Deep Tangency Portfolios *

Guanhao Feng   Liang Jiang   Junye Li   Yizhi Song

First version: Nov. 2021; This version: December 16, 2022

Abstract

We propose a parametric approach for directly estimating the tangency portfolio weights by combining fundamental finance theory and deep learning techniques. The deep tangency portfolio is a combination of the market factor and a deep long-short factor constructed using a large number of characteristics. We apply our approach to the corporate bond market. Albeit acting as a market-hedged portfolio, the deep factor achieves a sizable risk premium with an out-of-sample annualized Sharpe ratio of 2.08. The deep tangency portfolio outperforms those constructed from commonly used observable or latent factors with an out-of-sample annualized Sharpe ratio of 2.90.

Keywords: Corporate Bond Returns, Deep Learning, Factor Models, Stochastic Discount Factor, Tangency Portfolios.

*We appreciate insightful comments from Hendrik Bessembinder, Seth Pruitt, Jianeng Xu, and Mao Ye. We are also grateful for helpful comments from the seminar and conference participants at Sun Yat-Sen University, Fudan University, 2022 Asian Finance Association Annual Conference, 2022 International Conference on Finance & Technology, 2022 China International Risk Forum, and FIRM 2022. Feng acknowledges the ECS grant (9048131) from Hong Kong Research Grants Council and the support from the InnoHK initiative and the Laboratory for AI-Powered Financial Technologies. 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

A fundamental theory in asset pricing is the equivalence of the stochastic discount factor (SDF) and the mean-variance efficient (MVE) portfolio. The minimum variance of the SDF equals the maximal squared Sharpe ratio of the MVE portfolio in the economy (Hansen and Jagannathan, 1991). Markowitz (1952) marks the beginning of modern portfolio theory, which formulates the solution to the optimal portfolio only using expected returns and covariance of individual assets ($\Sigma^{-1}\mu$). However, it is notoriously difficult to estimate the MVE portfolio in real-world situations accurately, making Cochrane (p.7, 2014) state "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."

A common alternative approach in the finance literature is to proxy the SDF as a linear function of a small number of characteristics-managed factors (e.g., Fama and French, 1996, 2015), hoping that those factors can span the minimum-variance SDF or the mean-variance frontier. However, the commonly used factors can hardly achieve the maximal Sharpe ratio of the asset universe (e.g., Kozak et al., 2018; Daniel et al., 2020; He et al., 2022; Lopez-Lira and Roussanov, 2020), leaving many "anomalies" unexplained. The literature has then examined and proposed a large number of factors (Harvey, Liu, and Zhu, 2016; Hou, Xue, and Zhang, 2020), leading to an issue of "factor zoo" (Cochrane, 2011).

This paper proposes a deep learning framework for constructing the optimal or tangency portfolio without relying on expected returns and covariance matrix estimates. Firm characteristics 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; Kozak, Nagel, and Santosh, 2020). Therefore, we directly parameterize the tangency portfolio weights as a nonlinear function of a large number of characteristics. Indeed, using a large set of characteristics and their nonlinear combinations is crucial, as the existing studies on machine learning (ML) show that there does not exist clear-cut evidence of sparsity of characteristics (see, e.g., Kozak, Nagel, and Santosh, 2020; Giannone, Lenza, and Primiceri, 2021), and nonlinearity is important (see, e.g., Freyberger, Neuhierl, and Weber, 2020; Gu, Kelly, and Xiu, 2020; Cong, Feng, He, and He, 2022).[1]

We aim to construct the tangency portfolio using a large panel of individual assets and a benchmark portfolio, such as the market factor. Our parametric portfolio policy mimics the commonly used portfolio sorting approach in empirical asset pricing through an economically-guided deep learning model, which extends and generalizes the one proposed by Feng et al. (2022). The multi-layer nonlinear deep neural network provides a supervised dimension reduction by transforming a large number of characteristics into one deep characteristic for each asset, based on which a deep factor is constructed as a long-short portfolio of individual assets using a nonlinear ranking scheme. The deep tangency portfolio is formed by combining the deep factor and the benchmark factor by maximizing its squared Sharpe ratio. Our deep learning framework is flexible enough to introduce multiple deep factors to augment the benchmark.

The economically-guided deep factor plays two important roles: (i) under the risk-adjusted objective of the tangency portfolio, the deep factor should have a low or even negative correlation with the benchmark factor, providing us with a potential hedge portfolio; and (ii) the deep factor spans to a large extent any missing risk factors other than the benchmark factor that should enter the pricing kernel. Our deep parametric portfolio policy only relies on the Sharpe ratio improvement over the benchmark without using any test assets, similar to the factor selection of Barillas and Shanken (2017) and Barillas et al. (2020). All the above features make our deep learning model more economically interpretable and largely alleviate the "black-box" criticism.

To demonstrate 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. The literature has proposed some observable factors for explaining time-series comovements and cross-sectional variations of corporate bond

[1]See the latest textbook (Negal, 2021) and review (Giglio et al., 2022), as well as references therein.

returns. For example, Fama and French (1993) argue that two factors based on bond term and default, together with the bond market portfolio, can capture common variation in both equity and bond returns. Bai, Bali, and Wen (2019) (BBW hereafter) propose an alternative bond factor model based on downside risk, credit risk, and liquidity risk. However, these models impose a strong ad hoc sparsity when constructing the factors.[2] Therefore, those observable factor models may be incapable of competing with latent factor models that take into account a large number of characteristics (see, e.g., Kelly, Palhares, and Pruitt, 2022; Kelly and Pruitt, 2022).

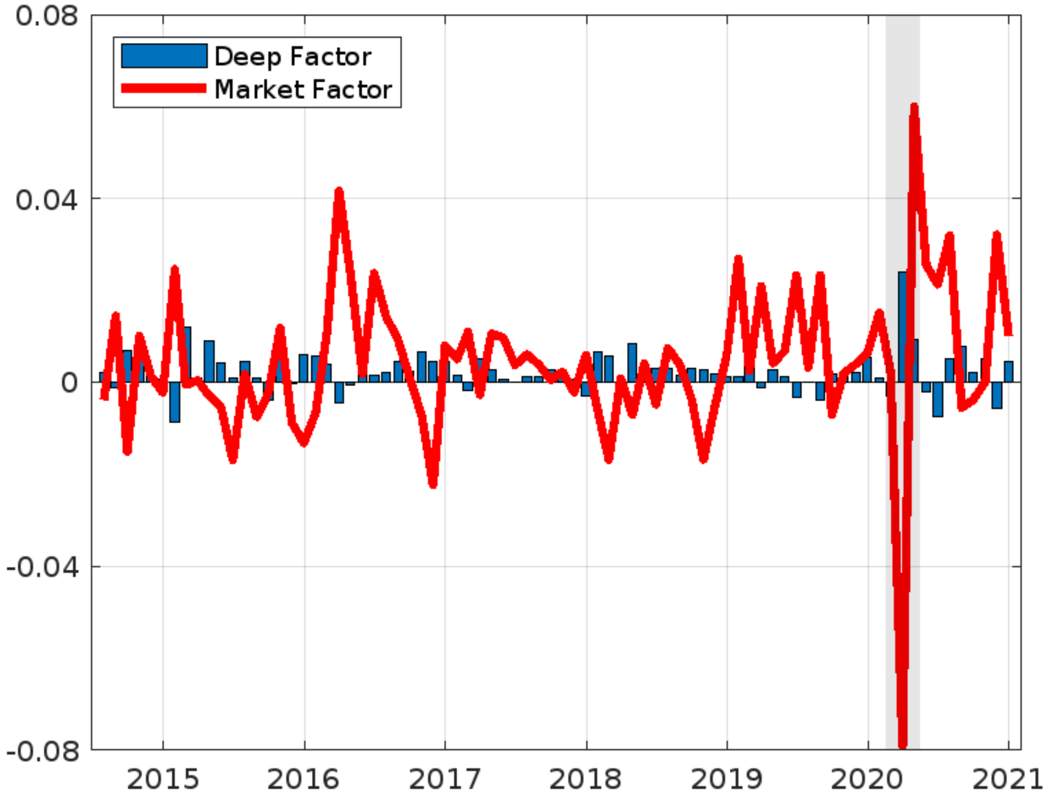
Empirical Highlights. We construct the monthly corporate bond returns using transaction data on corporate bond prices from the enhanced version of the Trade Reporting and Compliance Engine (TRACE). Three types of characteristics are taken into account. The first type is bond characteristics. Combining the data of TRACE and the Mergent Fixed Income Securities Database (FISD), we construct a set of 41 bond characteristics. Second, given that both bond and stock prices are contingent on firm fundamentals, we also collect 61 equity characteristics that are commonly used in the literature (see, e.g., Freyberger, Neuhierl, and Weber, 2020; Feng, He, Polson, and Xu, 2022). Lastly, given that the recent literature has found that equity option-related variables contain information on future corporate bond returns (see, e.g., Cao, Goyal, Xiao, and Zhan, 2022; Chung, Wang, and Wu, 2019; Huang, Jiang, and Li, 2021), we construct 30 equity option-related characteristics. Altogether, 132 characteristics are fed into our deep learning model. The sample period ranges from July 2004 to December 2020. We take the subsample from July 2004 to June 2014 for model training and validating and the subsample from July 2014 to December 2020 for out-of-sample testing.

Our empirical findings can be briefly summarized as follows. First, while it earns a relatively small mean excess return compared to the bond market factor and some other traded factors (e.g., BBW factors), the deep factor only slightly varies over time, resulting in a much higher Sharpe ratio than all other factors under consideration for

[2]Other corporate bond 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).

Figure 1: Bond Market Factor v.s. Deep Factor

This figure displays the time series of deep factor and market factor returns in the out-of-sample period(2014.07 to 2020.12).



both in-sample and out-of-sample periods. Furthermore, the deep factor negatively correlates with the bond market factor, and they hardly go down simultaneously, providing us with a market-hedge portfolio. Figure 1 presents time series of returns on the deep factor and the market factor for the out-of-sample period. During those market downturn periods, the deep factor always keeps positive, and this is particularly striking during the outbreak of the Covid-19 pandemic.

Second, the deep tangency portfolio constructed from the market portfolio and the deep factor earns an out-of-sample annualized Sharpe ratio of as large as 2.90, much higher than that of the market portfolio (0.86), of the tangency portfolio constructed from the BBW four factors (1.07), and of the tangency portfolio constructed from the Fama-French three equity factors (MKTRF, SMB, and HML) plus two bond factors (term and default factors) (1.26). Consistently, we find that neither of these factor models can explain excess returns on the deep factor and the deep tangency portfolios in the factor-spanning regressions. The deep factor loads negatively on the bond market factor, and the deep tangency portfolio has negligible exposure to the bond market factor, further highlighting the deep factor's market-hedging role.

Third, we further show that it is crucial to consider various types of characteristics in constructing the deep tangency portfolio. We find that when we exclude the option-related variables, the out-of-sample annualized Sharpe ratio of the deep tangency portfolio decreases to 1.91, and it has a further decrease when only bond characteristics are used. Indeed, for largely spanning expected returns and covariance, it is essential to introduce as many characteristics as possible. This evidence is in stark contrast to previous studies that argue 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 provides further empirical evidence in support of integration between bond and equity markets (see, e.g., Schaefer and Strebulaev, 2008; Kelly, Palhares, and Pruitt, 2022).

Finally, our deep parametric portfolio policy provides an alternative latent factor construction. We then make additional analysis by comparing the performance of our deep tangency portfolio with those constructed from two recently developed latent-factor methods, i.e., risk-premium principal component analysis (RP-PCA) of Lettau and Pelger (2020) and instrumental principal component analysis (IPCA) of Kelly, Pruitt, and Su (2019) and Kelly, Palhares, and Pruitt (2022). Our deep tangency portfolio outperforms: for the same out-of-sample period, the tangency portfolio from the five RP-PCA factors earns an annualized Sharpe ratio of only 0.91, and that from the five IPCA factors achieves an annualized Sharpe ratio of 2.32.

Literature. Our paper contributes to several strands of literature. First, it contributes to the recent literature on ML methods constructing latent factors to approximate the SDF by considering a large number of characteristics. Kelly, Pruitt, and Su (2019) propose an instrumental principal component analysis (IPCA) that allows for timevarying factor loadings depending on characteristics. Kozak, Nagel, and Santosh (2020) assume that the SDF loading is a linear function of characteristics, and find no clear evidence of sparsity of characteristics in the SDF loading. Bryzgalova, Pelger, and Zhu (2019) propose penalized regressions on tree-basis portfolios for constructing the maximal Sharpe ratio SDF, and Cong, Feng, He, and He (2022) develop a panel-tree approach for generating latent factors and constructing the SDF. In addition, our paper also relates to recent attempts that develop nonlinear deep neural networks for latent factor models, such as the auto-encoder (Gu, Kelly, and Xiu, 2021), generative adversarial network (Chen, Pelger, and Zhu, 2022), and characteristics-sorted factor approximation (Feng, He, Polson, and Xu, 2022). Differently, based on a fundamental economic theory of the equivalence between the SDF and the MVE portfolio, our paper develops a flexible and interpretable methodology, aiming to create a tangency portfolio without estimating expected returns and covariance.

Second, our paper contributes to the literature that investigates the cross-sectional predictability of characteristics to corporate bond returns.[3] However, most of those papers impose a strong ad hoc sparsity in modeling. Bali, Goyal, Huang, Jiang, and Wen (2021) and He et al. (2021b) investigate bond return predictability via machine learning methods. In the same vein as our paper, Kelly, Palhares, and Pruitt (2022) examine the cross-sectional pricing of corporate bonds relying on the IPCA method, showing that a five-factor model outperforms commonly used observable factor models on the ICE corporate bond dataset. Kelly and Pruitt (2022) further confirm that the main analysis of Kelly, Palhares, and Pruitt (2022) is robust to using the TRACE dataset. However, our method is more flexible and allows for modeling nonlinearity and interactions of characteristics. As discussed above, our deep tangency portfolio has a higher out-of-sample Sharpe ratio than the IPCA tangency portfolio within our data context.

Finally, our paper contributes to robust portfolio construction that sidesteps direct estimation of the covariance matrix and average returns. Brandt (1999) and Ait-Sahalia

[3] See, for example, Bai, Bali, and Wen (2019), Lin, Wang, and Wu (2011), Jostova, Nikolova, Philipov, and Stahel (2013), Chung, Wang, and Wu (2019), Huang, Jiang, and Li (2021), and He et al. (2021a).

and Brandt (2001) propose a nonparametric approach for estimating portfolio weights from the Euler first-order conditions, thus bypassing the estimation of return covariance and averages. Brandt, Santa-Clara, and Valkanov (2009) provide a parametric approach by estimating the portfolio weights as a linear function of characteristics (size, value, and momentum). Based on the same approach as Brandt, Santa-Clara, and Valkanov (2009), Brandt and Santa-Clara (2006) examine a market-timing problem involving stocks, bonds, and cash, and DeMiguel et al. (2020) show the economic rationale of transaction cost by utilizing multiple characteristics. Raponi, Uppal, and Zaffaroni (2021) combine an "alpha" portfolio and a "beta" portfolio relying on a factor model for the robust portfolio choice. Our paper provides a parametric approach for estimating the portfolio weights directly, where the unique design of long-short portfolio weights reflects the nonlinear risk-return relationship on the deep characteristics generated from the multi-layer deep neural network.

The remainder of the paper is organized as follows. Section 2 presents our model and the deep learning algorithms. Section 3 presents data on corporate bond returns and characteristics. Section 4 provides empirical findings. Section 5 concludes the paper. Additional materials and empirical results are reported in the Internet Appendix.

2 Methodology

2.1 Maximal Sharpe Ratio Portfolio

There exists a duality between the SDF variance and Sharpe ratios. We start with the minimum-variance SDF in the economy that spans N individual asset excess returns, $r_t = [r_{1,t}, ..., r_{N,t}]'$, as constructed by Hansen and Jagannathan (1991),

$$m_{t+1} = 1 - w_t'\Big(r_{t+1} - \mu_t\Big), \tag{1}$$

where $\mu_t = E_t[r_{t+1}]$ represents the conditional expectation of asset excess returns. By plugging in the linear SDF in Equation (1) into the fundamental pricing relation,

$E_t[m_{t+1} r_{t+1}] = 0$, the solution to the SDF loading w_t takes the form of

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

where Σ_t is the conditional variance-covariance matrix of excess returns, $\Sigma_t = Cov_t(r_{t+1})$. The conditional variance of the SDF is then given by

$$Var_t(m_{t+1}) = \mu_t' \Sigma_t^{-1} \mu_t, \tag{3}$$

which equals the maximal conditional squared Sharpe ratio of the tangency portfolio,

$$R_{t+1}^{opt} = w_t' r_{t+1}, \tag{4}$$

whose weights are the same as the SDF loadings in Equation (2).

In practice, it is challenging to estimate expected returns and covariance matrix. The number of individual assets, N, is usually large, making it difficult to estimate the large covariance matrix (Σ_t). Moreover, as a general observation, mean estimates (μ_t) 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 out-of-sample performance of optimal portfolios (see, e.g., DeMiguel, Garlappi, and Uppal, 2009).

A common approach in the finance literature is to adopt a factor model to reduce the dimensionality of the SDF by approximating it with a small number of factors (e.g., Fama and French, 1996, 2015). Assume that the SDF loading w_t in Equation (1) can be largely captured in a linear form by J characteristics, z_t, a $N \times J$ matrix for $J \ll N$, such that

$$w_t = \tilde{w}_t + z_t \kappa, \tag{5}$$

where following the convention of the finance literature, $\tilde{w}_t$ is normalized weights on market capitalization of firms, z_t is usually cross-sectionally standardized to have zero mean, 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 factors, which are zero-investment characteristic-managed long-short portfolios. Equation (5) suggests that the SDF takes the form of

$$m_{t+1} = 1 - \delta' \big(F_{t+1} - \mu_{F,t} \big), \tag{6}$$

where $\delta = [1, \kappa']'$, and $F_t = [R_{m,t}, f_t']'$.

Such a dimension reduction aims to use those small number of factors (F_t) to approximate the SDF (see Equation (1)) and span the MVE portfolio (see Equation (3)). 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 sparsity of characteristics (e.g., Kozak et al., 2020; Giannone et al., 2021) and many characteristics and their nonlinear combinations contain information on the joint distribution of asset returns for characterizing the cross-sectional variation (e.g., Freyberger et al., 2020; Gu et al., 2020; Cong et al., 2022).

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

$$w_{i,t} = \tilde{w}_{i,t} + \theta \, w_d(z_{i,t}; \Phi), \quad i = 1, \ldots, N, \tag{7}$$

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 (see, e.g., Brandt, Santa-Clara, and Valkanov, 2009).

To be explained clearly in the next subsection, the function, $w_d(\cdot)$, produces weights with economically-guided nonlinearity for forming a zero-cost long-short portfolio with the sums of weights for long and short legs normalized to 1 and -1, respectively. The tangency portfolio return in Equation (4) can then be represented by

$$R_{t+1}^{opt} = \sum_{i=1}^{N} \tilde{w}_{i,t} r_{i,t+1} + \theta \sum_{i=1}^{N} w_d(z_{i,t}; \Phi) r_{i,t+1} = R_{m,t+1} + \theta R_{d,t+1}, \tag{8}$$

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. The parameterization of Equation (7) suggests a two-factor reduced-form SDF with factors of $R_{m,t}$ and $R_{d,t}$. 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 (5).

When we have *a priori* knowledge that a particular set of observable factors is helpful in spanning the efficient portfolio frontier, we can introduce these factors by inserting them into Equation (7) and construct the portfolio weights, $w_{i,t}$, as follows,

$$w_{i,t} = \tilde{w}_{i,t} + \tilde{w}_{i,t}^{p} \theta_p + \theta_d w_d(z_{i,t}; \Phi), \quad i = 1, \ldots, N, \tag{9}$$

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 tangency portfolio return is then given by

$$R_{t+1}^{opt} = R_{m,t+1} + \theta_p' R_{p,t+1} + \theta_d R_{d,t+1}, \tag{10}$$

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 minimum-variance SDF or a tangency portfolio that delivers the maximal Sharpe ratio. For this purpose, we search for the 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^{opt}_{t+1},

$$\max_{\theta,\Phi} \frac{1}{T}\sum_{t=0}^{T-1} SR_t^2(R^{opt}_{t+1}). \tag{11}$$

The long-short portfolio, $R_{d,t+1}$, plays two fundamental roles: (i) according to the principle of diversification, Equation (11) suggests it should have a low or even negative correlation with the market (and other benchmark factors), providing us with a potential hedge portfolio; and (ii) 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, implying that it may have a sizable market price of risk.

Our approach can also be interpreted as a dimension reduction of characteristics and risk factors. Empirically, many studies have proved the failure of CAPM, i.e., other than the market factor, more factors need to be introduced to the pricing kernel for explaining time-series comovements of asset returns and expected return spreads across assets. The most popular factors are characteristic-managed portfolios, such as the Fama-French factors (Fama and French, 1996, 2015) and the BBW corporate bond factors (Bai, Bali, and Wen, 2019). 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. The dimension reduction in constructing characteristic-managed portfolios only relies on 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).

In what follows, we propose a deep learning method for constructing the portfolio weights of $w_d(\cdot)$ in Equations (7) and (9) and the characteristic-managed long-short portfolio $R_{d,t}$. While many popular characteristic-managed factors have sidestepped the high-dimensional problem by focusing on only a small number of characteristics, our approach can easily consider many potential characteristics and their nonlinear combinations.

2.2 Deep Factor and Deep Tangency Portfolio

Our construction of the long-short portfolio, $R_{d,t}$, relies on a deep learning model, aiming to construct the tangency portfolio by complementing the benchmark factors. Rather than specifically relying on average returns and covariance matrix of high-dimensional individual assets, we retain the conventional sorting scheme in deep learning based on information of many characteristics. For this purpose, we extend and generalize the deep learning method proposed by Feng et al. (2022).

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 lagged characteristics of N assets;
- $\{R_{b,t}\}_{b=1}^{P+1}$, 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, $i = 1, \dots, N$, our deep learning model works as follows,

$$Z_{i,t-1}^{(0)} = [z_{1,i,t-1}, \cdots, z_{K,i,t-1}]', \tag{12}$$

$$Z_{i,t-1}^{(l)} = G(A^{(l)} Z_{i,t-1}^{(l-1)} + b^{(l)}), \tag{13}$$

for $l = 1, \dots, L$, 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, $Z^{(L)}_{t-1}$. The parameters to be trained in this part are deep learning weights A and biases b, namely,

$$\left\{\left(A^{(l)}, b^{(l)}\right) : A^{(l)} \in \mathbb{R}^{K_l \times K_{l-1}}, b^{(l)} \in \mathbb{R}^{K_l}\right\}_{l=1}^{L}. \tag{14}$$

Different from the stepwise sorting approach that is commonly used to construct zero-cost long-short factors (e.g., the Fama-French SMB and HML factors), we adopt a nonlinear approach to proxy for the long-short portfolio weights as follows,

$$w_d(z_{t-1}) \equiv W_{t-1} = h(Z^{(L)}_{t-1}), \tag{15}$$

where the function $h(\cdot)$ uses the *softmax* activation and calculates the portfolio weights based on the ranking of the deep characteristic. For the $1 \times N$ vector of $x = Z^{(L)}_{t-1}$, it 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 we define $x^+ := -a_1 e^{-a_2 x}$ and $x^- := -a_1 e^{a_2 x}$, and a_1 and a_2 are two hyperparameters. 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 weights of zero, similar to the traditional sorting procedure. Furthermore, we normalize the portfolio weights such that the sum of weights in the long leg is equal to 1 and that in the short leg is equal to -1. As discussed in Feng et al. (2022), such a nonlinear rank-weighting scheme depends not only on the cross-sectional rank information but also on the distributional properties of characteristics. Nevertheless, in implementation, we also construct the deep portfolio weights using the standard sorting approach that longs top 30% individual assets and shorts bottom 30% individual assets on the basis of individual deep characteristics.

The deep factor portfolio weights, W_{t-1}, in Equations (15) and (16), sum to zero by construction. Our deep factor, $R_{d,t}$, can be computed as

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

which can be combined with the market or other benchmark factors to form the deep tangency portfolio as in Equation (8). Note that more than one deep factor can be constructed in an iterative fashion by treating the previous one as a new benchmark factor in our algorithm. As a result, the additional deep factor should capture pricing information not contained in the previous one.

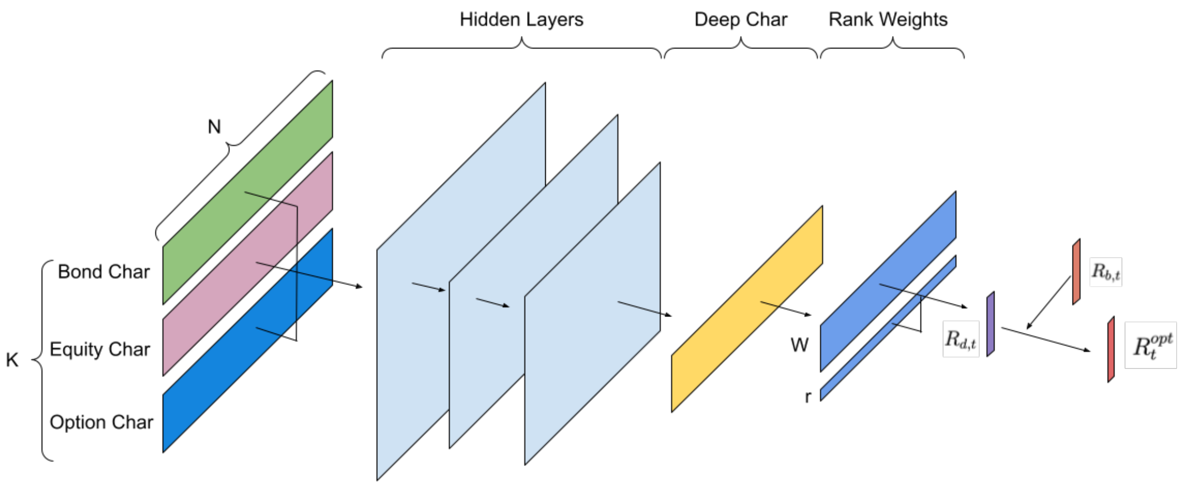
Given that all parameters in our model are time-invariant and that we implicitly assume that characteristics fully capture all aspects of expected returns and covariance relevant to optimal portfolios, the conditional model becomes an unconditional one, and the objective function in Equation (11) can be replaced by the unconditional squared Sharpe Ratio of optimal portfolio R_t^{opt} on $\tilde{F}_t = [R'_{b,t}, R_{d,t}]'$,

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

There are usually a large number of parameters for modeling a multi-layer neural network. To avoid overfitting and improve the model's out-of-sample performance, we augment the objective function by introducing L_1- and L_2-norm penalties and min-

Figure 2: Deep Learning Network Architecture

This figure provides a visualization of the deep learning architecture. The equity, bond, and option characteristics $Z^{(0)}$ are transformed via the multi-layer neural network to deep characteristics $Z^{(L)}$, which are nonlinearly ranked to calculate deep factor portfolio weights W. Combining W and individual bond returns, we construct the long-short deep factors, R_d, which are used with the benchmark factors, R_b, to compose an optimal portfolio R^{opt}.



imize the following loss function,

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

where the L_1-norm penalizes the off-diagonal weights, aiming to stabilize the model and make interactions of characteristics sparse, and the L_2-norm penalizes the complexity of the neural network, avoiding overfitting. The hyperparameters, γ_1 and γ_2, need to be tuned through training and validation. Figure 2 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 learning model in Subsection 3.2 and present the benchmark factor and competing factor models in Subsection 3.3.

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. The importance of 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 follow Bai, Bali, and Wen (2019) to 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, 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. As in Bai, Bali, and Wen (2019), 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 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, as in Feng et al. (2022), we make a balanced panel by only keeping 3,200 bonds with the largest size each month.[4] The final sample of corporate bond returns spans the period from July 2004 to December 2020.

Table 1 presents the summary statistics of excess corporate bond returns and some typical bond characteristics. The sample includes 24,789 corporate bonds issued by 3,383 unique firms and a total of 633,600 bond-month return observations. As shown in Panel A, the mean monthly excess bond return is about 0.63% with a standard deviation of 5.37%. The sample contains bonds with a mean size of about 803 million, a mean rating of 8.78, which is a BBB+ rating [5]. 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

[4] To avoid the volatility and liquidity effect of small market-value bonds, we select the largest 3200 bonds among all available bonds each month.

[5] 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.

have a smaller average monthly excess return (0.52% vs. 0.94%), a lower standard deviation (3.00% vs. 9.38%), and a higher rating level (7.04 vs. 13.95). Similarly, 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 are relevant to corporate bond return predictability. The first 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 some equity characteristics, which have been 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 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, invest-

Table 1: Summary Statistics

Our final data sample includes 633,600 monthly return observations of 24,789 unique corporate bonds from July 2004 to December 2020. We report the summary statistics not only around the whole Trace Data(ALL) but also sub-set separated by Rating type(Investment Grade(IG) & Non-Investment Grade(NIG)) and ownership(Public & Private).

Panel A: Cross-sectional statistics

	ALL	IG	NIG	Public	Private
Bond-month observations	633,600	474,225	159,375	485,992	147,608
Ret mean (%)	0.63	0.52	0.94	0.55	0.87
Ret std (%)	5.37	3.00	9.38	3.77	8.79
Rating mean	8.78	7.04	13.95	8.32	10.29
Duration mean	3.96	4.25	3.09	4.06	3.62
Age mean	4.30	4.40	4.01	4.28	4.39
Size mean (million)	803	858	640	831	711

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

Maturity	AAA	AA	A	B	Junk	ALL
2	0.15	0.71	2.63	2.71	1.45	7.66
3	0.19	0.78	3.00	3.36	2.02	9.35
4	0.18	0.77	3.02	3.57	2.55	10.09
5	0.15	0.75	2.99	3.70	3.13	10.72
6	0.10	0.39	1.93	2.86	3.44	8.72
7	0.09	0.38	1.88	2.89	3.41	8.66
8	0.08	0.34	1.76	2.79	2.61	7.58
9	0.07	0.34	1.72	2.78	1.90	6.81
10	0.07	0.34	1.66	2.63	1.32	6.01
$\geq$11	0.58	1.51	8.62	10.35	3.33	24.40
ALL	1.67	6.33	29.20	37.64	25.15	100

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

Ownership	AAA	AA	A	B	Junk	ALL
Private	0.12	1.22	4.53	8.50	8.93	23.30
Public	1.56	5.10	24.68	29.14	16.22	76.70
ALL	1.67	6.33	29.20	37.64	25.15	100.00

ment, profitability, frictions or size, and intangibles, most of which have been used in empirical asset pricing (see, e.g., Green, Hand, and Zhang, 2017; Freyberger, Neuhierl, and Weber, 2020).

In addition, the recent literature has found that a number of option-related variables have predictive power for corporate bond returns (see, e.g., Cao et al. (2022); Chung et al. (2019), Huang, Jiang, and Li (2021)). 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 help forecast corporate bond returns as well.

Altogether, we have a large number of characteristics, 132. The bond, equity, and option characteristics are listed in Table A1, Table A2, and Table A3, respectively, in Appendix. Before we feed those characteristics into our deep learning model, we cross-sectionally rank and standardize them each month such 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 Market Factor. Given that there do not exist well-established characteristic-managed factors in the corporate bond market, we just take the corporate bond market portfolio as our benchmark. Similar to Kelly, Palhares, and Pruitt (2022), our benchmark market portfolio is constructed simply as the equal-weighted average of excess corporate bond returns in our sample, i.e., $\tilde{w}_{i,t} = 1/N$.

Competing Factor Models. We consider two corporate bond factor models: one is the BBW four-factor model (Bai, Bali, and Wen, 2019), and the other is a Fama-French five-factor model that combines three equity factors and two bond factors (Fama and French, 1993, 1996):

(i) The BBW four factors. Bai, Bali, and Wen (2019) propose a four-factor model for the corporate bond market. Those factors include the bond market factor, the downside risk factor (DRF), the credit risk factor (CRF), and the liquidity factor (LRF). The downside risk factor is the value-weighted average return difference between the highest-VaR portfolio minus the lowest-VaR portfolio within each rating portfolio; the credit risk factor is the value-weighted average return difference between the highest credit risk portfolio minus the lowest credit risk portfolio within each VaR portfolio, and the liquidity risk factor is the value-weighted average return difference between the highest illiquidity portfolio minus the lowest illiquidity portfolio within each rating portfolio. Given that the BBW factors are constructed using the TRACE data, we obtain those factors from the authors, instead of constructing them ourselves.

(ii) The Fama-French five factors (FF5). We combine the Fama-French three equity factors, i.e., MKT, SMB, and HML (Fama and French, 1996), and two bond factors, i.e., the term and default factors (Fama and French, 1993). The term factor is defined as the difference between the long-term government bond returns and the one-month Treasury bill rate, and the default factor is defined as the difference between the long-term corporate bond returns and the long-term government bond returns.

4 Empirical Findings

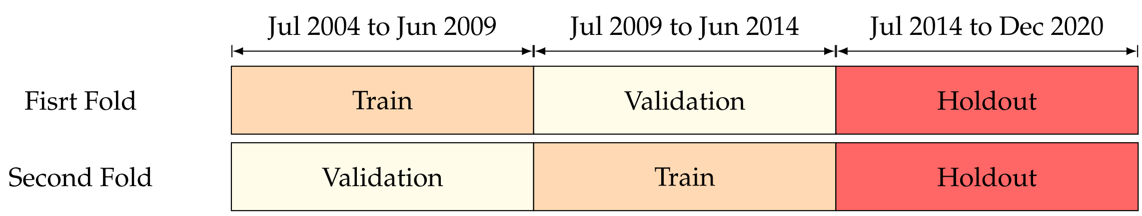
In our empirical implementation, we split the sample into two parts: the subsample from July 2004 to June 2014 for model training and validating and the subsample from July 2014 to December 2020 for out-of-sample testing. All out-of-sample results are based on in-sample parameter estimates. We adopt a two-fold deterministic cross-validation scheme using the first subsample as shown in Figure 3 to determine the penalty parameters and learning rate for a given number of neural network layers, which range from 1 to 3 Below presents our main empirical findings and examines how much improvement the deep factors can make over the benchmark and competing models.

4.1 Deep Corporate Bond Factors

Table 2 presents summary statistics of deep factors in terms of mean return, volatility, annualized Sharpe ratio, and maximal drawdown in the past 12 months (Max DD). When constructing deep factors, we take the equal-weighted corporate bond market factor as the benchmark and restrict the sums of weights of the long and short legs to

Figure 3: Two-Fold Cross Validation

This figure demonstrates the deterministic two-fold cross-validation scheme. We determine the hyper-parameters for the first 120 months (Jul 2004 to Jun 2014). Specifically, the deterministic design divides the sample into two consecutive parts (Jul 2004 to Jun 2009 and Jul 2009 to Jun 2014). We train our neural network separately on the train set and then compare the result with different parameter settings on the validation set. For example, for the first fold case, we train the model using the data from Jul 2004 to Jun 2009 and evaluate the out-of-sample performance on the data set from Jul 2009 to Jun 2014. After the same operation for the second fold case, we average the out-of-sample loss and choose the parameter pair with the best performance on this criterion.



1 and -1, respectively, such that the deep factor weights sum to zero. Panel A presents deep factors constructed from 1-, 2-, and 3-layer neural networks. In-sample training evidence shows that the shallow neural network works well enough because the 1-layer deep factor outperforms: it has the highest annualized Sharpe ratio (1.62). Such a result is further confirmed in the out-of-sample tests: the 1-layer deep factor has monthly mean return and volatility of 0.24% and 0.44%, respectively, resulting in an annualized Sharpe ratio of 1.86, which is much larger than that of the 2- or 3-layer deep factor.

We present the same summary statistics for the two competing factor models for comparison. Panel B is for the BBW factors. We find that in the in-sample period, the DRF and LRF factors earn significant average returns and annualized Sharpe ratios that are larger than 1.00, but smaller than that of the 1-layer deep factor; we further see that in the out-of-sample period, while both the DRF and LRF factors still earn significant average returns, only the LRF factor earns an annualized Sharpe ratio of larger than 1.00, which is also much smaller than that of the 1-layer deep factor (1.17 vs. 1.86). None of the Fama-French factors earns an annualized Sharpe ratio of larger than one both in the in-sample and out-of-sample periods (Panel C).

To sum up, while the premium of our deep factor is restrained, its volatility is

Table 2: Descriptive Statistics of Deep and Competing Factors

This table reports the descriptive statistics in percentage containing the mean of return, Newey West standard error (Newey and West, 1994) adjusted t-statistics , Standard Deviation(Std), annualized Sharpe Ratio (SR), and Maximal Drawdown (Max DD) of deep portfolios obtained from Neural Network and the constructed long-short factors contained by BBW, and Fama-French factors.

	In Sample Period (2004.7–2014.6)					Out of Sample Period (2014.7–2020.12)				
	Mean	tstat	Std	SR	Max DD	Mean	tstat	Std	SR	Max DD
					Panel A. Deep Factors					
R_d^1	0.29	(4.78)	0.63	1.62	6.78	0.24	(5.93)	0.44	1.86	0.97
R_d^2	0.23	(3.88)	0.62	1.30	7.11	0.12	(2.60)	0.42	0.98	1.61
R_d^3	0.34	(4.70)	0.75	1.58	7.53	0.18	(3.10)	0.58	1.06	2.03
					Panel B. BBW Four Factors					
MKT_C	0.60	(2.36)	2.56	0.81	19.85	0.41	(2.28)	1.67	0.86	7.92
DRF	0.78	(2.83)	2.39	1.13	18.62	0.55	(2.02)	2.17	0.88	13.15
CRF	0.50	(2.08)	1.94	0.90	18.57	0.10	(0.43)	1.81	0.18	14.45
LRF	0.54	(2.57)	1.50	1.25	7.66	0.29	(2.42)	0.86	1.17	4.94
					Panel C. FF Five Factors					
MKT_E	0.64	(1.34)	4.36	0.51	46.34	1.07	(2.62)	4.41	0.84	20.52
SMB	0.18	(0.95)	2.28	0.28	10.44	−0.10	(0.31)	2.84	−0.12	16.05
HML	0.07	(0.25)	2.53	0.09	24.33	−0.77	(1.94)	3.00	−0.88	33.74
TRM	0.49	(1.71)	3.36	0.51	14.40	0.54	(1.60)	2.87	0.65	11.97
DEF	0.03	(0.15)	2.30	0.05	19.41	0.08	(0.40)	1.93	0.14	13.53

small, leading to a high market price of risk and delivering relatively stable returns, particularly during market downturns, a fact further supported by its maximal drawdown. When we examine the maximal drawdowns of all single factors, our 1-layer deep factor has a maximal drawdown of 6.78 in the in-sample period and a maximal drawdown of 0.97 in the out-of-sample period, both of which are much smaller than those of all the other factors.

4.2 Deep Tangency Portfolios

We now move on to examine the portfolio performance. Table 3 presents the Sharpe ratios of our deep tangency portfolios and various optimal portfolios constructed from the competing factors. In Panel A, we see that our deep tangency portfolio earns an annualized in-sample Sharpe ratio of as high as 11.65 when using the 1-layer neural network, in stark contrast to the corresponding Sharpe ratio of the benchmark market factor (0.81). Such a high in-sample Sharpe ratio is not surprising as our deep learning model is trained to maximize the Sharpe ratio of the tangency portfolio formed by the market factor and the deep factor. Panel A of Figure 4 presents the in-sample scatter plot between the benchmark market factor and the 1-layer deep factor; we see that they are highly negatively correlated, resulting in a high Sharpe ratio of the deep tangency portfolio according to the principle of diversification. It seems that our deep factor plays the role of a market-hedge portfolio. Increasing the depth of the neural network does not help improve the performance of deep tangency portfolios.

We are, in fact, more interested in the out-of-sample performance of the deep tangency portfolios and other optimal portfolios. Note that all portfolios' weights are determined by the in-sample estimations. In Panel B, we see that the deep tangency portfolio that combines the market portfolio and the 1-layer deep factor has an annualized out-of-sample Sharpe ratio of 2.90, much higher than that of the market portfolio (0.86) and that of the deep factor itself (1.86, see Table 2). The scatter plot in Panel B of Figure 4 suggests that the deep factor negatively correlates with the market portfolio in the out-of-sample period. Interestingly, we find that the deep factor and the market portfolio hardly go down simultaneously, as very few returns are positioned in the lower-left coordinate (also see Figure 1). To further check this point, Panel C of Figure 4 plots the cumulative returns over time of the market portfolio and the deep factor for the in-sample and out-of-sample periods, respectively. We see that the deep factor goes the opposite whenever there is a market downturn. For example, during the 2008 global financial crisis (in sample) and the outbreak of the Covid-19 pandemic (out of sample), the cumulative returns of the deep factor increased over time much more smoothly than those of the market portfolio, consistent with the evidence found in the previous subsection. Those results provide further evidence in support of the deep factor as a market-hedge portfolio.

The optimal portfolios constructed from the competing factors perform much worse

Table 3: Performance of the Deep Tangency Portfolio

This table presents the Sharpe ratios of the tangency portfolios constructed using the multiple factors with the deep portfolio. The deep learning model is trained with the market factor as the only benchmark. For the deep learning model, we consider the 1-3 layers in the neural network architecture. We take the sample from Jul 2004 to Jun 2014 for model training and validation, and from Jul 2014 to Dec 2020 for out-of-sample tests. We follow Barillas and Shanken (2017) to statistically test the significance of the Sharpe ratio increase of one strategy over the other. $***$, $**$, and $*$ denote the level of significance of 1%, 5%, and 10%, respectively.

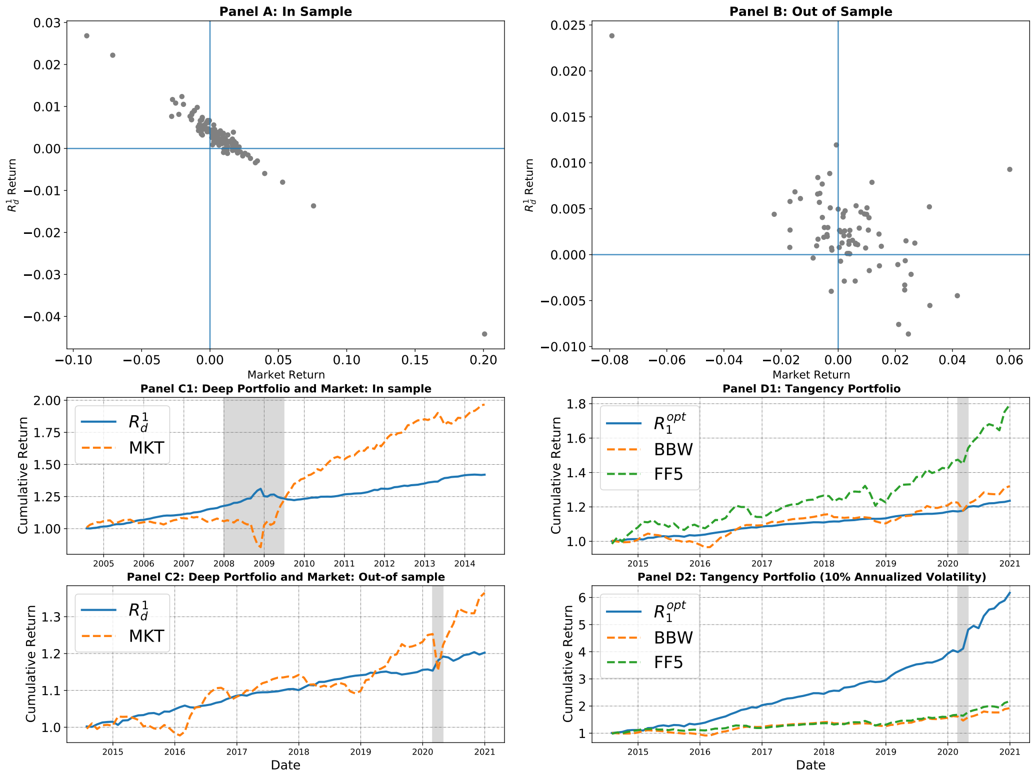
Benchmark and Competing Factors		L_1	L_2	L_3
		Panel A. In Sample Period (2004.7–2014.6)		
MKT_C	0.81	11.65***	10.39***	10.39***
BBW4	1.50	11.72***	10.45***	10.55***
FF5	0.89	3.49***	2.99***	3.52***
		Panel B. Out of Sample Period (2014.7–2020.12)		
MKT_C	0.86	2.90***	1.78***	2.38***
BBW4	1.07	2.88***	1.79***	2.42***
FF5	1.26	2.86***	1.94***	2.58***

than the deep tangency portfolio both in the in-sample and out-of-sample periods. The optimal portfolio constructed from the BBW four factors has an annualized in-sample Sharpe ratio of only 1.50 and an annualized out-of-sample Sharpe ratio of 1.07. The portfolio constructed from the Fama-French five factors has a smaller in-sample Sharpe ratio of 0.89, but a larger out-of-sample Sharpe ratio of 1.26. Note that the FF five factors contain three equity factors (MKT, SMB, and HML) plus two bond factors (Term and Default factors) as in Fama and French (1993).

What happens when we combine the competing factors and our deep factors? Table 3 also presents the Sharpe ratios of the portfolios constructed using various competing factors and a deep factor. When we combine the deep factors with BBW four factors, both in-sample and out-of-sample Sharpe ratios are very similar to those of our deep tangency portfolios. For example, the out-of-sample Sharpe ratio of the optimal portfolio constructed from the BBW four factors and the 1-layer deep factor is 2.88, which is almost the same as our tangency portfolio (2.90). Given that our deep factors are constructed by taking the bond market factor as a benchmark and using

Figure 4: Correlations and Cumulative Returns

This figure displays the deep portfolio R_d^1's return distribution and investment curve. Panel A and B are the scatter plots of bond market return and deep portfolio's return in and out of the sampling period. Panels C and D show the cumulative return of deep portfolio R_d^1 and bond market and the cumulative return of tangency portfolio R_1^{opt} of $[R_m, R_d^1]$ pairs and other competing models: BBW4, and FF5.



all firm characteristics, it should already contain non-market information of BBW factors; therefore, including those factors would not improve the Sharpe ratio over our deep tangency portfolio. In fact, we notice that the portfolio weights on the three non-market BBW factors are negligible. A notable result is that the in-sample performance of the optimal portfolio constructed from the Fama-French five factors and a deep factor is much worse; the reason is that our deep factors are constructed by taking the bond market factor, not the equity market factor, as a benchmark; however, its out-of-sample performance is still on par with our deep tangency portfolio with a Sharpe ratio of 2.86 in a one-layer neural network. Panel D1 of Figure 4 presents the cumulative returns of our 1-layer deep tangency portfolio and optimal portfolios constructed from the competing factor models in the out-of-sample period. We see that the cumulative returns of our deep tangency portfolio increase monotonically over time, and market downturns do not have any impacts on its returns; however, in spite that the cumulative returns on the competing optimal portfolios increase over time, their variations are very large, and notably, those portfolios usually suffer big losses in periods of market downturns. To further examine the performance of various portfolios, we normalize all the above optimal portfolios to have the same annual volatility of 10% and present the cumulative returns of those normalized portfolio returns. We see from Panel D2 of Figure 4 that our deep tangency portfolio, benefiting from its low volatility, has much higher cumulative returns in the out-of-sample period.

While 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, together with the market factor. Table 4 presents the performance of deep tangency portfolios constructed from the benchmark market factor and 1-3 deep factors. In-sample training suggests using more deep factors; however, the out-of-sample evidence shows that the first deep factor extracted from the one-layer neural network performs already well as the Sharpe ratio improves from using 2 or 3 deep factors is negligible and statistically insignificant. Therefore, we focus on the first deep factor from the 1-layer neural network in what follows.

4.3 Factor Spanning Regressions

The key findings up to now are that our deep factor constructed from a nonlinear combination of firm characteristics captures missing risks other than the market factor and plays a role as a hedge portfolio to market downturns. It seems that commonly used observable factors do not contain extra pricing information regarding Sharpe ratio improvement combined with the deep factor. In this part, we further examine these issues by implementing the simple factor-spanning regressions of the form,

$$R_{d,t} = \alpha + \beta' f_t + \epsilon_t, \tag{22}$$

Table 4: Multiple Deep Factors

This table presents the selection of the number of factors by the Sharpe ratios test of the MVE portfolio following Barillas and Shanken (2017). We take the sample from Jul 2004 to Jun 2014 for model training and validation, and from Jul 2014 to Dec 2020 for out-of-sample tests. We first sequentially add the number of deep portfolios in our model with three choices of the number of layers. We compare the out-of-sample Sharpe Ratio increment within the same layer (L_1,L_2 and L_3) across the different numbers of the deep portfolio (D_1,D_2 and $D3$). The deep portfolio will be taken into the model only if it has significant improvement on the mean-variance portfolio's Sharpe Ratio compared to the formal one. $***$, $**$, and $*$ denote the level of significance of 1%, 5%, and 10%, respectively.

	MKT_C	D_1	D_2	D_3
		Panel A. In Sample Period (2004.7–2014.6)		
L_1	0.81	11.65***	12.28***	12.49**
L_2	0.81	10.39***	12.73***	13.65***
L_3	0.81	10.39***	14.31***	15.17***
		Panel B. Out of Sample Period (2014.7–2020.12)		
L_1	0.86	2.90***	2.96*	2.99
L_2	0.86	1.78***	1.68	1.75**
L_3	0.86	2.38***	2.32	2.31

where f_t is a set of observable factors (e.g., BBW factors), and $R_{d,t}$ is our deep factor. Given that the 1-layer neural network performs the best in terms of maximum Sharpe ratios, we focus on the 1-layer deep factor. We also run a regression of

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

where R_t^{opt} represents the deep tangency portfolio constructed from the bond market factor and the 1-layer deep factor. Such a regression provides further evidence of whether the small number of observable factors can span the optimal portfolio.

Table 5 presents the spanning regression results for the out-of-sample period. Panel A reports alphas and betas from the spanning regressions of the deep factor and the deep tangency portfolio on the BBW four factors, respectively. We see that the BBW four factors are incapable of explaining excess returns on both the deep factor and the deep tangency portfolio: the alpha estimate is about 0.30% in the regression of the deep factor, and it is 0.24% in the regression of the deep tangency portfolio; both alpha

Table 5: Spanning Regressions

This table reports the monthly alphas in basis points and statistical significance based on Newey West standard error (Newey and West, 1994) for the bond factor spanning test. Specifically, we regress the factor and trading strategies (portfolios) in the rows against the factor models (BBW4, and FF5) in the column. R_1^{opt} is the mean-variance portfolio constructed by the bond market and the deep portfolio R_d^1. For t-statistics $*$, $**$, and $***$ indicate significance at the 10%, 5%, and 1% levels, respectively.

Panel A. BBW Four Factors							
	α	β_{MKT_C}	β_{DRF}	β_{CRF}	β_{LRF}		R^2
R_d^1	0.30	−0.17	−0.00	0.06	0.02		36.05
	(8.86)	(−3.01)	(−0.10)	(2.06)	(0.28)		
R_1^{opt}	0.24	0.06	−0.00	0.05	0.02		22.82
	(8.86)	(1.29)	(−0.10)	(2.06)	(0.28)		

Panel B. FF Five Factors							
	α	β_{MKT_E}	β_{SMB}	β_{HML}	β_{TRM}	β_{DEF}	R^2
R_d^1	0.26	0.03	−0.04	−0.02	−0.11	−0.17	50.80
	(8.62)	(1.81)	(−2.16)	(−1.40)	(−5.22)	(−3.87)	
R_1^{opt}	0.22	0.05	−0.03	−0.01	−0.00	−0.01	29.42
	(7.20)	(2.76)	(−2.04)	(−0.83)	(−0.14)	(−0.32)	

estimates are highly statistically significant. The loading of the deep factor on the bond market factor is negative and is highly statistically significant, -0.17 ($t = -3.01$), whereas the loading of the deep tangency portfolio on the bond market factor is almost zero, both further suggesting that the deep factor acts as a market-hedge portfolio.

Similar results in the spanning regressions on the Fama-French five factors (Panel B). The alpha estimate is about 0.26% ($t = 8.62$) in the regression of the deep factor and is about 0.22% ($t = 7.20$) in the regression of the deep tangency portfolio. Interestingly, we find that both the deep factor and the deep tangency portfolio load positively (though small) on the equity market factor, but negatively on the SMB (size) factor; we also find that the deep factor negatively loads on both the term and default factors, and the deep tangency portfolio does not expose to those two factors, both further suggesting that the deep factor is a bond market hedge portfolio.

4.4 Interpreting Deep Characteristics

By combining an economically well-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 types of characteristics contribute to the deep factor.

The nonlinear activation of neural networks in our methodology transforms characteristics into the deep one, which is a highly nonlinear combination of raw characteristics, whose exact functional form is unknown to us in principle. We first evaluate the linear contribution of each characteristic to the deep characteristic by running the Fama-MacBeth cross-sectional regressions (Fama and MacBeth, 1973) of the deep characteristic $Z_{i,t}^{(L)}$ on raw characteristics $z_{k,i,t}$ as Feng et al. (2022),

$$Z_{i,t}^{(L)} = a_t + b_{1,t}z_{1,i,t} + \cdots + b_{k,t}z_{k,i,t} + \cdots + b_{K,t}z_{K,i,t} + \epsilon_{i,t}, \tag{24}$$

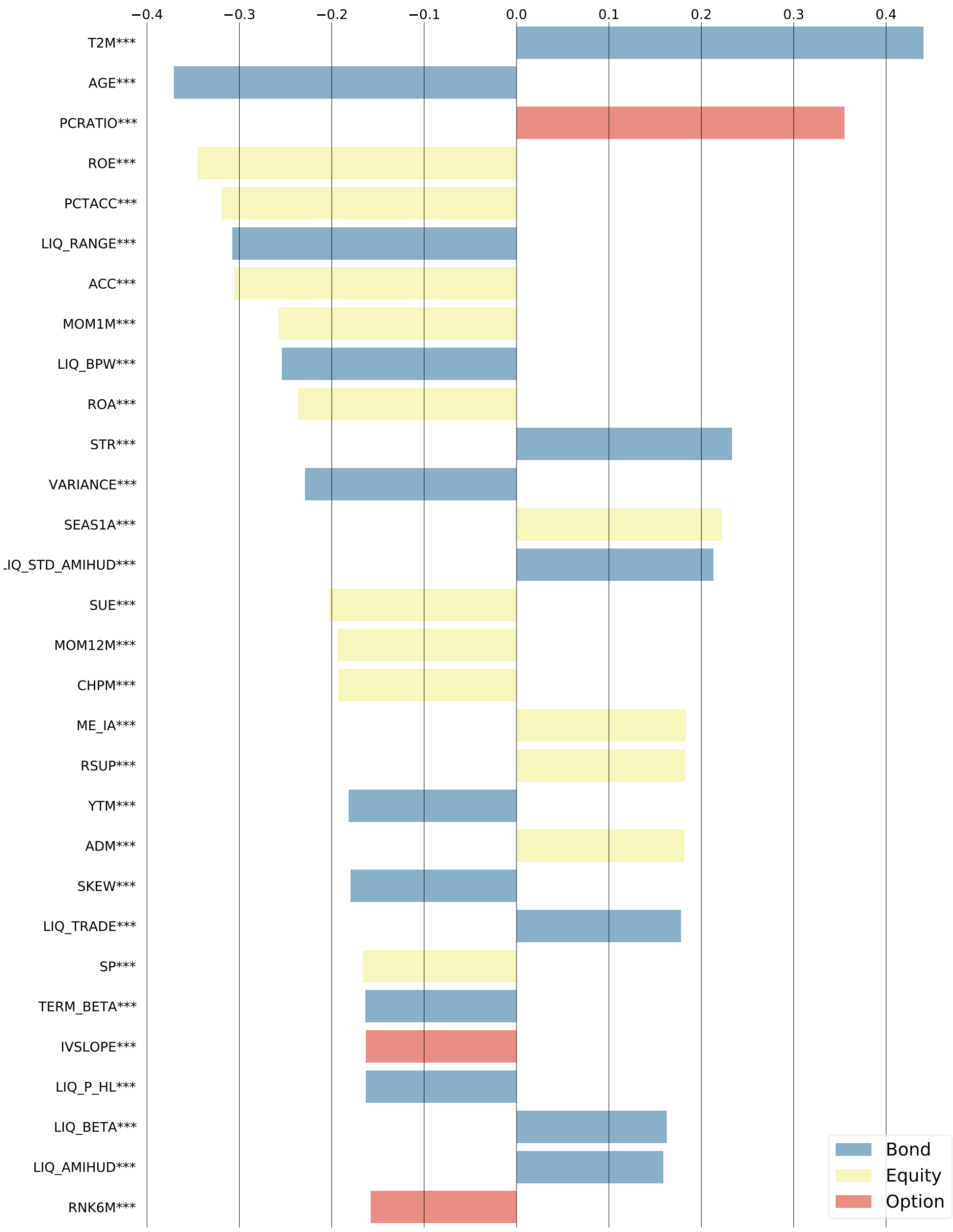
for $i = 1, \ldots, N$. Given that all characteristics are cross-sectionally normalized, we then evaluate each characteristic's contribution by the explained variation using the time-series average of $\hat{b}_{k,t}$, for $k = 1, \ldots, K$.

Figure 5 presents the top 30 most important characteristics. The bond, equity, and option characteristics are classified by the blue, yellow, and red bars, respectively. We report both the coefficient signs in brackets and significance levels. We find that the top 10 most important variables include four bond characteristics, namely, time to maturity (T2M), age (AGE), daily high-minus-low (LIQ_RANGE), and illiquidity (LIQ_BPW, Bao, Pan, and Wang, 2011), five equity characteristics, namely, return on equity (ROE, Hou, Xue, and Zhang, 2015), return on asset (ROA, Balakrishnan, Bartov, and Faurel, 2010), operating accruals (PCTACC and ACC), and one-month momentum (MOM1M), and one option characteristic, namely, put-call ratio (PCRATIO).

To further examine the importance of different types of characteristics, we reconstruct the deep tangency portfolios using bond characteristics alone or using equity and bond characteristics with option-related variables removed. Table 6 presents Sharpe

Figure 5: Variable Importance: Linear Contributions

This figure presents the variable importance via the Fama-MacBeth cross-sectional regressions of deep characteristics $Z_{i,t}^{(L)}$ on raw characteristics $z_{k,i,t}$ in the in-sample period. We report the normalized averaged coefficient $\hat{\beta}_{k,t}$ in different colors, representing the types of characteristics.



ratios obtained from different types of characteristics. Even though the in-sample training results are more or less similar, their out-of-sample performance is different. When we use all characteristics, as we have before, the one-layer neural network

Table 6: Importance of Characteristics

This table presents the affection of characteristic selection by the out-of-sample Sharpe ratios comparison of the mean-variance portfolio following Barillas and Shanken (2017). We take the sample from Jul 2004 to Jun 2014 for model training and validation, and from Jul 2014 to Dec 2020 for out-of-sample tests. We compare the one factor added mean-variance portfolio's Sharpe Ratio of neural network with different layers (L_1,L_2, and $L3$) when we have three sets of characteristics: Bond+Equity+Option, Bond+Equity, and Bond only.

	MKT_C	Bond+Equity+Option	Bond+Equity	Bond
		Panel A. In Sample Period (2004.7–2014.6)		
L_1	0.81	11.65***	11.48***	10.83***
L_2	0.81	10.39***	10.27***	11.52***
L_3	0.81	10.39***	10.76***	11.86***
		Panel B. Out of Sample Period (2014.6–2020.12)		
L_1	0.86	2.90***	1.71***	-0.38
L_2	0.86	1.78***	1.39***	1.46***
L_3	0.86	2.38***	1.91***	1.14***

works quite well, and the deep tangency portfolio earns an out-of-sample annualized Sharpe ratio of 2.90. However, when we exclude the option-related variables, we find that the deep tangency portfolio's out-of-sample Sharpe ratio reaches the largest value from the 3-layer neural network, but it is only 1.91. What is even worse is that when we use bond characteristics alone, the performance of the deep tangency portfolio further deteriorates, and its best out-of-sample Sharpe ratio is only about 1.46, from the 2-layer neural network.

To sum up, we find that all three types of characteristics are importantly weighted in deep characteristics and therefore are necessary for constructing the deep tangency portfolio. This finding is, in fact, in stark contrast to previous studies that argue those characteristics that predict equity returns do not necessarily forecast corporate bond returns (see, e.g., Chordia et al., 2017; Bali et al., 2021). But it provides further empirical evidence in support of integration between the bond and equity markets (Schaefer and Strebulaev, 2008; Kelly, Palhares, and Pruitt, 2022).

4.5 Additional Analyses

4.5.1 Latent Factors and Deep Factors

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 out-of-sample 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, Palhares, and Pruitt (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 out-of-sample split as before and extract the out-of-sample IPCA factors by fixing model parameters at the in-sample estimates.

Panel A of Table 7 presents a summary of in-sample and out-of-sample Sharpe ratios of the optimal portfolio constructed from the IPCA factors. While 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 out-of-sample 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 in-sample Sharpe ratio of 3.19 and an out-of-sample Sharpe ratio of 2.32, both of which are smaller than the corresponding values of our deep tangency portfolio (see Table 3). There are two reasons why we find such a weaker out-of-sample 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 out-of-sample IPCA factors, whereas we extract out-of-sample IPCA factors by fixing model parameters at the in-sample estimates to make

Table 7: Latent Factors and Deep Factors

This table's Panel A compares the Sharpe Ratio of the MVE portfolio between the competing factors and one factor(from 1- to 3-layer models) added competing factors. Panel B reports the factor-spanning regression similar to Table 5.

Panel A. Sharpe Ratios

	In Sample Period (2004.7–2014.6)				Out of Sample Period (2014.7–2020.12)			
	TP	L_1	L_2	L_3	TP	L_1	L_2	L_3
IPCA5	3.19	11.74***	10.51***	10.82***	2.32	2.86***	1.75	2.29
RP-PCA5	1.23	12.09***	10.90***	11.67***	0.91	2.97***	1.91***	2.59***

Panel B. Spanning Regressions

	α	β_1	β_2	β_3	β_4	β_5	R^2
B.1. IPCA Five Factors							
R_d^1	0.15	0.01	−0.01	0.07	0.08	0.23	60.17
	(4.57)	(0.12)	(−0.32)	(2.25)	(2.74)	(9.39)	
R_1^{opt}	0.11	0.05	0.08	0.05	0.12	0.15	54.85
	(4.18)	(1.09)	(2.73)	(2.10)	(4.96)	(7.77)	
B.2. RP-PCA Five Factors							
R_d^1	0.31	−0.01	−0.03	−0.02	0.02	−0.01	38.18
	(9.59)	(−0.53)	(−0.68)	(−0.46)	(0.42)	(−0.37)	
R_1^{opt}	0.24	0.02	−0.01	−0.01	0.01	−0.00	25.07
	(8.95)	(1.14)	(−0.47)	(−0.23)	(0.22)	(−0.14)	

it comparable with our methodology. We further find that including the deep factor in the IPCA five factors improves the out-of-sample Sharpe ratio of the optimal portfolio to 2.86, which is similar to that of our deep tangency portfolio; such Sharpe ratio improvement over the IPCA optimal portfolio is statistically significant.

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. Panel B presents the spanning regression results, which 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 0.15% and 0.11%, respectively, which are highly statistically significant in both regressions.

In addition, a recent paper by Lettau and Pelger (2020) proposes a risk-premium principal component analysis (RP-PCA) model for estimating latent asset pricing factors. Lettau and Pelger (2020) show that the RP-PCA performs much better than the PCA method, in particular, in identifying the weak factors. Table 7 also examines how the five RP-PCA factors perform compared to our deep factor. Again, we find that the out-of-Sharpe ratio of the RP-PCA tangency portfolio is much smaller than that of the deep tangency portfolio (0.91 vs. 2.90), and the five RP-PCA factors are unable to explain excess returns on both the deep factor and the deep tangency portfolio.

4.5.2 Importance of Nonlinearity

There are two places in our deep learning model where nonlinearity plays important roles: one is that we use a nonlinear ranking scheme to construct weights of the deep factor, and the other is that we apply a nonlinear activation function to transform raw characteristics to a deep characteristic. In what follows, we examine how important those two types of nonlinearity are in constructing the deep tangency portfolio.

Instead of relying on the softmax nonlinear ranking scheme, when constructing the deep factor, we simply follow the standard sorting approach that longs top 30% corporate bonds and shorts bottom 30% corporate bonds on the basis of individual deep characteristics. Based on such a deep factor, we construct the deep tangency portfolio as before. Panel A of Table 8 presents summary statistics of the deep factor and the deep tangency portfolio. We see that the annualized out-of-sample Sharpe ratios are 1.04 and 2.54, respectively, for the deep factor and the deep tangency portfolios, which are smaller than those obtained from using the softmax nonlinear ranking scheme (see Table 2 and Table 3).

Moreover, we examine what would happen if we remove the nonlinear *tanh* activation function and simply use a linear combination of raw characteristics in deep learning. We see from Panel B of Table 8 that without the nonlinear activation function, we need a deeper neural network and more deep factors to complement the benchmark

Table 8: Importance of Nonlinearity

This table reports the model performance of two linear models. Panel A reports the descriptive statistics similar to Table 2. Panel B presents the affection of removing nonlinear activation function by the in-sample and out-of-sample Sharpe ratios comparison of the mean-variance portfolio following Barillas and Shanken (2017). The training and validation process is the same as described in Table 4.

Panel A. Standard Sorting

	In Sample Period (2004.7–2014.6)					Out of Sample Period (2014.7–2020.12)				
	Mean	tstat	Std	SR	Max DD	Mean	tstat	Std	SR	Max DD
R_{ls}	0.32	(3.73)	0.91	1.24	7.59	0.19	(3.09)	0.62	1.04	1.98
R_{ls}^{opt}	0.40	(12.98)	0.21	6.63	0.21	0.24	(5.33)	0.33	2.54	0.54

Panel B. Linear Activation

	In Sample Period (2004.7–2014.6)				Out of Sample Period (2004.7–2014.6)			
	MKT_C	D_1	D_2	D_3	MKT_C	D_1	D_2	D_3
L_1	0.81	0.88**	9.76***	9.76	0.86	0.97**	1.32***	1.32
L_2	0.81	0.96***	9.44***	13.01***	0.86	1.08***	1.10	1.15
L_3	0.81	8.98***	10.38***	10.45	0.86	1.10***	1.04	1.03

factor and that the out-of-sample performance of the deep tangency portfolio becomes much worse, compared to the case with the nonlinear activation function.

To sum up, both types of nonlinearity in deep learning play important roles in constructing the deep tangency portfolio. Such a finding is largely consistent with what the literature has found on nonlinear effects of characteristics on expected returns (see, e.g., Freyberger et al., 2020; Gu et al., 2020; Cong et al., 2022).

5 Conclusion

The problem of constructing the optimal portfolio has economic importance: the stochastic discount factor is equivalent to the mean-variance efficient portfolio. While 70 years ago, Markowitz (1952) provides a simple solution to the optimal portfolio, it is almost practically useless in real-world situations, particularly when there are many assets. A common practice is to proxy the SDF as a function of a small number of characteristic-managed factors. However, the most commonly used factors can hardly span the mean-variance efficient portfolio of the economy, and leave many "anomalies" unexplained, leading to an issue of "factor zoo".

In this paper, we propose a parametric approach for directly estimating the optimal portfolio weights, thus sidestepping the estimation of mean returns and covariance by employing deep learning techniques. The deep tangency portfolio is a combination of the market portfolio and a zero-cost deep long-short factor constructed using a large number of characteristics. The deep factor constructed in this way plays two fundamental roles: (i) it has a low or even negative correlation with the market portfolio, thus providing us with a potential hedge portfolio; and (ii) it spans to a large extent any missing risk factors that should enter the pricing kernel.

To demonstrate 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 find that the deep tangency portfolio earns an out-of-sample annualized Sharpe ratio of 2.90, outperforming those portfolios constructed from commonly used observable factors. We further show that the recently developed latent-factor models, such as RP-PCA and IPCA, cannot span the deep factor and the deep tangency portfolio.

We further show that it is crucial to consider various types of characteristics in constructing the deep tangency portfolio. Excluding any type of characteristics would worsen the performance of the deep tangency portfolio. This evidence is in stark contrast to previous studies that argue those characteristics that predict equity returns do not necessarily forecast corporate bond returns (see, e.g., Chordia et al., 2017; Bali et al., 2021). Still, it provides further empirical evidence in support of integration between the bond and equity markets (see, 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.

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.

Balakrishnan, K., E. Bartov, and L. Faurel (2010). Post loss/profit announcement drift. *Journal of Accounting and Economics 50*(1), 20–41.

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.

Bao, J., J. Pan, and J. Wang (2011). The illiquidity of corporate bonds. *Journal of Finance 66*, 911–946.

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 (2019). Forest through the trees: Building cross-sections of stock returns. Technical report.

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 (2022). Deep learning in asset pricing. *Management Science, Forthcoming*.

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 (2022). Asset pricing with panel tree under global split criteria. Technical report, City University of Hong Kong.

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.

DeMiguel, V., A. Martin-Utrera, F. J. Nogales, and R. Uppal (2020). A transaction-cost perspective on the multitude of firm characteristics. *Review of Financial Studies 33*(5), 2180–2222.

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.

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 J. D. MacBeth (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy 81*(3), 607–636.

Feng, G., J. He, N. Polson, and J. Xu (2022). Deep learning of characteristics-sorted factor models. 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.

Giglio, S., B. Kelly, and D. Xiu (2022). Factor models, machine learning, and asset pricing. *Annual Review of Financial Economics 14*.

Green, J., J. R. Hand, and X. F. Zhang (2017). The characteristics that provide independent information about average us monthly stock returns. *Review of Financial Studies 30*(12), 4389–4436.

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

Gu, S., B. Kelly, and D. Xiu (2021). Autoencoder asset pricing models. *Journal of Econometrics 222*(1), 429–450.

Hansen, L. and R. Jagannathan (1991). Implications of security market data for models of dynamic economies. *Journal of Political Economy 99*, 225–262.

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

He, A., S. He, D. Huang, and G. Zhou (2022). Testing asset pricing models using pricing error information. Technical report, Washington University in St. Louis.

He, X., G. Feng, J. Wang, and C. Wu (2021a). Benchmarking individual corporate bonds. Technical report, City University of Hong Kong.

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

Hou, K., C. Xue, and L. Zhang (2015). Digesting anomalies: An investment approach. *Review of Financial Studies 28*(3), 650–705.

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 (2021). Downside variance premium, firm fundamentals, and expected corporate bond returns. Technical report, Fudan University.

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. 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., 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*(3), 628–650.

Lopez-Lira, A. and N. L. Roussanov (2020). Do common factors really explain the cross-section of stock returns? Technical report, University of Pennsylvania.

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

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

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

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 (1994). Automatic lag selection in covariance matrix estimation. *Review of Economic Studies 61*(4), 631–653.

Raponi, V., R. Uppal, and P. Zaffaroni (2021). Robust portfolio choice. Technical report, mperial College Business School.

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.

Appendices

A Characteristics

Table A1: Bond Characteristics

Characteristics	Description
AGE	Time since issuance in years
RATING	Rating
T2M	Time to maturity
SIZE	Amount outstanding
DUR	Duration
VAR5	Value-at-risk 5% over past 3 years
VAR10	Value-at-risk 10% over past 3 years
LIQ_BPW	Liquidity measure of Bao, Pan, and Wang (2011)
LIQ_ROLL	Roll's liquidity
LIQ_P_HL	Liquidity, high-low spread estimator
LIQ_P_FHT	Modified illiquidity measure based on zero returns
LIQ_AMIHUD	Amihud liqudity
LIQ_STD_AMIHUD	Standard devidation 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 Lin Wang Wu (2011)
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
SKEW	Skewness
KURT	Kurtosis
COSKEW	Systematic skewness
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	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: Equity Characteristics

This table lists the description for 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

Continue: Equity Characteristics

Characteristics	Description
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)
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: 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