V^{-1}X)^{-1}X'V^{-1}y = 28/14 = 2$. Estimated variance $= (\varepsilon^{GLS'}V^{-1}\varepsilon^{GLS}/2)(X'V^{-1}X)^{-1} = (3/2)/14 = 3/28$.

(c) $(X'V^{-1}X)^{-1}/(X'X)^{-1}(X'VX)(X'X)^{-1} = (1/14)/(794/98 \times 98) = 0.86$, so GLS is about 14% more efficient.

KK5 The OLS estimate is the sample mean, 2. To correct for the heteroskedasticity, divide each observation by the square root of its variance. Applying OLS to the transformed data we get 1 (the average of 1, 1, and 1).

LL1 Both estimators are unbiased. $V(K^*) = E(1/4)(\varepsilon_1 + \varepsilon_2 + \varepsilon_3 + \varepsilon_4)^2 = (1/16)E(u_{-1} + 2u_0 + 3u_1 + 3u_2 + 2u_3 + u_4)^2 = (28/16)(1/3) = 7/12$.
$V(K^{**}) = E(1/2)(\varepsilon_1 + \varepsilon_4)^2 = (1/4)E(u_{-1} + u_0 + u_1 + u_2 + u_3 + u_4)^2 = (6/4)(1/3) = 1/2$. Thus prefer K^{**} since it has a smaller variance.

LL3 **(a)** $\beta^{OLS} = 24/5$.

(b) The $V(\varepsilon)$ matrix is 9 times a matrix with rows 1, 0.5 and 0.5, 1. Putting this into the GLS formula produces $\beta^{GLS} = 5$.

(c) $V(\beta^{OLS}) = (X'X)^{-1}(X'VX)(X'X)^{-1} = 63/25$. $V(\beta^{GLS}) = (X'V^{-1}X)^{-1} = 9/4$.

LL5 GLS formula is $(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}$y where X is a 2×1 vector of ones, Ω is a 2×2 matrix with 2 and 3 on the diagonal and 1 in the off-diagonal positions, and y is a 2×1 vector containing 10 and 7. This yields the estimate 9.

MM1 Write the two relationships as one so that the vector of observations on the dependent variable consists of the N observations on y_1 followed by the N observations on y_2, and the column of observations on the regressor is N ones followed by N zeroes. The error vector's variance–covariance matrix is $V \otimes I$ where V is a matrix with rows 2, 1 and 1, 2 and I is a $N \times N$ identity. Using the GLS formula produces $\beta^{GLS} = \bar{y}_1 - (1/2)\bar{y}_2$.

MM3 **(a)** Write the two relationships as one so that the vector of observations on the dependent variable consists of the observations on y followed by the observations on q, and the column of observations on the first regressor is the x observations followed by a vector of zeroes; the column of observations on the second regressor is a vector of zeroes followed by the w observations. The error vector's variance–covariance matrix is $V \otimes I$ where V is a matrix with rows 2, 1 and 1, 3. Using the GLS formula produces $\alpha^{GLS} = 2/5$ and $\beta^{GLS} = 9/10$.

(b) From the GLS calculation earlier we can also calculate $V(\alpha^{GLS}) = 0.5$, $V(\beta^{GLS}) = 3.0$, and $C(\alpha^{GLS}, \beta^{GLS}) = 0.5$. So $V(2\alpha^{GLS} - \beta^{GLS}) = 2.0 + 3.0 - 1.0 = 4.0$. The required t value is $-0.1/2 = -0.05$, below any reasonable critical value, so the null is accepted.

(c) Regress y on x to get $\alpha^{OLS} = 3/4$; $SSE = \Sigma y^2 - \alpha^{OLS}\Sigma xy = 22.75$, so estimate $V(u)$ as 2.275. Regress q on w to get $\beta^{OLS} = 1$; SSE = 32, so estimate $V(v)$ as 3.2. Estimate covariance by $\Sigma\,(y - \hat{y})\,(q - \hat{q})\,/\,10 = (\Sigma yq - \Sigma q\hat{y} - \Sigma y\hat{q} + \Sigma\hat{y}\hat{q})\,/\,10 = 1.24$. This last calculation exploits the coefficient estimates, so that, for example, $\Sigma q\,\hat{y} = (2/5)\,\Sigma qx$.

NN1 Uncertain. If β^* is a better (i.e., smaller variance) estimate of β than that produced by regressing y on x and w, then the statement is true, but if β^* is a bad estimate, this suggestion will produce a worse estimate of δ.

NN3 **(a)** The stochastic prior information is incorporated as an additional observation, namely a y-value of 6 and an x-value of 1. To correct for heteroskedasticity this extra observation should be multiplied through by 2. The new Σxy is $186 + 24 = 210$, and the new $\Sigma x^2 = 30$, so the β estimate is 7.

(b) Variance is 16/30. Variance without incorporating additional information is 16/26, about 15% higher.

NN5 Add an extra, artificial observation to the bottom of the current data. Add a one to the bottom of the y vector, a zero to the bottom of the column of ones for the intercept term, and ones to the bottom of the columns of observations on K and L. From the original regression, find s^2 the estimate of the variance of the error term. Adjust the row of artificial observations by multiplying them all by $s/0.1$. Run OLS on the extended data.

NN7 If we have $\text{VALUE}_i = \alpha + \beta\text{AREA}_i + \gamma\text{YEAR}_i + \varepsilon_i$ for the ith property, then we know $750\,000 = \alpha + 900\beta + 1965\gamma$ so we have exact extraneous information. This can be incorporated in the usual fashion – solve for one of the unknown parameters in the second equation and substitute into the first equation to find a new estimating equation with one less parameter.

OO1 Minimize the generalized sum of squared errors $(y-X\beta)'\Omega^{-1}(y-X\beta)$ subject to the constraints to obtain $\beta^* = \beta^{GLS} + (X'\Omega^{-1}X)^{-1}R'[R(X'\Omega^{-1}X)^{-1}R']^{-1}(r-R\beta^{GLS})$.

OO3 **(a)** The W matrix is block diagonal with a different X_i matrix in each block. When Σ is diagonal then $\Sigma^{-1}\otimes I$ is block diagonal with $\sigma_i^{-2}I$ in each block. When multiplied out this produces $(X_i'X_i)^{-1}X_i'y_i$ for the ith equation's parameter estimates.

(b) The W matrix is block diagonal with X in each block and so can be written as $I\otimes X$. So $[W'(\Sigma^{-1}\otimes I)W]^{-1} = [(I\otimes X')(\Sigma^{-1}\otimes I)(I\otimes X)]^{-1} = [(I\otimes X')(\Sigma^{-1}\otimes X)]^{-1} = [\Sigma^{-1}\otimes X'X]^{-1} = \Sigma\otimes(X'X)^{-1}$. Similarly $W'(\Sigma^{-1}\otimes I)y = (\Sigma^{-1}\otimes X')y$ so SURE $= [I\otimes(X'X)^{-1}]y$.

PP1 From a regression of y on X, the regression sum of squares is $(\hat{y}-\bar{y})'(\hat{y}-\bar{y}) = (X\beta^{OLS})'.X\beta^{OLS} - N(\bar{y})^2 = y'X(X'X)^{-1}X'y - N(\bar{y})^2$. This implies that from a regression of $y-\bar{y}$ on X, the regression sum of squares is $(y-\bar{y})'X(X'X)^{-1}X'(y-\bar{y})$, since the average of $y-\bar{y}$ y- is zero.

$(e^2-s^{*2})Z(Z'Z)^{-1}Z'(e^2-s^{*2})/2s^{*4} = (1/2)[(e^2/s^{*2}-1]Z(Z'Z)^{-1}Z'[(e^2/s^{*2}-1]$ which is one-half the regression sum of squares of $[(e^2/s^{*2}-1]$ regressed on Z, since the average of the $[(e^2/s^{*2}-1]$ is zero.

PP3 **(a)** To allow this formulation to include a case of homoskedasticity.

(b) Only the first element of α is nonzero.

(c) $\exp(\ln k + \theta\ln w) = kw^\theta$; $\alpha_1 = \ln k$, $\alpha_2 = \theta$ and $x_2 = \ln w$.

PP5 **(a)** $\beta^{IV} = (W'X^*)^{-1}W'y^* = (X'P^{-1}PX)^{-1}X'P^{-1}Py = \beta^{OLS}$.

(b) Estimated $V(\beta^{IV}) = s^2(W'X^*)^{-1}W'W(X^{*\prime}W)^{-1} = s^2(X'X)^{-1}X'P^{-2}X(X'X)^{-1}$.

(c) This is s^2 times the heteroskedasticity-consistent estimate of the variance–covariance matrix of the OLS estimator.

QQ1 **(a)** Arrange the observations so that the two observations on the first household appear, then the two observations on the second household, and so on. The variance–covariance matrix is then seen to be the variance of ε times a block-diagonal matrix with 2 by 2 blocks containing ones down the diagonal and ρ on the off-diagonal.

(**b**) Transform each household's observations as though you had data on just that household. The first observation is multiplied by $\sqrt{(1-\rho^2)}$ and the second observation has ρ times the first observation subtracted from it.

QQ3 $p\lim\beta^{OLS} = \beta + p\lim(\Sigma y_{t-1}\varepsilon_t/N)/p\lim(\Sigma y_{t-1}^2/N)$. Numerator is σ^2 and denominator is $(2\beta\sigma^2 + 2\sigma^2)/(1-\beta^2)$ so bias is $(1-\beta)/2$.

QQ5 The GLS estimator $\mu^* = (X'W^{-1}X)^{-1}X'W^{-1}y$ is the BLUE, where X is column of ones, y is a vector with observations $x1$, $x2$, and $x3$, and W is a 3×3 matrix with ones down the diagonal and zeros on the off-diagonal except for the (1,2) and (2,1) positions in which there is 0.5. This produces $\mu^* = (0.5\times1 + 0.5\times2 + .75\times3)/1.75$. The variance of the sample mean is $(X'X)^{-1}X'WX(X'X)^{-1} = 4/9$. The variance of m^* is $(X'W^{-1}X)^{-1} = 0.75/1.75$, which is smaller.

RR1 Setting the derivative with respect to y equal to zero and solving for y gives
$y_t = \theta y^*_t + (1-\theta)y_{t-1}$ where $\theta = \alpha_1/(\alpha_1+\alpha_2)$. This can be rewritten as $y_t = y_{t-1} + \theta(y^*_t - y_{t-1})$ which is a partial adjustment model.

RR3 (**a**) Use L without a subscript to denote the lag operator and write YP_t as $Y_t/[1-(1-\lambda)L]$. Thus $C_t = (\beta-\alpha\theta)Y_t/[1-(1-\lambda)L] + \alpha L_{t-1} + \varepsilon_t$ which when multiplied through by $[1-(1-\lambda)L]$ yields the estimating equation
$C_t = (1-\lambda)C_{t-1} + (\beta-\alpha\theta)Y_t + \alpha L_{t-1} - \alpha(1-\lambda)L_{t-2} + \varepsilon_t - (1-\lambda)\varepsilon_{t-1}$.
(**b**) Main estimating problem is that θ and β are not identified. Also the MA error term is contemporaneously correlated with C_{t-1}.

RR5 (**a**) From the partial adjustment equation $K_t = (1-\lambda)K_{t-1} + \lambda\theta Y_t + \varepsilon_t$, which when substituted into the gross investment equation yields
$I_t = (\delta-\lambda)K_{t-1} + \lambda\theta Y_t + \varepsilon_t$. If δ is known, λ and θ are identified; if δ is not known, none of the parameters is identified.
(**b**) $K_t = (\lambda\theta Y_t + \varepsilon_t)/[1-(1-\lambda)L]$ where L is the lag operator. Thus
$I_t = (\delta-\lambda)\lambda\theta Y_{t-1}/[1-(1-\lambda)L] + \lambda\theta Y_t + \varepsilon_t/[1-(1-\lambda)L]$ and thus
$I_t = (1-\lambda)I_{t-1} + (\delta-\lambda)\lambda\theta Y_{t-1} + \lambda\theta Y_t - (1-\lambda)\lambda\theta Y_{t-1} + \varepsilon_t$ so
$I_t = (1-\lambda)I_{t-1} + (\delta-1)\lambda\theta Y_{t-1} + \lambda\theta Y_t + \varepsilon_t$.
Both λ and θ are over-identified if δ is known; if δ is not known, all parameters are just identified.
(**c**) Setting $Y_t = Y_{t-1}$ and $I_t = I_{t-1}$, the coefficient on Y in the long-run relationship is $\delta\theta$. It makes sense for λ to drop out in the long run, since it reflects the speed of short-run adjustment. $\delta\theta$ makes sense because it just says that in the long run an increase in income requires an increase in investment to cover the depreciation on the higher capital stock.

RR7 (**a**) $y_t - \alpha y_t = \eta - \alpha\Delta y_t + \beta_0 x_t + \beta_1 x_t - \beta_1 x_t + \beta_1 x_{t-1} + \varepsilon_t$
$y_t = \eta/(1-\alpha) - [\alpha/(1-\alpha)]\Delta y_t + [(\beta_0 + \beta_1)/(1-\alpha)]x_t - [\beta_1/(1-\alpha)]\Delta x_t + \varepsilon_t/(1-\alpha)$.
(**b**) $y_t - y_{t-1} = \eta - (1-\alpha)y_{t-1} + \beta_0 x_t + \beta_1 x_t - \beta_1 x_t + \beta_1 x_{t-1} + \varepsilon_t$
$\Delta y_t = (1-\alpha)\{\eta/(1-\alpha) + [(\beta_0 + \beta_1)/(1-\alpha)]x_t - y_{t-1}\} - \beta_1\Delta x_t + \varepsilon_t$
(**c**) y and x are growing at the same rate.

SS1 Must recognize that the αs are not identified, so only examine estimation of the βs. (i) Set values for the parameters, including $V(\varepsilon_1)$ and $V(\varepsilon_2)$. (ii) Select sample size, say 35, and choose 35 values for the exogenous variables Y and A. (iii) Solve for the reduced form to obtain equilibrium expressions for P and Q in terms of the parameters, the exogenous variables, and the error terms. (iv) Get computer to draw 35 ε_1 and 35 ε_2 values. (v) Use these errors and the reduced form expressions to calculate the corresponding 35 values for P and for Q. (vi) Use the data to run OLS to get β^{OLS} and 2SLS to get β^{2SLS}. Save these estimates. (vii) Repeat from (iv) until you have, say, 800 sets of these estimates. (viii) Use the 800 OLS estimates to estimate the bias, variance, and MSE of OLS and use the 800 2SLS estimates to estimate the bias, variance, and MSE of 2SLS. Compare.

TT1 Uncertain. Both estimators are biased in small samples, the sample mean ratio is asymptotically unbiased but has a larger variance than β^{OLS} which is asymptotically biased.

TT3 (**a**) With variables expressed as deviations about their means, the equation of interest is $y^* = \beta x^*$ or in terms of measured variables, $y = \beta x - \beta\varepsilon - \varepsilon$.

plimβ^{OLS} = plim($\Sigma xy/N$)/plim($\Sigma x^2/N$) = β – (β+1) σ^2/(Q+ σ^2)
where Q is the plim of Σx^{*2}/N.
(b) The bias is negative, which tends to discredit the argument in question.

TT5 **(a)** Slope coefficient estimates are still BLUE, but intercept estimate has bias $-2\alpha_2$.
(b) Estimates are BLUE except that the estimate of α_2 is actually an estimate of $\alpha_2/1.15$.
(c) All estimates are biased, even asymptotically. (The intercept estimate is a biased estimate of $\alpha_0-2\alpha_2$.)

UU1 **(a)** Predicted X is $W = Z(Z'Z)^{-1}Z'X$, so $(W'W)^{-1}W'y = [X'Z(Z'Z)^{-1}Z'Z(Z'Z)^{-1}Z'X]^{-1}X'Z(Z'Z)^{-1}Z'y = (Z'X)^{-1}Z'Z\ (X'Z)^{-1}X'Z(Z'Z)^{-1}Z'y = (Z'X)^{-1}Z'y = \beta^{IV}$.
(b) Suggests $\sigma^2(W'W)^{-1} = \sigma^2[X'Z(Z'Z)^{-1}Z'Z(Z'Z)^{-1}Z'X]^{-1} = \sigma^2(Z'X)^{-1}Z'Z\ (X'Z)^{-1}$.

UU3 **(a)** Both estimators are unbiased. MSE of β^{OLS} is $\sigma^2/\Sigma x^2 = \sigma^2/6$. MSE of β^{IV} is $\sigma^2\Sigma w^2/(\Sigma xw)^2 = 14\sigma^2/49$. Their ratio is 7/12.
(b) β^{IV} is 10, so numerator is 2. Residuals are -11, 4 and 1, so $s^2 = 69$. Thus denominator is square root of 69(14/49).

UU5 **(a)** Using m as an instrument for i produces $\beta^* = \Sigma m^2/\Sigma mi$. Regressing i on m produces a coefficient estimate $\Sigma mi/\Sigma m^2$. The reverse regression estimate of β is the inverse of this, so these two estimates are identical.
(b) Use an instrument for i that takes the i values when i is determined exogenously and the m values when m is determined exogenously.

UU7 **(a)** Regress h on w and z and obtain the predicted h values, *h*hat. Perform a Hausman test by regressing h on w, z, and *h*hat and testing the coefficient of *h*hat against zero.
(b) Regress y on i and *h*hat to produce the iv estimator.

VV1 False. OLS provides the best fit. Other methods outperform OLS on other criteria, such as consistency.

VV3 **(a)** The suggested estimator is 2SLS.
(b) No. The equation is not identified.

VV5 **(a)** β^{OLS} is 100/50 = 2. β^{2SLS} is 90/30 = 3. β^{ILS} is (90/80)/(30/80) = 3.
(b) None; the first equation is not identified.

VV7 **(a)** Use OLS since x is exogenous.
(b) 2SLS is $\Sigma xy_2/\Sigma xy_1$. Reduced form is $y_2 = \alpha\beta x$ plus error, so ILS is $(\Sigma xy_2/\Sigma x^2)/(\Sigma xy_1/\Sigma x^2) = \Sigma xy_2/\Sigma xy_1$ = 2SLS.
(c) Second equation is just identified.
(d) We know from theoretical considerations that 2SLS is asymptotically unbiased. Should u_2 change, u_1 will likely change because they are not independent, causing y_1 to change. This implies that y_1 and u_2 are not independent. Consequently, OLS is asymptotically biased. OLS will have a smaller variance, however.
(e) The result that recursive simultaneous equations can be estimated unbiasedly by OLS requires that the errors be independent across equations.

VV9 Under (i) supply equation is a linear function of P_{t-1}. Under (ii) it is a linear function of P_{t-1} and lagged Q. Under (iii) it has only an intercept. Choose by testing the coefficients on lagged P and lagged Q against zero.

WW1 Ridiculous test – since the OLS residuals are orthogonal to X by construction, the coefficient vector in question is always zero.

WW3 **(a)** Zero, since in this case both estimators are unbiased.
(b) $E(X'X)^{-1}X'\varepsilon\ [(X'X)^{-1}X'\varepsilon - (Z'X)^{-1}Z'\varepsilon]' = E(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1} - E(X'X)^{-1}X'\varepsilon\varepsilon'Z(X'Z)^{-1} = 0$.
(c) $\beta^{IV} = \beta^{OLS} - q$, so $V(\beta^{IV}) = V(\beta^{OLS}) - 2C(\beta^{OLS},q) + V(q) = V(\beta^{OLS}) + V(q)$ so that $V(q) = V(\beta^{IV}) - V(\beta^{OLS})$.
(d) $q'[V(q)]^{-1}q$ is distributed as a chi-square with degrees of freedom = dimension of q.
(e) Null is that X and ε are independent so that under null both OLS and IV are unbiased. Alternative is that X and ε are not independent, so that OLS is biased in both large and small samples, and IV is asymptotically unbiased.

WW5 **(a)** OLS is 6.5 with variance 100/220 = 0.45. IV is 7.2 with variance 0.5. Difference is 0.7 with variance 0.05. Chi-square is 0.49/0.05 = 9.8.
(b) Regressing y on x and z gives coefficient estimate for z of 15.4 with variance 22. Square of resulting normal statistic is 10.78.

(c) Regressing x on z gives coefficient estimate of 2 which produces w observations 2, 4, −4, 8, and −10. Regressing y on x and w gives w coefficient estimate of 7.7 with variance 5.5. Square of resulting normal statistic is 10.78.

WW7 Under the null hypothesis that Y is exogenous, researcher B's estimates play the role of OLS in a standard Hausman test, and researcher A's estimates play the role of IV.

XX1 **(i)** Choose values for the intercept α and slope β. (ii) Set sample size, say 35, and select 35 x values. (iii) For each x value calculate $y^* = \alpha + \beta x + \varepsilon$, where ε is a drawing from a standard normal. Choose some critical value, say zero, and set $y = 1$ if y^* is greater than zero, otherwise set $y = 0$. (Choose this critical value so that your coefficient values and x values do not create y observations that are almost all zero or almost all one.) (iv) Regress y on x and an intercept, saving the slope coefficient estimate β^*. (v) Repeat from (iii) until you have, say, 500 β^*s. (vi) Average these 500 estimates to estimate the expected value of the OLS estimate; compare this to $\beta f(\alpha + \beta x)$ evaluated at the mean of the x values, where f is the density of a standard normal. Alternatively, the comparison could be with $F(\alpha + \beta x + 1) - F(\alpha + \beta x)$ evaluated at the mean of the x values, where F is the cumulative standard normal.

XX3 **(a)** Find the log-odds ratio of the dependent variable (i.e., $\ln(f/(1-f))$), which is then regressed on the explanatory variables.

(b) (i) Select values for β and the variance of the error term, choose a sample size N and N values for the explanatory variables. (ii) Draw N errors e. (iii) Use $X\beta + e$ to create N log-odds ratios q. (iv) Calculate N f values as $e^q/(1+e^q)$. (v) Regress f on X to obtain your estimates β^*, and regress q on X to obtain your friend's estimates β^{**}. (vi) The influence of an explanatory variable on f is the partial of f with respect to that explanatory variable. For you this is just β^*, but for your friend it is the partial of the logit function, estimated by $\beta^{**}f^*(1-f^*)$. This depends on the X values and so the comparison will depend on the X values chosen. Choose the average X values. (vii) Repeat from (ii) to obtain, say 500 β^* and 500 $\beta^{**}f^*(1-f^*)$ values. (viii) Average both of these 500 values and see which is closest to $\beta f\#(1-f\#)$ where $f\#$ is the value of f calculated using the true β and the average X values.

(c) Regress the log-odds ratio on the explanatory variables to obtain the parameter estimates and N residuals. Using these parameter estimates and draws with replacement from the residuals, create new log-odds values. Run the regression again and calculate the estimated first partial as in part (b) above. Repeat to obtain, say, 900 such estimated first partials. Compute the sample variance of these 900 estimates.

YY1 **(a)** prob(getting a loan) = $e^w/(1+e^w)$ where
$w = \alpha + \beta\text{GPA} + \delta\text{AGE} + \theta\text{SEX} + \phi\text{ MA} + \eta\text{PHD}$
where SEX is one for males and zero for females, MA is one for MA students and zero otherwise, and PHD is one for PhD students and zero otherwise.

(b) The likelihood function is the product of 45 terms. For the 25 students who were offered a loan, these terms take the form $e^w/(1+e^w)$ and for the 20 students who were not offered a loan, these terms take the form $1/(1+e^w)$.

(c) Probability estimated as $e^w/(1+e^w)$ where w is $\alpha^* + 3.2\beta^* + 23\delta^* + \theta^*$ where * denotes the MLE estimate.

(d) LR is easiest. Estimation with and without the restriction is easy to accomplish with the logit package. For W and LM tests, variances and partial derivatives are awkward to compute.

(e) Wish to test $\phi = \eta = 0$. Run logit unrestricted to obtain $\ln L_U$ the unrestricted maximized log-likelihood. Run restricted (by not including the last two explanatory variables) to obtain $\ln L_R$ the restricted maximized log-likelihood. Calculate $LR = 2(\ln L_U - \ln L_R)$, which is distributed (asymptotically) as a chi-square with 2 degrees of freedom.

YY3 **(a)** Model the probability of being at the limit as $e^w/(1+e^w)$ where $w = \alpha + \beta$SIZE + δIN1 + θIN2 where SIZE is firm size, IN1 is firm size for industry type 1, zero otherwise, and IN2 is firm size for industry type 2, zero otherwise.
(b) Use an LR test to test $\delta = \theta = 0$.

YY5 For the *i*th individual, with row vector x_i of characteristics, prob(yes) = prob($x_i\beta + \varepsilon_i \geq w_i$) = prob($\varepsilon_i \geq w_i - x_i\beta$)
Estimate as a probit or logit with the coefficient on w restricted to equal one.

YY7 **(a)** It tests if all the slope parameters are jointly equal to zero. It is calculated as twice the difference between the maximized log-likelihood using all the explanatory variables and the maximized log-likelihood using only the intercept.
(b) (i) Calculate xb^* where x is a row vector of values of the characteristics of the new observation (first element unity for the intercept) and b^* is the estimated parameter vector from the probit estimation procedure. The probability of this observation being a one is given by the integral from minus infinity to xb^* of the standard normal; this can be found from the normal tables. Predicting a one or a zero for this observation requires knowledge of the appropriate loss function to combine with this estimated probability. (ii) xb^* is now calculated using the estimated parameter values from the logit estimation procedure. The probability of this observation being a one is calculated as $\exp(xb^*)/(1+\exp(xb^*))$. Predicting a one or a zero for this observation requires knowledge of the appropriate loss function to combine with this estimated probability.
(c) 60 $y = 1$ predictions, 48 of which were correct.
(d) Computer has 66 of 100 correct for 66% correct. The $y = 1$ competitor gives 70 of 100 correct for 70% correct.
(e) One possibility is to compare the average of the percentage of 1s that are correct and the percentage of 0s that are correct. On this criterion the computer scores (48/60 + 18/40)/2 = 62.5% and the $y = 1$ competitor scores (60/60 + 0)/2 = 50%.

YY9 **(a)** Regressions without an intercept produce unreliable R^2 values, possibly outside the zero–one range.
(b) Use ordered logit or ordered probit.

YY11 None of the suggestions is appropriate because the slope is an estimate of the change in log-odds ratio, not an estimate of the change in probability. For a given x value you can calculate the probability of a one, then repeat this calculation for that x value increased by one unit. The difference between these two probabilities measures the desired influence of the x variable. But because of the nonlinearity this measure will depend on the x value chosen. One tradition is to choose the average of the explanatory variables. A competing method is to calculate this measure for every observation in the sample and then average the results.

YY13 Greater than, because the *S* curve is steeper at 0.4 than at 0.8.

YY15 **(a)** The likelihood is $\exp(K\alpha)^*[1+\exp(\alpha)]^{-N}$ and the log-likelihood is $K\alpha - N^*\ln(1+\exp(\alpha))$.
(b) First partial is $K\alpha - N^*\ln(1+\exp(\alpha))$, so $\alpha^{MLE} = \ln[K/(N-K)]$ and $p^{MLE} = K/N$. No surprise, the estimate of the probability of a one is the sample proportion of ones.
(c) Second partial is $-N^*[\exp(\alpha)/(1+\exp(\alpha))]^*[1 - \exp(\alpha)/(1+\exp(\alpha))]$ so Cramer–Rao lower bound is $V(\alpha^{MLE}) = [Np(1-p)]^{-1}$. The variance of p^{MLE} is $(\partial p/\partial\alpha)^{2*}V(\alpha^{MLE}) = p^2(1-p)^2 = p(1-p)/N$. No surprise, this is the traditional estimate of the variance of the sample proportion statistic.
(d) LR is $2^*\{[K\alpha^{MLE} - N^*\ln(1+\exp(\alpha^{MLE}))] - [0.3K - N^*\ln(1+\exp(0.3))]\}$
LM is $[0.3K - N^*\ln(1+\exp(0.3))]^{2*}N^*0.3^*0.7$.
W is $(\alpha^{MLE} - 0.3)^{2*}N^*0.3^*0.7$.

ZZ1 No. There is a selection bias problem – to enter the sample of male professors one has to live long enough to become a professor, invalidating the comparison.

ZZ3 Suppose the regression is true rent $y = X\beta + \varepsilon$. For those units under controls we know only that the y value is greater than or equal to the measured rent y_m so that

$y_m \le X\beta + \varepsilon$ or $\varepsilon \ge y_m - X\beta$. Estimation is via the Tobit model, in which observations on rent controlled units enter the likelihood function as the integral from $y_m - X\beta$ to infinity.

ZZ5 These data are censored from above and truncated from below. To accommodate the latter, each term in the likelihood must be divided by one minus the probability that the error term is such as to make income less than $5000. To accommodate the former, each observation with income of $100 000 must enter the likelihood as an integral from 100 000 − $X\beta$ to infinity.

ZZ7 **(a)** No adjustment is needed for a limited independent variable, although you may want to warn readers that the results may not apply to those under age 25.
(b) Need to avoid the perfect multicollinearity with household size by dropping household size as a regressor. May want to warn readers that the results may not apply to other than childless married couples.
(c) A sample selection bias correction procedure should be considered here, such as maximum likelihood or Heckman two-stage.

ZZ9 There is no sample selection problem here – calculation of the IMR and use of the Heckman two-stage procedure is inappropriate. The split between high and low income observations gives rise to a dummy which can be used in the usual way to check if the intercept and/or slopes differ between the two categories.

AB1 **(a)** The cumulative distribution of w is $1 - e^{-\lambda w}$, so the probability that someone's unemployment duration is greater than w is $e^{-\lambda w}$. The likelihood function is the product of $\lambda e^{-\lambda x}$ for the Nx individuals currently employed, times the product of $e^{-\lambda y}$ for the individuals still unemployed. The log-likelihood is thus $Nx\ln\lambda - \lambda\Sigma x - \lambda\Sigma y$ which is maximized at $\lambda^{MLE} = Nx/(\Sigma x+\Sigma y)$.
(b) The second derivative of the log-likelihood with respect to λ is $-Nx/\lambda^2$, so the Cramer–Rao lower bound is λ^2/Nx. Thus estimate variance by $Nx/(\Sigma x+\Sigma y)^2$.
(c) $Ew = 1/\lambda$. Model $1/\lambda$ as a function of variables representing local conditions, then replace λ in the likelihood by the inverse of this function and maximize with respect to the parameters of this function.

AC1 In reality, $y = \eta + \lambda x + \theta z + \varepsilon$. Suppose there is some collinearity between x and z so that $z = \delta + \gamma x + u$ where u is a random error with mean zero. Substituting for z we get $y = \eta + \theta\delta + (\lambda + \theta\gamma)x + \varepsilon + \theta u$ so that in the regression of y on only x, the coefficient on x is actually estimating $\lambda+\theta\gamma$. This will be biased downward as an estimate of λ if $\theta\gamma$ is negative. This will happen if z and x are positively correlated and the influence of z on y is negative, or if z and x are negatively correlated and the influence of z on y is positive.

AC3 The relevant substitute for Brazilian coffee is other brands of coffee, not tea, so there is a major misspecification.

AC5 The specification is inappropriate; the influence of inches of rainfall, for example, depends on the number of acres. A more appropriate specification is to regress output per acre on seed per acre, inches of rainfall, and hours of sunshine.

AC7 High collinearity between the two price variables can be addressed by using the ratio of the two prices. High collinearity between GDP and aggregate consumption can be addressed by dropping GDP; both are capturing the same influence, with the latter making more sense.

AC9 Individual income is a linear function of age and whether or not the individual has a university education. Aggregating over a census tract and dividing by the population of the census tract we get average census tract income a linear function of average age and the fraction of the census tract population with a university degree. The error term for this relationship is an average error for the people in the census tract and so will be heteroskedastic because pop is different in each census tract. Transform the data by multiplying all observations (including the intercept!) by the square root of pop.

AC11 The impact of percent body fat on weight should be different for a large person than for a small person (because it is measured as a percent), so a better specification would be to enter *F* interactively with *H*.

AC13 State GDP should be per capita. For cross-sectional data, the DW statistic has nothing to do with autocorrelated errors unless the data have been ordered in some special way. The heteroskedasticity correction should be accomplished by dividing the data by the square root of GDP, although this heteroskedasticity problem may disappear once per capita GDP is used. The dummy coefficients being nonzero in this specification without an intercept simply means that the intercepts for all the regions are different from zero; it does not mean the intercepts are different across regions. A significant coefficient on advertising does not mean advertising should be increased; we have to know how much increasing advertising costs relative to how much extra profit is created by the resulting increase in demand.

AC15 **(a)** If a dummy for each year (except the first) is introduced, the specification says that the log price is determined by the quality variables plus an intercept amount unique to each year. The required price index is calculated by tracking this intercept over time. In particular, if $a75$ is the coefficient estimate for the 1975 dummy and $a76$ the coefficient estimate for the 1976 dummy, then the quality-adjusted percentage change in price during 1976 is estimated by $\exp(a76 - a75) - 1$. Normalizing the first period price index to 100, the rest of the index is calculated by adjusting the index from year to year by the percentage changes calculated from year to year as indicated above. An easier way to do this would be to define the dummies differently, so that, for example, the dummy for 1976 would be unity for 1976 and all subsequent years, rather than just for 1976. This means that the estimated coefficient on the 1975 dummy now becomes an estimate of the difference between the 1975 and 1976 intercepts, facilitating the calculation given above.

(b) A constant quality-adjusted percentage price change would require a time trend in the specification instead of the dummies. This would produce the restricted sum of squared errors to contrast with the unrestricted sum of squared errors from the specification with the dummies, enabling an *F* test. The numerator degrees of freedom would be the number of parameters estimated in the unrestricted version less the number of parameters estimated in the restricted version, namely the number of dummies less one.

AC17 Put in a dummy equal to one for close to the dump, zero otherwise, and a dummy equal to one for 1991, zero otherwise. Also add an interaction dummy which is the product of these two dummies. The interpretation of the coefficient on the interaction dummy is the impact on price of the dump, beyond an overall price change over time. This is sometimes called the difference in differences method.

AC19 The low DW value has probably resulted from an incorrect functional form. If the data were ordered from smallest to largest *D* values (or reverse), a linear functional form would give rise to a low DW because for small *D* there should be practically 100% putts made and for large *D* practically 0% putts made. A logistic functional form would be more suitable.

AC21 Could consider a logistic functional form. Conclusions (a), (b), and (c) are OK. Conclusion (d) requires a test. A female scores an extra 10 points if in the tech group. A comparable male scores an extra 10, less 6 because of the interaction term tech*male, plus 5 for being male. The difference is 1, in favor of the female. The variance of this difference is the variance of the sum of the estimated coefficient of male and the estimated coefficient of tech*male.

AC23 **(a)** With coding one for male and zero for female, the percentage difference of males relative to females would be estimated as $\exp(b^*) - 1$ where b^* is the gender dummy coefficient estimate. The new coding gives rise to an estimate $\exp(-b^*) - 1$, derived in the same way in which the original formula was derived.

(b) Use a Box–Cox regression and test the null that the Box–Cox parameter lambda equals unity (linearity) and test the null that lambda equals zero (semilogarithmic), in both cases against an alternative of an unrestricted lambda, using two LR tests.
(c) Run the regression using the female data and obtain *s*2*f*, the estimate of the variance of the error. Run the regression using the male data and obtain *s*2*m*. Calculate *s*2*f*/*s*2*m* and compare it to the *F* value that cuts off the upper 5% (for example) of the *F* distribution with d*f* 67 and 55.
(d) Transform all the male observations (including the intercept) by multiplying them by the square root of *s*2*f*/*s*2*m*. Run OLS on all the observations (the original female observations and the transformed male observations). The benefit of the EGLS procedure is that it increases efficiency and avoids inference problems characterizing OLS in the presence of heteroskedasticity.
(e) Use mixed estimation. In this case there would be five additional, artificial observations, one for each of the five parameter estimates. An EGLS procedure is necessary, with the full estimated variance covariance matrix from the earlier results appearing as a block in the lower right of the error variance–covariance matrix for the artificial regression.

AC25 She is regressing a stationary variable on a nonstationary variable (Canadian GDP is growing and so is nonstationary) which does not have available another nonstationary variable with which it can cointegrate. By putting in the time trend this problem is solved – after removing the time trend, Canadian GDP must be stationary, allowing a sensible regression via this cointegration. Interpretation should take the form of “Newfoundland unemployment decreases whenever Canadian GDP exceeds its equilibrium as given by its time trend.”

AC27 The negative sign probably is due to sample selection – the less productive firms were quick to apply for the grants. One way of proceeding would be to take the difference between the firms’ productivity measures in the first and second years and use that as the dependent variable and the second year’s observations on grant, sales, and employees as the explanatory variables. If it is believed that the grant only increases productivity with a lag, the difference between the firms’ productivity measures in the first and third years could be used as the dependent variable and the third year’s observations on grant, sales, and employees as the explanatory variables. Even better here would be to exploit the panel nature of the data by using fixed or random effects estimation. In this case, the productivity levels would be retained as the dependent variable observations.

AC29 **(a)** The D67 coefficient is the change in the intercept from 1966 to 1967.
(b) The D67 coefficient is the difference between the 1967 intercept and the “base” common intercept from 1959 through 1964.
(c) Not equivalent. The first specification has the intercept increasing to 1968 and steady thereafter at the 1968 level; the second specification has the intercept dropping back to the 1959–1964 base level after 1968.

AC31 The difference between these two definitions is that the former is looking at long-run causality whereas the latter is looking at both long- and short-run causality. If $\beta1 = \beta2 = 0$ is rejected but $\beta1+\beta2 = 0$ accepted, there is short-run causality but no long-run causality. It is not obvious that one definition is more appropriate than another; it depends on the purpose of the analysis.

AC33 **(a)** The concern about spending double one’s income on food is misplaced. The squared income term has a negative coefficient, so that with appropriate measurement units for income the results could be reasonable.
(b) In its current form, the coefficient on NC provides directly a measure of the difference between the cost of feeding an adult versus a child. The proposed method will produce exactly the same measure by taking the difference between the coefficient on number of adults and the coefficient on NC. But calculating its variance will be more awkward, and

because it is the same measure, there is no gain in efficiency.

AC35 Should suggest using a logistic functional form because the dependent variable varies between zero and one. Should suggest using police expenditures per capita. If OLS is to be used should suggest a correction for heteroskedasticity because the dependent variable will be calculated using different county populations. Finally, need to worry about simultaneity, requiring IV estimation: higher crime may induce higher expenditure on police.

AC37 The long-run impact of x is smaller than its short-run impact.

AC39 The easiest way is to create the following dummies:
DD = 1 for observations on hours during the experiment, otherwise zero.
DE = 1 for observations on individuals in the experiment, otherwise zero.
Regress hours on an intercept, DD, DE, and DD × DE. The slope on DB × DE is the difference in differences measure.

AD1 **(a)** This program is producing 2000 t values that can be used to estimate the sampling distribution, under the null $\beta = 2$, of the t statistic for testing $\beta = 2$.
(b) The t value obtained from the actual data is 2.50; because it lies within the two-tailed 5% critical values of –2.634 and 2.717, we fail to reject the null at this significance level.

AD3 **(a)** The bias is estimated by $av - 2$.
(b) The square root of var estimates the standard error.
(c) The t statistic for testing the null is the estimated bias divided by the square root of its estimated variance, namely sqrt of var/4000.
(d) The confidence interval is given by the interval between the 200th and the 3800th r values, adjusted for any bias (i.e., if bias is estimated to be .01, then .01 should be subtracted from the 200th and 3800th r values.

AD5 The following describes how to bootstrap with specification A as the null; it must be repeated with specification B as the null to complete the J test.
Perform the J test with A as the null. This requires obtaining the predicted y from running regression B and then running regression A with this predicted y as an additional regressor. Call the t value associated with this additional regressor tA. Ordinarily tA would be compared to a t table to accept or reject specification A. The bootstrapping procedure creates new critical values tailored to this specific application of the J test.
Run regression A to obtain parameter estimates and residuals eA.
Using these parameter estimates and errors drawn with replacement from eA, create new observations for y.
Using these new y values, run regression B and obtain predicted y values. Continuing to use these new y values, run regression A with this predicted y as an additional regressor. Call the t value associated with this additional regressor $t1$.
Repeat from (3) above to obtain a thousand of these t values, $t1$–$t1000$.
Order these t values from smallest to largest and find the new 5% critical values, the 25th and the 975th of these t values. If tA falls inside (outside) this interval, accept (reject) specification A.

AD7 **(a)** A Hausman test. The null is that the fixed and random effects estimates are equal. Sometimes this is described in terms of the x variables being uncorrelated with the "random" intercept.
(b) Misleading means that the type I error is not what it is supposed to be (typically chosen to be 5%).
(c) The errors may not be normally distributed, or the sample size may be quite small.
(d) Begin by performing the Hausman test to obtain the test statistic $H0$. Then do the following:

1. Perform fixed effects estimation, saving the slope estimates, the intercept estimates ints and the residuals resids.
2. Using these slope estimates, and drawing with replacement from ints and from resids,

create new y values. (Drawing the intercepts randomly ensures that the intercepts are not correlated with the explanatory variables, so the null is true.)

3. Perform the Hausman test using these new y observations, obtaining test statistic $H1$.
4. Repeat from 1 to obtain a thousand H values, $H1$–$H1000$.
5. Order the H values from smallest to largest and see if $H0$ is smaller than $H950$, in which case the null is accepted at the 5% level.