Portfolios that Contain Risky Assets

Portfolio Models 2.

Covariance Matrices

C. David Levermore
University of Maryland, College Park

Math 420: *Mathematical Modeling*
January 31, 2017 version

Portfolios that Contain Risky Assets
Part I: Portfolio Models

1. Risk and Reward
2. Covariance Matrices
3. Markowitz Portfolios
4. Solvent Portfolios
5. Leveraged Portfolios
6. Basic Markowitz Portfolio Theory
7. Unlimited Portfolios with Risk-Free Assets
8. Long Portfolios without Risk-Free Assets
9. Long Portfolios with a Safe Investment
10. Limited Leverage Portfolios

Portfolio Models 2. Covariance Matrices

1. Introduction
2. Symmetry and Definiteness
3. Eigenpairs and Diagonalization
4. Statistical Interpretation
5. Principle Component Analysis

Covariance Matrices

Introduction. Suppose that we are considering return histories $\{r_i(d)\}_{d=1}^{D}$ for assets $i = 1, \cdots, N$ over a period of D trading days and assign day d a weight $w(d) > 0$ such that the weights $\{w(d)\}_{d=1}^{D}$ satisfy

$$\sum_{d=1}^{D} w(d) = 1 \,.$$

Then the return means, variances, and covariances are given by

$$\begin{aligned} m_i &= \sum_{d=1}^{D} w(d)\, r_i(d) \,, \\ v_{ij} &= \sum_{d=1}^{D} w(d) \Big(r_i(d) - m_i \Big) \Big(r_j(d) - m_j \Big) \,. \end{aligned} \tag{1}$$

The return history can be expressed as $\{\mathbf{r}(d)\}_{d=1}^{D}$ where

$$\mathbf{r}(d) = \begin{pmatrix} r_1(d) \\ \vdots \\ r_N(d) \end{pmatrix} .$$

The N-vector of return means $\mathbf{m}$ and the $N \times N$-matrix of return variances and covariances $\mathbf{V}$ then can be expressed as

$$\mathbf{m} = \begin{pmatrix} m_1 \\ \vdots \\ m_N \end{pmatrix} = \sum_{d=1}^{D} w(d)\, \mathbf{r}(d) \,,$$

$$\mathbf{V} = \begin{pmatrix} v_{11} & \cdots & v_{1N} \\ \vdots & \ddots & \vdots \\ v_{N1} & \cdots & v_{NN} \end{pmatrix} = \sum_{d=1}^{D} w(d) \Big(\mathbf{r}(d) - \mathbf{m}\Big) \Big(\mathbf{r}(d) - \mathbf{m}\Big)^{\mathsf{T}} .$$

We call $\mathbf{V}$ the *covariance matrix*. It also is called the *variance/covariance matrix* or the *variance matrix*. Here we give some properties of $\mathbf{V}$ that will be used to extract statistical information from it.

Symmetry and Definiteness. The most important properties of $\mathbf{V}$ are that *it is always symmetric* and that *it is almost always positive definite*. These properties are taught in elementary linear algebra courses, but are so important that we review them here.

Definition 1. A real $N \times N$-matrix $\mathbf{A}$ is said to be symmetric if $\mathbf{A}^{\mathsf{T}} = \mathbf{A}$. It is said to be nonnegative definite if

$$\mathbf{x}^{\mathsf{T}} \mathbf{A} \mathbf{x} \geq 0 \quad \text{for every } \mathbf{x} \in \mathbb{R}^N .$$

It is said to be positive definite if

$$\mathbf{x}^{\mathsf{T}} \mathbf{A} \mathbf{x} > 0 \quad \text{for every nonzero } \mathbf{x} \in \mathbb{R}^N .$$

Remarks. Clearly, every positive definite matrix is nonnegative definite. A nonnegative matrix is positive definite if and only if

$$\mathbf{x}^{\mathsf{T}} \mathbf{A} \mathbf{x} = 0 \quad \Longrightarrow \quad \mathbf{x} = \mathbf{0} . \tag{2}$$

Fact 1. The covariance matrix $\mathbf{V}$ is symmetric.

Proof. It is clear from (1) that $v_{ij} = v_{ji}$, whereby $\mathbf{V} = \mathbf{V}^\mathsf{T}$. $\square$

Fact 2. The covariance matrix $\mathbf{V}$ is nonegative definite.

Proof. Let $\mathbf{x} \in \mathbb{R}^N$ be arbitrary. Then

$$\begin{aligned}
\mathbf{x}^\mathsf{T}\mathbf{V}\mathbf{x} &= \mathbf{x}^\mathsf{T}\left(\sum_{d=1}^{D} w(d)\left(\mathbf{r}(d)-\mathbf{m}\right)\left(\mathbf{r}(d)-\mathbf{m}\right)^\mathsf{T}\right)\mathbf{x} \\
&= \sum_{d=1}^{D} w(d)\,\mathbf{x}^\mathsf{T}\left(\mathbf{r}(d)-\mathbf{m}\right)\left(\mathbf{r}(d)-\mathbf{m}\right)^\mathsf{T}\mathbf{x} \\
&= \sum_{d=1}^{D} w(d)\left(\left(\mathbf{r}(d)-\mathbf{m}\right)^\mathsf{T}\mathbf{x}\right)^2 \geq 0\,.
\end{aligned}$$

$\square$

Fact 3. The covariance matrix $\mathbf{V}$ is positive definite if and only if the vectors $\{\mathbf{r}(d) - \mathbf{m}\}_{d=1}^{D}$ span $\mathbb{R}^N$.

Proof. Because $w(d) > 0$, the calculation in the previous proof shows that $\mathbf{x}^{\mathsf{T}}\mathbf{V}\mathbf{x} = 0$ if and only if

$$\big(\mathbf{r}(d) - \mathbf{m}\big)^{\mathsf{T}}\mathbf{x} = 0 \quad \text{for every } d = 1, \cdots, D\,. \tag{3}$$

First, suppose that $\mathbf{V}$ is not positive definite. Then by (2) there exists an $\mathbf{x} \in \mathbb{R}^N$ such that $\mathbf{x}^{\mathsf{T}}\mathbf{V}\mathbf{x} = 0$ and $\mathbf{x} \neq \mathbf{0}$. This implies by (3) that the vectors $\{\mathbf{r}(d) - \mathbf{m}\}_{d=1}^{D}$ lie in the hyperplane orthogonal (normal) to $\mathbf{x}$. Therefore the vectors $\{\mathbf{r}(d) - \mathbf{m}\}_{d=1}^{D}$ do not span $\mathbb{R}^N$.

Conversely, suppose that the vectors $\{\mathbf{r}(d) - \mathbf{m}\}_{d=1}^{D}$ do not span $\mathbb{R}^N$. Then there must be a nonzero vector $\mathbf{x}$ that is orthogonal to their span. This implies that $\mathbf{x}$ satisfies (3), whereby $\mathbf{x}^{\mathsf{T}}\mathbf{V}\mathbf{x} = 0$. Therefore $\mathbf{V}$ is not positive definite by (2). $\square$

Remark. The set of vectors $\{\mathbf{r}(d) - \mathbf{m}\}_{d=1}^{D}$ can span $\mathbb{R}^N$ only if $D \geq N$. Therefore we require that $D \geq N$.

Remark. In practice D will be much larger than N. In the homework and projects for this course usually $N \leq 10$ while $D \geq 42$ (often $D = 252$). When D is so much greater than N the covariance matrix $\mathbf{V}$ will almost always be positive definite.

Remark. If $\{\mathbf{r}(d) - \mathbf{m}\}_{d=1}^{D}$ spans $\mathbb{R}^N$ then $\{\mathbf{r}(d)\}_{d=1}^{D}$ also spans $\mathbb{R}^N$. However, the converse need not hold. A counterexample for $N = 2$ and any $D \geq 2$ can be constructed as follows. Let $\{\mathbf{m}, \mathbf{n}\}$ span $\mathbb{R}^2$. Let $\mathbf{r}(d) = \mathbf{m} + h(d)\mathbf{n}$ where $h(d) \neq 0$ and

$$\sum_{d=1}^{D} w(d)h(d) = 0\,.$$

Then $\{\mathbf{r}(d)\}_{d=1}^{D}$ spans $\mathbb{R}^2$ while $\{\mathbf{r}(d) - \mathbf{m}\}_{d=1}^{D}$ does not span $\mathbb{R}^2$.

Eigenpairs and Diagonalization. Let us recall from linear algebra that an *eigenpair* $(\lambda, \mathbf{q})$ of a real $N \times N$ matrix $\mathbf{A}$ is a scalar λ (possibly complex) and a nonzero vector $\mathbf{q}$ (possibly with complex entries) such that

$$\mathbf{A}\mathbf{q} = \lambda \mathbf{q} \,. \tag{4}$$

An eigenpair is called a real eigenpair when λ and every entry of $\mathbf{q}$ is real.

An important fact from linear algebra is that if $\mathbf{A}$ is symmetric then it has N real eigenpairs

$$(\lambda_1, \mathbf{q}_1)\,, \quad (\lambda_2, \mathbf{q}_2)\,, \quad \cdots \quad (\lambda_N, \mathbf{q}_N)\,, \tag{5}$$

such that the eigenvectors $\{\mathbf{q}_i\}_{i=1}^{N}$ are an orthonormal set. This means that they satisfy the orthonormality conditions

$$\mathbf{q}_i^{\top} \mathbf{q}_j = \delta_{ij} \equiv \begin{cases} 1 & \text{if } i = j\,, \\ 0 & \text{if } i \neq j\,. \end{cases} \tag{6}$$

Because the $\{\mathbf{q}_i\}_{i=1}^N$ satisfy the orthonormality conditions (6), they form an orthonormal basis of $\mathbb{R}^N$. Every $\mathbf{x} \in \mathbb{R}^N$ can be expanded as

$$\mathbf{x} = \mathbf{q}_1 \, \mathbf{q}_1^{\mathsf{T}} \mathbf{x} + \mathbf{q}_2 \, \mathbf{q}_2^{\mathsf{T}} \mathbf{x} + \cdots + \mathbf{q}_N \, \mathbf{q}_N^{\mathsf{T}} \mathbf{x} \,. \tag{7}$$

The numbers $\{\mathbf{q}_i^{\mathsf{T}} \mathbf{x}\}_{i=1}^N$ are called the coordinates of $\mathbf{x}$ for the orthonormal basis $\{\mathbf{q}_i\}_{i=1}^N$. The square of the Euclidean norm of $\mathbf{x}$ is given by

$$\|\mathbf{x}\|^2 = \mathbf{x}^{\mathsf{T}} \mathbf{x} = \left(\mathbf{q}_1^{\mathsf{T}} \mathbf{x}\right)^2 + \left(\mathbf{q}_2^{\mathsf{T}} \mathbf{x}\right)^2 + \cdots + \left(\mathbf{q}_N^{\mathsf{T}} \mathbf{x}\right)^2 . \tag{8}$$

Because the $\{\mathbf{q}_i\}_{i=1}^N$ are eigenvectors of $\mathbf{A}$, we see from (7) that

$$\begin{aligned} \mathbf{A}\mathbf{x} &= \mathbf{A}\mathbf{q}_1 \, \mathbf{q}_1^{\mathsf{T}} \mathbf{x} + \mathbf{A}\mathbf{q}_2 \, \mathbf{q}_2^{\mathsf{T}} \mathbf{x} + \cdots + \mathbf{A}\mathbf{q}_N \, \mathbf{q}_N^{\mathsf{T}} \mathbf{x} \\ &= \lambda_1 \mathbf{q}_1 \, \mathbf{q}_1^{\mathsf{T}} \mathbf{x} + \lambda_2 \mathbf{q}_2 \, \mathbf{q}_2^{\mathsf{T}} \mathbf{x} + \cdots + \lambda_N \mathbf{q}_N \, \mathbf{q}_N^{\mathsf{T}} \mathbf{x} \,. \end{aligned} \tag{9}$$

Hence, the $\{\lambda_i \mathbf{q}_i^{\mathsf{T}} \mathbf{x}\}_{i=1}^N$ are the coordinates of $\mathbf{A}\mathbf{x}$ for the orthonormal basis $\{\mathbf{q}_i\}_{i=1}^N$. Therefore by (8) we have

$$\|\mathbf{A}\mathbf{x}\|^2 = \lambda_1^2 \left(\mathbf{q}_1^{\mathsf{T}} \mathbf{x}\right)^2 + \lambda_2^2 \left(\mathbf{q}_2^{\mathsf{T}} \mathbf{x}\right)^2 + \cdots + \lambda_N^2 \left(\mathbf{q}_N^{\mathsf{T}} \mathbf{x}\right)^2 . \tag{10}$$

Moreover, $\mathbf{A}$ can be expressed in the factored form $\mathbf{A} = \mathbf{Q}\mathbf{D}\mathbf{Q}^{\mathsf{T}}$ where $\mathbf{D}$ and $\mathbf{Q}$ are the real $N \times N$ matrices constructed from the eigenpairs (5) as

$$\mathbf{D} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_N \end{pmatrix}, \qquad \mathbf{Q} = \begin{pmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_N \end{pmatrix}. \tag{11}$$

Because the matrix $\mathbf{D}$ is a *diagonal matrix*, this factorization of $\mathbf{A}$ is called diagonalization. The orthonormality conditions (6) satisfied by the vectors $\{\mathbf{q}_i\}_{i=1}^{N}$ imply that $\mathbf{Q}$ is an orthogonal matrix. This means that $\mathbf{Q}$ satisfies

$$\mathbf{Q}^{\mathsf{T}}\mathbf{Q} = \mathbf{I} = \mathbf{Q}\mathbf{Q}^{\mathsf{T}}.$$

The relation $\mathbf{Q}^{\mathsf{T}}\mathbf{Q} = \mathbf{I}$ is a recasting of the orthonormality conditions (6). The relation $\mathbf{I} = \mathbf{Q}\mathbf{Q}^{\mathsf{T}}$ is equivalent to $\mathbf{x} = \mathbf{Q}\mathbf{Q}^{\mathsf{T}}\mathbf{x}$, which is a recasting of expansion (7). These relations show that $\mathbf{Q}$ and $\mathbf{Q}^{\mathsf{T}}$ are inverses of each other — i.e. that $\mathbf{Q}^{-1} = \mathbf{Q}^{\mathsf{T}}$ and that $\mathbf{Q}^{-\mathsf{T}} = \mathbf{Q}$.

Other important facts from linear algebra are that if $\mathbf{A}$ is a real symmetric matrix then:

- it is nonnegative definite if and only if all its eigenvalues are nonnegative;

- it is positive definite if and only if all its eigenvalues are positive.

Proof. The ($\Longrightarrow$) directions of these characterizations follow from the fact that if $(\lambda, \mathbf{q})$ is an eigenpair of $\mathbf{A}$ that is normalized so that $\mathbf{q}^\mathsf{T}\mathbf{q} = 1$ then

$$\lambda = \lambda \mathbf{q}^\mathsf{T}\mathbf{q} = \mathbf{q}^\mathsf{T}(\lambda \mathbf{q}) = \mathbf{q}^\mathsf{T}(\mathbf{A}\mathbf{q}) = \mathbf{q}^\mathsf{T}\mathbf{A}\mathbf{q}\,.$$

The ($\Longleftarrow$) directions of these characterizations use the full power of the orthonormality conditions (6) as embodied by expansion (9),

$$\mathbf{A}\mathbf{x} = \lambda_1 \mathbf{q}_1 \, \mathbf{q}_1^{\mathsf{T}}\mathbf{x} + \lambda_2 \mathbf{q}_2 \, \mathbf{q}_2^{\mathsf{T}}\mathbf{x} + \cdots + \lambda_N \mathbf{q}_N \, \mathbf{q}_N^{\mathsf{T}}\mathbf{x} \,.$$

By taking the scalar product of this expansion with $\mathbf{x}$ we obtain

$$\mathbf{x}^{\mathsf{T}}\mathbf{A}\mathbf{x} = \lambda_1 \big(\mathbf{q}_1^{\mathsf{T}}\mathbf{x}\big)^2 + \lambda_2 \big(\mathbf{q}_2^{\mathsf{T}}\mathbf{x}\big)^2 + \cdots + \lambda_N \big(\mathbf{q}_N^{\mathsf{T}}\mathbf{x}\big)^2 \,.$$

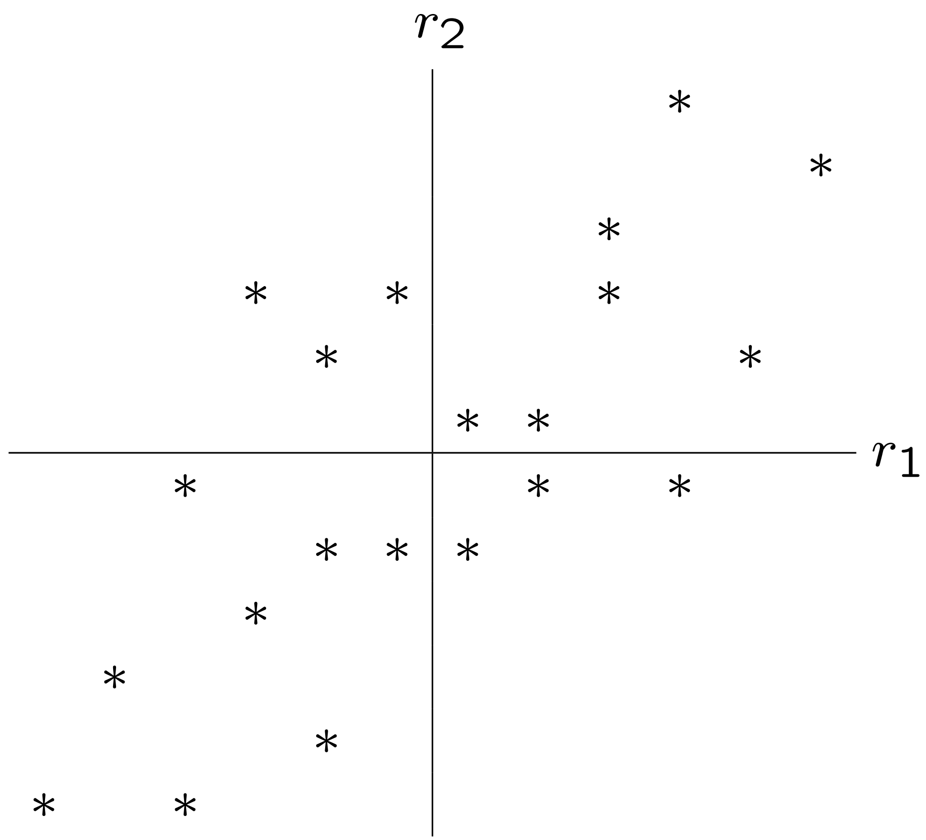
It is thereby clear that:

- if $\lambda_i \geq 0$ for every $i = 1, \cdots, N$ then $\mathbf{A}$ is nonnegative definite;

- if $\lambda_i > 0$ for every $i = 1, \cdots, N$ then $\mathbf{A}$ is positive definite.

This proves the ($\Longleftarrow$) directions of the characterizations. $\square$

Statistical Interpretation. Let us consider the case $N = 2$ and $D = 21$. Suppose that when the return history $\{(r_1(d), r_2(d))\}_{d=1}^{21}$ is plotted as points in the r_1r_2-plane we obtain



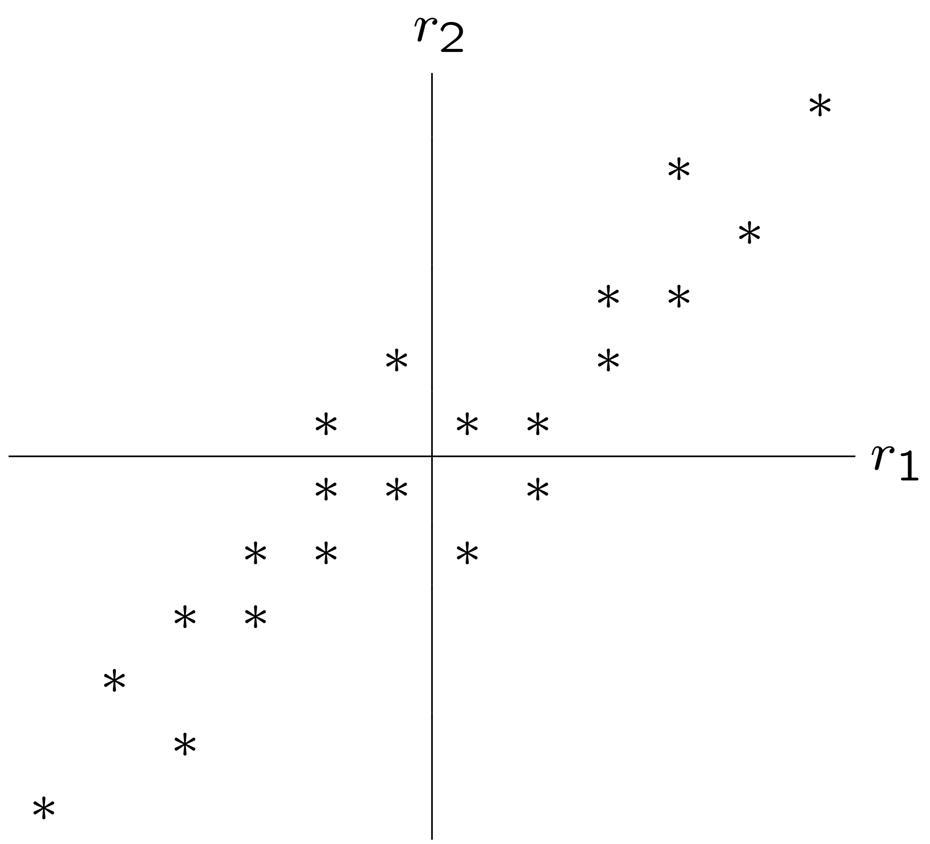
This so-called *scatter plot* shows a distribution of points clustered about the origin in a way that favors the first and third quadrants. These properties are quantified by the return mean vector $\mathbf{m}$ and return covariance matrix $\mathbf{V}$ computed with uniform weights as

$$\mathbf{m} = \begin{pmatrix} m_1 \\ m_2 \end{pmatrix} = \frac{1}{21} \sum_{d=1}^{21} \begin{pmatrix} r_1(d) \\ r_2(d) \end{pmatrix} ,$$

$$\mathbf{V} = \begin{pmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{pmatrix} = \frac{1}{21} \sum_{d=1}^{21} \begin{pmatrix} \tilde{r}_1(d)^2 & \tilde{r}_1(d)\tilde{r}_2(d) \\ \tilde{r}_2(d)\tilde{r}_1(d) & \tilde{r}_2(d)^2 \end{pmatrix} ,$$

where $\tilde{r}_1(d) = r_1(d) - m_1$ and $\tilde{r}_2(d) = r_2(d) - m_2$. The vector $\mathbf{m}$ gives the center of the cluster. It lies in the third quadrant close to the origin. The matrix $\mathbf{V}$ will have eigenvectors that are roughly parallel to $\nearrow$ and to $\nwarrow$. The eigenvalue associated with $\nearrow$ will be greater than the one associated with $\nwarrow$. This is how $\mathbf{m}$ and $\mathbf{V}$ tell us that the points are clustered about the origin in a way that favors the first and third quadrants.

Now suppose that when the return history $\{(r_1(d), r_2(d))\}_{d=1}^{21}$ is plotted as points in the r_1r_2-plane we obtain



In this scatter plot r_1 and r_2 are more highly correlated than in the first. The vector $\mathbf{m}$ is almost the same as it was for the first scatter plot. It lies in the third quadrant close to the origin. The matrix $\mathbf{V}$ again has eigenvectors that are roughly parallel to $\nearrow$ and to $\nwarrow$. However now the eigenvalue associated with $\nearrow$ is *very much greater* than the one associated with $\nwarrow$.

Both the scatter plot and the analysis of $\mathbf{m}$ and $\mathbf{V}$ suggest that the points $\{(r_1(d), r_2(d))\}_{d=1}^{21}$ cluster along a line. Let $\mathbf{q}$ designate the eigenvector associated with the largest eigenvalue of $\mathbf{V}$. If $\mathbf{q}$ is proportional to $(1, s)$ then the line in the r_1r_2-plane that the points cluster along is

$$r_2 - m_2 = s(r_1 - m_1)\,.$$

This suggests $r_2(d)$ could be modeled as

$$r_2(d) - m_2 = s\Big(r_1(d) - m_1\Big) + z(d)\,,$$

where $z(d)$ are small random numbers that on average sum to zero.

Principle Component Analysis. Scatter plots become harder to visualize as N grows beyond 3. However the analysis of $\mathbf{m}$ and $\mathbf{V}$ can be carried out easily for much larger N. In statistics the eigenpair analysis of the covariance matrix $\mathbf{V}$ is called *Principle Component Analysis (PCA)*.

A principle component analysis of $\mathbf{V}$ yields N eigenpairs

$$(\lambda_1, \mathbf{q}_1)\,, \quad (\lambda_2, \mathbf{q}_2)\,, \quad \cdots\,, \quad (\lambda_N, \mathbf{q}_N)\,. \tag{12}$$

The eigenvalues will almost always be distinct, in which case we will order them as

$$\lambda_1 > \lambda_2 > \cdots > \lambda_N > 0\,. \tag{13}$$

In this case the eigenvectors will be unique up to a nonzero factor. If they are normalized so that $\|\mathbf{q}_i\| = 1$ then they are unique up to a factor of ± 1 and $\{\mathbf{q}_i\}_{i=1}^N$ will be an orthonormal basis of $\mathbb{R}^N$.

Let $\mathbf{D}$ and $\mathbf{Q}$ be the diagonal and orthogonal matrices constructed from the eigenpairs (12) as in (11). Then $\mathbf{V} = \mathbf{Q}\mathbf{D}\mathbf{Q}^{\mathsf{T}}$ and $\mathbf{Q}^{\mathsf{T}}\mathbf{Q} = \mathbf{Q}\mathbf{Q}^{\mathsf{T}} = \mathbf{I}$.

Then the underlying return history $\{\mathbf{r}(d)\}_{d=1}^{D}$ can be transformed into the history $\{\mathbf{p}(d)\}_{d=1}^{D}$ where $\mathbf{p}(d) = \mathbf{Q}^{\mathsf{T}}\mathbf{r}(d)$. The entries of $\mathbf{p}(d)$ are called the *principle components* of $\mathbf{r}(d)$. Their mean vector is given by

$$\sum_{d=1}^{D} w(d)\mathbf{p}(d) = \sum_{d=1}^{D} w(d)\mathbf{Q}^{\mathsf{T}}\mathbf{r}(d) = \mathbf{Q}^{\mathsf{T}}\left(\sum_{d=1}^{D} w(d)\mathbf{r}(d)\right) = \mathbf{Q}^{\mathsf{T}}\mathbf{m}\,.$$

Similarly, their covariance matrix is given by

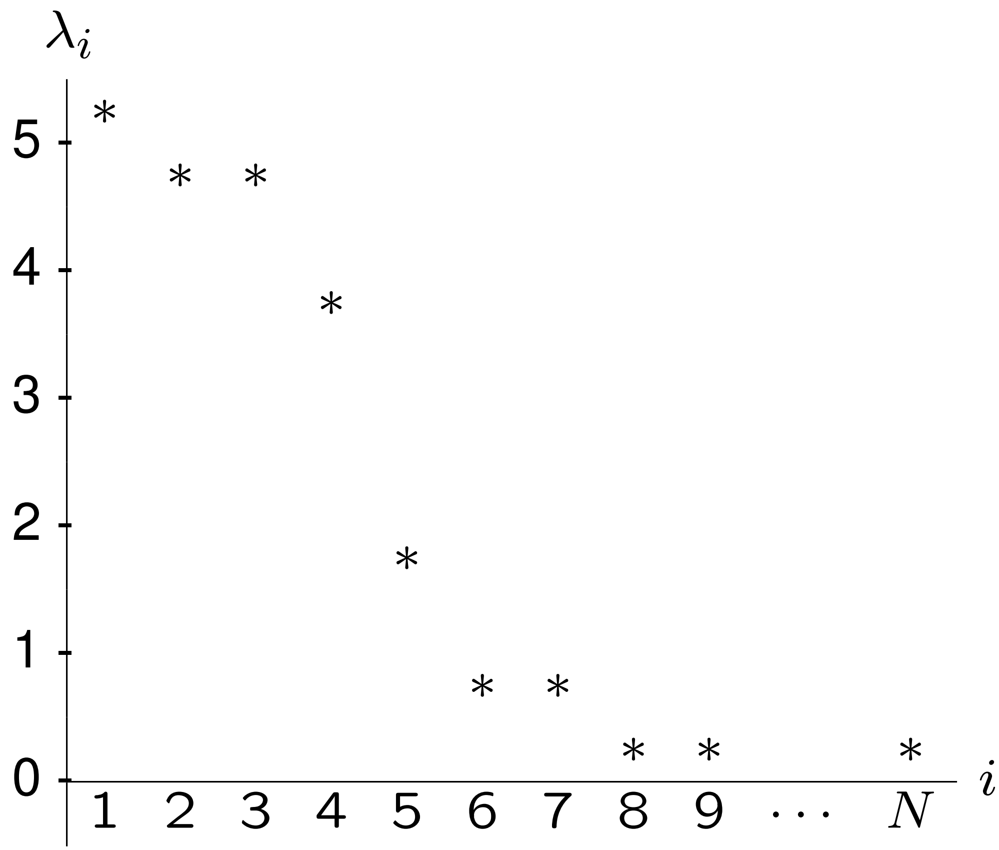
$$\begin{aligned}
&\sum_{d=1}^{D} w(d)\big(\mathbf{p}(d) - \mathbf{Q}^{\mathsf{T}}\mathbf{m}\big)\big(\mathbf{p}(d) - \mathbf{Q}^{\mathsf{T}}\mathbf{m}\big)^{\mathsf{T}} \\
&= \mathbf{Q}^{\mathsf{T}}\left(\sum_{d=1}^{D} w(d)\big(\mathbf{r}(d) - \mathbf{m}\big)\big(\mathbf{r}(d) - \mathbf{m}\big)^{\mathsf{T}}\right)\mathbf{Q} \\
&= \mathbf{Q}^{\mathsf{T}}\mathbf{V}\mathbf{Q} = \mathbf{Q}^{\mathsf{T}}\big(\mathbf{Q}\mathbf{D}\mathbf{Q}^{\mathsf{T}}\big)\mathbf{Q} = \big(\mathbf{Q}^{\mathsf{T}}\mathbf{Q}\big)\mathbf{D}\big(\mathbf{Q}^{\mathsf{T}}\mathbf{Q}\big) = \mathbf{D}\,.
\end{aligned}$$

Because $\mathbf{D}$ is a diagonal matrix, the covariance of distinct entries of $\mathbf{p}(d)$ vanishes. Because the i^{th} entry of $\mathbf{p}(d)$ is $\mathbf{q}_i^{\top}\mathbf{r}(d)$, its variance is λ_i. *Therefore PCA can be viewed as an orthogonal coordinate transformation that maps the data into new coordinates (the principle components) that are uncorrelated and such that the first entry has the largest variance, the second entry has the second largest variance, and so on.*

Remark. The vectors $\mathbf{q}_i$ are called the *principle component coefficients* because they are the vectors whose scalar product with the data $\mathbf{r}(d)$ gives the principle components. They are also called *loadings*.

One application of PCA is to identify possible lower dimensional models that capture the bulk of the variation in the data. The dimension of such a model is read off by selecting a subset of the largest eigenvalues of $\mathbf{V}$.

For example, a plot of λ_i versus i might look like the figure below. It shows that the underlying data has four major dimensions and that eigenvalues are negligible for $i \geq 8$. This suggests that the data might be captured well with a 4, 5, or 7 dimensional model.



Remark. The dimension obtained in this way gives an upper bound on the actual dimension of the data, which can be lower when it satisfies an approximate nonlinear relationship. Such a relationship is illustrated below for two dimensional data.