Deep Tangency Portfolio*

Guanhao Feng   Liang Jiang   Junye Li   Yizhi Song

First version: Nov. 2021; This version: November 13, 2024

Abstract

We propose a parametric approach that directly estimates the tangency portfolio weights on thousands of individual assets by integrating fundamental finance theory with deep learning models. Our deep tangency portfolio combines the market factor with a long-short deep factor constructed by an economic-guided structural neural network. The deep factor plays two fundamental roles: (i) market-hedged portfolio and (ii) spanning missing risk factors on high-dimensional firm characteristics. Applying this method to the corporate bond market, we find that the deep factor achieves an out-of-sample annualized Sharpe ratio of 1.79, and the deep tangency portfolio outperforms those constructed from commonly used observable or latent factors, with an out-of-sample annualized Sharpe ratio of 2.29. Finally, we also find evidence supporting the integration between the bond and equity markets.

Keywords: Tangency Portfolio, Deep Learning, Corporate Bonds, IPCA, RP-PCA

JEL Classification: C45, G11, G12.

*We appreciate insightful comments from Doron Avramov, Tarun Bali, Zhiguo He, Hendrik Bessembinder, Yong Chen, Hong Liu, Robert Macrae, Andreas Neuhierl, Seth Pruitt, Kuntara Pukthuanthong, Jianeng Xu, Dacheng Xiu, and Mao Ye. We are also grateful for helpful comments from the seminar and conference participants at University of Missouri, University of Chinese Academy of Sciences, Sun Yat-Sen University, Fudan University, Xi'an Jiaotong-Liverpool University, City University of Hong Kong, 2022 AsianFA, 2022 International Conference on Finance & Technology, 2022 China International Risk Forum, FIRM 2022, and 2023 SoFiE annual conference. Feng (Email: gavin.feng@cityu.edu.hk) and Song (Email: yizhisong2-c@my.cityu.edu.hk) are at the City University of Hong Kong, Li (Email: li_junye@fudan.edu.cn) and Jiang (Email: jiangliang@fudan.edu.cn) are at Fudan University.

1 Introduction

The modern portfolio theory, pioneered by Markowitz (1952), formulates an elegant solution to the mean-variance efficient (MVE) portfolio using only expectations and covariance matrix of asset returns ($\Sigma^{-1}\mu$). However, it is notoriously difficult to use such a formula when the number of assets becomes large, making Cochrane (2014, p.7) conclude, *"But this formula is essentially useless in practice. The hurdles of estimating large covariance matrices, overcoming the curse of $\sigma/\sqrt{T}$ in estimating mean returns, and dealing with parameter uncertainty and drift are not minor matters."*

The asset pricing literature usually relies on a small number of characteristic-managed factors (e.g., Fama and French, 1996, 2015), hoping that they can span the efficient frontier. However, the commonly used factors can hardly achieve the maximum Sharpe ratio of the asset universe of either basis portfolios or individual assets (e.g., Kozak et al., 2018; Daniel et al., 2020; Cong et al., 2024). A large number of factors have been proposed to improve the spanning of the efficient frontier and to explain "anomalies" (Harvey et al., 2016; Hou et al., 2020), leading to a "factor zoo" issue (Cochrane, 2011), and the curse of high dimensionality. A recent study by Kozak and Nagel (2024) shows that characteristics-managed factors hardly span the efficient frontier unless a large number of characteristics are used simultaneously.

This paper proposes a deep learning approach for constructing individual assets' optimal or tangency portfolios without estimating their expected returns and the covariance matrix. Unlike the regularized portfolio literature (see, e.g., Ao, Li, and Zheng, 2019; Bryzgalova, Pelger, and Zhu, 2024) that typically uses dozens of assets, our deep tangency portfolio is constructed using thousands of individual assets. The key to constructing our deep tangency portfolio is to utilize high-dimensional firm characteristics, which contain rich information on the joint distribution of asset returns. Cochrane (2011) asserts that expected returns, variances, and covariances are stable functions of characteristics (also see, e.g., Kelly, Pruitt, and Su, 2019). Therefore, we directly parameterize the portfolio weights as a nonlinear function of a large number of firm characteristics – a crucial feature of our model. This aligns with recent machine learning (ML) finance studies that question the sparsity of characteristics (e.g., Kozak et al., 2020; Giannone et al., 2021), and emphasize the importance of nonlinearity (e.g., Freyberger, Neuhierl, and Weber, 2020; Gu, Kelly, and Xiu, 2020).[1]

Specifically, when constructing the tangency portfolio, we consider many individual assets and a few benchmark portfolios, such as the market factor. Using a divide-and-conquer strategy, we estimate the tangency portfolio by combining a deep factor (from individual assets) and the benchmark portfolios. A nonlinear neural network, guided by an economically motivated loss function, provides supervised dimension reduction by transforming high-dimensional characteristics into a deep characteristic for each asset, based on which a single deep factor is formed as a long-short characteristic-sorted portfolio of individual assets. This endogenous latent factor construction, which uses a nonlinear ranking scheme, mimics the widely used characteristic-sorted factor approach for dimension reduction in empirical asset pricing. This economically guided deep factor plays two important roles: (i) it has a low or even negative correlation with the benchmark portfolios under the maximum Sharpe ratio objective, providing a potential hedge portfolio; and (ii) it is constructed using information from high-dimensional characteristics, potentially spanning any missing risk factors other than the benchmark needed for the efficient frontier estimation and tangency portfolio construction.

Our method can be viewed as a customized latent factor construction for portfolio optimization, adapting the framework of the deep characteristics-sorted factor model of Feng et al. (2023). The endogenous deep parametric portfolio approach avoids extreme long and short positions (Avramov, Cheng, and Metzker, 2023) and is flexible

[1]See the latest textbook survey of Negal (2021) and Kelly and Xiu (2023), as well as references therein.

to incorporate various benchmark factors and multiple deep factors. Additionally, it can be readily adapted to other economic objectives, such as the minimum variance portfolio or utility maximization, with different economic constraints, such as the no short-sale or leverage constraint. All these features make our deep parametric portfolio approach a valuable contribution to portfolio construction and asset allocation.

We apply our method to the U.S. corporate bond market, where cross-sectional pricing is less understood than the equity market. The literature has proposed observable factors to explain time-series comovement and cross-sectional variations in corporate bond returns. For example, Fama and French (1993) find that two factors based on bond term and default are important to explain common variation in bond returns. A recent paper by Dickerson, Mueller, and Robotti (2023) argues that the CAPM remains the best model in the corporate bond market.[2] However, most of these models impose strong and ad hoc sparsity by using only a few factors or characteristics, which may suffer from model misspecification and omitted factors. Therefore, observable-factor models may not perform as well as latent factor models that account for high-dimensional characteristics (see, e.g., Kelly, Palhares, and Pruitt, 2022).

We construct monthly corporate bond returns using transaction data on corporate bond prices from the enhanced Trade Reporting and Compliance Engine (TRACE). To employ as many characteristics as possible, we consider three types of characteristics. First, we construct a set of 41 corporate bond characteristics by combining TRACE and the Mergent Fixed Income Securities Database (FISD) data. Second, since bond and stock prices are contingent on firm fundamentals, we collect 61 equity characteristics frequently used in the literature. Third, recent literature suggests that equity option-related variables provide information about future corporate bond returns (e.g., Huang, Jiang, and Li, 2023). Therefore, we construct 30 characteristics related

[2]Other corporate bond risk factors include liquidity (Lin, Wang, and Wu, 2011), momentum (Jostova, Nikolova, Philipov, and Stahel, 2013), volatility (Chung, Wang, and Wu, 2019), and long-term reversal (Bali, Subrahmanyam, and Wen, 2021).

to equity options. To our knowledge, our paper is the first to utilize option-related characteristics for cross-sectional predictability of corporate bond returns. Our deep parametric portfolio approach incorporates these 132 characteristics and considers either the equal- or value-weight bond market factor as the single benchmark. The sample period ranges from July 2004 to December 2020, with a subsample from July 2004 to June 2014 for model training and validation, and a subsample from July 2014 to December 2020 for out-of-sample (OOS) testing.

Our empirical findings are summarized as follows. First, the deep factor can achieve an annualized Sharpe ratio of 1.79 for the OOS period, compared to 0.86 for the equal-weighted market factor. Importantly, it negatively correlates with the bond market factor and rarely declines simultaneously, particularly during market downturns, providing a market hedge. The deep tangency portfolio, combining the market and deep factors, can achieve a Sharpe ratio of 2.29.

Second, we find that the popular Fama-French five-factor model (Fama and French, 1993), which includes the bond market, bond term, bond default, equity size, and value factors, fails to explain excess returns of our deep factor and deep tangency portfolio in the factor-spanning regressions. We further compare the performance of our deep tangency portfolio with those constructed using two recently developed latent factor asset pricing models: risk-premium principal component analysis (RP-PCA) of Lettau and Pelger (2020) and instrumental principal component analysis (IPCA) of Kelly, Pruitt, and Su (2019). Our deep tangency portfolio outperforms both: for the same OOS period, the tangency portfolio from either the five RP-PCA factors or the five IPCA factors achieves a Sharpe ratio of about 1.7.

Third, we further show that it is crucial to consider various types of characteristics when constructing the deep tangency portfolio. The OOS Sharpe ratio of the deep tangency portfolio decreases to 1.83 when option-related variables are excluded, and it further declines when only bond characteristics are used. Therefore, it is crucial to incorporate as many characteristics as feasible to better span the efficient frontier of corporate bonds. This finding is in stark contrast to previous studies that argue that those characteristics that predict equity returns do not necessarily forecast corporate bond returns (see, e.g., Chordia et al., 2017; Bali et al., 2021). However, it is consistent with the recent finding that equity characteristics could provide incremental value for bond return predictability (see, e.g., Dickerson, Mueller, and Robotti, 2023) and offers further evidence in support of the integration of the bond and equity markets (see, e.g., Schaefer and Strebulaev, 2008; Kelly, Palhares, and Pruitt, 2022). Our new findings are from the tangency portfolio perspective that considers both return and risk.

Our paper mainly contributes to the literature on the robust portfolio construction that sidesteps the direct estimation of expected returns and covariance matrix. Early studies include Brandt (1999), Ait-Sahalia and Brandt (2001), and Brandt, Santa-Clara, and Valkanov (2009). Brandt (1999) and Ait-Sahalia and Brandt (2001) propose a non-parametric approach for estimating portfolio weights from the Euler first-order conditions, thereby bypassing the need to estimate the return joint distribution. Brandt, Santa-Clara, and Valkanov (2009) provide a parametric approach by estimating the portfolio weights as a linear function of a few characteristics (also see Brandt and Santa-Clara, 2006). However, their approach cannot accommodate a large number of characteristics or assets, nor handle the nonlinear relationships – two key features of our deep parametric portfolio approach.

Recent studies employ ML techniques with various economic constraints. Guijarro-Ordonez, Pelger, and Zanotti (2024) develop a framework for statistical arbitrage using a convolutional transformer. Jensen, Kelly, Malamud, and Pedersen (2024) propose a framework for constructing an efficient frontier by integrating trading-cost-aware portfolio optimization with ML. Other recent advancements include Liu et al. (2024), which combines the genetic programming with the Sharpe ratio objective in portfolio construction, as well as Chen et al. (2024), Cong et al. (2022), and Simon et al.

(2023) who employ deep learning methods. Our paper contributes to this growing body of literature by focusing on the corporate bond market, where the understanding of pricing factors remains limited. Bali et al. (2021) and Feng et al. (2021) examine the predictability of corporate bond returns using ML methods. Our paper differs by proposing a flexible and interpretable methodology that combines fundamental economic theory with deep learning to construct the tangency portfolio.

The remainder of the paper is organized as follows. Section 2 presents our deep parametric portfolio approach. Section 3 introduces the data on corporate bond returns and firm characteristics. Section 4 discusses our empirical performance and comparison to existing methods. Section 5 concludes the paper. Additional materials and empirical results are provided in the Internet Appendix.

2 Methodology

2.1 Maximum Sharpe Ratio Portfolios

Assume that there are N primitive assets in the economy with excess returns, $r_t = [r_{1,t}, ..., r_{N,t}]'$. The MVE portfolio of Markowitz (1952) takes the form of

$$R^{opt}_{t+1} = w'_t r_{t+1}, \tag{1}$$

with the portfolio weight, w_t, given by

$$w_t = \Sigma_t^{-1} \mu_t, \tag{2}$$

where $\mu_t = E_t[r_{t+1}]$ represents the conditional expected asset returns, and $\Sigma_t = Cov_t(r_{t+1})$ is the conditional covariance matrix of individual assets. It turns out that the maximum conditional squared Sharpe ratio of the optimal portfolio is

$$SR^2_t = \mu'_t \Sigma_t^{-1} \mu_t. \tag{3}$$

In practice, however, estimating both the conditional expected returns (μ_t) and the covariance matrix (Σ_t) is challenging. The number of individual assets, N, is usually very large, making it difficult to estimate the large covariance matrix. Moreover, as a general observation, the estimates of conditional expected returns are often imprecise even with long samples and a high frequency of excess returns (see, e.g., Merton, 1980; Cochrane, 2014). Both issues yield a very inaccurate estimate of w_t in Equation (2), resulting in the poor OOS performance of optimal portfolios (see, e.g., DeMiguel, Garlappi, and Uppal, 2009).

Instead of estimating expected returns and covariance matrix separately, an approach in the finance literature is to directly estimate the portfolio weight (w_t) using a small number of firm characteristics (see, e.g., Brandt, Santa-Clara, and Valkanov, 2009). Assume that the portfolio weight, w_t, in (2) can be largely captured in a linear form by J characteristics, z_t, an $N \times J$ matrix for $J \ll N$, such that

$$w_t = \tilde{w}_t + z_t \kappa, \tag{4}$$

where following the convention of the finance literature, $\tilde{w}_t$ is normalized weights of firm market capitalization, z_t is usually cross-sectionally standardized to have zero means, and κ is a $J \times 1$ vector of coefficients. Define $R_{m,t+1} = \tilde{w}_t' r_{t+1}$, representing the market portfolio, and $f_{t+1} = z_t' r_{t+1}$, representing J characteristic-managed portfolios. The optimal portfolio in Equation (1) then becomes

$$R_{t+1}^{opt} = R_{m,t+1} + \kappa' f_{t+1} = \delta F_{t+1}, \tag{5}$$

where $\delta = [1, \kappa']'$, and $F_t = [R_{m,t}, f_t']'$. The maximum conditional squared Sharpe ratio in Equation (3) is then equal to the maximum conditional squared Sharpe ratio of factors, i.e., $SR_t^2 = \mu_{F,t}' \Sigma_{F,t}^{-1} \mu_{F,t}$.

Such a dimension reduction aims to use those small number of factors (F_t) to span the efficient frontier. Building on the intertemporal capital asset pricing model of Merton (ICAPM, 1973), Fama and French (1996, 2015) also interpret those factors of f_t as *"[they] are just diversified portfolios that provide different combinations of exposure to the unknown state variables"*. However, the literature has found that there does not exist clear-cut evidence of the sparsity of characteristics (e.g., Kozak, Nagel, and Santosh, 2020; Giannone, Lenza, and Primiceri, 2021) and many characteristics and their nonlinear combinations contain information on the joint distribution of asset returns (see, e.g., Freyberger, Neuhierl, and Weber, 2020; Gu, Kelly, and Xiu, 2020). A recent study by Kozak and Nagel (2024) formally shows that the characteristics-managed factors span the efficient portfolio or the SDF only if a large number of characteristics are used simultaneously.

Therefore, in this paper, we sidestep direct estimation of μ_t and Σ_t, or simple reduction of dimension with a few firm characteristics, but instead approximate the tangency portfolio weights by parameterizing w_t as a nonlinear function of a large number of K firm characteristics, z_t, an $N \times K$ matrix for $K \gg J$. We formulate w_t as

$$w_{i,t} = \tilde{w}_{i,t} + \theta\, w_d(z_{i,t}; \Phi), \quad i = 1, \ldots, N, \tag{6}$$

where, as before, $\tilde{w}_{i,t}$ is the weight of asset i in the market portfolio, $w_d(\cdot)$ is a function of $z_{i,t}$ that can account for any potential nonlinear relations among a large number of characteristics of asset i, Φ is the required parameters, and θ is a scalar controlling the relative weight in the tangency portfolio. We estimate the portfolio weights as a single function of characteristics that applies to all assets as in Brandt, Santa-Clara, and Valkanov (2009).

The function, $w_d(\cdot)$, generates the weights for constructing a zero-cost long-short portfolio, as will be explained in detail in the next subsection. The optimal portfolio return in Equation (1) can then be represented by

$$R_{t+1}^{opt} = \sum_{i=1}^{N} \tilde{w}_{i,t} r_{i,t+1} + \theta \sum_{i=1}^{N} w_d(z_{i,t};\Phi) r_{i,t+1} = R_{m,t+1} + \theta R_{d,t+1}, \tag{7}$$

where $R_{m,t+1}$, as before, is the market portfolio return, and given that $w_d(\cdot)$ cross-sectionally sums to zero, $R_{d,t+1}$ is, in fact, the returns on a long-short portfolio constructed based on non-linear combinations of characteristics. When the function $w_d(\cdot)$ takes a linear form, and the number of characteristics is small, our parameterization becomes the standard approach as in Equation (4).

When we have *a priori* knowledge that a particular set of observable factors helps span the efficient frontier, we can introduce these factors by inserting them into Equation (6) and construct the portfolio weights, $w_{i,t}$, as follows,

$$w_{i,t} = \tilde{w}_{i,t} + \tilde{w}_{i,t}^{p} \theta_p + \theta_d w_d(z_{i,t};\Phi), \quad i = 1, \ldots, N, \tag{8}$$

where $\tilde{w}_t^p$ is a $N \times P$ vector of weights on individual assets for constructing the P observable factors, and θ_p is a $P \times 1$ vector of coefficients. The optimal portfolio return is then given by

$$R_{t+1}^{opt} = R_{m,t+1} + \theta_p' R_{p,t+1} + \theta_d R_{d,t+1}, \tag{9}$$

where $R_{p,t+1}$ is a $P \times 1$ vector of returns on P observable factors at time $t+1$. Now denote $\theta = [\theta_p', \theta_d]'$.

The main objective of our model is to find the optimal portfolio that delivers the maximum Sharpe ratio. For this purpose, we search for the possible functional form of $w_d(\cdot)$ and estimate the model parameters θ and Φ by maximizing the average conditional squared Sharpe ratio of the portfolio R_{t+1}^{opt},

$$\max_{\theta,\Phi} \frac{1}{T} \sum_{t=0}^{T-1} SR_t^2(R_{t+1}^{opt}), \tag{10}$$

which suggests that according to the principle of diversification, the long-short portfolio, $R_{d,t+1}$, should have a low or even negative correlation with the market factor (and other benchmark factors), providing us with a potential market hedge portfolio. Therefore, the deep learning algorithm aims to find the optimal transformation of high-dimensional characteristics used to construct the characteristics-sorted factor, which negatively correlates with the market factor, such that the Sharpe ratio of the deep tangency portfolio can be maximized.

Our approach can also be interpreted as a dimension reduction of characteristics and risk factors. Empirically, many studies have proved the failure of CAPM. In addition to the market factor, more factors need to be introduced to the pricing kernel to explain the time-series comovement of asset returns and expected return spreads across individual assets. The most popular factors are characteristic-managed portfolios, such as the Fama-French factors (Fama and French, 1996, 2015). Our framework aims to find such characteristic-managed portfolios based on a fundamental economic theory: the MVE portfolio is equivalent to the SDF. The proposed nonlinear modeling approximates the long-short factor construction using a large number of characteristics and reflects the underlying risk-return relationship. Therefore, when the market (and other benchmark factors) alone cannot capture all systematic risk, the deep factor spans to a large extent any missing risk factors that should enter the pricing kernel. In addition, the dimension reduction in constructing characteristic-managed portfolios relies on the Sharpe ratio improvement over the market or other benchmark factors without using any test assets. Such irrelevance of test assets in factor model comparison has been discussed by Barillas and Shanken (2017, 2018) and Barillas et al. (2020).

In what follows, we propose a deep learning method for constructing the portfolio weights of $w_d(\cdot)$ in Equations (6) and (8) and a deep long-short portfolio $R_{d,t}$. While many popular characteristic-managed factors have sidestepped the high-dimensional problem by focusing on only a few characteristics, we try to consider as many potential characteristics and their nonlinear combinations as possible.

2.2 Deep Factor and Deep Tangency Portfolio

Our long-short portfolio construction, $R_{d,t}$, relies on a deep learning model with an economically motivated target, aiming to construct the tangency portfolio by complementing the benchmark factors. Rather than specifically relying on average returns and a covariance matrix of thousands of individual assets, we retain the conventional sorting scheme in our deep parametric portfolio approach based on information about a large number of characteristics.

Deep Characteristic. We follow the standard modeling approach of a neural network for dimensional reduction of a large number of characteristics (see, e.g., Gu et al., 2020; Feng et al., 2023). We first clarify notations. A typical training observation indexed by time t includes the following types of data:

- $\{r_{i,t}\}_{i=1}^{N}$, excess returns of N individual assets;
- $\{z_{k,i,t-1} : 1 \leq k \leq K\}_{i=1}^{N}$, K characteristics of N assets observed at time $t-1$;
- $\{R_{b,t}\}_{b=1}^{P+1}$, a $(P+1) \times 1$ vector of excess returns on the market factor and P observable factors.

We design a L-layer neural network that transforms K characteristics to one deep characteristic that is relatively interpretable. At each time t and for each asset i, $i = 1, \ldots, N$, our deep parametric portfolio approach works as follows,

$$Z_{i,t-1}^{(0)} = [z_{1,i,t-1}, \cdots, z_{K,i,t-1}]', \tag{11}$$

$$Z_{i,t-1}^{(l)} = G(A^{(l)} Z_{i,t-1}^{(l-1)} + b^{(l)}), \tag{12}$$

for $l = 1, \ldots, L$, where $Z_{i,t-1}^{(l)}$ is the i-th column of the $K_l \times N$ matrix of $Z_{t-1}^{(l)}$, for $1 \leq K_l \leq K$, and $G(\cdot)$ is a univariate activation function, which is chosen to be the *tanh* function in the paper, $G(x) = (e^x - e^{-x})/(e^x + e^{-x})$. $A^{(l)}$ and $b^{(l)}$ are deep learning weight and bias parameters, respectively, and need to be trained in the algorithm.

The algorithm performs the transformation and dimension reduction for each asset without interactions among different assets through the univariate activation function. In the end, we have a $1 \times N$ matrix of deep characteristics, $Z_{t-1}^{(L)}$. The parameters to be trained in this part are deep learning weights A and biases b, namely,

$$\left\{ \left(A^{(l)}, b^{(l)}\right) : A^{(l)} \in \mathbb{R}^{K_l \times K_{l-1}}, b^{(l)} \in \mathbb{R}^{K_l} \right\}_{l=1}^{L}. \tag{13}$$

Deep Factors. The deep characteristics, $Z_{t-1}^{(L)}$, are then used to form weights of a deep portfolio (factor) as follows,

$$w_d(z_{t-1}) \equiv W_{t-1} = h(Z_{t-1}^{(L)}), \tag{14}$$

where the function, $h(\cdot)$, needs to be differentiable.

Following the literature, our first choice of the function $h(\cdot)$ is simply a linear function, resulting in a deep characteristic-managed portfolio, i.e.,

$$W_{t-1} = \frac{a}{N} Z_{t-1}^{(L)}, \tag{15}$$

where a is a scaling parameter.

Moreover, to mimic the commonly used portfolio sort approach (i.e., undifferentiable step function), following Feng et al. (2023), we adopt the *softmax* function and calculate the portfolio weights as follows. For $x = Z_{t-1}^{(L)}$, the function $h(\cdot)$ takes the form of,

$$h(x) = \begin{bmatrix} softmax(x_1^+) \\ softmax(x_2^+) \\ \vdots \\ softmax(x_N^+) \end{bmatrix} - \begin{bmatrix} softmax(x_1^-) \\ softmax(x_2^-) \\ \vdots \\ softmax(x_N^-) \end{bmatrix}, \tag{16}$$

where $x^+ := -a_1 e^{-a_2 x}$ and $x^- := -a_1 e^{a_2 x}$, and a_1 and a_2 are two tuning parameters.

The nonlinear *softmax* function is an increasing function,

$$softmax(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{N} e^{x_j}}, \tag{17}$$

and $\sum_{i=1}^{N} softmax(x_i) = 1$. On the right-hand side of Equation (16), the first term represents the long position weights of assets, and the second term is for the symmetric short position. In implementation, we choose $a_1 = 50$ and $a_2 = 8$ such that at each time, about 50% to 70% assets are in the middle rank and have zero weights, similar to the traditional sorting procedure (see Figure A2 in the Internet Appendix). Furthermore, 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. As discussed in Feng et al. (2023), such a nonlinear ranking scheme depends on the cross-sectional rank information and the distributional properties of characteristics. Such construction of the deep factor avoids extreme positions in both long and short legs (Avramov, Cheng, and Metzker, 2023).

In what follows, we refer to Equation (15) as the linear ranking and to Equation (16) as the softmax ranking. The deep factor portfolio weights, W_{t-1}, in Equation (15) or (16), sum to zero by construction. Our deep factor, $R_{d,t}$, can then be computed as

$$R_{d,t} = W_{t-1} r_t. \tag{18}$$

Loss Function. The deep factor in Equation (18) can be combined with the market or other benchmark factors to form the deep tangency portfolio as in Equation (7). Note that more than one deep factor can be constructed iteratively by treating the previous one as a new benchmark factor in our algorithm. As a result, the additional deep factor may capture pricing information not contained in the previous one.

Given that all parameters in our model are time-invariant and that we implicitly assume that characteristics fully capture all aspects of expected returns and covariance relevant to optimal portfolios, the conditional model becomes an unconditional one, and the objective function in Equation (10) can be replaced by the unconditional squared Sharpe ratio of the optimal portfolio R_t^{opt} on $\tilde{F}_t = [R'_{b,t}, R_{d,t}]'$,

$$SR^2(R_t^{opt}) \equiv SR^2(\tilde{F}_t) = \mathbf{E}(\tilde{F}_t)'\mathbf{Cov}(\tilde{F}_t)^{-1}\mathbf{E}(\tilde{F}_t). \tag{19}$$

There are usually a large number of parameters for modeling a multi-layer neural network. To avoid overfitting and improve the model's OOS performance, we augment the objective function by introducing the regularization penalties and minimizing the following loss function,

$$\mathcal{L}_{\gamma_1,\gamma_2} = \exp\left\{-SR^2(R_t^{opt})\right\} + \underbrace{\gamma_1 \sum_{l=1}^{L-1} \left|A^{(l)}\right| + \gamma_2 \sum_{l=1}^{L-1} ||A^{(l)}||^2}_{\text{penalties}}, \tag{20}$$

where the L_1-norm and L_2-norm penalties aim to restrict the complexity of the neural network, stabilize parameters, and thus avoid overfitting.[3] The tuning parameters, γ_1 and γ_2 need to be tuned through training and validation. Figure 1 presents a visualization of our deep learning architecture and summarizes the critical stages for constructing the deep factor and the deep tangency portfolio.

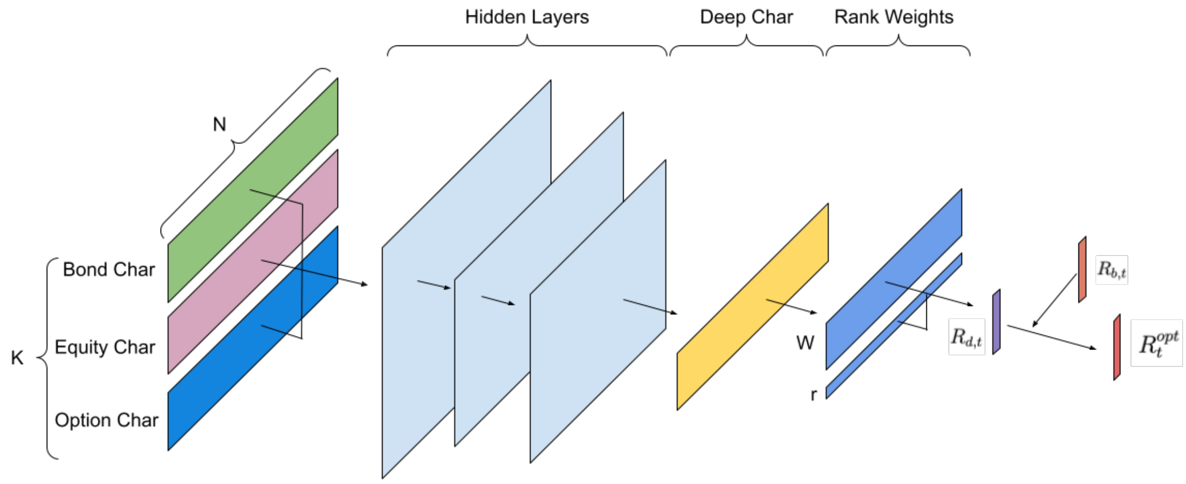
3 Data

To illustrate the performance of our methodology, we apply it to the corporate bond market, given that relative to the equity market, studies on the cross-sectional pricing of corporate bonds remain limited. We first construct the corporate bond returns based on the TRACE data in Subsection 3.1; we then introduce various types of characteristics that will be fed into our deep parametric portfolio approach in Subsection 3.2 and present the benchmark factor and competing factor models in Subsection 3.3.

[3]We want to emphasize that our deep parametric portfolio approach is not merely a return prediction model. Instead, it functions as a long-short factor automatic generation model, where the spreads between the long and short tails of about 30% of the observations are not easily overfitted by transforming the high-dimensional characteristics into a single one for univariate bond sorting.

Figure 1: Deep Learning Network Architecture

This figure provides a visualization of the deep learning architecture. Different types of characteristics, $Z^{(0)}$ (e.g., equity, bond, and option characteristics) are transformed via the multi-layer neural network to deep characteristics, $Z^{(L)}$, based on which the deep portfolio (factor) weights, W, are formed. The deep tangency portfolio, R^{opt}, is constructed by combining the deep factor, R_d, and the benchmark factors, R_b.



3.1 Corporate Bond Returns and Summary Statistics

We obtain corporate bond intraday transaction data from the enhanced version of TRACE, which offers the best-quality data on corporate bond prices, trading volume, and buy-sell indicators. Using TRACE transaction data to measure abnormal corporate bond performance is emphasized in Bessembinder et al. (2009). We merge the TRACE dataset with the FISD to obtain bond characteristics such as offering date, offering amount, maturity date, coupon type and rate, bond type and rating, interest payment frequency, and issuer information.

Following the standard procedures in Dick-Nielsen (2009, 2014), we exclude duplicates, withdrawn, and erroneous trade entries in the TRACE data. Additionally, we apply several filters to the data such that we remove: (i) bonds that are not listed or traded in the U.S. public market; (ii) bonds that are structured notes, mortgage-backed, asset-backed, agency-backed, or equity-linked; (iii) convertible bonds whose option feature distorts the return calculation and makes it impossible to compare the returns of convertible and nonconvertible bonds; (iv) bonds with time to maturity of fewer than two years; and (v) bonds that trade under \$5 or above \$1,000. We then calculate the daily bond price as the trading-volume-weighted average of intraday prices, as in Bessembinder et al. (2009). In line with the literature (see, e.g., Dickerson, Mueller, and Robotti, 2023), for each corporate bond i, its return at month t is calculated as follows:

$$\tilde{r}_{i,t} = \frac{Pr_{i,t} + AI_{i,t} + C_{i,t}}{Pr_{i,t-1} + AI_{i,t-1}} - 1, \tag{21}$$

where $Pr_{i,t}$ is its transaction price in month t, $AI_{i,t}$ is its accrued interest, and $C_{i,t}$ is its coupon payment in month t. We identify two scenarios to calculate a realized return at the end of the month t: (i) from the end of the month $t-1$ to the end of the month t and (ii) from the beginning of month t to the end of the month t. The end (beginning) of the month refers to the last (first) five trading days in that month, and if there is more than one trading record in this five-day window, we use the last (first) observation of the month. If a return at the end of a month is realized in both scenarios, we use the realized return from the end of the month $t-1$ to the end of the month t. The excess bond return is then defined as the difference between the bond return and the risk-free rate, $r_{i,t} = \tilde{r}_{i,t} - r_{f,t}$, where the risk-free rate, $r_{f,t}$, is proxied by the one-month Treasury bill rate obtained from CRSP. Furthermore, similar to Feng et al. (2023), to avoid the volatility and liquidity effect of small market value bonds, we make a balanced panel by only keeping 3,200 bonds with the largest size each month. The final sample of corporate bond returns spans from July 2004 to December 2020.

Table 1 presents the summary statistics of excess corporate bond returns and typical bond characteristics. Our sample includes 16,188 corporate bonds and 633,600 bond-month return observations. As shown in Panel A, the mean monthly excess bond return is about 0.49% with a standard deviation of 4.32%. The sample contains bonds with an average size of about 809 million, an average rating of 8.78, which is a BBB+ rating[4]. Panel A also reports the cross-sectional statistics of investment grade (IG) bonds, which takes about 74.8% of all observations, and non-investment grade (NIG) bonds. Compared to the NIG bonds, the IG bonds have a smaller average return (0.43% vs. 0.69%), a lower standard deviation (2.75% vs. 7.19%), and a higher rating level (7.04 vs. 13.95). The last two columns report summary statistics of the public and private bonds. The public bonds take about 76.7% of all the bond-month observations, their returns are smaller on average, and their ratings are higher on average, compared to private bonds. Both IG and Public bonds have much larger average sizes than their counterparts. Panel B and C report the sample distributions by Rating & Maturity and Ownership & Rating, respectively. A general observation is that most bonds with high ratings are long-maturity bonds.

3.2 Characteristics

We consider three types of characteristics that contain useful information for corporate bond return predictability. The first type of characteristics includes 41 bond characteristics that can be classified into three major categories: basis characteristics (e.g., rating, duration, liquidity), return-distribution characteristics (e.g., momentum, reversal, variance, skewness), and covariances with common risk factors (e.g., market beta, TERM beta, DEF beta).

Furthermore, given that both bond and stock prices are contingent on firm fundamentals, we also consider those equity characteristics shown helpful in predicting equity returns. Recent studies have shown that bond and equity markets are largely integrated. Choi and Kim (2018) argue that market integration suggests different markets should share common factors. Schaefer and Strebulaev (2008) show that bond and equity returns are related through the capital structure hedge ratio. By approximating

[4]Ratings are represented in numerical scores, where 1 refers to an AAA rating, 2 refers to an AA+ rating, . . . , and 21 refers to a C rating. Investment-grade bonds have ratings from 1 (AAA) to 10 (BBB-), and non-investment-grade bonds have ratings of 11 or above. Similar to Bai, Bali, and Wen (2019), we use the ratings of Standard & Poor's (S&P) or Moody's to determine a bond's rating. When both rating companies rate a bond, we use the average of their ratings.

Table 1: Summary Statistics

The table reports summary statistics of the whole sample (ALL) and several subsamples constructed based on Rating (Investment Grade(IG) & Non-Investment Grade(NIG)), ownership (Public & Private), and/or Maturity. The data contain 633,600 month-return observations of 16,188 unique corporate bonds from July 2004 to December 2020.

Panel A: Cross-sectional statistics

	ALL	IG	NIG	Public	Private
Bond-month observations	633,600	474,105	159,495	486,143	147,457
Ret mean (%)	0.49	0.43	0.69	0.45	0.64
Ret std (%)	4.32	2.75	7.19	3.43	6.44
Rating mean	8.78	7.04	13.95	8.32	10.3
Duration mean	3.97	4.26	3.09	4.07	3.63
Age mean	4.25	4.33	4.00	4.22	4.33
Size mean (million)	809	865	644	836	719

Panel B: Sample Distribution(%) by Maturity and Rating

Maturity	AAA	AA	A	B	Junk	ALL
2	0.15	0.71	2.61	2.65	1.44	7.57
3	0.19	0.77	3.00	3.28	2.01	9.25
4	0.18	0.77	3.02	3.54	2.54	10.06
5	0.15	0.74	3.02	3.73	3.11	10.75
6	0.10	0.40	1.92	2.84	3.44	8.69
7	0.09	0.38	1.89	2.88	3.42	8.66
8	0.08	0.34	1.75	2.77	2.61	7.55
9	0.07	0.35	1.75	2.81	1.92	6.90
10	0.07	0.35	1.72	2.74	1.34	6.22
$\geq$11	0.58	1.53	8.59	10.33	3.33	24.35
ALL	1.66	6.33	29.27	37.56	25.17	100.00

Panel C: Sample Distribution(%) by Ownership and Rating

Ownership	AAA	AA	A	B	Junk	ALL
Private	0.11	1.21	4.55	8.43	8.97	23.27
Public	1.55	5.12	24.72	29.13	16.21	76.73
ALL	1.66	6.33	29.27	37.56	25.17	100.00

the hedge ratio with a Merton model for debt, they find that the sensitivity of debt returns to equity is close to that predicted by the Merton model. Building on Schaefer and Strebulaev (2008) and Choi and Kim (2018), Kelly, Palhares, and Pruitt (2022) find that debt and equity markets are more integrated than previous estimates suggest, and that these markets are substantially more integrated in terms of their systematic risks than their idiosyncratic risks. Therefore, the second type includes a total of 61 equity characteristics that cover six major categories: momentum, value, investment, profitability, frictions or size, and intangibles, which are also used in Feng et al. (2023).

In addition, the recent literature has found that several option-related variables have predictive power for corporate bond returns (see, e.g., Cao et al. 2022; Chung et al. 2019; Huang et al. 2023). We, therefore, construct a total of 30 option-related characteristics. Many of those option-related variables have been shown to have predictive power for equity returns (see, e.g., Neuhierl et al., 2021); here, we examine whether they also help forecast corporate bond returns.

Altogether, we have a large number of characteristics (in total, 132). The bond, equity, and option characteristics are listed in Table A1, Table A2, and Table A3, respectively, in the Internet Appendix. Before feeding those characteristics into our deep parametric portfolio approach, we cross-sectionally rank and standardize them each month so that they are in the range of $[-1, 1]$, and their cross-sectional averages are equal to 0. Any missing values are imputed to be 0. One advantage of using the cross-sectional ranks of characteristics is that the impact of potential data errors and outliers in individual characteristics can be largely alleviated (see, e.g., Kelly, Pruitt, and Su, 2019; Freyberger, Neuhierl, and Weber, 2020; Kozak, Nagel, and Santosh, 2020).

3.3 Benchmark Market Factor and Competing Factors

Benchmark Factors. Recent studies in the corporate bond market show that the market factor is the most important one and explains most of the characteristic-managed bond factors (see, e.g., Dickerson, Mueller, and Robotti, 2023). Therefore, we take both the equal- and value-weight corporate bond market portfolios as our benchmarks.

Competing Factors. We consider one 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), and two latent-factor models, i.e., the Risk-Premium PCA (RP-PCA) (Lettau and Pelger, 2020) and the instrumental PCA (IPCA) (Kelly, Pruitt, and Su, 2019).

(i) The Fama-French five factors (FF5). We combine the Fama-French three bond factors, i.e., the value-weight bond market factor, the term factor, and the default factor (Fama and French, 1993), and two equity factors, i.e., the size factor (SMB) and the value factor (HML) (Fama and French, 1996). The term factor is constructed as the difference between the long-term government bond returns and the one-month Treasury bill rate, and the default factor is the difference between the long-term corporate bond returns and the long-term government bond returns. The equity size and value factors are obtained from Ken French's online data library.[5]

(ii) IPCA Factors. A recent paper by Kelly, Palhares, and Pruitt (2022) shows that a five-factor model based on the IPCA of Kelly, Pruitt, and Su (2019) outperforms commonly used observable factor models in pricing corporate bonds. We follow their IPCA approach and construct five corporate bond factors using our TRACE data and all three types of characteristics.

(iii) RP-PCA Factors. Following the procedure of (Lettau and Pelger, 2020), we construct five RP-PCA factors using 48 tercile portfolios based on 16 characteristics, which turn out to be the most important variables contributing to the deep characteristic.[6] See Figure 4 in Section 4.5 for these 16 characteristics.

[5]See https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. We thank him for making these data available online.

[6]Given that our sample size is relatively small (120 months in sample and 78 months out of sample), the number of portfolios used for constructing RP-PCA factors cannot be too large.

4 Empirical Findings

In our empirical implementation, we split the sample into two parts: the subsample from July 2004 to June 2014 for model training and validating, and the subsample from July 2014 to December 2020 for OOS testing. This split is based on the consideration that the in-sample period is long enough to cover various types of market conditions, including the 2008 global financial crisis. We employ a two-fold deterministic cross-validation scheme to determine the tuning parameters and the learning rate. The Internet Appendix (Section III) presents results from alternative split and validation schemes for robustness checks.[7] The number of neural network layers is also a tuning parameter in the range of [1, 3].[8] See Section II in the Internet Appendix for implementation details.

In what follows, we present our main empirical findings and examine how much improvement the deep factor can make over the benchmark, and make a comparison with the competing observable and latent factor models.

4.1 Deep Corporate Bond Factors

Table 2 presents summary statistics of the OOS deep factors. We take both the equal- and value-weight corporate bond market factors as our benchmarks when constructing deep factors and restrict the weights of the long and short legs to be 1 and -1, respectively. In implementation, we normalize the in-sample annualized volatility of the deep factor to 10% and adjust its OOS returns accordingly. All OOS results are

[7] To be specific, we consider two alternative splits. The first is to use the sample from January 2011 to December 2020 for in-sample training and validation and the sample from July 2004 to December 2010 for OOS testing, with the same two-fold validation scheme (Section III.1). The other is a random split scheme (Section III.2). Following Fama and French (2018), we randomly split the data into three groups sequentially over time; that is, for each quarter, the three-month data in this quarter are randomly assigned to three groups, and this fashion is repeated over time up to the end of the sample. We use the data of any two groups for training and validation and the data of the other for the OOS test.

[8] To be specific, the two-fold deterministic design divides the sample from July 2004 to June 2014 into two equal-length consecutive fold samples. We train our neural network separately on one fold and then calculate the fitted results with different tuning parameters on the other. We average the loss from the validation samples and choose the parameter pair that results in the smallest loss. Finally, we refit the model with the selected tuning parameters.

based on the in-sample parameter estimates. The optimal number of layers is trained and validated using the in-sample data. The 1-layer shallow neural network is always the preferred choice, regardless of whether the linear ranking in Equation (15) or the softmax ranking in Equation (16) is considered. However, for comparison, Panels A and B present deep factors constructed from the 1-, 2-, and 3-layer neural networks for the OOS period.

Although in-sample training and validation show that the shallow neural network works well and generates the highest Sharpe ratios for both the linear and softmax ranking, the OOS tests suggest slightly different results. While the softmax ranking-based deep factor continues to outperform in the 1-layer neural network, generating a Sharpe ratio of 1.79 when using the equal-weight market factor as the benchmark and 1.41 when using the value-weight market factor as the benchmark, the linear ranking-based deep factor achieves the highest Sharpe ratio in the 2-layer neural network regardless of whether the equal- or the value-weight market factor is used as the benchmark. The softmax ranking-based deep factor performs much better than the linear ranking-based one for both benchmarks. Therefore, in what follows, we implement our empirical analysis mainly relying on the softmax ranking-based deep learning model.

The table also presents the mean and its Newey-West t-statistics (in parentheses), minimum, maximum, and maximum drawdown of the deep factor. Taking the softmax ranking-based deep factor as an example, we see that the deep factor returns vary from -5.99% (min) to 10.4% (max), with the mean return of 1.80% (t = 4.41) when taking the equal-weight market factor as the benchmark, and vary from -7.96% (min) to 7.85% (max), with the mean return of 1.41% (t = 3.27) when taking the value-weight market factor as the benchmark. The max drawdowns are about 10.2% and 11.5%, respectively, for both benchmarks, which are relatively small compared to the standard bond or equity market factor.

Table 2: Deep Corporate Bond Factors

The table presents descriptive statistics of the deep factor returns, including mean and its Newey-West t-statistics (Newey and West, 1987), minimum, maximum, maximal dropdown, and the annualized Sharpe ratio. The factor's in-sample annualized volatility is normalized to 10%, and its OOS returns are adjusted accordingly. Both equal- and value-weight market factors are considered as benchmarks, along with linear and softmax ranking schemes. R^i denotes the deep factor return from a neural network with i layers. The sample from July 2004 to June 2004 is used for model training and validation, while the sample from July 2014 to December 2020 is for OOS testing.

	Equal-Weight Market Factor					Value-Weight Market Factor				
	Mean	SR	Min	Max	Max DD	Mean	SR	Min	Max	Max DD
					Panel A. Deep Factors: Linear Ranking					
R_l^1	0.77 (1.61)	0.77	−10.85	9.75	23.44	0.91 (2.73)	1.22	−5.59	6.71	12.50
R_l^2	0.71 (2.68)	1.33	−4.21	5.93	6.34	1.09 (3.82)	1.37	−7.62	9.37	10.06
R_l^3	0.44 (3.04)	0.96	−3.64	7.00	7.50	0.78 (3.17)	1.14	−5.00	13.29	7.81
					Panel B. Deep Factors: Softmax Ranking					
R_s^1	1.80 (4.41)	1.79	−5.99	10.38	10.20	1.21 (3.27)	1.41	−7.96	7.85	11.45
R_s^2	0.89 (3.29)	1.30	−5.61	7.28	7.93	1.53 (3.87)	1.25	−14.10	13.44	17.72
R_s^3	0.35 (1.11)	0.50	−7.21	9.21	22.57	1.89 (3.42)	1.41	−10.55	16.58	17.80

4.2 Deep Tangency Portfolios

We now move on to examine the performance of deep tangency portfolios. Panel A of Table 3 presents annualized Sharpe ratios of our deep tangency portfolios for the OOS period. Note that all portfolios' weights are determined by the in-sample estimates, and the optimal number of neural network layers from training and validation is 1. When taking the equal-weight market factor as the benchmark, the deep tangency portfolio achieves the highest annualized OOS Sharpe ratio of 2.29 in the 1-layer softmax ranking-based deep learning model, with the maximum drawdown of 23.3% (in the square brackets). When we employ the value-weight market factor as the benchmark, the deep tangency portfolio performs almost the same. The Sharpe ratios of the deep tangency portfolios are much higher than those of the corresponding market factors, and the Sharpe ratio improvements are highly statistically significant according

Table 3: Deep Tangency Portfolios and Competing Portfolios

The table presents the Sharpe ratios of various portfolios for the out-of-sample period. Panel A is for the deep tangency portfolio that takes either the equal-weight or value-weight market factor as the benchmark and employs the softmax ranking. L_i represents a neural network with i layers. Panel B presents the Sharpe ratios of competing portfolios spanned by the Fama-French five factors, IPCA five factors, and RP-PCA five factors, respectively (SR), and of portfolios spanned by these different types of factors together with the deep factor (SR (+DF)) constructed using the equal-weight market factor as the benchmark and the softmax ranking. In square brackets, the maximum drawdowns in percentage are reported. We follow Barillas and Shanken (2017) to test the significance of the Sharpe ratio improvement. In Panel A, we test the Sharpe ratio improvement of the deep tangency portfolio (L_1) over the benchmark market factor, and in Panel B, we test the improvement of SR (+DF) over SR. $***$, $**$, and $*$ denote the level of significance of 1%, 5%, and 10%, respectively. The sample from July 2004 to June 2014 is used for model training and validating, and the sample from July 2014 to December 2020 is for out-of-sample testing.

Panel A. Deep Tangency Portfolios					Panel B. Competing Portfolios			
	MKT	L_1	L_2	L_3		FF5	IPCA5	RP-PCA5
Equal-Weight	0.86	2.29***	1.48	1.13	SR	0.78	1.66	1.67
		[23.3]				[10.1]	[10.2]	[9.38]
Value-Weight	0.92	2.12***	1.50	1.55	SR (+DF)	2.29***	2.22***	2.38***
		[19.3]				[23.1]	[25.2]	[24.2]

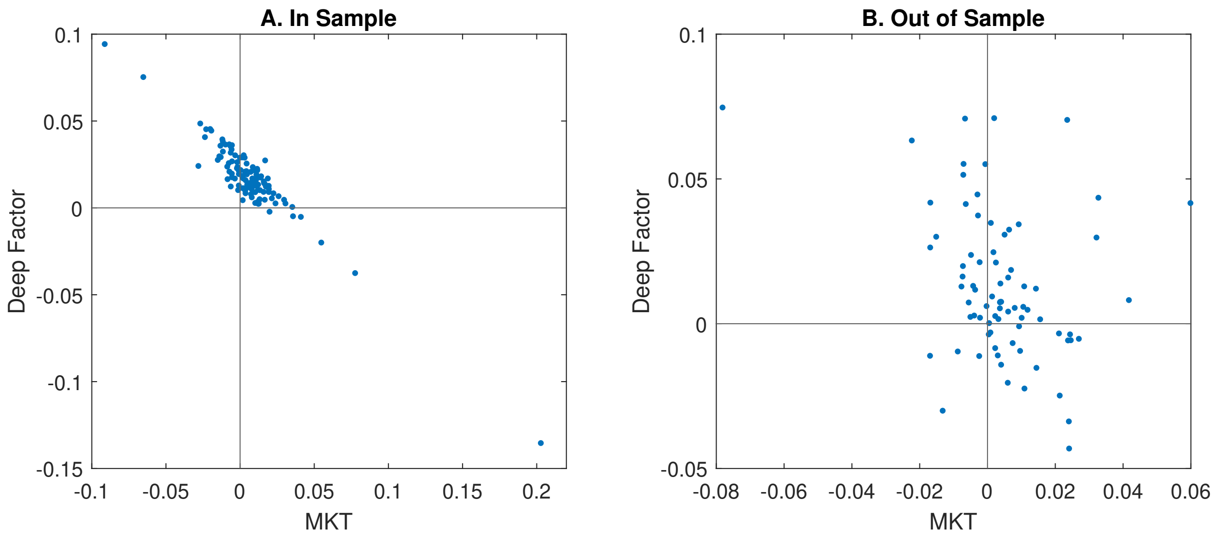
to Barillas and Shanken (2017).[9]

The high Sharpe ratio of the deep tangency portfolio stems from the negative correlation between the market factor and the deep factor in our trained deep learning model. Panel A of Figure 2 presents the scatter plot between the equal-weight market factor and the 1-layer softmax ranking-based deep factor for the in-sample period. We see that they are highly negatively correlated. Panel B of Figure 2 presents the same scatter plot for the OOS period, indicating that the deep factor is negatively correlated with the market factor. Notably, the deep factor and the market factor hardly drop down simultaneously, as few observations are in the lower-left quadrant. Such a negative correlation results in a high Sharpe ratio of the deep tangency portfolio according to the principle of diversification.

[9]Section IV in the Internet Appendix presents the performance of the deep tangency portfolio accounting for transaction costs, proxied by 50% of the bid-ask spread.

Figure 2: **Correlations between the Market and Deep Factors**

Panels A and B present scatter plots of the bond market factor and the deep factor for the in-sample and out-of-sample periods. The deep factor is constructed using the equal-weight market factor as the benchmark and the softmax ranking. The in-sample period ranges from July 2004 to June 2014, and the out-of-sample from July 2014 to December 2020.

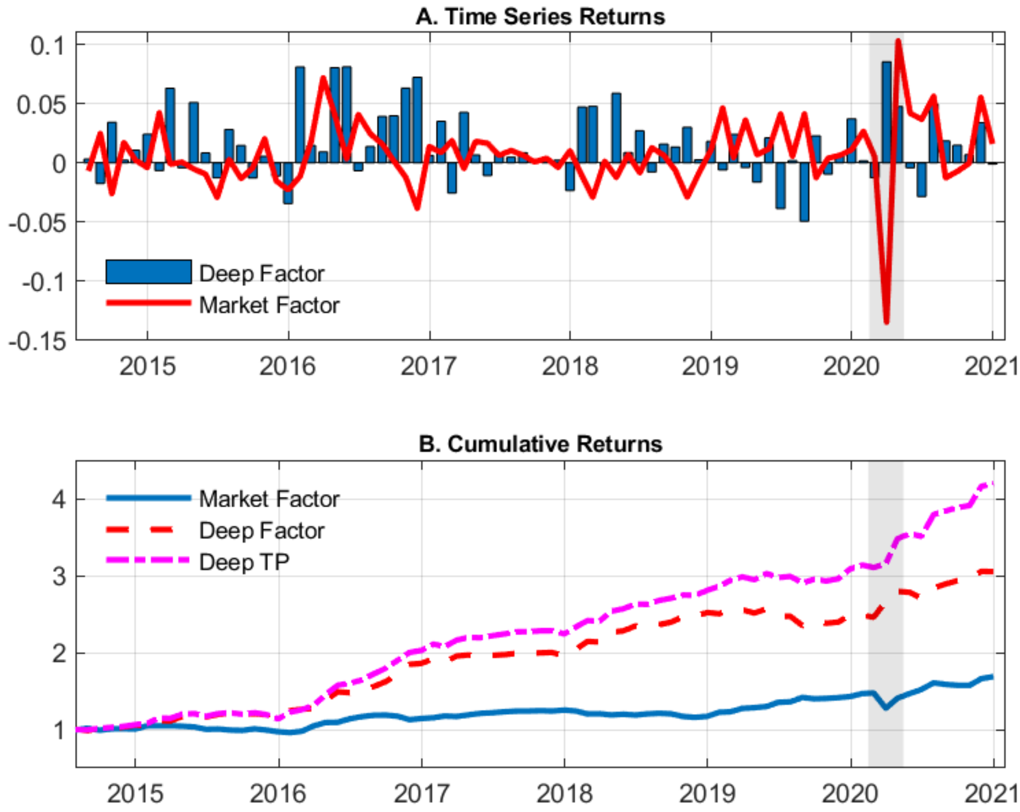


To further verify this point, Panel A of Figure 3 presents the time series of excess returns on the equal-weight market factor and the deep factor over the OOS period (both normalized to have an annualized volatility of 10%). We see that the deep factor remains positive during periods of market downturns. More importantly, we note that during the outbreak of the COVID-19 pandemic, the market factor drops dramatically, whereas our deep factor generates a high positive return. It turns out that our deep factor plays the role of a market-hedge portfolio. Panel B of Figure 3 displays cumulative returns of the equal-weight market factor, the deep factor, and the deep tangency portfolio for the OOS period, respectively. Both the deep factor and the deep tangency portfolio perform much better than the market factor. We see that during the outbreak of the Covid-19 pandemic, the cumulative returns of the deep factor and the deep tangency portfolio continue to increase, whereas the market factor suffers a great loss, which is consistent with the findings in Panel A.

Panel B of Table 3 shows that the optimal portfolios constructed from competing factors perform much worse than the deep tangency portfolio for the same OOS pe-

Figure 3: **Time Series and Cumulative Returns**

Panel A presents the times series returns of the market factor and the deep factor for the out-of-sample period, and Panel B presents the cumulative returns of the market factor, the deep factor, and the deep tangency portfolio for the out-of-sample period. All portfolios are normalized to have 10% annualized volatility. The deep factor and the deep tangency portfolio are constructed using the equal-weight market factor as the benchmark and the softmax ranking. The out-of-sample period ranges from July 2014 to December 2020.



riod. As before, the portfolio weights are determined by the in-sample estimates. The optimal portfolio constructed from the Fama-French five factors has a Sharpe ratio of 0.78, much smaller than our deep tangency portfolio. As expected, its maximum drawdown also becomes lower. Note that the Fama-French five factors contain two equity factors (SMB and HML). Then what happens when we combine the Fama-French factors and our deep factor? We see that the Sharpe ratio of the optimal portfolio spanned by the Fama-French five factors and our deep factor (taking equal-weight market factor as the benchmark) increases to 2.29, with the maximum drawdown of 23.1%, both nearly the same as those of our deep tangency portfolios. Given that our deep factor is constructed by taking the bond market factor as a benchmark and using all firm characteristics, it should already contain non-market information of the Fama-French factors; therefore, including those factors should not improve the Sharpe ratio over our deep tangency portfolio.

A recent paper by Kelly, Palhares, and Pruitt (2022) shows that a five-factor model based on the instrumental principle component analysis (IPCA, Kelly, Pruitt, and Su, 2019) outperforms commonly used observable factor models in pricing corporate bonds. They find that a tangency portfolio constructed from their five IPCA factors using the ICE corporate bond return data can earn an annualized OOS Sharpe ratio of as large as 6.23. We note that in another paper, Kelly and Pruitt (2022) shows that the core analysis of Kelly et al. (2022) is robust to using the TRACE data. We follow their IPCA approach and construct five corporate bond factors using our TRACE data and all three types of characteristics. To be consistent with our primary empirical analysis, we use the same in-sample, and OOS split as before and extract the OOS IPCA factors using in-sample model parameter estimates and OOS characteristics.

Even though both Kelly, Palhares, and Pruitt (2022) and Kelly and Pruitt (2022) find that an optimal portfolio constructed using the five IPCA factors can earn an OOS Sharpe ratio of larger than 6 in using both ICE and TRACE corporate bond data, we find that such a portfolio can only earn an OOS Sharpe ratio of 1.66, which is much smaller than that of our deep tangency portfolio (Panel B of Table 3). There are two reasons why we find such a weaker OOS Sharpe ratio. First, the sample size in our paper is much larger than that in Kelly and Pruitt (2022): the total number of bond-month observations in our paper is 633,600, whereas it is only 144,933 in Kelly and Pruitt (2022). Second, both Kelly, Palhares, and Pruitt (2022) and Kelly and Pruitt (2022) adopt an expanding window procedure to construct the OOS IPCA factors, whereas we extract OOS IPCA factors by fixing model parameters at the in-sample estimates to make it comparable with our methodology.[10] We further find that combining the deep factor with the IPCA five factors improves the OOS Sharpe ratio of the optimal portfolio to 2.22, similar to that of our deep tangency portfolio, with a slightly higher maximum drawdown of 25.2%.

In addition, Lettau and Pelger (2020) proposes a risk-premium principal component analysis (RP-PCA) model for estimating latent asset pricing factors. They show that the RP-PCA performs much better than the PCA method, particularly in identifying the weak factors. We also examine how the five RP-PCA factors perform compared to our deep tangency portfolio. We construct the RP-PCA factors relying on 48 tercile portfolios based on 16 bond characteristics as discussed in Subsection 3.3. Again, we find that the OOS Sharpe ratio of the RP-PCA tangency portfolio is much smaller than that of the deep tangency portfolio (1.67 vs. 2.29). When we combine the RP-PCA five factors with our deep factor, we find that the Sharpe ratio of the resulted optimal portfolio is about 2.38, which is slightly larger than that of our deep tangency portfolio, but this Sharpe ratio improvement is not statistically significant.

Although we have just used one deep factor in our previous analysis, our methodology is flexible enough to introduce multiple deep factors if necessary. This can be done by simply iterating the algorithm by taking the deep factor extracted as another benchmark. Table 4 presents the performance of the softmax ranking-based deep tangency portfolios constructed from the market factor and 1-3 deep factors. Both in-sample training and OOS testing indicate that the first deep factor extracted from the one-layer neural network performs well. The improvement in the Sharpe ratio from using 2 or 3 deep factors over the tangency portfolio with one deep factor is negligible and statistically insignificant, regardless of whether the equal-weight or value-weight market factor is used as the benchmark. Therefore, we focus on the first deep factor

[10] We also implement a recursive expanding-window approach similar to Kelly, Palhares, and Pruitt (2022) and Kelly and Pruitt (2022) using our TRACE data and find an almost identical OOS Sharpe ratio of the IPCA optimal portfolio. The results show a mean of 0.39 and a standard deviation of 0.81, resulting in a Sharpe ratio of 1.67.

and the corresponding deep tangency portfolio from the 1-layer neural network.

Table 4: Multiple Deep Factors

This table presents Sharpe ratios of the deep tangency portfolios constructed using multiple deep factors. L_i represents a neural network with i layers, and D_i represents i deep factors included in the model. We sequentially add one additional deep factor for each choice of the number of neural network layers. The test of the Sharpe ratio improvement by including an additional deep factor is based on Barillas and Shanken (2017). $***$, $**$, and $*$ denote the level of significance of 1%, 5%, and 10%, respectively. We take the sample from July 2004 to June 2014 for model training and validating and the sample from July 2014 to December 2020 for out-of-sample testing.

	MKT	D_1	D_2	D_3
		Panel A. Equal-Weight Market Factor		
L_1	0.86	2.29***	2.31	2.30
L_2	0.86	1.48***	1.48	1.46
L_3	0.86	1.13***	0.79	0.79
		Panel B. Value-Weight Market Factor		
L_1	0.92	2.12***	2.12	1.94
L_2	0.92	1.50***	1.50	1.50
L_3	0.92	1.55***	1.53	1.48

4.3 How Important Are the Economic Loss and Nonlinearity?

The deep factor in our deep parametric portfolio approach is formed on a deep characteristic – a highly nonlinear combination of raw characteristics based on an economically motivated target. A key question is how important this economic target is compared to the commonly used statistical target.

For this purpose, following the literature (see, e.g., Gu, Kelly, and Xiu, 2020), we construct long-short portfolios based on return forecasts obtained from various ML methods. In addition to neural networks with different layers, which serve as natural choices for comparison with our deep parametric portfolio approach, we also consider Lasso, Ridge, and PCA in our analysis. To be specific, at each time t, we form the return forecasts using all characteristics relying on these ML models as follows,

$$W_t \equiv E_t[r_{t+1}] = g(z_t, b), \tag{22}$$

where $g(\cdot)$ is the function to be learned by these ML methods, and b collects all model

Table 5: ML Long-Short Portfolios and Linear Activation

Panel A reports annualized Sharpe ratios of the long-short portfolios based on return forecasts from the neural networks with 1-3 layers (NN1, NN2, and NN3), Lasso, Ridge, and PCA. Panel B presents annualized Sharpe ratios of the softmax ranking-based deep tangency portfolios constructed using the equal-weight market factor as the benchmark and replacing the nonlinear activation function with a linear one in the deep learning model. L_i represents a neural network with i layers, and D_i represents i deep factors included in the model. We take the sample from July 2004 to June 2014 for model training and validating, and from July 2014 to December 2020 for out-of-sample testing.

Panel A. ML Long-Short Portfolios					
NN1	NN2	NN3	LASSO	Ridge	PCA
0.74	0.83	0.86	0.81	1.13	0.64
Panel B. Linear Activation					
L_1/D_1	L_1/D_2	L_1/D_3	L_2/D_1	L_2/D_2	L_2/D_3
0.31	0.31	0.30	0.56	0.56	0.57

parameters. We then construct an equal-weight long-short portfolio using the standard sorting approach that longs top 30% and shorts bottom 30% corporate bonds based on W_t. As before, we use the sample from July 2004 to June 2014 for model training and validation and the sample from July 2014 to December 2020 for OOS tests. See Section V in the Internet Appendix for implementation details.

Panel A of Table 5 presents the Sharpe ratios of the ML portfolios for the OOS period. We see that none of them achieves a Sharpe ratio that is larger than that of our deep tangency portfolios. In particular, even though our model also employs the neural network, the long-short portfolios based on neural network (NN) return forecasts perform much worse than our deep tangency portfolios. The portfolio Sharpe ratios of NN1, NN2, and NN3 are 0.74, 0.83, and 0.86, respectively. The portfolio based on Ridge return forecasts performs well, delivering a Sharpe ratio larger than 1, which remains much smaller than that of our deep tangency portfolio.

We also examine the impact of replacing the nonlinear *tanh* activation function and use a linear combination of characteristics in the deep learning model. We see from Panel B of Table 5 that a deeper neural network and more deep factors are needed to complement the benchmark factor without the nonlinear activation function. However, more importantly, the OOS performance of the deep tangency portfolio now performs much worse compared to the case with the nonlinear activation.

To sum up, the economically motivated loss function and the nonlinear activation in our deep parametric portfolio approach play important roles in constructing the deep tangency portfolio. Such findings are largely consistent with the literature on the importance of economic restrictions (see, e.g., Jensen, Kelly, Malamud, and Pedersen, 2024; Avramov, Cheng, and Metzker, 2023; Liu, Zhou, and Zhu, 2024) and of nonlinear effects of characteristics in portfolio construction (see, e.g., Freyberger, Neuhierl, and Weber, 2020; Gu, Kelly, and Xiu, 2020).

4.4 Factor-Spanning Regressions

We find that our deep factor, derived from a nonlinear combination of firm characteristics, plays a role as a market-hedge portfolio. Moreover, commonly used observable and latent factors do not provide additional pricing information for improving the Sharpe ratio when combined with the deep factor. In this part, we further examine whether these commonly used observable and latent factors can explain excess returns on our deep factor and deep tangency portfolio by implementing factor-spanning regressions of the form,

$$R_t = \alpha + \beta' f_t + \epsilon_t, \tag{23}$$

where f_t is a set of observable or latent factors, and R_t represents returns on either the deep factor or the deep tangency portfolio. We take the equal-weight bond market factor as the benchmark and employ the 1-layer neural network with the softmax ranking in our deep parametric portfolio approach.

Table 6 presents the factor-spanning regression results for the OOS period. Panel A takes the Fama-French five factors discussed in Subsection 3.3 as competing factors. It is observed that the Fama-French factors cannot explain excess returns on both the

Table 6: Factor-Spanning Regressions

This table reports factor-spanning regression results for the out-of-sample period. Specifically, we regress the deep factor and the deep tangency portfolios on the Fama-French five factors, the IPCA five factors, and the RP-PCA five factors, respectively. The deep factor and the deep tangency portfolio are constructed using the equal-weight market factor as the benchmark and the softmax ranking. The alpha estimates are in percentage, and the Newey-West t-statistics (Newey and West, 1987) are reported in brackets. The out-of-sample period ranges from July 2014 to December 2020.

Panel A. FF Five Factors

	α	β_{MKT}	β_{TRM}	β_{DEF}	β_{SMB}	β_{HML}	Adj R^2
R_d^1	0.21	0.06	−0.07	−0.12	−0.01	0.01	16.5
	(5.22)	(1.02)	(−2.16)	(−1.75)	(−1.05)	(0.83)	
R_1^{opt}	0.19	0.15	−0.07	−0.11	−0.01	0.01	12.6
	(5.22)	(2.86)	(−2.16)	(−1.75)	(−1.05)	(0.83)	

Panel B. IPCA Five Factors

	α	β_1	β_2	β_3	β_4	β_5	Adj R^2
R_d^1	0.17	0.03	−0.04	−0.08	0.03	0.14	35.4
	(4.51)	(0.47)	(−1.90)	(−1.88)	(0.80)	(4.36)	
R_1^{opt}	0.16	0.05	0.01	−0.05	0.04	0.09	33.6
	(4.44)	(0.85)	(0.37)	(−1.37)	(1.08)	(3.14)	

Panel C. RP-PCA Five Factors

	α	β_1	β_2	β_3	β_4	β_5	Adj R^2
R_d^1	0.21	0.00	−0.08	−0.08	0.09	−0.01	27.4
	(5.19)	(0.46)	(−2.15)	(−0.81)	(1.33)	(−0.07)	
R_1^{opt}	0.19	0.02	−0.07	−0.07	0.10	−0.01	24.5
	(4.93)	(2.05)	(−2.13)	(−0.78)	(1.60)	(−0.08)	

deep factor and the deep tangency portfolio. The alpha estimate is 0.21% in the regression of the deep factor and is about 0.19% in the regression of the deep tangency portfolio. Both alpha estimates are highly statistically significant. The deep factor and the deep tangency portfolio load negatively on the term and default factors. However, while the loadings on the term factor are statistically significant, the loadings on the default factor are only marginally significant. The loadings on both equity factors are not statistically significant. The adjusted R^2s are only about 16.5% and 12.6%, respectively, in these two spanning regressions.

Given that the IPCA factors are also estimated by taking into account all firm characteristics (in a linear form), and that Kelly, Palhares, and Pruitt (2022) and Kelly and Pruitt (2022) show that the IPCA factors extremely outperform popular observable factors, we examine whether they can span our deep factor and deep tangency portfolio. The spanning regression results in Panel B show that the five IPCA factors cannot explain excess returns on both the deep factor and the deep tangency portfolio, as the alpha estimates are about 0.17% and 0.16%, respectively, which are highly statistically significant in both regressions, and there is only one IPCA factor that seems important. The adjusted R^2s are about 35% and 34%, respectively, improving a little compared to that using the Fama-French factor.

Similarly, we further see in Panel C that the RP-PCA five factors are also unable to explain excess returns on the deep factor and the deep tangency portfolio. The alpha estimates are about 0.21% (t = 5.19) and 0.19% (t = 4.93) and the adjusted R^2s are 27% and 25%, respectively, in the two spanning regressions. This marginal improvement of our deep tangency portfolio over two PCA-based approaches could be linked to the nonlinear characteristics-return modeling of deep learning, which is not straightforward for PCA methods.

4.5 Interpreting Deep Characteristics

By combining an economically motivated loss function with deep learning and constructing the deep factor as long-short portfolio returns, we aim to improve the transparency and interpretability of our methodology. Therefore, a natural next step is to understand how different characteristics contribute to the deep factor.

The nonlinear activation function in our deep parametric portfolio approach transforms firm characteristics into a deep characteristic, a highly nonlinear combination of raw characteristics whose exact functional form is unknown to us. To examine variable importance, we rely on average absolute gradients. Following Chen, Pelger, and

Zhu (2024), we assess the significance of a firm characteristic by analyzing its sensitivity to the deep characteristic, $Z_t^{(L)}$ (see Section 2.2). To be specific, the sensitivity of a given variable (z_j) is defined as the average absolute derivative of $Z_t^{(L)}$ with respect to that variable, given by

$$\text{Sensitivity}(z_j) = \frac{1}{C}\sum_{i=1}^{N}\sum_{t=1}^{T}\left|\frac{\partial Z_{i,t}^{(L)}}{\partial z_j}\right|, \tag{24}$$

for $j = 1, \ldots, K$, where C is a normalization constant.

Figure 4 presents average absolute gradients of all characteristics, with bond, equity, and option characteristics classified by the blue, yellow, and red bars, respectively. Two observations are in order. First, the deep characteristic is not dominated by a small particular set of characteristics, and most characteristics contribute to the deep characteristic, suggesting that there isn't cut-off evidence of the sparsity of characteristics. Second, the top 10 most important characteristics are all bond-related variables. The most important equity characteristic is corporate investment (CINVEST), which ranks 12th, while the most important option variable is the put-call ratio (PCRATIO), which ranks 14th.

To further examine the importance of different types of characteristics, we reconstruct the softmax ranking-based deep tangency portfolios using equity and bond characteristics with option-related variables removed, or using bond characteristics alone. Table 7 presents the Sharpe ratios of the deep tangency portfolios obtained from different types of characteristics by taking the equal-weight market factor as the benchmark. When using all characteristics, the one-layer neural network performs well, and the deep tangency portfolio earns an OOS annualized Sharpe ratio of 2.29. However, excluding the option-related variables results in a smaller OOS Sharpe ratio of 1.83 for the deep tangency portfolio, despite still being the highest value from the 1-layer neural network. This suggests that option-related variables contain valuable

Figure 4: Variable Importance: Average Absolute Gradients

This figure presents the average absolute gradients of all characteristics using Equation (24) for the out-of-sample period. The blue bars stand for bond characteristics, the yellow for equity characteristics, and the red for option characteristics. The OOS period ranges from July 2014 to December 2020.

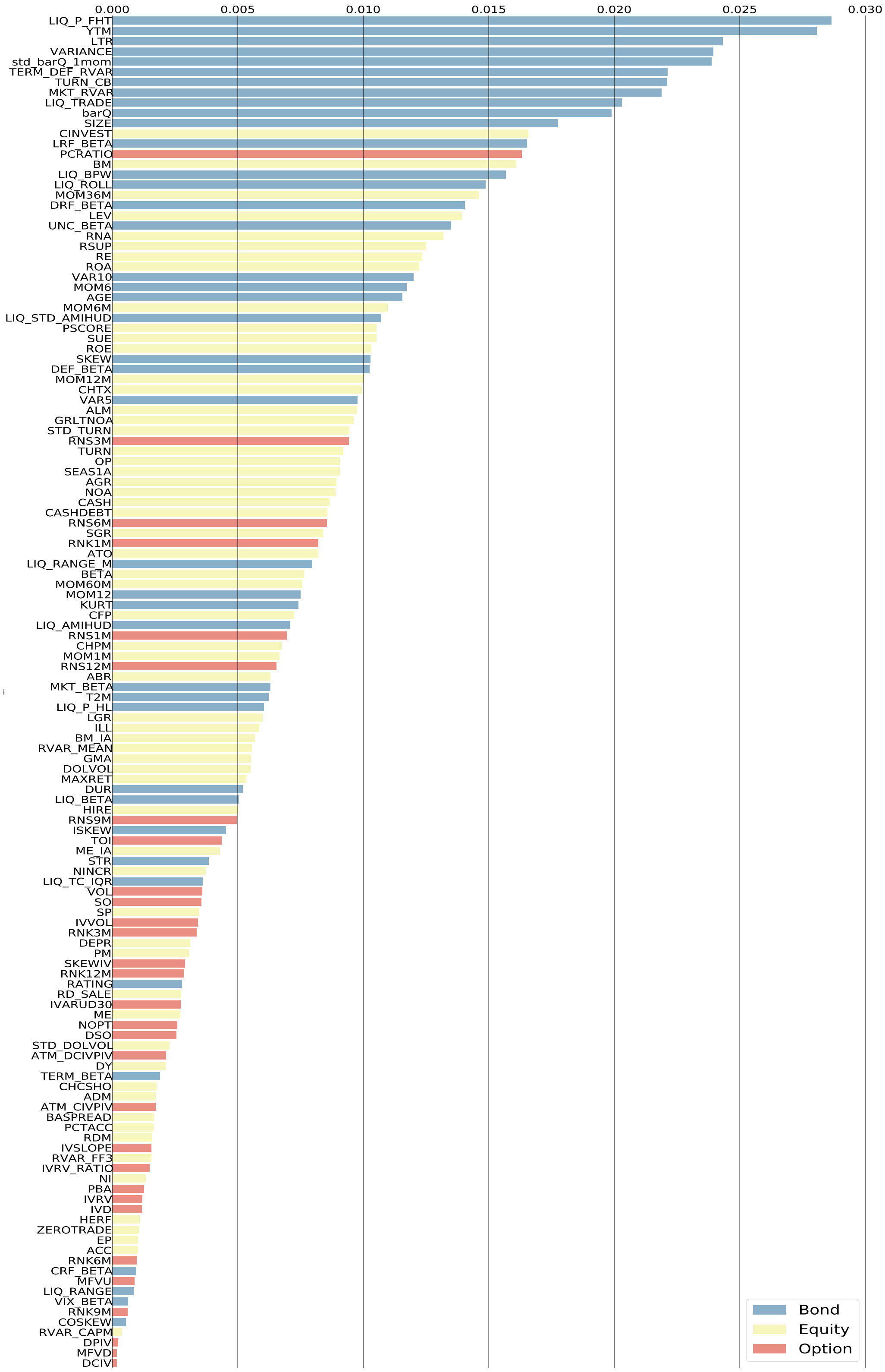


Table 7: Importance of Characteristics

This table presents annualized Sharpe ratios obtained from different types of characteristics. We consider three sets of characteristics: all 132 characteristics, 102 bond and equity characteristics, and 41 bond characteristics. The test of the Sharpe ratio improvement of the deep tangency portfolio over the market factor is based on Barillas and Shanken (2017). $***$, $**$, and $*$ denote the level of significance of 1%, 5%, and 10%, respectively. L_i represents a neural network with i layers. We use the equal-weight market factor as the benchmark and employ the softmax ranking. The sample from July 2004 to June 2014 is for model training and validating, and the sample from July 2014 to December 2020 is for out-of-sample testing.

	MKT	Bond+Equity+Option	Bond+Equity	Bond
L_1	0.86	2.29***	1.83***	0.71
L_2	0.86	1.48***	0.49	0.53
L_3	0.86	1.13***	1.70***	0.61

information on future corporate bond returns. What is even worse is that when we use bond characteristics alone, the performance of the deep tangency portfolio further deteriorates. Its best OOS Sharpe ratio is only about 0.71 from the 1-layer neural network, which is smaller than the market portfolio.

To sum up, all three types of characteristics are important and heavily weighted in the deep characteristic. Therefore, they are necessary for constructing the deep tangency portfolio. This finding is, in fact, in stark contrast to previous studies that argue that those characteristics that predict equity returns do not necessarily forecast corporate bond returns (see, e.g., Chordia et al., 2017; Bali et al., 2021). However, it is in line with Dickerson, Mueller, and Robotti (2023) who find that equity characteristics provide incremental value for bond return predictability and provide further empirical evidence in support of the integration between the bond and equity markets (Schaefer and Strebulaev, 2008; Kelly, Palhares, and Pruitt, 2022).

5 Conclusion

The tangency portfolio holds both economic and practical importance. However, the mean-variance efficient portfolio of Markowitz (1952) is hardly implementable with a large number of individual assets. Although asset pricing literature often relies on factors based on a few firm characteristics, these commonly used factors cannot fully span the efficient frontier, leading to the "factor zoo" issue and the curse of high dimensionality. The literature has not found clear-cut evidence of the sparsity of characteristics (Kozak, Nagel, and Santosh, 2020), and Kozak and Nagel (2024) show that a large number of characteristics are necessary to span the efficient frontier.

This paper proposes a parametric approach to directly estimating the tangency portfolio weights, thereby bypassing the need to estimate expected returns and the covariance matrix, using deep learning models. The deep learning model estimates a deep factor composed of thousands of individual assets and then combines it with the market factor. The endogenous deep factor construction mimics the widely used characteristic-sorted factor approach and employs the nonlinear transformation of high-dimensional firm characteristics. The economically guided deep factor plays two important roles: (i) under the maximum Sharpe ratio objective of the tangency portfolio, the deep factor has a low or even negative correlation with benchmark factors, providing a potential market hedge portfolio, and (ii) the deep factor is constructed using information from high-dimensional characteristics, potentially spanning any missing risk factors that should be included in the optimal portfolio construction.

We apply our method to the corporate bond market, where the understanding of pricing factors remains limited. Our deep tangency portfolio achieves an annualized Sharpe ratio of 2.29, outperforming portfolios spanned by commonly used observable or latent factors, such as the Fama-French five factors, the instrumental PCA factors, and the Risk-Premium PCA factors. Additionally, we emphasize the importance of incorporating various types of characteristics in constructing the deep tangency portfolio. Excluding any type of characteristics (equity, bond, or options) deteriorates the performance of the deep tangency portfolio. From the tangency portfolio perspective that considers both return and risk, our findings starkly contrast with previous studies suggesting that characteristics predicting equity returns may not necessarily forecast corporate bond returns (e.g., Chordia et al., 2017). Instead, our results provide evidence in support of the integration between the bond and equity markets (e.g., Schaefer and Strebulaev, 2008; Kelly, Palhares, and Pruitt, 2022).

References

Ait-Sahalia, Y. and M. W. Brandt (2001). Variable selection for portfolio choice. *Journal of Finance 56*, 1297–1351.

Ao, M., Y. Li, and X. Zheng (2019). Approaching mean-variance efficiency for large portfolios. *The Review of Financial Studies 32*(7), 2890–2919.

Avramov, D., S. Cheng, and L. Metzker (2023). Machine learning vs. economic restrictions: Evidence from stock return predictability. *Management Science 69*(5), 2587–2619.

Bai, J., T. G. Bali, and Q. Wen (2019). Common risk factors in the cross-section of corporate bond returns. *Journal of Financial Economics 131*(3), 619–642.

Bali, T. G., A. Goyal, D. Huang, F. Jiang, and Q. Wen (2021). The cross-sectional pricing of corporate bonds using big data and machine learning. Technical report, Georgetown University.

Bali, T. G., A. Subrahmanyam, and Q. Wen (2021). Long-term reversals in the corporate bond market. *Journal of Financial Economics 139*(2), 656–677.

Barillas, F., R. Kan, C. Robotti, and J. Shanken (2020). Model comparison with sharpe ratios. *Journal of Financial and Quantitative Analysis 55*(6), 1840–1874.

Barillas, F. and J. Shanken (2017). Which alpha? *Review of Financial Studies 30*(4), 1316–1338.

Barillas, F. and J. Shanken (2018). Comparing asset pricing models. *Journal of Finance 73*, 715–754.

Bessembinder, H., K. Kahle, W. Maxwell, and D. Xu (2009). Measuring abnormal bond performance. *Review of Financial Studies 22*, 4219–4258.

Brandt, M. W. (1999). Estimating portfolio and consumption choice: A conditional euler equations approach. *Journal of Finance 54*, 1609–1646.

Brandt, M. W. and P. Santa-Clara (2006). Dynamic portfolio selection by augmenting the asset space. *Journal of Finance 61*(5), 2187–2217.

Brandt, M. W., P. Santa-Clara, and R. Valkanov (2009). Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. *Review of Financial Studies 22*, 3411–3447.

Bryzgalova, S., M. Pelger, and J. Zhu (2024). Forest through the trees: Building cross-sections of stock returns. *Journal of Finance, Forthcoming*.

Cao, J., A. Goyal, X. Xiao, and X. Zhan (2022). Implied volatility changes and corporate bond returns. *Management Science, Forthcoming*.

Chen, L., M. Pelger, and J. Zhu (2024). Deep learning in asset pricing. *Management Science 55*, 714–750.

Choi, J. and Y. Kim (2018). Anomalies and market (dis)integration. *Journal of Monetary Economics 100*, 16–34.

Chordia, T., A. Goyal, Y. Nozawa, A. Subrahmanyam, and Q. Tong (2017). Are capital market anomalies common to equity and corporate bond markets? an empirical investigation. *Journal of Financial and Quantitative Analysis 52*(4), 1301–1342.

Chung, K. H., J. Wang, and C. Wu (2019). Volatility and the cross-section of corporate bond returns. *Journal of Financial Economics 133*(2), 397–417.

Cochrane, J. H. (2011). Presidential address: Discount rates. *Journal of Finance 66*(4), 1047–1108.

Cochrane, J. H. (2014). A mean-variance benchmark for intertemporal portfolio theory. *Journal of Finance 69*, 1–49.

Cong, L. W., G. Feng, J. He, and X. He (2024). Growing the efficient frontier on panel trees. *Journal of Financial Economics*.

Cong, L. W., K. Tang, J. Wang, and Y. Zhang (2022). Alphaportfolio: Direct construction through deep reinforcement learning and interpretable ai. Technical report, Cornell University.

Daniel, K., L. Mota, S. Rottke, and T. Santos (2020). The cross-section of risks and returns. *Review of Financial Studies 33*, 1927–1979.

DeMiguel, V., L. Garlappi, and R. Uppal (2009). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *Review of Financial Studies 22*(5), 1915–1953.

Dick-Nielsen, J. (2009). Liquidity biases in trace. *Journal of Fixed Income 19*(2), 43–55.

Dick-Nielsen, J. (2014). How to clean enhanced trace data. Technical report, Copenhagen Business School.

Dickerson, A., P. Mueller, and C. Robotti (2023). Priced risk in corporate bonds. *Journal of Financial Economics 150*, 103707.

Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics 33*(1), 3–56.

Fama, E. F. and K. R. French (1996). Multifactor explanations of asset pricing anomalies. *Journal of Finance 51*(1), 55–84.

Fama, E. F. and K. R. French (2015). A five-factor asset pricing model. *Journal of Financial Economics 116*(1), 1–22.

Fama, E. F. and K. R. French (2018). Choosing factors. *Journal of Financial Economics 128*(2), 234–252.

Feng, G., J. He, N. Polson, and J. Xu (2023). Deep Learning of Characteristics-Sorted Factor Models. *Journal of Financial and Quantitative Analysis, Forthcoming*.

Feng, G., X. He, Y. Wang, and C. Wu (2021). Predicting individual corporate bond returns. Technical report, City University of Hong Kong.

Freyberger, J., A. Neuhierl, and M. Weber (2020). Dissecting characteristics nonparametrically. *Review of Financial Studies 33*, 2326–2377.

Giannone, D., M. Lenza, and G. E. Primiceri (2021). Economic predictions with big data: The illusion of sparsity. *Econometrica 89*, 2409–2437.

Gu, S., B. Kelly, and D. Xiu (2020). Empirical asset pricing via machine learning. *Review of Financial Studies 33*, 2223–2273.

Guijarro-Ordonez, J., M. Pelger, and G. Zanotti (2024). Deep learning statistical arbitrage. *Management Science Forthcoming*.

Harvey, C. R., Y. Liu, and H. Zhu (2016). ... and the cross-section of expected returns. *Review of Financial Studies 29*(1), 5–68.

Hou, K., C. Xue, and L. Zhang (2020). Replicating anomalies. *Review of Financial Studies 33*(5), 2019–2133.

Huang, T., L. Jiang, and J. Li (2023). Downside variance premium, firm fundamentals, and expected corporate bond returns. *Journal of Banking and Finance, Forthcoming*.

Jensen, T. I., B. T. Kelly, S. Malamud, and L. H. Pedersen (2024). Machine learning and the implementable efficient frontier. *Review of Financial Studies Forthcoming*.

Jostova, G., S. Nikolova, A. Philipov, and C. W. Stahel (2013). Momentum in corporate bond returns. *Review of Financial Studies 26*(7), 1649–1693.

Kelly, B. and D. Xiu (2023). Financial machine learning. *Foundations and Trends® in Finance 13*, 205–363.

Kelly, B. T., D. Palhares, and S. Pruitt (2022). Modeling corporate bond returns. *Journal of Finance, Forthcoming*.

Kelly, B. T. and S. Pruitt (2022). Reconciling trace bond returns. Technical report, Yale University.

Kelly, B. T., S. Pruitt, and Y. Su (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics 134*(3), 501–524.

Kozak, S. and S. Nagel (2024). When do cross-sectional asset pricing factors span the stochastic discount factor? Technical report, University of Maryland.

Kozak, S., S. Nagel, and S. Santosh (2018). Interpreting factor models. *Journal of Finance 73*(3), 1183–1223.

Kozak, S., S. Nagel, and S. Santosh (2020). Shrinking the cross-section. *Journal of Financial Economics 135*(2), 271–292.

Lettau, M. and M. Pelger (2020). Factors that fit the time series and cross-section of stock returns. *Review of Financial Studies 33*, 2274–2325.

Lin, H., J. Wang, and C. Wu (2011). Liquidity risk and expected corporate bond returns. *Journal of Financial Economics 99*, 628–650.

Liu, Y., G. Zhou, and Y. Zhu (2024). Maximizing the sharpe ratio: A genetic programming approach. Technical report, Tsinghua Univeristy.

Markowitz, H. (1952). Portfolio selection. *Journal of Finance 7*, 77–99.

Merton, R. C. (1973). An intertemporal capital asset pricing model. *Econometrica 41*, 867–887.

Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics 8*, 323–361.

Negal, S. (2021). *Machine Learning in Asset Pricing*. Princeton University Press.

Neuhierl, A., X. Tang, R. Varneskov, and G. Zhou (2021). Option characteristics as cross-sectional predictors. Technical report, Washington University in St. Louis.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*, 703–708.

Schaefer, S. M. and I. Strebulaev (2008). Structural models of credit risk are useful: Evidence from hedge ratios on corporate bonds. *Journal of Financial Economics 90*, 1–19.

Simon, F., S. Weibels, and T. Zimmermann (2023). Deep parametric portfolio policies. Technical report, University of Cologne.

Internet Appendix for

"Deep Tangency Portfolios"
(not for publication)

Summary of Contents

- Section I presents detailed definitions of bond, equity, and option characteristics used in our empirical studies.
- Section II presents the implementation details of our deep learning algorithm and some robustness checks.
- Section III presents the results from two alternative sample split schemes.
- Section IV presents the performance of the deep tangency portfolio accounting for transaction costs.
- Section V presents the implementation details of machine learning (ML) algorithms.

I Bond, Equity, and Option Characteristics

Table A1: Description of 41 Bond Characteristics

Characteristics	Description
AGE	Time since issuance in years
RATING	Bond credit rating
T2M	The number of years to maturity
SIZE	Amount outstanding
DUR	Bond duration
VAR5	Value-at-risk 5% over past 3 years
VAR10	Value-at-risk 10% over past 3 years
LIQ_BPW	Liquidity measure of transitory price movements.
LIQ_ROLL	Roll's liquidity
LIQ_P_HL	High-low spread estimator
LIQ_P_FHT	Modified illiquidity measure based on zero returns
LIQ_AMIHUD	Amihud liquidity
LIQ_STD_AMIHUD	Standard deviation of Amihud daily liquidity
LIQ_TC_IQR	Interquartile range
MKT_BETA	Market beta
DEF_BETA	DEF factor beta
TERM_BETA	TERM factor beta
LIQ_BETA	Liquidity beta of bond illiquidity factor
DRF_BETA	Downside risk beta controlling bond market factor
CRF_BETA	Credit risk beta controlling bond market factor
LRF_BETA	Liquidity risk beta controlling bond market factor
VIX_BETA	VIX index beta
UNC_BETA	Macroeconomic Uncertainty Beta
STR	Short-term reversal t-1
VARIANCE	Variance of raw returns
SKEW	Skewness of raw returns
KURT	Kurtosis of raw returns
COSKEW	Systematic skewness with bond market
ISKEW	Idiosyncratic skewness
LIQ_RANGE	Simple high-low spread
LIQ_TRADE	Number of trades
MKT_RVAR	Market residual variance
TERM_DEF_RVAR	TERM DEF residual variance
TURN_CB	Bond Turnover
YTM	Yield-to-maturity
MOM6	Momentum from t-2 to t-6
MOM12	Momentum from t-7 to t-12
LTR	Long-term reversal from t-13 to t-48
barQ	average daily dollar volume in the 1-month period
std_barQ_1mom	standard deviation of dollar volume in the 1-month period
LIQ_RANGE_M	Simple high-low spread

Table A2: Description of 61 Equity Characteristics

Characteristics	Description
ABR	Abnormal returns around earnings announcement
ACC	Operating Accruals
ADM	Advertising Expense-to-market
AGR	Asset growth
ALM	Quarterly Asset Liquidity
ATO	Asset Turnover
BASPREAD	Bid-ask spread (3 months)
BETA	Beta (3 months)
BM	Book-to-market equity
BM_IA	Industry-adjusted book to market
CASH	Cash holdings
CASHDEBT	Cash to debt
CFP	Cashflow-to-price
CHCSHO	Change in shares outstanding
CHPM	Industry-adjusted change in profit margin
CHTX	Change in tax expense
CINVEST	Corporate investment
DEPR	Depreciation / PP&E
DOLVOL	Dollar trading volume
DY	Dividend yield
EP	Earnings-to-price
GMA	Gross profitability
GRLTNOA	Growth in long-term net operating assets
HERF	Industry sales concentration
HIRE	Employee growth rate
ILL	Illiquidity rolling (3 months)
LEV	Leverage
LGR	Growth in long-term debt
MAXRET	Maximum daily returns (3 months)
ME	Market equity
ME_IA	Industry-adjusted size
MOM12M	Cumulative Returns in the past (2-12) months
MOM1M	Previous month return
MOM36M	Cumulative Returns in the past (13-35) months
MOM60M	Cumulative Returns in the past (13-60) months
MOM6M	Cumulative Returns in the past (2-6) months
NI	Net Equity Issue
NINCR	Number of earnings increases
NOA	Net Operating Assets
OP	Operating profitability
PCTACC	Percent operating accruals
PM	profit margin
PS	Performance Score
RD_SALE	R&D to sales
RDM	R&D Expense-to-market
RE	Revisions in analysts' earnings forecasts
RNA	Return on Net Operating Assets
ROA	Return on Assets
ROE	Return on Equity
RSUP	Revenue surprise
RVAR_CAPM	Residual variance - CAPM (3 months)
RVAR_FF3	Res. var. - Fama-French 3 factors (3 months)
RVAR_MEAN	Return variance (3 months)

Description of 61 Equity Characteristics (continued)

Characteristics	Description
SEAS1A	1-Year Seasonality
SGR	Sales growth
SP	Sales-to-price
STD_DOLVOL	Std of dollar trading volume (3 months)
STD_TURN	Std. of Share turnover (3 months)
SUE	Unexpected quarterly earnings
TURN	Shares turnover
ZEROTRADE	Number of zero-trading days (3 months)

Table A3: Description of 30 Equity Option Characteristics

Characteristics	Description
IVSLOPE	Implied Volatility Slope
IVVOL	Volatility of atm implied volatility
IVRV	Implied and historical volatility spread
IVRV_RATIO	Ratio of implied to historical volatility
ATM_CIVPIV	Implied volatility spread
SKEWIV	Implied volatility skew
IVD	Implied volatility duration
DCIV	Change of implied volatility of atm call
DPIV	Change of implied volatility of atm put
ATM-DCIVPIV	Change of implied volatility spread
NOPT	Number of traded options
SO	Stock-option volume ratio
DSO	Stock-option dollar volume ratio
VOL	Option Trading Volume
PCRATIO	Put-call ratio
PBA	Proportional bid-ask spread
TOI	Total open interest
MFVU	Option-implied upside semivariance
MFVD	Option-implied downside semivariance
RNS1M	1-month risk-neutral skewness
RNK1M	1-month risk-neutral kurtosis
IVARUD30	Option-implied variance asymmetry
RNS3M	3-month risk-neutral skewness
RNK3M	3-month risk-neutral kurtosis
RNS6M	6-month risk-neutral skewness
RNK6M	6-month risk-neutral kurtosis
RNS9M	9-month risk-neutral skewness
RNK9M	9-month risk-neutral kurtosis
RNS12M	12-month risk-neutral skewness
RNK12M	12-month risk-neutral kurtosis

II Implementation Details

II.1 Data Clearning

As shown in Table 1, we provide the summary statistics for all our Bond-Month observations (3200 monthly observations). Since the raw bond return data started in July 2002 and at least one year of data is needed to initialize risk characteristics, we start with the data from July 2004. After standardizing the raw characteristics, we filter the bonds whose maturity is less than two years and impute the missing values with 0. The equity and option characteristics are merged into bond data by the 'PERMNO'. If the firm has issued multiple stocks simultaneously, we only keep and merge the earliest issued stock.

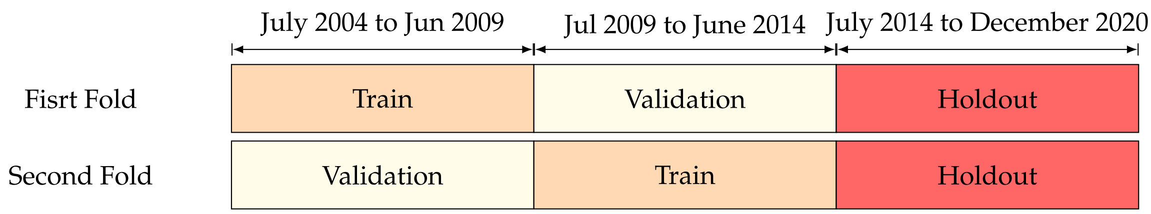
II.2 Model Training

We divide the TRACE dataset into two parts: we perform all the training on the sample from July 2004 to June 2014 and test the sample from July 2014 to December 2020. We train the neural network structure and parameters during the in-sample period and determine the mean-variance portfolio's weights based on the in-sample factors' statistics. To avoid the effect of outliers during training, we also winsorize the in-sample bond returns within the $ret_{low,t}$ and $ret_{up,t}$ bounds at each month t, where $ret_{low,t}$ and $ret_{up,t}$ are the cross-sectional 2.5% and 97.5% quantiles of month t, respectively. Each month, bond returns that are lower/higher than the monthly low/up quantile of cross-sectional returns will be revised as the $ret_{low,t}$ and $ret_{up,t}$. It is important to emphasize that winsorization is only applied to in-sample data. All out-of-sample (OOS) results are tested on non-winsorized data with potential extreme values.

We use both equal- and value-weight market factors as our benchmarks. For the IPCA/RP-PCA model, we train the PCA structure using the balanced individual bond returns from July 2004 to June 2014 and compute the factor value for both in-sample

Figure A1: Two-Fold Cross Validation

This figure demonstrates the deterministic two-fold cross-validation scheme to choose the tuning parameters. Specifically, we determine the tuning parameters using the sample from July 2004 to June 2014. The deterministic design divides the sample into two consecutive fold samples. We train our neural network separately on one and then calculate the fitting result with different tuning parameters on the other. We average the out-of-sample loss and choose the parameter pair with the best performance on this criterion.



and OOS periods based on this trained structure. Similarly, our neural network takes the individual bond data from July 2004 to June 2014 as a training set for our neural network structure, and we determine the tuning parameters through a two-fold cross-validation process as shown in Figure A1, for which we will provide more details in Appendix II.3.

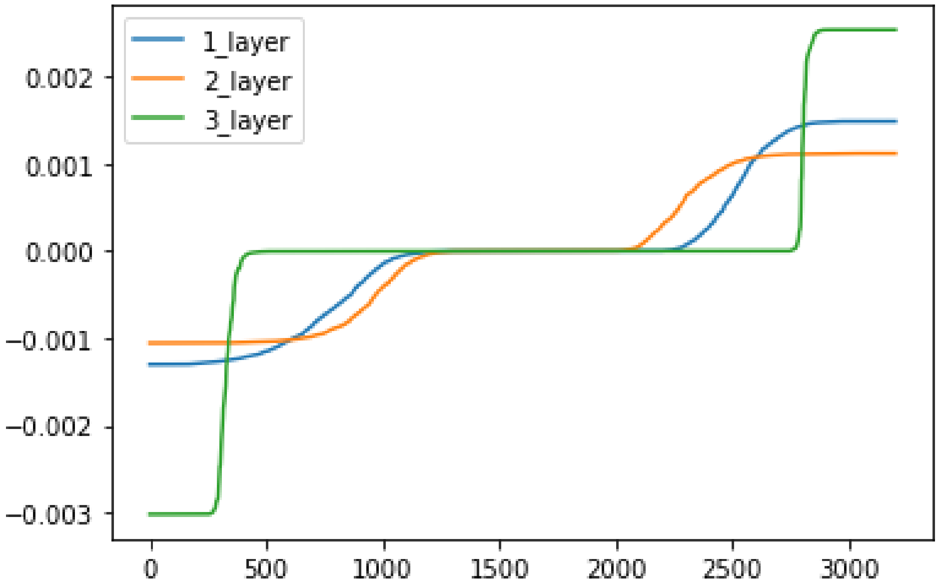
In ranking the bond returns, we choose $a1 = 50$ and $a2 = 8$ in Equation (16) such that at each time, about 50% to 70% of assets are in the middle rank and have zero weights, similar to the traditional sorting procedure (see Figure A2).

After training, we use the fixed network structure and input the pairwise characteristics to generate in-sample and OOS estimates. All results in section II are computed using PyTorch 1.10.2 and parallelized across a server with 96 Intel(R) Xeon (R) Gold 6230 @ 2.10GHz CPUs and 314 GB of RAM.

II.3 Tuning Parameters Selection

In this subsection, we outline our procedure for selecting tuning parameters. To determine the optimal tuning parameters for our network, we employ a two-fold cross-validation approach (in Figure A1) as follows. We first divide the sample period from July 2004 to June 2014 into two consecutive fold samples of equal length. We train

Figure A2: Softmax Ranking: $a_1 = 50$, and $a_2 = 8$



our neural network on one fold and evaluate the fitted results using different tuning parameters on the other fold. We compute the average loss from the validation samples and select the parameter pairs resulting in the smallest loss. Finally, we retrain the model using the chosen tuning parameters. This approach ensures that the data used is completely in-sample, eliminating any look-ahead bias that may affect our OOS trading results discussed in the main text. When implementing, we start with a reasonable set of initial tuning parameters and test additional points adjacent to these sets. We evaluate a total of 16 combinations of tuning parameters, which are shown in Table A4.

III Alternative Sample Splits

III.1 Reversed Sample

To ensure robustness in OOS evaluation, we reverse the training, validation, and forecasting periods in this subsection. In other words, we use the end periods of the original sample as training and validating sets and the beginning periods as the OOS

Table A4: Tuning parameter selection in the empirical analysis

This table presents the network tuning parameters used in our empirical analysis.

Notation	Tuning Parameters	Candidates	Chosen
a_1	Value in equation (16)	50	50
a_2	Value in equation (16)	5,8	8
HDN	Number of nodes in the hidden layer	66	66
BTCH	Batch size, in months	120	120
LR	Learning rate	1e-3,5*1e-3,1e-2,5*1e-2,1e-1,5*1e-1	5*1e-2
L^1 Penalty	L^1 penalty in objective function	1e-9,1e-8,1e-7,1e-6	1e-8
L^2 Penalty	L^2 penalty in objective function	1e-9,1e-8.1e-7,1e-6	1e-8
EPCH	Number of optimization epochs	400	400
OPT	Optimization method	Adam	Adam

evaluation periods.

We use the same approach as in subsection 2.2 to build the neural network. Table A5 shows that, 1) the deep tangent portfolios have significantly higher OOS Sharpe ratios than the market portfolio; 2) the layer-1 deep tangent portfolios achieve the highest OOS Sharpe ratio in most cases, both of which are consistent with our previous findings.

Table A5: The out-of-sample Sharpe ratios in the reversed sample

This table presents Sharpe ratios of the deep tangency portfolios constructed using multiple deep factors in the reversed sample. L_i represents a neural network with i layers, and D_i represents i deep factors included in the model. The test of the Sharpe ratio improvement by including an additional deep factor is based on Barillas and Shanken (2017). $***$, $**$, and $*$ denote the level of significance of 1%, 5%, and 10%, respectively. We take the sample from January 2011 to December 2020 for model training and validating and from July 2004 to December 2010 for out-of-sample testing.

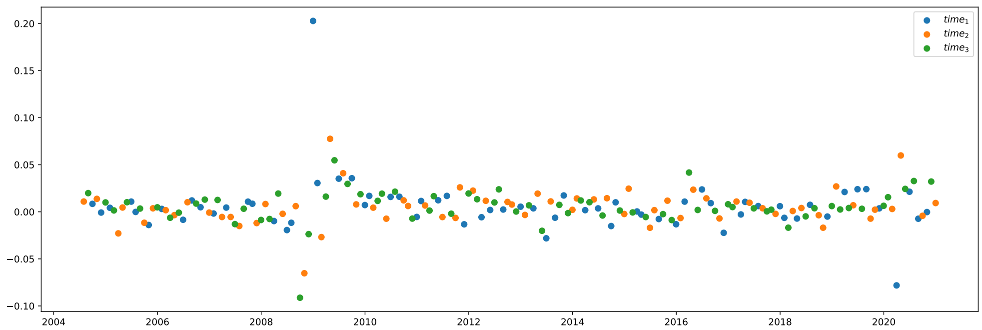
	MKT	D_1	D_2	D_3
		Panel A. Equal-Weight Market Factor		
L_1	0.73	2.13***	2.04	0.22
L_2	0.73	2.03***	2.06	2.04
L_3	0.73	0.82**	0.8	0.82
		Panel B. Value-Weight Market Factor		
L_1	0.67	2.27***	2.06	0.78
L_2	0.67	2.02***	2.07	2.13
L_3	0.67	0.77**	0.77	0.6

III.2 Sample Randomly Split

To further ensure robustness in OOS evaluation, we follow Fama and French (2018) to randomly split the dataset into three parts sequentially. Specifically, in every quarter (three months), the data is randomly assigned to one of three groups. This process continues until each month's data is assigned to a group. Figure A3 illustrates how we divide the dataset into three groups, represented by different colors. We then use the same approach as in subsection 2.2 to build the neural network and conduct a two-fold cross-validation scheme: for each test group, the other two groups are used for training and validation. Table A6 shows the OOS performance for all three groups. We find that: 1) the deep tangent portfolios have significantly higher OOS Sharpe ratios than the market portfolio; 2) the layer-1 deep tangent portfolios achieve the highest OOS Sharpe ratio in most cases. These results are consistent with our previous findings.

Figure A3: Bond market return for the sample randomly split

This figure plots the bond market return from July 2004 to December 2020. Different colors represent three different groups randomly split, denoted by $time_1, time_2$, and $time_3$ respectively.



IV Deep Tangency Portfolios Accounting for Transaction Costs

To account for transaction costs, we use bid-ask spreads to adjust each bond's return in the portfolio. Specifically, for the bond i in the long leg of the portfolio, the net-of-costs return is calculated as $r_{i,t} - \frac{1}{2}\text{bid-ask spread}_{i,t}$; for the bond i in the short leg of the portfolio, the net-of-costs return is calculated as $-r_{i,t} - \frac{1}{2}\text{bid-ask spread}_{i,t}$. We use the same approach as in subsection 2.2 to build the model. Table A7 shows the

Table A6: The out-of-sample Sharpe ratios in the sample randomly split

This table shows the out-of-sample Sharpe ratios in the sample randomly split. We follow Fama and French (2018) to randomly split the dataset into three parts sequentially. Specifically, in every quarter (three months), the data is randomly assigned to one of three groups. We use the same approach as in subsection 2.2 to build the neural network and conduct a two-fold cross-validation scheme: for each test group, the other two groups are used for training and validation. For example, the $time_1$column shows the out-of-sample performance when groups 2 and 3 are used for training and validation but group 1 for the out-of-sample evaluation. The *all* column shows the out-of-sample Sharpe ratios when we evaluate all three groups' out-of-sample performance.

	Equal-Weight Market Factor				**Value-Weight Market Factor**			
	$time_1$	$time_2$	$time_3$	all	$time_1$	$time_2$	$time_3$	all
MKT	0.69	0.90	1.07	0.83	0.67	0.89	1.14	0.88
L_1	2.78	2.95	3.10	2.80	2.11	3.23	2.08	2.39
L_2	0.91	1.91	2.23	1.40	2.32	1.94	3.36	2.18
L_3	1.60	2.09	1.69	1.78	2.68	2.39	1.66	2.20

OOS Sharpe ratios for the deep tangency portfolios incorporating transaction costs, which indicates that: 1) deep tangent portfolios consistently outperform the market portfolio with higher OOS Sharpe ratios; and 2) layer-1 deep tangent portfolios often achieve the highest OOS Sharpe ratios. These results align with our previous findings.

Table A7: Deep Tangency Portfolios Accounting for Transaction Costs

This table presents Sharpe ratios of the deep tangency portfolios constructed using multiple deep factors accounting for transaction cost. L_i represents a neural network with i layers, and D_i represents i deep factors included in the model. We sequentially add one additional deep factor for each choice of the number of neural network layers. The test of the Sharpe ratio improvement by including an additional deep factor is based on Barillas and Shanken (2017). $***$, $**$, and $*$ denote the level of significance of 1%, 5%, and 10%, respectively. We take the sample from July 2004 to June 2014 for model training and validating and the sample from July 2014 to December 2020 for out-of-sample testing.

	MKT	D_1	D_2	D_3
		Panel A. Equal-Weight Market Factor		
L_1	0.86	1.12***	0.25	-0.04
L_2	0.86	0.79	0.64	0.36
L_3	0.86	0.88	1.14***	0.94
		Panel B. Value-Weight Market Factor		
L_1	0.92	1.57***	0.84	0.74
L_2	0.92	1.16***	0.38	0.35
L_3	0.92	1.21***	0.98	0.55

V The Implementations of Standard Machine Learning Methods

In section 4.3, we present a forecast-implied long-short portfolio based on predicted bond returns from four ML methods: Lasso, Ridge, principal component analysis regression (PCA), and standard neural networks (NN) to predict the asset returns for comparison directly. Our implementation follows Feng, He, Wang, and Wu (2021).

PCA is a common dimension reduction technique used in empirical asset pricing to address the bias-variance trade-off problem. It involves using a low-dimension version of linearly transformed predictors to construct a predictive model. The procedure consists of two steps: the first combines predictors with a small set of linear combinations that best preserves the covariance structure, and the second involves using the first K components in multiple regressions. In our applications, K is set to 5 for PCA.

Lasso and ridge are linear regularized regressions used in ML finance to maintain interpretability. They add a penalty over least squares to preserve the predictors without transforming them. Lasso performs variable selection, while ridge shrinks the regression coefficients of useless predictors to small values. A tuning parameter controls the penalty weight, with a larger penalty weight imposing more shrinkage on the coefficients. A two-fold cross-validation determines all tuning parameters.

For the standard neural network (NN), we use the same input characteristics and two-fold cross-validation scheme as in our deep learning network. We also add regularization penalty terms to the loss function to limit the complexity of the neural networks, stabilize estimates, and avoid overfitting.