Linear compositional regression

Josef Bukac
Bulharska 298, Jaromer-Josefov 55102
Czech Republic

Abstract: We study the properties of regression coefficients when the sum of the dependent variables is one, that is, the dependent variables are compositional. We show that the sum of intercepts is equal to one and the sum of other corresponding regression coefficients is zero. We do it for simple linear regressions and also for a more general case using matrix notation. The last part treats the case when the dependent variables do not sum up to one. We simplify the well known formula derived by the use of Lagrange multipliers.

Keywords: compositional data, equality, regression coefficients, restricted regression, simplified formula.

Introduction

Compositional data describe quantities that are parts related to some total or whole. They represent relative information, they have to be nonnegative, their sum has to be equal to one. We may point out that the conditions that are satisfied by compositional data may, Goldberger (1964, p.257), improve on efficiency of estimates in linear regression.

If we consider regression of dependent compositional variables y and z on some independent variable x, we use observations y_t and z_t at some point x_t assuming $y_t+z_t = 1$ for each $t = 1, 2, \ldots, T$, where T is the number of datapoints x_t, y_t, z_t. This is what we call a compositional regression.

Independent variables cannot contain compositional data. If there were more than one independent variables and their sum would be equal to one for each t, that is, such independent variables would be compositional, we could not use a regression model with an intercept because the design matrix would be singular. We don't call this a compositional regression.

Our assumption is satisfied in the case when y_t and z_t are relative frequencies, $y_t + z_t = 1$. It is not satisfied when y_t and z_t are realizations of random variables $\mathbf{Y}$ and $\mathbf{Z}$ and the expected value of their sum is $E(\mathbf{Y} + \mathbf{Z}) = 1$. In such a case we do not assume $y_t + z_t = 1$ but we still require that $a_y + b_yx + a_z + b_zx = 1$. In a case like this we may call it compositional on the average and we will use partitioned matrices and Lagrange multipliers. This may happen, for example, when we measure concentrations and their sum is not equal to one due to errors in measurements.

Simple linear compositional regression

If we want the sum of two functions $a_y + b_yx$ and $a_z + b_zx$ to be equal to one for all x, we write it as $a_y + a_z + b_yx + b_zx = 1$, where 1 stands for the constant function of x, we have to have $b_y + b_z = 0$, thus the equation is $a_y + a_z = 1$.

We start with simple linear regression of y on x and that of z on x. The models look like $y = a_y + b_y x + \epsilon_y$ and $z = a_z + b_z x + \epsilon_z$. When the least squares method is used, the formulas for b_y and b_z are well known. We want to calculate their sum. We assume there are observations on one independent variable $x_1, x_2, \ldots, x_T$ and two dependent variables with their sums being equal to one for each t. We discuss only two dependent variables y_t and z_t just for the sake of simplicity. If $y_t + z_t = 1$ is true for all $t = 1, 2, \ldots, T$, then so is $\bar{y} + \bar{z} = 1$ because

$$\bar{y} + \bar{z} = \frac{\sum y_t}{T} + \frac{\sum z_t}{T} = \frac{\sum (y_t + z_t)}{T} = \frac{T}{T} = 1.$$

We calculate the sum of the two slopes as

$$b_y + b_z = \frac{\sum (y_t - \bar{y})(x_t - \bar{x})}{\sum (x_t - \bar{x})^2} + \frac{\sum (z_t - \bar{z})(x_t - \bar{x})}{\sum (x_t - \bar{x})^2} =$$

$$\frac{\sum y_t (x_t - \bar{x})}{\sum (x_t - \bar{x})^2} - \frac{\sum \bar{y}(x_t - \bar{x})}{\sum (x_t - \bar{x})^2} + \frac{\sum z_t (x_t - \bar{x})}{\sum (x_t - \bar{x})^2} - \frac{\sum \bar{z}(x_t - \bar{x})}{\sum (x_t - \bar{x})^2} =$$

$$\frac{\sum (y_t + z_t)(x_t - \bar{x})}{\sum (x_t - \bar{x})^2} - \frac{\sum (\bar{y} + \bar{z})(x_t - \bar{x})}{\sum (x_t - \bar{x})^2} = \frac{\sum (x_t - \bar{x})}{\sum (x_t - \bar{x})^2} - \frac{\sum (\bar{y} + \bar{z})(x_t - \bar{x})}{\sum (x_t - \bar{x})^2}$$

We have discarded $y_t + z_t$ because it is equal to 1. We also use the fact $\bar{y} + \bar{z} = 1$ and obtain the required result

$$b_y + b_z = \frac{\sum (x_t - \bar{x})}{\sum (x_t - \bar{x})^2} - \frac{\sum (\bar{y} + \bar{z})(x_t - \bar{x})}{\sum (x_t - \bar{x})^2} = \frac{\sum (x_t - \bar{x})}{\sum (x_t - \bar{x})^2} - \frac{\sum (x_t - \bar{x})}{\sum (x_t - \bar{x})^2} = 0$$

To calculate the intercepts a_y and a_z we use the well known formulas $\bar{y} = a_y + b_y \bar{x}$ and $\bar{z} = a_z + b_z \bar{x}$. In our cases the sum of intercepts is

$$a_y + a_z = \bar{y} - b_y \bar{x} + \bar{z} - b_z \bar{x} = \bar{y} + \bar{z} - (b_y + b_z)\bar{x} = \bar{y} + \bar{z} = 1.$$

Interesting situations may occure. What if somebody says that he or she does not want to hear anything about compositional data and any stuff like that because simple linear regression will do for the purpose of calculations. Such a person is actually doing it right without knowing about it.

Another point is that observations of some dependent variables may be missing completely. If this is the situation, simple linear regression delivers the right result anyway.

Compositional regression with matrix notation

It is easy to see that the generalisation to more than two dependent variables is obvious. Now we want to generalize this result for more than one independent variable.

A set of T values of K independent variables is given, $x_{t1},\ x_{t2},\ \ldots,\ x_{tK},$ for $t = 1, 2, \ldots, T$. We assume $x_{t1} = 1$ to indicate the presence of the intercept. To be specific, we define the $T \times K$ design matrix $\mathbf{X}$ as

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1K} \\ \vdots & \vdots & \ddots & \vdots \\ x_{t1} & x_{t2} & \dots & x_{tK} \\ \vdots & \vdots & \ddots & \vdots \\ x_{T1} & x_{T2} & \dots & x_{TK} \end{pmatrix} = \begin{pmatrix} 1 & x_{12} & \dots & x_{1K} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{t2} & \dots & x_{tK} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{T2} & \dots & x_{TK} \end{pmatrix}.$$

Theorem. Let $\mathbf{X}$ be a matrix of independent variables with the first column consisting of ones. Let $\mathbf{y} = (\mathrm{y_1, y_2, \dots, y_t})'$ and $\mathbf{z} = (\mathrm{z_1, z_2, \dots, z_t})'$ satisfy $y_t + z_t = 1$ for $t = 1, 2, \dots, T$. Let $\mathbf{X'X}$ be nonsingular and let $\mathbf{b}^{\mathrm{y}}$, and $\mathbf{b}^{\mathrm{z}}$ be the solutions of equations $\mathbf{X'Xb}^{\mathrm{y}} = \mathbf{X'y}$ and $\mathbf{X'Xb}^{\mathrm{z}} = \mathbf{X'z}$ respectively. If $\mathbf{b} = \mathbf{b}^{\mathrm{y}} + \mathbf{b}^{\mathrm{z}}$, then $b_1 = 1$ and $b_k = 0$ for $k = 2, \dots, K$.

Proof. Write down the equations for $\mathbf{b}^{\mathrm{y}}$ and $\mathbf{b}^{\mathrm{z}}$ and form their sum $\mathbf{X'Xb}^{\mathrm{y}} + \mathbf{X'Xb}^{\mathrm{z}} = \mathbf{X'y} + \mathbf{X'z}$ which is $\mathbf{X'Xb} = \mathbf{X}'(\mathbf{y} + \mathbf{z})$. Since $\mathbf{y} + \mathbf{z} = \mathbf{e}$ where $\mathbf{e}' = (1, 1, \dots, 1)'$, the wright hand side is $\mathbf{X'e} = (\mathrm{T}, \sum_{\mathrm{t=1}}^{\mathrm{T}} \mathrm{x_{t2}}, \dots, \sum_{\mathrm{t=1}}^{\mathrm{T}} \mathrm{x_{tK}})'$.

If $b_1 = 1$ and $b_k = 0$ for $k = 2, \dots, K$, we check what the first column of $\mathbf{X'Xb}$ looks like and see that it is $(\mathrm{T}, \sum_{\mathrm{t=1}}^{\mathrm{T}} \mathrm{x_{t2}}, \dots, \sum_{\mathrm{t=1}}^{\mathrm{T}} \mathrm{x_{tK}})'$ which is the same as $\mathbf{X'e}$ if we multiply it by $b_1 = 1$. The remaining columns of the matrix $\mathbf{X'X}$ do not matter because they are multiplied by zeroes.

We see the equation $\mathbf{X'Xb} = \mathbf{X'e}$ is satisfied but the nonsingularity of $\mathbf{X'X}$ means that the solution is unique.

General linear compositional regression

Suppose we have T observations on each of M dependent variables $y_{t1}, \dots, y_{tM}$ for each $t = 1, \dots, T$ and K independent variables $x_{t1}, \dots, x_{tK}$. We define the $T \times M$ regressand matrix $\mathbf{Y}$ as

$$\mathbf{Y} = (\mathbf{y_1} | \dots | \mathbf{y_m} | \dots | \mathbf{y_M}).$$

This style of notation for sets of regressions is taken from Goldberger (1964, p. 201) and this is why we use this book as a reference.

To introduce the intercept, we assume $x_{t1} = 1$ for all $t = 1, \dots, T$ and minimize the sum of squares

$$\sum_{m=1}^{M} \sum_{t=1}^{T} \left(\sum_{k=1}^{K} x_{tk} b_{km} - y_{tm} \right)^2$$

with respect to b_{km}, $k = 1, 2, \dots, K$, $m = 1, 2, \dots, M$, subject to K constraints

$$\sum_{m=1}^{M} b_{1m} - 1 = 0, \quad \text{and} \quad \sum_{m=1}^{M} b_{km} = 0 \qquad \text{for} \quad k = 2, \dots, K.$$

A matrix notation is convenient when we deal with a set of dependent variables. We use the notation introduced in Goldberger (1964), p.201. A set of T

values of K independent variables is given, $x_{t1,}$ $x_{t2,}$ $\ldots$, $x_{tK,}$ for $t = 1, 2, \ldots, T$. We assume $x_{t1} = 1$ to indicate the presence of the intercept.

A set of T values of M dependent variables is given as y_{t1}, y_{t2}, $\ldots$, y_{tM} for $t = 1, 2, \ldots, T$. We define the regressand vectors as $\mathbf{y}_1 = (y_{11}, y_{21}, \ldots, y_{T1})'$, $\ldots$, $\mathbf{y}_m = (y_{1m}, y_{2m}, \ldots, y_{Tm})'$, $\ldots$, $\mathbf{y}_M = (y_{1M}, y_{2M}, \ldots, y_{TM})'$.

The regression coefficients of the classical unrestricted regression analysis are $\mathbf{a}_m = (a_{m1}, a_{m2}, \ldots, a_{mK})'$ where $\mathbf{a}_m = (\mathbf{X'X})^{-1}\mathbf{X'y}_m$. To derive this formula Goldberger (1964) uses matrix differentiation, Maindonald (1984, p.23) does not.

Since we want to impose certain restrictions on the regression coefficients, it is more convenient to define an equivalent joint regression model by using a partitioned matrix

$$\mathbf{X}_P = \begin{pmatrix} \mathbf{X} & | & \mathbf{0} & | & \ldots & | & \mathbf{0} \\ \mathbf{0} & | & \mathbf{X} & | & \ldots & | & \mathbf{0} \\ \ldots & \ldots & & \ddots & \ldots & & \\ \mathbf{0} & | & \mathbf{0} & | & \ldots & | & \mathbf{X} \end{pmatrix},$$

in which $\mathbf{X}$ occurs M times in $\mathbf{X}_{\mathrm{P}}$ which is of type $MT \times MK$. If the rank of $\mathbf{X}$ is K, the rank of $\mathbf{X}_P$ is MK. We also define a partitioned column vectors $\mathbf{y}_P = (\mathbf{y}_1'| \quad \mathbf{y}_2'| \quad \ldots \quad |\mathbf{y}_M')'$ and $\mathbf{a}_P = (\mathbf{a}_1'| \quad \mathbf{a}_2'| \quad \ldots \quad |\mathbf{a}_M')'$. The joint model is $\mathbf{y}_P = \mathbf{X}_P\mathbf{a}_P + \mathbf{e}_P$. We minimize the unrestricted sum of squares by calculating $\mathbf{a}_P = (\mathbf{X}_P'\mathbf{X}_P)^{-1}\mathbf{X}_P'\mathbf{y}_P$.

In this joint model we have a square matrix

$$\mathbf{X}_P'\mathbf{X}_P = \begin{pmatrix} \mathbf{X'X} & | & \mathbf{0} & | & \ldots & | & \mathbf{0} \\ \mathbf{0} & | & \mathbf{X'X} & | & \ldots & | & \mathbf{0} \\ \ldots & \ldots & & \ddots & \ldots & & \\ \mathbf{0} & | & \mathbf{0} & | & \ldots & | & \mathbf{X'X} \end{pmatrix},$$

the type of $\mathbf{X}_P'\mathbf{X}_P$ is $MK \times MK$ which makes it computationally feasible because it is sufficient to calculate $(\mathbf{X'X})^{-1}$ to get the inverse

$$(\mathbf{X}_P'\mathbf{X}_P)^{-1} = \begin{pmatrix} (\mathbf{X'X})^{-1} & | & \mathbf{0} & | & \ldots & | & \mathbf{0} \\ \mathbf{0} & | & (\mathbf{X'X})^{-1} & | & \ldots & | & \mathbf{0} \\ \ldots & \ldots & & \ddots & \ldots & & \\ \mathbf{0} & | & \mathbf{0} & | & \ldots & | & (\mathbf{X'X})^{-1} \end{pmatrix}.$$

We get the same result as we would obtain in M separate models.

The advantage of the joint model becomes clear when we impose restrictions on the model. The regression coefficients in the restricted model are denoted $\mathbf{b}_P = (\mathbf{b}_1'| \quad \mathbf{b}_2'| \quad \ldots \quad |\mathbf{b}_M')'$. We write linear restrictions on the coefficients in the form

$$\mathbf{R}_P\mathbf{b}_P = \mathbf{r}_P.$$

$\mathbf{R}_P$ is a partitioned matrix, $\mathbf{R}_P = \big(\mathbf{I}_K|\mathbf{I}_K|\ldots|\mathbf{I}_K\big)$ where $\mathbf{I}_{\mathrm{K}}$ is a $K \times K$ identity matrix. Thus $\mathbf{R}_P$ is a $K \times KM$ matrix of rank K. The right hand side is $\mathbf{r}_P =$

$(1, 0, \dots, 0)'$of type $1 \times M$ which is a way of saying that the sum of intercepts is $\sum_{m=1}^{M} b_{1m} = 1$ and that the meaning of zero components is $\sum_{m=1}^{M} b_{km} = 0$ for $k = 2, 3, \dots, K$.

The formula we can now use is well known, Goldberger (1964), p.256-7.

When $\mathbf{a}_P = (\mathbf{X}'_P\mathbf{X}_P)^{-1}\mathbf{X}'_P\mathbf{y}_P$, is obtained by the unconstrained classical least squares method, then the formula, when the restrictions are applied, is

$$\mathbf{b}_P = \mathbf{a}_P + (\mathbf{X}'_P\mathbf{X}_P)^{-1}\mathbf{R}'_P(\mathbf{R}_P(\mathbf{X}'_P\mathbf{X}_P)^{-1}\mathbf{R}'_P)^{-1}(\mathbf{r}_P - \mathbf{R}_P\mathbf{a}_P)$$

It is easy to verify that

$$(\mathbf{R}_P(\mathbf{X}'_P\mathbf{X}_P)^{-1}\mathbf{R}'_P)^{-1} = \mathbf{X}'\mathbf{X}/\mathrm{M}$$

and therefore

$$\mathbf{b}_P = \mathbf{a}_P + (\mathbf{X}'_P\mathbf{X}_P)^{-1}\mathbf{R}'_P(\mathbf{X}'\mathbf{X})(\mathbf{r}_P - \mathbf{R}_P\mathbf{a}_P)/\mathrm{M}$$

Since

$$(\mathbf{X}'_P\mathbf{X}_P)^{-1}\mathbf{R}'_{\mathrm{P}} = \begin{pmatrix} (\mathbf{X}'\mathbf{X})^{-1} & | & \mathbf{0} & | & \dots & | & \mathbf{0} \\ \mathbf{0} & | & (\mathbf{X}'\mathbf{X})^{-1} & | & \dots & | & \mathbf{0} \\ \dots & \dots & \ddots & \dots & & & \\ \mathbf{0} & | & \mathbf{0} & | & \dots & | & (\mathbf{X}'\mathbf{X})^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{I}_{\mathrm{K}} \\ \dots \\ \mathbf{I}_{\mathrm{K}} \end{pmatrix},$$

we have

$$(\mathbf{X}'_P\mathbf{X}_P)^{-1}\mathbf{R}_{\mathrm{P}} = \begin{pmatrix} (\mathbf{X}'\mathbf{X})^{-1} \\ \dots \\ (\mathbf{X}'\mathbf{X})^{-1} \end{pmatrix}$$

Thus

$$(\mathbf{X}'_P\mathbf{X}_P)^{-1}\mathbf{R}_{\mathrm{P}}\mathbf{X}'\mathbf{X} = \begin{pmatrix} (\mathbf{X}'\mathbf{X})^{-1} \\ \dots \\ (\mathbf{X}'\mathbf{X})^{-1} \end{pmatrix} \mathbf{X}'\mathbf{X} = \begin{pmatrix} \mathbf{I}_{\mathrm{K}} \\ \dots \\ \mathbf{I}_{\mathrm{K}} \end{pmatrix}$$

and

$$\mathbf{b}_{\mathrm{P}} = \mathbf{a}_{\mathrm{P}} + \begin{pmatrix} \mathbf{I}_{\mathrm{K}} \\ \dots \\ \mathbf{I}_{\mathrm{K}} \end{pmatrix} (\mathbf{r}_{\mathbf{P}} - \mathbf{R}_{\mathbf{P}}\mathbf{a}_{\mathbf{P}})/\mathrm{M} = \mathbf{a}_{\mathrm{P}} + \frac{1}{\mathrm{M}} \begin{pmatrix} \mathbf{r}_P - \mathbf{R}_P\mathbf{a}_P \\ \dots \\ \mathbf{r}_P - \mathbf{R}_P\mathbf{a}_P \end{pmatrix}$$

We can see that the difference of the restricted estimator $\mathbf{b}_{\mathrm{P}}$ from the unrestricted one $\mathbf{a}_{\mathrm{P}}$ is

$$\frac{1}{M} \begin{pmatrix} \mathbf{r}_P - \mathbf{R}_P\mathbf{a}_P \\ \dots \\ \mathbf{r}_P - \mathbf{R}_P\mathbf{a}_P \end{pmatrix},$$

that is, the difference $\mathbf{r}_P - \mathbf{R}_P\mathbf{a}_P$ is divided into M equal ammounts to correct the ordinary regression coefficients.

It is also worth mentioning that we got rid of calculating inverse matrices. It is certainly an advantage from the numerical point of view.

7. Numerical example

Our approach was originally designed to study concentrations of certain substances. Since such an example would not be obvious, we took the data about the cause of death in the US. The number of inhabitants may be different each year but the proportions may follow some other pattern.

http://www.cdc.gov/nchs/nvss/mortality/lcwk9.htm
is the address from which only a small part of data was used.

The most prominent causes of death are cardiovascular deseases (denoted as Heart) and cancer, other specific causes have frequency less than six percent. The data are presented in the table below. We indicate that we used the year minus 2000. Coefficients of the restricted linear regression follow.

```
  Original data                 Linear regression restricted or not
Year  Heart  Cancer  Other      Coeff Heart    Cancer    Other
1999 0.30327 0.22987 0.46686    a    0.29627  0.22894  0.47479
2000 0.29578 0.23009 0.47412    b  -0.005830 0.0002348 0.005595
2001 0.28969 0.22911 0.48120        y=a+b(Year-2000)
2002 0.28526 0.22802 0.48672    Overall sum of squares 7.038E-05
2003 0.27984 0.22746 0.49270
2004 0.27214 0.23099 0.49688
2005 0.26638 0.22848 0.50515
2006 0.26037 0.23075 0.50888
2007 0.25423 0.23221 0.51356
2008 0.24957 0.22878 0.52165
2009 0.24591 0.23293 0.52116
```

The coefficients are the same as the coefficients for unrestricted regressions. This is because we used the catch-all category Other.

8. What is it good for.

It is true that the equalities derived in this paper may be interesting. Especially the part when somebody does not like compositional approach but uses it without realizing it.

Compositional linear regression yields linear unbiased estimates with variances less than or equal to those unrestricted estimates. But there is a serious disadvantage of linear functions. They eventually take on values that are negative or greater than one when arguments are allowed to be arbitrarily large.

On the other hand, we need not worry about some observations equal to zero. Restricted regression does not care but we have to be carefull about the domain of definition not to obtain values that are negative or greater than one.

The main purpose of the linear compositional regression is to obtain the values of functions and their partial derivatives that will be used for interpolation by generalized logistic functions, Agresti (1990). The values of parameters we will obtain by interpolation will be used as starting values for iterative processes to find the minimal sum of squares numerically. Iterations will be difficult but the generalized logistic functions are intrinsically positive and less than zero which is their main advantage as compared with linear functions.

As we mentioned above the values of linear functions are eventually outside the interval (0,1). The arguments for which such values are within the interval [0,1] form a polytope. Only an interior point of a polytope may be used for interpolation by logistic functions. That is why the following has been included.

9. Interior point of the polytope

Fact. Let $p_1, \ldots, p_M$, where $M > 1$, be real numbers with $\sum_{m=1}^{M} p_m = 1$. Then $0 < p_m < 1$ for each $m = 1, \ldots, M$ if and only if $0 < p_m$ for each $m = 1, \ldots, M$.

Proof. Obviously $0 < p_m < 1$ implies $0 < p_m$. But if $0 < p_m$ for all m and

$$p_m = 1 - \sum_{i=1, i \neq m}^{M} p_i < 1.$$

It is good to write this fact down because in our case of restricted linear regression it will suffice to examine merely if $0 < p_m$ for each $m = 1, \ldots, M$. Now we repalce p_m by a function $f_m(\mathbf{x})$ of $\mathbf{x}$ where $\mathbf{x} = (\mathrm{x}_1, \ldots, \mathrm{x}_\mathrm{K})'$ and check if there is a domain of definition for which $0 < f_m(\mathbf{x})$ for all $m = 1, \ldots, M$. When we consider practical applications, we calculate the means of the independent variable data $\bar{x_k} = \sum_{t=1}^{T} x_{tk}$, for $t = 1, \ldots, T$, $\bar{\mathbf{x}} = (\bar{x_1}, \ldots, \bar{x_K})'$ and check if these means satisfy $0 < f_m(\bar{\mathbf{x}})$ for all $m = 1, \ldots, M$. Unfortunately there is no guarantee that the means will work.

To find the values of $\mathbf{x}$ that satisfy $0 < f_m(\mathbf{x})$ for all $m = 1, \ldots M$ we introduce a new variable y and examine $f_m(\mathbf{x}) - \mathbf{y} \geq \mathbf{0}$. in our case of restricted linear regression we have a linear programming model: Maximize y subject to $y \geq 0$ and $f_m(\mathbf{x}) - \mathrm{y} \geq 0$ for $m = 1, \ldots, M$, $\mathbf{x}$ unrestricted as to the signs. This task may be solved easily by the simplex method because the number of variables is small.

The only trouble with the simplex algorithm may arise due to rounding errors. That is why we prefer the use of this simple idea instead of other methods of finding an interior point or some optimal center of a polytope.

References

Agresti, A., 1990,*Categorical Data Analysis,* John Wiley and Sons, New York, p.313.

Aitchison, J., 1986, *The Statistical Analysis of Compositional Data,* Chapman and Hall; reprinted in 2003 by The Blackburn Press.

Bukac (2021). *Minimum with inequality constraint applied to increasing cubic, logistic and Gomperz or convex quartic and biexponential regressions.* Vixra-2021-12-02.

Goldberger, A.S., (1964), Econometric Theory, John Wiley and Sons, New York, London, Sydney.

Maindonald, J.H., (1984), Statistical Computation, John Wiley and Sons Inc, N.Y., etc.