

# Big Data in Finance

## Lecture 5: Unsupervised Learning

Dr Daniele Bianchi

Spring 2026

### Contents

<b>Overview</b>	<b>2</b>
<b>1 From Supervised to Unsupervised Learning</b>	<b>2</b>
1.1 Why Unsupervised Learning Matters in Finance . . . . .	3
<b>2 Principal Component Analysis (PCA)</b>	<b>3</b>
2.1 The Core Idea . . . . .	3
2.2 Geometric Intuition . . . . .	3
2.3 The Mathematics . . . . .	4
2.4 Variance Explained . . . . .	4
2.5 Choosing the Number of Components . . . . .	5
2.6 Standardization . . . . .	5
2.7 Interpreting Principal Components . . . . .	5
2.8 Limitations of PCA . . . . .	5
<b>3 Clustering Methods</b>	<b>6</b>
3.1 K-Means Clustering . . . . .	6
3.2 Choosing K . . . . .	6
3.3 Hierarchical Clustering . . . . .	6
3.4 Linkage Methods . . . . .	7
3.5 Important Considerations . . . . .	7
<b>4 PCA for Statistical Factor Models</b>	<b>7</b>
4.1 Statistical vs. Economic Factors . . . . .	7
4.2 PCA on Stock Returns . . . . .	8
4.3 Interpreting the Components . . . . .	8
4.4 Applications in Risk Management . . . . .	8
<b>5 Clustering Applications in Finance</b>	<b>9</b>
5.1 Clustering Stocks for Portfolio Construction . . . . .	9
5.2 Regime Detection . . . . .	10
<b>6 PCA for Dimensionality Reduction in Prediction</b>	<b>10</b>
6.1 Principal Component Regression (PCR) . . . . .	10
6.2 PCR vs. Ridge Regression . . . . .	10
6.3 Diffusion Index Forecasting . . . . .	11
<b>7 Summary and Key Takeaways</b>	<b>11</b>
<b>Readings</b>	<b>12</b>

## Overview

In Lectures 2 through 4, we studied **supervised learning**: given labeled data  $(X_i, Y_i)$ , learn to predict  $Y$  from  $X$ . Whether predicting returns (regression) or defaults (classification), we always had a target variable to guide our learning.

This lecture introduces **unsupervised learning**, where we have no target variable. We observe only features  $X_1, X_2, \dots, X_p$  and seek to discover structure in the data. The two main tasks are dimensionality reduction (finding a smaller set of variables that captures most of the information) and clustering (grouping similar observations together).

Sections 1 and 2 develop Principal Component Analysis (PCA), the fundamental technique for dimensionality reduction. PCA finds new variables—linear combinations of the originals—that capture maximum variance while being uncorrelated with each other. This is the mathematical foundation behind statistical factor models in finance.

Sections 3 and 4 introduce clustering methods. K-means partitions observations into  $K$  groups by minimizing within-cluster variation. Hierarchical clustering builds a tree structure that reveals relationships at multiple scales. Both require choosing the number of clusters and handling the challenge that there is no “ground truth” against which to evaluate.

Section 5 connects PCA to factor models in finance, showing how statistical factors from PCA relate to economic factors like Fama-French. Section 6 applies clustering to portfolio construction and regime detection. Section 7 discusses Principal Component Regression, which uses PCA for dimensionality reduction before prediction.

By the end of this lecture, you will understand how to use unsupervised methods to discover structure in financial data—extracting factors from returns, grouping stocks by behavior, and identifying market regimes.

## 1 From Supervised to Unsupervised Learning

In supervised learning, we had a clear objective: minimize prediction error for some target  $Y$ . The labels guided the learning process. In regression, we minimized squared errors; in classification, we maximized likelihood or minimized misclassification.

In unsupervised learning, there is no target. We observe only the features  $X_1, X_2, \dots, X_p$  for each observation. The goal is to discover **structure**—patterns, groupings, or low-dimensional representations that help us understand the data.

This is both liberating and challenging. It is liberating because we can analyze data without needing labels (which are often expensive or impossible to obtain). It is challenging because without a target, there is no obvious way to evaluate success. What makes one clustering “better” than another? What makes one set of principal components more useful than another?

The two main tasks in unsupervised learning are:

**Dimensionality reduction:** Find a smaller set of variables that captures most of the information. If we have 100 predictors, perhaps 5 principal components explain 90% of the variation. We can then work with 5 variables instead of 100, reducing computational burden and noise.

**Clustering:** Group similar observations together. If we have 500 stocks, perhaps they naturally fall into 10 groups based on return behavior. We can then analyze or trade at the group level rather than the individual stock level.

## 1.1 Why Unsupervised Learning Matters in Finance

Financial data exhibits strong structure. Stock returns are highly correlated—when the market rises, most stocks rise. This correlation structure means that returns can be well-approximated by a small number of common factors. PCA extracts these factors from the data.

Stocks within industries tend to move together, but industry classifications are static and sometimes arbitrary. Is Amazon a tech company or a retailer? Clustering based on return behavior provides a data-driven alternative that updates automatically as relationships change.

Market behavior changes over time. Bull markets, bear markets, and crisis periods have different characteristics. Clustering time periods by market conditions can identify these regimes and inform conditional trading strategies.

## 2 Principal Component Analysis (PCA)

PCA is the fundamental technique for dimensionality reduction. The goal is to find new variables—**principal components**—that are linear combinations of the original variables, capture as much variance as possible, and are uncorrelated with each other.

### 2.1 The Core Idea

Suppose we have  $p$  variables, each measured for  $n$  observations. The data form an  $n \times p$  matrix  $\mathbf{X}$ . We seek new variables  $Z_1, Z_2, \dots, Z_p$  that are linear combinations of the originals:

$$Z_1 = w_{11}X_1 + w_{12}X_2 + \dots + w_{1p}X_p$$

The first principal component  $Z_1$  is chosen to have maximum variance. The second principal component  $Z_2$  is chosen to have maximum variance subject to being uncorrelated with  $Z_1$ . And so on.

Why variance? Variables that do not vary carry no information. If a variable is constant across observations, it tells us nothing about differences between observations. Variance is a measure of information content.

The constraint that principal components be uncorrelated ensures they capture different aspects of the data. If  $Z_1$  captures the “market effect,” then  $Z_2$  captures something orthogonal—perhaps a size effect or sector rotation.

### 2.2 Geometric Intuition

Imagine data points scattered in two dimensions, forming an elliptical cloud. The first principal component is the direction along the major axis of the ellipse—the direction of maximum spread. The second principal component is perpendicular to the first, along the minor axis.

If the ellipse is elongated, the first principal component captures most of the spread. Projecting the data onto just this one direction loses little information. The second component adds only the residual spread.

In higher dimensions, the same logic applies. PCA finds the  $p$  orthogonal directions in which the data vary most, ordered by importance. The first few directions typically capture most of the variation; the remaining directions capture mostly noise.

## 2.3 The Mathematics

Assume the variables are centered (mean zero). The sample covariance matrix is:

$$\mathbf{S} = \frac{1}{n-1} \mathbf{X}^\top \mathbf{X}$$

The first principal component maximizes:

$$\mathbf{w}_1 = \arg \max_{\|\mathbf{w}\|=1} \mathbf{w}^\top \mathbf{S} \mathbf{w}$$

This is a constrained optimization problem. The solution comes from the eigenvalue decomposition of  $\mathbf{S}$ :

$$\mathbf{S} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^\top$$

where  $\mathbf{V}$  contains the eigenvectors and  $\mathbf{\Lambda}$  is diagonal with eigenvalues  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ .

The principal components are the eigenvectors, and the variance of each component equals its eigenvalue. The first eigenvector (with largest eigenvalue) is the first principal component. The second eigenvector is the second principal component. And so on.

## 2.4 Variance Explained

The total variance in the data equals the sum of all eigenvalues:

$$\text{Total variance} = \sum_{j=1}^p \lambda_j$$

The proportion of variance explained by the  $j$ -th principal component is:

$$\text{PVE}_j = \frac{\lambda_j}{\sum_{k=1}^p \lambda_k}$$

The cumulative proportion explained by the first  $k$  components is:

$$\text{CPVE}_k = \frac{\sum_{j=1}^k \lambda_j}{\sum_{j=1}^p \lambda_j}$$

Consider a concrete example: PCA applied to 10 stock returns.

	Variance Explained	Cumulative
PC1	45%	45%
PC2	20%	65%
PC3	10%	75%
PCs 4–10	25%	100%

Using just 3 principal components instead of 10 original variables, we retain 75% of the information. The last 7 components add little—mostly idiosyncratic noise.

In financial applications, a typical finding is that the first principal component of stock returns explains 35–45% of total variance, the first three explain 50–60%, and the first five explain 60–70%. The remaining components capture mostly idiosyncratic noise.

## 2.5 Choosing the Number of Components

How many principal components should we keep? Common approaches include:

**Scree plot:** Plot eigenvalues against component number. Look for an “elbow” where the curve flattens—beyond this point, additional components add little variance explained.

**Cumulative variance threshold:** Keep enough components to explain 80–90% of total variance. This is subjective but widely used.

**Cross-validation:** If using PCA for prediction, choose the number of components that minimizes out-of-sample prediction error.

**Domain knowledge:** In finance, economic theory often suggests a small number of factors. Three to five components typically capture the systematic variation in stock returns.

## 2.6 Standardization

An important choice is whether to standardize variables before applying PCA. Without standardization, PCA uses the covariance matrix; variables with larger variance dominate. With standardization (each variable scaled to variance 1), PCA uses the correlation matrix; all variables are treated equally.

The rule of thumb is: if variables are in different units or have very different variances, standardize first. If variables are comparable (e.g., all stock returns in percentage terms), the covariance matrix may be preferred because it preserves the relative importance of more volatile stocks.

## 2.7 Interpreting Principal Components

The **loadings**—the weights  $w_{jk}$  in the linear combination—reveal what each principal component represents. Consider a concrete example of PCA applied to stock returns:

Stock	PC1 Loading	PC2 Loading	PC3 Loading
Apple	0.15	0.35	−0.10
Microsoft	0.14	0.32	−0.12
JPMorgan	0.12	−0.25	0.40
Goldman Sachs	0.11	−0.28	0.38
ExxonMobil	0.10	−0.05	−0.30

**First PC:** All stocks load positively with similar magnitudes. This is the “market factor”—when PC1 is high, most stocks have high returns.

**Second PC:** Tech stocks (Apple, Microsoft) load positively while banks (JPMorgan, Goldman) load negatively. This captures a “growth vs. value” or sector rotation effect.

**Third PC:** Banks load positively, energy (ExxonMobil) loads negatively. This might capture another form of sector rotation.

The interpretation is not guaranteed. PCA finds statistical patterns that maximize variance, not economically meaningful factors. But in practice, the first few principal components often have intuitive interpretations.

## 2.8 Limitations of PCA

PCA is powerful but has limitations:

**Linearity:** PCA finds only linear combinations. If the underlying structure is nonlinear, PCA will miss it. Kernel PCA extends the method to nonlinear relationships.

**Orthogonality:** Principal components must be uncorrelated. But real economic factors may be correlated—value and momentum, for instance, are not orthogonal.

**Variance focus:** PCA maximizes variance, not predictive power. A variable with low variance might be highly predictive of returns, but PCA will give it little weight.

**Interpretability:** Each principal component is a linear combination of all original variables. This makes interpretation difficult. Sparse PCA forces some loadings to zero, improving interpretability at the cost of optimality.

### 3 Clustering Methods

While PCA reduces the number of variables, clustering groups observations. Given  $n$  observations with  $p$  features each, we seek to partition them into clusters such that observations within a cluster are similar and observations in different clusters are dissimilar.

#### 3.1 K-Means Clustering

K-means is the simplest and most widely used clustering algorithm. The goal is to partition  $n$  observations into  $K$  clusters by minimizing within-cluster variation:

$$\min_{C_1, \dots, C_K} \sum_{k=1}^K \sum_{i \in C_k} \|x_i - \mu_k\|^2$$

where  $\mu_k = \frac{1}{|C_k|} \sum_{i \in C_k} x_i$  is the centroid of cluster  $k$ .

The algorithm iterates between two steps:

**Assignment step:** Given current centroids, assign each observation to the nearest centroid.

**Update step:** Given current assignments, recompute each centroid as the mean of its assigned observations.

This process is guaranteed to decrease the objective at each step and converges to a local minimum. Because the solution depends on initialization, K-means is typically run multiple times with different random starting points.

#### 3.2 Choosing K

K-means requires specifying  $K$  in advance. Common approaches for choosing  $K$  include:

**Elbow method:** Plot total within-cluster sum of squares against  $K$ . As  $K$  increases, the sum of squares decreases (more clusters means tighter clusters). Look for an “elbow” where the improvement slows—this suggests a natural number of clusters.

**Silhouette score:** For each observation, compute how well it fits its cluster compared to other clusters. The silhouette score ranges from  $-1$  (wrong cluster) to  $+1$  (well-clustered). Choose  $K$  to maximize the average silhouette score.

**Domain knowledge:** In finance, we often have prior beliefs about the number of groups. Sectors, risk buckets, or market regimes suggest natural values of  $K$ .

#### 3.3 Hierarchical Clustering

Hierarchical clustering builds a tree structure (dendrogram) that shows relationships at multiple scales. The agglomerative (bottom-up) approach works as follows:

1. Start with each observation as its own cluster.
2. Find the two closest clusters and merge them.
3. Repeat until all observations are in one cluster.

The result is a dendrogram that can be cut at any height to produce a clustering. High cuts produce few large clusters; low cuts produce many small clusters. This flexibility is an advantage over K-means: we do not need to specify  $K$  in advance.

### 3.4 Linkage Methods

Hierarchical clustering requires defining “distance” between clusters. Common choices are:

**Single linkage:** Distance between clusters is the minimum distance between any pair of points. Tends to produce long, stringy clusters.

**Complete linkage:** Distance is the maximum distance between any pair. Produces compact clusters.

**Average linkage:** Distance is the average of all pairwise distances.

**Ward’s method:** Minimizes the increase in within-cluster variance when merging. Tends to produce compact, equal-sized clusters and is often the best choice.

### 3.5 Important Considerations

Several practical issues affect clustering quality:

**Feature scaling:** Clustering algorithms use distances, so scale matters. A variable measured in millions will dominate one measured in decimals. Standardize features before clustering unless the scale differences are meaningful.

**Dimensionality:** In high dimensions, distances become less meaningful (the “curse of dimensionality”). Consider applying PCA before clustering to reduce noise and focus on systematic variation.

**Outliers:** K-means is sensitive to outliers because they distort centroids. Consider removing outliers first or using robust alternatives.

**Validation:** Without ground truth labels, validating clusters is difficult. Use multiple methods, check stability across subsamples, and rely on domain expertise to assess whether clusters are meaningful.

## 4 PCA for Statistical Factor Models

In asset pricing, returns are driven by common factors. The factor model representation is:

$$R_{it} - R_f = \alpha_i + \beta_{i1}F_{1t} + \beta_{i2}F_{2t} + \cdots + \beta_{iK}F_{Kt} + \varepsilon_{it}$$

The factors  $F_{jt}$  represent common sources of risk; the loadings  $\beta_{ij}$  measure each asset’s exposure; the residual  $\varepsilon_{it}$  is idiosyncratic.

### 4.1 Statistical vs. Economic Factors

There are two approaches to identifying factors:

**Economic factors:** Based on theory. Fama-French factors (market, size, value) are constructed from sorted portfolios with economic motivations. They are interpretable but may miss important sources of covariation.

**Statistical factors:** Based on data. PCA on returns extracts the directions of maximum covariation. They are optimal for variance explanation but may lack economic interpretation.

## 4.2 PCA on Stock Returns

When we apply PCA to a matrix of stock returns (stocks as variables, time periods as observations), we extract statistical factors. The first principal component is the direction of maximum return covariation. Empirically, this is almost always a “market” factor—all stocks load positively with similar magnitudes.

Typical findings for US equities:

Component	Variance Explained	Cumulative
PC1	35–45%	35–45%
PC2	8–12%	45–55%
PC3	4–6%	50–60%
PC4–PC5	2–4% each	58–68%

The first principal component dominates. The first three to five components explain the majority of covariation. Remaining components capture idiosyncratic noise.

## 4.3 Interpreting the Components

Examining loadings reveals economic interpretation:

**PC1:** All positive, similar loadings. This is the market factor. Correlation with the Fama-French MKT factor exceeds 0.95.

**PC2:** Often shows a size pattern. Small stocks load positively, large stocks negatively (or vice versa). Correlation with SMB is typically 0.6–0.8.

**PC3:** Often shows a value pattern. High book-to-market stocks load differently from low book-to-market stocks. Correlation with HML is 0.5–0.7.

Statistical factors from PCA are related to but not identical to Fama-French factors. PCA factors capture size and value effects plus other sources of covariation that Fama-French may miss.

## 4.4 Applications in Risk Management

PCA factors are useful for several risk management tasks:

**Covariance estimation:** With 500 stocks and 60 months of data, the sample covariance matrix is noisy (more parameters than observations). Using only the first few principal components produces a smoother, more stable estimate.

**Portfolio optimization:** Instead of estimating a  $500 \times 500$  covariance matrix, estimate exposures to 5 factors. Fewer parameters means less estimation error and more stable portfolios.

**Risk decomposition:** Factor models decompose portfolio variance into systematic and idiosyncratic components. How much risk comes from the market? From sector bets? From stock-specific positions?

## 5 Clustering Applications in Finance

### 5.1 Clustering Stocks for Portfolio Construction

Traditional portfolio construction groups stocks by sector (GICS codes, SIC codes). But sector classifications are static, and stocks within sectors can behave quite differently. Is Amazon a technology company, a retailer, or a cloud services provider?

Clustering based on return behavior provides a data-driven alternative. We group stocks that actually move together, regardless of their official sector labels.

**What features to use?** Three options:

1. **Raw returns:** Cluster based on return time series. High-dimensional and noisy.
2. **Correlations:** Convert correlations to distances:  $d_{ij} = \sqrt{2(1 - \rho_{ij})}$ . Stocks with high correlation are “close.”
3. **PCA scores:** First reduce dimension with PCA, then cluster based on factor loadings. This denoises the data.

The correlation-based distance is intuitive. The following table shows how correlations map to distances:

If two stocks have...	Correlation	Distance
Identical returns	$\rho = 1.0$	0 (very close)
Similar movements	$\rho = 0.8$	0.63
No relationship	$\rho = 0.0$	1.41
Opposite movements	$\rho = -1.0$	2.0 (very far)

Apple and Microsoft, with high correlation, would have a small distance and end up in the same cluster. A tech stock and a utility, with low correlation, would have a large distance and end up in different clusters.

The third approach—PCA followed by clustering—often works best. PCA removes noise and focuses on systematic variation.

**Cluster-based allocation:** Instead of equal-weighting stocks, allocate equally across clusters, then equal-weight within clusters. This ensures true diversification across different sources of risk, not just different sector labels.

Consider a concrete example. Suppose clustering 9 stocks reveals 3 groups:

Cluster 1 (Tech)	Cluster 2 (Financials)	Cluster 3 (Energy)
Apple	JPMorgan	ExxonMobil
Microsoft	Bank of America	Chevron
Google	Goldman Sachs	ConocoPhillips

A simple cluster-based allocation would: (1) allocate equally across clusters (33.3% each), then (2) within each cluster, equal-weight the stocks (11.1% per stock).

Compare this to naive equal-weighting (11.1% each). If the Tech cluster is highly correlated internally, naive equal-weighting puts 33% in one correlated “bet” without making this explicit. Cluster-aware allocation recognizes that the three tech stocks represent essentially one source of risk and allocates accordingly.

## 5.2 Regime Detection

Market behavior changes over time. Bull markets have low volatility, positive drift, and low correlations. Bear markets have high volatility, negative drift, and high correlations. Crisis periods are extreme versions of bear markets.

We can use clustering to identify these **regimes** by clustering time periods based on market characteristics:

**Features for each period:** Market return, realized volatility, average correlation, credit spread, term spread, VIX.

**Process:** Standardize features, apply K-means or hierarchical clustering, label periods by cluster.

Typical findings with  $K = 3$ :

Feature	Regime 1 (Bull)	Regime 2 (Normal)	Regime 3 (Crisis)
Market return	High (+)	Moderate	Low (-)
Volatility	Low	Medium	High
Avg correlation	0.2–0.3	0.3–0.5	0.6–0.8
Frequency	30%	55%	15%

Regime detection enables conditional strategies: different allocations or models depending on the detected regime. When a crisis regime is detected, reduce risk exposure. When a bull regime is detected, increase it.

## 6 PCA for Dimensionality Reduction in Prediction

PCA can also be used as a preprocessing step for prediction. When we have many predictors that are correlated, we can reduce them to a smaller set of principal components and then regress the target on these components.

### 6.1 Principal Component Regression (PCR)

The procedure is:

1. Apply PCA to the predictors  $\mathbf{X}$ , obtaining  $k$  principal components  $\mathbf{F}$ .
2. Regress the target  $\mathbf{y}$  on  $\mathbf{F}$ :

$$y_t = \gamma_0 + \gamma_1 F_{1t} + \gamma_2 F_{2t} + \cdots + \gamma_k F_{kt} + \varepsilon_t$$

This helps for several reasons:

**Fewer parameters:** We estimate  $k$  coefficients instead of  $p$ , reducing overfitting.

**No multicollinearity:** Principal components are uncorrelated by construction.

**Noise reduction:** Dropping low-variance components removes noise from the predictors.

### 6.2 PCR vs. Ridge Regression

Both PCR and Ridge shrink the effective number of parameters, but differently:

	PCR	Ridge
Approach	Drop low-variance PCs	Shrink all directions
Shrinkage pattern	Discrete (keep or drop)	Continuous
Uses target?	No (unsupervised)	Yes (supervised)

A key insight is that Ridge shrinks more in low-variance directions—exactly where estimation error is highest. PCR drops these directions entirely, which may discard predictive information if a low-variance direction happens to predict well.

In practice, Ridge often outperforms PCR because it uses the target variable to guide shrinkage. PCR's advantage is interpretability: the principal components may have meaningful interpretations (market factor, size factor, etc.).

### 6.3 Diffusion Index Forecasting

Stock and Watson (2002) popularized PCR for macroeconomic forecasting under the name “diffusion index forecasting.” The idea is to summarize many macro predictors (100+ series) into a few principal components (“diffusion indices”) and use these for forecasting inflation, output, and other targets.

The name “diffusion” reflects that the indices aggregate “diffuse” information across many series. No single series dominates; each contributes.

In finance, Rapach, Strauss, and Zhou (2013) apply this approach to international return prediction. They extract principal components from predictors across many countries and show that these improve forecast accuracy compared to using predictors from a single country.

## 7 Summary and Key Takeaways

This lecture has covered unsupervised learning methods and their applications in finance. The key points:

**Unsupervised learning discovers structure without labels.** The two main tasks are dimensionality reduction (PCA) and clustering (K-means, hierarchical). Evaluation is challenging because there is no target against which to measure success.

**PCA finds directions of maximum variance.** Principal components are linear combinations of original variables that are uncorrelated and ordered by variance explained. The first few components often capture most of the systematic variation.

**PCA extracts statistical factors from returns.** The first component is almost always a market factor. Subsequent components often correlate with size, value, and other economic factors, but are not identical to Fama-French.

**Clustering groups similar observations.** K-means requires specifying  $K$  and minimizes within-cluster variation. Hierarchical clustering builds a dendrogram and does not require pre-specifying  $K$ . Feature scaling is critical.

**Cluster stocks by return behavior, not sector labels.** Correlation-based clustering provides a data-driven alternative to traditional sector classifications. This can improve portfolio diversification.

**Cluster time periods to detect regimes.** Bull markets, bear markets, and crises have different characteristics. Regime detection enables conditional strategies.

**PCR reduces dimensions before prediction.** Replace  $p$  correlated predictors with  $k$  uncorrelated principal components. Ridge regression often outperforms PCR because it uses the target to guide shrinkage.

In the next lecture, we turn to model interpretability and explainability—methods for understanding what machine learning models have learned and why they make the predictions they do.

## Readings

### Required:

- James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). *An Introduction to Statistical Learning with Applications in Python*, Chapter 12.

### Supplementary:

- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*, Chapters 14.3 and 14.5. More mathematical treatment of PCA and clustering.
- Jolliffe, I. T., & Cadima, J. (2016). “Principal Component Analysis: A Review and Recent Developments.” *Philosophical Transactions of the Royal Society A*, 374(2065). Comprehensive review of PCA methods and extensions.
- López de Prado, M. (2016). “Building Diversified Portfolios that Outperform Out-of-Sample.” *Journal of Portfolio Management*, 42(4), 59–69. Hierarchical clustering for portfolio construction.
- Stock, J. H., & Watson, M. W. (2002). “Forecasting Using Principal Components from a Large Number of Predictors.” *Journal of the American Statistical Association*, 97(460), 1167–1179. Foundational paper on diffusion index forecasting.
- Rapach, D. E., Strauss, J. K., & Zhou, G. (2013). “International Stock Return Predictability: What Is the Role of the United States?” *Journal of Finance*, 68(4), 1633–1662. Application of PCA to international return prediction.