Appendix A

Big-Oh and Little-Oh

The notation *Big-Oh* and *Little-Oh* is used to compare the relative values of two functions, $f(x)$ and $g(x)$, as x approaches ∞, or 0, depending on which of these two cases is being considered. We will suppose that g is positive-valued and that $x > 0$.

A.1 Big-Oh

The case $\mathbf{x} \to \infty$

f is $\mathcal{O}(g)$ if there exist constants $A > 0$ and $x_0 > 0$ such that $\dfrac{|f(x)|}{g(x)} < A$ for all $x > x_0$.

The case $\mathbf{x} \to \mathbf{0}$

f is $\mathcal{O}(g)$ if there exist constants $A > 0$ and $x_0 > 0$ such that $\dfrac{|f(x)|}{g(x)} < A$ for all $x < x_0$.

Example A1

The function $f(x) = 3x^3 + 4x^2$ is $\mathcal{O}(x^3)$ as $x \to \infty$. (Here $g(x) = x^3$.) We have that

$$\frac{|f(x)|}{g(x)} = \frac{3x^3 + 4x^2}{x^3} = 3 + \frac{4}{x} \, .$$

There exist (infinitely) many pairs $A, x_0 > 0$ that show that f is $\mathcal{O}(x^3)$, for example, $\frac{|f(x)|}{g(x)} < 4.1$ for all $x > 40$.

Example A2

The function $f(x) = 3x^3 + 4x^2$ is $\mathcal{O}(x^2)$ as $x \to 0$. Here

$$\frac{|f(x)|}{g(x)} = \frac{3x^3 + 4x^2}{x^2} = 3x + 4 \, ,$$

and, for example, $\frac{|f(x)|}{g(x)} < 4.3$ for all $x < 0.1$.

A.2 Little-Oh

The case $\mathbf{x \to \infty}$: f is $o(g)$ if $\displaystyle\lim_{x\to\infty} \frac{|f(x)|}{g(x)} = 0\,.$

The case $\mathbf{x \to 0}$: f is $o(g)$ if $\displaystyle\lim_{x\to 0} \frac{|f(x)|}{g(x)} = 0\,.$

Example A3

The function $f(x) = 3x^3 + 4x^2$ is $o(x^4)$ as $x \to \infty$ because

$$\lim_{x\to\infty} \frac{|f(x)|}{g(x)} = \lim_{x\to\infty} \frac{3x^3 + 4x^2}{x^4} = \lim_{x\to\infty} \left(\frac{3}{x^2} + \frac{4}{x} \right) = 0\,.$$

Example A4

The function $f(x) = 3x^3 + 4x^2$ is $o(x)$ as $x \to 0$ because

$$\lim_{x\to 0} \frac{|f(x)|}{g(x)} = \lim_{x\to 0} \frac{3x^3 + 4x^2}{x} = \lim_{x\to 0}(3x^2 + 4x) = 0\,.$$

Example A5

The function $f(x) = 3x^2 + 4x$ is $o(1)$ as $x \to 0$ because

$$\lim_{x\to 0} \frac{|f(x)|}{g(x)} = \lim_{x\to 0} \frac{3x^2 + 4x}{1} = \lim_{x\to 0}(3x^2 + 4x) = 0\,.$$

Appendix B

Taylor expansions

This appendix gives a brief justification of the expansions used in Section 1.3.2. Details can be found in standard calculus texts, for example, Courant R. and John F. (1965) *Introduction to Calculus and Analysis*, Wiley, New York.

Suppose that the function f has $n+1$ continuous derivatives in the interval $[a, a+h]$, if $h > 0$, or $[a+h, a]$, if $h < 0$. The n-term Taylor approximation for $f(a+h)$ is given by

$$f(a+h) = f(a) + \frac{h}{1!}f^{(1)}(a) + \frac{h^2}{2!}f^{(2)}(a) + \cdots + \frac{h^n}{n!}f^{(n)}(a) + R_n(a,h), \qquad \text{(B.1)}$$

where $f^{(r)}(a)$ denotes the r-th derivative of f at the point a, and $R_n(a,h)$, called the remainder, is the error in approximating $f(a+h)$ by the n-degree polynomial (in h) on the right-hand side.

The remainder can be represented in different forms. The following result is based on a version of the *Lagrange form*[1]: If there exists a positive constant M, such that $|f^{(n+1)}(t)| \leq M$ for all $t \in [a, a+h]$ (or alternatively $t \in [a+h, a]$, in the case $h < 0$) then

$$|R_n(a,h)| \leq \frac{|h|^{n+1}}{(n+1)!} M \, .$$

Thus, regarded as a function of n, for fixed a and h, the term $|R_n(a,h)|$ becomes small as n increases. Alternatively, regarded as a function of h, for fixed a and n, $|R_n(a,h)|$ becomes small as $h \to 0$. In the notation explained in Appendix A, $R_n(a,h)$ is $o(h^n)$ as $h \to 0$.

Example B1

This example relates to the material in Section 1.3.2. Specifically, we wish to investigate the behaviour of a two-term Taylor approximation to $f(x - zh)$ as h becomes small, for fixed values x and z. We assume that f is three times differentiable, and that, in some

[1] After the Italian-French mathematician Joseph-Louis Lagrange (1736-1813).

closed interval containing x, the absolute value of it's third derivative is bounded by some positive constant M. Applying the expansion (B.1) yields

$$f(x - hz) = f(x) + \frac{(-hz)}{1!}f'(x) + \frac{(-hz)^2}{2!}f''(x) + R_2(x, -zh),$$

where $|R_2(x, -zh)| \leq \frac{|zh|^3}{3!} M$. Thus as, h becomes small, we have that

$$f(x - hz) = f(x) - hzf'(x) + \frac{h^2 z^2}{2}f''(x) + o(h^2)$$

Similarly it follows that

$$f(x - hz) = f(x) - hzf'(x) + o(h)$$

and that

$$f(x - hz) = f(x) + o(1)$$

Appendix C

The Method of Weighted Least Squares

The purpose of this appendix is to derive a general formula for the estimator of the parameters of the linear model using the method of weighted least squares. We begin with a very simple case based on the method of ordinary least squares.

Ordinary least squares: A simple case

Considering the model

$$y_i = \theta + e_i \qquad i = 1, \ldots, n \ ,$$

where $\mathrm{E}(e_i) = 0$ and $\mathrm{Var}(e_i) = \sigma^2$. The least squares estimator of the parameter θ is the value which mimimizes the sum of squares of the residuals. A formula for the estimator can be derived by setting the derivative of the sum of residual squares equal to zero.

$$\begin{aligned}
\mathrm{RSS}(\theta) &= \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \theta)^2 \\
\frac{d\mathrm{RSS}(\theta)}{d\theta} &= -2\sum_{i=1}^{n} y_i + 2n\hat{\theta} \stackrel{!}{=} 0 \\
\hat{\theta} &= \frac{1}{n}\sum_{i=1}^{n} y_i
\end{aligned}$$

Weighted least squares: The simple case

The extension to weighted least squares is performed by defining weights w_1, w_2, …, w_n (given constants) and minimizing over the residual weighted sum of squares, RWSS, instead of the residual sum of squares, RSS:

$$\begin{aligned}
\text{RWSS}(\theta) &= \sum_{i=1}^{n} w_i e_i^2 = \sum_{i=1}^{n} w_i (y_i - \theta)^2 \\
\frac{d\text{RWSS}(\theta)}{d\theta} &= -2 \sum_{i=1}^{n} w_i (y_i - \hat{\theta}) \stackrel{!}{=} 0 \\
\hat{\theta} &= \frac{\sum_{i=1}^{n} w_i y_i}{\sum_{i=1}^{n} w_i}
\end{aligned}$$

Ordinary least squares: The general case

A more convenient way of estimating the parameters in linear regression (especially when dealing with more parameters) is using a matrix notation.

Consider the model

$$y_i = \theta_1 + \theta_2\, x_i + e_i \qquad i = 1, \ldots, n$$

This can be written in matrix form as follows:

$$y = X\theta + e$$

$$\text{with} \quad y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \; X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \; \theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \; e = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

The residual sum of squares can be minimized by setting the derivative with respect to θ equal to zero.

$$\begin{aligned}
\text{RSS}(\theta) &= e'e = (y - X\theta)'(y - X\theta) \\
&= y'y - y'X\theta - \theta'X'y + \theta'X'X\theta = y'y - 2\theta'X'y + \theta'X'X\theta \\
\frac{\partial \text{RSS}(\theta)}{\partial \theta} &= -2X'y + 2X'X\hat{\theta} \stackrel{!}{=} 0 \\
X'y &= X'X\hat{\theta}
\end{aligned}$$

The ordinary least squares estimator of θ is thus given by

$$\hat{\theta} = (X'X)^{-1}X'y \tag{C.1}$$

Note that the above derivation is also applicable to the case in which there are $p > 1$ covariates. Consider the model

$$y_i = \theta_1 + \theta_2\, x_{1\,i} + \theta_3\, x_{2\,i} \ldots \theta_{p+1}\, x_{p\,i} + e_i \qquad i = 1, \ldots, n$$

This model can also be written in matrix form $y = X\theta + e$, where:

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \; X = \begin{pmatrix} 1 & x_{1\,1} & x_{2\,1} & \ldots & x_{p\,1} \\ 1 & x_{2\,2} & x_{2\,2} & \ldots & x_{p\,2} \\ \vdots & \vdots & & \ldots & \\ 1 & x_{1\,n} & x_{2\,n} & \ldots & x_{p\,n} \end{pmatrix}, \; \theta = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_{p+1} \end{pmatrix}, \; e = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

Weighted least squares: The general case

The extension to weighted sums of squares is performed by defining a diagonal matrix W, whose diagonal elements comprise the weights w_i, $i = 1, 2, \ldots, n$:

$$W = \begin{pmatrix} w_1 & & & \mathbf{0} \\ & w_2 & & \\ & & \ddots & \\ \mathbf{0} & & & w_n \end{pmatrix}$$

The residual weighted sum of squares (RWSS) is given by $\sum_{i=1}^{n} w_i e_i^2 = e'We$. This is minimized by setting the derivative of RWSS with respect to θ equal to zero.

$$\begin{aligned} \text{RWSS}(\theta) &= (y - X\theta)'W(y - X\theta) \\ &= y'Wy - y'WX\theta - \theta'X'Wy + \theta'X'WX\theta = y'Wy - 2\theta'X'Wy + \theta'X'WX\theta \\ \frac{\partial \text{RWSS}(\theta)}{\partial \theta} &= -2X'Wy + 2X'WX\hat{\theta} \stackrel{!}{=} 0 \\ X'Wy &= X'WX\hat{\theta} \end{aligned}$$

$$\hat{\theta} = (X'WX)^{-1}X'Wy \tag{C.2}$$

The special case in which $w_i = 1$, $i = 1, 2, \ldots, n$ reduces to the ordinary least squares estimator (C.1).