Principal Components and Factor Models

Kerry Back
BUSI 521/ECON 505
Rice University

Open companion notebook in Colab

Factor Models

Setup

Let R_t be an N–vector of excess returns with mean μ and covariance matrix Σ.

A K–factor model is

$$R_t = \mu + Bf_t + e_t$$

- f_t is a K–vector of factors (zero mean, $\text{Cov}(f_t) = I$)
- $B = \text{Cov}(R_t, f_t)$ is the $N \times K$ loading matrix
- e_t is uncorrelated with f_t

What Makes It a Factor Model?

The key restriction: the elements of e_t are mutually uncorrelated.

So $\text{Cov}(e_t) = D$ for a diagonal matrix D.

This implies

$$\Sigma = BB' + D$$

All return covariances are explained by common factor exposures.
Residual risks are idiosyncratic.

Factor Analysis

Rotational Indeterminacy

The decomposition $\Sigma = BB' + D$ does not uniquely determine B.

For any $K \times K$ orthogonal matrix Q (so $QQ' = I$), define $\tilde{B} = BQ$. Then

$$\tilde{B}\tilde{B}' = BQQ'B' = BB'$$

So $\tilde{B}$ and $\tilde{f}_t = Q'f_t$ give the same covariance structure. The loading matrix is only identified up to an orthogonal rotation.

The Identification Restriction

To pin down B, maximum likelihood factor analysis imposes:

$$B'D^{-1}B \quad \text{is diagonal}$$

Interpretation. In the pre-whitened space, define $\tilde{B} = D^{-1/2}B$. Then

$$\tilde{B}'\tilde{B} = B'D^{-1}B$$

So the restriction says the columns of $\tilde{B}$ (the whitened loadings) are orthogonal. This is natural because $B'D^{-1}B$ appears in the GLS formula for factor inference.

Why It Resolves Indeterminacy

A $K \times K$ orthogonal matrix Q has $K(K-1)/2$ free parameters (one angle per pair of factors). The restriction "$B'D^{-1}B$ is diagonal" kills the $K(K-1)/2$ off-diagonal elements—a perfect match.

- Existence. Given any B, the symmetric matrix $B'D^{-1}B$ has an eigendecomposition $Q\Lambda Q'$. Setting $\tilde{B} = BQ$ gives $\tilde{B}'D^{-1}\tilde{B} = \Lambda$ (diagonal).
- Uniqueness. If the diagonal entries of Λ are distinct, the only orthogonal Q satisfying $Q'\Lambda Q = \Lambda$ is a signed permutation matrix (sign flips and reordering of factors), which is a trivial relabeling.

We also need enough data moments. The model has $NK + N - K(K-1)/2$ free parameters in B and D, which must not exceed the $N(N+1)/2$ distinct elements of Σ.

ML Estimation

Given the identification restriction, ML estimation recovers $\hat{B}$ and $\hat{D}$ from the sample covariance matrix.

But we still need the factor realizations f_t at each date.

Inferring the Factors

At each date t, the model says

$$\underbrace{R_t - \mu}_{N\times 1} = \underbrace{B}_{N\times K}\ \underbrace{f_t}_{K\times 1} + \underbrace{e_t}_{N\times 1}$$

This is a cross-sectional regression: N observations (one per asset), K unknowns (the factor values f_t), and B plays the role of the design matrix.

- OLS estimate: $\hat{f}_t = (B'B)^{-1}B'(R_t - \mu)$
- GLS estimate: $\hat{f}_t = (B'D^{-1}B)^{-1}B'D^{-1}(R_t - \mu)$

GLS is preferred because $\text{Cov}(e_t) = D$ is heteroskedastic (different assets have different idiosyncratic variances).

GLS Is Maximum Likelihood

Under the model, $R_t - \mu \mid f_t \sim N(Bf_t,\ D)$. The GLS estimate at each date t,

$$\hat{f}_t = (B'D^{-1}B)^{-1}B'D^{-1}(R_t - \mu)$$

maximizes the conditional likelihood $p(R_t \mid f_t)$. Equivalently, it solves

$$\min_f \sum_{i=1}^{N} \frac{(R_{it} - \mu_i - b_i' f)^2}{d_i}$$

- Assets whose variation is mostly factor-driven (small d_i) receive more weight.
- Assets with large idiosyncratic variance receive less weight.

The inferred factors are the ones that make the observed returns most probable under the assumed factor structure at each date.

GLS Creates Traded Factors

The GLS formula

$$f_t = (\hat{B}'\hat{D}^{-1}\hat{B})^{-1}\hat{B}'\hat{D}^{-1}(R_t - \hat{\mu})$$

shows that each factor is a linear combination of returns—a portfolio return (up to a constant).

- The companion notebook confirms this: regressing any inferred factor on the N returns gives $R^2 = 1$ (exact spanning).
- This exact spanning is what makes the FMP construction work. The regression coefficients, normalized to sum to one, are the portfolio weights.

In python: `sklearn.decomposition.FactorAnalysis`

→ Open companion notebook in Colab

Principal Components

Eigenvectors of the Covariance Matrix

Let C be the matrix of eigenvectors of Σ, ordered by eigenvalue (largest first).

Partition $C = (C_1 \mid C_2)$ where C_1 contains the first K eigenvectors.

The principal component factors are

$$f_t = (R_t - \hat{\mu})\, C_1$$

The eigenvectors are orthonormal: $C'C = CC' = I$.

PCA as a Factor Model

Since $(f_t, \text{extras}_t) = (R_t - \hat{\mu})\, C$ and $CC' = I$:

$$R_t = \hat{\mu} + f_t\, C_1' + \text{extras}_t\, C_2'$$

- The loading matrix is C_1' (the transposed eigenvectors)
- The "residuals" are $\text{extras}_t \cdot C_2'$
- Factors and residuals are orthogonal

PCA does not force residuals to be uncorrelated. Instead, it makes residuals small by dropping eigenvectors with small eigenvalues.

How Much Variance Is Explained?

Each eigenvalue λ_i equals the variance of the *i*th principal component.

The fraction of total variance explained by K components is

$$\frac{\lambda_1 + \cdots + \lambda_K}{\lambda_1 + \cdots + \lambda_N}$$

For the 25 size/BM portfolios, the first eigenvalue alone accounts for about 82% of total variance. Three components capture about 93%.

In python: `sklearn.decomposition.PCA`

Factor-Mimicking Portfolios

Why Factor-Mimicking Portfolios?

The raw factors (from FA or PCA) are linear combinations of returns, but the weights don't sum to one.

A factor-mimicking portfolio (FMP) rescales the weights so $\iota' w = 1$.

The FMP excess return is then a traded excess return that tracks the factor.

Constructing FMPs: PCA

For PCA, the weights are the eigenvector entries. To make an FMP:

$$w_i = \frac{C_{1,i}}{\iota' C_{1,i}}$$

That is, divide each eigenvector's entries by their sum.

This gives K portfolios whose excess returns are the PCA factor-mimicking portfolios.

Constructing FMPs: Factor Analysis

For factor analysis, regress each factor $f_t^{(i)}$ on a constant and the N excess returns:

$$f_t^{(i)} = \alpha + w' R_t + \text{error}$$

The regression coefficients w give the spanning weights. Normalize so $\iota' w = 1$.

The FMP return is $w' R_t$.

The R^2 is 1 (exact spanning) because the factors were constructed from the returns via the GLS formula.

Optimization Perspectives

PCA = Maximum-Variance Portfolios

The first principal component solves

$$\max_{w} \frac{1}{2} w' \Sigma w \quad \text{subject to} \quad w'w = 1$$

The FOC is $\Sigma w = \lambda w$ — the eigenvector equation.

The second component solves the same problem with the additional constraint $w' w_1 = 0$.

Successive components are maximum-variance portfolios orthogonal to all previous ones.

FA = Minimum Residual-Variance Portfolios

The factor-analysis FMPs solve

$$\min_{w} \frac{1}{2} w' D w \quad \text{subject to} \quad B' w = e_i$$

This minimizes idiosyncratic risk while targeting a unit loading on factor i and zero loadings on all other factors.

The FOC gives $w = D^{-1} B (B' D^{-1} B)^{-1} e_i$, and the resulting portfolio returns reproduce exactly the GLS factor estimates.

Sharpe Ratios

Maximum Sharpe Ratios

Given K excess-return factors with mean μ_f and covariance Σ_f, the maximum Sharpe ratio from combining them is

$$\text{SR}^* = \sqrt{\mu_f' \Sigma_f^{-1} \mu_f}$$

This is the Sharpe ratio of the tangency portfolio of the K factors.

Using monthly excess returns on the 25 Fama–French size/BM sorted portfolios from 1980–2021, with $K = 3$:

- Fama–French factor SR ≈ 0.178
- FA mimicking-portfolio SR ≈ 0.183
- PCA mimicking-portfolio SR ≈ 0.177

All three are similar—the statistical factors capture roughly the same pricing information as the Fama–French factors.

Summary

- Factor analysis: ML estimation of loadings and diagonal residual covariance; factors inferred by GLS. FMPs minimize idiosyncratic risk.
- PCA: Eigenvectors of the covariance matrix; factors are maximum-variance linear combinations. FMPs are rescaled eigenvectors.
- Both methods produce traded factor-mimicking portfolios by normalizing weights to sum to one.
- For the 25 size/BM portfolios, three statistical factors deliver Sharpe ratios comparable to Fama–French.