Financial Econometrics
Lecture 11: Factor Models and Principle Component Analysis (PCA)

Prof Hamed Ghoddusi
2019

Big Picture

- The goal is to explain the variation in the return of financial asset
- Dimension reduction
  - Unobserved (latent) factors: Factor Analysis (FA)
  - Common modes of variation: Principle Component Analysis (PCA)
- Instead of using the past history of return, we use other variables (factors) as the regressors

Factor Models

- Factor models: statistical models that try to explain complex phenomena through a small number of basic causes or factors.
- Two main purposes of factor models:
  1. Reducing the dimensionality of models to make estimation possible;
  2. Finding the true causes that drive data.
- Introduced by Charles Spearman in 1904.

Factor Models: Example

- Explains intellectual abilities through one common factor, the famous "general intelligence" g factor, plus another factor s which is specific to each distinct ability
- Louis Leon Thurstone developed the first true multifactor model of intelligence, where were identified the following seven primary mental abilities:

Verbal Comprehension	Word Fluency
Number Facility	Spatial Visualization
Associative Memory	Perceptual Speed
Reasoning.	

Factor Models

- Social science and psychology: the objective is to explain behavior or traits as probability distributions conditional on the value of one or more factors.
  - Example: one can make predictions of the future success of young individuals in different activities.
- Economics and finance: factor models are typically applied to time-series. The objective is to explain the behavior of a large number of stochastic processes, typically price, returns, or rate processes, in terms of a small number of factors.
  - The factors are themselves stochastic processes.
  - Examples: market factor, HML, SMB, Momentum, ...
- In finance: to explain all pairwise correlations in terms of a much smaller number of correlations between factors.

Factor Models

- Suppose there are k assets (most often stocks), and T periods. Let r_{it} be the (excess) return of asset i at time t. Throughout this note we assume r_{it} is stationary.
- A general form of the factor model is

$$r_{it} = \beta_{0i} + \beta_{1i} f_{1t} + \cdots + \beta_{mi} f_{mt} + e_{it} \tag{1}$$

  where we assume there are m factors, and f_{jt} is the j-th factor at time t.
- To distinguish various factor models, the key is to pay attention to the subscript.

Remarks

1. f_{jt} has the time subscript t. That means the factors are time-varying
2. f has no subscript i. That means the factors are common for all assets.
3. β has the subscript i. So the beta is asset-specific. That is, we need to run regression (1) separately for each asset return.
4. If we put k regressions together, it becomes a special case of seemingly unrelated regression (SUR). Because each regression has the same regressors (factors), the GLS estimator for this special SUR is the same as the equation-by-equation OLS estimator.

Example: one-factor model

- There are $k = 13$ monthly excess returns of stocks in the data file

  `672_2014_factor1.txt`

  from Jan 1990 to Dec 2003 (so T=168)
- The single factor (so $m = 1$) is the excess return of the S&P500 index, which approximates the market return.
- We regress each return series onto the S&P500 index. There are in total 13 regressions. The coefficients in each regressions are different.
- We want to find the association between the return of each individual asset and the market.

R Code

```
data = read.table("672_2014_factor1.txt", header=T)
n = 168
c = rep(1, n) # intercept term
X = as.matrix(cbind(c, data[,14]))
Y = as.matrix(data[,1:13])
bet = matrix(0, 2, 13); ehat = matrix(0, n, 13)
for (i in 1:13) {
bet[,i] = solve(t(X)%*%X)%*%(t(X)%*%Y[,i])
ehat[,i] = Y[,i] - X%*%bet[,i]
}
rss = diag(crossprod(ehat)) # RSS
rsq = 1 - rss/diag(var(Y)*(n-1)) # R-Squared
```


Remarks

The key is the a loop of OLS estimation

```
for (i in 1:13) {
bet[,i] = solve(t(X)%*%X)%*%(t(X)%*%Y[,i])
ehat[,i] = Y[,i] - X%*%bet[,i]
}
```

1. For the return of the i-th stock, the coefficient estimate is

$$\beta_i = (X'X)^{-1}(X'Y_i)$$

   where the dimension of X is $168 \times 2; Y_i$ is 168×1; and β_i is 2×1.

2. The residual is

$$E_i = Y_i - X\beta_i$$

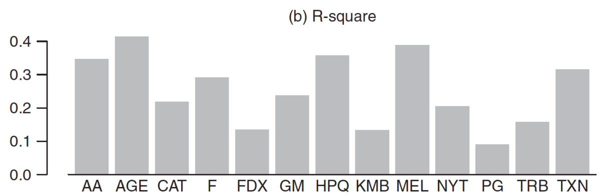
Result

	Beta	Sigma	R-Squared
AA	1.292	7.694	0.347
AGE	1.514	7.807	0.415
CAT	0.941	7.724	0.219
F	1.219	8.241	0.292
FDX	0.805	8.854	0.135
GM	1.046	8.130	0.238
HPQ	1.628	9.469	0.358
KMB	0.550	6.070	0.134
MEL	1.123	6.120	0.388
NYT	0.771	6.590	0.205
PG	0.469	6.459	0.090
TRB	0.718	7.215	0.157
TXN	1.796	11.474	0.316

Example: Factor Model

```
> t(rbind(beta.hat,sqrt(D.hat),r.square))
     beta.hat  sigma(i)  r.square
 AA     1.292     7.694     0.347
AGE     1.514     7.808     0.415
CAT     0.941     7.725     0.219
  F     1.219     8.241     0.292
FDX     0.805     8.854     0.135
 GM     1.046     8.130     0.238
HPQ     1.628     9.469     0.358
KMB     0.550     6.070     0.134
MEL     1.123     6.120     0.388
NYT     0.771     6.590     0.205
 PG     0.469     6.459     0.090
TRB     0.718     7.215     0.157
TXN     1.796    11.474     0.316
```

R2 of Factor Model



Remarks

For example, for the stock of Alcoa (AA is its Tick or Ticker symbol), we find

1. its return is positively correlated with the market return because $\hat{\beta}_{1,AA} = 1.292 > 0$
2. the estimated standard error of regression (SER) σ is 7.694
3. the R^2 is 0.347. So the market return can explain around 34% of the variation of AA stock return.

GMVP

- Using *global minimum variance portfolio* (GMVP) to compare the covariance matrix implied by a fitted factor model with the sample covariance matrix of the returns.
- For a given covariance matrix Σ, the global minimum variance portfolio is the portfolio ω that solves

$$\min_{\omega} \sigma^2_{P,\omega} = \omega' \Sigma \omega \quad \text{such that} \quad \omega' \mathbf{1} = 1$$

and is given by

$$\omega = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}},$$

where $\mathbf{1}$ is the k-dimensional vector of ones.

GMVP

- For the market model considered, the GMVP for the fitted model and the data are as follows:

```
> w.gmin.model=solve(cov.r)%*%rep(1,nrow(cov.r))
> w.gmin.model=w.gmin.model/sum(w.gmin.model)
> t(w.gmin.model)
      AA AGE CAT F FDX GM
[1,] 0.0117 -0.0306 0.0792 0.0225 0.0802 0.0533
      HPQ KMB MEL NYT PG TRB TXN
[1,] -0.0354 0.2503 0.0703 0.1539 0.2434 0.1400 -0.0388
> w.gmin.data=solve(var(rtn))%*%rep(1,nrow(cov.r))
> w.gmin.data=w.gmin.data/sum(w.gmin.data)
> t(w.gmin.data)
      AA AGE CAT F FDX GM
[1,] -0.0073 -0.0085 0.0866 -0.0232 0.0943 0.0916
      HPQ KMB MEL NYT PG TRB TXN
[1,] 0.0345 0.2296 0.0495 0.1790 0.2651 0.0168 -0.0080
```

GMVP

- Comparing the two GMVPs, the weights assigned to TRB stock differ markedly.
- The two portfolios, however, have larger weights for KMB, NYT, and PG stocks.
- Finally, we examine the residual covariance and correlation matrices to verify the assumption that the special factors are not correlated among the 13 stocks.
- The first four columns of the residual correlation matrix are given below and there exist some large values in the residual cross-correlations, for example, Cor(CAT,AA) = 0.45 and Cor(GM,F) = 0.48.

GMVP

```
> resi.cov=t(E.hat)%*%E.hat/(168-2)
> resi.sd=sqrt(diag(resi.cov))
> resi.cor=resi.cov/outer(resi.sd,resi.sd)
> print(resi.cor,digits=1,width=2)
       AA    AGE   CAT      F
AA   1.00  -0.13  0.45   0.22
AGE -0.13   1.00 -0.03  -0.01
CAT  0.45  -0.03  1.00   0.23
F    0.22  -0.01  0.23   1.00
FDX  0.00   0.14  0.05   0.07
GM   0.14  -0.09  0.15   0.48
HPQ  0.24  -0.13 -0.07  -0.00
KMB  0.16   0.06  0.18   0.05
MEL -0.02   0.06  0.09   0.10
NYT  0.13   0.10  0.07   0.19
PG  -0.15  -0.02 -0.01  -0.07
TRB  0.12  -0.02  0.25   0.16
TXN  0.19  -0.17  0.09  -0.02
```


BARRA Factor Model

1. The BARRA factor model is
$$r_t = \beta f_t + e_t \tag{2}$$
where $r_t = (r_{1t}, \ldots, r_{kt})'$. β is given, and is a set of industry dummy variables.
2. f_t is called factor realization, and is unknown here, and needs to be estimated.
3. In short, we need to run the above regression repeatedly for each period. For a given period, the dependent variable is the returns of all assets at that period. The regressors are a set of time-invariant industry-dummies. The estimated coefficient is the factor realization at that period.

Example: BARRA Factor Model

- There are $k = 10$ monthly excess returns of assets in the data file
  `672_2014_factor2.txt`
  from Jan 1990 to Dec 2003 (so T=168)
- The regressors are three industry dummy variables:

```
d.fin = c(rep(1,4),rep(0,6))
d.it = c(rep(0,4),rep(1,3),rep(0,3))
d.ot = c(rep(0,7),rep(1,3))
```

For example, d.fin equals one if a company is in the financial sector. In this case, AGE, C, MWD and MER are in financial sector, and other firms are not. So d.fin has four ones and six zeros.

OLS Estimator

```
Y = as.matrix(data[,1:10])
X = cbind(d.fin, d.it, d.ot)
f.o = solve(crossprod(X))%*%(t(X)%*%t(Y)) # OLS
f.o[,1:2]
```

Note that

1. you get transpose of the data matrix using $t(Y)$.
2. so each row of Y (or each column of $t(Y)$) is used as the dependent variable
3. in the first period $t = 1$, for instance, the coefficient of d.fin is -10.80. In the second period, the coefficient of d.fin is 2.212500. There are in total 168 coefficients of d.fin. That coefficient series is the factor realization for the financial dummy.

OLS and Industrial Average

Because X includes three industry dummies (and no intercept term). The estimated f_t is

$$f_t = \begin{pmatrix} \frac{AGE_t+Ct+MWD_t+MER_t}{4} \\ \frac{DELL_t+HPQ_t+IBM_t}{3} \\ \frac{AA_t+CAT_t+PG_t}{3} \end{pmatrix}$$

So each component of f_t is the average of returns of firms in that specific industry in the t-th period.

WLS Estimator

1. The OLS estimator is inefficient because it ignores the heteroskedasticity

$$\Omega = Ee_te_t' = \text{diag}(\sigma_1^2, \ldots, \sigma_k^2) \neq \sigma^2 I$$

where σ_k^2 is the variance of the k-th asset return.

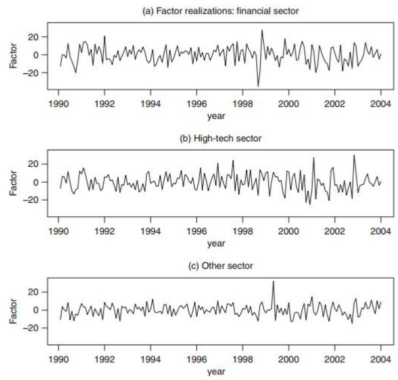
2. We can estimate Ω using the OLS residual. Then the more efficient weighted least squares (WLS) estimator is

$$f_t^{\text{WLS}} = (\beta'\widehat{\Omega}^{-1}\beta)^{-1}(\beta'\widehat{\Omega}^{-1}r_t)$$

R Code

```
ehat.o = t(Y) - X%*%f.o
omega = var(t(ehat.o))
weight = diag(omega^(-1))
w.m = Diagonal(10, x = weight)
x.wls = t(X)%*%w.m%*%X
y.wls = t(X)%*%w.m%*%t(Y)
f.wls = solve(x.wls)%*%y.wls # WLS
```

Factor Mimicking Portfolio



Estimated factor realizations of a BARRA industrial factor model for 10 monthly stock returns in three industrial sectors.

Factor Mimicking Portfolio

- Consider the special case of BARRA factor models with a single factor.
- Consider a portfolio $\omega = (\omega_1, \ldots, \omega_k)'$ of the k assets that solves

$$\min_{\omega}(\frac{1}{2}\omega' D\omega) \quad \text{such that} \quad \omega'\beta = 1.$$

- It turns out that the solution to this portfolio problem is given by

$$\omega' = (\beta' D^{-1}\beta)^{-1}(\beta' D^{-1}).$$

- Thus, the estimated factor realization is the portfolio return

$$\hat{f}_t = \omega' r_t.$$

Factor Mimicking Portfolio

- If the portfolio ω is normalized such that $\sum_{i=1}^{k}\omega_i = 1$, it is referred to as a $factor\ mimicking\ portfolio$. For multiple factors, one can apply the idea to each factor individually.

  Remark. In practice, the sample mean of an excess return is often not significantly different from zero. Thus, one may not need to remove the sample mean before fitting a BARRA factor model.

Fama-French Approach

- For a given asset fundamental (e.g., ratio of book-to-market value), Fama and French (1992) determined factor realizations using a two-step procedure.
  - First, they sorted the assets based on the values of the observed fundamental
  - Then they formed a hedge portfolio which is long in the top quintile (1/3) of the sorted assets and short in the bottom quintile of the sorted assets.
- The observed return on this hedge portfolio at time t is the observed factor realization for the given asset fundamental.
- The procedure is repeated for each asset fundamental under consideration.

Fama-French Approach

- Finally, given the observed factor realizations $\{f_t | t = 1, \ldots, T\}$, the betas for each asset are estimated using a time series regression method.
- These authors identify three observed fundamentals that explain high percentages of variability in excess returns.
- The three fundamentals used by Fama and French are:
  (a) the overall market return (market excess return)
  (b) the performance of small stocks relative to large stocks (SMB, small minus big)
  (c) the performance of value stocks relative to growth stocks (HML, high minus low)
- The size sorted by market equity and the ratio of book equity to market equity is used to define value and growth stocks with value stocks having high book equity to market equity ratio.

Principal Components Analysis

- Introduced in 1933 by Harold Hotelling as a way to determine factors with statistical learning techniques when factors are not exogenously given.
- One way to imagine the concept of PCA is the following.
  - Consider a set of n stationary time series X_i.
  - Consider next a linear combination of these series, that is, a portfolio of securities.
  - Each portfolio P is identified by an n-vector of weights ω_P and is characterized by a variance σ_P^2 .
  - Consider a normalized portfolio, which has the largest possible variance. In this context, a normalized portfolio is a portfolio such that the squares of the weights sum to one.

Principal Components Analysis

- If we assume that returns are IID sequences, jointly normally distributed with variance-covariance matrix σ, a lengthy direct calculation demonstrates that each portfolios return will be normally distributed with variance
$$\sigma_P^2 = \omega_P^T \sigma \omega_P$$
- The normalized portfolio of maximum variance can therefore be determined in the following way: Maximize $\omega_P^T \sigma \omega_P$ subject to the normalization condition
$$\omega_P^T \omega_P = 1$$
where the product is a scalar product.
- It can be demonstrated that the solution of this problem is the eigenvector ω_2 corresponding to the largest eigenvalue λ_2 of the variance-covariance matrix σ.

Statistical Factor Model

1. In the statistical factor model, the factors are the principle components (PC) of the return series.
2. Consider a linear combination (portfolio) of k returns at the t-th period
$$\sum_{i=1}^{k} w_i r_{it}$$
where the weight for the i-th asset is w_i.
3. PC is a special linear combination so that
   (a) each PC is uncorrelated with each other
   (b) each PC will explain the maximum amount of remaining variance-covariance of r_t after the previous PC.

Eigenvalue and Eigenvector

1. For a square matrix A, its eigenvalue solves the equation

$$A\lambda = \lambda c,$$

   where c is a (nonzero) column vector, called the eigenvector

2. Mathematically, there are k (possibly duplicate or complex) eigenvalues if the dimension of A is $k \times k$. They are obtained as the roots of

$$\text{determinant}(A - \lambda I) = 0$$

Principle Component (PC)

1. Denote the variance-covariance matrix of k returns by Ω.
2. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k$ be the eigenvalues of Ω in descending order
3. Let $c_1, c_2, \ldots, c_k$ be the corresponding eigenvectors
4. We can prove that the j-th PC is the j-th eigenvector multiplied by the return series, and the variance of the j-th PC is λ_j.

Principal Components Analysis

- PCA works either on the variance-covariance matrix or on the correlation matrix. The technique is the same but results are generally different.
- PCA applied to the variance-covariance matrix is sensitive to the units of measurement, which determine variances and covariances. If PCA is applied to prices and not to returns, the currency in which prices are expressed matters; one obtains different results in different currencies. In these cases, it might be preferable to work with the correlation matrix.
- PCA is a generalized dimensionality reduction technique applicable to any set of multidimensional observations. It admits a simple geometrical interpretation which can be easily visualized in the three-dimensional case.

Example: Statistical Factor Model

- There are $k = 5$ monthly excess returns of assets in the data file

  `672_2014_factor3.txt`

  from Jan 1990 to Dec 1999 (so T=120)

- The eigenvalues and eigenvectors are

```
data = read.table("672_2014_factor3.txt", header=T)
data = data[1:120,]
v.m = var(data)
e.value = eigen(v.m)$values
e.vector = eigen(v.m)$vector
```

Principle Components

```
pc1 = as.matrix(data)%*%e.vector[,1]
pc2 = as.matrix(data)%*%e.vector[,2]
pc3 = as.matrix(data)%*%e.vector[,3]
var(pc1)
var(pc2)
var(pc3)
cor(cbind(pc1, pc2, pc3))
```

We can verify the three PCs are mutually uncorrelated, and the variance of each PC is the corresponding eigenvalues.

Statistical Factor Model

```
summary(lm(data[,1]~pc1+pc2+pc3))
summary(lm(data[,5]~pc1+pc2+pc3))
lm(formula = data[, 5] ~ pc1 + pc2 + pc3)
```

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.32290    0.40212 -0.803     0.424
pc1          -0.46613    0.02456 -18.981   <2e-16 ***
pc2          -0.48495    0.03759 -12.902   <2e-16 ***
pc3           0.03609    0.04771   0.756    0.451
```

For instance, we find all three PCs are significant in explaining the first asset. However, for the 5th asset, only the first two PCs are significant.

Standard Outputs of any PCA Code

- A set of eigenvalues: information of the variability in the data
- A table with the scores or Principal Components (PCs): the structure of the observations.
- A table of loadings (or correlations between variables and PCs): allow you to get a sense of the relationships between variables, as well as their associations with the extracted PCs.

PCA Using the Covariance Matrix

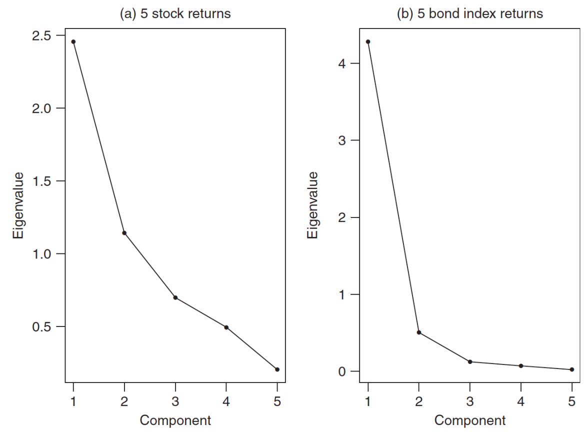
Using Sample Covariance Matrix					
Eigenvalue	256.16	116.14	64.91	46.82	22.11
Proportion	0.506	0.229	0.128	0.093	0.044
Cumulative	0.506	0.736	0.864	0.956	1.000
Eigenvector	0.246	0.327	0.586	−0.700	0.018
	0.461	0.360	0.428	0.687	−0.050
	0.409	0.585	−0.683	−0.153	0.033
	0.522	−0.452	−0.082	−0.115	−0.710
	0.536	−0.467	−0.036	−0.042	0.701

PCA Using the Correlation Matrix

Using Sample Correlation Matrix

Eigenvalue	2.456	1.145	0.699	0.495	0.205
Proportion	0.491	0.229	0.140	0.099	0.041
Cumulative	0.491	0.720	0.860	0.959	1.000
Eigenvector	0.342	0.525	0.691	−0.362	−0.012
	0.474	0.314	−0.043	0.820	0.050
	0.387	0.405	−0.717	−0.414	−0.034
	0.503	−0.481	0.052	−0.147	0.701
	0.505	−0.481	0.071	−0.062	−0.711

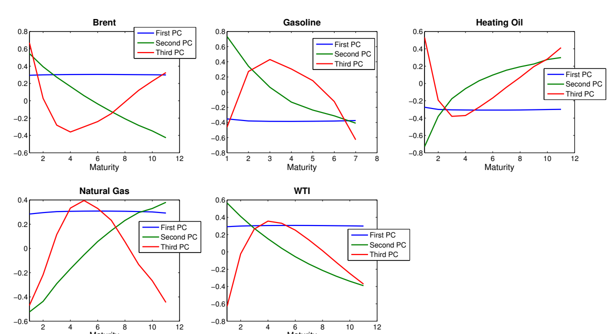
Eigen Values from PCA



PCA's in Financial Data

- PCA applied to term-structure data (e.g. futures prices of a commodity or interests rates),
- It is common to observe three major factors
  - Level factor
  - Slope factor
  - Curvature factor

Example: PCA of Commodity Roll Returns



Applying PCA to Bond Portfolio Management

- There are two applications in bond portfolio management where PCA has been employed.
- Application 1: explaining the movement or dynamics in the yield curve and then applying the resulting principal components to measure and manage yield curve risk.
- Application 2: identify risk factors beyond changes in the term structure.
- Given historical bond returns and factors that are believed to affect bond returns, PCA can be used to obtain principal components that are linear combinations of the variables that explain the variation in returns.

Using PCA to Control Interest Rate Risk

- Using PCA, several studies have investigated the factors that have affected the historical returns on Treasury portfolios.
- Robert Litterman and Jose Scheinkman: three factors explained historical bond returns for U.S. Treasuries zero-coupon securities:
  1. The changes in the level of rates;
  2. The changes in the slope of the yield curve;
  3. The changes in the curvature of the yield curve.
- After identifying the factors, Litterman and Scheinkman use regression analysis to assess the relative contribution of these three factors in explaining the returns on zero-coupon Treasury securities of different maturities.

Using PCA to Control Interest Rate Risk

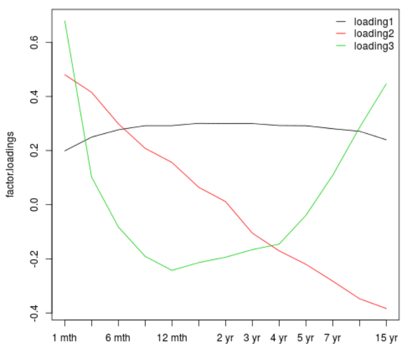
- On average, the first principal component explained about 90% of the returns, the second principal component 8%, and the third principal component 2%.
- Thus, only three principal components were needed to fully explain the dynamics of the yield curve.
- There have been several studies that have examined the yield curve movement using PCA and reported similar results.

Using PCA to Control Interest Rate Risk

Once yield curve risk is described in terms of principal components, the factor loadings can be used to:

- Construct hedges that neutralize exposure to changes in the direction of interest rates.
- Construct hedges that neutralize exposure to changes in nonparallel shifts in the yield curve.
- Structure yield curve trades.

Typical Loading of a Bond PCA



Factor Rotation

- As mentioned before, for any $m \times m$ orthogonal matrix P ,

$$r_t - \mu = \beta f_t + \epsilon_t = \beta^* f_t^* + \epsilon_t,$$

where $\beta^* = \beta P$ and $f_t^* = P' f_t$. In addition,

$$\beta\beta' + D = \beta P P' \beta' + D = \beta^* (\beta^*)' + D.$$

- This result indicates that the communalities and the specific variances remain unchanged under an orthogonal transformation.
- It is then reasonable to find an orthogonal matrix P to transform the factor model so that the common factors have nice interpretations.

Factor Rotation

- Such a transformation is equivalent to rotating the common factors in the m-dimensional space.
- In fact, there are infinite possible factor rotations available.
- Kaiser (1958) proposes a *varimax* criterion to select the rotation that works well in many applications.
- Denote the rotated matrix of factor loadings by $\beta^* = [\beta^*_{ij}]$ and the ith communality by c_i^2.
- Define $\tilde{\beta}^*_{ij} = \beta^*_{ij}/c_i$ to be the rotated coefficients scaled by the (positive) square root of communalities.

Factor Rotation

- The varimax procedure selects the orthogonal matrix P that maximizes the quantity

$$V = \frac{1}{k} \sum_{j=1}^{m} \left[\sum_{i=1}^{k} (\tilde{\beta}^*_{ij})^4 - \frac{1}{k} \Big(\sum_{i=1}^{k} \tilde{\beta}^{*2}_{ij} \Big)^2 \right].$$

- This complicated expression has a simple interpretation. Maximizing V corresponds to spreading out the squares of the loadings on each factor as much as possible.
- Consequently, the procedure is to find groups of large and negligible coefficients in any column of the rotated matrix of factor loadings.

Factor Rotation

- In a real application, factor rotation is used to aid the interpretations of common factors.
- It may be helpful in some applications, but not informative in others.
- There are many criteria available for factor rotation.

PCA and Factor Analysis Compared

- The two illustrations of PCA and FA are relative to the same data and will help clarify the differences between the two methods.
- Lets first observe that PCA does not imply any specific restriction on the process. Given a nonsingular covariance matrix, we can always perform PCA as an exact linear transformation of the series. When we consider a smaller number of principal components, we perform an approximation which has to be empirically justified.
- Factor analysis, on the other hand, assumes that the data have a strict factor structure in the sense that the covariance matrix of the data can be represented as a function of the covariances between factors plus idiosyncratic variances.

PCA and Factor Analysis Compared

- PCA tends to be a dimensionality reduction technique that can be applied to any multivariate distribution and that yields incremental results. This means that there is a trade-off between the gain in estimation from dimensionality reduction and the percentage of variance explained.
- Factor analysis, on the other hand, tends to reveal the exact factor structure of the data. That is, FA tends to give an explanation in terms of what factors explain what processes.
- Factor rotation can be useful both in the case of PCA and FA.