Deep Tangency Portfolios

March 8, 2026

Background in Tangency Portfolio

- Markowitz (1952 JF) marks the beginning of modern portfolio theory, which formulates the solution to the optimal portfolio by relying on covariance and averages of individual asset returns ($\Sigma^{-1}\mu$).
- Constructing the tangency portfolio has economic importance: the tangency portfolio is equivalent to the stochastic discount factor (SDF).
- However, estimating such a tangency portfolio is notoriously difficult (Cochrane, 2014 JF)!

Background in Tangency Portfolio

- It is challenging to estimate expected returns and covariance matrix:
  - # individual assets, N is usually larger than # observations (T), making it difficult to estimate the large covariance matrix (Σ_t);
  - Expected returns (μ_t) are often imprecise even with long samples and high frequency of observations (Merton, 1980 JFE and Cochrane, 2014 JF)
  - Both issues yield a very inaccurate estimate, resulting in the poor out-of-sample performance (see, e.g.,DeMiguel et al., 2009 RFS)
- Alternative studies that bypass estimating return covariance and averages
  - Brandt (1999 JF) and Ait-Sahalia and Brandt (2001 JF) propose nonparametric approaches for portfolio estimation from the Euler first-order conditions.
  - Brandt and Santa-Clara (2006 JF) and Brandt et al. (2009 RFS) provide a parametric approach by estimating portfolio weights as a linear function of characteristics (size, value, and momentum).

Background in Multifactor Efficiency

- A common practice in the literature focuses on alternative representations of the SDF by proxying it as a linear function of a small number of characteristics-managed factors (Fama and French, 1996 JF; Fama and French, 2015 JFE),
  - Hoping that those factors can span the minimum-variance SDF or the mean-variance efficient frontier,
  - i.e., $\mu_F' \Sigma_F^{-1} \mu_F = \mu' \Sigma^{-1} \mu$!?
- However, the commonly used factors can hardly achieve the same maximal Sharpe ratio of the asset universe (see, e.g.Kozak et al., 2018 JF; Daniel et al., 2020 RFS; Lopez-Lira and Roussanov, 2020)
- The finance literature has thus proposed and examined a large number of factors (Harvey et al., 2016 RFS; Hou et al., 2020 RFS) , leading to an issue of “factor zoo” (Cochrane 2011 JF).

Background in Machine Learning

- Machine learning (ML) provides useful tools for handling high-dimensional asset-pricing predictors with flexible models.
  - See a textbook survey by (Negal, 2021) and a review by (Giglio et al., 2022 RFE) , as well as references therein.
- In addition to forecasting, ML can be adapted to many empirical asset pricing methods, e.g., cross-sectional factor models.
- Recent ML methods have been developed for generating latent factors.
  - PCA: Instrumental PCA (Kelly et al., 2019 JFE) , Risk-Premium PCA (Lettau and Pelger, 2018) , etc.
  - Deep Learning: a structural feed-forward network of Feng et al. (2022) and generative adversarial network of Chen et al. (2022 MS)
  - Regression Tree: panel tree (Cong et al., 2022)

Research Question and Takeaways

- This paper endeavors in those directions and proposes a deep learning framework for constructing an Enhancement Device:
  - Tangency portfolio not rely on return covariance and average estimates;
  - We directly parameterize the tangency portfolio weights as a nonlinear function of a large number of characteristics;
  - Using a large set of characteristics is important in spanning expected returns and covariance and revealing the underlying risk-return relationship.
- Our model provides an economic-guided dimension reduction by transforming many characteristics into a deep one.
- A deep factor can be created by approximating a long-short portfolio of individual assets sorted on the deep characteristic.
- Deep tangency portfolio combines the deep factor and benchmarks.

Tangency Portfolio

Mean-Variance Efficiency

The MVE portfolio of Markowitz (1952 JF) takes the form of

$$R_t^{opt} = w_{t-1}' r_t, \tag{1}$$

with the portfolio weight, w_t, given by

$$w_{t-1} = \Sigma_{t-1}^{-1} \mu_{t-1}, \tag{2}$$

Multifactor Efficiency

- Assume that the portfolio weight, w_{t-1}, in (2) can be largely captured in a linear form by J characteristics, z_{t-1}, an $N \times J$ matrix for $J \ll N$, such that

$$w_{t-1} = \widetilde{w}_{t-1} + z_{t-1}\kappa, \tag{3}$$

  - $\widetilde{w}_{t-1}$ is market-capitalization weights;
  - z_{t-1} is the cross-sectionally standardized characteristics.

- Define $R_{m,t} = \widetilde{w}'_{t-1} r_t$, representing the market portfolio, and $f_t = z'_{t-1} r_t$, representing J characteristics-managed long-short portfolios.

Maximal Sharpe Ratio Portfolio

- In this paper, we approximate the tangency portfolio weights by parameterizing w_{t-1} as a nonlinear function of a large set of K assets' characteristics, $z_{i,t-1}$, a $N \times K$ matrix with $K \gg J$.
- We formulate w_t as

$$w_{i,t-1} = \widetilde{w}_{i,t-1} + \theta \; w_d(z_{i,t-1}; \Phi), \quad i = 1, \ldots, N, \tag{4}$$

  — $\widetilde{w}_{i,t-1}$ is the weight of asset i in the market portfolio;

  — $w_d(\cdot)$ is a function of $z_{i,t-1}$ that can account for any potential nonlinear relations among a large number of characteristics of asset i and generates the weights for constructing a zero-cost long-short portfolio;

  — We estimate the portfolio weights as a single function of high-dimensional characteristics that applies to all assets, which extends the low-dimensional and linear approach in Brandt et al. (2009 RFS).

Maximal Sharpe Ratio Portfolio

- The tangency portfolio return in Equation (1) can then be represented by

$$R_t^{opt} = \sum_{i=1}^{N} \widetilde{w}_{i,t-1} r_{i,t} + \theta \sum_{i=1}^{N} w_d(z_{i,t-1}; \Phi) r_{i,t} = R_{m,t} + \theta R_{d,t}, \qquad (5)$$

  - $R_{m,t}$, as before, is the market portfolio return;
  - $w_d(\cdot)$ cross-sectionally sums to 0; $R_{d,t}$ is the returns on a long-short portfolio $\Rightarrow$ deep factor;
  - When the function $w_d(\cdot)$ takes a linear form, and the number of characteristics is set to be small, our parameterization becomes the conventional approach as in Equation (3).

Maximal Sharpe Ratio Portfolio

- When we have *a priori* knowledge that a particular set of observable factors helps span the efficient portfolio frontier, we can introduce these factors by

$$w_{i,t-1} = \widetilde{w}_{i,t-1} + \widetilde{w}^{p}_{i,t-1}\theta_p + \theta_d w_d(z_{i,t-1};\Phi), \quad i = 1,\ldots,N, \tag{6}$$

- The tangency portfolio return is then given by

$$R_t^{opt} = R_{m,t} + \theta_p' R_{p,t} + \theta_d R_{d,t}, \tag{7}$$

- We define the function $w_d(\cdot)$ and estimate model parameters (θ, Φ) by maximizing the average conditional SR^2 of the optimal portfolio R_t^{opt},

$$\max_{\theta,\Phi} \frac{1}{T}\sum_{t=1}^{T} SR^2_{t-1}(R_t^{opt}). \tag{8}$$

Deep Factor Interpretation

- The long-short deep factor, $R_{d,t}$, plays two fundamental roles:
  - According to the principle of diversification, Equation (8) suggests that it should have low or even negative correlations with the market factor, providing us with a potential hedge portfolio.
  - When the market (and benchmark factors) alone can not capture all systematic risk, it spans to a large extent any missing risk factors
- The construction of the deep factor, $R_{d,t}$, only relies on Sharpe ratio improvement over the market or other benchmark factors without using any test assets, similar to Barillas and Shanken (2017 RFS, 2018 JF).

Deep Learning

Deep Learning Framework

- Our construction of the deep factor, $R_{d,t}$, relies on a deep learning model, aiming to utilize high-dimensional characteristics and construct the tangency portfolio by complementing the benchmark factors (market).
- At any time t, three types of data need to be fed into our deep learning:
  - $\{r_{i,t}\}_{i=1}^{N}$, excess returns on N individual assets;
  - $\left\{z_{k,i,t-1} : 1 \leq k \leq K\right\}_{i=1}^{N}$, K lagged characteristics of N assets;
  - $\{R_{p,t}\}_{p=1}^{P+1}$, a $(P+1) \times 1$ vector of excess returns on the market factor and P observable factors.

Deep Characteristics

- We design a L-layer neural network that transforms K characteristics to one deep characteristic.
- For each asset i, $i = 1, \ldots, N$, and at each time t, our deep learning model works as follows,

$$
\begin{aligned}
Z_{i,t-1}^{(0)} &= [z_{1,i,t-1}, \cdots, z_{K,i,t-1}]', &(9)\\
Z_{i,t-1}^{(l)} &= G(A^{(l)} Z_{i,t-1}^{(l-1)} + b^{(l)}), \quad l = 1, \ldots, L &(10)
\end{aligned}
$$

- $G(\cdot)$ is a univariate activation function (e.g., we use *tanh* function),

$$G(x) = (e^x - e^{-x})/(e^x + e^{-x}). \tag{11}$$

- $Z_{i,t-1}^{(l)}$ is the i-th column of the $K_l \times N$ matrix of $Z_{t-1}^{(l)}$, for $1 \leq K_l \leq K$.
- In the end, we have a $1 \times N$ matrix of deep characteristics, $Z_{t-1}^{(L)}$.

Long-Short Deep Factor

- By approximating security sorting, we adopt a nonlinear rank approach to construct the long-short portfolio weights as follows,

$$w_d(z_{t-1}) \equiv W_{t-1} = h(Z_{t-1}^{(L)}), \tag{12}$$

- The function h takes the form of, for the $1 \times N$ vector of $x = Z_{t-1}^{(L)}$,

$$h(z) = \underbrace{\begin{bmatrix} softmax(x_1^+) \\ \vdots \\ softmax(x_N^+) \end{bmatrix}}_{\text{long weights}} - \underbrace{\begin{bmatrix} softmax(x_1^-) \\ \vdots \\ softmax(x_N^-) \end{bmatrix}}_{\text{short weights}}, \tag{13}$$

- Here, $x^+ := -a_1 e^{-a_2 x}$ and $x^- := -a_1 e^{a_2 x}$, where a_1 and a_2 are two hyperparameters in the *deterministic* softmax rank weights.

Deep Factor

- In the implementation, we choose $a_1 = 50$ and $a_2 = 8$ such that, about 50% to 70% assets are in the middle rank and have zero weights.
- We normalize the portfolio weights such that the sum of weights in the long leg equals 1 and that in the short leg equals -1.
- Such a nonlinear rank-weighting scheme depends not only on the cross-sectional rank information, but also on the distributional properties.
- The weights, W_{t-1}, in Equations (12) and (13), sum to zero by construction. The deep factor, $R_{d,t}$, is then computed as

$$R_{d,t} = W_{t-1} r_t. \tag{14}$$

Loss Function

- All parameters in our model are time-invariant, and we implicitly assume that characteristics fully capture all aspects of expected returns and covariance relevant to optimal portfolios;
- The objective function in Equation (8) can be replaced by the unconditional SR^2 of optimal portfolio R_t^{opt} on $\widetilde{F} = [R_{p,t}, R_{d,t}]$,

$$SR^2(R_t^{opt}) \equiv SR^2(\widetilde{F}_t) = \mathbf{E}(\widetilde{F}_t)'\mathbf{Cov}(\widetilde{F}_t)^{-1}\mathbf{E}(\widetilde{F}_t). \tag{15}$$

- The loss function is,

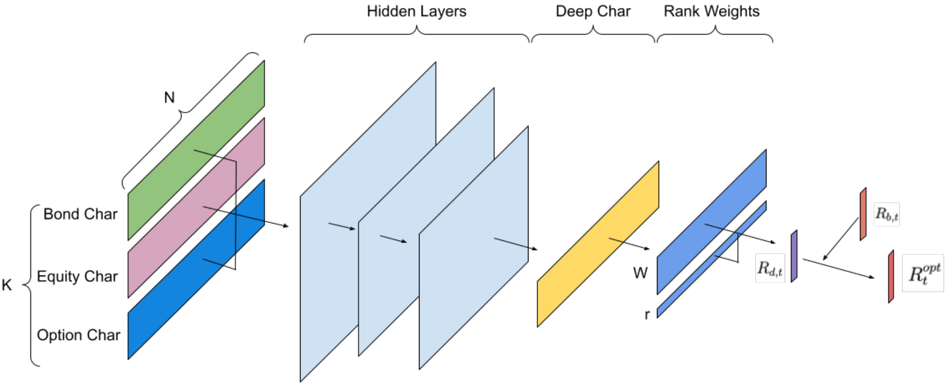
$$\mathcal{L} = \mathop{\mathbb{E}}_{B\sim\mathcal{D}_B, D\sim\text{Dropout}(p)} \left\{ \exp\left[-SR^2\left(R_t^{opt}(B, D) \right) \right] \right\}, \tag{16}$$

where

— $B \sim \mathcal{D}_B$ denotes a randomly sampled mini-batch from the training data;

— $D \sim \text{Dropout}(p)$ represents a stochastic dropout mask with probability p

Deep Learning Network Architecture



Empirical Results

Corporate Bond Returns

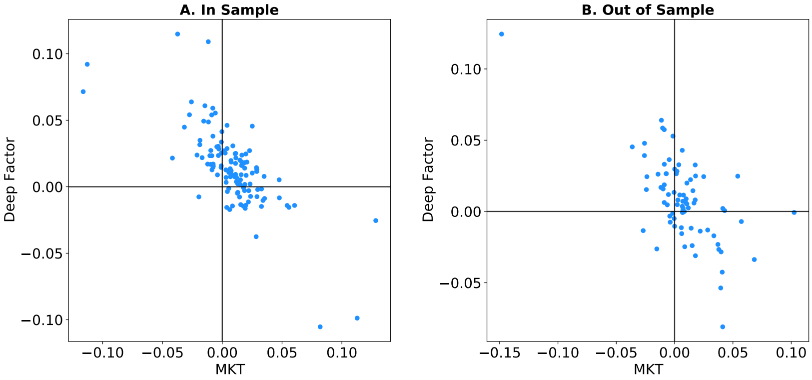
- We apply our method to the cross-section of corporate bonds.
- TRACE Data for constructing corporate bond returns.
- Monthly frequency, July 2004 — December 2020.
- Investment grade (IG): 75.7%, and Public: 77.3%.
- 132 firm characteristics: 41 bond + 61 equity + 30 equity option.

	ALL	IG	NIG	Public	Private
Bond-month observations	590,991	447,603	143,388	456,873	134,118
Return mean (%)	0.45	0.42	0.56	0.44	0.51
Return std (%)	3.41	2.61	5.13	3.15	4.14
Rating mean	8.70	7.02	13.95	8.27	10.18
Duration mean	3.98	4.27	3.09	4.07	3.69
Age mean	4.16	4.24	3.93	4.13	4.26
Size mean ($ million)	846	900	678	871	764

Benchmark and Competing Factors

- Benchmark Market Factor: equal-weighted average of excess corporate bond returns in our sample, i.e., $\widetilde{w}_{i,t} = 1/N$.
- Corporate bond observable-factor model, i.e., the Fama-French five-factor model that combines three bond factors and two equity factors (Fama and French, 1993, 1996)
- Two latent-factor models, i.e., the Risk-Premium PCA (RP-PCA) (Lettau and Pelger, 2020 RFS) and the instrumental PCA (IPCA) (Kelly, Pruitt and Su, 2019 JFE).

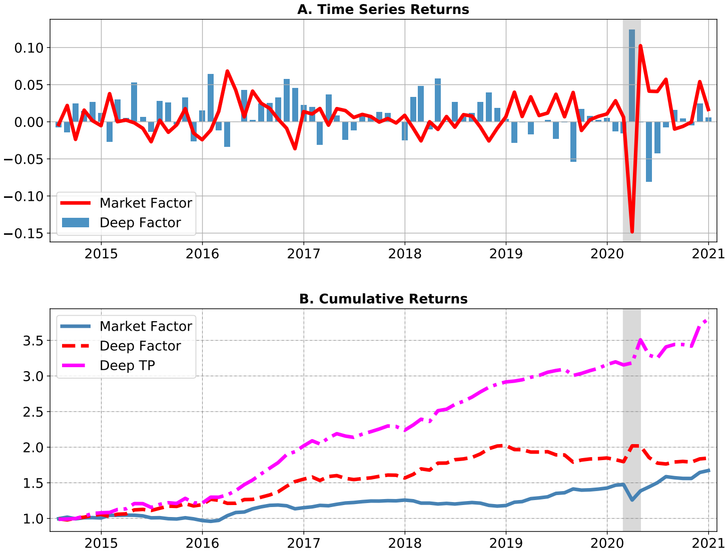
Deep Factor vs. Market Factor



The first fundamental role played by deep factors:

- Market-hedge portfolio (negative correlation).

Time Series and Cumulative Returns



Deep Tangency Portfolios

Panel A: Deep Tangency Portfolios				
	MKT	L_1	L_2	L_3
EW	0.85	0.80	2.13***	0.95***
			[0.89]	
VW	0.90	1.43***	2.21***	1.08***
			[0.88]	

Panel B: Alternative Portfolios			
	FF5	IPCA5	RP-PCA5
SR	0.88	1.48	0.93
	[2.98]	[3.07]	[4.35]
SR (+DF)	1.89***	1.91***	2.04***
	[0.93]	[1.09]	[1.04]

The second fundamental role played by deep factors:

- Enhancement Device: spanning missing risk factors with high SR (2-layer), better than other alternatives.

Variable Importance – Bond/Equity/Option Chars.

	MKT	Bond+Equity +Option	Bond+Equity	Bond	Equity
L_1	0.85	0.80	1.04***	0.76	0.83
L_2	0.85	2.13***	1.01***	0.70	0.57
L_3	0.85	0.95**	0.83	0.24	0.59

- All three types of characteristics are important in constructing the deep tangency portfolio.
- In stark contrast to previous studies that argue those equity characteristics do not necessarily forecast corporate bond returns (see, e.g., Chordia et al., 2017 JFQA; Bali et al., 2021).
- Further empirical evidence in support of integration between bond and equity markets (Schaefer and Strebulaev, 2008 JFE; Choi and Kim, 2018 JME; Kelly et al., 2022 JF).

Conclusion

- A parametric approach for directly estimating the tangency portfolio weights, thus sidestepping average returns and covariance estimation.
- A economic-guided structural neural network that approximates the security sorting and constructs a long-short deep factor.
- The deep factor plays two fundamental roles: (i) market-hedged portfolio; and (ii) enhancement device.
- In the corporate bond market, we find that the deep tangency portfolio earns an out-of-sample annualized Sharpe ratio of 2.13, outperforming alternative methods that used observable and latent factors.
- Finally, we find it important to consider various types of characteristics (bond, equity, option) in corporate bond investing.