

*Discussion of:*  
Shrinking the Cross Section

Serhiy Kozak, Stefan Nagel and Shrihari Santosh

Kent Daniel<sup>†</sup>

<sup>†</sup>Columbia Business School & NBER

New Methods for the Cross Section of Returns Conference  
September 29, 2018



# Outline

- Much of this discussion is stolen from either directly from other discussions:
  - John Campbell's discussion of Haddad, Kozak, and Santosh (2018) at the 2018 NBER-SI
  - Lars Hansen's discussion of this paper at the 2018 RedRock conference.
- ...or from the insights of coauthors and colleagues.

# Outline

- We have too many characteristics that forecast returns.
  - however, these characteristics are correlated, and not all of them work out-of-sample.
- challenge is dimensionality reduction
  - Cochrane (2011): “factor zoo” or “the multidimensional challenge”.
- What we are interested in is:
  - what characteristics (or combination of characteristics) forecast  $\mathbb{E}[r]$ 's?
  - how are these associated with the covariance structure?
  - or, equivalently, how can we map these characteristics into a reasonable SDF?
- Standard methods lead to overfitting (e.g., linear regression).
- Idea here is to do:
  - ① bayesian approach with a new prior specification.
  - ② estimation using PCA.
- Resulting portfolio, based on the first few PC's, works well out-of-sample

# Background—Factor Structures and PCA

- Ross (1976): in a large economy, only exposures to common factors can be priced.
  - If this were not true, an asymptotic arbitrage opportunity would arise.
  - Vanishingly rare exceptions are allowed (average  $\alpha^2 \rightarrow 0$  as  $N \rightarrow \infty$ ).
- In a finite economy, the corresponding empirical insight is that a characteristic that predicts average return must also predict comovement of stocks with that characteristic.
  - If this were not true, an investor could form a characteristic-sorted portfolio, hedge out the factor risk, and earn near-arbitrage profits (ie,  $SR_p \rightarrow \infty$ )
- How to find the common factors?
  - If there are  $K$  common factors then, asymptotically, the first  $K$  principal components will span the space of the common factors (Chamberlain and Rothschild, 1983)
  - Note that this *doesn't tell us anything about which of the  $K$  will earn the highest SR.*

# PCA

- Several papers in the 1980s used statistical methods such as PC analysis to extract common factors from the covariance matrix of *individual firm* stock returns.
  - Roll and Ross (1980), Connor and Korajczyk (1986, 1988), Lehmann and Modest (1988)
- The approach fell out of favor because the factors:
  - ① did not appear to be low-order, or stable
  - ② other than the first PC, they didn't command a premium.
    - The first PC was always close to a scaled EW Mkt portfolio: ( $\mathbf{q}_1 \approx \mathbf{1}/\sqrt{N}$ )
    - $\Rightarrow \mathbf{q}'_k \mathbf{1} = 0$  for  $k > 1$  (i.e., other portfolios are zero-investment)
  - ③ they were hard to interpret.
  - ④ concerns about the effects of asset repackaging (Bray, 1994)
- KNS are getting better results because they start with characteristic-sorted long-short portfolios, not individual stocks.

# Structure

- KNS start with  $N$  excess returns  $\mathbf{R}_t$
- They then form  $H(= 50, 80 \text{ or } 1, 375)$  \$1 long-\$1 short, characteristic sorted portfolios.
  - Given the  $N \times H$  matrix of scaled, demeaned characteristic ranks  $\mathbf{Z}_{t-1}$ , the  $H \times 1$  vector of returns on the “factor” portfolios is:

$$\mathbf{F}_t = \mathbf{Z}_{t-1} \mathbf{R}_t$$

- Given an estimated (symmetric, positive definite)  $H \times H$  covariance matrix  $\Sigma$  (assumed constant):

$$\Sigma = \mathbf{Q} \mathbf{D} \mathbf{Q}'$$

where

- $\mathbf{D}$  is an ordered diagonal matrix of eigenvalues.
- The  $H$  eigenvectors (the columns of  $\mathbf{Q}$ ) form an orthonormal basis for the  $H$ -dimensional return space.
- The “Principal component” portfolio returns  $\mathbf{P}_t$  are given by:

$$\mathbf{P}_t = \mathbf{Q}' \mathbf{F}_t$$

# PCA

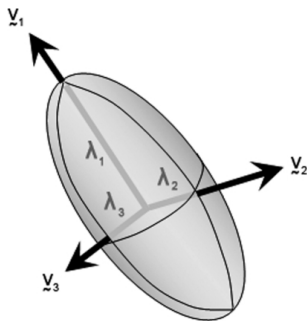
- The first PC is the portfolio with weight vector  $\mathbf{q}_1$  given by:

$$\mathbf{q}_1 = \arg \max_{\mathbf{q}} \mathbf{q}' \Sigma \mathbf{q} \quad s.t. \quad \mathbf{q}' \mathbf{q} = 1$$

- where  $\mathbf{q}_1' \Sigma \mathbf{q}_1 = \mathbf{q}_1' \mathbf{Q} \mathbf{D} \mathbf{Q}' \mathbf{q}_1 = d_1$

- The second PC ( $\mathbf{q}_2$ ) solves the same problem in the  $H-1$  dimensional manifold  $\perp \mathbf{q}_1$

- *NB:* – in the picture on the right,  $\mathbf{q}_i \rightarrow \mathbf{v}_i$  and  $d_i \rightarrow \lambda_i$



# Asset Repackaging

- Suppose you have two LS portfolios:  $R_1 = f_1$  and  $R_2 = f_2$ 
  - Suppose  $f_1 \perp f_2$  and  $\sigma_1 > \sigma_2$
  - $\Rightarrow Q = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$  and  $D = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$
- How can you change the ordering of the PC's?
  - 1 Add enough idiosyncratic noise to  $R_2$  enough so that its volatility is  $> \sigma_1$ .
    - This will decrease its SR, but make it the first PC.
  - 2 Add assets: Suppose you in place of  $R_2$ , you have N assets 2-(N+1) each of which has return

$$R_i = f_2 + \epsilon_i, \quad \text{where } \sigma_\epsilon^2 \text{ is small \& } \epsilon_i \perp \epsilon_j$$

the  $f_2$  eigenvector will have weights  $1/\sqrt{N}$  on each of these N assets, and corresponding eigenvalue  $N\sigma_2^2$ .



# Asset Repackaging

- Suppose you have two LS portfolios:  $R_1 = f_1$  and  $R_2 = f_2$ 
  - Suppose  $f_1 \perp f_2$  and  $\sigma_1 > \sigma_2$
  - $\Rightarrow Q = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$  and  $D = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$
- How can you change the ordering of the PC's?
  - 1 Add enough idiosyncratic noise to  $R_2$  enough so that its volatility is  $> \sigma_1$ .
    - This will decrease its SR, but make it the first PC.
  - 2 Add assets: Suppose you in place of  $R_2$ , you have N assets 2-(N+1) each of which has return

$$R_i = f_2 + \epsilon_i, \quad \text{where } \sigma_\epsilon^2 \text{ is small \& } \epsilon_i \perp \epsilon_j$$

the  $f_2$  eigenvector will have weights  $1/\sqrt{N}$  on each of these N assets, and corresponding eigenvalue  $N\sigma_2^2$ .

# Asset Repackaging

- Suppose you have two LS portfolios:  $R_1 = f_1$  and  $R_2 = f_2$ 
  - Suppose  $f_1 \perp f_2$  and  $\sigma_1 > \sigma_2$
  - $\Rightarrow Q = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$  and  $D = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$
- How can you change the ordering of the PC's?
  - 1 Add enough idiosyncratic noise to  $R_2$  enough so that its volatility is  $> \sigma_1$ .
    - This will decrease its SR, but make it the first PC.
  - 2 Add assets: Suppose you in place of  $R_2$ , you have N assets 2-(N+1) each of which has return

$$R_i = f_2 + \epsilon_i, \quad \text{where } \sigma_\epsilon^2 \text{ is small \& } \epsilon_i \perp \epsilon_j$$

the  $f_2$  eigenvector will have weights  $1/\sqrt{N}$  on each of these N assets, and corresponding eigenvalue  $N\sigma_2^2$ .

# Asset Repackaging

- Suppose you have two LS portfolios:  $R_1 = f_1$  and  $R_2 = f_2$ 
  - Suppose  $f_1 \perp f_2$  and  $\sigma_1 > \sigma_2$
  - $\Rightarrow Q = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$  and  $D = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$
- How can you change the ordering of the PC's?
  - ① Add enough idiosyncratic noise to  $R_2$  enough so that its volatility is  $> \sigma_1$ .
    - This will decrease its SR, but make it the first PC.
  - ② Add assets: Suppose you in place of  $R_2$ , you have N assets 2-(N+1) each of which has return

$$R_i = f_2 + \epsilon_i, \quad \text{where } \sigma_\epsilon^2 \text{ is small \& } \epsilon_i \perp \epsilon_j$$

the  $f_2$  eigenvector will have weights  $1/\sqrt{N}$  on each of these N assets, and corresponding eigenvalue  $N\sigma_2^2$ .

# Cross-Sectional Risk and Priced Risk

- The PC structure tells you about the structure of cross-sectional risk in the stock market.
  - Economically, it makes sense that the biggest risk might be the market.
- However, at least for individual stocks, after that first PC there is no reason to believe that the way that risks are packaged in the stock market should be related to their SR (or equivalently, their correlation with  $\tilde{m}$ ).
  - Repackaging will alter the cross-section of risk in an arbitrary way, without affecting  $m^*$  (the MVE portfolio).
- Further, there may very well be priced risks that can't be hedged (if  $m^* \neq m$ )
- The *only* thing that will determine whether the first few PCs priced is the way the test assets are packaged.

# Priors

KNS consider priors of the form:

$$\begin{aligned}\boldsymbol{\mu} &\sim \mathcal{N}\left(\mathbf{0}, \frac{\kappa^2}{\tau} \mathbf{D}^\eta\right) \\ \mathbf{D}^{1/2} \boldsymbol{\mu} &\sim \mathcal{N}\left(\mathbf{0}, \frac{\kappa^2}{\tau} \mathbf{D}^{\eta-1}\right)\end{aligned}$$

- What  $\eta$  is most reasonable here?
- Other shrinkage-estimators (Pástor, 2000; Pástor and Stambaugh, 2000, e.g.,) use  $\eta = 1$ 
  - As KNS point out, this is equivalent to assuming that the distribution of SRs is independent of the risk of the portfolio.
- KNS choose  $\eta = 2$ .
  - The logic is very cool, but I am not sure that I buy it!

# Prior Variances

$$D^{1/2}\boldsymbol{\mu} \sim \mathcal{N}\left(\mathbf{0}, \frac{\kappa^2}{\tau} D^{\eta-1}\right)$$

- KNS argue that:

$$\mathbb{E}[\mathbf{b}'\mathbf{b}] = \frac{\kappa^2}{\tau} \sum_{i=1}^H d_i^{\eta-2}$$

- they state that:

*... with  $\eta < 2$  the prior would imply that the  $i$  optimal portfolio of a rational investor is likely to place huge bets on the lowest-eigenvalue PCs. Setting  $\eta \geq 2$  avoids such unrealistic portfolio weights. (p. 15)*

- Indeed, this seems to me a very good reason for why we should expect higher SRs for portfolios that require extreme weights.
  - Seeing an a very high SR for a portfolio that requires very high leverage seems consistent with what we know about market frictions and limits to arbitrage

## Prior Means = 0?

$$D^{1/2}\boldsymbol{\mu} \sim \mathcal{N}\left(\mathbf{0}, \frac{\kappa^2}{\tau} D^{\eta-1}\right)$$

- Longer term, it would be nice to move away from centering the prior distribution at zero.
- We have some reasonable—rational and behavioral—theories, about which characteristics should be linked to  $\mathbb{E}[R]$ 's, and the direction of this link:
  - The PV relationship in Fama and French (2015)
  - The short- and long-horizon relationships exploited in Daniel, Hirshleifer, and Sun (2018).
- Why not put a higher mean on those characteristics that show up in such relationships?

# References I

- Bray, Margaret, 1994, *The Arbitrage Pricing Theory is not Robust 1: Variance Matrices and Portfolio Theory in Pictures* (LSE Financial Markets Group).
- Chamberlain, Gary, and Michael Rothschild, 1983, Arbitrage, factor structure, and mean-variance analysis on large asset markets, *Econometrica* 51, 1281–1304.
- Cochrane, John H, 2011, Presidential address: Discount rates, *The Journal of Finance* 66, 1047–1108.
- Connor, Gregory, and Robert A. Korajczyk, 1986, Performance measurement with the arbitrage pricing theory: A new framework for analysis, *Journal of Financial Economics* 17, 255–289.
- , 1988, Risk and return in an equilibrium APT: Application of a new test methodology, *Journal of Financial Economics* 21, 255–289.
- Daniel, Kent, David Hirshleifer, and Lin Sun, 2018, Short- and long-horizon behavioral factors, Columbia Business School working paper.
- Fama, Eugene F., and Kenneth R. French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.
- Lehmann, Bruce, and David Modest, 1988, Empirical foundations of the arbitrage pricing theory, *Journal of Financial Economics* 21, 213–254.
- Pástor, Ľuboš, 2000, Portfolio selection and asset pricing models, *The Journal of Finance* 55, 179–223.
- , and Robert F Stambaugh, 2000, Comparing asset pricing models: an investment perspective, *Journal of Financial Economics* 56, 335–381.
- Roll, Richard W., and Steven A. Ross, 1980, An empirical investigation of the arbitrage pricing theory, *Journal of Finance* 35, 1073–1103.
- Ross, Stephen A., 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory* 13, 341–360.