IEOR E4602: Quantitative Risk Management

Dimension Reduction Techniques

Martin Haugh

Department of Industrial Engineering and Operations Research
Columbia University
Email: martin.b.haugh@gmail.com

Reference: Chapter 18 of 2^{nd} ed. of SDAFA by Ruppert and Matteson.

Outline

Principal Components Analysis

Let $\mathbf{Y} = (Y_1 \ldots Y_n)^\top$ denote an n-dimensional random vector with variance-covariance matrix, $\mathbf{\Sigma}$.

$\mathbf{Y}$ represents (normalized) changes of risk factors over some appropriately chosen time horizon.

These risk factors might be:

- Security price **returns**
- **Returns** on futures contracts of varying maturities or
- **Changes** in spot interest rates, again of varying maturities.

Goal of PCA is to construct linear combinations of the Y_i's

$$P_i := \sum_{j=1}^{n} w_{ij} \, Y_j \quad \text{for} \quad i = 1, \ldots, n$$

in such a way that ....

Principal Components Analysis

(1) The P_i's are orthogonal so that $\mathsf{E}[P_i P_j] = 0$ for $i \neq j$

and

(2) The P_i's are ordered in such a way that:

(i) P_1 explains the largest percentage of the total variability in the system

and

(ii) each P_i explains the largest percentage of the total variability in the system that has **not** already been explained by $P_1, \ldots, P_{i-1}$.

Principal Components Analysis

In practice common to apply PCA to normalized random variables so that $\mathrm{E}[Y_i] = 0$ and $\mathrm{Var}(Y_i) = 1$

- can normalize by subtracting the means from the original random variables and then dividing by their standard deviations.

We normalize to ensure no one component of $\mathbf{Y}$ can influence the analysis by virtue of that component's measurement units.

Will therefore assume the Y_i's have already been normalized

- but common in financial applications to also work with non-normalized variables if clear that components of $\mathbf{Y}$ all on similar scale.

Key tool of PCA is the spectral decomposition or (more generally) the singular value decomposition (SVD) of linear algebra.

Spectral Decomposition

The spectral decomposition states that any symmetric matrix, $\mathbf{A} \in \mathbb{R}^{n \times n}$, can be written as

$$\mathbf{A} = \boldsymbol{\Gamma}\,\boldsymbol{\Delta}\,\boldsymbol{\Gamma}^\top \tag{1}$$

where:

(i) $\boldsymbol{\Delta}$ is a diagonal matrix, diag$(\lambda_1, \ldots, \lambda_n)$, of the eigen values of $\mathbf{A}$

- without loss of generality ordered so that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$.

(ii) $\boldsymbol{\Gamma}$ an orthogonal matrix with i^{th} column of $\boldsymbol{\Gamma}$ containing i^{th} standardized eigen-vector, $\boldsymbol{\gamma}_i$, of $\mathbf{A}$.

- "Standardized" means $\boldsymbol{\gamma}_i^\top \boldsymbol{\gamma}_i = 1$
- Orthogonality of $\boldsymbol{\Gamma}$ implies $\boldsymbol{\Gamma}\,\boldsymbol{\Gamma}^\top = \boldsymbol{\Gamma}^\top\,\boldsymbol{\Gamma} = \mathbf{I}_n$.

Spectral Decomposition

Since $\mathbf{\Sigma}$ is symmetric can take $\mathbf{A} = \mathbf{\Sigma}$ in (1).

Positive semi-definiteness of $\mathbf{\Sigma}$ implies $\lambda_i \geq 0$ for all $i = 1, \ldots, n$.

The principal components of $\mathbf{Y}$ then given by $\mathbf{P} = (P_1, \ldots, P_n)$ satisfying

$$\mathbf{P} = \mathbf{\Gamma}^\top \mathbf{Y}. \tag{2}$$

Note that:

(a) $\mathsf{E}[\mathbf{P}] = \mathbf{0}$ since $\mathsf{E}[\mathbf{Y}] = \mathbf{0}$
and

(b) $\mathsf{Cov}(\mathbf{P}) = \mathbf{\Gamma}^\top \mathbf{\Sigma} \mathbf{\Gamma} = \mathbf{\Gamma}^\top (\mathbf{\Gamma} \mathbf{\Delta} \mathbf{\Gamma}^\top) \mathbf{\Gamma} = \Delta$

So components of $\mathbf{P}$ are uncorrelated and $\mathsf{Var}(P_i) = \lambda_i$ are decreasing in i as desired.

Factor Loadings

The matrix $\mathbf{\Gamma}^\top$ is called the matrix of factor loadings.

Can invert (2) to obtain

$$\mathbf{Y} = \mathbf{\Gamma}\,\mathbf{P} \tag{3}$$

- so easy to go back and forth between $\mathbf{Y}$ and $\mathbf{P}$.

Explaining The Total Variance

Can measure the ability of the first few principal components to explain the total variability in the system:

$$\sum_{i=1}^{n} \mathsf{Var}(P_i) = \sum_{i=1}^{n} \lambda_i = \mathsf{trace}(\boldsymbol{\Sigma}) = \sum_{i=1}^{n} \mathsf{Var}(Y_i). \tag{4}$$

If we take $\sum_{i=1}^{n} \mathsf{Var}(P_i) = \sum_{i=1}^{n} \mathsf{Var}(Y_i)$ to measure the total variability then by (4) can interpret

$$\frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{n} \lambda_i}$$

as the percentage of total variability explained by first k principal components.

Explaining The Total Variance

Can also show that first principal component, $P_1 = \gamma_1^\top \mathbf{Y}$, satisfies

$$\mathsf{Var}(\gamma_1^\top \mathbf{Y}) \;=\; \max\left\{\mathsf{Var}(\mathbf{a}^\top \mathbf{Y}) \;:\; \mathbf{a}^\top \mathbf{a} \;=\; 1\right\}.$$

And that each successive principal component, $P_i = \gamma_i^\top \mathbf{Y}$, satisfies the same optimization problem but with the added constraint that it be orthogonal, i.e. uncorrelated, to $P_1, \ldots, P_{i-1}$.

Financial Applications of PCA

In financial applications, often the case that just two or three principal components are sufficient to explain anywhere from 60% to 95% or more of the total variability

- and often possible to interpret the first two or three components.

e.g. If $\mathbf{Y}$ represents (normalized) changes in the spot interest rate for n different maturities, then:

1. 1^{st} principal component can usually be interpreted as the (approximate) change in overall **level** of the yield curve
2. 2^{nd} component represents change in **slope** of the curve
3. 3^{rd} component represents change in **curvature** of the curve.

In equity applications, first component often represents a systematic market factor whereas the second (and possibly other) components may be identified with industry specific factors.

But generally less interpretability with equity portfolios and 2 or 3 principal components often **not** enough to explain most of overall variance.

Empirical PCA

In practice do not know true variance-covariance matrix but it may be estimated using historical data.

Suppose then we have multivariate observations, $\mathbf{X}_1, \ldots \mathbf{X}_m$

- $\mathbf{X}_t = (X_{t1} \ldots X_{tn})^\top$ represents the date t sample observation.

Important (why?) that these observations are from a stationary time series

- **e.g.** asset **returns** or yield **changes**
- But **not** price levels which are generally non-stationary.

If μ_j and σ_j are sample mean and standard deviation, respectively, of $\{X_{tj} \ : \ t = 1, \ldots, m\}$, then can **normalize** by setting

$$Y_{tj} \ = \ \frac{X_{tj} - \mu_j}{\sigma_j} \quad \text{for} \quad t = 1, \ldots, m \quad \text{and} \quad j = 1, \ldots, n.$$

Empirical PCA

Let $\mathbf{\Sigma}$ be the sample variance-covariance matrix so that

$$\mathbf{\Sigma} \;=\; \frac{1}{m} \sum_{t=1}^{m} \mathbf{Y}_t \, \mathbf{Y}_t^{\top}.$$

Principal components, $\mathbf{P}_t$, then computed using this covariance matrix.

From (3), see that original data obtained from principal components as

$$\begin{aligned} \mathbf{X}_t &= \text{diag}(\sigma_1, \ldots, \sigma_n) \, \mathbf{Y}_t \;+\; \boldsymbol{\mu} \\ &= \text{diag}(\sigma_1, \ldots, \sigma_n) \, \mathbf{\Gamma} \, \mathbf{P}_t \;+\; \boldsymbol{\mu} \end{aligned} \tag{5}$$

where $\mathbf{P}_t \;:=\; (P_{t1} \ldots P_{tn})^{\top} \;=\; \mathbf{\Gamma}^{\top} \mathbf{Y}_t$ is the t^{th} sample principal component vector.

Applications of PCA: Building Factor Models

If first k principal components explain sufficiently large amount of total variability then may partition the $n \times n$ matrix $\mathbf{\Gamma}$ according to $\mathbf{\Gamma} = [\mathbf{\Gamma}_1 \ \mathbf{\Gamma}_2]$ where $\mathbf{\Gamma}_1$ is $n \times k$ and $\mathbf{\Gamma}_2$ is $n \times (n-k)$.

Similarly can write $\mathbf{P}_t = [\mathbf{P}_t^{(1)} \ \mathbf{P}_t^{(2)}]^\top$ where $\mathbf{P}_t^{(1)}$ is $k \times 1$ and $\mathbf{P}_t^{(2)}$ is $(n-k) \times 1$.

May then use (5) to write

$$\mathbf{X}_{t+1} \quad = \quad \boldsymbol{\mu} \ + \ \text{diag}(\sigma_1, \ldots, \sigma_n) \, \mathbf{\Gamma}_1 \, \mathbf{P}_{t+1}^{(1)} \ + \ \boldsymbol{\epsilon}_{t+1} \tag{6}$$

where

$$\epsilon_{t+1} \ := \ \text{diag}(\sigma_1, \ldots, \sigma_n) \, \mathbf{\Gamma}_2 \, \mathbf{P}_{t+1}^{(2)} \tag{7}$$

represents an error term.

Can interpret (6) as a k-factor model for the changes in risk factors, $\mathbf{X}$

- but then take ϵ_{t+1} as an uncorrelated noise vector and ignore (7).

Applications of PCA: Scenario Generation

Easy to generate scenarios using PCA.

Suppose today is date t and we want to generate scenarios over the period $[t, t+1]$.

Can then use (6) to apply stresses to first few principal components, either singly or jointly, to generate loss scenarios.

Moreover, know that $\text{Var}(P_i) = \lambda_i$ so can easily control severity of the stresses.

Applications of PCA: Estimating VaR and CVaR

Can use the k-factor model and Monte-Carlo to simulate portfolio returns.

Could be done by estimating joint distribution of first k principal components

e.g. Could assume (why?)

$$\mathbf{P}_{t+1}^{(1)} \sim \mathsf{MN}_k(\mathbf{0},\ \mathsf{diag}(\lambda_1, \ldots, \lambda_k)).$$

but other heavy-tailed distributions may be more appropriate.

If we want to estimate the conditional loss distribution (as we usually do) of $\mathbf{P}_{t+1}^{(1)}$ then time series methods such as GARCH models should be used.

E.G. An Analysis of (Risk-Free) Yield Curves

- We use daily yields on U.S. Treasury bonds at 11 maturities: $T = 1$, 3, and 6 months and 1, 2, 3, 5, 7, 10, 20, and 30 years.
- Time period is January 2, 1990, to October 31, 2008.
- We use PCA to study how the curves change from day to day.
- To analyze daily changes in yields, all 11 time series were **differenced**.
- Daily yields were missing from some values of T for various reasons
  - e.g. the 20-year constant maturity series was discontinued at the end of 1986 and reinstated on October 1, 1993.
- All days with missing values of the differenced data were omitted.
  - this left 819 days of data starting on July 31, 2001, when the one-month series started and ending on October 31, 2008, with the exclusion of the period February 19, 2002 to February 2, 2006 when the 30-year Treasury was discontinued.
- The covariance matrix rather than the correlation matrix was used
  - which is fine here because the variables are comparable and in the same units.

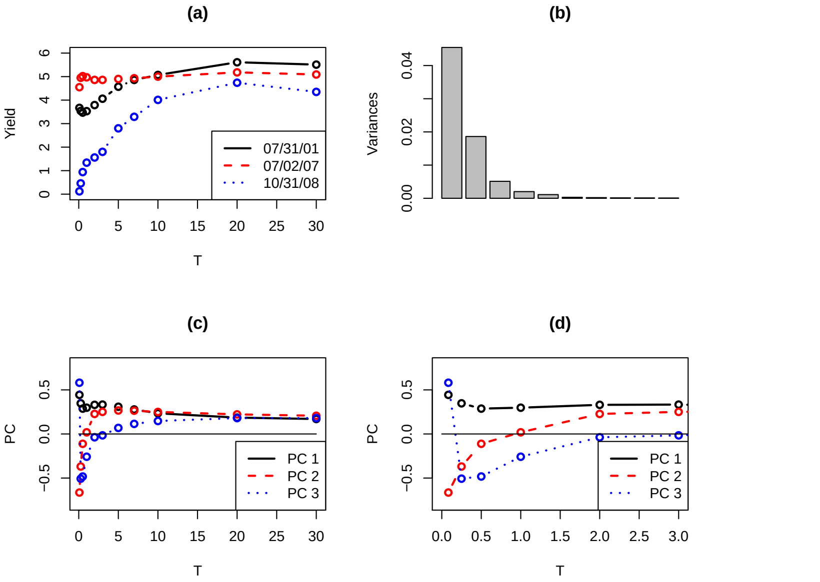

Figure 18.1 from Ruppert and Matteson: (a) Treasury yields on three dates. (b) Scree plot for the changes in Treasury yields. Note that the first three principal components have most of the variation, and the first five have virtually all of it. (c) The first three eigenvectors for changes in the Treasury yields. (d) The first three eigenvectors for changes in the Treasury yields in the range $0 \le T \le 3$.

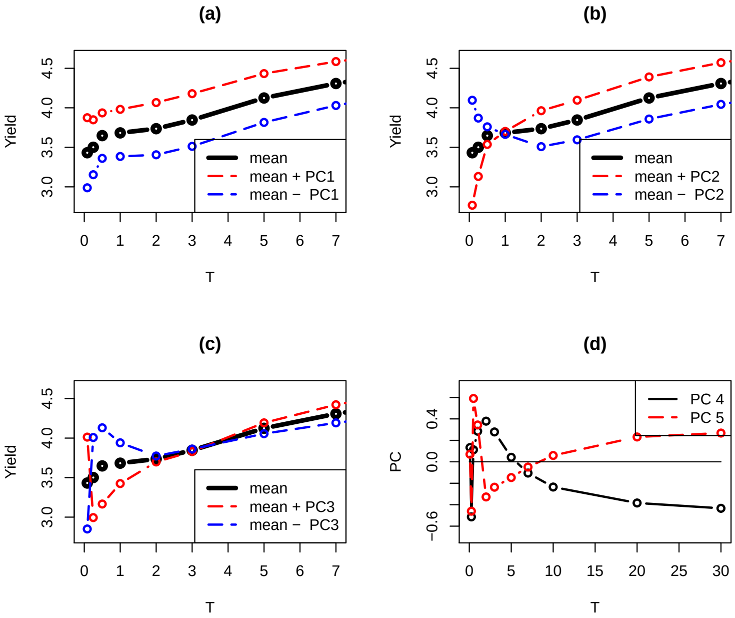

Figure 18.2 from Ruppert and Matteson: (a) The mean yield curve plus and minus the first eigenvector. (b) The mean yield curve plus and minus the second eigenvector. (c) The mean yield curve plus and minus the third eigenvector. (d) The fourth and fifth eigenvectors for changes in the Treasury yields.

E.G. An Analysis of (Risk-Free) Yield Curves

- Would actually be interested in the behavior of the yield changes over time.
- But time series analysis based on the changes in the 11 yields would be problematic.
  - better approach would be to use first three principal components.
- Their time series and **auto**- and **cross-correlation** plots are shown in Figs. 18.3 and 18.4, respectively.
- Notice that lag-0 cross-correlations are **zero**; this is not a coincidence! Why?
- Cross-correlations at nonzero lags are not zero, but in this example they are small
  - practical implication is that parallel shifts, changes in slopes, and changes in convexity are nearly **uncorrelated** and could be analyzed **separately**.
- The time series plots show substantial volatility clustering which could be modeled using GARCH models.

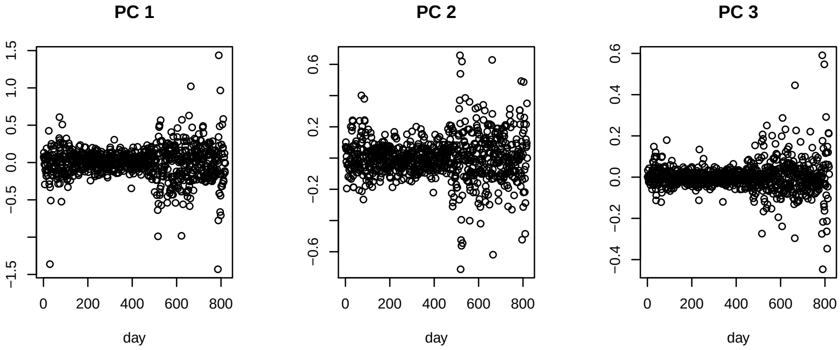

Figure 18.3 from Ruppert and Matteson: Time series plots of the first three principal components of the Treasury yields. There are 819 days of data, but they are not consecutive because of missing data; see text.

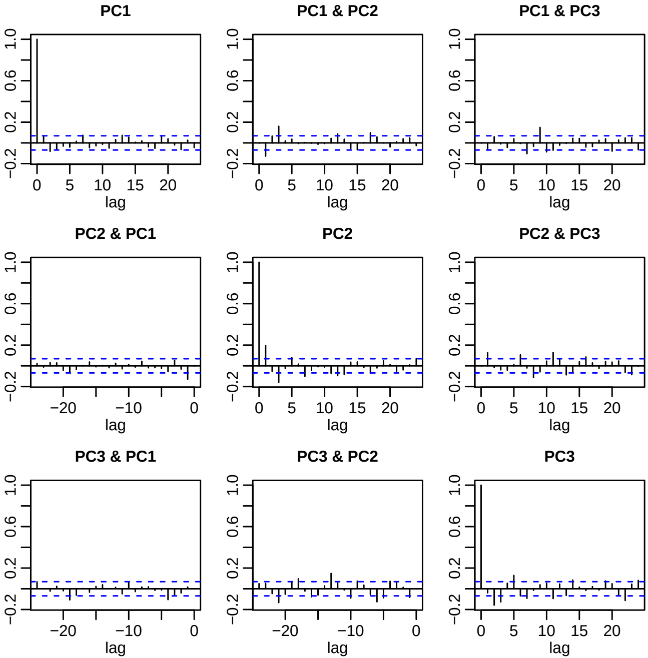

Figure 18.4 from Ruppert and Matteson: Sample auto- and cross-correlations of the first three principal components of the Treasury yields.

Applications of PCA: Portfolio Immunization

Also possible to hedge or immunize a portfolio against moves in the principal components.

e.g. Suppose we wish to hedge value of a portfolio against movements in the first k principal components.

Let $V_t =$ time t value of portfolio and assume our hedge will consist of positions, ϕ_{ti}, in the securities with time t prices, S_{ti}, for $i = 1, \ldots, k$.

Let $Z_{(t+1)j}$ be date $t+1$ level of the j^{th} risk factor

- so $\Delta Z_{(t+1)j} \;=\; X_{(t+1)j}$ using our earlier notation.

If change in value of **hedged** portfolio between dates t and $t+1$ is denoted by ΔV^*_{t+1}, then we have ....

Applications of PCA: Portfolio Immunization

$$
\begin{aligned}
\Delta V^*_{t+1} &\approx \sum_{j=1}^{n} \left(\frac{\partial V_t}{\partial Z_{tj}} + \sum_{i=1}^{k} \phi_{ti} \frac{\partial S_{ti}}{\partial Z_{tj}} \right) \Delta Z_{(t+1)j} \\
&= \sum_{j=1}^{n} \left(\frac{\partial V_t}{\partial Z_{tj}} + \sum_{i=1}^{k} \phi_{ti} \frac{\partial S_{ti}}{\partial Z_{tj}} \right) X_{(t+1)j} \\
&\approx \sum_{j=1}^{n} \left(\frac{\partial V_t}{\partial Z_{tj}} + \sum_{i=1}^{k} \phi_{ti} \frac{\partial S_{ti}}{\partial Z_{tj}} \right) \left(\mu_j + \sigma_j \sum_{l=1}^{k} \mathbf{\Gamma}_{jl} P_l \right) \\
&= \sum_{j=1}^{n} \left(\frac{\partial V_t}{\partial Z_{tj}} + \sum_{i=1}^{k} \phi_{ti} \frac{\partial S_{ti}}{\partial Z_{tj}} \right) \mu_j \\
&\quad + \sum_{l=1}^{k} \left(\sum_{j=1}^{n} \left(\frac{\partial V_t}{\partial Z_{tj}} + \sum_{i=1}^{k} \phi_{ti} \frac{\partial S_{ti}}{\partial Z_{tj}} \right) \sigma_j \mathbf{\Gamma}_{jl} \right) P_l \qquad (8)
\end{aligned}
$$

Applications of PCA: Portfolio Immunization

Can now use (8) to hedge the risk associated with the first k principal components.

In particular, we solve for the ϕ_{tl}'s so that the coefficients of the P_l's in (8) are zero

- a system of k linear equations in k unknowns so it is easily solved.

If we include an additional hedging asset then could also ensure that total value of hedged portfolio is equal to value of original un-hedged portfolio.

Factor Models

Definition: We say the random vector $\mathbf{X} = (X_1 \ldots X_n)^\top$ follows a linear k-factor model if it satisfies

$$\mathbf{X} \;=\; \mathbf{a} \;+\; \mathbf{B}\,\mathbf{F} \;+\; \boldsymbol{\epsilon} \tag{9}$$

where

(i) $\mathbf{F} = (F_1 \ldots F_k)^\top$ is a random vector of common factors with $k < n$ and with a positive-definite covariance matrix;

(ii) $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)$ is a random vector of idiosyncratic error terms which are uncorrelated and have mean zero;

(iii) $\mathbf{B}$ is an $n \times k$ constant matrix of factor loadings, and $\mathbf{a}$ is an $n \times 1$ vector of constants;

(iv) $\text{Cov}(F_i, \epsilon_j) = 0$ for all i, j.

Factor Models

If $\mathbf{X} \sim \mathsf{MN}(\cdot,\cdot)$ and follows (9) then possible to find a version of the model where $\mathbf{F} \sim \mathsf{MN}(\cdot,\cdot)$ and $\boldsymbol{\epsilon} \sim \mathsf{MN}(\cdot,\cdot)$.

- In this case the error terms, ϵ_i, are independent.

If $\boldsymbol{\Omega}$ is the covariance matrix of $\mathbf{F}$ then covariance matrix, $\boldsymbol{\Sigma}$, of $\mathbf{X}$ satisfies (why?)

$$\boldsymbol{\Sigma} = \mathbf{B}\,\boldsymbol{\Omega}\,\mathbf{B}^\top + \boldsymbol{\Upsilon}$$

where $\boldsymbol{\Upsilon}$ is a diagonal matrix of the variances of $\boldsymbol{\epsilon}$.

Exercise

Show that if (9) holds then there is also a representation

$$\mathbf{X} = \mu + \mathbf{B}^* \ \mathbf{F}^* + \boldsymbol{\epsilon} \tag{10}$$

where

$$\begin{aligned} \mathsf{E}[\mathbf{X}] &= \boldsymbol{\mu} \quad \text{and} \\ \mathsf{Cov}(\mathbf{F}^*) &= \mathbf{I}_k \end{aligned}$$

so that $\boldsymbol{\Sigma} = \mathbf{B}^* \ (\mathbf{B}^*)^\top + \boldsymbol{\Upsilon}$.

Example: Factor Models Based on Principal Components

Factor model of (6) may be interpreted as a k-factor model with

$$\begin{aligned} \mathbf{F} &= \mathbf{P}^{(1)} \quad \text{and} \\ \mathbf{B} &= \text{diag}(\sigma_1, \ldots, \sigma_n)\, \mathbf{\Gamma}_1. \end{aligned}$$

As constructed, covariance of ϵ in (6) is not diagonal

- so it does not satisfy part (ii) of definition above.

Nonetheless, quite common to construct factor models in this manner and to then make the assumption that ϵ is a vector of uncorrelated error terms.

Calibration Approaches

Three different types of factor models:

1. **Observable Factor Models**
   - Factors, $\mathbf{F}_t$, have been identified in advance and are observable.
   - They typically have a fundamental economic interpretation e.g. a 1-factor model where market index plays role of the single factor.
   - These models are usually calibrated and tested for goodness-of-fit using multivariate or time-series regression techniques.

   e.g. A model with factors constructed from change in rate of inflation, equity index return, growth in GDP, interest rate spreads etc.

2. **Cross-Sectional Factor Models**
   - Factors are unobserved and therefore need to be estimated.
   - The factor loadings, $\mathbf{B}_t$, are observed, however.

   e.g. A model with dividend yield, oil and tech factors. We assume the factor returns are unobserved but the loadings are known. Why?

   e.g. `BARRA`'s factor models are generally cross-sectional factor models.

3. **Statistical Factor Models**
   - Both factors and loadings need to be estimated.
   - Two standard methods for doing this: factor analysis and PCA.

Factor Models in Risk Management

Straightforward to use a factor model to manage risk.

For a given portfolio composition and fixed matrix, $\mathbf{B}$, of factor loadings, the sensitivity of the total portfolio value to each factor, F_i for $i = 1, \ldots, k$, is easily computed.

Can then adjust portfolio composition to achieve desired overall factor sensitivity.

Process easier to understand and justify when the factors are easy to interpret.

When this is not the case then the model is purely statistical.

- Tends to occur when statistical methods such as factor analysis or PCA are employed
- But still possible even then for identified factors to have an economic interpretation.

E.G. Scenario Analysis for an Options Portfolio

Mr. Smith has a portfolio consisting of various stock positions as well as a number of equity options.

He would like to perform a basic scenario analysis with just two factors:

1. An equity factor, F_{eq}, representing the equity market.
2. A volatility factor, F_{vol}, representing some general implied volatility factor.

Can perform the scenario analysis by stressing combinations of the factors and computing the P&L resulting from each scenario.

Of course using just two factors for such a portfolio will result in a scenario analysis that is quite coarse but in many circumstances this may be sufficient.

Question: How do we compute the value of each security in each scenario?

Question: This is easy if we adopt a factor model framework

- see lecture notes for further details.