Revision Notes

M300 'Econometric Methods'

Povilas Lastauskas*

8th May 2014

Contents

*Errors and typos to be reported to pl312@cam.ac.uk. For the most recent version, see www.lastauskas.com.

1 Maximum Likelihood

A random sample y_i, $i = (1, \ldots, n)$ is taken from the exponential distribution

$$f(y;\theta) = \frac{1}{\theta}\exp\left(-y/\theta\right).$$

i. What is the ML estimator of θ?
ii. Obtain the (scalar) information matrix
iii. Find the asymptotic distribution of the ML estimator of θ.

Further, consider a random sample $(y_i, \underline{x}_i')$, $i = (1, \ldots, n)$ that is taken from the joint distribution of $(y, \underline{x}')$. The conditional distribution of y given $\underline{x}'$ is the exponential distribution

$$f(y;\boldsymbol{\beta}) = \frac{1}{\theta\left(\underline{x}'\right)}\exp\left(-y/\theta\left(\underline{x}'\right)\right).$$

where $\theta\left(\underline{x}'\right) = \underline{x}_i'\boldsymbol{\beta}$.

(a) Show how this model for y given $\underline{x}'$ can be interpreted as a heteroskedastic regression model.

(b) Differentiate the conditional log-likelihood with respect to $\boldsymbol{\beta}$. Hence, show that the ML estimator of $\boldsymbol{\beta}$ can be computed by a weighted least squares recursion of the form

$$\hat{\boldsymbol{\beta}}_j = \left(\sum_{i=1}^{n}\frac{\underline{x}_i\underline{x}_i'}{(\underline{x}_i'\hat{\boldsymbol{\beta}}_{j-1})^2}\right)^{-1}\sum_{i=1}^{n}\frac{\underline{x}_i y_i}{(\underline{x}_i'\hat{\boldsymbol{\beta}}_{j-1})^2},\ j = 1, 2, \ldots.$$

(c) Find the asymptotic variance matrix of the ML estimator of $\boldsymbol{\beta}$.

1.1 Suggested solution.

From the probability density function of a single observation y_i

$$p(y_i;\theta) = \frac{1}{\theta}e^{-y_i/\theta}, \quad \text{for} \quad y_i > 0\,,$$

we can derive the likelihood function for a single observation

$$L(\theta; y_i) = \frac{1}{\theta}e^{-y_i/\theta},$$

and, assuming that observations are independently distributed, the likelihood function for the whole sample is

$$L\left(\theta; y_1, y_2, \ldots, y_n\right) = \prod_{i=1}^{n} L\left(\theta; y_i\right).$$

Taking logs gives

$$\begin{aligned}\log L\left(\theta; y_1, y_2, \ldots, y_n\right) &= \log\prod_{i=1}^{n} L\left(\theta; y_i\right) \\ &= \sum_{i=1}^{n}\log L\left(\theta; y_i\right) = \sum_{i=1}^{n}\left(\frac{-y_i}{\theta} - \log\theta\right) = -\frac{1}{\theta}\sum_{i=1}^{n} y_i - n\log\theta\,.\end{aligned}$$

To find the maximum likelihood estimator $\hat{\theta}_{ML}$ calculate the first derivative and set it equal to zero

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

This holds for $\theta = \bar{y}$. Now check that the second derivative is negative at that point,

$$\frac{\partial^2 \log L(\theta)}{\partial \theta^2} = -\frac{2}{\theta^3}\sum_{i=1}^{n} y_i + \frac{n}{\theta^2}$$

and

$$\left.\frac{\partial^2 \log L(\theta)}{\partial \theta^2}\right|_{\theta=\bar{y}} = -2n^3\left(\sum_{i=1}^{n} y_i\right)^{-2} + n^3\left(\sum_{i=1}^{n} y_i\right)^{-2} = -n^3\left(\sum_{i=1}^{n} y_i\right)^{-2} < 0.$$

Therefore, the function is maximised at

$$\hat{\theta}_{ML} = \frac{\sum_{i=1}^{n} y_i}{n} = \bar{y}.$$

By definition, the information matrix is given by

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

Note that the second derivative is calculated *at the true* (unknown) *value* of θ, not at the estimator. Use the fact that for an exponentially distributed random variable $E[y_i] = \theta$ to find

$$\boldsymbol{I}(\theta) = E\left[\frac{2}{\theta^3}\sum_{i=1}^{n} y_i - \frac{n}{\theta^2}\right] = \frac{2n}{\theta^2} - \frac{n}{\theta^2} = \frac{n}{\theta^2}.$$

The asymptotic distribution of the ML estimator is given in general by

$$\hat{\theta}_{ML} \sim AN\left(\theta,\, \boldsymbol{I}^{-1}(\theta)\right),$$

so, in this case we have $\hat{\theta}_{ML} \sim AN\left(\theta, \frac{\theta^2}{n}\right)$. Equivalently,

$$\sqrt{n}\left(\hat{\theta}_{ML} - \theta\right) \xrightarrow{d} N(0, \boldsymbol{I}_A^{-1}(\theta) \quad \text{where} \quad \boldsymbol{I}_A(\theta) = \lim_{n\to\infty}\frac{\boldsymbol{I}(\theta)}{n}.$$

In this case, $\sqrt{n}\left(\hat{\theta}_{ML} - \theta\right) \xrightarrow{d} N(0, \theta^2)$.

1.1.1 Vector-valued case

Now assume that θ depends on k explanatory variables (that we observe) through an unknown parameter vector $\boldsymbol{\beta}$,

$$\theta_i = \underline{\mathrm{x}}_i'\beta$$

and so $y_i \sim \exp(\underline{x}_i'\beta)$.

Furthermore, assume that the observations y_i can be described as the sum of the nonstochastic term $\underline{x}_i'\beta$ and a random disturbance u_i,

$$y_i = \underline{x}_i'\beta + u_i, \qquad \text{so} \qquad u_i = y_i - \underline{x}_i'\beta.$$

What is the distribution of u_i? We know the distribution of y_i, the $\underline{x}$'s are fixed and so we can derive it. The p.d.f. of u_i has the same shape as $p(y_i)$ but it is shifted to the left by $\underline{x}_i'\beta$. The moments are

$$\begin{aligned} E[u_i] &= E[y_i] - \underline{x}_i'\beta = \theta_i - \underline{x}_i'\beta = 0, \\ \operatorname{var}(u_i) &= \operatorname{var}(y_i) + 0 = \theta_i^2 = (\underline{x}_i'\beta)^2. \end{aligned}$$

This model can be interpreted as a heteroskedastic regression in which the disturbances are non-normal and the variances are unknown because they depend on β. The loglikelihood of β is

$$\log L(\boldsymbol{\beta}; \underline{y}) = -\sum_{i=1}^{n} \left(\frac{y_i}{\underline{x}_i'\boldsymbol{\beta}} + \log(\underline{x}_i'\boldsymbol{\beta}) \right).$$

Differentiation gives

$$\frac{\partial \log L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \left(\frac{y_i}{(\underline{x}_i'\boldsymbol{\beta})^2}\underline{x}_i - \frac{1}{\underline{x}_i'\boldsymbol{\beta}}\underline{x}_i \right) = \sum_{i=1}^{n} \frac{\underline{x}_i y_i}{(\underline{x}_i'\boldsymbol{\beta})^2} - \left(\sum_{i=1}^{n} \frac{\underline{x}_i\underline{x}_i'}{(\underline{x}_i'\boldsymbol{\beta})^2} \right) \boldsymbol{\beta}. \tag{1}$$

To find the ML estimator we should set the gradient (1) equal to zero, solve for $\boldsymbol{\beta}$ and check that the Hessian is negative definite. In this case, though, the equation is so complicated that an iterative procedure might be more convenient. Suppose the value $\boldsymbol{\beta}$ that maximises the likelihood is $\tilde{\boldsymbol{\beta}}$. The condition

$$\tilde{\boldsymbol{\beta}} = \left(\sum_{i=1}^{n} \frac{\underline{x}_i\underline{x}_i'}{(\underline{x}_i'\tilde{\boldsymbol{\beta}})^2} \right)^{-1} \sum_{i=1}^{n} \frac{\underline{x}_i y_i}{(\underline{x}_i'\tilde{\boldsymbol{\beta}})^2} \tag{2}$$

must hold. If we knew the value of $\underline{x}_i'\tilde{\boldsymbol{\beta}}$ it would be easy to calculate $\tilde{\boldsymbol{\beta}}$ using the last expression. What we can do is start from an inefficient estimate $\hat{\boldsymbol{\beta}}_0$ (we can take the OLS estimator which is still unbiased here) and apply (2) iteratively. This can be written as

$$\hat{\boldsymbol{\beta}}_{j+1} = \left(\sum_{i=1}^{n} \frac{\underline{x}_i\underline{x}_i'}{(\underline{x}_i'\hat{\boldsymbol{\beta}}_j)^2} \right)^{-1} \sum_{i=1}^{n} \frac{\underline{x}_i y_i}{(\underline{x}_i'\hat{\boldsymbol{\beta}}_j)^2}. \tag{3}$$

Each of the estimators $\hat{\beta}_{j+1}$ can be thought of as a Weighted Least Squares estimator with weights $(\underline{x}_i'\hat{\boldsymbol{\beta}}_j)^{-2}$. The name of this estimator comes from the fact that in general if we minimise *a weighted sum* of the squares of the residuals in a linear regression model

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{n} w_i (y_i - \underline{x}_i'\boldsymbol{\beta})^2$$

with respect to $\boldsymbol{\beta}$ we get $\hat{\boldsymbol{\beta}}_{WLS} = \left(\sum w_i \underset{\sim}{x}_i \underset{\sim}{x}_i'\right)^{-1} \sum w_i y_i \underset{\sim}{x}_i$. If the weights are all equal ($w_i = w$), $\hat{\boldsymbol{\beta}}_{WLS}$ is just the OLS estimator. In a heteroskedastic model where the disturbances are uncorrelated and the variances are known, the optimal weights are given by $w_i = \sigma_i^{-2}$ which gives the GLS estimator.

The Hessian of the log-likelihood function is

$$\frac{\partial^2 \log L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = \sum_{i=1}^{n} \left(\frac{-2y_i}{(\underset{\sim}{x}_i' \boldsymbol{\beta})^3} \underset{\sim}{x}_i \underset{\sim}{x}_i' + \frac{1}{(\underset{\sim}{x}_i' \boldsymbol{\beta})^2} \underset{\sim}{x}_i \underset{\sim}{x}_i' \right) = \sum_{i=1}^{n} \frac{-2y_i + \underset{\sim}{x}_i' \boldsymbol{\beta}}{(\underset{\sim}{x}_i' \boldsymbol{\beta})^3} \underset{\sim}{x}_i \underset{\sim}{x}_i',$$

and so, using again the fact that if y_i is exponentially distributed then $E[y_i] = \theta_i = \underset{\sim}{x}_i' \boldsymbol{\beta}$,

$$\boldsymbol{I}(\boldsymbol{\beta}) = E\left[-\frac{\partial^2 \log L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \right] = \sum_{i=1}^{n} \frac{2\underset{\sim}{x}_i' \boldsymbol{\beta} - \underset{\sim}{x}_i' \boldsymbol{\beta}}{(\underset{\sim}{x}_i' \boldsymbol{\beta})^3} \underset{\sim}{x}_i \underset{\sim}{x}_i' = \sum_{i=1}^{n} \left(\frac{1}{\underset{\sim}{x}_i' \boldsymbol{\beta}} \right)^2 \underset{\sim}{x}_i \underset{\sim}{x}_i'.$$

This gives $\mathrm{Avar}(\hat{\boldsymbol{\beta}}_{ML}) = \boldsymbol{I}^{-1}(\boldsymbol{\beta})$.

2 Probit

Consider the following probit regression model

$$\Pr(Y_i = 1 \mid X_i,\, Z_i) = \Phi(\beta_1 + \beta_2 X_i + \beta_3 Z_i),$$

where Y_i is a binary variable, assuming value 1 if i satisfies a specified condition, X_i is a continuous variable, and Z_i is a dummy variable. You are given the following numerical results

$$\hat{\beta}^{ML} = \begin{pmatrix} -2.2587 \\ 2.7416 \\ 0.7082 \end{pmatrix}, \ \mathrm{Var}\left(\hat{\beta}^{ML}\right) = \begin{pmatrix} 0.0169 & -0.0445 & -0.0010 \\ -0.0445 & 0.1293 & -0.0018 \\ -0.0010 & -0.0018 & 0.0070 \end{pmatrix}.$$

(a) Estimate the marginal effect of a change in X_i on the probability of $Y_i = 1$ for a case when $Z_i = 1$ with $X_i = 0.3$.

(b) Compute the standard error of your estimate in (a) using the Delta method.

(c) Estimate the effect of the change from $Z_i = 1$ to $Z_i = 0$ on the probability of $Y_i = 1$ given that $X_i = 0.3$.

2.1 Suggested solution.

(a) In the probit model

$$\Pr(Y_i = 1 \mid X_i,\, Z_i) = \Phi(\beta_1 + \beta_2 X_i + \beta_3 Z_i),$$

the marginal effect of X_i on the probability of $Y_i = 1$ equals

$$ME_i = \phi(\beta_1 + \beta_2 X_i + \beta_3 Z_i)\, \beta_2.$$

For the case $Z_i = 1$ and $X_i = 0.3$, this effect would be

$$ME_i = \phi\left(\beta_1 + \beta_2 \times 0.3 + \beta_3\right)\beta_2.$$

By the invariance property of the ML estimator,

$$\hat{ME}_i^{ML} = \phi\left(\beta_1^{ML} + \beta_2^{ML} 0.3 + \beta_3^{ML}\right)\beta_2^{ML}.$$

Since $\hat{\beta}^{ML} = \begin{pmatrix} \text{-2.2587} \\ 2.7416 \\ 0.7082 \end{pmatrix}$, we have that

$$\hat{ME}_i^{ML} = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\text{-2.2587} + 2.7416 \times 0.3 + 0.7082\right)^2\right\} 2.7416 = 0.839.$$

(b) The first order Taylor approximation of ME viewed as a non-linear function of β_1, β_2 and β_3 is

$$\hat{ME}_i^{ML} \approx ME + \begin{pmatrix} \frac{\partial ME}{\partial \beta_1} & \frac{\partial ME}{\partial \beta_2} & \frac{\partial ME}{\partial \beta_3} \end{pmatrix} \begin{pmatrix} \hat{\beta}_1^{ML} - \beta_1 \\ \hat{\beta}_2^{ML} - \beta_2 \\ \hat{\beta}_3^{ML} - \beta_3 \end{pmatrix},$$

so, according to the Delta method,

$$\text{Var}\left(\hat{ME}_i^{ML}\right) = \begin{pmatrix} \frac{\partial ME}{\partial \beta_1} & \frac{\partial ME}{\partial \beta_2} & \frac{\partial ME}{\partial \beta_3} \end{pmatrix} \text{Var}\left(\hat{\beta}^{ML}\right) \begin{pmatrix} \frac{\partial ME}{\partial \beta_1} \\ \frac{\partial ME}{\partial \beta_2} \\ \frac{\partial ME}{\partial \beta_3} \end{pmatrix}, \tag{4}$$

where the partial derivatives are evaluated at $\beta = \hat{\beta}^{ML}$. Therefore,

$$\begin{aligned} \frac{\partial ME}{\partial \beta_1} &= \phi'\left(\beta_1 + \beta_2 \times 0.3 + \beta_3\right)\beta_2, \\ \frac{\partial ME}{\partial \beta_2} &= 0.3\phi'\left(\beta_1 + \beta_2 \times 0.3 + \beta_3\right)\beta_2 + \phi\left(\beta_1 + \beta_2 \times 0.3 + \beta_3\right), \\ \frac{\partial ME}{\partial \beta_3} &= \phi'\left(\beta_1 + \beta_2 \times 0.3 + \beta_3\right)\beta_2. \end{aligned}$$

Recall that $\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$, so that $\phi'(x) = -\frac{x}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) = -x\phi(x)$. Hence,

$$\begin{aligned} \frac{\partial ME}{\partial \beta_1} &= -\phi\left(\beta_1 + \beta_2 \times 0.3 + \beta_3\right)\beta_2\left(\beta_1 + \beta_2 \times 0.3 + \beta_3\right), \\ \frac{\partial ME}{\partial \beta_2} &= \phi\left(\beta_1 + \beta_2 \times 0.3 + \beta_3\right)\left[-0.3\left(\beta_1 + \beta_2 \times 0.3 + \beta_3\right)\beta_2 + 1\right], \\ \frac{\partial ME}{\partial \beta_3} &= -\phi\left(\beta_1 + \beta_2 \times 0.3 + \beta_3\right)\beta_2\left(\beta_1 + \beta_2 \times 0.3 + \beta_3\right). \end{aligned}$$

Evaluating these expressions at $\beta = \hat{\beta}^{ML}$, we get

$$\begin{aligned} \left.\frac{\partial ME}{\partial \beta_1}\right|_{\beta = \hat{\beta}^{ML}} &= -0.6109, \\ \left.\frac{\partial ME}{\partial \beta_2}\right|_{\beta = \hat{\beta}^{ML}} &= 0.4893, \\ \left.\frac{\partial ME}{\partial \beta_3}\right|_{\beta = \hat{\beta}^{ML}} &= -0.6109. \end{aligned}$$

Substituting these values and the value of $\text{Var}\left(\hat{\beta}^{ML}\right)$ into (4), we get

$$\text{Var}\left(M\hat{E}_i^{ML}\right) = \begin{pmatrix} -0.6109 & 0.4893 & -0.6109 \end{pmatrix} \begin{pmatrix} 0.0169 & \text{-}0.0445 & \text{-}0.0010 \\ \text{-}0.0445 & 0.1293 & \text{-}0.0018 \\ \text{-}0.0010 & \text{-}0.0018 & 0.0070 \end{pmatrix} \begin{pmatrix} -0.6109 \\ 0.4893 \\ -0.6109 \end{pmatrix}$$
$$= 0.0668.$$

So the standard error of $M\hat{E}_i^{ML}$ is $\sqrt{\text{Var}\left(M\hat{E}_i^{ML}\right)} = \sqrt{0.0668} = 0.2585.$

(c) The effect of the change from $Z_i = 1$ to $Z_i = 0$ on the probability of $Y_i = 1$ is

$$\Phi\left(\beta_1 + \beta_2 X_i + \beta_3 \times 0\right) - \Phi\left(\beta_1 + \beta_2 X_i + \beta_3 \times 1\right).$$

Provided $X_i = 0.3$, this effect is equal to $\Phi\left(\beta_1 + \beta_2 \times 0.3\right) - \Phi\left(\beta_1 + \beta_2 \times 0.3 + \beta_3 \times 1\right).$ By the invariance property of the ML estimator,

$$\Phi\left(\beta_1^{ML} + \beta_2^{ML} X_i\right) - \Phi\left(\beta_1^{ML} + \beta_2^{ML} X_i + \beta_3^{ML} \times 1\right).$$

Since $\hat{\beta}^{ML} = \begin{pmatrix} \text{-}2.2587 \\ 2.7416 \\ 0.7082 \end{pmatrix}$, we have that

$$\Phi\left(\text{-}2.2587 + 2.7416 \times 0.3\right) - \Phi\left(\text{-}2.2587 + 2.7416 \times 0.3 + 0.7082\right)$$
$$= 0.0755 - 0.2333 = -0.1578.$$

3 LR vs LM tests

Consider a linear regression

$$Y = X_1\beta_1 + X_2\beta_2 + u,$$

where $u \mid X_1,\, X_2 \sim N\left(0,\, \sigma^2 I_n\right)$. We would like to test $H_0 :\ \beta_1 = 0$ against $H_1 :\ \beta_1 \neq 0$. Suppose X_1 is $n \times k_1$ and X_2 is $n \times k_2$, and let $\hat{u}$ and $\tilde{u}$ be the $n \times 1$ vectors of residuals from the OLS regression of Y on X_1 and X_2, and Y on X_2, respectively.

(a) Show that the likelihood ratio statistic LR equals

$$n \log \frac{\tilde{u}'\tilde{u}}{\hat{u}'\hat{u}}.$$

(b) Show that the Lagrange multiplier statistic LM equals

$$n \frac{\tilde{u}' X_1 \left(X_1' M_{X_2} X_1\right)^{-1} X_1' \tilde{u}}{\tilde{u}'\tilde{u}}.$$

(c) Using the result from (b) or otherwise, show that

$$LM = n \frac{\tilde{u}'\tilde{u} - \hat{u}'\hat{u}}{\tilde{u}'\tilde{u}}.$$

3.1 Suggested solution.

(a) Let $\hat{\beta}$ and $\hat{\sigma}^2$ be the unrestricted ML estimates of $\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$ and σ^2. Under normality, they are $\hat{\beta} = \hat{\beta}_{OLS}$ and $\hat{\sigma}^2 = \hat{u}'\hat{u}/n$. To derive an explicit form of the LR statistic, note that

$$\begin{aligned} \log L\left(\hat{\beta}, \hat{\sigma}^2, Y, X\right) &= -\tfrac{n}{2}\log(2\pi) - \tfrac{n}{2}\log\hat{\sigma}^2 - \tfrac{1}{2\hat{\sigma}^2}\hat{u}'\hat{u} + \log L(X) \\ &= -\tfrac{n}{2}\log(2\pi) - \tfrac{n}{2}\log\tfrac{\hat{u}'\hat{u}}{n} - \tfrac{n}{2} + \log L(X). \end{aligned}$$

Similarly, denoting the restricted ML estimates of $\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$ and σ^2 as $\tilde{\beta}$ and $\tilde{\sigma}^2$, we get

$$\log L\left(\tilde{\beta}, \tilde{\sigma}^2, Y, X\right) = -\tfrac{n}{2}\log(2\pi) - \tfrac{n}{2}\log\tfrac{\tilde{u}'\tilde{u}}{n} - \tfrac{n}{2} + \log L(X).$$

Therefore,

$$LR = 2\left(-\frac{n}{2}\log\frac{\hat{u}'\hat{u}}{n} + \frac{n}{2}\log\frac{\tilde{u}'\tilde{u}}{n}\right) = n\log\frac{\tilde{u}'\tilde{u}}{\hat{u}'\hat{u}}.$$

(b) Recall that $\left(\hat{I}_1^{11}\left(\hat{\theta}_{ML}\right)\right)^{-1} = n\hat{\sigma}^2\left(X_1'M_{X_2}X_1\right)^{-1}$ (this is the conditional variance of $\hat{\beta}_1$ in the unrestricted regression, multiplied by the number of observations n).[1] Similarly, $\left(\hat{I}_1^{11}\left(\tilde{\theta}_{ML}\right)\right)^{-1} = n\tilde{\sigma}^2\left(X_1'M_{X_2}X_1\right)^{-1}$ and therefore

$$\begin{aligned} LM &= \tfrac{1}{n}\tfrac{d}{d\theta_1}\log L\left(\tilde{\theta}_{ML}\right)'\left(\hat{I}_1^{11}\left(\tilde{\theta}_{ML}\right)\right)^{-1}\tfrac{d}{d\theta_1}\log L\left(\tilde{\theta}_{ML}\right) \\ &= \tfrac{1}{n}\tfrac{\tilde{u}'X_1}{\tilde{\sigma}^2}n\tilde{\sigma}^2\left(X_1'M_{X_2}X_1\right)^{-1}\tfrac{X_1'\tilde{u}}{\tilde{\sigma}^2} = \tfrac{\tilde{u}'X_1\left(X_1'M_{X_2}X_1\right)^{-1}X_1'\tilde{u}}{\tilde{u}'\tilde{u}/n}. \end{aligned}$$

(c) First notice that $\tilde{u} = M_{X_2}\tilde{u}$. Therefore, $\left(X_1'M_{X_2}X_1\right)^{-1}X_1'M_{X_2}\tilde{u}$ which is $\hat{\beta}_1$. We have

$$\begin{aligned} LM &= \tfrac{\tilde{u}'X_1\left(X_1'M_{X_2}X_1\right)^{-1}X_1'\tilde{u}}{\tilde{u}'\tilde{u}/n} = \tfrac{\tilde{u}'X_1\hat{\beta}_1}{\tilde{u}'\tilde{u}/n} \\ &= \tfrac{\tilde{u}'\left(X_1\hat{\beta}_1 + X_2\left(\tilde{\beta}_2 - \hat{\beta}_2\right)\right)}{\tilde{u}'\tilde{u}/n}. \end{aligned}$$

The latter equality holds because $\tilde{u}$ is orthogonal to X_2. Further, $X_1\hat{\beta}_1 + X_2\left(\tilde{\beta}_2 - \hat{\beta}_2\right) = \tilde{u} - \hat{u}$. Hence, to finalise

$$LM = \frac{\tilde{u}'\left(\tilde{u} - \hat{u}\right)}{\tilde{u}'\tilde{u}/n} = \frac{\tilde{u}'\tilde{u} - \hat{u}'\hat{u}}{\tilde{u}'\tilde{u}/n}.$$

---

[1] Recall the residual maker and projection matrices: residuals $= \left(I - X\left(X'X\right)^{-1}X'\right)y = (I - P_X)y = M_X y$. Therefore, $y = P_X y + M_X y = (P_X + M_X)y$. Note that both P_X and M_X are symmetric and idempotent, ($P_X P_X = P_X$ and $M_X M_X = M_X$). Therefore, $\left(M_{X_2}X_1\right)'M_{X_2}X_1 = X_1'M_{X_2}X_1$ and $M_{X_2} = I - P_{X_2}$. Hence, $\hat{\beta}_1$ is the result of regressing y on $M_{X_2}X_1$, $\hat{\beta}_1 = \left(X_1'M_{X_2}'M_{X_2}X_1\right)^{-1}X_1'M_{X_2}'y$ where $M_{X_2}X_1$ is the matrix of residuals from the regression of X_1 on X_2 (therefore, we are regressing two sets of residuals). Observe further that replacing y by $M_{X_2}y$ leads to exactly the same result.

4 2SLS

The simultaneous equation

$$y_1 = -\beta_{12}y_2 - \beta_{13}y_3 - \gamma_{11}x_1 + u_1$$

is part of a three equation model which contains three other exogenous variables x_2, x_3 and x_4. Observations give the following cross-product matrices

$$\boldsymbol{Y}'\boldsymbol{Y} = \begin{pmatrix} 20 & 15 & -5 \\ 15 & 60 & -45 \\ -5 & -45 & 70 \end{pmatrix}, \boldsymbol{Y}'\boldsymbol{X} = \begin{pmatrix} 2 & 2 & 4 & 5 \\ 0 & 4 & 12 & -5 \\ 0 & -2 & -12 & 10 \end{pmatrix}, \boldsymbol{X}'\boldsymbol{X} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 5 \end{pmatrix}.$$

(a) Obtain 2SLS estimates of the parameters of this equation.
(b) How would you construct standard errors for 2SLS estimators?

4.1 Suggested solution.

The equation

$$y_1 = -\beta_{12}y_2 - \beta_{13}y_3 - \gamma_{11}x_1 + u_1, \tag{5}$$

should not be estimated via OLS due to the presence of endogenous variables (y_2 and y_3 are correlated with u_1) in the RHS. Since we have information on 4 exogenous variables and we only care about estimating the parameters in (5), the 2SLS procedure comes on handy.

In the first stage of the 2SLS method we capture exogenous variation in y_2 and y_3 by regressing them on the full set of available exogenous variables. In the second stage, we use the fitted values from the first stage as regressors. 2SLS can be regarded in many senses as an instrumental variable method, where the x's are the instruments for the RHS y's.

Let $\underset{\sim}{\mathrm{y}}_i$ $(i = 1, 2, 3)$ be the $T \times 1$ stacked vector of the observations of the endogenous variable y_{it} and define $\underset{\sim}{\mathrm{x}}_j$ $(j = 1, 2, 3, 4)$ similarly. For the *first step* it is also useful to define the $T \times 2$ matrix $\underset{\sim}{\mathrm{Y}}_1 = [\,\underset{\sim}{\mathrm{y}}_2\ \underset{\sim}{\mathrm{y}}_3\,]$ that contains the endogenous variables in (5) and $\underset{\sim}{\mathrm{X}} = [\,\underset{\sim}{\mathrm{x}}_1\ \underset{\sim}{\mathrm{x}}_2\ \underset{\sim}{\mathrm{x}}_3\ \underset{\sim}{\mathrm{x}}_4\,]$ as the $T \times 4$ matrix of exogenous regressors. We are to find the fitted values of the regressions of $\underset{\sim}{\mathrm{y}}_2$ and $\underset{\sim}{\mathrm{y}}_3$ on $\underset{\sim}{\mathrm{X}}$, and we can do this directly with a multivariate regression. The fitted $T \times 2$ matrix is

$$\widehat{\underset{\sim}{\mathrm{Y}}}_1 = \underset{\sim}{\mathrm{X}}\widehat{\underset{\sim}{\mathrm{B}}} = \underset{\sim}{\mathrm{X}}(\underset{\sim}{\mathrm{X}}'\underset{\sim}{\mathrm{X}})^{-1}\underset{\sim}{\mathrm{X}}'\underset{\sim}{\mathrm{Y}}_1\,. \tag{6}$$

In the *second stage* we regress $\underset{\sim}{\mathrm{y}}_1$ on the $T \times 3$ matrix $\underset{\sim}{\mathrm{Z}} = [\,\widehat{\underset{\sim}{\mathrm{Y}}}_1\ \underset{\sim}{\mathrm{x}}_1\,]$. The estimated vector of coefficients is

$$-\begin{bmatrix} \widehat{\beta}_{12} \\ \widehat{\beta}_{13} \\ \widehat{\gamma}_{11} \end{bmatrix} = (\underset{\sim}{\mathrm{Z}}'\underset{\sim}{\mathrm{Z}})^{-1}\,\underset{\sim}{\mathrm{Z}}'\underset{\sim}{\mathrm{y}}_1 = \begin{bmatrix} \widehat{\underset{\sim}{\mathrm{Y}}}_1'\widehat{\underset{\sim}{\mathrm{Y}}}_1 & \widehat{\underset{\sim}{\mathrm{Y}}}_1'\underset{\sim}{\mathrm{x}}_1 \\ \underset{\sim}{\mathrm{x}}_1'\widehat{\underset{\sim}{\mathrm{Y}}}_1 & \underset{\sim}{\mathrm{x}}_1'\underset{\sim}{\mathrm{x}}_1 \end{bmatrix}^{-1} \begin{bmatrix} \widehat{\underset{\sim}{\mathrm{Y}}}_1'\underset{\sim}{\mathrm{y}}_1 \\ \underset{\sim}{\mathrm{x}}_1'\underset{\sim}{\mathrm{y}}_1 \end{bmatrix}. \tag{7}$$

Note from (7) that:

- $\widehat{\underline{\mathrm{Y}}}_1'\widehat{\underline{\mathrm{Y}}}_1 = \underline{\mathrm{Y}}_1'\underline{\mathrm{X}}(\underline{\mathrm{X}}'\underline{\mathrm{X}})^{-1}\underline{\mathrm{X}}'\underline{\mathrm{X}}(\underline{\mathrm{X}}'\underline{\mathrm{X}})^{-1}\underline{\mathrm{X}}'\underline{\mathrm{Y}}_1 = \underline{\mathrm{Y}}_1'\underline{\mathrm{X}}(\underline{\mathrm{X}}'\underline{\mathrm{X}})^{-1}\underline{\mathrm{X}}'\underline{\mathrm{Y}}_1 = (\underline{\mathrm{X}}'\underline{\mathrm{Y}}_1)'(\underline{\mathrm{X}}'\underline{\mathrm{X}})^{-1}(\underline{\mathrm{X}}'\underline{\mathrm{Y}}_1)$
- $\widehat{\underline{\mathrm{Y}}}_1'\underline{\mathrm{y}}_1 = \underline{\mathrm{Y}}_1'\underline{\mathrm{X}}(\underline{\mathrm{X}}'\underline{\mathrm{X}})^{-1}\underline{\mathrm{X}}'\underline{\mathrm{y}}_1 = (\underline{\mathrm{X}}'\underline{\mathrm{Y}}_1)'(\underline{\mathrm{X}}'\underline{\mathrm{X}})^{-1}(\underline{\mathrm{X}}'\underline{\mathrm{y}}_1)$
- $\widehat{\underline{\mathrm{Y}}}_1'\underline{\mathrm{x}}_1 = \underline{\mathrm{Y}}_1'\underline{\mathrm{X}}(\underline{\mathrm{X}}'\underline{\mathrm{X}})^{-1}\underline{\mathrm{X}}'\underline{\mathrm{x}}_1 = \underline{\mathrm{Y}}_1'\underline{\mathrm{x}}_1$[2]

Now, from the information provided we have that

$$\underline{\mathrm{x}}_1'\underline{\mathrm{Y}}_1 = \begin{bmatrix} 0 & 0 \end{bmatrix} \qquad \underline{\mathrm{x}}_1'\underline{\mathrm{x}}_1 = 1 \qquad \underline{\mathrm{x}}_1'\underline{\mathrm{y}}_1 = 2$$

$$\underline{\mathrm{y}}_1'\underline{\mathrm{X}} = \begin{bmatrix} 2 & 2 & 4 & 5 \end{bmatrix} \qquad \underline{\mathrm{Y}}_1'\underline{\mathrm{X}} = \begin{bmatrix} 0 & 4 & 12 & -5 \\ 0 & -2 & -12 & 10 \end{bmatrix} \qquad (\underline{\mathrm{X}}'\underline{\mathrm{X}})^{-1} = \mathrm{diag}\left(1, \tfrac{1}{2}, \tfrac{1}{4}, \tfrac{1}{5}\right)$$

If follows that $\widehat{\underline{\mathrm{Y}}}_1'\widehat{\underline{\mathrm{Y}}}_1 = \begin{bmatrix} 49 & -50 \\ -50 & 58 \end{bmatrix}$ and $\widehat{\underline{\mathrm{Y}}}_1'\underline{\mathrm{y}}_1 = \begin{bmatrix} 11 \\ -4 \end{bmatrix}$. Plugging all these results in (7),

$$-\begin{bmatrix} \widehat{\beta}_{12} \\ \widehat{\beta}_{13} \\ \widehat{\gamma}_{11} \end{bmatrix} = \begin{bmatrix} 49 & -50 & 0 \\ -50 & 58 & 0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 11 \\ -4 \\ 2 \end{bmatrix} \cong \begin{bmatrix} 1.28 \\ 1.04 \\ 2 \end{bmatrix}. \qquad (8)$$

An estimator of the (asymptotic) covariance matrix of the estimated parameter vector is

$$\mathrm{Avar}\left(\begin{bmatrix} \widehat{\beta}_{12} \\ \widehat{\beta}_{13} \\ \widehat{\gamma}_{11} \end{bmatrix}\right) = s^2\left(\underline{\mathrm{Z}}'\underline{\mathrm{Z}}\right)^{-1} = s^2\begin{bmatrix} \underline{\mathrm{Y}}_1'\underline{\mathrm{X}}\left(\underline{\mathrm{X}}'\underline{\mathrm{X}}\right)^{-1}\underline{\mathrm{X}}'\underline{\mathrm{Y}}_1 & \underline{\mathrm{Y}}_1'\underline{\mathrm{x}}_1 \\ \underline{\mathrm{x}}_1'\underline{\mathrm{Y}}_1 & \underline{\mathrm{x}}_1'\underline{\mathrm{x}}_1 \end{bmatrix}^{-1}.$$

The important point to make here is that s^2 must be based on the residuals

$$\widehat{u} = y_1 + \widehat{\beta}_{12}y_2 + \widehat{\beta}_{13}y_3 + \widehat{\gamma}_{11}x_1.$$

If you compute the 2SLS estimator by carrying out two OLS estimations on any software package, you will see that the estimator s^2 printed by the program after the second stage is wrong. This happens because the PC does not know that our regression is the second stage of a two-stage procedure, so the residuals that are used to compute s^2 are

$$\widetilde{u} = y_1 + \widehat{\beta}_{12}\widehat{y}_2 + \widehat{\beta}_{13}\widehat{y}_3 + \widehat{\gamma}_{11}x_1,$$

as if the regressors in the second stage ($\widehat{y}_1$ and $\widehat{y}_2$) were the independent variables. This results in an inconsistent estimator.

[2]Note that $\underline{\mathrm{x}}_1$ is a column of $\underline{\mathrm{X}}$. Therefore, $\underline{\mathrm{X}}'\underline{\mathrm{x}}_1$ is a column of $\underline{\mathrm{X}}'\underline{\mathrm{X}}$. Now, $(\underline{\mathrm{X}}'\underline{\mathrm{X}})^{-1}\underline{\mathrm{X}}'\underline{\mathrm{X}} = \underline{\mathrm{I}}$, so $(\underline{\mathrm{X}}'\underline{\mathrm{X}})^{-1}\underline{\mathrm{X}}'\underline{\mathrm{x}}_1$ is just one column of $\underline{\mathrm{I}}$. When you multiply $\underline{\mathrm{X}}$ by this matrix you select a column of $\underline{\mathrm{X}}$, which is precisely $\underline{\mathrm{x}}_1$.

5 GMM

Consider a linear regression

$$y_i = x_i\beta + \varepsilon_i,$$

where

$$E(x_i\varepsilon_i) = E(z_i\varepsilon_i) = 0.$$

(a) Show that the asymptotic variance of the GMM estimator for the moment equation

$$E\begin{bmatrix} x_i(y_i - x_i\beta) \\ z_i(y_i - x_i\beta) \end{bmatrix} = 0$$

is equal to

$$\sigma^2\left(E\begin{bmatrix} x_i^2 & x_iz_i \end{bmatrix}\left(E\begin{bmatrix} x_i^2 & x_iz_i \\ x_iz_i & z_i^2 \end{bmatrix}\right)^{-1}E\begin{bmatrix} x_i^2 \\ x_iz_i \end{bmatrix}\right)^{-1}$$

if it happens that

$$E\begin{bmatrix} x_i^2\varepsilon_i^2 & x_iz_i\varepsilon_i^2 \\ x_iz_i\varepsilon_i^2 & z_i^2\varepsilon_i^2 \end{bmatrix} = \sigma^2 E\begin{bmatrix} x_i^2 & x_iz_i \\ x_iz_i & z_i^2 \end{bmatrix}.$$

(b) Show that

$$E\begin{bmatrix} x_i^2 & x_iz_i \end{bmatrix}\left(E\begin{bmatrix} x_i^2 & x_iz_i \\ x_iz_i & z_i^2 \end{bmatrix}\right)^{-1}E\begin{bmatrix} x_i^2 \\ x_iz_i \end{bmatrix} = E\left[x_i^2\right].$$

(c) Demonstrate that the asymptotic variance is equal to

$$\frac{\sigma^2}{E\left[x_i^2\right]}$$

and convince yourself that it is exactly the asymptotic variance of the OLS estimator.

5.1 Suggested solution.

GMM estimator solves the following problem.

$$\min_\beta \left(\frac{1}{n}\sum_{i=1}^n \begin{bmatrix} x_i(y_i - x_i\beta) \\ z_i(y_i - x_i\beta) \end{bmatrix}\right)' W \left(\frac{1}{n}\sum_{i=1}^n \begin{bmatrix} x_i(y_i - x_i\beta) \\ z_i(y_i - x_i\beta) \end{bmatrix}\right),$$

where W is the optimal weight matrix

$$W = \left[n\mathrm{Var}\left(\tfrac{1}{n}\textstyle\sum_{i=1}^n \begin{bmatrix} x_i(y_i - x_i\beta) \\ z_i(y_i - x_i\beta) \end{bmatrix}\right)\right]^{-1}$$
$$= \left[\tfrac{1}{n}\mathrm{Var}\left(\textstyle\sum_{i=1}^n \begin{bmatrix} x_i\varepsilon_i \\ z_i\varepsilon_i \end{bmatrix}\right)\right]^{-1} = \left[\mathrm{Var}\begin{pmatrix} x_i\varepsilon_i \\ z_i\varepsilon_i \end{pmatrix}\right]^{-1}$$
$$= \left[E\left[\begin{pmatrix} x_i\varepsilon_i \\ z_i\varepsilon_i \end{pmatrix}\begin{pmatrix} x_i\varepsilon_i & z_i\varepsilon_i \end{pmatrix}\right]\right]^{-1} = \left[E\begin{bmatrix} x_i^2\varepsilon_i^2 & x_iz_i\varepsilon_i^2 \\ x_iz_i\varepsilon_i^2 & z_i^2\varepsilon_i^2 \end{bmatrix}\right]^{-1} = \left[\sigma^2 E\begin{bmatrix} x_i^2 & x_iz_i \\ x_iz_i & z_i^2 \end{bmatrix}\right]^{-1}.$$

Returning to the FOC of the minimisation problem

$$2\left(\frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix} -x_i^2 \\ -z_i x_i \end{bmatrix}\right)' W \left(\frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix} x_i\left(y_i - x_i\beta\right) \\ z_i\left(y_i - x_i\beta\right) \end{bmatrix}\right) = 0.$$

Arranging the terms, the GMM estimator of β is

$$\hat{\beta}^{GMM} = \left(\left(\frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix} x_i^2 \\ z_i x_i \end{bmatrix}\right)' W \left(\frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix} x_i^2 \\ z_i x_i \end{bmatrix}\right)\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix} x_i^2 \\ z_i x_i \end{bmatrix}\right)' W \left(\frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix} x_i y_i \\ z_i y_i \end{bmatrix}\right).$$

Using $y_i = x_i\beta + \varepsilon_i$

$$\begin{aligned}\hat{\beta}^{GMM} &= \left(\left(\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\begin{bmatrix} x_i^2 \\ z_i x_i \end{bmatrix}\right)' W \left(\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\begin{bmatrix} x_i^2 \\ z_i x_i \end{bmatrix}\right)\right)^{-1} \\ &\times \left(\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\begin{bmatrix} x_i^2 \\ z_i x_i \end{bmatrix}\right)' W \left(\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\begin{bmatrix} x_i^2\beta \\ z_i x_i\beta \end{bmatrix} + \tfrac{1}{n}\textstyle\sum_{i=1}^{n}\begin{bmatrix} x_i\varepsilon_i \\ z_i\varepsilon_i \end{bmatrix}\right) = \beta \\ &+ \left(\left(\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\begin{bmatrix} x_i^2 \\ z_i x_i \end{bmatrix}\right)' W \left(\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\begin{bmatrix} x_i^2 \\ z_i x_i \end{bmatrix}\right)\right)^{-1} \left(\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\begin{bmatrix} x_i^2 \\ z_i x_i \end{bmatrix}\right)' W \left(\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\begin{bmatrix} x_i\varepsilon_i \\ z_i\varepsilon_i \end{bmatrix}\right),\end{aligned}$$

thereby

$$\sqrt{n}\left(\hat{\beta}^{GMM} - \beta\right) = \left(\left(\frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix} x_i^2 \\ z_i x_i \end{bmatrix}\right)' W \left(\frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix} x_i^2 \\ z_i x_i \end{bmatrix}\right)\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix} x_i^2 \\ z_i x_i \end{bmatrix}\right)' W \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{bmatrix} x_i\varepsilon_i \\ z_i\varepsilon_i \end{bmatrix}\right).$$

Therefore

$$\sqrt{n}\left(\hat{\beta}^{GMM} - \beta\right) \xrightarrow{d} N\left(0,\, V\right),$$

where $V = \left(D'WD\right)^{-1} D'W\mathrm{Var}\begin{pmatrix} x_i\varepsilon_i \\ z_i\varepsilon_i \end{pmatrix} WD\left(D'WD\right)^{-1} = \left(D'WD\right)^{-1} D'WW^{-1}WD\left(D'WD\right)^{-1} = \left(D'WD\right)^{-1}$ and $D \equiv E\begin{bmatrix} x_i^2 \\ z_i x_i \end{bmatrix}$.

Hence, the asymptotic variance is given by

$$\begin{aligned}V &= \left(D'WD\right)^{-1} \\ &= \left(E\begin{bmatrix} x_i^2 & z_i x_i \end{bmatrix}\left[\sigma^2 E\begin{bmatrix} x_i^2 & x_i z_i \\ x_i z_i & z_i^2 \end{bmatrix}\right]^{-1} E\begin{bmatrix} x_i^2 \\ z_i x_i \end{bmatrix}\right)^{-1} \\ &= \sigma^2\left(E\begin{bmatrix} x_i^2 & z_i x_i \end{bmatrix}\left[E\begin{bmatrix} x_i^2 & x_i z_i \\ x_i z_i & z_i^2 \end{bmatrix}\right]^{-1} E\begin{bmatrix} x_i^2 \\ z_i x_i \end{bmatrix}\right)^{-1}.\end{aligned}$$

Simple inversion yields

$$\left[E\begin{bmatrix} x_i^2 & x_i z_i \\ x_i z_i & z_i^2 \end{bmatrix}\right]^{-1} = \tfrac{1}{Ex_i^2 Ez_i^2 - \left(Ex_i z_i\right)^2}\begin{bmatrix} Ez_i^2 & -Ex_i z_i \\ -Ex_i z_i & Ex_i^2 \end{bmatrix}.$$

To conclude,

$$V=\sigma^2\left(\tfrac{1}{Ex_i^2Ez_i^2-(Ex_iz_i)^2}E\left[\begin{array}{cc} x_i^2 & z_ix_i\end{array}\right]\left[\begin{array}{cc} Ez_i^2 & -Ex_iz_i \\ -Ex_iz_i & Ex_i^2\end{array}\right]E\left[\begin{array}{c} x_i^2 \\ z_ix_i\end{array}\right]\right)^{-1}$$
$$=\sigma^2\left(\tfrac{1}{Ex_i^2Ez_i^2-(Ex_iz_i)^2}\left[\begin{array}{cc} Ex_i^2Ez_i^2-(Ex_iz_i)^2 & -Ex_i^2Ez_ix_i+Ex_i^2Ez_ix_i\end{array}\right]E\left[\begin{array}{c} x_i^2 \\ z_ix_i\end{array}\right]\right)^{-1}$$
$$=\sigma^2\left(\tfrac{1}{Ex_i^2Ez_i^2-(Ex_iz_i)^2}\left[\begin{array}{cc} Ex_i^2Ez_i^2-(Ex_iz_i)^2 & 0\end{array}\right]E\left[\begin{array}{c} x_i^2 \\ z_ix_i\end{array}\right]\right)^{-1}=\sigma^2\left(\tfrac{\left[Ex_i^2Ez_i^2-(Ex_iz_i)^2\right]Ex_i^2}{Ex_i^2Ez_i^2-(Ex_iz_i)^2}\right)^{-1}=\tfrac{\sigma^2}{Ex_i^2}.$$

It is easy to see that this is exactly the same as the the asymptotic variance of the OLS estimator. Since the OLS estimator is $\hat{\beta}^{OLS}=\left(\frac{1}{n}\sum_{i=1}^n x_i^2\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n x_iy_i\right)$, its asymptotic variance is

$$E\left(\hat{\beta}^{OLS}-\beta\right)\left(\hat{\beta}^{OLS}-\beta\right)'=\left(Ex_i^2\right)^{-1}\operatorname{Var}\left(x_i\varepsilon_i\right)\left(Ex_i^2\right)^{-1}$$
$$=\left(Ex_i^2\right)^{-1}\sigma^2\left(Ex_i^2\right)\left(Ex_i^2\right)^{-1}=\sigma^2\left(Ex_i^2\right)^{-1}.$$

6 Unit roots and cointegration

Consider the autoregressive distributed lag model

$$y_t=\alpha y_{t-1}+\beta_0x_t+\beta_1x_{t-1}+\varepsilon_t,\ 0<\alpha<1,\ (t=1,\ldots,T),$$

where x_t is exogenous $I(1)$ process and the error terms ε_t, $(t=1,\ldots,T)$, are distributed independently of each other with mean zero and variance σ^2.

(a) Suppose that $\beta_0\neq-\beta_1$. Is it possible that y_t is stationary? Explain.

(b) Is it possible that y_t is $I(2)$? Explain.

(c) Write down the error correction form of the model, showing the relationship between the new parameters and those in the original model. Argue that y_t and x_t are co-integrated. What is the co-integrating vector?

6.1 Suggested solution.

(a) First, let's re-write the model as

$$y_t=\alpha y_{t-1}+(\beta_0+\beta_1)x_t+\beta_1(x_{t-1}-x_t)+\varepsilon_t$$
$$=\alpha y_{t-1}+(\beta_0+\beta_1)x_t+\beta_1\triangle x_t+\varepsilon_t.$$

Since $\beta_0+\beta_1\neq0$, we have

$$x_t=\frac{y_t-\alpha y_{t-1}-\beta_1\triangle x_t-\varepsilon_t}{\beta_0+\beta_1}.$$

Had y_t been stationary, the right hand side of the above equality would have been stationary, whereas the left hand side is non-stationary. This is impossible. Thus, y_t cannot be stationary.

(b) If y_t is $I(2)$, Δy_t must be $I(1)$ and $\Delta^2 y_t$ must be $I(0)$. Applying the first difference operator to both sides of equality $y_t = \alpha y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t$, we get

$$\begin{gathered}\Delta y_t = \alpha \Delta y_{t-1} + \beta_0 \Delta x_t + \beta_1 \Delta x_{t-1} + \Delta \varepsilon_t \\ \Delta y_t - \Delta y_{t-1} = (\alpha - 1)\, \Delta y_{t-1} + \beta_0 \Delta x_t + \beta_1 \Delta x_{t-1} + \Delta \varepsilon_t \\ \Delta^2 y_t = (\alpha - 1)\, \Delta y_{t-1} + \beta_0 \Delta x_t + \beta_1 \Delta x_{t-1} + \Delta \varepsilon_t,\end{gathered}$$

and noting that $\alpha < 1$, we have

$$\Delta y_{t-1} = \frac{\Delta^2 y_t - \beta_0 \Delta x_t - \beta_1 \Delta x_{t-1} - \Delta \varepsilon_t}{\alpha - 1}.$$

Similar to the first part, we conclude that the right hand side is stationary, whereas the left hand side is non-stationary, which cannot be true. Hence, y_t cannot be $I(2)$. In fact, $\Delta y_t = \alpha \Delta y_{t-1} + \beta_0 \Delta x_t + \beta_1 \Delta x_{t-1} + \Delta \varepsilon_t$ implies that y_t is stationary so that y_t is $I(1)$.

(c) Lastly, rearranging $y_t = \alpha y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t$ into

$$\Delta y_t = (\alpha - 1)\, y_{t-1} + \beta_0 \triangle x_t + (\beta_0 + \beta_1)\, x_{t-1} + \varepsilon_t.$$

That is,

$$\Delta y_t = (\alpha - 1) \left(y_{t-1} - \frac{\beta_0 + \beta_1}{1 - \alpha} x_{t-1} \right) + \beta_0 \triangle x_t + \varepsilon_t,$$

which is the error correction form. Since y_t and x_t are $I(1)$, it must be the case that $y_{t-1} - \frac{\beta_0+\beta_1}{1-\alpha} x_{t-1}$ is $I(0)$. Hence, y_t and x_t are co-integrated and the cointegration vector is $\left[1, -\frac{\beta_0+\beta_1}{1-\alpha}\right]$.