SHRINKING THE CROSS-SECTION

Serhiy Kozak\textsuperscript{1}, Stefan Nagel\textsuperscript{2}, Shri Santosh\textsuperscript{3}

September 26, 2018

\textsuperscript{1}University of Michigan
\textsuperscript{2}University of Chicago, NBER & CEPR
\textsuperscript{3}University of Maryland
Cross-sectional asset pricing: general framework

- The goal of empirical asset pricing – find an empirical SDF
- Project the “true” SDF onto the space of excess returns,
  \[ M_t = 1 - b_{t-1}' (R_t - \mathbb{E}[R_t]) , \]
- Parameterize \( b_{t-1} \) as linear in “derived” characteristics \( \phi(Z_{t-1}) \):
  \[ b_{t-1} = \phi(Z_{t-1}) b , \]
  - i.e., pricing relevant information (\( \mathbb{E}[R], \text{var}, \text{covar} \)) about a stock captured by characteristics
- Plug in:
  \[ M_t = 1 - b' (F_t - \mathbb{E}[F_t]) , \]
  - where \( F_t = \phi(Z_{t-1})' R_t \) is a vector of characteristics-based factors
Factor models and the cross-section of expected stock returns

• Much research looking for characteristics-sparse SDF: \( F \) contains factors based on small number (ad-hoc) stock characteristics

• Example: Fama and French (1993)

\[ M = 1 - b_1 F_M - b_2 F_{SMB} - b_3 F_{HML} \]

• More recently proposed reduced-form factor SDFs
  • Hou et al. (2015): 4 factors
  • Fama and French (2016): 5 factors
  • Barillas and Shanken (2017): 6 factors
The high-dimensionality challenge

- Does such a characteristics-sparse SDF really suffice to describe cross-section of expected stock returns?

- Statistically daunting task: High-dimensional problem!
  - How to take into account, jointly, all anomalies that have been discovered so far?
  - Plus potentially interactions and nonlinearities?

- **Our paper**: Use machine-learning inspired tools designed for high-dimensional setting
  - Bayesian foundation to bring in economically motivated prior beliefs
The high-dimensionality challenge

- Consider large number of anomaly factors $F$: test assets and potential SDF factors at the same time
- Why not just throw in hundreds of factor portfolio returns into vector $F$ and estimate price-of-risk coefficients $b$ in

$$M = 1 - b'(F - \mathbb{E}[F])$$

- Naive approach:
  - Using $\mathbb{E}[MF] = 0$, solve for $b = \Sigma^{-1} \mu$, where $\mu = \mathbb{E}[F]$
  - Estimate with sample equivalent

$$\hat{b} = \Sigma^{-1} \frac{1}{T} \sum_{t=1}^{T} F_t,$$

- Naive approach would result in extreme overfitting of noise \Rightarrow Terrible out-of-sample performance.
  - Main problem: Means $\mu$ estimated imprecisely
Economically motivated priors

- Prior (known $\Sigma$):
  \[
  \mu \sim \mathcal{N}(0, \kappa \Sigma^2)
  \]
  i.e., economic restriction that links first ($\mu$) and second moments ($\Sigma$) of factors
  - $\kappa$ controls strength of the prior
  - For intuition, consider diagonal $\Sigma$. Then,
    \[
    \frac{\mu_i}{\sigma_i} \sim \mathcal{N}(0, \kappa \sigma_i^2)
    \]

- Implies that only high-variance PCs are likely to earn high Sharpe Ratios
  - True in rational asset pricing models with “macro” factors
  - True in behavioral model with sentiment-driven demand and arbitrageurs (Kozak, Nagel, Santosh 2017a)
SDF coefficient estimator

- Combining prior with sample data on mean factor returns $\bar{\mu}$ we get the posterior mean of the SDF coefficients $b$:
  \[
  \hat{b} = (\Sigma + \gamma I_K)^{-1} \bar{\mu} \tag{1}
  \]

- *Interpretation 1*: C-S regression of $\bar{\mu}$ on covariances of returns and factors, with a penalty on the maximum squared SR $b' \Sigma b$:
  \[
  \hat{b} = \arg \min_b \left\{ (\bar{\mu} - \Sigma b)' (\bar{\mu} - \Sigma b) + \gamma b' \Sigma b \right\}
  \]

- *Interpretation 2*: HJ distance s.t. an $L^2$-norm penalty on $b' b$:
  \[
  \hat{b} = \arg \min_b \left\{ (\bar{\mu} - \Sigma b)' \Sigma^{-1} (\bar{\mu} - \Sigma b) + \gamma b' b \right\}
  \]
  
  - Penalty term effectively down-weights contributions of low-variance PCs to the overall max. SR
Allowing for sparsity

- Useful to have a prior that reflects possibility that many elements of the $b$ vector may be zero:
  - Laplace prior $\Rightarrow L^1$ norm penalty
- Dual-penalty specification

\[
\hat{b} = \arg \min_b (\bar{\mu} - \Sigma b)' \Sigma^{-1} (\bar{\mu} - \Sigma b) + \gamma_2 b' b + \gamma_1 \sum_{i=1}^{H} |b_i|
\]

- Similar to elastic net, but not quite
- Penalty parameters $\gamma_1, \gamma_2$ chosen to maximize out-of-sample performance
Empirical application

• Initial sanity check: Fama-French 25 size-B/M sorted portfolios
  • If method works, it should recover SDF with SMB and HML factors
• Main analysis: Stock characteristics portfolios
  • 50 anomaly characteristics portfolios
  • 80 WRDS financial ratios & past return portfolios
• Questions:
  • Can we find an SDF sparse in characteristics?
  • Can we find an SDF sparse in PCs?
Empirical evaluation: 3-fold cross-validation

- Sample split in 3 blocks
- 2 blocks used for estimation of $b$ given values for the penalties $\gamma_1, \gamma_2$
- Remaining block used for out-of-sample evaluation: $R^2$ in explaining average returns

$$R^2 = 1 - \frac{(\bar{\mu}_2 - \bar{\Sigma}_2 \hat{b})' (\bar{\mu}_2 - \bar{\Sigma}_2 \hat{b})}{\bar{\mu}_2' \bar{\mu}_2},$$

- Choose different blocks for estimation and evaluation and repeat
- Record average OOS $R^2$ from all three evaluation samples
- Look for $\gamma_1, \gamma_2$ that maximize average OOS $R^2$

Lewellen et al. (2010): FF25 ex market return \(\approx\) linear combinations of SMB and HML factor \(\Rightarrow\) any selection of 2-3 out of 25 portfolios sufficient to span SDF, i.e., sparsity

Kozak et al. (2017): SMB and HML \(\approx\) PC1 and PC2 of FF25 ex market return \(\Rightarrow\) 2 PCs should be sufficient to span SDF
FF25: OOS $R^2$ from dual-penalty specification

Figure 1: OOS $R^2$ from dual-penalty specification (Fama-French 25 ME/BM portfolios).

(a) Raw Fama-French 25 portfolios

(b) PCs of Fama-French 25 portfolios
FF25: $L^2$ Model Selection and Sparsity

Figure 2: $L^2$ Model Selection and Sparsity (Fama-French 25 ME/BM portfolios).
## List of 50 Anomaly Characteristics

<table>
<thead>
<tr>
<th>Size</th>
<th>Investment</th>
<th>Long-term Reversals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>Inv/Cap</td>
<td>Value (M)</td>
</tr>
<tr>
<td>Profitability</td>
<td>Investment Growth</td>
<td>Net Issuance (M)</td>
</tr>
<tr>
<td>Value-Profitability</td>
<td>Sales Growth</td>
<td>SUE</td>
</tr>
<tr>
<td>F-score</td>
<td>Leverage</td>
<td>Return on Book Equity</td>
</tr>
<tr>
<td>Debt Issuance</td>
<td>Return on Assets (A)</td>
<td>Return on Market Equity</td>
</tr>
<tr>
<td>Share Repurchases</td>
<td>Return on Equity (A)</td>
<td>Return on Assets</td>
</tr>
<tr>
<td>Net Issuance (A)</td>
<td>Sales/Price</td>
<td>Short-term Reversals</td>
</tr>
<tr>
<td>Accruals</td>
<td>Growth in LTNOA</td>
<td>Idiosyncratic Volatility</td>
</tr>
<tr>
<td>Asset Growth</td>
<td>Momentum (6m)</td>
<td>Beta Arbitrage</td>
</tr>
<tr>
<td>Asset Turnover</td>
<td>Industry Momentum</td>
<td>Seasonality</td>
</tr>
<tr>
<td>Gross Margins</td>
<td>Value-Momentum</td>
<td>Industry Rel. Reversals</td>
</tr>
<tr>
<td>D/P</td>
<td>Value-Prof-Momentum</td>
<td>Industry. Rel. Rev. (LV)</td>
</tr>
<tr>
<td>E/P</td>
<td>Short Interest</td>
<td>Industry Momentum-Rev</td>
</tr>
<tr>
<td>CF/P</td>
<td>Momentum (12m)</td>
<td>Composite Issuance</td>
</tr>
<tr>
<td>Net Operating Assets</td>
<td>Momentum-Reversals</td>
<td>Stock Price</td>
</tr>
</tbody>
</table>
50 anomalies: OOS $R^2$ from dual-penalty specification

**Figure 3:** Sparse Model Selection (50 anomaly portfolios).
50 anomalies: $L^2$ model selection and Sparsity

(a) $L^2$ model selection

(b) Sparsity

Figure 4: $L^2$ Model Selection and Sparsity (50 anomaly portfolios).
## 50 anomalies: Coefficient estimates and $t$-statistics

<table>
<thead>
<tr>
<th></th>
<th>$b$</th>
<th>$t$-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Industry Rel. Rev. (L.V.)</td>
<td>-0.88</td>
<td>3.53</td>
</tr>
<tr>
<td>Ind. Mom-Reversals</td>
<td>0.48</td>
<td>1.94</td>
</tr>
<tr>
<td>Industry Rel. Reversals</td>
<td>-0.43</td>
<td>1.70</td>
</tr>
<tr>
<td>Seasonality</td>
<td>0.32</td>
<td>1.29</td>
</tr>
<tr>
<td>Earnings Surprises</td>
<td>0.32</td>
<td>1.29</td>
</tr>
<tr>
<td>Value-Profitability</td>
<td>0.30</td>
<td>1.18</td>
</tr>
<tr>
<td>Return on Market Equity</td>
<td>0.30</td>
<td>1.18</td>
</tr>
<tr>
<td>Investment/Assets</td>
<td>-0.24</td>
<td>0.95</td>
</tr>
<tr>
<td>Return on Equity</td>
<td>0.24</td>
<td>0.95</td>
</tr>
<tr>
<td>Composite Issuance</td>
<td>-0.24</td>
<td>0.95</td>
</tr>
<tr>
<td>Momentum (12m)</td>
<td>0.23</td>
<td>0.91</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$b$</th>
<th>$t$-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>PC 4</td>
<td>1.01</td>
<td>4.25</td>
</tr>
<tr>
<td>PC 1</td>
<td>-0.54</td>
<td>3.08</td>
</tr>
<tr>
<td>PC 2</td>
<td>-0.56</td>
<td>2.65</td>
</tr>
<tr>
<td>PC 9</td>
<td>-0.63</td>
<td>2.51</td>
</tr>
<tr>
<td>PC 15</td>
<td>0.32</td>
<td>1.27</td>
</tr>
<tr>
<td>PC 17</td>
<td>-0.30</td>
<td>1.18</td>
</tr>
<tr>
<td>PC 6</td>
<td>-0.29</td>
<td>1.18</td>
</tr>
<tr>
<td>PC 11</td>
<td>-0.19</td>
<td>0.74</td>
</tr>
<tr>
<td>PC 13</td>
<td>-0.17</td>
<td>0.65</td>
</tr>
<tr>
<td>PC 23</td>
<td>0.15</td>
<td>0.56</td>
</tr>
<tr>
<td>PC 7</td>
<td>0.14</td>
<td>0.56</td>
</tr>
</tbody>
</table>

**(a)** 50 anomaly portfolios

**(b)** PCs of 50 anomaly portfolios
50 anomalies: Summary

- Very little redundancy in original characteristics space: Characteristics-sparse SDF not achievable
- But PC-sparse SDF based on a few (high-variance) PCs prices well
- Result could be partly a consequence of looking at a set of data-mined anomalies ⇒ we examine alternative data sets in the paper...
Out-of-sample test with fully-withheld sample

- Constructing SDF is equivalent to finding the MVE portfolio
- Pre-2006 data yield SDF coefficient estimates $\hat{b} = \text{MVE portfolio weights}$
- Apply $\hat{b}$ to post 2005 returns: OOS MVE portfolio return
- Test alphas relative to restricted benchmarks
  - Construct MVE portfolio weights from sparse characteristics-based models (e.g., FF 5-factor model) in pre-2006 data and apply weights to post-2005 returns
  - Regress OOS MVE portfolio return on MVE portfolio return of sparse characteristics-based factors to estimate OOS abnormal return
Out-of-sample test with fully-withheld sample

**Table 1:** MVE portfolio’s annualized OOS $\alpha$ (%) in the withheld sample (2005-2017)

<table>
<thead>
<tr>
<th>SDF factors</th>
<th>Benchmark</th>
<th>CAPM</th>
<th>FF 6-factor</th>
<th>Char.-sparse</th>
<th>PC-sparse</th>
</tr>
</thead>
<tbody>
<tr>
<td>50 anomaly portfolios</td>
<td>12.35</td>
<td>8.71</td>
<td>9.55</td>
<td>4.60</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5.26)</td>
<td>(4.94)</td>
<td>(3.95)</td>
<td>(2.22)</td>
<td></td>
</tr>
<tr>
<td>80 WFR portfolios</td>
<td>20.05</td>
<td>19.77</td>
<td>17.08</td>
<td>3.63</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5.26)</td>
<td>(5.29)</td>
<td>(5.05)</td>
<td>(2.93)</td>
<td></td>
</tr>
<tr>
<td>1,375 interactions of anomalies</td>
<td>25.00</td>
<td>22.79</td>
<td>21.68</td>
<td>12.41</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5.26)</td>
<td>(5.18)</td>
<td>(5.03)</td>
<td>(3.26)</td>
<td></td>
</tr>
</tbody>
</table>
Conclusions

- Characteristics-sparse SDF elusive
  - Not much redundancy among “anomaly” characteristics
  ⇒ Debate whether we need 3, 4, 5, or 6 characteristics-based factors in SDF not helpful
- Instead, construct an SDF directly, or extract few factors that aggregate information from all characteristics
  - E.g., principal component factors
    - Risk premia earned mostly by major sources of return covariance
      ⇒ Makes economic sense both in asset pricing models with rational and in models with imperfectly rational investors
- PCs could be used as test assets and to look for correlations with macro risk factors, sentiment, intermediary risk factors, ...