

APPENDIX C

Miscellaneous Data Used in this Book

Contents

C.1 <i>Introduction</i>	119
C.2 <i>Key Matrices in PAMD Process Cycles</i>	119
C.3 <i>PCA of the States2007data.csv Dataset</i>	125
C.4 <i>Employment Data in 26 European Countries in 1979</i>	128

C.1 Introduction

The purpose of this appendix is to provide more information about data and datasets that were used as examples throughout this book. In section C.2, I provide more data related to the PAMD procedure for calculating principal components previously discussed in section 2.2.2. Section C.3 shows all 18 principal components associated with the 2007 States' expenditure dataset, whereas section C.4 presents a detailed description of 1979 employment data in 26 European countries.

C.2 Key Matrices in PAMD Process Cycles

This section contains more details about the calculation of principal components presented in section 2.2.2 and based on the iterative PAMD process of Figure 2.4. Presented in this section are key matrices calculated in each of the 6 cycles necessary to compute all 7 principal components (PC) associated with the correlation matrix Σ of Table 2.2.

Each of the 6 PAMD process cycles associated with the 7×7 correlation matrix Σ of Table 2.2 requires the calculation of 4 key matrices and 2 vectors. An arbitrary cycle t leads to the calculation of the following 4 matrices:

- The Householder matrix \mathbf{H}_t , the construction of which is discussed in details in section 2.2.3.
- Matrix \mathbf{H}_t^* . This matrix is used in each step (5) of the t^{th} PAMD cycle to translate the eigenvector \mathbf{v}_{t+1} of matrix \mathbf{K}_{t+1} into an eigenvector \mathbf{u}_{t+1} of the initial correlation matrix Σ .
- Matrix $\mathbf{G}_{t+1}^* = \mathbf{H}_t \mathbf{K}_t \mathbf{H}_t^\top$ is used in each step (2) of the t^{th} PAMD cycle to produce the deflated matrix \mathbf{K}_{t+1} .
- Matrix \mathbf{K}_{t+1}^* is the deflated matrix to which the power algorithm is applied to compute a new eigenvalue λ_{t+1} and the associated eigenvector \mathbf{v}_{t+1} .

The 2 vectors calculated during the t^{th} cycle of the PAMD process are the following:

- \mathbf{v}_{t+1} is the eigenvector of the $(p - t) \times (p - t)$ matrix \mathbf{K}_{t+1} .
- \mathbf{u}_{t+1} is the eigenvector of the initial correlation matrix Σ .