# Lecture 11
# Multi-factor Models:
# Statistical Factor Models (PCA)

Lars A. Lochstoer
UCLA Anderson School of Management

Winter 2025

# Overview of Lecture 11

1. Statistical Factor Models: Principal Components Analysis
    1. Math
    2. Properties
    3. Estimation
    4. Application: Modeling the Yield Curve
    5. Application: Factors in stock returns

# Motivation

- use factor models to **fight** the curse of dimensionality
- find a couple of factors that explain most of the variation
- factor models can naturally arise from economic theory (next lecture)
- up to this slide, factors are observable.
  - market index
  - portfolios sorted on accounting variables
- **statistical factor models**: choice of factors determined by data.
  - principal component analysis (PCA)

# Factor Model

- model returns

$$r_{it} = \alpha_i + \boldsymbol{\beta}_i' \mathbf{f}_t + \varepsilon_{it}, \qquad i = 1, \ldots, N, \quad t = 1, \ldots, T$$

- using matrix notation

$$\mathbf{r}_t = \boldsymbol{\alpha} + \boldsymbol{\beta} \mathbf{f}_t + \boldsymbol{\varepsilon}_t, \qquad t = 1, \ldots, T$$

- $\mathbf{r}_t$ is a $N \times 1$ vector of returns.
- $\boldsymbol{\beta}$ is a $N \times K$ matrix of factor loadings.
- the covariance matrix of returns:

$$\text{Cov}(\mathbf{r}_t) \; = \; \Sigma_{\mathbf{r}} \; = \; \boldsymbol{\beta} \boldsymbol{\Sigma}_f \boldsymbol{\beta}' + \boldsymbol{\Sigma}_{\varepsilon}$$

- assume $\Sigma_{\varepsilon}$ is diagonal.

# Principal Components Analysis (PCA)

- given a vector of returns $\mathbf{r}_t = (r_{1t}, r_{2t}, \ldots, r_{Nt})$, PCA tries to use a few linear combinations to explain the structure of $\mathbf{\Sigma}_r$
- a few means less than $N$
- let $\mathbf{w}_i$ be a $N \times 1$ vector of weights, $i = 1, 2, \ldots, N$
- we construct linear combinations of the random vectors:

$$y_{it} = \mathbf{w}_i' \mathbf{r}_t = \sum_{j=1}^{N} w_{ij} r_{jt}$$

- in financial applications, we can conveniently interpret $y_{it}$ as a portfolio return.
- we normalize the weights such that $\mathbf{w}_i' \mathbf{w}_i = 1$

# Properties

- moments of $y_{it}$:

$$V\left(y_{it}\right) = \mathbf{w}_i' \boldsymbol{\Sigma}_{\mathbf{r}} \mathbf{w}_i, \qquad i = 1, 2, \ldots, N$$

$$\text{Cov}\left(y_{it}, y_{jt}\right) = \mathbf{w}_i' \boldsymbol{\Sigma}_{\mathbf{r}} \mathbf{w}_j, \qquad i, j = 1, 2, \ldots, N$$

- idea of PCA: find linear combinations (the vectors $\mathbf{w}_i$) that are uncorrelated and with variances of $y_{it}$ that are as large as possible

# PCA (cont'd)

1. 1st principal component: maximizes the variance of $y_1$ subject to $\mathbf{w}_1'\mathbf{w}_1 = 1$
2. 2nd principal component: maximizes the variance of $y_2$ subject to $\mathbf{w}_2'\mathbf{w}_2 = 1$ and $\text{Cov}(y_1, y_2) = 0$
3. 3rd principal component: maximizes the variance of $y_3$ subject to $\mathbf{w}_3'\mathbf{w}_3 = 1$ and $\text{Cov}(y_3, y_i) = 0,\ i = 1, 2$
4. $k$-th principal component: maximizes the variance of $y_k$ subject to $\mathbf{w}_k'\mathbf{w}_k = 1$ and $\text{Cov}(y_r, y_i) = 0,\ i = 1, 2, \ldots, k-1$

# Spectral Decomposition: Covariance Matrix

## Definition

A real, symmetric $m \times m$ matrix $\mathbf{B}$ has a spectral decomposition given by

$$\mathbf{B} \;=\; \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}'$$

where

- $\boldsymbol{\Lambda}$ is a diagonal matrix with eigenvalues $\boldsymbol{\lambda}$ on the diagonal that are all real and positive
- $\mathbf{P}$ is an $m \times m$ **orthogonal** matrix consisting of the $m$ eigenvectors

- any **orthogonal** matrix $C$ satisfies $C'C = I$ and $C' = C^{-1}$

# Eigenvalues and -vectors

- let $\mathbf{e}_i$ denote the $i$-th column of $\mathbf{P}$. This is the eigenvector associated with the $i$-th eigenvalue $\lambda_i$.
- each eigenvector/eigenvalue pair are solutions to the equation

$$\mathbf{B}\mathbf{e}_i = \lambda_i \mathbf{e}_i$$

- in general, the $m$ eigenvalues $\boldsymbol{\lambda}$ may be real or complex numbers.
- if $\lambda_i$ is a complex eigenvalue, then its associated eigenvector $\mathbf{e}_i$ is also complex, but for a covariance matrix eigenvalues are always real.
- each eigenvector $\mathbf{e}_i$ is not unique up to scale, i.e. $\gamma\mathbf{e}_i$ is also a eigenvector where $\gamma$ is any real number.
- consequently, the eigenvectors are typically normalized to have unit length: $\mathbf{e}_i'\mathbf{e}_i = 1$.

# Principal Components Result

Let $(\lambda_i, \mathbf{e}_i)$ be the i-th (eigenvalue, eigenvector) pair for the covariance matrix $\mathbf{\Sigma}_r$. The i-th principal component is given by:

$$y_{it} = \mathbf{e}_i' \mathbf{r}_t = \sum_{j=1}^{N} e_{ij} r_{jt}$$

The variance of the *i*-th principal component is given by:

$$\mathbf{e}_i' \mathbf{\Sigma}_r \mathbf{e}_i = \lambda_i, \qquad i = 1, 2, \ldots, N$$

because $\mathbf{e}_i' \mathbf{e}_i = 1$ and $\mathbf{e}_i' \mathbf{e}_j = 0$

- choose the vectors $\mathbf{w}_i = \mathbf{e}_i$

# Variance Accounting

the total variance accounted for by the $i$-th principal component is

$$\frac{V(y_{it})}{\sum_{i=1}^{N} V(y_{it})} = \frac{\lambda_i}{\lambda_1 + \ldots + \lambda_N}$$

where $\sum_{i=1}^{N} V(y_{it}) = \lambda_1 + \ldots + \lambda_N$

- Typically, we order eigenvalues and -vectors according to the magnitude of the $\lambda_i$

# Principal Components (R and MatLab)

- **princomp()** is the R function that corresponds exactly to the Spectral Decomposition we have discussed here.
- **prcomp()** can also be used. It relies on the *Singular Value Decomposition*, which is numerically more stable. For the typical covariance matrix (symmetric positive definite) this shouldn't matter, but R on a general basis recommends using prcomp().
- Implementation in R will be further discussed in the TA session
- In MatLab, you can use the **eig()** function

# Statistical Factor Model

## Definition

Consider a $N$-dimensional time series of returns $\mathbf{r}_t = (r_{1t}, r_{2t}, \ldots, r_{Nt})$. and assume the series is covariance stationary. For a covariance stationary time series we can define its mean

$$E[\mathbf{r}_t] = \boldsymbol{\mu}, \qquad \boldsymbol{\Sigma}_r = E\left[(\mathbf{r}_t - \boldsymbol{\mu})(\mathbf{r}_t - \boldsymbol{\mu})'\right]$$

where $\boldsymbol{\Sigma}_r$ is a $N \times N$ matrix. The statistical factor model postulates that the return is linearly dependent on a few unobservable random variables:

$$\mathbf{r}_t \; = \; \boldsymbol{\mu} + \boldsymbol{\beta}\mathbf{f}_t + \boldsymbol{\varepsilon}_t$$

where $\mathbf{f}_t = (f_{1t}, f_{2t}, \ldots, f_{Kt})$ with $K < N$, and $\boldsymbol{\beta}$ is a $N \times K$ matrix of factor loadings

- The factors can be found using PCA
- Stock $i$'s beta with respect to factor $j$ is the $i$'th element of eigenvector $j$
  - See PCA_Example.xls
  - This depends on normalization of eigenvectors. In general, betas are proportional to this element.

# Statistical Factor Model: Identification

- the covariance matrix of returns is

$$\boldsymbol{\Sigma}_r = \boldsymbol{\beta}\boldsymbol{\Sigma}_f\boldsymbol{\beta}' + \boldsymbol{\Sigma}_\varepsilon$$

- assume $\boldsymbol{\Sigma}_\varepsilon$ is diagonal.

- the factors $\mathbf{f}_t$ are latent and not directly observable. We estimate them from the data $\mathbf{r}_t = (r_{1t}, r_{2t}, \ldots, r_{Nt})$.

- we cannot identify both $\boldsymbol{\Sigma}_f$ and $\boldsymbol{\beta}$ from the data. Why?
- note that for any $K \times K$ orthogonal matrix $\mathbf{P}$

$$\mathbf{r}_t - \boldsymbol{\mu} = \boldsymbol{\beta}\mathbf{f}_t + \boldsymbol{\varepsilon}_t = \boldsymbol{\beta}^\star\mathbf{f}_t^\star + \boldsymbol{\varepsilon}_t$$

  where $\boldsymbol{\beta}^\star = \boldsymbol{\beta}\mathbf{P}$ and $\mathbf{f}_t^\star = \mathbf{P}'\mathbf{f}_t$. We can't tell if it is $\mathbf{f}_t$ or $\mathbf{f}_t^\star$.
- Thus: PCA-derived factors do not by themselves have economic meaning
  - However, we may derive some meaning by investigating the eigenvectors
  - more on this in a couple of slides (yield example)

# Applications of PCA

There are many applications of PCA in economics and finance

1. yield curves: Litterman and Scheinkman (1991), Cochrane and Piazzesi (2005)
2. macroeconomic indicators: Stock and Watson (1999), Stock and Watson (2002)
3. currency returns: Lustig, Roussanov, and Verdelhan (2011)

# Factor Models and the Yield Curve

There are lots of different yields and interest rates.

It would be helpful if we could summarize the variation bond yields using a couple of factors.

Litterman and Scheinkman (1991) and Knez, Litterman and Scheinkman (1994) find, using principal component analysis, that there is a level, slope and curvature factor in bond yields and bond returns.

# Bond Notation

- an $n$-period zero coupon bond pays one dollar $n$ periods from now
- notation:
  - $P_t^{(n)}$ denotes the price of an $n$-period zero-coupon bond
  - $p_t^{(n)}$ denotes the log price of an $n$-period zero-coupon bond

$$p_t^{(n)} = \log\left(P_t^{(n)}\right)$$

  - the log yield of an $n$-period zero-coupon bond is:

$$y_t^{(n)} \;=\; -\frac{1}{n} p_t^{(n)}$$

  - the log holding period return:

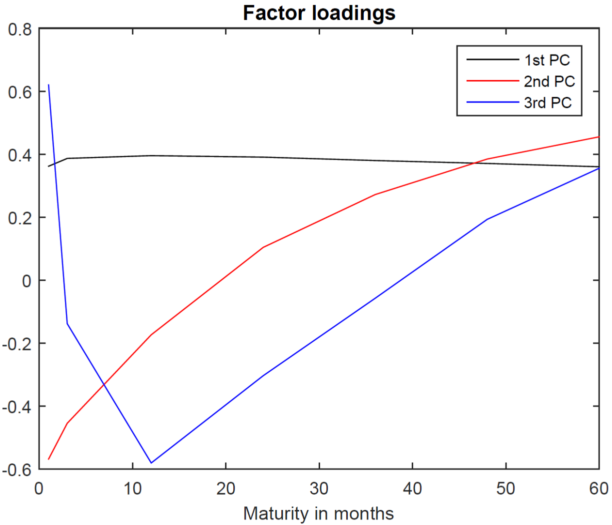
$$hpr_{t+1}^{(n)} = p_{t+1}^{(n-1)} - p_t^{(n)}$$

# Principal Components of Yields

| *Maturity* | Principal component | | |
|---|---|---|---|
| 1 | 0.362 | -0.569 | 0.620 |
| 3 | 0.387 | -0.455 | -0.139 |
| 12 | 0.395 | -0.184 | -0.581 |
| 24 | 0.391 | 0.104 | -0.304 |
| 36 | 0.380 | 0.272 | -0.058 |
| 48 | 0.371 | 0.385 | 0.193 |
| 60 | 0.360 | 0.455 | 0.355 |
| % *Var*. | 98.01 | 1.71 | 0.19 |

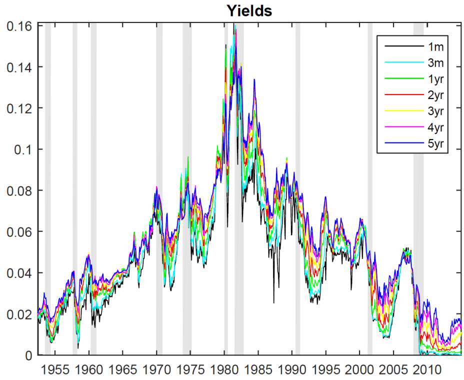
The sample period is 6/1952-12/2014. Zero coupon bonds with maturity in months. CRSP Fama-Bliss Data.

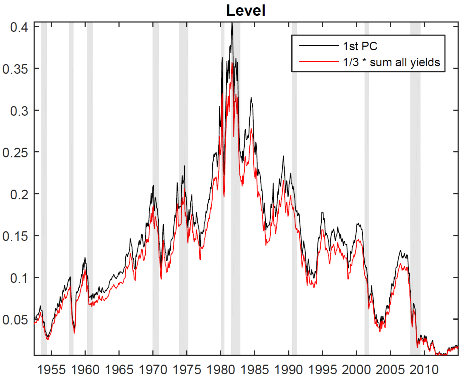
# Factor Loadings




The sample period is 6/1952-12/2014. CRSP Fama-Bliss Data.

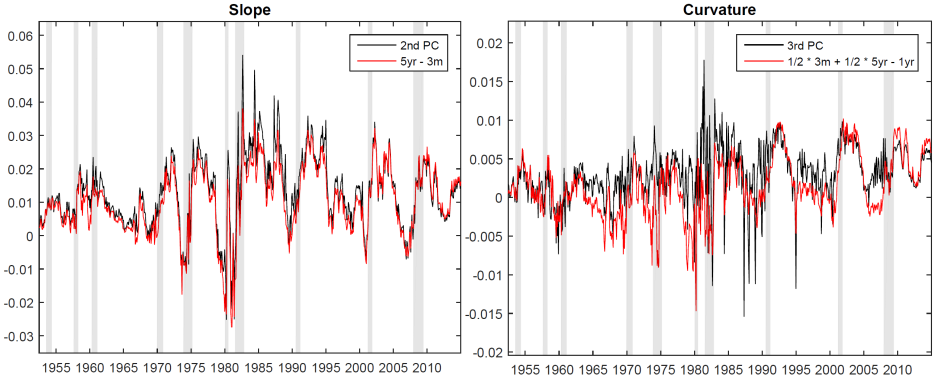
# Factors in the Yield Curve




The sample period is 6/1952-12/2014. Yields (left) and level factor (right). CRSP Fama-Bliss Data.

# Factors in the Yield Curve (cont'd)

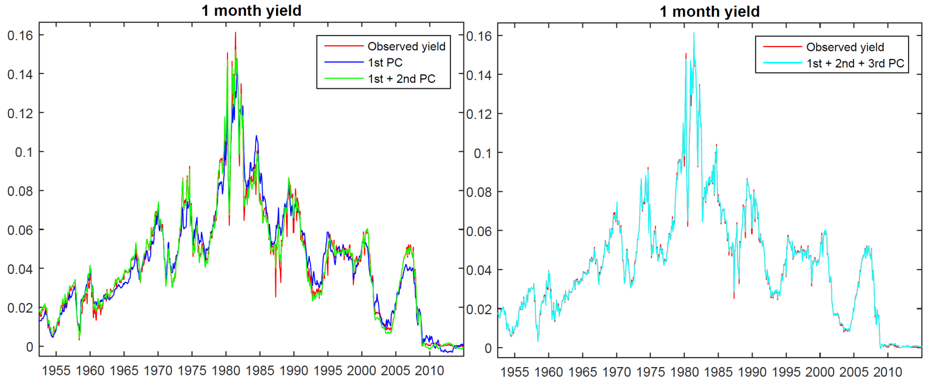



The sample period is 6/1952-12/2014. Yields (left) and level factor (right). CRSP Fama-Bliss Data.

# Replicating Yields: 1 month

Replicate log yield as linear in first 3 PCs:

$$\hat{y}_t^{(1)} = a_1 + \beta_{1,1} PC_{1,t} + \beta_{1,2} PC_{2,t} + \beta_{1,3} PC_{3,t}$$

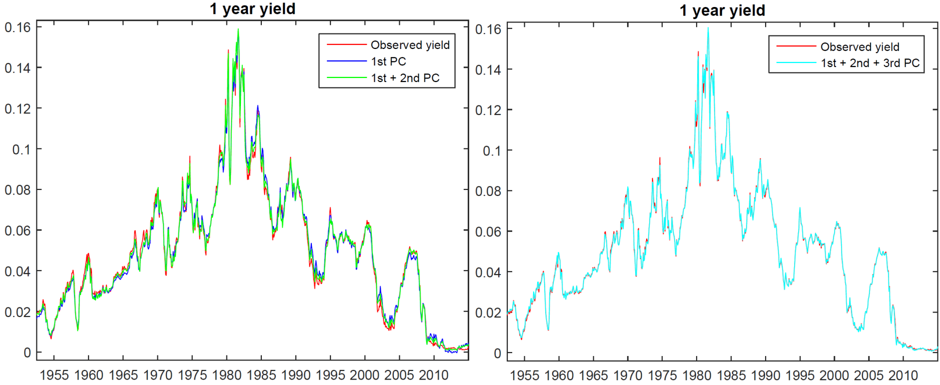



The sample period is 6/1952-12/2014. CRSP Fama-Bliss Data.

# Replicating Yields: 1 year

Replicate log yield as linear in first 3 PCs:

$$\hat{y}_t^{(12)} = a_{12} + \beta_{12,1} PC_{1,t} + \beta_{12,2} PC_{2,t} + \beta_{12,3} PC_{3,t}$$

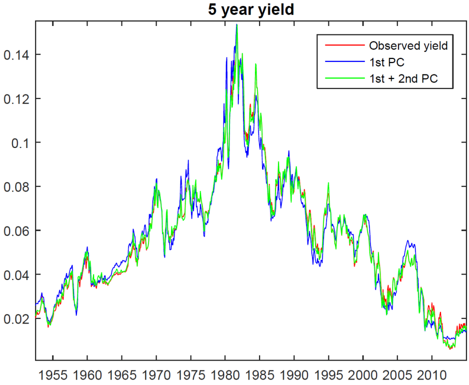


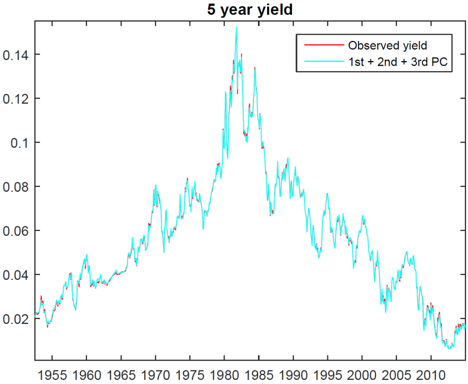

The sample period is 6/1952-12/2014. CRSP Fama-Bliss Data.

# Replicating Yields: 5 years

Replicate log yield as linear in first 3 PCs:

$$\hat{y}_t^{(60)} = a_{60} + \beta_{60,1} PC_{1,t} + \beta_{60,2} PC_{2,t} + \beta_{60,3} PC_{3,t}$$




The sample period is 6/1952-12/2014. CRSP Fama-Bliss Data.

# PCs in the cross-section of stock returns

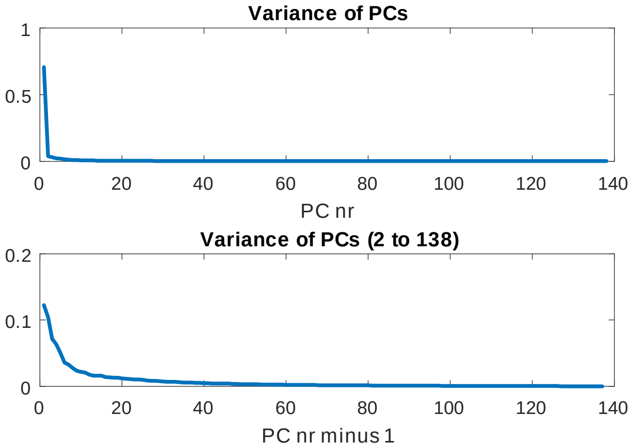
Monthly data from July 1969 through December 2016 on 138 portfolios from Kenneth French's webpage

- Industry and characteristic sorted portfolios (nearly all)

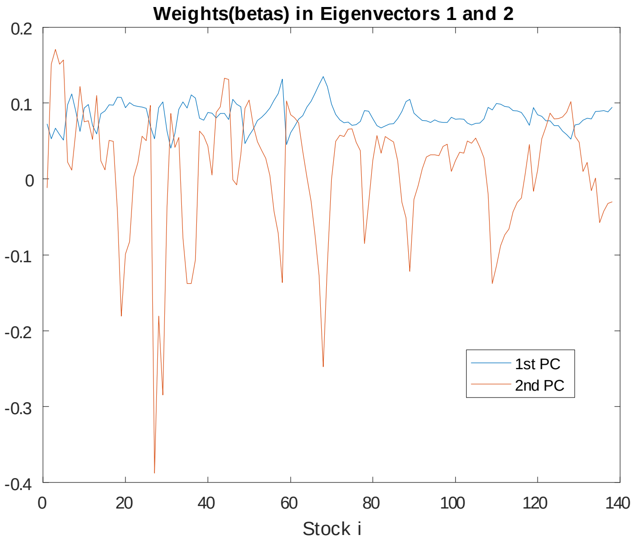
Next slides show:

1. The first PC is by far the most important and it has correlation 0.985 with the market
2. The next PCs decline in importance (variance) quite slowly.
3. The first 20 PCs accounts for 91% of the variation in the return of these portfolios (the first 10 accounts for 86%)

# PCs in the cross-section of stock returns



# PCs in the cross-section of stock returns



# References Lecture Note 10 and 11


- Black, F. and R. Litterman (1992). Global portfolio optimization. Financial Analysts Journal 48(5), pp. 28-43.
- Blume, M. E. (1975). Betas and their regression tendencies. The Journal of Finance 30(3), 115-143.
- Campbell, J. Y., A. W. Lo, and A. C. MacKinlay (1997). The Econometrics of Financial Markets. Princeton, NJ: Princeton University Press.
- Chen, N.-F., R. Roll, and S. A. Ross (1986). Economic forces and the stock market. The Journal of Business 59(3), pp. 383-403.
- Cochrane, J. H. and M. Piazzesi (2005). Bond risk premia. American Economic Review 95(1), 138-160.
- DeMiguel, V., L. Garlappi, and R. Uppal (2009). Optimal versus naive diversification: How ineffcient is the 1/N portfolio strategy? The Review of Financial Studies 22(5), 1915-1953.
- Fama, E. F. and K. French (1992). The cross-section of expected stock returns. The Journal of Finance 47(2), 427-465.

# References (cont'd)


- Fama, E. F. and K. French (1993). Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33(1), 23-49.
- Hastie, T., R. Tibshirani, and J. Friedman (2009). The Elements of Statistical Learning (Second ed.). New York, NY: Springer Press.
- Knez, P. J., R. Litterman, and J. Scheinkman (1994). Explorations into factors explaining money market returns. The Journal of Finance 49(5), pp. 1861-1882.
- Lintner, J. (1965). The valuation of risky assets and the selection of risky investments in stock portfolios and capital budgets. The Review of Economics and Statistics 47(1), 13-37.
- Litterman, R. and J. Scheinkman (1991). Common factors affecting bond returns. The Journal of Fixed Income 1(1), 54-61.
- Lustig, H., N. Roussanov, and A. Verdelhan (2011). Common risk factors in currency markets. The Review of Financial Studies 24(11), 3731-3777.
- Markowitz, H. M. (1952). Portfolio selection. The Journal of Finance 7(1), 77-91.

# References (cont'd)


- Sharpe, W. F. (1964). Capital asset prices: a theory of market equilibrium under conditions of risk. The Journal of Finance 19(3), 425-442.
- Stock, J. H. and M. W. Watson (1999). Forecasting inflation. Journal of Monetary Economics 44(2), 293-335.
- Stock, J. H. and M. W. Watson (2002). Forecasting using principal components from a large number of predictors. Journal of the American Statistical Association 97, 1167-79.
- Vasicek, O. A. (1973). A note on using cross-sectional information in Bayesian estimation of security betas. The Journal of Finance 28(5), 1233-1239.