Canonical Correlation Analysis

Rodrigo Loureiro Malacarne

This article performs a canonical correlation analysis on financial data of country-specific Exchange Traded Funds (ETFs) to analyze the relationship between stock markets in developed and developing countries. We conclude, using Bartlett's statistic, that there is a significant relationship between these two datasets.

■ Introduction

The aim of canonical correlation analysis is to find the best linear combination between two multivariate datasets that maximizes the correlation coefficient between them. This is particularly useful to determine the relationship between *criterion measures* and the set of their *explanatory factors*. This technique involves, first, the reduction of the dimensions of the two multivariate datasets by projection, and second, the calculation of the relationship (measured by the correlation coefficient) between the two projections of the datasets.

While the correlation coefficient measures the relationship between two simple variables, canonical correlation analysis measures the relationship between two *sets* of variables. Although the correlation measure employed for both techniques is the same, namely

$$\operatorname{corr}(\mathbf{X}, \mathbf{Y}) = \frac{\operatorname{cov}(\mathbf{X}, \mathbf{Y})}{\sqrt{\operatorname{var}(\mathbf{X}) \cdot \operatorname{var}(\mathbf{Y})}}, \tag{1}$$

the distinction between the two techniques must be clear: while for the correlation coefficient $\mathbf{X}$ and $\mathbf{Y}$ must be n-dimensional vectors containing realizations of the random variables, for canonical correlation analysis (CCA) $\mathbf{X}$ has to be an $n \times p$ and $\mathbf{Y}$ an $n \times q$ matrix, with p and q at least 2. In the latter case, n is the number of realizations for all $p + q$ random variables, where p is the number of random variables contained in the set $\mathbf{X}$ and q is the number of random variables in the set $\mathbf{Y}$.

Here **X** and **Y** are used for correlation coefficient calculation.

$$\mathbf{X} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad \mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}.$$

Here **X** and **Y** are used for canonical correlation analysis (CCA).

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{pmatrix},$$

$$\mathbf{Y} = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1q} \\ y_{21} & y_{22} & \cdots & y_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nq} \end{pmatrix} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_n \end{pmatrix}.$$

This article calculates, through CCA, the relationship between stock markets of developed and developing countries and performs Bartlett's test for the statistical significance of the canonical correlation found.

For an introduction to statistics in financial markets, see [1].

■ The Data

The data employed for the CCA in the present work was obtained directly from *Mathematica*'s `FinancialData[]` function. The variables are divided into two groups: the ETFs representing developed nations and the ETFs representing developing countries. The first group is treated as independent variables and the second group as dependent variables. The idea here is to analyze the relationship between stock markets in these two groups of countries through ETFs traded at the New York Stock Exchange (NYSE).

Although there are several country-specific ETFs traded on the NYSE, not all of them were chosen. The idea is to select, for each group, those ETFs representing countries with large stock markets according to a market capitalization criterion. The market capitalization of all stock markets was obtained from the website of the World Federation of Exchanges (www.world-exchanges.org/statistics). All countries with stock markets greater than 500 billion US dollars in December 2012 were chosen, and only one ETF per country was selected.

These six ETFs were included in the group of developed nations: EWA (Australia), EWC (Canada), EWG (Germany), EWJ (Japan), EWU (UK), and SPY (USA).

Eight ETFs were included in the group of developing countries: EWZ (Brazil), FXI (China), EPI (India), EWW (Mexico), RSX (Russia), EWS (Singapore), EWY (South Korea), and EWT (Taiwan).

These are the monthly returns for the five-year period between March 2008 and February 2013 (60 months).

```
ETFs =
  FinancialData[#, "Return",
     {{2008, 2}, {2013, 2}, "Month"}][[All, 2]] & /@
   {"EWA", "EWC", "EWG", "EWJ", "EWU", "SPY", "EWZ",
    "FXI", "EPI", "EWW", "RSX", "EWS", "EWY", "EWT"};
```
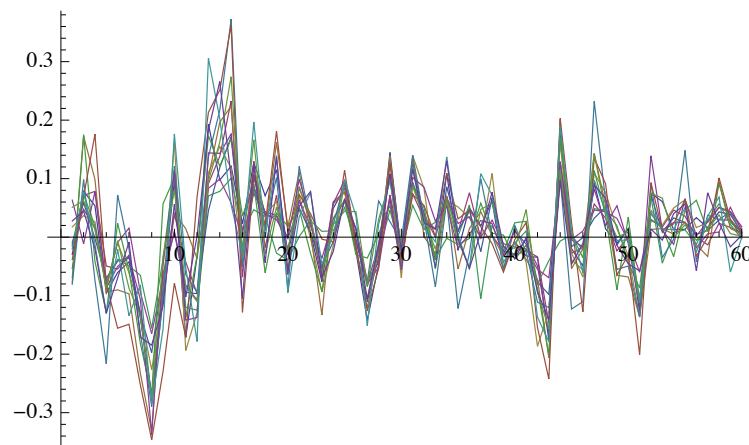
This checks the number of observations for each variable. Evaluate the previous command again if the lengths are not all 60.

```
Length[#] & /@ ETFs
```

```
{60, 60, 60, 60, 60, 60, 60, 60, 60, 60, 60, 60, 60, 60}
```
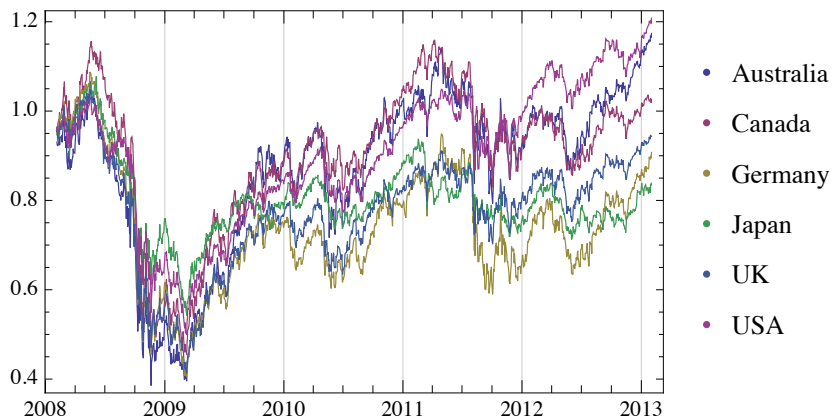
This plots the data for all the variables.
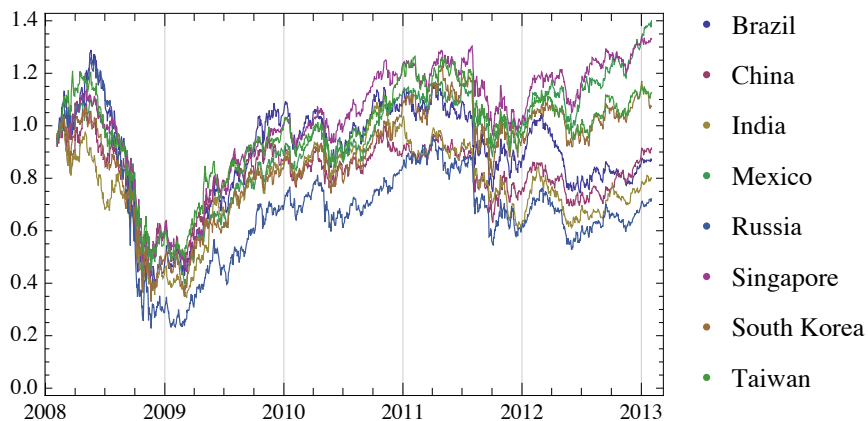
```
ListLinePlot[ETFs]
```

This plots the price behavior of the six ETFs representing developed countries for the 60-month period.

```
DateListPlot[
 Tooltip[FinancialData[#, "CumulativeFractionalChange",
    {{2008, 2}, {2013, 2}}], #] & /@
  {"EWA", "EWC", "EWG", "EWJ", "EWU", "SPY"}, Joined → True,
 PlotLegends → {"Australia", "Canada", "Germany",
   "Japan", "UK", "USA"}]
```


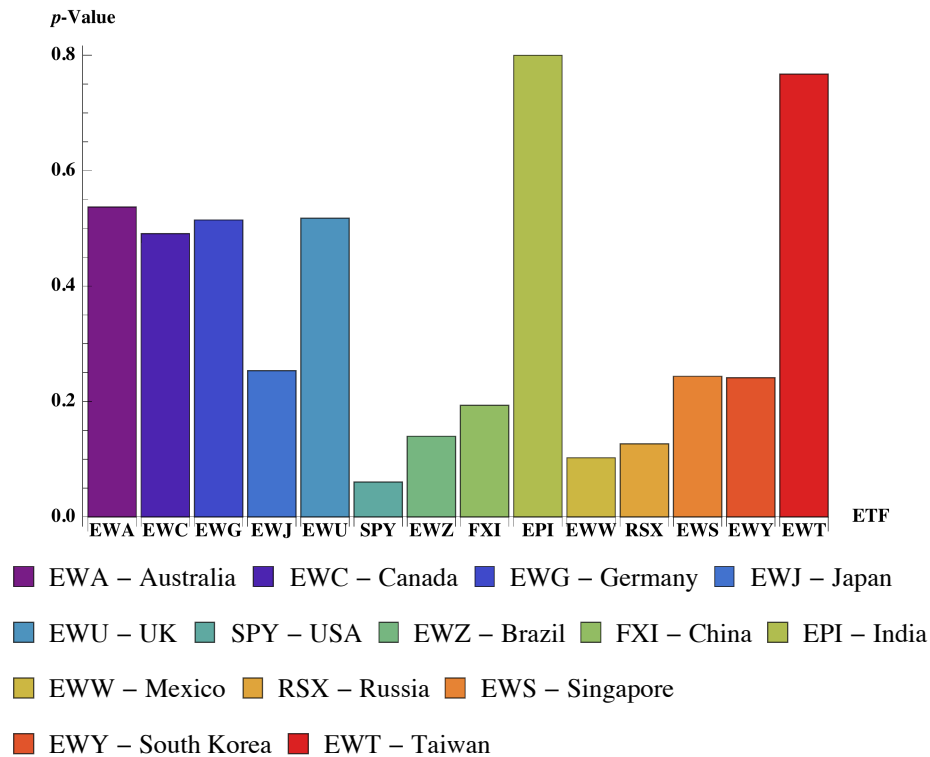This plots the price behavior of the eight ETFs representing developing countries for the 60-month period.

```
DateListPlot[
 Tooltip[FinancialData[#, "CumulativeFractionalChange",
    {{2008, 2}, {2013, 2}}], #] & /@
  {"EWZ", "FXI", "EPI", "EWW", "RSX", "EWS", "EWY", "EWT"},
 Joined → True,
 PlotLegends → {"Brazil", "China", "India", "Mexico",
   "Russia", "Singapore", "South Korea", "Taiwan"}]
```

According to [2], "to use canonical correlation analysis safely for descriptive purposes requires no distributional assumptions." However, they still state that "to test the significance of the relationships between canonical variates, (…), the data should meet the requirements of multivariate normality and homogeneity of variance" ([2], p. 339). Is the data normally distributed in this sense?

```
BarChart[DistributionFitTest[#] & /@ ETFs,
 ChartLabels -> {"EWA", "EWC", "EWG", "EWJ", "EWU",
   "SPY", "EWZ", "FXI", "EPI", "EWW", "RSX", "EWS",
   "EWY", "EWT"},
 ChartLegends →
  Placed[{"EWA - Australia", "EWC - Canada",
    "EWG - Germany", "EWJ - Japan", "EWU - UK",
    "SPY - USA", "EWZ - Brazil", "FXI - China",
    "EPI - India", "EWW - Mexico", "RSX - Russia",
    "EWS - Singapore", "EWY - South Korea",
    "EWT - Taiwan"}, Below],
 LabelStyle -> { Bold, FontSize -> 8},
 AxesLabel -> {"ETF", Row[{Style["p", Italic], "-Value"}]},
 ChartStyle -> "Rainbow"]
```

As can be seen, the null hypothesis of normality cannot be rejected for all variables at the 5% confidence level.

In order to perform the canonical correlation analysis, it is necessary to organize the data into two groups of variables: $\mathbf{X}$ (representing the developed countries) and $\mathbf{Y}$ (representing the developing countries);

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \\ \mathbf{x}_4 \\ \mathbf{x}_5 \\ \mathbf{x}_6 \end{pmatrix}, \mathbf{Y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \\ \mathbf{y}_4 \\ \mathbf{y}_5 \\ \mathbf{y}_6 \\ \mathbf{y}_7 \\ \mathbf{y}_8 \end{pmatrix},$$

where $\mathbf{x}_1$ to $\mathbf{x}_6$ represent the developed countries' ETFs and $\mathbf{y}_1$ to $\mathbf{y}_8$ represent the developing countries' ETFs.

■ The Theory

In canonical correlation analysis, $\mathbf{X} \in \mathbb{R}^p$ and $\mathbf{Y} \in \mathbb{R}^q$, and the problem is to find the "most interesting" linear combinations

$\mathbf{a}^\mathrm{T}\mathbf{X}$ and $\mathbf{b}^\mathrm{T}\mathbf{Y}$

for the two sets of variables, that is, those values that maximize

$$\rho = \mathrm{Corr}(\mathbf{a}^\mathrm{T}\mathbf{X}, \mathbf{b}^\mathrm{T}\mathbf{Y}). \tag{2}$$

Let $\mathbf{Z}$ be the concatenation of the matrices $\mathbf{X}$ and $\mathbf{Y}$,

$\mathbf{Z} = (\mathbf{X}^\mathrm{T}, \mathbf{Y}^\mathrm{T}),$

so

$$\mathbf{Z} \sim \left((\boldsymbol{\mu}_\mathbf{X}\ \boldsymbol{\mu}_\mathbf{Y}), \begin{pmatrix} \mathbf{S_{XX}} & \mathbf{S_{YX}} \\ \mathbf{S_{XY}} & \mathbf{S_{YY}} \end{pmatrix} \right),$$

where $\mathbf{S_{XX}}$ and $\mathbf{S_{YY}}$ are the (empirical) variance-covariance matrices and $\boldsymbol{\mu}_\mathbf{X}$ and $\boldsymbol{\mu}_\mathbf{Y}$ are the mean vectors of $\mathbf{X}$ and $\mathbf{Y}$, respectively. $\mathbf{S_{XY}}$ represents the covariance matrix of $\mathbf{X}$ and $\mathbf{Y}$, and $\mathbf{S_{YX}}$ is its transpose.

From equation (1) and from the properties

$$\mathrm{Var}(\mathbf{U}\,\mathbf{X} + c) = \mathbf{U}\,\mathrm{Var}(\mathbf{X})\,\mathbf{U}^\mathrm{T}, \tag{3}$$

$$\mathrm{Cov}(\mathbf{U}\,\mathbf{X}, \mathbf{V}\,\mathbf{Y}) = \mathbf{U}\,\mathrm{Cov}(\mathbf{X}, \mathbf{Y})\,\mathbf{V}^\mathrm{T}, \tag{4}$$

where $\mathbf{U}$ and $\mathbf{V}$ are conformable and c is a constant,

$$\text{Corr}(\mathbf{A}^{\mathsf{T}}\mathbf{X}, \mathbf{B}^{\mathsf{T}}\mathbf{Y}) = \frac{\mathbf{A}^{\mathsf{T}}\mathbf{S_{XY}}\mathbf{B}}{(\mathbf{A}^{\mathsf{T}}\mathbf{S_{XX}}\mathbf{A})^{1/2}\,(\mathbf{B}^{\mathsf{T}}\mathbf{S_{YY}}\mathbf{B})^{1/2}}\,, \tag{5}$$

where

$$\mathbf{A} = \begin{pmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{p1} & \cdots & a_{pk} \end{pmatrix} = (\mathbf{a}_1, \ldots, \mathbf{a}_k),$$

$$\mathbf{B} = \begin{pmatrix} b_{11} & \cdots & b_{1k} \\ \vdots & \ddots & \vdots \\ b_{q1} & \cdots & b_{qk} \end{pmatrix} = (\mathbf{b}_1, \ldots, \mathbf{b}_k),$$

$$k = \text{rank}\,(\mathbf{S_{XY}}) = \text{rank}\,(\mathbf{S_{YX}}).$$

◻ Solution Procedure I

CCA can be performed either on variance-covariance matrices or on correlation matrices. If the random variables $\mathbf{X}$ and $\mathbf{Y}$ are standardized to have unit variance, the variance-covariance matrix becomes a correlation matrix.

After partitioning the variance-covariance matrix, and given equation (5), the main objective is to solve

$$\max_{\mathbf{A},\mathbf{B}} = \mathbf{A}^{\mathsf{T}}\mathbf{S_{XY}}\mathbf{B} \tag{6}$$

subject to

$$\mathbf{A}^{\mathsf{T}}\mathbf{S_{XX}}\mathbf{A} = \mathbf{I}$$

$$\mathbf{B}^{\mathsf{T}}\mathbf{S_{YY}}\mathbf{B} = \mathbf{I}.$$

To solve this problem, define:

$$\mathbf{K} = \mathbf{S_{XX}}^{-1/2}\,\mathbf{S_{XY}}\,\mathbf{S_{YY}}^{-1/2}. \tag{7}$$

A singular value decomposition of $\mathbf{K}$ gives

$$\mathbf{K} = \Gamma\,\Lambda\,\Delta^{\mathsf{T}} \tag{8}$$

where

$$\Gamma = (\gamma_1, \ldots, \gamma_k), \tag{9}$$

$$\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_k), \tag{10}$$

$$\Delta = (\delta_1, \ldots, \delta_k), \tag{11}$$

Γ and Δ are column orthonormal matrices ($\Gamma^{\mathsf{T}}\Gamma = \Delta^{\mathsf{T}}\Delta = \mathbf{I}$), and Λ is a diagonal matrix with positive elements, namely, the eigenvalues of $\mathbf{K}$. (For detailed information about singular value decomposition, see [3].) From the property

$$\text{rank}(\mathbf{U}\,\mathbf{V}\,\mathbf{W}) = \text{rank}(\mathbf{V}), \text{ for nonsingular } \mathbf{U},\,\mathbf{W},$$

and from equation (7),

$$k = \text{rank}(\mathbf{K}) = \text{rank}(\mathbf{S_{XY}}) = \text{rank}(\mathbf{S_{YX}}).$$

For this solution procedure, the largest eigenvalue of $\mathbf{K}$ is the canonical correlation of our analysis. $\mathbf{A}$ and $\mathbf{B}$ can also be found through

$$\mathbf{A} = \mathbf{S_{XX}}^{-1/2}\, \Gamma, \tag{12}$$

$$\mathbf{B} = \mathbf{S_{YY}}^{-1/2}\, \Delta. \tag{13}$$

□ Solution Procedure II

The problem in this case is to solve the following canonical equations [2, 4]:

$$\left(\mathbf{S_{XX}}^{-1}\, \mathbf{S_{XY}}\, \mathbf{S_{YY}}^{-1}\, \mathbf{S_{YX}} - \lambda\, \mathbf{I}\right)\mathbf{A} = \mathbf{0} \tag{14}$$

and

$$\left(\mathbf{S_{YY}}^{-1}\, \mathbf{S_{YX}}\, \mathbf{S_{XX}}^{-1}\, \mathbf{S_{XY}} - \lambda\, \mathbf{I}\right)\mathbf{B} = \mathbf{0}, \tag{15}$$

where $\mathbf{I}$ is the identity matrix and λ is the largest eigenvalue for the characteristic equations

$$\left|\, \mathbf{S_{XX}}^{-1}\, \mathbf{S_{XY}}\, \mathbf{S_{YY}}^{-1}\, \mathbf{S_{YX}} - \lambda\, \mathbf{I}\,\right| = 0, \tag{16}$$

and

$$\left|\, \mathbf{S_{YY}}^{-1}\, \mathbf{S_{YX}}\, \mathbf{S_{XX}}^{-1}\, \mathbf{S_{XY}} - \lambda\, \mathbf{I}\,\right| = 0. \tag{17}$$

The largest eigenvalue of the product matrices

$$\mathbf{S_{XX}}^{-1}\, \mathbf{S_{XY}}\, \mathbf{S_{YY}}^{-1}\, \mathbf{S_{YX}} \text{ or } \mathbf{S_{YY}}^{-1}\, \mathbf{S_{YX}}\, \mathbf{S_{XX}}^{-1}\, \mathbf{S_{XY}}$$

is the squared canonical correlation coefficient. Furthermore, it can be shown that

$$\mathbf{A} = \frac{\mathbf{S_{XX}}^{-1}\, \mathbf{S_{XY}}\, \mathbf{B}}{\sqrt{\lambda}} \tag{18}$$

and

$$\mathbf{B} = \frac{\mathbf{S_{YY}}^{-1}\, \mathbf{S_{YX}}\, \mathbf{A}}{\sqrt{\lambda}}, \tag{19}$$

which means that only one of the characteristic equations needs to be solved in order to find $\mathbf{A}$ or $\mathbf{B}$.

■ The Technique

This transposes the data.

```
Z = Transpose[ETFs];
```

This checks the dimensions of Z; it has 60 rows (months) and 14 columns (ETFs).

```
Dimensions[Z]
```

```
{60, 14}
```

There are 14 random variables (six in the first set and eight in the second); the dimensions of the submatrices are 60×6 for $\mathbf{X}$, 60×8 for $\mathbf{Y}$, 6×6 for $\mathbf{S_{XX}}$, 6×8 for $\mathbf{S_{XY}}$, 8×6 for $\mathbf{S_{YX}}$, and 8×8 for $\mathbf{S_{YY}}$.

Define $M1$ to be the variance-covariance matrix of Z. Here are the first seven columns of $M1$.

```
M1 = Covariance[Z];
Take[#, 7] & /@ M1 // MatrixForm
```

$$
\begin{pmatrix}
0.00795939 & 0.0060745 & 0.00725258 & 0.0037696 & 0.00555587 & 0.0043357 & 0.00807816 \\
0.0060745 & 0.00612215 & 0.00586968 & 0.00325367 & 0.00482703 & 0.00371566 & 0.00737054 \\
0.00725258 & 0.00586968 & 0.0083465 & 0.00418995 & 0.00566203 & 0.00446104 & 0.00748621 \\
0.0037696 & 0.00325367 & 0.00418995 & 0.00322637 & 0.00309875 & 0.00241995 & 0.0037935 \\
0.00555587 & 0.00482703 & 0.00566203 & 0.00309875 & 0.00481028 & 0.00347672 & 0.00616873 \\
0.0043357 & 0.00371566 & 0.00446104 & 0.00241995 & 0.00347672 & 0.00295862 & 0.00459984 \\
0.00807816 & 0.00737054 & 0.00748621 & 0.0037935 & 0.00616873 & 0.00459984 & 0.0108334 \\
0.00656412 & 0.00537637 & 0.00634277 & 0.00321835 & 0.00478931 & 0.00359436 & 0.00773334 \\
0.00752131 & 0.00700981 & 0.00801951 & 0.00423608 & 0.00597372 & 0.00450287 & 0.00942523 \\
0.0068966 & 0.00592205 & 0.00697301 & 0.0039565 & 0.00517899 & 0.00419915 & 0.00731397 \\
0.00969121 & 0.00887666 & 0.00920613 & 0.00486632 & 0.00752961 & 0.0054498 & 0.0116731 \\
0.00648799 & 0.00594467 & 0.00673921 & 0.00357384 & 0.00520314 & 0.00383307 & 0.00767873 \\
0.00803397 & 0.00654152 & 0.00823685 & 0.0042539 & 0.00567634 & 0.00473121 & 0.0085958 \\
0.00620696 & 0.00538466 & 0.00647598 & 0.00321678 & 0.00480555 & 0.00382077 & 0.00716238
\end{pmatrix}
$$

Partition $M1$ into the four submatrices $\mathbf{S_{XX}}$, $\mathbf{S_{XY}}$, $\mathbf{S_{YX}}$, and $\mathbf{S_{YY}}$.

```
S_XX = Take[M1, {1, 6}, {1, 6}];
S_XX // MatrixForm
```

$$
\begin{pmatrix}
0.00795939 & 0.0060745 & 0.00725258 & 0.0037696 & 0.00555587 & 0.0043357 \\
0.0060745 & 0.00612215 & 0.00586968 & 0.00325367 & 0.00482703 & 0.00371566 \\
0.00725258 & 0.00586968 & 0.0083465 & 0.00418995 & 0.00566203 & 0.00446104 \\
0.0037696 & 0.00325367 & 0.00418995 & 0.00322637 & 0.00309875 & 0.00241995 \\
0.00555587 & 0.00482703 & 0.00566203 & 0.00309875 & 0.00481028 & 0.00347672 \\
0.0043357 & 0.00371566 & 0.00446104 & 0.00241995 & 0.00347672 & 0.00295862
\end{pmatrix}
$$

```
SYY = Take[M1, {7, 14}, {7, 14}]; SYY // MatrixForm
```

$$\begin{pmatrix} 0.0108334 & 0.00773334 & 0.00942523 & 0.00731397 & 0.0116731 & 0.00767873 & 0.0085958 & 0.00716238 \\ 0.00773334 & 0.00778011 & 0.00781391 & 0.00608476 & 0.00838883 & 0.00637501 & 0.00704273 & 0.00538344 \\ 0.00942523 & 0.00781391 & 0.0124785 & 0.00758933 & 0.0107761 & 0.0083224 & 0.00896888 & 0.00739509 \\ 0.00731397 & 0.00608476 & 0.00758933 & 0.00759882 & 0.00898833 & 0.00668777 & 0.0074871 & 0.00571778 \\ 0.0116731 & 0.00838883 & 0.0107761 & 0.00898833 & 0.0167653 & 0.00915405 & 0.0101488 & 0.00906655 \\ 0.00767873 & 0.00637501 & 0.0083224 & 0.00668777 & 0.00915405 & 0.00741912 & 0.00763834 & 0.00607639 \\ 0.0085958 & 0.00704273 & 0.00896888 & 0.0074871 & 0.0101488 & 0.00763834 & 0.0112516 & 0.007834 \\ 0.00716238 & 0.00538344 & 0.00739509 & 0.00571778 & 0.00906655 & 0.00607639 & 0.007834 & 0.0073981 \end{pmatrix}$$

```
SXY = Take[M1, {1, 6}, {7, 14}]; SXY // MatrixForm
```

$$\begin{pmatrix} 0.00807816 & 0.00656412 & 0.00752131 & 0.0068966 & 0.00969121 & 0.00648799 & 0.00803397 & 0.00620696 \\ 0.00737054 & 0.00537637 & 0.00700981 & 0.00592205 & 0.00887666 & 0.00594467 & 0.00654152 & 0.00538466 \\ 0.00748621 & 0.00634277 & 0.00801951 & 0.00697301 & 0.00920613 & 0.00673921 & 0.00823685 & 0.00647598 \\ 0.0037935 & 0.00321835 & 0.00423608 & 0.0039565 & 0.00486632 & 0.00357384 & 0.0042539 & 0.00321678 \\ 0.00616873 & 0.00478931 & 0.00597372 & 0.00517899 & 0.00752961 & 0.00520314 & 0.00567634 & 0.00480555 \\ 0.00459984 & 0.00359436 & 0.00450287 & 0.00419915 & 0.0054498 & 0.00383307 & 0.00473121 & 0.00382077 \end{pmatrix}$$

```
SYX = Transpose[SXY];
SYX // MatrixForm
```

$$\begin{pmatrix} 0.00807816 & 0.00737054 & 0.00748621 & 0.0037935 & 0.00616873 & 0.00459984 \\ 0.00656412 & 0.00537637 & 0.00634277 & 0.00321835 & 0.00478931 & 0.00359436 \\ 0.00752131 & 0.00700981 & 0.00801951 & 0.00423608 & 0.00597372 & 0.00450287 \\ 0.0068966 & 0.00592205 & 0.00697301 & 0.0039565 & 0.00517899 & 0.00419915 \\ 0.00969121 & 0.00887666 & 0.00920613 & 0.00486632 & 0.00752961 & 0.0054498 \\ 0.00648799 & 0.00594467 & 0.00673921 & 0.00357384 & 0.00520314 & 0.00383307 \\ 0.00803397 & 0.00654152 & 0.00823685 & 0.0042539 & 0.00567634 & 0.00473121 \\ 0.00620696 & 0.00538466 & 0.00647598 & 0.00321678 & 0.00480555 & 0.00382077 \end{pmatrix}$$

To better understand the relationship between the random variables, here is M2, the correlation matrix of Z.

```
M2 = Correlation[Z];
Take[#, 7] & /@ M2 // MatrixForm
```

$$\begin{pmatrix} 1. & 0.8702 & 0.889816 & 0.743873 & 0.897897 & 0.893459 & 0.86994 \\ 0.8702 & 1. & 0.821126 & 0.732091 & 0.889494 & 0.873052 & 0.905032 \\ 0.889816 & 0.821126 & 1. & 0.807421 & 0.893582 & 0.897716 & 0.787275 \\ 0.743873 & 0.732091 & 0.807421 & 1. & 0.786582 & 0.78326 & 0.641653 \\ 0.897897 & 0.889494 & 0.893582 & 0.786582 & 1. & 0.921596 & 0.854529 \\ 0.893459 & 0.873052 & 0.897716 & 0.78326 & 0.921596 & 1. & 0.812484 \\ 0.86994 & 0.905032 & 0.787275 & 0.641653 & 0.854529 & 0.812484 & 1. \\ 0.834149 & 0.779013 & 0.787107 & 0.642367 & 0.782879 & 0.749178 & 0.842347 \\ 0.754697 & 0.801998 & 0.785805 & 0.667615 & 0.771043 & 0.741079 & 0.810639 \\ 0.886792 & 0.868254 & 0.875578 & 0.799064 & 0.856617 & 0.885612 & 0.806114 \\ 0.838943 & 0.876178 & 0.77825 & 0.661665 & 0.838458 & 0.773804 & 0.866155 \\ 0.844295 & 0.882063 & 0.856407 & 0.730469 & 0.870971 & 0.818138 & 0.856505 \\ 0.848952 & 0.788169 & 0.849967 & 0.706029 & 0.77157 & 0.820013 & 0.778566 \\ 0.808871 & 0.800104 & 0.824125 & 0.658421 & 0.805559 & 0.816669 & 0.800044 \end{pmatrix}$$

This defines K.

```
K = MatrixPower[S_XX, -1/2].S_XY.MatrixPower[S_YY, -1/2];
K // MatrixForm
```

$$\begin{pmatrix} 0.271151 & 0.352789 & -0.0325728 & 0.219503 & 0.259628 & -0.0684817 & 0.328458 & 0.0434131 \\ 0.411041 & -0.0397389 & 0.240208 & 0.0732172 & 0.361416 & 0.261652 & 0.0239136 & 0.0136539 \\ -0.0814398 & 0.175156 & 0.268967 & 0.178162 & 0.0728574 & 0.197723 & 0.328236 & 0.252641 \\ -0.129041 & 0.0123515 & 0.130073 & 0.335202 & 0.0631399 & 0.0576363 & 0.135943 & -0.0334651 \\ 0.216228 & 0.0905682 & 0.0738786 & -0.0142152 & 0.266901 & 0.341796 & -0.182569 & 0.139534 \\ 0.11749 & -0.0402141 & -0.000160231 & 0.387394 & -0.104841 & -0.141786 & 0.195696 & 0.274237 \end{pmatrix}$$

This performs the singular value decomposition on K.

```
({Γ, Λ, Δ} = SingularValueDecomposition[K,
     Min[Dimensions[K]]]);
{Γ, Λ, Transpose[Δ]} // (Map[MatrixForm, #] &)
```

$$\left\{ \begin{pmatrix} -0.538801 & -0.1968 & -0.702829 & 0.398369 & -0.00318537 & 0.135228 \\ -0.519439 & 0.499902 & -0.015898 & -0.323312 & -0.379338 & -0.481248 \\ -0.462462 & -0.270999 & 0.593072 & 0.397906 & 0.31726 & -0.31933 \\ -0.224782 & -0.280031 & 0.347581 & -0.00471202 & -0.70685 & 0.500584 \\ -0.321826 & 0.539419 & 0.167422 & -0.0837616 & 0.417322 & 0.629485 \\ -0.268154 & -0.518206 & -0.0721247 & -0.755913 & 0.285734 & 0.036135 \end{pmatrix}, \right.$$

$$\begin{pmatrix} 0.971955 & 0. & 0. & 0. & 0. & 0. \\ 0. & 0.749561 & 0. & 0. & 0. & 0. \\ 0. & 0. & 0.470167 & 0. & 0. & 0. \\ 0. & 0. & 0. & 0.381178 & 0. & 0. \\ 0. & 0. & 0. & 0. & 0.293977 & 0. \\ 0. & 0. & 0. & 0. & 0. & 0.212468 \end{pmatrix},$$

$$\left. \begin{pmatrix} -0.405401 & -0.27942 & -0.292792 & -0.425274 & -0.445793 & -0.283333 & -0.376016 & -0.265693 \\ 0.354977 & -0.0940912 & 0.0761927 & -0.476497 & 0.387497 & 0.443459 & -0.506428 & -0.170308 \\ -0.558381 & -0.25753 & 0.502337 & 0.0774503 & -0.150622 & 0.529001 & -0.0722986 & 0.236204 \\ -0.429189 & 0.644942 & 0.0254614 & -0.415979 & 0.189321 & 0.118254 & 0.31598 & -0.276569 \\ 0.110197 & 0.296267 & -0.227366 & -0.354204 & -0.265377 & 0.0853083 & -0.0760133 & 0.799654 \\ -0.279462 & 0.341885 & -0.443743 & 0.519615 & 0.158806 & 0.190922 & -0.52577 & -0.00180777 \end{pmatrix} \right\}$$

This is the largest eigenvalue of K.

```
Max[Λ]
```

```
0.971955
```

This checks by computing the square root of the eigenvalues of

$$N_1 = S_{XX}^{-1} S_{XY} S_{YY}^{-1} S_{YX} \text{ and } N_2 = S_{YY}^{-1} S_{YX} S_{XX}^{-1} S_{XY}$$

according to the second solution procedure. (Chop replaces numbers that are close to zero by the exact integer 0.)

```
N1 = MatrixPower[S_XX, -1].S_XY.MatrixPower[S_YY, -1].S_YX;
N1 // MatrixForm
```

$$\begin{pmatrix} 0.494204 & 0.218224 & 0.294593 & 0.142871 & 0.17855 & 0.197417 \\ 0.371325 & 0.565149 & 0.316036 & 0.164735 & 0.402222 & 0.192645 \\ 0.163371 & 0.107079 & 0.357208 & 0.170383 & 0.140408 & 0.120919 \\ 0.0442545 & -0.023319 & 0.11238 & 0.151472 & -0.0127134 & 0.0561372 \\ -0.187836 & 0.160555 & -0.233617 & -0.190563 & 0.146297 & -0.172369 \\ 0.143677 & -0.171878 & 0.209731 & 0.179497 & -0.104109 & 0.290126 \end{pmatrix}$$

```
Sqrt[Eigenvalues[N1]]
```

```
{0.971955, 0.749561, 0.470167, 0.381178, 0.293977, 0.212468}
```

```
N2 = MatrixPower[S_YY, -1].S_YX.MatrixPower[S_XX, -1].S_XY;
N2 // MatrixForm
```

$$\begin{pmatrix} 0.40665 & 0.1484 & 0.157544 & 0.127227 & 0.407786 & 0.159026 & 0.0854459 & 0.141849 \\ 0.00626537 & 0.130915 & -0.0155567 & 0.042418 & 0.0148242 & 0.00533221 & 0.135856 & 0.0396727 \\ -0.0452259 & -0.0288341 & 0.0621691 & 0.0172187 & -0.0274181 & 0.0253373 & 0.0221429 & 0.00626727 \\ 0.222424 & 0.244493 & 0.279946 & 0.50056 & 0.224105 & 0.211473 & 0.564517 & 0.337392 \\ 0.182219 & 0.129076 & 0.142737 & 0.06247 & 0.25516 & 0.134922 & 0.0525406 & 0.0483432 \\ 0.162043 & 0.0126256 & 0.226237 & -0.118596 & 0.306917 & 0.22083 & -0.252554 & 0.000535197 \\ 0.0417883 & 0.135274 & 0.0827455 & 0.164469 & 0.0356844 & 0.0397634 & 0.291767 & 0.101405 \\ -0.00816529 & 0.0143104 & 0.0415299 & 0.0816778 & -0.049616 & 0.0351104 & 0.111236 & 0.136405 \end{pmatrix}$$

```
Sqrt[Eigenvalues[N2] // Chop]
```

```
{0.971955, 0.749561, 0.470167,
 0.381178, 0.293977, 0.212468, 0, 0}
```

Performing a spectral decomposition on $\mathbf{N}_3 = \mathbf{K}\,\mathbf{K}^T$ and $\mathbf{N}_4 = \mathbf{K}^T\,\mathbf{K}$ and calculating the square roots of their eigenvalues is another check of the canonical correlation coefficient.

```
N3 = K.Transpose[K];
N3 // MatrixForm
```

$$\begin{pmatrix} 0.429092 & 0.190045 & 0.194212 & 0.0943536 & 0.0770348 & 0.161384 \\ 0.190045 & 0.433436 & 0.126582 & 0.0429495 & 0.285418 & 0.0116512 \\ 0.194212 & 0.126582 & 0.357365 & 0.159541 & 0.077945 & 0.150209 \\ 0.0943536 & 0.0429495 & 0.159541 & 0.172992 & -0.0148756 & 0.116811 \\ 0.0770348 & 0.285418 & 0.077945 & -0.0148756 & 0.301479 & -0.0576627 \\ 0.161384 & 0.0116512 & 0.150209 & 0.116811 & -0.0576627 & 0.310093 \end{pmatrix}$$

```
Sqrt[Eigenvalues[N3]]
```

```
{0.971955, 0.749561, 0.470167, 0.381178, 0.293977, 0.212468}
```

```
N4 = Transpose[K].K;
N4 // MatrixForm
```

$$
\begin{pmatrix}
0.32632 & 0.078325 & 0.0671696 & 0.0742909 & 0.250268 & 0.122688 & 0.0381334 & 0.0635185 \\
0.078325 & 0.166691 & 0.0343782 & 0.0930089 & 0.119162 & 0.0374444 & 0.149693 & 0.0602204 \\
0.0671696 & 0.0343782 & 0.153481 & 0.100845 & 0.125902 & 0.151033 & 0.0874929 & 0.0757295 \\
0.0742909 & 0.0930089 & 0.100845 & 0.34792 & 0.0731874 & -0.0011138 & 0.256303 & 0.148577 \\
0.250268 & 0.119162 & 0.125902 & 0.0731874 & 0.289551 & 0.200921 & 0.0571727 & 0.0409904 \\
0.122688 & 0.0374444 & 0.151033 & -0.0011138 & 0.200921 & 0.252495 & -0.0336499 & 0.0574329 \\
0.0381334 & 0.149693 & 0.0874929 & 0.256303 & 0.0571727 & -0.0336499 & 0.306304 & 0.121155 \\
0.0635185 & 0.0602204 & 0.0757295 & 0.148577 & 0.0409904 & 0.0574329 & 0.121155 & 0.161695
\end{pmatrix}
$$

```
Sqrt[Eigenvalues[N4] // Chop]
```

```
{0.971955, 0.749561, 0.470167,
 0.381178, 0.293977, 0.212468, 0, 0}
```

The checks agree.

The last step in this analysis is to find the canonical correlation vectors, which maximize the correlation between the canonical variates. According to equations (12) and (13), this computes the canonical correlation vectors.

```
A = MatrixPower[S_XX, -1 / 2].Γ;
A // MatrixForm
```

$$
\begin{pmatrix}
-4.35566 & -6.8927 & -24.711 & 14.9376 & -3.83951 & 5.08049 \\
-5.81308 & 14.8372 & 1.91355 & -6.86934 & -14.6082 & -20.5991 \\
-2.30863 & -4.13699 & 18.646 & 14.1762 & 8.4946 & -16.5887 \\
-0.439443 & -8.02144 & 6.71823 & 0.597798 & -24.7938 & 15.5012 \\
0.723582 & 28.2739 & 7.17402 & 2.3701 & 18.2707 & 29.5904 \\
-0.40495 & -32.3557 & -7.09957 & -42.5372 & 12.8216 & -2.01286
\end{pmatrix}
$$

The canonical correlation matrix **B** is computed using $S_{YY}^{-1/2} \triangle^{\mathsf{T}}$, not $S_{YY}^{-1/2} \triangle$, because *Mathematica*'s singular value decomposition gives $\triangle$ transposed.

```
B = MatrixPower[S_YY, -1 / 2].Δ;
B // MatrixForm
```

$$
\begin{pmatrix}
-2.57549 & 6.86444 & -13.7419 & -14.7903 & 1.90278 & -8.46949 \\
-0.575906 & -4.09041 & -5.90749 & 16.6482 & 8.22683 & 9.63799 \\
0.0633956 & -0.666331 & 7.51492 & -0.202924 & -6.8425 & -9.20899 \\
-4.59606 & -15.0944 & -0.712017 & -14.5928 & -8.76194 & 13.1409 \\
-1.31747 & 5.08685 & -1.86366 & 7.63898 & -7.33872 & 2.30219 \\
0.1799 & 18.8801 & 16.7062 & 4.78416 & 4.58196 & 5.53023 \\
-1.58413 & -8.62959 & -4.31395 & 8.70756 & -6.81483 & -12.7175 \\
-0.574675 & -4.52373 & 5.67938 & -9.42162 & 22.6216 & 4.78577
\end{pmatrix}
$$

Given that

$$\mathbf{a}_i = \mathbf{S_{XX}}^{-1/2} \gamma_i, \, i = (1, \dots, k),$$

$$\mathbf{b}_i = \mathbf{S_{YY}}^{-1/2} \delta_i, \, i = (1, \dots, k),$$

the canonical correlation vectors $\mathbf{a}_i$ and $\mathbf{b}_i$ are the columns of $\mathbf{A}$ and $\mathbf{B}$.

In terms of the canonical correlation vectors, the canonical variates are

$$\eta_i = \mathbf{a}_i{}^\mathsf{T} \mathbf{X},$$

$$\varphi_i = \mathbf{b}_i{}^\mathsf{T} \mathbf{Y},$$

where, as before,

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \\ \mathbf{x}_4 \\ \mathbf{x}_5 \\ \mathbf{x}_6 \end{pmatrix} \text{ and } \mathbf{Y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \\ \mathbf{y}_4 \\ \mathbf{y}_5 \\ \mathbf{y}_6 \\ \mathbf{y}_7 \\ \mathbf{y}_8 \end{pmatrix}.$$

Given that

$$\mathbf{A}^\mathsf{T} \mathbf{S}_{\mathbf{XY}} \mathbf{B} = \left(\mathbf{a}_i{}^\mathsf{T} \mathbf{S}_{\mathbf{XY}} \mathbf{b}_j \right)_{i,j=1,\ldots,6} = \begin{pmatrix} \lambda_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_4 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda_5 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda_6 \end{pmatrix}, \tag{20}$$

only $\mathbf{a}_1$ and $\mathbf{b}_1$ are needed in order to find λ_1. Thus, the only canonical variates needed are η_1 and φ_1.

```
a₁ = Transpose[A][[1]]
```

```
{-4.35566, -5.81308, -2.30863, -0.439443, 0.723582, -0.40495}
```

```
b₁ = Transpose[B][[1]]
```

```
{-2.57549, -0.575906, 0.0633956, -4.59606,
 -1.31747, 0.1799, -1.58413, -0.574675}
```

```
η₁ = a₁.(Subscript[x, #] & /@ Range[6])
```

```
-4.35566 x₁ - 5.81308 x₂ - 2.30863 x₃ -
  0.439443 x₄ + 0.723582 x₅ - 0.40495 x₆
```

```
φ₁ = b₁.(Subscript[y, #] & /@ Range[8])
```

$-2.57549\, y_1 - 0.575906\, y_2 + 0.0633956\, y_3 - 4.59606\, y_4 -$
$1.31747\, y_5 + 0.1799\, y_6 - 1.58413\, y_7 - 0.574675\, y_8$

■ The Interpretation

The interpretation of canonical correlation coefficients, canonical correlation vectors, and canonical variates is one of the most difficult tasks in the whole analysis. CCA would be better understood relating the original data matrix to the matrix computed using the canonical correlation vectors, which is simply a reduction of the data matrix through linear combinations of its elements. It should be easier to understand that the canonical correlation coefficient is merely the ordinary Bravais–Pearson correlation between the two columns of the reduced matrix.

In principle, one can say that the highest canonical correlation coefficient that was found is the maximum possible correlation between the two columns of the reduced matrix. In this case, it is usual to say that this coefficient represents the relationship between the two datasets, **X** and **Y**, in the sense of a correlation measure. Thus, if **X** is the matrix containing the explanatory factors of **Y**, the matrix containing the criterion measures (or criterion variables), it is possible to say that the explanatory factors would perfectly explain the criterion variables if $\lambda_1 = 1$. If $\lambda_1 = 0$, the explanatory factors have no influence on the criterion variables, and any value between 1 and 0 is merely an interpolation of these extreme cases.

In the next inputs we will compute and show (partially) the reduced data matrix. In order to demonstrate the validity of the CCA theory, we also compute the correlation for the other (not so interesting for our analysis) canonical variates. We start by defining ψ_2 and ϕ_2.

```
ψ₂ = Transpose[A][[2]].Take[Transpose[Z], 6]
```

```
{-0.701634, -0.0862203, -0.0194891, 1.10675, -0.919005,
 -0.980257, -0.667874, -0.22305, 0.167482, -1.78318,
 2.33614, 1.08783, -2.43577, 0.291475, 4.24827, -1.66435,
 0.160344, -1.15281, 0.15165, 0.364005, -0.0349483,
 -0.284747, -0.408797, -0.707029, -0.407211, -0.630875,
 0.462998, 0.166161, 1.20209, 0.993131, -1.03681,
 -0.235525, -0.976946, -0.281728, -0.384766, 0.37312,
 -0.00341568, -0.050043, 0.157003, -0.937606, 0.370199,
 0.757301, -0.643029, -0.166161, -0.511661, 0.176578,
 -1.22379, -0.252588, -1.91354, 1.02647, -0.724215,
 -0.0831884, -0.0377708, 0.450653, -0.0333892,
 0.525277, -0.508017, -0.0658863, -1.19839, -0.316336}
```

```
ϕ₂ = Transpose[B][[2]].Drop[Transpose[Z], 6]
```

```
{-0.33371, 1.24548, 1.04997, 0.986147, -0.420766,
 -1.10299, -1.4832, 0.397619, 0.235873, -2.52815,
 1.25937, 1.84652, -2.81063, -0.0695078, 3.99543,
 -0.863144, -0.0147283, -0.507002, 0.489669, 0.781387,
 0.416829, -0.479017, -0.0978382, -0.164522, -0.967378,
 0.490543, 0.102533, 0.117317, 0.854026, 0.683749,
 -0.608303, -0.898438, -0.774125, -0.440444, 0.0447816,
 0.0249161, -0.414276, -0.293167, 0.0538587, -0.213599,
 0.922356, -0.281703, 0.0175409, -0.218042, -0.740662,
 -0.798377, 1.26853, 0.435227, -1.07086, -0.156835,
 -0.334811, -0.71894, 1.10498, 0.167045, -1.06676,
 -0.13703, -1.07255, 0.0134996, 0.166383, -0.217133}
```

The first column of our reduced data matrix is ψ_1.

```
ψ₁ = Transpose[A][[1]].Take[Transpose[Z], 6]
```

```
{0.396743, -0.848148, -0.767434, 1.00521, 0.591439, 0.49632,
 1.63962, 3.27098, 1.24836, -1.07389, 1.44933, 1.03787,
 -1.6588, -1.57095, -2.09588, 0.372608, -1.57278, -0.295276,
 -1.03401, 0.649415, -0.932334, -0.189607, 0.939148,
 -0.551109, -1.02872, 0.0670338, 1.27282, 0.545789, -1.20293,
 0.390801, -1.29956, -0.499904, 0.162531, -1.00948,
 -0.0779906, -0.740676, -0.0421858, -0.509284, 0.506751,
 0.154219, 0.36837, 0.778653, 1.89, -1.74277, 0.29713,
 0.480178, -1.00121, -0.426809, 0.117258, 0.0392344,
 1.47932, -0.594257, -0.391835, -0.429013, -0.41206,
 -0.166701, -0.0584444, -0.448645, -0.370089, -0.111208}
```

The first value of ψ_1, for instance, refers to the linear combination between EWA, EWC, EWG, EWJ, EWU, and SPY for March 2008, such that

$$-4.35914\,x_1 - 5.81353\,x_2 - 2.28865\,x_3 - \\ 0.450599\,x_4 + 0.7212\,x_5 - 0.420922\,x_6 = 0.396013.$$

We can also define ϕ_1.

```
ϕ₁ = Transpose[B][[1]].Drop[Transpose[Z], 6]
```

```
{0.0202821, -0.715098, -0.792045, 1.14504, 0.729336, 0.8394,
 1.71655, 3.50744, 0.970058, -0.882707, 1.03891, 0.90983,
 -1.86765, -1.99771, -2.03299, 0.292375, -1.25016, -0.18,
 -1.0991, 0.113514, -0.927486, -0.355003, 0.767698,
 -0.402027, -1.0113, 0.0565531, 1.13518, 0.220337, -1.04349,
 0.48455, -1.19754, -0.529427, -0.0386316, -0.83475,
 0.178636, -0.0178977, -0.461011, -0.250112, 0.368992,
 0.0398996, 0.118652, 0.910772, 2.06065, -1.79901, 0.116565,
 0.401803, -1.15496, -0.470715, 0.0717579, 0.278235,
 1.41343, -0.868437, -0.122433, -0.0115514, -0.576028,
 0.101217, -0.136712, -0.779001, -0.186774, -0.111408}
```

Thus, after assigning the values to the canonical variates, ψ_1, ψ_2, ϕ_1, and ϕ_2, we have four vectors with the values of the linear combinations of **X** and **Y**. Now we can simply compute the Bravais–Pearson correlation between all the canonical variables.

```
Correlation[ψ₁, ϕ₁]
```

```
0.971955
```

```
Correlation[ψ₂, ϕ₂]
```

```
0.749561
```

```
Correlation[ψ₁, ϕ₂] // Chop
```

```
0
```

```
Correlation[ψ₂, ϕ₁] // Chop
```

```
0
```

```
Correlation[ψ₁, ψ₂] // Chop
```

```
0
```

```
Correlation[ϕ₁, ϕ₂] // Chop
```
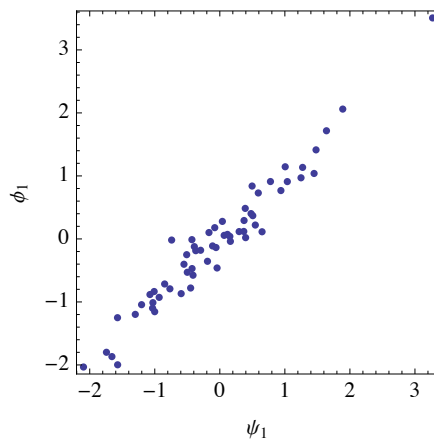
```
0
```

We also verify equation (20).

```
Transpose[A].S_XY.B // MatrixForm // Chop
```

$$\begin{pmatrix} 0.971955 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.749561 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.470167 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.381178 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.293977 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.212468 \end{pmatrix}$$
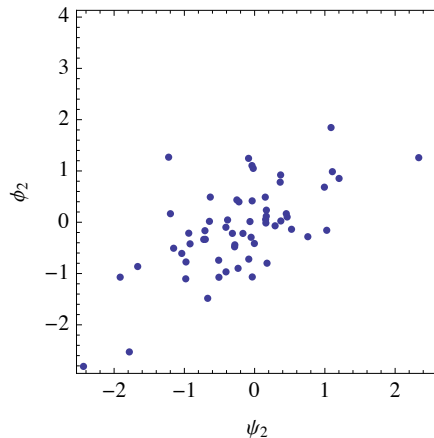
The correlation between the canonical variates can be better interpreted graphically. First we show the reduced matrix computed using the canonical correlation vectors a_1 and b_1, whose canonical correlation coefficient is $\lambda_1 = 0.971939$.

```
ListPlot[Transpose[{ψ_1, φ_1}], FrameLabel → {"ψ_1", "φ_1"},
  Axes → False, Frame → True, AspectRatio → 1, ImageSize → Small]
```

Now we show the reduced matrix computed using the canonical correlation vectors a_2 and b_2, whose canonical correlation coefficient is $\lambda_2 = 0.746103$.

```
ListPlot[Transpose[{ψ₂, ϕ₂}], FrameLabel → {"ψ₂", "ϕ₂"},
  Axes → False, Frame → True, AspectRatio → 1, ImageSize → Small]
```


Finally, we compute the *canonical loadings*, that is, the correlation between every single ETF and its respective canonical variate.

```
Correlation[ψ₁, #] & /@ ETFs[[1 ;; 6]]
```

```
{-0.965254, -0.959044, -0.925253,
 -0.795082, -0.93171, -0.928958}
```

```
Correlation[ϕ₁, #] & /@ ETFs[[7 ;; 14]]
```

```
{-0.932935, -0.8619, -0.841685, -0.947216,
 -0.905548, -0.928495, -0.889568, -0.870467}
```

We can also compute the *canonical cross-loadings*, that is, the correlation between every single ETF and its opposite canonical variate.

```
Correlation[ϕ₁, #] & /@ ETFs[[1 ;; 6]]
```

```
{-0.938184, -0.932148, -0.899305,
 -0.772784, -0.905581, -0.902905}
```

```
Correlation[ψ₁, #] & /@ ETFs[[7 ;; 14]]
```

```
{-0.906771, -0.837729, -0.81808, -0.920652,
 -0.880153, -0.902456, -0.86462, -0.846055}
```

It might be of interest to compute the canonical loadings for the *second canonical variate*, that is, the linear combination of variables with correlation coefficient $\lambda_2 = 0.746103$.

```
Correlation[ψ₂, #] & /@ ETFs[[1 ;; 6]]
```

```
{-0.0916125, 0.189661, -0.167379,
  -0.204344, 0.123393, -0.184749}
```

```
Correlation[ϕ₂, #] & /@ ETFs[[7 ;; 14]]
```

```
{0.228935, 0.0239593, 0.0981016, -0.148301,
  0.250836, 0.155292, -0.240297, -0.0501894}
```

Finally, we compute the canonical cross-loadings for the second canonical variate, that is, the linear combination of variables with correlation coefficient $\lambda_2 = 0.746103$.

```
Correlation[ϕ₂, #] & /@ ETFs[[1 ;; 6]]
```

```
{-0.0686692, 0.142162, -0.125461,
  -0.153169, 0.0924909, -0.138481}
```

```
Correlation[ψ₂, #] & /@ ETFs[[7 ;; 14]]
```

```
{0.1716, 0.017959, 0.0735331, -0.111161,
  0.188017, 0.116401, -0.180117, -0.03762}
```

It is possible to compute canonical loadings and cross-loadings for all the six canonical variates. However, only the first two are shown here for descriptive purposes.

■ Statistical Testing

In this section we test the hypothesis of no correlation between the two sets **X** and **Y**. An approximation for large n was provided in [5]:

$$-\left\{n - \frac{(p+q+3)}{2}\right\} \log \prod_{i=1}^{k} (1 - L_i) \sim \chi^2_{pq}, \tag{21}$$

where

n is the number of observations

p is the number of rows of **X**

q is the number of rows of **Y**

k is the rank (**K**)

L_i is the canonical correlation coefficients.

We can also test the hypothesis that the individual canonical correlation coefficients are different from zero:

$$-\left\{n - \frac{(p+q+3)}{2}\right\} \log \prod_{i=s+1}^{k} (1 - L_i) \sim \chi^2_{(p-s)(q-s)}, \qquad (22)$$

where *s* is a parameter to select the canonical correlation coefficient to be tested.

This defines the Bartlett variable.

```
Bartlett[n_, k_, p_, q_, s_] :=
    -(n - (p + q + 3) / 2) * Log[∏ (1 - L[i]²)]
                                   i=s+1
```

This assigns values to the L_i.

```
L[1] = Λ[[1, 1]];
L[2] = Λ[[2, 2]];
L[3] = Λ[[3, 3]];
L[4] = Λ[[4, 4]];
L[5] = Λ[[5, 5]];
L[6] = Λ[[6, 6]];
```

We calculate Bartlett's statistic (equation (21)) to test if the two sets of variables **X** and **Y** are uncorrelated. Our hypotheses are:

$H_0 : \text{Corr}(X, Y) = 0,$
$H_1 : \text{Corr}(X, Y) = 0.$

```
Bartlett[Length[Z], MatrixRank[K], Length[X], Length[Y],
  0]

249.415
```

This computes the 99% quantile of the chi-square distribution with 48 ($p\,q = 6 \times 8 = 48$) degrees of freedom, χ^2_{48}.

```
Quantile[ChiSquareDistribution[48], 0.99]

73.6826
```

Test Conclusion: The hypothesis of no correlation between the two sets has to be rejected once the Bartlett statistic (here 249.415) is greater than the 99% quantile of the chi-square distribution with 48 degrees of freedom (here 73.6826).

■ Conclusion

This article analyzed the relationship between two sets of variables, namely financial assets represented by NYSE-traded country-specific ETFs. The ETFs were divided into two sets representing developed and developing countries. In the first set a total of six ETFs (representing developed countries) were included, while in the second set a total of eight ETFs were included (representing developing countries). Using monthly return data for a five-year period it was possible to show, through canonical correlation analysis (CCA), that there is a significant relationship between these two sets of ETFs. The highest correlation coefficient found in the present study was $\lambda_1 = 0.971939$ and, in an analogous manner to R^2 statistics in regression analysis, we could interpret its squared value $\lambda_1{}^2 = 0.942723$ as the explanatory power of the canonical correlation analysis. In other words, the squared canonical correlation coefficient $\lambda_1{}^2$ indicates the proportion of variance a dependent variable linearly shares with the independent variable generated from the observed variable's set (i.e., the canonical variates).

■ References

[1] J. Franke, W. Härdle, and C. Hafner, *Einführung in die Statistik der Finanzmärkte*, Berlin: Springer Verlag, 2001.

[2] W. R. Dillon and M. Goldstein, Chap. 9 in *Multivariate Analysis*: *Methods and Applications*, New York: Wiley, 1984.

[3] K. V. Mardia, J. T. Kent, and J. M. Bibby, *Multivariate Analysis*, London: Academic Press, 1979.

[4] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, 3rd ed., New York: Wiley, 2003.

[5] M. S. Bartlett, "A Note on Tests of Significance in Multivariate Analysis," *Proceedings of the Cambridge Philosophical Society*, **35**(2), pp. 180–185, 1939. doi:10.1017/S0305004100020880.

About the Author

Rodrigo Loureiro Malacarne is a professor of financial mathematics and financial management at the Faculdades Integradas Espirito Santenses (FAESA). His areas of research include statistics of financial markets and financial time series analysis.

Rodrigo Loureiro Malacarne
School of Management and Business, School of Industrial Engineering
Faculdades Integradas Espirito Santenses (FAESA)
Av. Vitória, 2.220 - Monte Belo
Vitória, ES, Brazil - CEP 29.053-360
malacarne@gmail.com