Section 1

Principal Component Analysis
Background

- Principal Component Analysis (PCA) is a tool for looking at multivariate data.
- General setup: we observe several variables for each of several cases.
- In our context, the variables are financial:
  - interest rates for various maturities;
  - log returns for various stocks;
  - exchange rates between USD and various other currencies.
- Each case consists of the values of those variables on a given date.
• The general idea behind PCA (and *Factor Analysis*, FA) is that the way the variables covary can be attributed to common underlying forces.
  • For example, stock market returns are all affected by overall market sentiment.
• We look for:
  • common *modes* of variation (PCA);
  • unobserved (latent) *factors* (FA).
Matrix methods

- Write $y_{t,j}$ for the value of the $j^{th}$ variable on the $t^{th}$ date.
- Assemble these into a data matrix $X$, where $x_{t,j}$ might be:
  - raw data $y_{t,j}$;
  - centered data $y_{t,j} - \bar{y}_j$, where $\bar{y}_j$ is the average, over time, of the $j^{th}$ variable:
    \[ \bar{y}_j = \frac{1}{T} \sum_{t=1}^{T} y_{t,j}; \]
  - standardized (or scaled) data $\frac{y_{t,j} - \bar{y}_j}{s_j}$, where $s_j$ is the standard deviation, again over time, of the $j^{th}$ variable:
    \[ s_j = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (y_{t,j} - \bar{y}_j)^2}. \]
• The data are always centered by default.
  • But when all variables vary naturally around zero, such as log returns of tradable assets, it is not necessary.
• If the variables are in different units, they *must* be scaled to make them comparable.
• Even when they have common units, their variances may be very different, and scaling is again necessary.
  • Scaling by the standard deviation is convenient, but nothing more.
Modes of Variation

- Each *mode of variation* is a part of $X$ of the form

$$duv',$$

where:
- $d > 0$ is a scalar multiplier;
- $u$ is a column vector of length $T$, with one entry for each date;
- $v'$ is a row vector of length $J$, with one entry for each variable;
- in PCA, $u$ and $v'$ are normalized:

$$u'u = v'v = 1.$$
• Note that $duv'$ is a rank-1 matrix, and that any rank-1 matrix can be written in this form.

• Terminology:
  • The entries of the (normalized) row vector $v'$ are called the loadings for the mode.
  • The entries of the (unnormalized) column vector $du$ are called the scores for the mode.
Principal Component

- PCA and FA differ in how the loadings and scores are constructed.
- In PCA, the first (or dominant) component is defined to be the best approximation to $X$ in the Frobenius norm:

$$d_1 u_1 v_1' = \underset{d,u,v}{\text{argmin}} \|X - duv'\|_F,$$

where for any $T \times J$ matrix $A$,

$$\|A\|_F = \sqrt{\sum_{t=1}^{T} \sum_{j=1}^{J} a_{t,j}^2}.$$
• The next component is the one that gives the best rank-2 approximation:

\[ d_2 u_2 v'_2 = \underset{d,u,v}{\arg\min} \| X - d_1 u_1 v'_1 - d u v' \|_F. \]

• If, as here, we fix the first component and optimize over only the second, the solution can be shown to have the orthogonality properties

\[ u'_1 u_2 = v'_1 v_2 = 0. \tag{1} \]

• If, instead, we optimize over both components simultaneously, we need to impose a constraint like (1), and the solution is essentially the same.
Components 3 through $J$ are defined similarly, either:

- incrementally, in which case they automatically satisfy the generalization of (1);
- or simultaneously, constrained by (1).

Again, the solution is the same either way.

Note that for each component,

$$d_k u_k v'_k = (-d_k u_k)(-v'_k).$$

That is, the loadings and scores are determined only up to multiplication by $-1$.

You should feel free to change the sign if it simplifies interpretation, provided you change both the loadings and the scores.
Singular Value Decomposition

- PCA can be carried out using the *Singular Value Decomposition* (SVD).
- Any $T \times J$ matrix $X$, $T \geq J$, can be factorized as

$$X = UDV'$$  \hspace{1cm} (2)

where:
- $U$ is $T \times J$ with $U'U = I_J$;
- $D$ is $J \times J$ diagonal, with diagonal entries $d_1 \geq d_2 \geq \cdots \geq d_J \geq 0$;
- $V$ is $J \times J$ with $V'V = I_J$. 
• Equation (2) can also be written

\[ X = \sum_{k=1}^{J} d_k u_k v_k', \]

where \( u_k \) is the \( k^{th} \) column of \( U \) and \( v_k' \) is the \( k^{th} \) row of \( V' \).

• Easily shown: \( d_k u_k v_k' \) is the \( k^{th} \) PCA component.

• Terminology: \( d_k, u_k, \) and \( v_k' \) are the \( k^{th} \) singular value, left singular vector, and right singular vector, respectively.
Loadings and Scores

- Note that the SVD factorization

\[ X = UDV' \]

and the orthogonality conditions \( U'U = V'V = I_J \) imply that

\[ U = XD^{-1}, \]
\[ D = U'XV, \]

and

\[ V' = D^{-1}U'X. \]

- That is, any one of \( X, U, D, \) and \( V' \) can be calculated directly from the other three.
Covariance and Correlation

- PCA is often described in terms of the covariance or correlation matrix, rather than the data matrix.
- If $X$ is the centered data matrix, then
  \[
  \frac{1}{T}X'X
  \]
  is the sample covariance matrix.
- If $X$ is the standardized data matrix, then
  \[
  \frac{1}{T}X'X
  \]
  is the sample correlation matrix.
• In either case, the SVD shows that

$$\frac{1}{T} X'X = V \left( \frac{1}{T} D^2 \right) V'.$$

• That is, the eigenvectors of $\frac{1}{T} X'X$ are the columns of $V$, which are the transposes of the rows of loadings.

• Also, the eigenvalues of $\frac{1}{T} X'X$ are $\frac{1}{T} d_k^2$.

• So the loadings and singular values can be found from the spectral decomposition of the correlation matrix or covariance matrix, as appropriate.

• For the scores, you need the original data matrix:

$$UD = XV.$$
• Note that the variances of the variables are the diagonal entries of \( \frac{1}{T}X'X \).

• The total variance is

\[
\text{tr} \frac{1}{T}X'X = \text{tr}V \left( \frac{1}{T}D^2 \right) V' = \frac{1}{T} \text{tr}D^2
\]

• That is, each squared singular value measures the contribution of the component to the total variance.

• If the data were scaled, each variance is 1, and

\[
\text{tr} \frac{1}{T}X'X = \frac{1}{T} \text{tr}D^2 = J.
\]