Asset Pricing

Semyon Malamud

EPFL

Table of Contents

Mean-Variance Optimization: Unconditional i

- assets $i = 1, \cdots, N$ have prices $P_{i,t}$ and *excess* returns

$$R_{i,t+1} = \frac{P_{i,t+1} + D_{i,t+1}}{P_{i,t}} - \underbrace{R_{f,t}}_{risk\ free\ rate} \tag{1}$$

- if you invest fraction $\pi_{i,t}$ of your wealth W_t into security i, the rest stays on your bank account and grows at the rate $R_{f,t}$:

$$W_t = \sum_i \underbrace{\pi_{i,t} W_t}_{investment\ in\ stock\ i} + \underbrace{(W_t - \sum_i \pi_{i,t} W_t)}_{bank\ account} \tag{2}$$

Mean-Variance Optimization: Unconditional ii

and then you sell your investments at time t and collect dividends so that

$$\begin{aligned} W_{t+1} &= \sum_i W_t \pi_{i,t} \frac{P_{i,t+1} + D_{t+1}}{P_{i,t}} + (W_t - \sum_i \pi_{i,t} W_t) R_{f,t} \\ &= W_t R_{f,t} + W_t \sum_i \pi_{i,t} R_{i,t+1} \end{aligned} \tag{3}$$

- Thus, the excess return on your wealth is

$$\frac{W_{t+1}}{W_t} - R_{f,t} = \sum_i \pi_{i,t} R_{i,t+1} = \pi_t' R_{t+1} \tag{4}$$

- Thus, we want π_t that gives good returns. But what is the criterion?

Mean-Variance Optimization: Unconditional iii

- Intuitively, we like **high return** and **low variance,** hence, we might try to find a **static** portfolio that maximizes

$$\pi \;=\; \arg\max_{\pi} \left(E[\pi' R_{t+1}] \;-\; 0.5 \underbrace{\gamma}_{\textit{risk aversion}} \mathrm{Var}[\pi' R_{t+1}] \right) \tag{5}$$

- The solution is Markowitz

$$\pi \;=\; \mathrm{Var}[R]^{-1}\, E[R]\,. \tag{6}$$

- Alternatively, one could optimize

$$\pi \;=\; \arg\max_{\pi} \left(E[\pi' R_{t+1}] \;-\; 0.5 \underbrace{\gamma}_{\textit{risk aversion}} E[(\pi' R_{t+1})^2] \right) \tag{7}$$

and the solution is

$$\begin{aligned}\tilde{\pi} &= \gamma^{-1}(E[R_{t+1}R'_{t+1}])^{-1}\,E[R_{t+1}] \\ &= const\cdot\pi, \qquad const = \frac{1}{1+E[R_{t+1}]'\mathrm{Var}[R_{t+1}]^{-1}E[R_{t+1}]}\end{aligned} \tag{8}$$

where

$$E[R_{t+1}R'_{t+1}] = \mathrm{Var}[R_{t+1}] + E[R_{t+1}]E[R_{t+1}]' = (E[R_{i,t+1}R_{j,t+1}])_{i,j=1}^{N} \tag{9}$$

Why Are the Two Markowitz Portfolios Proportional? The Sherman-Morrison formula i

The magic behind is the

Lemma (Sherman-Morrison formula)

$$(A + xx')^{-1} = A^{-1} - \frac{A^{-1}xx'A^{-1}}{1 + x'A^{-1}x} \tag{10}$$

and

$$(A + xx')^{-1}x = \frac{A^{-1}x}{1 + x'A^{-1}x} \tag{11}$$

Proof[Proof of the Sherman-Morrison formula] Recall that

$$xx' = (x_i x_j)_{i,j=1}^{N}$$

Why Are the Two Markowitz Portfolios Proportional? The Sherman-Morrison formula ii

is a symmetric, positive, semi-definite, $rank - 1$ matrix (all columns are proportional to x). Then,

$$(A + xx')(A^{-1} - \frac{A^{-1}xx'A^{-1}}{1 + x'A^{-1}x})$$
$$= I - \frac{xx'A^{-1}}{1 + x'A^{-1}x} + xx'A^{-1} - xx'\frac{A^{-1}xx'A^{-1}}{1 + x'A^{-1}x} \quad (12)$$
$$= I - \frac{xx'A^{-1}}{1 + x'A^{-1}x} + xx'A^{-1} - xx'A^{-1}\frac{x'A^{-1}x}{1 + x'A^{-1}x} = I$$

and

$$(A + xx')^{-1}x = (A^{-1} - \frac{A^{-1}xx'A^{-1}}{1 + x'A^{-1}x})x = \frac{A^{-1}x}{1 + x'A^{-1}x} \quad (13)$$

(Very Big) Issues with Markowiz

- Markowitz **assumes that we know the truth! The true**

$$E[R] \;=\; (E[R_{i,t+1}])_{i=1}^{N_t},\ \mathrm{Var}[R] \;=\; (\mathrm{Cov}(R_{i,t+1}, R_{j,t+1}))_{i,j=1}^{N_t} \quad (14)$$

where N_t is the number of assets (stocks?) available at time t.

- The problem is that:
  - **expected stock returns move a lot over time**: Hence, using **static** portfolio is a **very bad idea**
  - we just **do not have enough data** to estimate $E[R]$ and $\mathrm{Var}[R]$. We can use naive

$$\bar{E}[R] \;=\; \frac{1}{T}\sum_{t=1}^{T} R_t,\ \overline{\mathrm{Var}}[R] \;=\; \frac{1}{T}\sum_{t=1}^{T} \underbrace{\underbrace{(R_t - \bar{E}[R])}_{N\times 1}\underbrace{(R_t - \bar{E}[R])'}_{1\times N}}_{N\times N}$$

Incorporating Conditional Information: The conditional expectation i

- We would like to incorporate conditional information.
- For the specific portfolio applications, we would need

$$
\begin{aligned}
E_t[R_{t+1}] &= \arg\min_{F:\mathbb{R}^P\to\mathbb{R}^N} E[\|R_{t+1} - F(S_t)\|^2] \\
E_t[R_{t+1}R'_{t+1}] &= \arg\min_{G:\mathbb{R}^P\to\mathbb{R}^{N\times N}} E[\|R_{t+1}R'_{t+1} - G(S_t)\|^2]
\end{aligned} \tag{15}
$$

- The reality is that **we still cannot compute** $E[\cdot]$ because we do not have enough data. So, we will still be doing

$$
E_t[X_{t+1}] = \arg\min_F \frac{1}{T}\sum_t |X_{t+1} - F(S_t)|^2 \tag{16}
$$

Incorporating Conditional Information: The conditional Markowitz i

- mean-variance optimization:

$$\pi_t = \arg\max_{\pi_t} \Big(E_t[\pi_t' R_{t+1}] - 0.5 \underbrace{\gamma}_{risk \ aversion} \mathrm{Var}_t[\pi_t' R_{t+1}] \Big) \quad (17)$$

and hence the **Mean-Variance Efficient (MVE) portfolio is**

$$\underbrace{\pi_t}_{conditional \ tangency \ portfolio} = \gamma^{-1} \underbrace{(\mathrm{Var}_t[R_{t+1}])^{-1}}_{N\times N \ covariance \ matrix} \underbrace{E_t[R_{t+1}]}_{N\times 1 \ expected \ returns} \quad (18)$$

Incorporating Conditional Information: The conditional Markowitz ii

- Similarly,

$$\begin{aligned}\tilde{\pi}_t &= \gamma^{-1}(E_t[R_{t+1}R'_{t+1}])^{-1}\,E_t[R_{t+1}] \\ &= \frac{1}{1+E_t[R_{t+1}]'\mathrm{Var}_t[R_{t+1}]^{-1}E_t[R_{t+1}]}\pi_t\end{aligned} \tag{19}$$

  where

$$E_t[R_{t+1}R'_{t+1}] = \mathrm{Var}_t[R_{t+1}] + E_t[R_{t+1}]E_t[R_{t+1}]' \tag{20}$$

Table of Contents

- Intuitively, we expect that

$$P_{i,t} = \underbrace{(R_{f,t})^{-1} E_t[P_{i,t+1} + D_{i,t+1}]}_{Definitely\ wrong\ in\ the\ data} \tag{21}$$

because the **discount factor** $(R_{f,t})^{-1}$ is too naive

- **We need a smart discount factor (SDF)**:

$$P_{i,t} = E_t[\underbrace{M_{t,t+1}}_{stochastic\ discount\ factor}(P_{i,t+1} + D_{i,t+1})] \tag{22}$$

- with a bit of algebra, this is equivalent to

$$E_t[R_{i,t+1} M_{t,t+1}] = 0 \tag{23}$$

- By direct calculation,

$$M_{t+1} \;=\; 1 \;-\; \tilde{\pi}_t' R_{t+1} \tag{24}$$

does the job:

$$\begin{aligned} E_t[R_{t+1} M_{t,t+1}] \;&=\; E_t[R_{t+1}\,(1 \;-\; R_{t+1}' \tilde{\pi}_t)] \\ =\; E_t[R_{t+1}] \;-\; E_t[R_{t+1} R_{t+1}']\,\tilde{\pi}_t \;&=\; 0 \end{aligned} \tag{25}$$

implies

$$\tilde{\pi}_t \;=\; E_t[R_{t+1} R_{t+1}']^{-1}\,E_t[R_{t+1}] \tag{26}$$

We now state

Theorem

Nothing Has Alpha Against $\tilde{\pi}_t' R_{t+1}$

Theorem

The following are equivalent:

- R^M_{t+1} *is the return on the conditionally efficient portfolio:*

$$R^M_{t+1} = a_t^{-1}\pi_t' R_{t+1}, \ \pi_t = \mathrm{Cov}_t(R_{t+1})^{-1} E_t[R_{t+1}] \text{ for some } a_t \in \mathbb{R}. \tag{27}$$

-

$$E_t[R_{i,t+1}] = \beta_{i,t} E_t[R^M_{t+1}] \ , \ \beta_{i,t} = \frac{\mathrm{Cov}_t(R_{i,t+1}, R^M_{t+1})}{\mathrm{Var}_t[R^M_{t+1}]} \tag{28}$$

- *Furthermore,* R^M_t *satisfies* (27) *with* $a_t = 1 + E_t[R_{t+1}]'\mathrm{Cov}_t(R_{t+1})^{-1} E_t[R_{t+1}]$ *if and only if it prices returns on any feasible portfolio unconditionally:*

$$E[R_{i,t}] = \beta_i E[R^M_t], \ \beta_i = \frac{\mathrm{Cov}(R_{i,t}, R^M_t)}{\mathrm{Var}[R^M_t]} \tag{29}$$

Proof i

Lemma (Sherman-Morrison formula)

$$\begin{aligned}(A+xx')^{-1} &= A^{-1} - A^{-1}xx'A^{-1}/(1+x'A^{-1}x) \\ (A+xx')^{-1}x &= A^{-1}x/(1+x'A^{-1}x)\end{aligned} \quad (30)$$

for any matrix $A \in \mathbb{R}^{P\times P}$ and any vector $x \in \mathbb{R}^{P}$.

[Proof of Theorem 3] First, if

$$R^M_{t+1} = \pi_t' R_{t+1}, \quad (31)$$

then

$$\mathrm{Var}_t[R^M_{t+1}] = \pi_t'\Sigma_t\pi_t, \quad (32)$$

where $\Sigma_t = \mathrm{Cov}_t(R_{t+1}), \mu_t = E_t[R_{t+1}]$, and

$$\begin{aligned} \pi_t &= a_t^{-1}\Sigma_t^{-1}E_t[R_{t+1}] \Leftrightarrow E_t[R_{t+1}] = a_t\Sigma_t\pi_t \\ &= a_t\mathrm{Cov}_t(R_{t+1}, R_{t+1}^M) = a_t\mathrm{Var}_t[R_{t+1}^M]\beta_t, \end{aligned} \tag{33}$$

where

$$a_t\mathrm{Var}_t[R_{t+1}^M] = a_t\pi_t'\Sigma_t\pi_t = a_t^{-1}\mu_t'\Sigma_t^{-1}\mu_t \tag{34}$$

while

$$E_t[R_{t+1}^M] = \pi_t'\mu_t = a_t^{-1}\mu_t'\Sigma_t^{-1}\mu_t\,. \tag{35}$$

Reversing the arguments, we get that the first two items of the theorem are, in fact, equivalent.

Now, suppose R_{t+1}^M is the efficient portfolio. Then,

$$E[R_{t+1}^Z] = E[Z_tR_{t+1}] = E[Z_tE_t[R_{t+1}]] = E[Z_t\beta_tE_t[R_t^M]]\,. \tag{36}$$

Proof iii

At the same time, by Lemma 4, we have that

$$\begin{aligned} E_t[R_{t+1}R'_{t+1}]^{-1}\mu_t &= (\Sigma_t + \mu_t\mu_t')^{-1}\mu_t \\ &= \Sigma^{-1}\mu_t/(1+\mu_t'\Sigma_t^{-1}\mu_t) \\ &= \ (1+\theta^2_{M,t})^{-1}\Sigma_t^{-1}\mu_t\,, \end{aligned} \tag{37}$$

where $\theta^2_{M,t} = \mu_t'\Sigma_t^{-1}\mu_t$. Hence, if $R^M_{t+1} = R'_{t+1}\pi_t$, the identity

$$E_t[R_{t+1}] \ = \ E_t[R_{t+1}R^M_{t+1}] \tag{38}$$

holds if and only if

$$\pi_t \ = \ (1+\theta^2_{M,t})^{-1}\Sigma_t^{-1}\mu_t \tag{39}$$

because

$$E_t[R_{t+1}R^M_{t+1}] \ = \ E_t[R_{t+1}R_{t+1}]\pi_t \tag{40}$$

Proof iv

Now, standard arguments imply that the conditional identity (38) is equivalent to the unconditional identity

$$E[R_{t+1}^Z] \;=\; E[R_{t+1}^Z R_{t+1}^M] \tag{41}$$

holding for any Z. Furthermore,

$$E[R_{t+1}^Z R_{t+1}^M] \;=\; \mathrm{Var}[R_{t+1}^M]\beta^Z \;+\; E[R_{t+1}^Z]E[R_{t+1}^M] \tag{42}$$

and hence (41) is equivalent to

$$E[R_{t+1}^Z] \;=\; \mathrm{Var}[R_{t+1}^M]\beta^Z \;+\; E[R_{t+1}^Z]E[R_{t+1}^M], \tag{43}$$

which is equivalent to

$$E[R_{t+1}^Z] \;=\; \frac{\mathrm{Var}[R_{t+1}^M]}{1 - E[R_{t+1}^M]}\beta^Z\,. \tag{44}$$

Proof v

Applying (41) to R^M_{t+1}, we get

$$E[(R^M_{t+1})^2] \;=\; E[R^M_{t+1}] \tag{45}$$

and, hence, after some algebra,

$$\frac{\mathrm{Var}[R^M_{t+1}]}{1 - E[R^M_{t+1}]} \;=\; E[R^M_{t+1}]\,. \tag{46}$$

The proof is complete.

Testing Conditional Efficiency

- We cannnot compute $E_t[\cdot]$
- Instead, we can build instruments Z_t and test that

$$E_t[M_{t+1}R_{t+1}] \;=\; 0 \;\Leftrightarrow\; E[Z_t\, M_{t+1}R_{t+1}] \;=\; 0$$

  for all instruments!
- Thus, we need to build infinitely many Z_t thought machine learning and then test

$$\frac{1}{T}\sum_t Z_t\, M_{t+1}R_{t+1} \;\approx\; 0$$

  Complexity is always there!

From Non-Tradable to Tradable SDFs i

- What about asset pricing theory?
- the SDF

$$\widetilde{M}_{t+1} = \underbrace{\frac{e^{-\rho} U'(C_{t+1})}{U'(C_t)}}_{=IMRS}$$

comes from the Euler equation (things get more complex with Epstein-Zin preferences, expectations, sentiments, etc)

$$E_t[\underbrace{\frac{e^{-\rho} U'(C_{t+1})}{U'(C_t)}}_{=IMRS}(R_{t+1} + R_{f,t})] \;=\; 1 \;\Leftrightarrow\; E_t[\widetilde{M}_{t+1} R_{t+1}] \;=\; 0 \quad (47)$$

because

$$R_{f,t} \;=\; E_t[\widetilde{M}_{t+1}]^{-1}\,. \quad (48)$$

From Non-Tradable to Tradable SDFs ii

- When markets are complete,

$$\widetilde{M}_{t+1} = \tilde{a}_t M_{t+1} \tag{49}$$

- In general, we need to **project**

$$\underbrace{\tilde{a}_t M_{t+1}}_{unique\ \ tradable} = Proj_t(\widetilde{M}_{t+1}) = \arg\min_{a,\pi} E_t[(\widetilde{M}_{t+1} - (a - \pi' R_{t+1}))^2] \tag{50}$$

- Note the scale $\tilde{a}_t$ is needed to catch the interest rate,

$$E_t[\tilde{a}_t M_{t+1}] = R_{f,t}^{-1} \tag{51}$$

Table of Contents

Panel Datasets: Leveraging the Power of Big Data i

- Now comes the big question: **How do we measure the conditional** expectations, $E_t[R_{t+1}]$ and $E_t[R_{t+1}R'_{t+1}]$?
- Running prediction models **per stock** is infeasible due to insufficient data:

$$E_t[R_{i,t+1}] \underbrace{=}_{bad\ \ idea} g_i(X_{i,t})$$

- use panel data

$$E_t[R_{i,t+1}] \underbrace{=}_{good\ \ idea} g(X_{i,t})$$

- **panel** means **same function** g for all stocks.
- **non-linear** g means machine learning.

Panel Datasets: Leveraging the Power of Big Data ii

- Conditional covariance matrix

$$E_t[R_{i,t+1}R_{j,t+1}] = \underbrace{X'_{i,t}\,\Sigma_F\,X_{j,t}}_{systematic\;\;covariance} + \underbrace{\delta_{i,j}\sigma^2_{i,t}}_{idiosyncratic\;\;variance}$$

  where Σ_F and $\sigma_{i,t}$ are to be estimated.

- **Can we avoid computing the conditional covariance matrix?**

Managed Portfolios and Rich Conditional Factor Structures i

- Suppose $R_{i,t+1} \;=\; \underbrace{S'_{i,t}}_{conditional\;\;betas} \cdot \underbrace{\tilde{F}_{t+1}}_{latent\;\;factors} + \varepsilon_{i,t+1}$

- $$E_t[\tilde{F}_{t+1}] \;=\; \underbrace{\lambda_F}_{latent\;\;factor\;\;risk\;\;premia}\,,\; E_t[\tilde{F}_{t+1}\tilde{F}'_{t+1}] \;=\; \underbrace{\Sigma_F}_{latent\;\;factor\;\;cov}$$

- $$M_{t+1} \;=\; 1 - \tilde{\pi}'_t R_{t+1} = \;1\; - \;W(S_t)' R_{t+1}\,, \tag{52}$$

  where $\tilde{\pi}_t = E_t[R_{t+1}R'_{t+1}]^{-1} E_t[R_{t+1}]$ and, hence,

  $$W(S_t) \;=\,=\; \underbrace{(S_t \Sigma_{F,t} S'_t + \Sigma_\varepsilon)^{-1}}_{conditional\;\;covariance} \quad \underbrace{S_t \lambda_F}_{conditional\;\;expectation} \tag{53}$$

Managed Portfolios and Rich Conditional Factor Structures ii

- Define **managed portfolios**

$$F_{t+1} \;=\; S_t' R_{t+1}. \tag{54}$$

and the unconditionally efficient portfolio

$$\lambda = E[F_{t+1}F_{t+1}']^{-1}E[F_{t+1}] \tag{55}$$

- By construction,

$$M_{t+1} = 1 - \lambda' F_{t+1} \tag{56}$$

prices factors unconditionally:

$$E[M_{t+1}F_{t+1}] \;=\; 0 \tag{57}$$

- However,

$$E_t[M_{t+1}R_{t+1}] \;\neq\; 0$$

- $$\lambda' S_t' R_{t+1} \neq \lambda_F{}' S_t' \Sigma_t^{-1} R_{t+1},$$

  with

  $$\Sigma_t = (S_t \Sigma_{F,t} S_t' + \Sigma_\varepsilon)$$

Theorem

Suppose that in the limit, as $P \to \infty$, the vector of latent risk premia λ_F satisfies

$$\lambda_F' A \lambda_F \ \to \ 0 \tag{58}$$

for any symmetric, positive definite A with uniformly bounded trace. Let

$$M_{t+1} = 1 - \lambda' F_{t+1}, \tag{59}$$

be the factor approximation for the SDF with λ. Then, M_{t+1} converges to $\tilde{M}_{t+1}$ and the Sharpe ratio of $\lambda' F_{t+1}$ converges to that of $W(S_t)' R_{t+1}$ as $P \to \infty$. In particular,

$$E_t[M_{t+1} R_{t+1}] \ \to \ 0$$

Sources of Complexity i

- We now know: If

$$R_{t+1} \;=\; \underbrace{S_t}_{N_t\times P\;\; signals}\;\; \underbrace{\tilde{F}_{t+1}}_{P\times 1\;\; latent\;\; factors} \;+\; \underbrace{\varepsilon_{t+1}}_{residuals} \tag{60}$$

then we build

$$F_{t+1} \;=\; S_t' R_{t+1} \;=\; (S_t' S_t)\tilde{F}_{t+1} \;+\; (S_t' \varepsilon_{t+1}) \tag{61}$$

- But where do S_t come from?
- Suppose

$$R_{i,t+1} = \beta(X_{i,t})' \underbrace{G_{t+1}}_{K\times 1} + u_{i,t+1}, \tag{62}$$

Sources of Complexity ii

- $$\beta(X_{i,t}) \approx \sum_{p=1}^{P} \xi_p S_{i,t,p} = \underbrace{\Xi}_{K\times P} \underbrace{S_{i,t}}_{P\times 1}, \tag{63}$$

  where

  $$S_{i,t} = A(\Omega X_{i,t}) = (A(\omega_p' X_{i,t}))_{p=1}^{P}. \tag{64}$$

- This gives

  $$\begin{aligned} &R_{t+1} \approx S_t \tilde{F}_{t+1} + u_{t+1},\ \textit{with} \\ &\tilde{F}_{t+1} = \Xi' G_{t+1},\ \nu = E[\tilde{F}_{t+1}] = \Xi' E[G_{t+1}]. \end{aligned} \tag{65}$$

- Thus, if β is non-linear, we need to go for high-dimensional S_t

Sources of Complexity iii

- This gives an SDF

$$M_{t+1} = 1 - \lambda' F_{t+1} = 1 - \lambda' A(X_t)' R_{t+1} = 1 - \sum_{i=1}^{P} w(X_{i,t})\, R_{i,t+1} \tag{66}$$

with

$$w(X_{i,t}) = \sum_{p} \lambda_p A(\omega_p' X_{i,t})$$

Complexity in the Cross Section: A Brief History i

- Most academic attempts to build an SDF assume

$$M^{\star}_{t+1} = 1 - \sum_{i=1}^{N} w(X_{i,t}) R_{i,t+1} \tag{67}$$

- Cross-sectional asset pricing is about $w_t = w(X_t)$
  - Explains differences in average returns
  - Defines the MVE portfolio
- Why does cross-section literature rarely start here? Because w must be estimated
  - This is a high-dimensional (*complex*) problem
  - We know: In-sample tangency portfolio behaves horribly out-of-sample
  - Why? Complexity ($n/T \not\to 0$) $\to$ LLN doesn't apply $\to$ IS and OOS diverge

Complexity in the Cross Section: A Brief History ii

- Standard solution: Restrict w
  - E.g., Fama-French: $w_{i,t} = b_0 + b_1 \text{Size}_{i,t} + b_2 \text{Value}_{i,t}$ (Brandt et al. 2007 generalize):

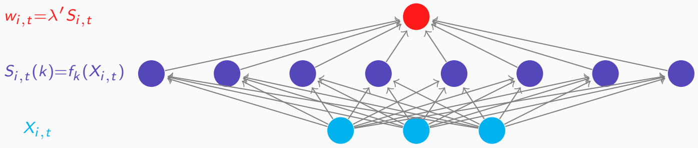
$$\begin{aligned}\sum_{i=1}^{N} w(X_{i,t}) R_{i,t+1} &= \sum_{i=1}^{N} (b_0 + b_1 \text{Size}_{i,t} + b_2 \text{Value}_{i,t}) R_{i,t+1} \\ &= b_0 \sum_{i=1}^{N} R_{i,t+1} + b_1 \sum_{i=1}^{N} \text{Size}_{i,t} R_{i,t+1} + b_2 \sum_{i=1}^{N} \text{Value}_{i,t} R_{i,t+1} \\ &= b_0 MKT_{t+1} + b_1 SMB_{t+1} + b_2 HML_{t+1}\,.\end{aligned} \tag{68}$$

  - Reduces parameters, implies factor model: $M_{t+1} = 1 - b_0 MKT - b_1 SMB - b_2 HML$
  - "Shrinking the cross-section" Kozak et al. (2020) — use a few PCs of anomaly factors

Complexity in the Cross Section: Machine Learning Perspective i

Rather than restricting $w(X_t)$....

- ...expand parameterization, saturate with conditioning information
- For example, approximate w with neural network: $w(X_{i,t}) \approx \lambda' S_{i,t}$
- $P \times 1$ vector $S_{i,t}$ is known nonlinear function of original predictors $X_{i,t}$



Complexity in the Cross Section: Machine Learning Perspective ii

- Implies that empirical SDF is a high-dimensional factor model

$$\sum_{i=1}^{N} w(X_{i,t}) R_{i,t+1} = \sum_{i=1}^{N} (\sum_{k} \lambda_k \underbrace{S_{i,t}(k))}_{S_{i,t}(k)=f_k(X_{i,t})} R_{i,t+1} = \sum_{k} \lambda_k \underbrace{\sum_{i=1}^{N} S_{i,t}(k) R_{i,t+1}}_{F_{k,t+1}} \quad (69)$$

$$M^{\star}_{t+1} \approx M_{t+1} = 1 - \lambda' S'_t R_{t+1} = 1 - \lambda' F_{t+1}$$

Complexity in the Cross Section: Machine Learning Perspective i

The Objective:

- Maximize out-of-sample Sharpe ratio (equivalently, minimize out-of-sample pricing errors) of SDF

The Choice:

- Fix T data points. Decide on "complexity" (number of factors P) to use in approximating model

The Tradeoff:

Complexity in the Cross Section: Machine Learning Perspective ii

- Simple SDF ($P << T$) has low variance (thanks to parsimony) but is a poor approximator of w
- Complex SDF ($P > T$) is a good approximator but may behave poorly (and requires shrinkage)

- Which P should the analyst opt for? Does the benefit of more factors justify their cost?

Answer:

- Use the largest factor model (largest P) that you can compute

Implementation i

- Build a bunch of features (random features if you want a shallow model; deep features (output layer) if you want a deep model).
- Call them $S_{i,t}(k) = f_k(X_{i,t}; \theta_k), \; k = 1, \cdots, P$
- Build the factors

$$F_{t+1}(k) \; = \; \sum_{i=1}^{N_t} S_{i,t}(k) R_{i,t+1} \tag{70}$$

- Take the vector of factors $F_{t+1} = (F_{t+1}(k))_{k=1}^{P}$ and minimize

$$\min_{\lambda} \frac{1}{T} \sum_{t=1}^{T} (1 - \lambda' F_{t+1})^2 \; + \; z \, \|\lambda\|^2 \tag{71}$$

This objective is known as the **Maximal Sharpe Ratio Regression (MSRR)**. For a deep model, you need to minimize this objective using GD

Implementation ii

- Why MSRR? Well,

$$\frac{1}{T}\sum_{t=1}^{T}(1-\lambda' F_{t+1})^2 \approx E[(1-\lambda' F_{t+1})^2] = 1-2E[\lambda' F_{t+1}]+E[(\lambda' F_{t+1})^2] \quad (72)$$

where

$$U(x) = x - 0.5x^2$$

- Now, $\tilde{\pi}_t = E_t[R_{t+1}R'_{t+1}]^{-1}E_t[R_{t+1}]$ solves

$$\max_{\pi} E_t[U(\pi'_t R_{t+1})] \quad (73)$$

It is conditionally efficient for a quadratic utility. By the law of iterated expectations,

$$E[E_t[U(\pi'_t R_{t+1})]] = E[U(\pi'_t R_{t+1})]$$

Implementation iii

and dynamic consistency gives

$$\max_{all\ \ policies\ \ \pi_t} E[U(\pi_t' R_{t+1})] \; = \; E[\max_{\pi} E_t[U(\pi_t' R_{t+1})]$$

- Thus, MSRR looks for conditional policies that maximize unconditional utility and hence, by consistency, are conditionally optimal.
- For a shallow model, you can do it in closed form:

$$\hat{\lambda}(z) \; = \; \left(zI + \frac{1}{T}\sum_{t=1}^{T} F_t F_t' \right)^{-1} \frac{1}{T}\sum_{t=1}^{T} F_t \underbrace{\not\approx}_{Complexity!} \lambda_*(z) \qquad (74)$$

where

$$\lambda_*(z) \; = \; (zI + E[FF'])^{-1} E[F] \qquad (75)$$

Implementation iv

- Limits To Learning

$$E[\hat{\lambda}(z)'F_{T+1}] \approx \frac{Z_*(z)}{z} E[\lambda_*(Z_*(z))'F_{T+1}], \tag{76}$$

where

$$Z_*(z) = z\,(1+\xi(z)) > z. \tag{77}$$

- In fact,

$$E_T[\hat{\lambda}(z)'F_{T+1}] = \frac{Z_*(z)}{z} E_T[\lambda_*(Z_*(z))'F_{T+1}]$$
$$= \frac{Z_*(z)}{z} \lambda_*(Z_*(z))'E[F] = \frac{Z_*(z)}{z} E[F]'(Z_*(z)I + E[FF'])^{-1}E[F] \tag{78}$$

Implementation v

Proof. Let $E[F]=\mu,\ E[FF']=\Psi$; everything is i.i.d. across t. Then,

$$E[\hat{\lambda}(z)'F_{T+1}] \;=\; E[\hat{\lambda}(z)'\mu] \;=\; E[\frac{1}{T}\sum_{t=1}^{T}F_t'\left(zI+\frac{1}{T}\sum_{t=1}^{T}F_tF_t'\right)^{-1}]\mu$$

$$\underbrace{=}_{symmetry}\; E[F_t'\left(zI+\frac{1}{T}\sum_{t=1}^{T}F_tF_t'\right)^{-1}]\mu$$

$$=\; E[F_t'(zI+\Psi_{T,t})^{-1}\frac{1}{1+T^{-1}F_t'\left(zI+\hat{\Psi}_{T,t}\right)^{-1}F_t}]\mu$$

$$\underbrace{\approx}_{F_t\ \text{is independent}}\; \mu'E[\left(zI+\hat{\Psi}_{T,t}\right)^{-1}]\mu(1+\xi(z;c))^{-1} \tag{79}$$

Implementation vi

where

$$\hat{\Psi}_{T,t} \;=\; \frac{1}{T}\sum_{\tau=1}^{T} F_\tau F_\tau' \;-\; F_t F_t'$$

where we have used that

$$T^{-1} F_t' \left(zI + \hat{\Psi}_{T,t}\right)^{-1} F_t \;\approx\; \xi(z; c) \tag{80}$$

The claim follows now from the Master Theorem:

$$z\,\mu' \left(zI + \hat{\Psi}_{T,t}\right)^{-1} \mu \;\approx\; Z_*\left(Z_* I + \Psi\right)^{-1} \tag{81}$$

Table of Contents

Empirical Analysis

- Analyze empirical analogues to theoretical comparative statics
- Study conventional setting with conventional data
  - Forecast target is monthly return of US stocks from CRSP 1963–2021
  - Conditioning info (X_t) is 130 stock characteristics from Jensen, Kelly, and Pedersen (2022)
- Out-of-sample performance metrics are:
  - SDF Sharpe ratio
  - Mean squared pricing errors (factors as test assets)

Empirical Analysis i

Random Fourier Features

- Empirical model: $M_{t+1} = 1 - \lambda' S_t' R_{t+1}$
- Need framework to smoothly transition from low to high complexity
- Adopt ML method known as "random Fourier features" (RFF)
  - Let $X_{i,t}$ be 130×1 predictors. RFF converts $X_{i,t}$ into

$$S_{\ell,i,t} = \sin(\gamma_\ell' X_{i,t}), \quad \gamma_\ell \sim iidN(0, \gamma I)$$

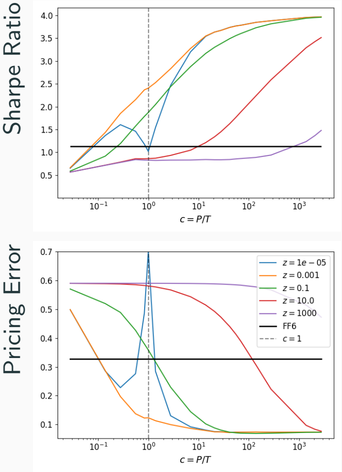
  - $S_{\ell,i,t}$: Random lin-combo of $X_{i,t}$ fed through non-linear activation
  - **we then rank the random features in the cross-section**
- For fixed inputs can create an arbitrarily large (or small) feature set
  - Low-dim model (say $P = 1$) draw a single random weight
  - High-dim model (say $P = 10{,}000$) draw many weights
- In fact, RFF is a two-layer neural network with fixed weights (γ) in the first layer and optimized weights (λ) in the second layer

Empirical Analysis

Training and Testing

- We estimate out-of-sample SDF with:
  i. Thirty-year rolling training window ($T = 360$)
  ii. Various shrinkage levels, $\log_{10}(z) = -12, ..., 3$
  iii. Various complexity levels $P = 10^2, ..., 10^6$
- For each level of complexity $c = P/T$, we plot
  i. Out-of-sample Sharpe ratio of the kernels and
  ii. Pricing errors on 10^6 "complex" factors: $F_{t+1} = S_t' R_{t+1}$
- Also report Sharpe ratio and pricing errors of FF6 to benchmark our results

Out-of-sample SDF Performance

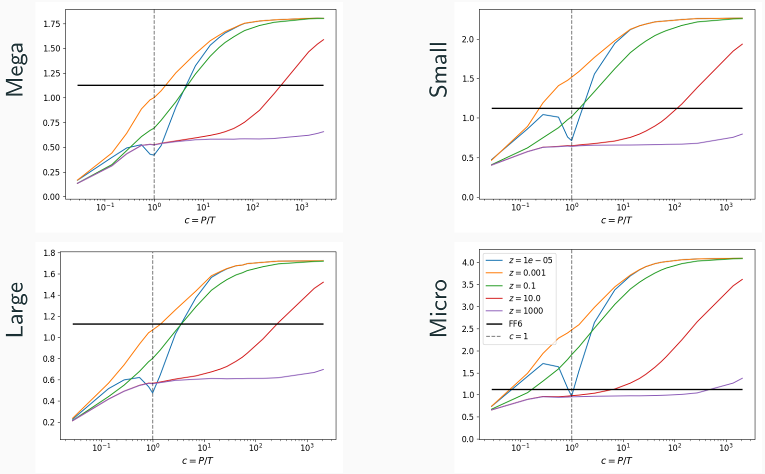


Main Empirical Result

- OOS behavior of ML-based SDF closely matches theory
- High complexity models
    - Improve over simple models by a factor of 3 or more
    - Dominate popular benchmarks like FF6

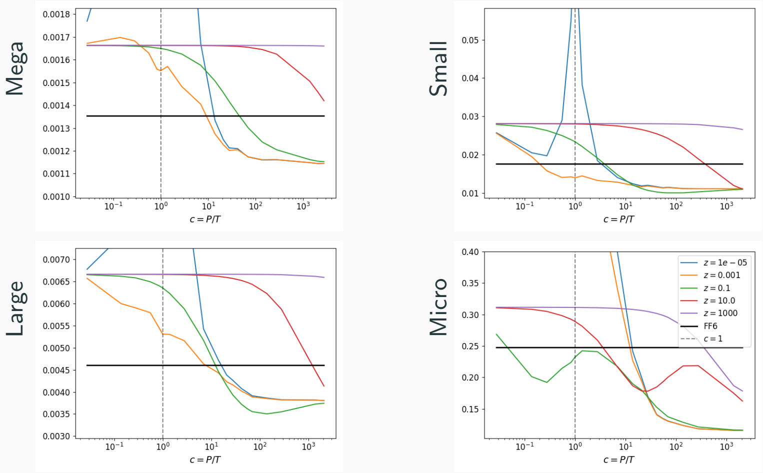
SDF Performance in Restricted Samples: Sharpe Ratio

Market Capitalization Subsamples

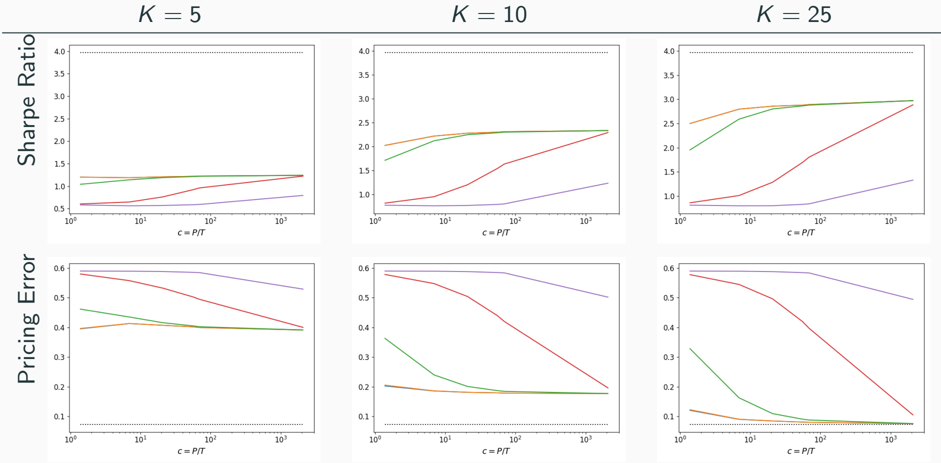


SDF Performance in Restricted Samples: Pricing Errors

Market Capitalization Subsamples



What About "Shrinking" With PCA?



Experiments with Managed Portfolios

Managed Portfolios Notebook