23 Principal Component Analysis in Action: From Theory to Real Data

23.1 From Ellipses to Real Data: Seeing PCA in Action

23.1.1 A Quick Recap: The Geometry Behind PCA

In the previous chapter, we explored Principal Component Analysis (PCA) from a geometric perspective, using an ellipse to illustrate how PCA finds the directions that best describe data variation, as illustrated in Figure 1. Conceptually, this means "straightening out" a rotated ellipse so that its axes align with the directions of greatest variance. In this section, we shift from intuition to practice by applying PCA to real financial data.

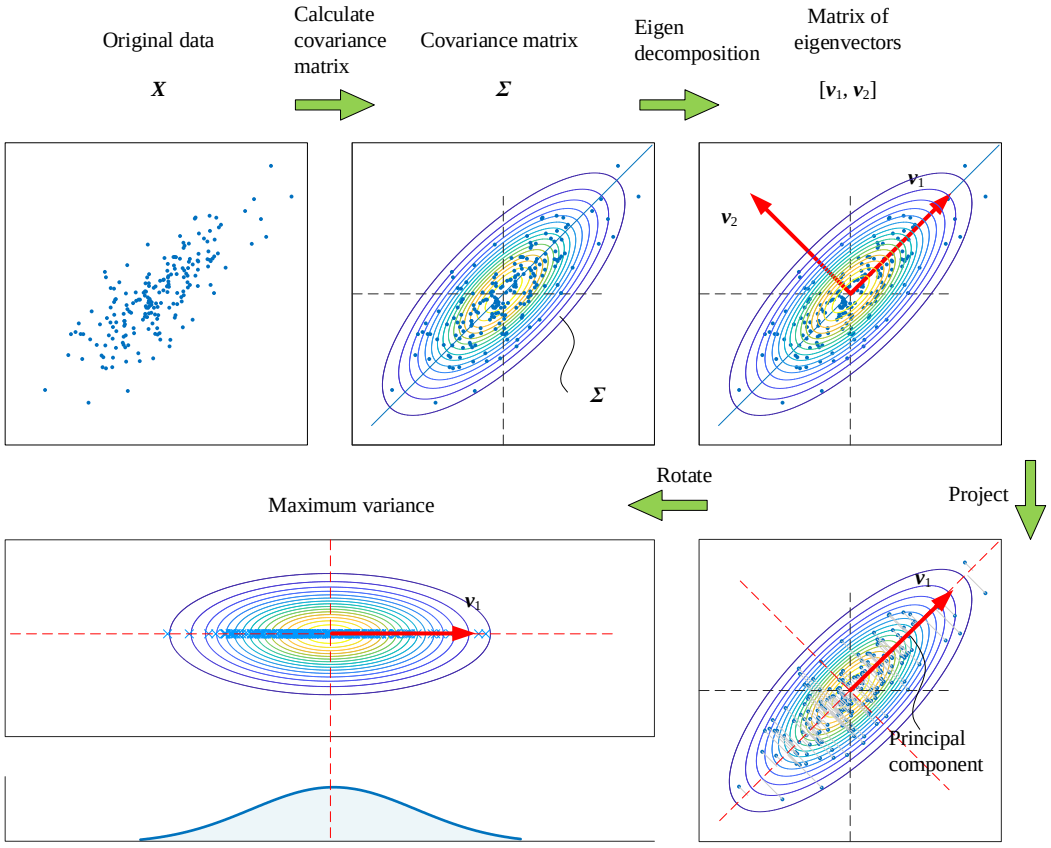

Figure 1. Principal Component Analysis through Eigenvalue Decomposition of the Covariance Matrix: A Geometric Perspective

23.1.2 Interest Rate Changes as a Real-World Dataset

Figure 2 shows the daily changes in interest rates for various maturities (for example, 0.5-year, 1-year, 2-year, and so on). Each column of the data matrix $\boldsymbol{X}$ corresponds to a maturity, and each row represents one day's change relative to the previous day.

The matrix $\boldsymbol{X}$ has 248 rows and 8 columns—tall and narrow—meaning we have many daily observations but only a few interest rate maturities. The values in Figure 2 are shown as decimal numbers; multiplying them by 100 would convert them to daily percentage changes. However, the plot alone does not clearly reveal whether different maturities move together or exhibit any common pattern.

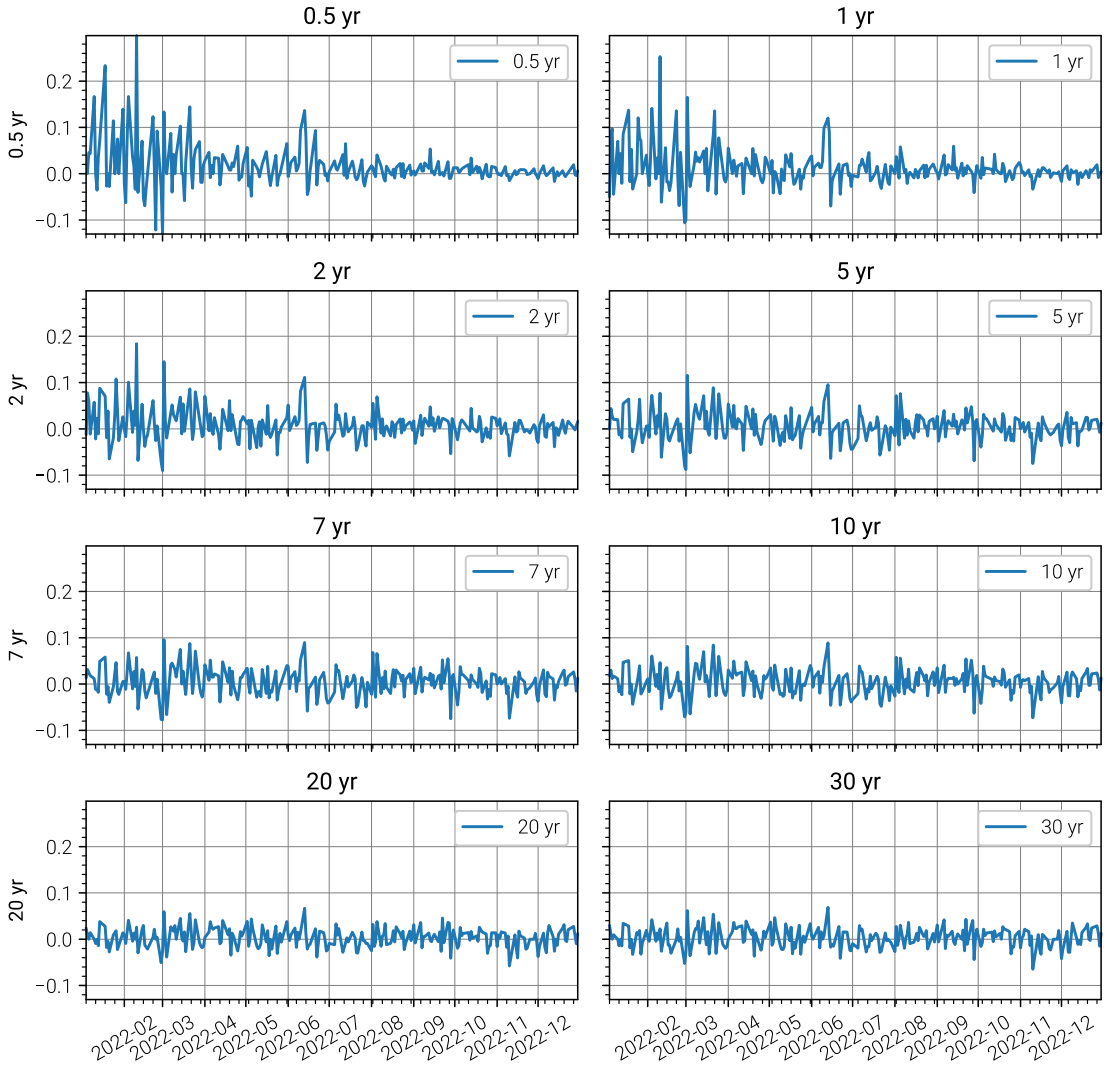

Figure 2. Daily changes in interest rates across different maturities. Figure generated by Ch23_01_PCA_IR_data.ipynb.

23.1.3 Do Interest Rates Move Together? First Visual Clues

To better understand their relationships, we turn to Figure 3. Figure 3 displays pairwise scatter plots of daily rate changes across maturities. We can immediately observe strong positive linear correlations—when short-term rates increase, longer-term rates tend to move in the same direction. This visual evidence suggests the presence of underlying common factors driving interest rate movements, which PCA can help uncover.

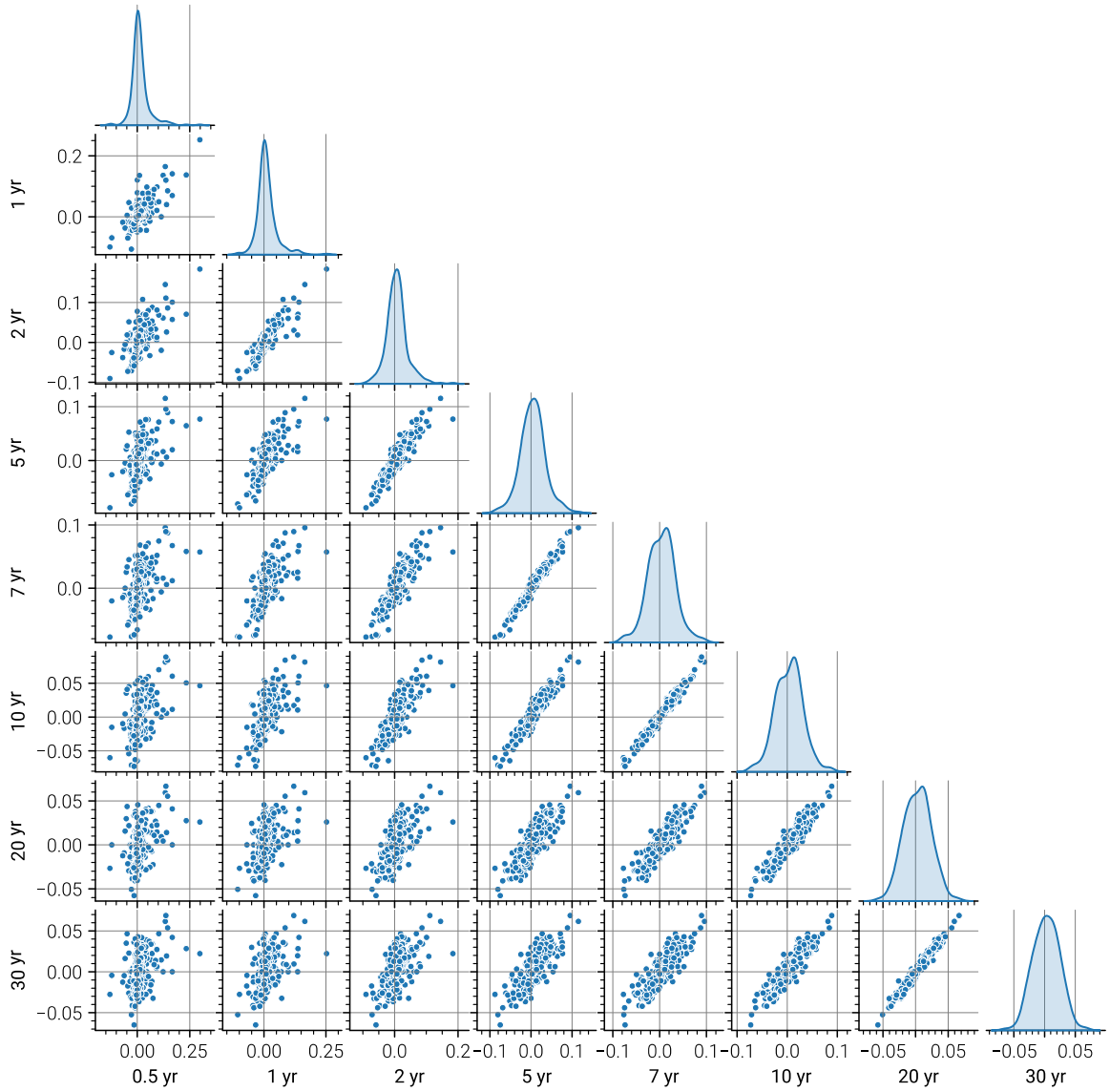

Figure 3. Pairwise scatter plots of daily rate changes across maturities (plots above the main diagonal omitted). Figure generated by Ch23_01_PCA_IR_data.ipynb.

23.2 The Covariance Matrix: Quantifying Co-Movements

23.2.1 Centering the Data and Building the Covariance Matrix

Before applying PCA, we first need to understand the structure of the covariance matrix, which captures how features vary together. After centering the data matrix $\boldsymbol{X}$ by subtracting its mean (column) vector $\boldsymbol{\mu}$, we obtain the centered matrix $\boldsymbol{X}_c$

$$\boldsymbol{X}_c = \boldsymbol{X} - \boldsymbol{\mu}^{\mathrm{T}} \tag{1}$$

This step shifts the data so that its mean lies at the origin, ensuring that PCA measures variation relative to the overall average rather than absolute values.

The covariance matrix can be viewed as a special type of **Gram matrix** computed from $\boldsymbol{X}_c$

$$\boldsymbol{\Sigma} = \frac{\boldsymbol{X}_c^{\mathrm{T}} \boldsymbol{X}_c}{n-1} \tag{2}$$

23.2.2 Variance, Covariance, and What They Mean

Mathematically, covariance matrix summarizes how much each pair of features co-varies. The diagonal elements represent the **variances** of individual features, indicating how much each one fluctuates over time, while the off-diagonal elements represent **covariances**, showing how strongly two features move together.

Figure 4 presents the covariance matrix of the interest rate change data. As we can see, the diagonal elements differ noticeably across maturities, suggesting that short-term and long-term interest rates exhibit different levels of volatility.

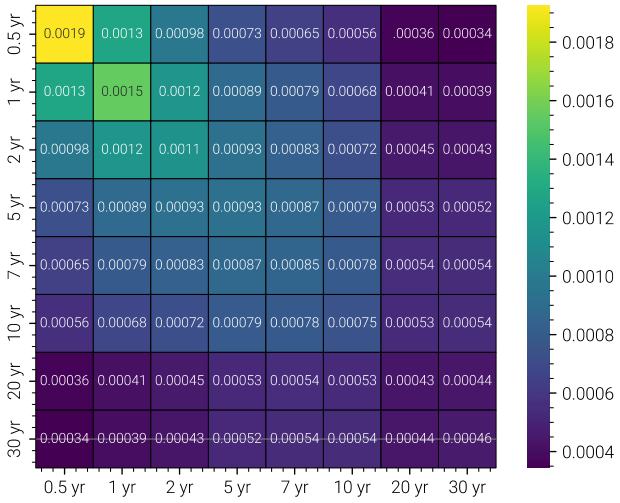

Figure 4. Heatmap of the covariance matrix for daily interest rate changes. Figure generated by Ch23_01_PCA_IR_data.ipynb.

23.2.3 Visualizing Volatility Across Maturities

To visualize this more clearly, Figure 5 shows the standard deviation (the square root of variance) of each maturity's rate changes. Some maturities fluctuate much more than others. As discussed earlier, when the scale of variation differs greatly among features, **standardizing** the data before performing PCA is essential. Standardization ensures that PCA focuses on shared patterns rather than being dominated by features with larger numerical ranges.

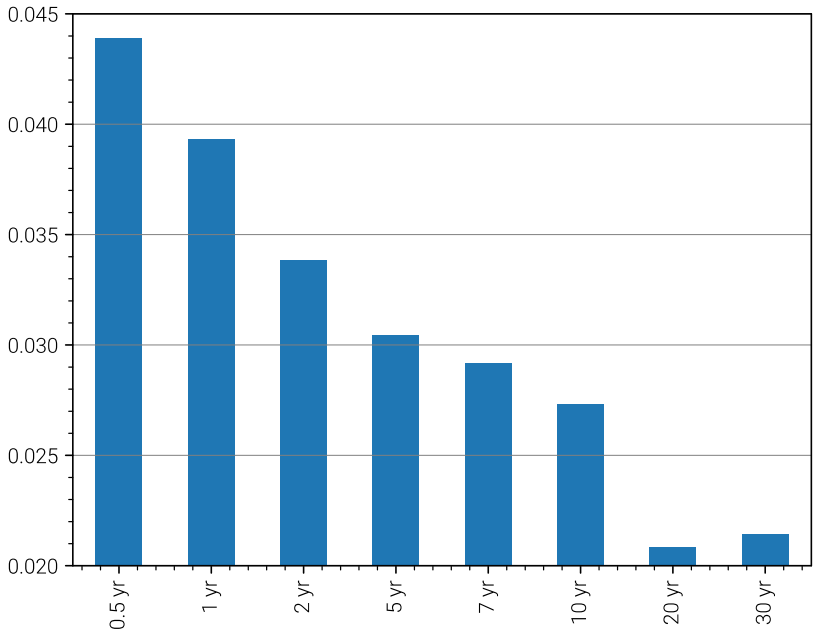

Figure 5. Standard deviation of daily rate changes across different maturities. Figure generated by Ch23_01_PCA_IR_data.ipynb.

23.3 Standardization: Putting All Features on Equal Footing

23.3.1 What Standardization Does and Why It's Necessary

Before performing PCA, it is often necessary to standardize the data so that all features are measured on the same scale. The standardized data matrix, denoted as $\boldsymbol{Z}$, is obtained by dividing each centered feature by its standard deviation

$$\boldsymbol{Z} = \boldsymbol{X}_c \boldsymbol{D}^{-1} = \left(\boldsymbol{X} - \boldsymbol{\mu}^{\mathrm{T}}\right)\boldsymbol{D}^{-1} \tag{3}$$

In other words, we remove the mean from each feature and then scale it so that its standard deviation equals one. The diagonal matrix $\boldsymbol{D}$ contains the standard deviations of each feature along its diagonal

$$\boldsymbol{D} = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_D \end{bmatrix} \tag{4}$$

The key idea behind standardization is to eliminate the effect of different units or magnitudes among features. For example, in financial data, short-term and long-term interest rate changes may have very different levels of volatility. If we do not standardize, PCA would give more weight to the features with larger numerical ranges, causing the first few principal components to be dominated by those variables.

23.3.2 Constructing the Standardized Data Matrix

Figure 6 shows the standardized time series. Each maturity's rate changes now have a mean of zero and a standard deviation of one. This transformation allows us to meaningfully compare patterns across maturities, since they are now on a common scale.

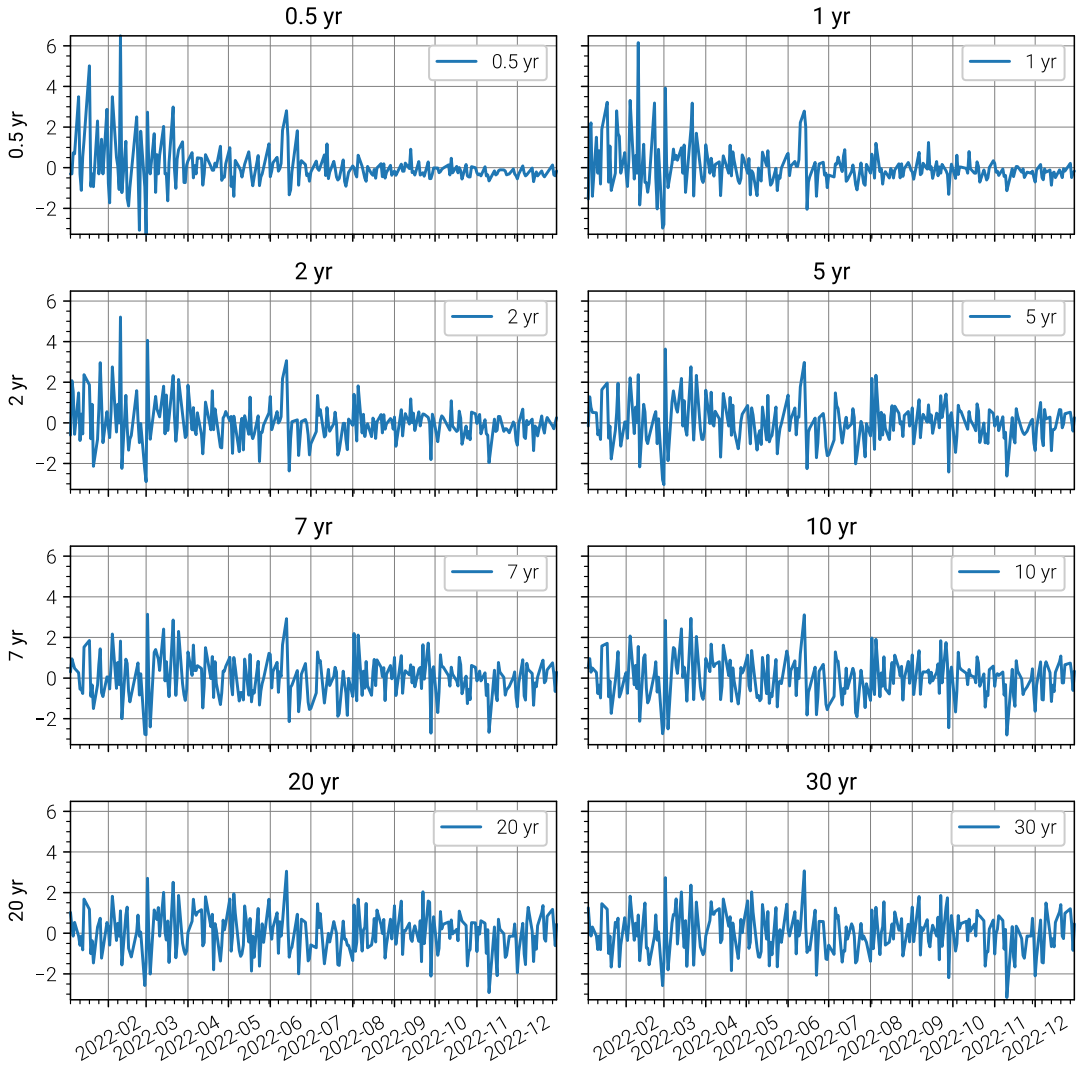

Figure 6. Standardized time series of daily rate changes (mean = 0, standard deviation = 1) . Figure generated by Ch23_01_PCA_IR_data.ipynb.

Figure 7 presents the pairwise scatter plots of the standardized data. Comparing Figure 3 and Figure 7, we can see that standardization does not alter the overall shape or structure of the data distribution—it simply shifts and rescales it so that all variables are centered at zero with the same spread. The relative relationships between samples remain unchanged. One important note is that standardization can be sensitive to outliers, since both the mean and standard deviation are influenced by extreme values. In practice, it is often helpful to detect and handle outliers before applying standardization.

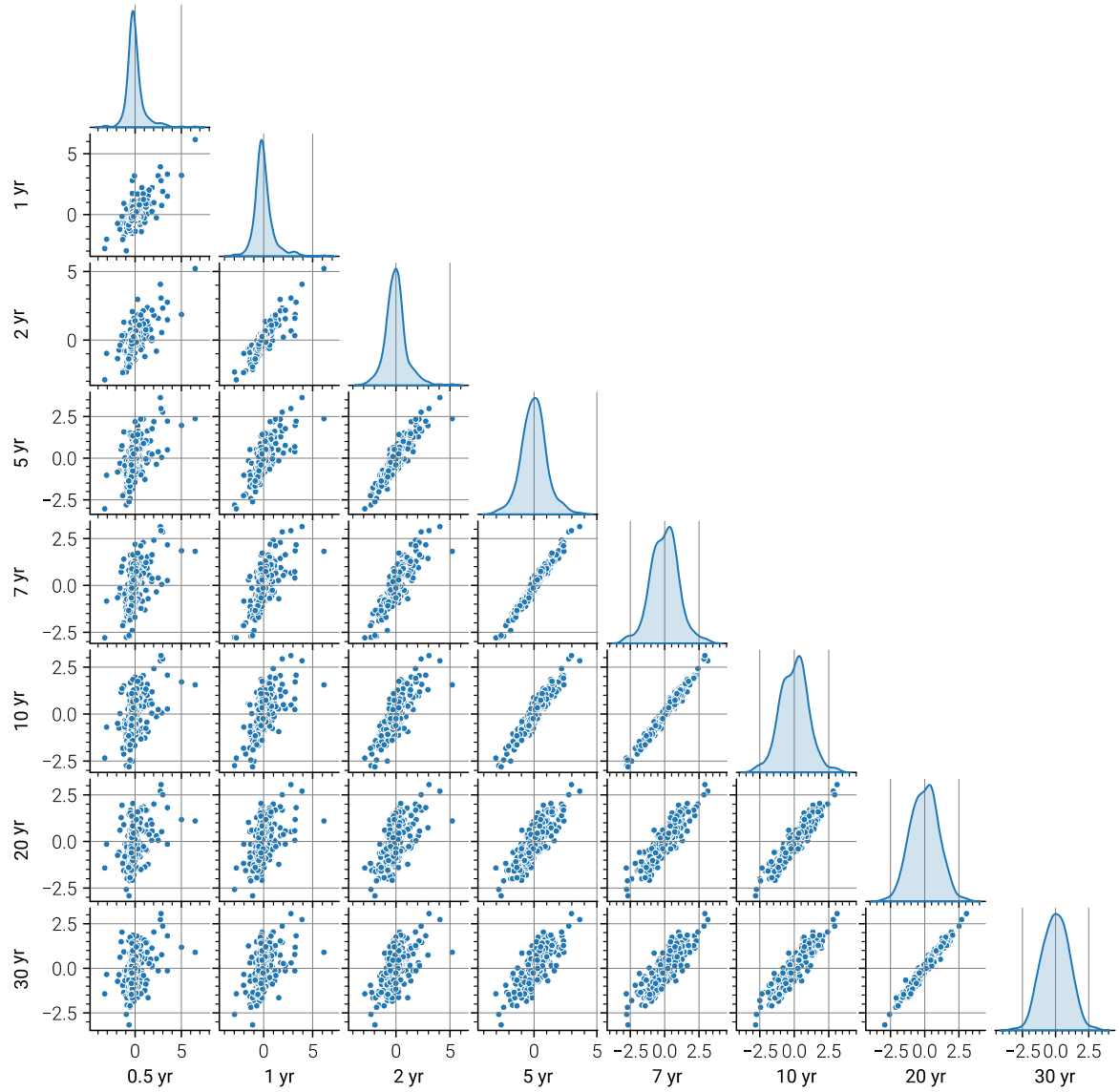

Figure 7. Pairwise scatter plots of standardized data (plots above the main diagonal omitted) . Figure generated by Ch23_01_PCA_IR_data.ipynb.

23.4 The Correlation Matrix: Measuring Linear Relationships

23.4.1 From Covariance to Correlation

The correlation matrix provides a standardized way to measure how strongly different variables move together. Each element in this matrix represents the linear correlation coefficient between two variables, which quantifies both the strength and direction of their linear relationship

$$\boldsymbol{P} = \begin{bmatrix} 1 & \rho_{1,2} & \cdots & \rho_{1,D} \\ \rho_{1,2} & 1 & \cdots & \rho_{2,D} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1,D} & \rho_{2,D} & \cdots & 1 \end{bmatrix} \tag{5}$$

The correlation coefficient ranges from −1 to 1: a value of 1 means the two variables move perfectly together in the same direction, −1 means they move perfectly in opposite directions, and 0 means there is no linear relationship between them.

Unlike covariance, the correlation coefficient is **unit-free** because it is computed after both variables have been standardized. This means that correlation reflects only the similarity of their movement patterns, regardless of their original units or scales.

Figure 8 displays the correlation matrix for the standardized interest rate data. Each off-diagonal value shows how two maturities co-move, while the diagonal elements are always 1, indicating that every feature is perfectly correlated with itself. In fact, the correlation matrix can be understood as the **covariance matrix of the standardized data**

$$\boldsymbol{P} = \boldsymbol{\Sigma}_z = \frac{\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z}}{n-1} \tag{6}$$

Gram matrix

Once each feature has been scaled to have zero mean and unit variance, the covariance between features becomes equivalent to their correlation coefficient.

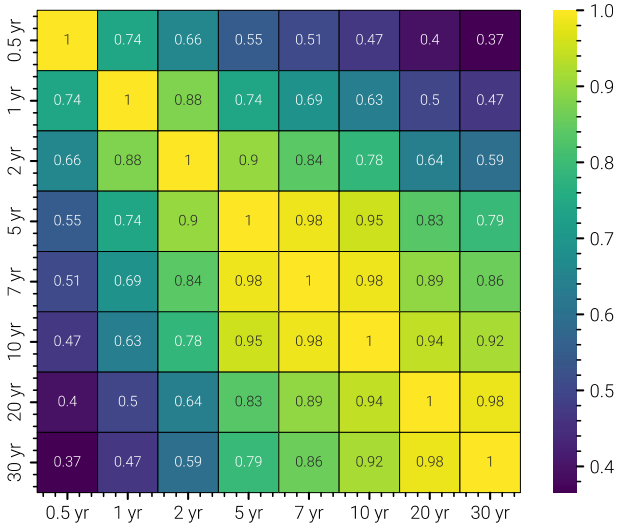

Figure 8. Correlation matrix of standardized interest rate changes. Figure generated by Ch23_01_PCA_IR_data.ipynb.

23.4.2 The Meaning of Correlation Coefficients

To build further intuition, consider projecting the standardized data $\boldsymbol{Z}$ onto one of the coordinate axes, say the first axis $\boldsymbol{e}_1$. This projection gives us a variable $\boldsymbol{z}_1$ representing the first feature

$$\boldsymbol{z}_1 = \boldsymbol{Z}\boldsymbol{e}_1 = \begin{bmatrix} \boldsymbol{z}_1 & \boldsymbol{z}_2 & \cdots & \boldsymbol{z}_D \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{7}$$

Since each column of $\boldsymbol{Z}$ has already been standardized, the variance of $\boldsymbol{z}_1$ equals 1

$$\mathrm{var}(\boldsymbol{z}_1) = \frac{\boldsymbol{z}_1^{\mathrm{T}}\boldsymbol{z}_1}{n-1} = \frac{(\boldsymbol{Z}\boldsymbol{e}_1)^{\mathrm{T}}\boldsymbol{Z}\boldsymbol{e}_1}{n-1} = \boldsymbol{e}_1^{\mathrm{T}}\frac{\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z}}{n-1}\boldsymbol{e}_1 = \boldsymbol{e}_1^{\mathrm{T}}\boldsymbol{P}\boldsymbol{e}_1 = 1 \tag{8}$$

and the same applies to all other features.

When we examine how $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ vary together, their covariance directly corresponds to the correlation coefficient between the first and second features

$$\operatorname{cov}(\boldsymbol{z}_1, \boldsymbol{z}_2) = \operatorname{cov}(\boldsymbol{z}_2, \boldsymbol{z}_1) = \frac{\boldsymbol{z}_2^{\mathrm{T}} \boldsymbol{z}_1}{n-1} = \frac{(\boldsymbol{Z}\boldsymbol{e}_2)^{\mathrm{T}} \boldsymbol{Z}\boldsymbol{e}_1}{n-1} = \boldsymbol{e}_2^{\mathrm{T}} \frac{\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z}}{n-1} \boldsymbol{e}_1 = \boldsymbol{e}_2^{\mathrm{T}} \boldsymbol{P} \boldsymbol{e}_1 = \rho_{1,2} \tag{9}$$

Thus, the correlation matrix summarizes all pairwise relationships in a compact, symmetric form. It serves as the foundation for PCA when we analyze standardized data, allowing us to identify common patterns among features without being influenced by differences in scale.

23.5 Eigenvalue Decomposition: Discovering the Directions of Maximum Variance

23.5.1 Spectral Decomposition of the Correlation Matrix

Once we have the correlation matrix $\boldsymbol{P}$, the next step in PCA is to perform eigenvalue decomposition. This step allows us to uncover the underlying directions in which the data varies most strongly. Because the correlation matrix is symmetric, its eigenvalue decomposition is also known as a spectral decomposition, which guarantees that all eigenvalues are real and that the eigenvectors form an orthogonal set.

Figure 9shows the eigenvalue decomposition of the correlation matrix $\boldsymbol{P}$:

$$\boldsymbol{P} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{\mathrm{T}} \tag{10}$$

Here, $\boldsymbol{V}$ is an orthogonal matrix, meaning that its columns (the eigenvectors) are mutually perpendicular and each has unit length. These eigenvectors define new coordinate axes—directions along which the data exhibits the greatest variability.

The diagonal matrix $\boldsymbol{\Lambda}$ contains the eigenvalues, each representing the amount of variance explained by its corresponding eigenvector.

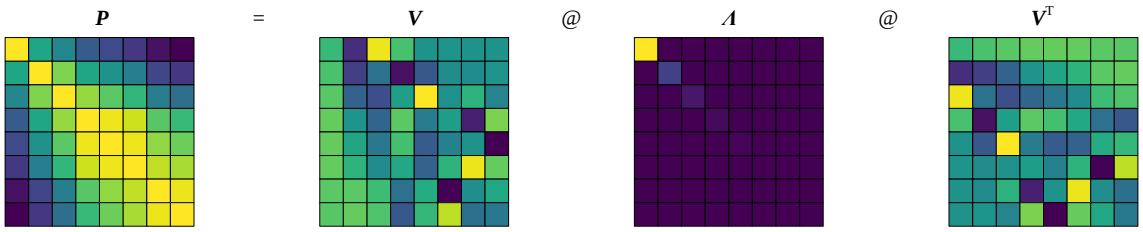

Figure 9. Eigenvalue decomposition of the correlation matrix (spectral decomposition form). Figure generated by Ch23_01_PCA_IR_data.ipynb.

23.5.2 Eigenvectors as New Coordinate Axes

When we project the standardized data $\boldsymbol{Z}$ onto one of these eigenvector directions (say $\boldsymbol{v}_1$), we obtain a new variable

$$\boldsymbol{y}_1 = \boldsymbol{Z}\boldsymbol{v}_1 \tag{11}$$

The variance of $\boldsymbol{y}_1$ equals the first eigenvalue,

$$\text{var}(\boldsymbol{y}_1) = \frac{\boldsymbol{y}_1^{\mathrm{T}}\boldsymbol{y}_1}{n-1} = \frac{(\boldsymbol{Z}\boldsymbol{v}_1)^{\mathrm{T}}\boldsymbol{Z}\boldsymbol{v}_1}{n-1} = \boldsymbol{v}_1^{\mathrm{T}}\frac{\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z}}{n-1}\boldsymbol{v}_1 = \boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{P}\boldsymbol{v}_1 = \lambda_1 \tag{12}$$

which is the largest among all. This means that $\boldsymbol{v}_1$ corresponds to the first principal component, the direction that captures the most variance in the data. In other words, PCA finds the rotation of the coordinate system that best aligns with the natural structure of the data, with the first few directions capturing the most important patterns.

The matrix $\boldsymbol{V}$ is sometimes called the loading matrix, and each of its column vectors is referred to as a loading vector. In some conventions, loadings are defined as $\boldsymbol{v}_i$ scaled by the square root of their corresponding eigenvalue, which adjusts for the magnitude of the variance explained by each component.

23.5.3 Loadings, Factor Scores, and the New Basis

The factor scores (also known as principal component scores) are the coordinates of the samples in this new basis. They are obtained by projecting the standardized data $\boldsymbol{Z}$ onto the eigenvector matrix $\boldsymbol{V}$, giving:

$$\boldsymbol{Y} = \boldsymbol{Z}\boldsymbol{V} \tag{13}$$

This operation transforms the data from the original feature space into the principal component space, where the new variables (the columns of $\boldsymbol{Y}$) are uncorrelated and arranged in order of decreasing variance.

The covariance matrix of $\boldsymbol{Y}$ is simply the diagonal matrix of eigenvalues,

$$\boldsymbol{\Sigma}_{\boldsymbol{Y}} = \frac{\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y}}{n-1} = \frac{(\boldsymbol{Z}\boldsymbol{V})^{\mathrm{T}}\boldsymbol{Z}\boldsymbol{V}}{n-1} = \boldsymbol{V}^{\mathrm{T}}\frac{\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z}}{n-1}\boldsymbol{V} = \boldsymbol{V}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{V} = \boldsymbol{\Lambda} \tag{14}$$

This means each principal component captures a distinct portion of the total variance, with no overlap or redundancy between components.

In summary, eigenvalue decomposition provides the mathematical foundation of PCA: it identifies orthogonal directions that best summarize the variability of the data and expresses the data in this new coordinate system for easier interpretation and dimensionality reduction.

23.6 Eigenvalues and Explained Variance: How Much Information Do We Keep?

23.6.1 Linking Total Variance to the Sum of Eigenvalues

The eigenvalues of the correlation matrix $\boldsymbol{P}$ tell us how much variance each principal component explains. Because $\boldsymbol{P}$ is computed from standardized data, its total variance equals the sum of all feature variances. Mathematically, this total variance is represented by the trace of $\boldsymbol{P}$, written as

$$\text{trace}(\boldsymbol{\Lambda}) = \lambda_1 + \lambda_2 + \cdots + \lambda_8 = 8 \tag{15}$$

since there are eight standardized features, each with variance equal to one.

Interestingly, the trace of the correlation matrix equals the trace of its eigenvalue matrix $\boldsymbol{\Lambda}$, that is, trace($\boldsymbol{P}$) = trace($\boldsymbol{\Lambda}$). This means the sum of all eigenvalues equals the total variance in the data. Each eigenvalue λ_j therefore represents how much of the total variance is captured by the j-th principal component.

23.6.2 Understanding the Scree Plot

Figure 10 shows the eigenvalues arranged in descending order. The first few eigenvalues are much larger than the rest, indicating that most of the variance can be explained by just a few principal components.

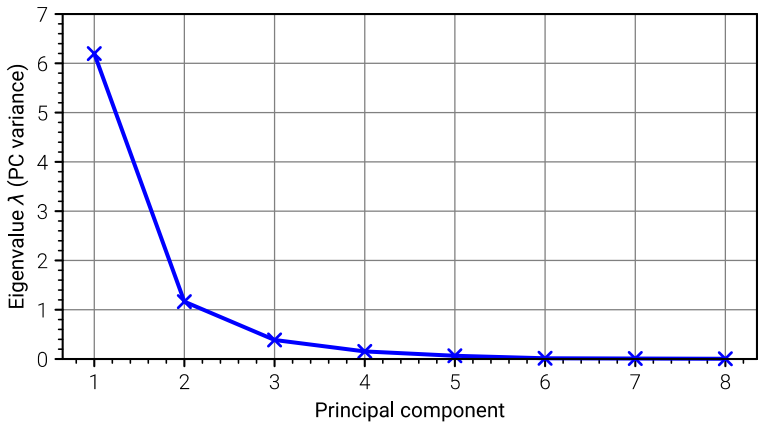

Figure 10. Eigenvalues of the correlation matrix sorted in descending order. Figure generated by Ch23_01_PCA_IR_data.ipynb.

To quantify this, we often calculate the percentage of explained variance for each component:

$$\frac{\lambda_j}{\sum_{j=1}^{D}\lambda_j}100\% \tag{16}$$

This ratio measures the importance of the *j*-th principal component in representing the overall variability of the dataset.

23.6.3 Cumulative Explained Variance and Dimensionality Choice

By summing these ratios for the first *p* components, we obtain the cumulative percentage of explained variance,

$$\frac{\sum_{j=1}^{p}\lambda_j}{\sum_{j=1}^{D}\lambda_j}100\% \tag{17}$$

which tells us how much of the total variance is retained when using the first *p* principal components

Figure 11 illustrates the cumulative explained variance curve. This plot is often used to decide how many components to keep.

For instance, if the first few components capture most of the variance, we can safely reduce the dimensionality of the data without losing much information.

In our dataset, the first principal component alone explains nearly 80 percent of the total variance, while the first two components together explain over 90 percent

$$\frac{\lambda_1+\lambda_2}{\lambda_1+\lambda_2+\cdots+\lambda_8}\times 100\% = 92.038\% \tag{18}$$

This means that even though the original data has eight variables, the majority of its information can be represented effectively in just two dimensions—a key motivation behind PCA.

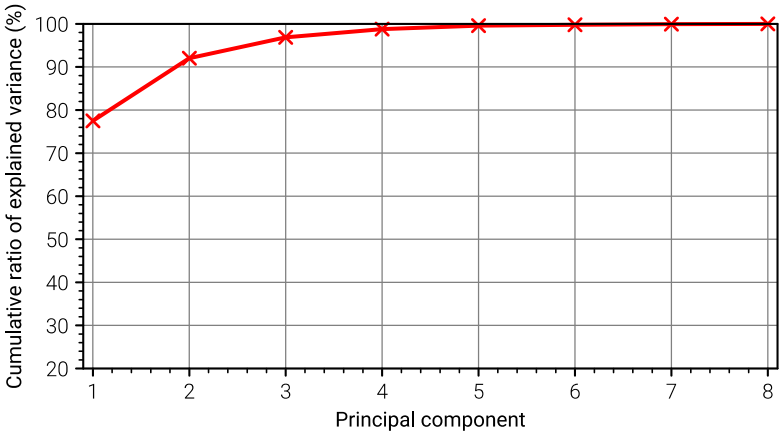

Figure 11. Cumulative percentage of total variance explained by the first *p* principal components. Figure generated by Ch23_01_PCA_IR_data.ipynb.

23.7 Reconstruction: Rebuilding the Data from Principal Components

23.7.1 The One-Component Reconstruction

When we reconstruct the standardized data $\boldsymbol{Z}$ using only the first principal component, we are projecting each sample onto the direction $\boldsymbol{v}_1$ and then using that one-dimensional projection to form an approximation $\boldsymbol{Z}_1$

$$\hat{\boldsymbol{Z}} = \boldsymbol{Z}_1 = \boldsymbol{y}_1 @ \boldsymbol{v}_1^{\mathrm{T}} = \boldsymbol{Z} @ \left(\boldsymbol{v}_1 @ \boldsymbol{v}_1^{\mathrm{T}} \right) \tag{19}$$

In matrix terms this is the orthogonal projection of $\boldsymbol{Z}$ onto the subspace spanned by $\boldsymbol{v}_1$, so the difference

$$\boldsymbol{Z} - \boldsymbol{Z}_1 = \boldsymbol{Z} - \boldsymbol{Z} @ \left(\boldsymbol{v}_1 @ \boldsymbol{v}_1^{\mathrm{T}} \right) = \boldsymbol{Z} \left(\boldsymbol{I} - \boldsymbol{v}_1 @ \boldsymbol{v}_1^{\mathrm{T}} \right) \tag{20}$$

is itself the residual left by that projection and is orthogonal to the retained subspace. Geometrically, using a single principal component means replacing each original high dimensional point with its foot on the one-dimensional line defined by $\boldsymbol{v}_1$.

To reconstruct the original (non-standardized) data from this one-component approximation we must reverse the standardization and centering steps.

First we scale $\boldsymbol{Z}_1$ by the original feature standard deviations (the diagonal matrix $\boldsymbol{D}$), which returns an approximation to the centered data $\boldsymbol{X}_c$.

Then we add back the mean vector $\boldsymbol{\mu}$ to recover an approximation to the original data matrix $\boldsymbol{X}$.

In other words, the full reconstruction is obtained by the sequence: translate (center) the data, scale (standardize), project onto $\boldsymbol{v}_1$, undo the scaling, and finally undo the translation

$$\hat{\boldsymbol{X}} = \boldsymbol{Z}_1 \boldsymbol{D} + \boldsymbol{\mu}^{\mathrm{T}} = \left(\boldsymbol{y}_1 @ \boldsymbol{v}_1^{\mathrm{T}} \right) \boldsymbol{D} + \boldsymbol{\mu}^{\mathrm{T}} = \left(\boldsymbol{Z} @ \left(\boldsymbol{v}_1 @ \boldsymbol{v}_1^{\mathrm{T}} \right) \right) \boldsymbol{D} + \boldsymbol{\mu}^{\mathrm{T}} \tag{21}$$

Because projection, scaling, and translation are linear (or affine) operations, they compose cleanly and the resulting reconstruction error for $\boldsymbol{X}$ can be written explicitly in matrix form

$$\hat{\boldsymbol{X}} = \left(\left(\boldsymbol{X} - \boldsymbol{\mu}^{\mathrm{T}} \right) \boldsymbol{D}^{-1} @ \left(\boldsymbol{v}_1 @ \boldsymbol{v}_1^{\mathrm{T}} \right) \right) \boldsymbol{D} + \boldsymbol{\mu}^{\mathrm{T}} \tag{22}$$

Plotting the original series together with the reconstruction from the first principal component makes the approximation quality intuitive, as shown in Figure 12. In such a plot, the original data appears as one curve while the reconstructed data from the first component tracks it more roughly; the difference between the two curves is the reconstruction error.

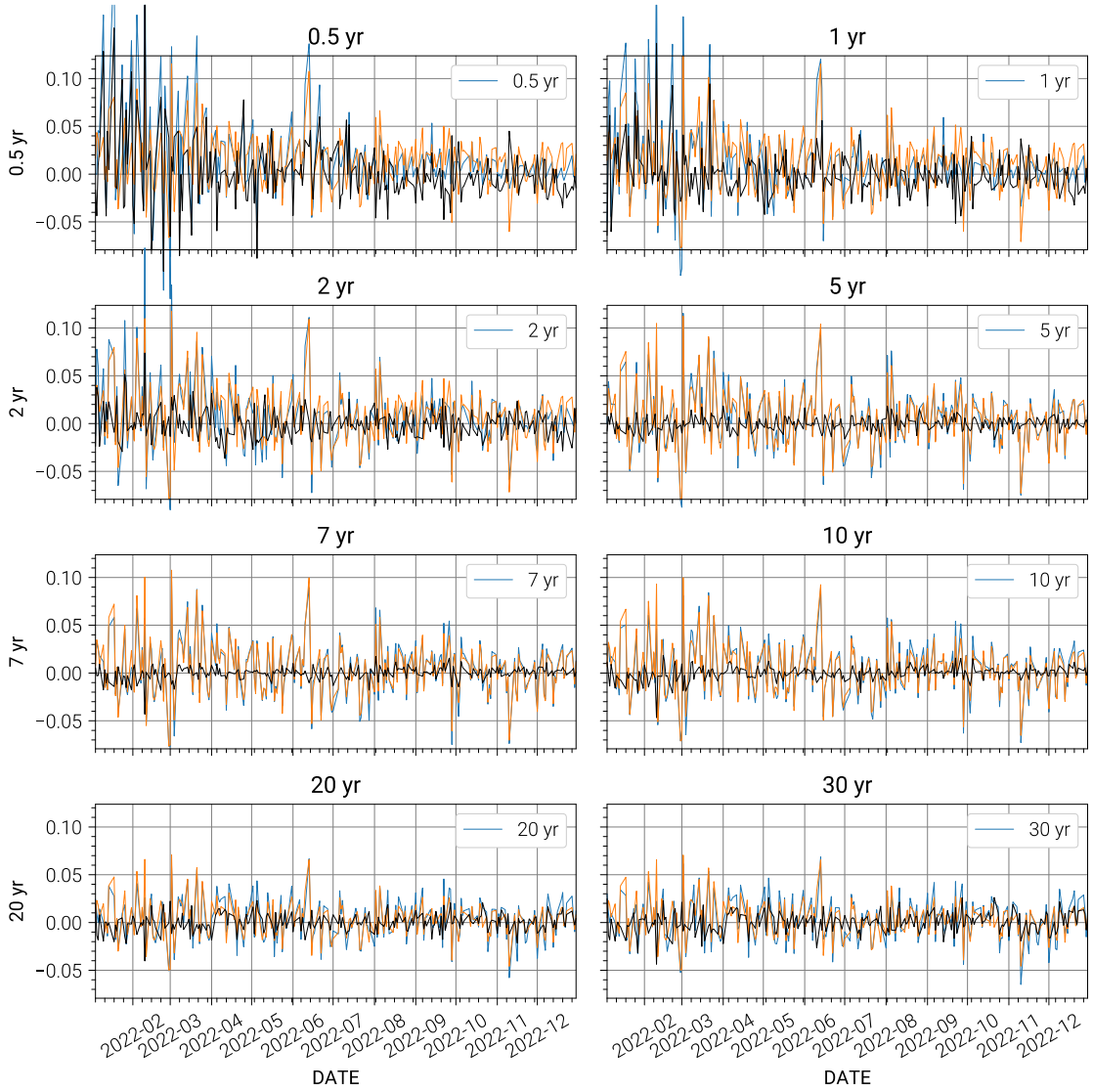

Figure 12. Original time series (blue), one-component reconstruction (orange), and reconstruction error (black). Figure generated by Ch23_01_PCA_IR_data.ipynb.

When we show pairwise scatter plots of the reconstructed series, as shown in Figure 13, all points fall on a single straight line in each subplot because a one-component reconstruction forces all variables to vary together according to their loadings on that single component.

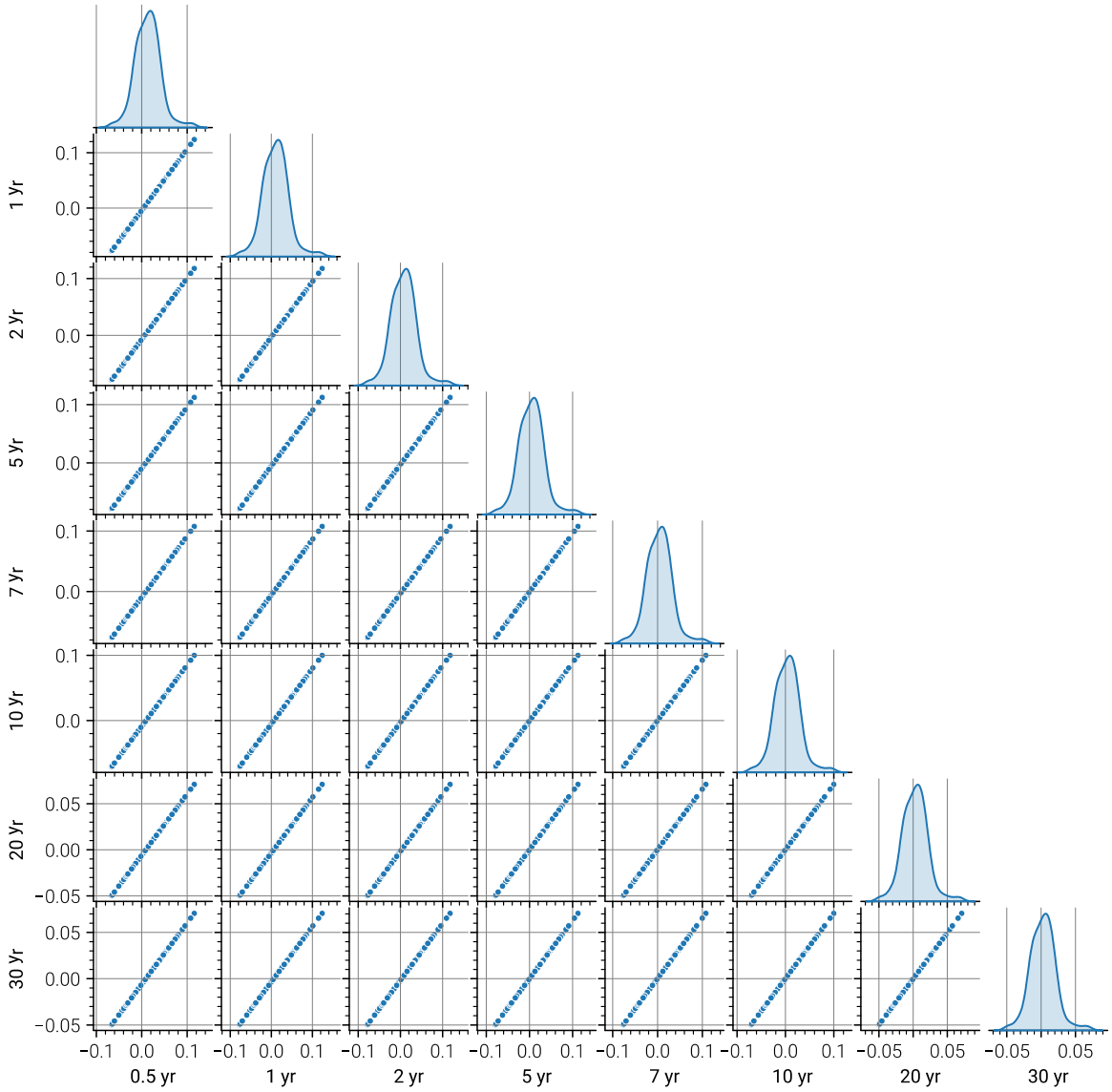

Figure 13. Pairwise scatter plots of one-component reconstructed data (plots above the main diagonal omitted). Figure generated by Ch23_01_PCA_IR_data.ipynb.

A convenient visual diagnostic is a scatter plot that compares reconstructed values on the horizontal axis with original values on the vertical axis, which gives us Figure 14. If the reconstruction were perfect, points would lie exactly on the diagonal $y = x$. The closer the cloud of points is to that diagonal, the better the reconstruction. Adding this reference line provides an immediate, graphical measure of similarity between the original and reconstructed data.

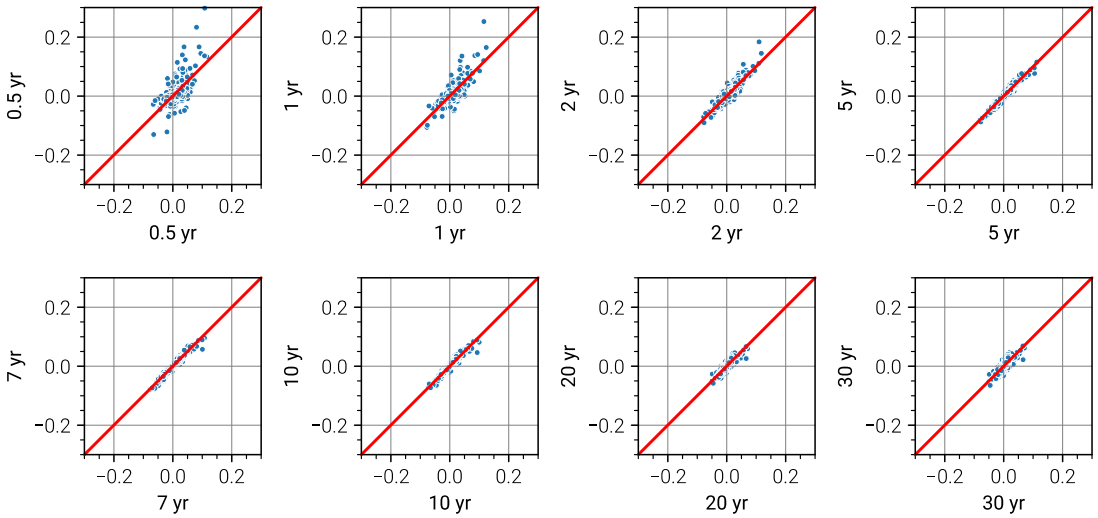

Figure 14. Scatter plot of reconstructed versus original values with reference diagonal $y = x$ to indicate perfect reconstruction. Figure generated by Ch23_01_PCA_IR_data.ipynb.

23.7.2 Two-Component Reconstruction: Recovering More Variance

If we keep the first two principal components instead of just the first, reconstruction follows the same sequence of geometric operations but projects onto the two-dimensional subspace spanned by $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$

$$\hat{\boldsymbol{X}} = \left(\boldsymbol{Z}_1 + \boldsymbol{Z}_2\right)\boldsymbol{D} + \boldsymbol{\mu}^{\mathrm{T}} = \left(\boldsymbol{y}_1 @ \boldsymbol{v}_1^{\mathrm{T}} + \boldsymbol{y}_2 @ \boldsymbol{v}_2^{\mathrm{T}}\right)\boldsymbol{D} + \boldsymbol{\mu}^{\mathrm{T}} = \boldsymbol{Z} @ \left(\boldsymbol{v}_1 @ \boldsymbol{v}_1^{\mathrm{T}} + \boldsymbol{v}_2 @ \boldsymbol{v}_2^{\mathrm{T}}\right)\boldsymbol{D} + \boldsymbol{\mu}^{\mathrm{T}} \tag{23}$$

This typically reduces the residual error because a two-dimensional subspace can capture more of the original variance. The reconstruction error for the two-component case can be computed from the residual covariance in the same way as for one component, and the linearly independent nature of $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ guarantees the residual remains orthogonal to the retained subspace.

Figure 15 ~ Figure 17 show the results of two-component reconstruction, which illustrate how much additional variance the second component recovers and how much the residual error shrinks.

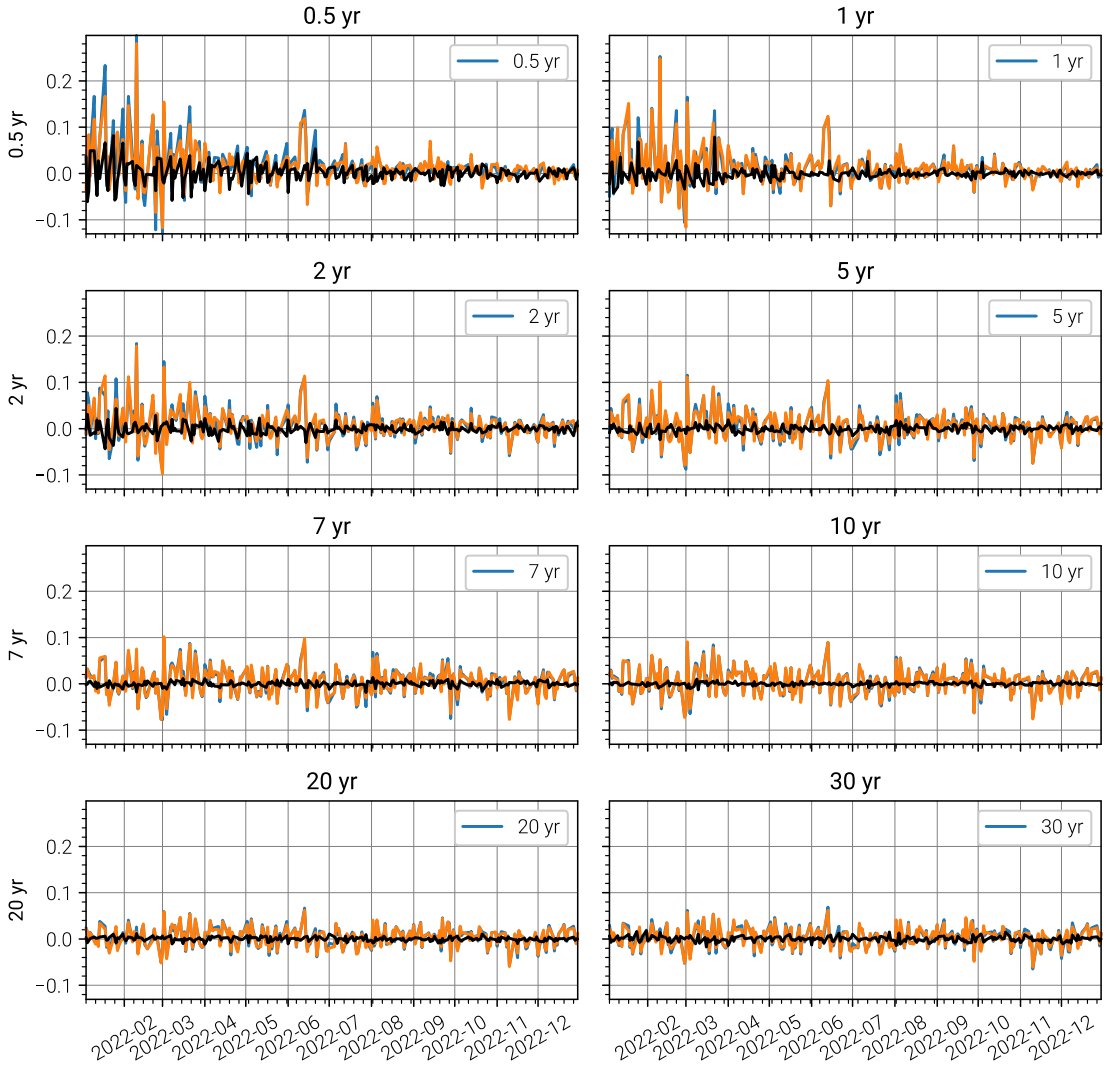

Figure 15. Original time series (blue) and two-component reconstruction (orange) comparison. Figure generated by Ch23_01_PCA_IR_data.ipynb.

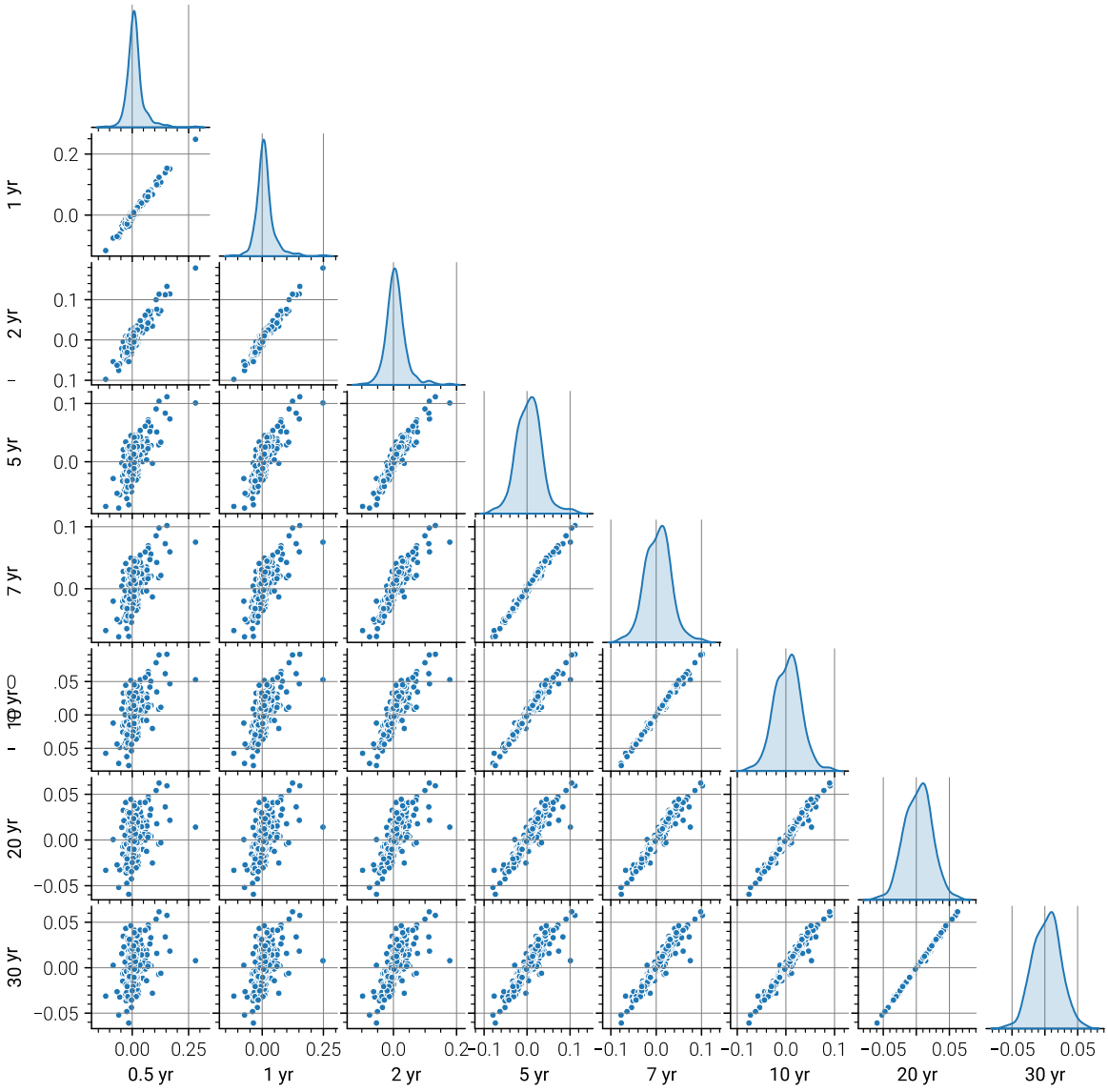

Figure 16. Scatter plot comparison of original and two-component reconstructed data (plots above the main diagonal omitted). Figure generated by Ch23_01_PCA_IR_data.ipynb.

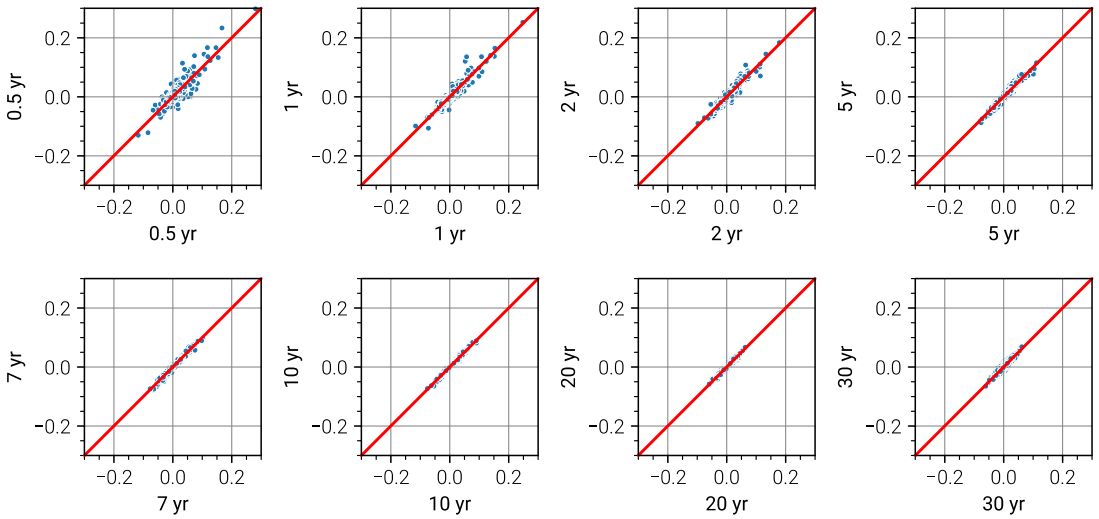

Figure 17. Reconstructed versus original scatter plot for the first two components with reference diagonal $y = x$. Figure generated by Ch23_01_PCA_IR_data.ipynb.

23.8 Conclusion

This chapter revisits Principal Component Analysis through a practical example using real financial data. It begins by exploring daily interest rate changes across different maturities and uses scatter plots to reveal strong correlations, suggesting common driving factors. The chapter explains how the covariance and correlation matrices capture these relationships, emphasizing the need for data standardization before applying PCA.

Through eigenvalue decomposition, PCA identifies orthogonal directions that capture the greatest variance, allowing complex data to be expressed in a simpler coordinate system. The explained variance plot shows that most of the variability can be summarized by the first few components.

Finally, the chapter demonstrates how to reconstruct the data using one or two principal components, illustrating how much structure can be preserved with dimensionality reduction. Together, these steps connect PCA's geometric intuition with its real-world application in analyzing and simplifying multivariate financial datasets.