Treasury Bond Price and Yield Curve Prediction via No Arbitrage Arguments and Machine Learning

Weiping Zhang[1], Qing Yang[2], Tingting Ye[3], Ruyan Tian[4], Weiliang Yao[5],

Liangliang Zhang[6]

Abstract

This paper proposes a novel bond return (price or yield curve) prediction methodology, unifying the classical no arbitrage pricing framework, which is ubiquitous and serves as the fundamental theoretical building block in mathematical finance, and empirical asset (bond) pricing methodologies, e.g., (Bianchi, et al., 2020) for treasury bonds and (Gu, et al., 2021) for equities. The methodology can be viewed as a unification of theoretical and empirical asset pricing frameworks. Our method is mathematically and theoretically rigorous, arbitrage-free and meantime enjoys the flexibility offered by the empirical asset pricing framework, i.e., a potentially rich factor structure and accurate function approximations via machine learning regression. Real market backtesting studies show that our predictions are accurate, in the sense that the formulated equally- or optimally-weighted treasury bond portfolios in China exchange-based markets bear significant positive returns. The hit rate for weekly yield curve prediction reaches 57.00% and the related long-only trading strategy based on the prediction results in an annualized absolute return as high as 15.85% with Calmar ratio achieving 5.17 for equally weighted portfolios. The risk-adjusted returns are even higher for mean-variance optimal portfolios. As a by-product of our prediction framework, spot

[1] Fudan University, School of Economics.
[2] Fudan University, School of Economics.
[3] University of Maine, Maine Business School.
[4] Fudan University, School of Economics. Ruyan Tian is the corresponding author is this paper.
[5] Fudan University, School of Economics.
[6] Huatai Securities, Financial Modeling Group.

yield curves can be predicted accurately in an arbitrage-free manner.

JEL Code: E12.

1 Introduction

1.1 Main Contents and Our Contributions

In this paper, we propose a novel bond asset price prediction framework, which can be extended to a generic asset that is traded in the financial market with abundant price observations and adequate liquidity. The framework combines the classical stochastic no-arbitrage pricing arguments (Black, et al., 1973) with the factor-based empirical asset pricing framework, see (Bianchi, et al., 2020) and (Gu, et al., 2021). The proposal is theoretically sound, mathematically rigorous, and empirically effective. The formulated long only treasury bond portfolios enjoy significant positive returns. Previously, theoretical, and empirical asset pricing theories are segmented with obvious gaps between them. For example, asset pricing theories, ranging from methods with simplest and least assumptions, to the methods with most sophisticated economic and financial structures, can be logically divided into three stages. As the first stage, empirical asset pricing framework aims at describing the variations of the cross-sections of asset returns with carefully chosen explanatory variables and it relies on regression techniques, e.g., (Leippold, et al., 2021), (Azevedo, et al., 2021), (Geertsema, et al., 2019), (He, et al., 2021), (Nard, et al., 2020), (Coqueret, et al., 2020), (Rapach, et al., 2022). Little theory is involved at this stage. At the second stage, no-arbitrage arguments, or preferably, the partial equilibrium analysis, are designed to price primitive and derivative assets in a relative manner. To be precise, in order to price a stock option, we need to assume exogenous dynamics, usually modeled by a stochastic differential equation system, for the underlying stock, and derive the option prices by taking the conditional expectations of the option payoff discounted by the risk-free rate under the Q-measure, or risk neutral measure. References can be found in (Black, et al., 1973) as a classical example and (DeLise, 2021), (Lotfinezhad, et al., 2021), (Lecuyer, et al., 2021), (Anderson, et al., 2022) and (Cohen, et al., 2021) as the latest developments. The mechanism guarantees that in the financial-economic system, there are no arbitrage opportunities, or more technically, no free lunch with vanishing risk (NFLVR), e.g., (Bouchard, et al., 2013). The third stage is represented by the general equilibrium asset pricing framework, see (Yang, et al., 2019), (Geng, et al., 2021), (Kubler, et al., 2021), (Guerrón-Quintana, et al., 2012) and references therein. Taking general equilibrium asset pricing theory in a pure exchange economy as an example, we assume that, in an arbitrage-free environment, the agents in the economy solve their optimization problems of making allocation decisions between savings, investment and consumption. In addition to the no-arbitrage theory, in a pure exchange economy, for example, the general equilibrium framework imposes two additional market clearing conditions, e.g., the goods and financial market clearing. The equilibrium asset prices are obtained by solving a complicated recursive equation (often PDE[7] or BSDE[8]) system in a stochastic market environment, see (Yang, et al., 2019) as a recent discussion. The general equilibrium consideration is an absolute pricing theory, in the sense that all the target financial prices are obtained simultaneously and endogenously. It is a common knowledge that a dynamic stochastic general equilibrium (DSGE) often results in linear factor structures for conditional expected asset returns, w.r.t. some exogenous and endogenous state variables as pricing factors, in equilibrium, see (Yang, et al., 2019). However, whether any empirical linear factor model is supported by a corresponding DSGE is unknown. Moreover, as the evolution

[7] PDE stands for Partial Differential Equations.
[8] BSDE stands for Backward Stochastic Differential Equations.

of empirical asset pricing theory, non-linear, especially machine learning, methods start to emerge, which adds to the difficulty in relating empirical asset pricing models to DSGE models. As a remedy, our theoretical framework brings three contributions to the literature. First, we bridge the gap between the no arbitrage asset pricing theory for primitive assets and financial derivatives, illustrated by treasury bonds as derived contracts on interest rates, and regression-based empirical asset pricing theory. Second, in terms of curve prediction, for example, spot yield curve prediction, we ensure that the predicted curves are arbitrage-free, meantime allowing a flexible factor-structure, where the factors can be any data that bear explanatory power. Third, in terms of spot yield curve fitting, using brute-force machine learning regression methods often results in exceptionally large search spaces, e.g., the universal approximation in the whole space of bounded continuous functions, see (Hornik, 1991), while our framework shrinks the search spaces greatly and finding solutions becomes simpler.

1.2 Literature Review

There are two approaches that one can model the term structure of interest rates in continuous time. The financial market equilibrium approach, pioneered by (Cox, et al., 1981), (Cox, et al., 1985) and (Cox, et al., 1985), starts from a description of the underlying economy, assumptions on the stochastic evolution of one or more exogenous factors, or, state variables in the economy and hypothesis on the preferences of a representative agent. General equilibrium considerations are used to endogenously derive the interest rate and price of all the contingent claims. The no arbitrage approach relies on assumptions on the stochastic evolution of interest rates movements and derives the prices of all contingent claims by imposing the condition that there are no arbitrage opportunities in the economy (for example, see (Vasicek, 1977), (Brennan, et al., 1979), (Langetieg, 1980), and (Artzner, et al., 1988)).

However, the aforementioned equilibrium models require estimates of the market prices of risk to price contingent claims. These quantities, being stochastic and nonstationary, are difficult to estimate. Strictly speaking, the work on "arbitrage-free" models of bond pricing was initiated by (Ho, et al., 1986). They study a discrete trading economy where bond prices fluctuate stochastically through time according to a single binomial process. Then using the martingale measure method, the authors provide arbitrage-free prices that do not explicitly depend on the "market price for risk", but rather on an exogenously specified initial forward rate curve.

In (Ho, et al., 1986), the authors work with a single factor model. One drawback of such model is that all points on the yield curve become perfectly correlated. For this reason, (Heath, et al., 1990) extends the model of Ho's by allowing for several stochastically independent random sources in the exogenous term structure. Besides this, they derived the risk-neutral description of the forward-rate process, which facilitates the estimation of the volatility parameters.

However, Markovian term structure models with normally distributed interest rates or alternatively, log-normally distributed bond prices like HJM model are far from being prefect: the discrete process chosen for estimation and computation implies negative interests rate with positive probability, and hence they are not arbitrage free in an economy with opportunities for riskless and costless storage of money. In follow-up studies, (Briys, et al., 1991) apply the Gaussian framework to derive closed-form solutions for caps, floors, and European zero-coupon bond options. To exclude the influence of negative forward rates on the pricing of zero-coupon bond options, they introduce an additional boundary condition. As shown by (Rady, 1994), these pricing formulas are only supported by a term structure model with an absorbing boundary for the forward rate at zero, where the absorbing probability is not negligible, which, for a term structure model, is quite problematic.

For these reasons, a new type of models for valuing interest rate derivatives has been developed by (Brace, et al., 1997), (Jamshidian, 1997) and (Miltersen, et al., 1997). It is usually referred to as the LIBOR and swap market model, which is based on lognormal assumption on forward rates and serves as an extension to the well-known HJM model. The main advantage of the LIBOR or swap market model over HJM is that it is easier to calibrate to the market prices of interest rate caps and European swap options. It also justifies the use of Black's formula for caplet and swaption prices, which has long been a standard market practice. Moreover, as shown by (Brace, et al., 1997), the LIBOR market model overcomes some technical difficulties, such as existence and uniqueness considerations, associated with the lognormal version of the HJM model.

The Black's formula establishes a relationship between option prices and local volatilities of the forward rates, and such relationship has enabled fast calibration of the standard model (Wu, 2003). Nevertheless, there are limitations even for the standard market model: it only generates flat implied volatility curves, whereas the implied volatility curves observed in LIBOR markets often exhibit a smile or skew. The implication is that, after being calibrated to at-the-money options, the model underprices off-the-money options. Because the benchmark role of caps and swaptions in the fixed-income derivatives markets, there have been great interests in extending the standard model to fix the problem of underpricing, or, speaking in terms of implied volatilities, to capture the smiles and skews.

From another perspective, a strand of literature strives to predict the future yield curve movements from data available at present. Literature such as (Diebold, et al., 2013) tries to use a dynamic version of Nelson-Siegel model to predict yield curve. In (Bianchi, et al., 2020), the authors apply machine learning factor model to predict treasury bond returns. But neither models exclude arbitrage opportunities.

1.3 Organization of This Paper

The organization of this paper is as following. Section 2 introduces the main methodology of this paper. Section 3 performs empirical studies and Section 4 concludes.

2 Methodology

2.1 No Arbitrage Bond Pricing

Under the risk neutral probability measure (the Q-measure hereafter), the price of a European financial derivative, whose payoff relies on a vector of primitive financial variables x, for example, market indexes, stock prices, commodity prices, interest rates or FX rates, with the maturity date T, can be evaluated by the following formula

Equation 1 $\quad V_t = E_t^Q\left[e^{-\int_t^T r_u du} V(T, x_T)\right],$

where we are standing at current time t, r is the risk-free rate and V is the payoff function of this financial product. For a pure discount bond, as the terminal payoff is always normalized to \$1, we can write

Equation 2 $\quad B_t^T = E_t^Q\left[e^{-\int_t^T r_u du}\right]$[9],

where the random process r (short rate) is modeled via a stochastic differential equation system (SDE) under the Q-measure

[9] The case of a coupon bond can be evaluated using the zero-coupon bond price expressions as building blocks.

Equation 3 $\quad dr_t = \mu^r(t, r_t, \vartheta)dt + \sigma^r(t, r_t, \vartheta)dW_t^Q, \quad r_0 = v.$

Here (μ^r, σ^r) are measurable functions, potentially nonlinear, such that the above SDE system has a unique strong solution. W_t is an n-dimensional standard Brownian motion under Q-measure and ϑ is the parameter of this SDE system. For a corporate bond, the short rate process $\bar{r}_t = r_t + s_t$, where s represents the credit spread process for this corporate bond and it is modeled, again, through an SDE system under the risk neural measure

Equation 4 $\quad ds_t = \mu^s(t, s_t, \theta)dt + \sigma^s(t, s_t, \theta)dW_t^Q.$

A zero-coupon corporate bond pricing formula can be written as

Equation 5 $\quad \bar{B}_t^T = E_t^Q\left[e^{-\int_t^T \bar{r}_u du}\right].$

This is a general short rate-based bond price equation and under the above Markov settings, the bond price solves a parabolic partial differential equation system of second order. When the riskless short rate r and credit spread process s are affine processes[10], for example, a Vasicek or a Cox-Ingersoll-Ross (CIR) process, the zero-coupon bond price has an exponential-linear structure

Equation 6 $\quad B_t^T = e^{a(t,T)+b(t,T)r_t}$

and

Equation 7 $\quad \bar{B}_t^T = e^{\bar{a}(t,T)+\bar{b}(t,T)r_t+\bar{c}(t,T)\bar{s}_t},$

where coefficients $(a, b, \bar{a}, \bar{b}, \bar{c})$ are deterministic and only depend on time (t, T) and can be solved via a Ricatti ODE[11] system. In general, the variables (r, s) cannot be directly observed in the financial market, although proxies are readily available. One way to recover these quantities is to calibrate (r, ϑ) from treasury bond data and then (s, θ) from corporate bond prices. Obtaining calibrated time-series (r, s) enables us to use factor models to predict their next-period values. The

[10] The definition of affine processes can be found in (Cuchiero, et al., 2008).
[11] ODE stands for Ordinary Differential Equation.

conditional expected log-returns of zero treasury and corporate bonds, under P-measure, can be written as

Equation 8 $\quad E_t[log(B_u^T)] = a(u,T) + b(u,T)E_t[r_u]$

and

Equation 9 $\quad E_t[log(\bar{B}_u^T)] = \bar{a}(u,T) + \bar{b}(u,T)E_t[r_u] + \bar{c}(u,T)E_t[s_u]$.

For a coupon treasury or corporate bond, we have

$$C_t^T = \sum_{i=1}^{N} c_i B_t^{T_i} + B_t^{T_N} = f(t,T,r_t)$$

and

$$\bar{C}_t^T = \sum_{i=1}^{N} c_i \bar{B}_t^{T_i} + \bar{B}_t^{T_N} = \bar{f}(t,T,\bar{r}_t).$$

For the prediction problem

$$E_t[\log C_u^T] = E_t[\log f(u,T,r_u)] = E_t\left[a(t,u,T) + b(t,u,T)r_u + e_{t,u}^Q\right]$$

$$= a(t,u,T) + b(t,u,T)E_t[r_u] + E_t\left[e_{t,u}^Q\right]^{12}$$

and

$$E_t[\log \bar{C}_u^T] = E_t\left[\log \bar{f}(u,T,\bar{r}_u)\right] = E_t\left[\bar{a}(t,u,T) + \bar{b}(t,u,T)\bar{r}_u + \bar{e}_{t,u}^Q\right] = \bar{a}(t,u,T) + \bar{b}(t,u,T)E_t[\bar{r}_u] + E_t\left[\bar{e}_{t,u}^Q\right].$$

Here $a(t,u,T)$ and $b(t,u,T)$ are linear regression coefficients of Q-simulated log-bond price $\log f(u,T,r_u)$ on r_u and likewise for $\bar{a}(t,u,T)$ and $\bar{b}(t,u,T)$. $e_{t,u}^Q$ and $\bar{e}_{t,u}^Q$ are the Q-error term and their conditional expectations under the P-measure can be evaluated by considering the Radon-Nikodym derivative. Therefore, the focus of the problem is the prediction of $E_t[r_u]$, $E_t[s_u]$ $E_t\left[e_{t,u}^Q\right]$ and $E_t\left[\bar{e}_{t,u}^Q\right]$ under P-measure[13]. Suppose that we have, under the real-world statistical

---

[12] This linear regression expansion formula can also be replaced by Ito's formula on $\log f(u,T,r_u)$ w.r.t. r_u. The readers can also try to predict every single term of the coupon bond price expression instead of using the linearization technique proposed here. The conditional expected value of a Q-Brownian motion under P measure is $E_t^P[dW_t^Q] = E_t^P[dW_t^P + \gamma_t dt] = \gamma_t dt$.

[13] In practice, we often assume that $E_t\left[e_{t,u}^Q\right] \approx 0$.

measure, determined a set of macro- and bond-specific (fundamental and technical) pricing factors f, we can rely on the non-linear regression representation, under P-measure,

Equation 10 $$r_{t_{j+1}} = \left[r_{t_j} + \left(m - kr_{t_j}\right)\Delta t\right] + \left[h\left(t_j, f_{t_j}\right)\Delta t + \epsilon_{t_j,t_{j+1}}\right].$$

Here $r_{t_j} + \left(m - kr_{t_j}\right)\Delta t$ is an auto-regressive term, projecting $r_{t_{j+1}}$ onto its lagged values and preserving the affine structure like Vasicek or CIR model

Equation 11 $$dr_t = k(m - r_t)dt + \sigma^r dW_t$$

or

Equation 12 $$dr_t = k(m - r_t)dt + \sigma^r \sqrt{r_t}\, dW_t.$$

$h\left(t_j, f_{t_j}\right)$ is the nonlinear regression term of r on factors f, with functional-form h potentially modeled through machine learning techniques. $\epsilon_{t_j,t_{j+1}}$, under P-measure, is a serially independent Gaussian innovation term by assumption[14], with zero mean. The whole expression $\epsilon_{t_j,t_{j+1}} - \left(m - kr_{t_j}\right)\Delta t + \left(m^Q - k^Q r_{t_j}\right)\Delta t + h\left(t_j, f_{t_j}\right)\Delta t$ under P-measure, can be viewed as the Gaussian process $\epsilon^Q_{t_j,t_{j+1}}$ (Brownian motion) under Q-measure. Here we refer to Girsanov theorem $dW_t^Q = dW_t^P + \gamma_t dt$, where W is the standard Brownian motion process and density γ_t induces a measure change from P to Q. In our case $\gamma_{t_j} = -\left(m - kr_{t_j}\right) + \left(m^Q - k^Q r_{t_j}\right) + h\left(t_j, f_{t_j}\right)$. Here parameters m^Q and k^Q can be recovered by the no arbitrage bond pricing formula[15] and m and k, as well as h, can be estimated by the empirical factor models. For credit spread process s, the regression relationship Equation 10 becomes a panel one, illustrated below

Equation 13 $$s^i_{t_{j+1}} = \left[s^i_{t_j} + \left(m^i - k^i s^i_{t_j}\right)\Delta t\right] + \left[g\left(t_j, f^i_{t_j}\right)\Delta t + \epsilon^i_{t_j,t_{j+1}}\right].$$

The superscript i denotes the i-th corporate bond. The above analysis is based on linear short rate

[14] This assumption means that the conditional serial correlation of short rate is zero, but it does not exclude non-zero unconditional serial correlation.

[15] Please note that, the bond pricing formula recovers Q-parameters and our conditional expected log-bond returns are evaluated under the physical P measure.

process. It is not difficult to see that the results still hold when we generalize the model space to general non-linear processes.

2.2 Yield Curve Prediction

Under Affine model assumption, the expected spot yield of a pure discount (bootstrapped) treasury bond can be expressed as

Equation 14 $$E_t[y(u,T)] = -\frac{a(u,T)}{T-u} - \frac{b(u,T)}{T-u}E_t[r_u]$$

and the expected spot yield to maturity for a corporate bond is

Equation 15 $$E_t[\bar{y}(u,T)] = -\frac{\bar{a}(u,T)}{T-u} - \frac{\bar{b}(u,T)}{T-u}E_t[r_u] - \frac{\bar{c}(u,T)}{T-u}E_t[s_u].$$

It can be observed that the shape of the expected curves $E_t[y(u,T)]$ and $E_t[\bar{y}(u,T)]$ across maturity time T, is completely determined by the functional dependency of $\left(a, b, \bar{a}, \bar{b}, \bar{c}\right)$ on T. While the intertemporal change in the level of the curve is determined by $E_t[r_u]$ and $E_t[s_u]$. When we have a set of coupon treasury bonds, we can use bootstrap method to recover the corresponding zero-bond prices and calibrate the yield curve.

2.3 Bond Portfolio Management

After we obtain the conditional expected bond price returns, we can apply the asset allocation methods documented in (Hong, et al., 2021) to create optimal bond portfolios on different assumptions on utility function (power or quadratic, etc.) and backtest the strategy in China exchange-based bond markets. In general, a bond portfolio optimization problem can be abstracted as

Equation 16 $$sup_{w\in W}[G(w,xr) - \gamma \cdot \delta(w,xr)]$$

where G is the portfolio reward, or gains process, δ is the portfolio risk which panelizes the gain process, γ denotes the risk aversion coefficient and w is the vector of portfolio weights. Space W restricts the values that w can take. As a special case of constraint W, we can, potentially, add constraints on duration and convexity of the bond portfolio. For example, in traditional mean-variance analysis, $G(w, xr) = w \cdot E[xr]$ and $(w, xr) = w \cdot COV[xr] \cdot w$, where $COV[xr]$ is the covariance matrix of the return vector xr.

2.4 Nonlinear Equity Pricing: An Illustration

To illustrate how to price an equity product, e.g., a common stock in a nonlinear model, we exploit the fundamental pricing relationship

Equation 17 $$S_t = E_t^Q\left[\int_t^{+\infty} e^{-\int_t^v r_u du} D_v dv\right] = h(t, r_t, D_t, l_t)$$

where h is a known function and the instantaneous and often hidden dividend process D is modeled by a stochastic volatility model

Equation 18 $$\frac{dD_t}{D_t} = r^d dt + \sigma^d \sqrt{l_t} dW_t$$

where

Equation 19 $$dl_t = (m^l - k^l l_t)dt + \sigma^l \sqrt{l_t} dZ_t$$

with

Equation 20 $$d\langle W, Z\rangle_t = \rho dt$$

where $(D_t, r^d, \sigma^d, l_t, m^l, k^l, \sigma^l, \rho)$ are unknowns to be recovered from high frequency (daily or minute-wise) equity prices S_t. In the same spirit, we can write, under the P-measure

Equation 21 $$D_{t+1} = D_t + r^d D_t dt + q(t, f_t)dt + \epsilon_{t,t+1}$$

and likewise, for l. This is the regression-based prediction of S_t is therefore

Equation 22 $$E_t[S_u] = E_t[h(u, r_u, D_u, l_u)].$$

To use the expectation $(E_t[r_u], E_t[D_u], E_t[l_u])$, we run a panel linear regression

Equation 23 $\quad h(u, r_u, D_u, l_u) = a(t,u) + b(t,u)r_u + c(t,u)D_u + e(t,u)l_u + \epsilon_{t,u}.$

Then, we have

Equation 24 $\quad E_t[S_u] = a(t,u) + b(t,u)E_t[r_u] + c(t,u)E_t[D_u] + e(t,u)E_t[l_u]$

where the conditional expectations $(E_t[r_u], E_t[D_u], E_t[l_u])$ are determined by AI-regressions and empirical factor data f_t.

2.5 Machine Learning

2.5.1 Artificial Neural Networks

An artificial neural network consists of different layers of neurons, including the input, hidden and output layers. The neurons of the input layer take input signals, which are processed by neurons of the hidden layers and the results are presented by the output layer.

A Back-Propagation (BP network hereafter) network is a simple neural network whose training procedure is based on back-propagation algorithm. According to the well-known universal approximation theorem (see (Hornik, et al., 1989)), as long as there is a linear output layer and at least one hidden layer with enough number of neurons, any bounded continuous function can be approximated at any level of accuracy in a compact domain of the canonical Euclidean space. The result serves as the theoretical foundation of the BP neural networks.

The BP network in this article is configurated as a classical 3-layer structure (input-hidden-output), there are 20 neurons for the hidden layer and the activation function is ReLU. The parameter estimation method is chosen as Adam (adaptive moment estimation), which is considered by us to be more reliable than the stochastic gradient descent (SGD) method.

2.5.2 XGBoost

Extreme Gradient Boosting (XGBoost) is proposed by (Chen, et al., 2016). Compared to the classical Gradient Boosting Decision Tree methods (GBDT), XGBoost can be viewed as a superior realization. XGBoost extends GBDT along the following dimensions. First, XGBoost introduces regularization to the objective function to better control the complexity of the model. Second, in addition to the gradient vector usually required by GBDT, XGBoost optimizes the parameters with both first and second order derivatives. Third, similar to Bagging, XGBoost supports feature sampling. This not only alleviates the problem of overfitting, but also reduces the time and computational power used in training.

In this paper, when we call the *XGBRegressor* method of the Python module xgboost, the learning rate is set to be 0.05, the number of decision trees is 50 and the maximum depth is 5.

3 Empirical Studies

3.1 The Data

The provisional bond information, clean and dirty prices and yield to maturity data are downloaded from the Wind terminal. Time ranges from 2015-01 to 2021-07, with 335 weeks in between and all our predictions are weekly. The treasury bonds are traded in the Shanghai and Shenzhen exchanges.

3.2 The Methodology

Consider a Vasicek short rate model in Equation 11, which can be written in an equivalent form, with a slightly abuse of notation

Equation 25 $$dr_t = (b - \beta r_t)dt + \sigma dW_t$$

The treasury bond price (Equation 2) can be expressed as (insert reference here)

Equation 26 $$a(t,T) = -\frac{1}{2}\sigma^2\left(\frac{1}{2\beta^3}\left(e^{-2\beta(T-t)} - 4e^{-\beta(T-t)} - 2\beta(T-t) + 3\right)\right) + b\,\frac{e^{-\beta(T-t)} - 1 + \beta(T-t)}{\beta^2}$$

Equation 27 $$b(t,T) = \frac{e^{-\beta(T-t)} - 1}{\beta}$$

and

Equation 28 $$B_t^T = e^{a(t,T) + b(t,T) r_t}.$$

The price of a coupon treasury bond with coupon rate C, coupon dates $\{T_i\}_{i=0}^n$ and face value \$1 can be expressed as

Equation 29 $$P_t^T = C\sum_{i=1}^n \delta_i B_t^{T_i} + B_t^{T_n}$$

where $\delta_i = T_i - T_{i-1}$ and $T_0 = t$. The price of a defaultable zero-coupon bond can be found in Chapter 7 of (Schonbucher, 2003). The first step, is to use the above equations, together with the market panel data for treasury bonds $\hat{P}(t_i, T_j)$ to recover the unknown parameters in Equation 28 via the objective function stated below

Equation 30 $$\left[\frac{1}{nm}\sum_{i=1}^n \sum_{j=1}^m \left(\hat{P}(t_i, T_j) - P(t_i, T_j)\right)^2\right]^{\frac{1}{2}}$$

where $P(t_i, T_j)$ is the model price and $\hat{P}(t_i, T_j)$ represent market observations of treasury bond prices. In what follows, we can follow the steps outlined in Section 2 to make predictions of next period bond returns and yield curves. The algorithm consists of the following steps

1) At each point in time, use treasury bond data traded in Shanghai and Shenzhen Exchange to calibrate a Nelson-Siegel curve.

2) Select key tenors along the curve and use linear regression / Lasso / ANN / XGB to predict the next period key tenor yields and fit the next-period Nelson-Siegel curve using the predicted key tenor yields.

3) Calibrate the Vasicek short-rate model to the Nelson-Siegel curve at each period and obtain the parameters $(k^r, \theta^r, \sigma^r, r_t)$.

4) After obtaining a time series of estimates of hidden short rates $\{\widehat{r_t}\}$, use machine learning models to run a one-factor regression $\widehat{r_{t+1}} = h(\widehat{r_t}) + e_{t+1}$. Therefore, $E_t[\widehat{r_{t+1}}] = h(\widehat{r_t})$ and we can use the method introduced in Section 2 to predict the next period bond returns for each treasury bonds in the pool.

5) Use the method discussed above in Section 2 to estimate the arbitrage-free yield curve.

6) Use the method documented in (Hong, et al., 2021) to preform mean-variance optimization.

7) Evaluate performance.

3.3 The Results

3.3.1 Yield Curve Prediction Results

Table 1 reports the spot YTM curve prediction errors and accuracy via root mean squared error (RMSE), mean absolute error (MAE), mean-absolute percentage error (MAPE) and hit rate (HR)[16]. The results in Table 1 report good performance of our method in predicting the yield curve movements. RMSE of all tenors are within 13bps, MAE within 8bps and MAPE is smaller than 2.50%. The average hit rate is 54.43% across all tenors and reaches 57.00% for short-end tenors. The prediction achieves better accuracy for tenors shorter than 30 years, with slightly decay for longer maturities. Globally, the

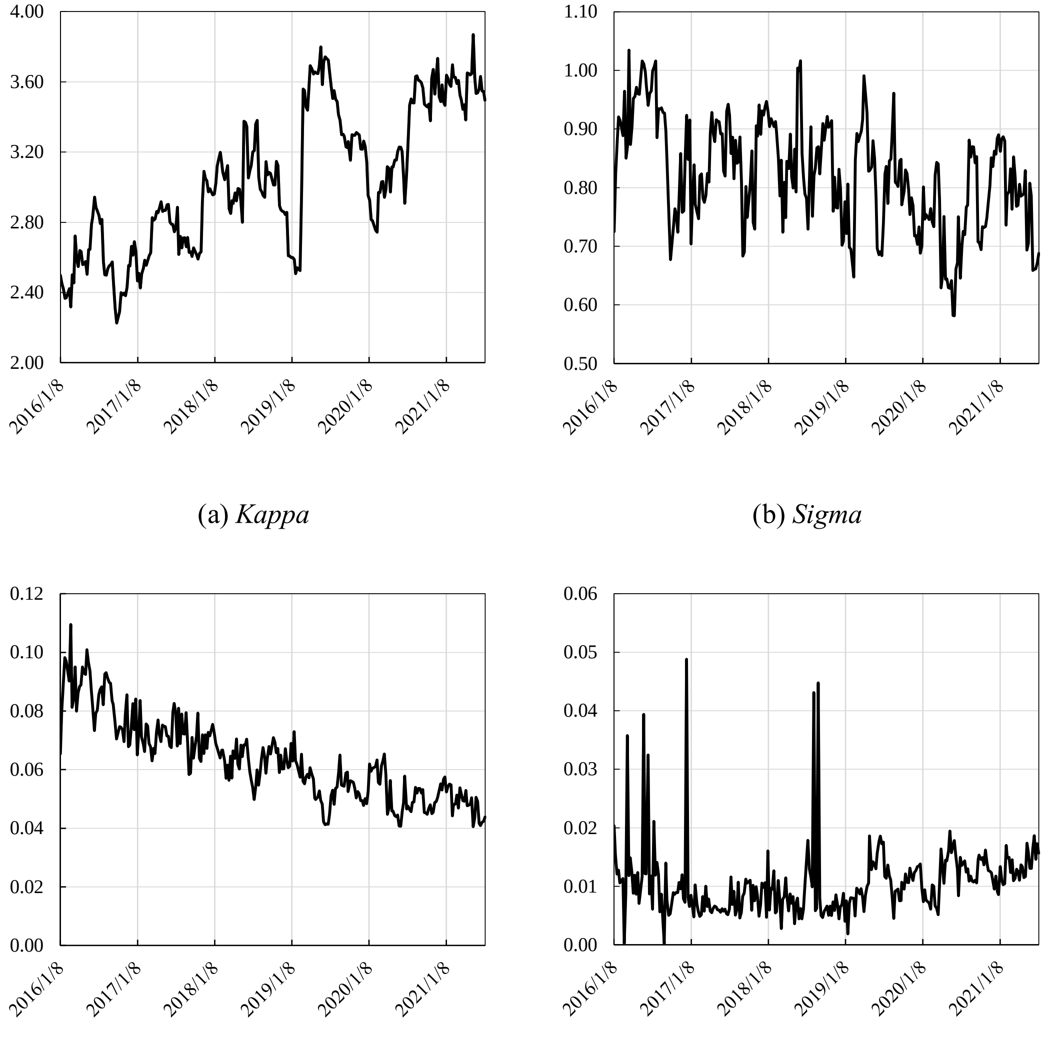
[16] All these metrics are standard in the literature.

prediction is stable and accurate.

Table 1: Accuracy Metrics for Yield Curve Predictions

Tenor	RMSE	MAE	MAPE	HitRate
0.500	0.0866	0.0563	2.30%	57.09%
0.750	0.0851	0.0555	2.24%	57.45%
1.000	0.0836	0.0548	2.19%	58.16%
1.500	0.0808	0.0534	2.09%	57.09%
2.000	0.0782	0.0522	2.00%	57.45%
5.000	0.0658	0.0459	1.60%	56.74%
7.000	0.0606	0.0433	1.44%	55.32%
10.00	0.0564	0.0411	1.29%	56.38%
12.00	0.0552	0.0403	1.22%	57.80%
15.00	0.0546	0.0396	1.16%	55.32%
17.00	0.0544	0.0393	1.12%	54.96%
20.00	0.0539	0.0389	1.08%	53.90%
25.00	0.0543	0.0395	1.08%	57.80%
30.00	0.0622	0.0455	1.23%	54.96%
31.00	0.0654	0.0475	1.29%	54.26%
32.00	0.0693	0.0498	1.35%	53.55%
33.00	0.0739	0.0523	1.42%	53.19%
34.00	0.0791	0.0552	1.51%	52.48%
35.00	0.0851	0.0584	1.60%	51.42%

36.00	0.0916	0.0620	1.70%	50.71%
37.00	0.0988	0.0658	1.82%	50.71%
38.00	0.1066	0.0699	1.94%	50.35%
39.00	0.1149	0.0743	2.08%	50.00%
40.00	0.1238	0.0789	2.22%	49.29%
Average	0.0792	0.0525	1.62%	54.43%



(a) *Kappa*

(b) *Sigma*

(c) *Theta*

(d) *Rate*

Figure 1: Dynamic Parameter Calibration for Vasicek Short Rate Models

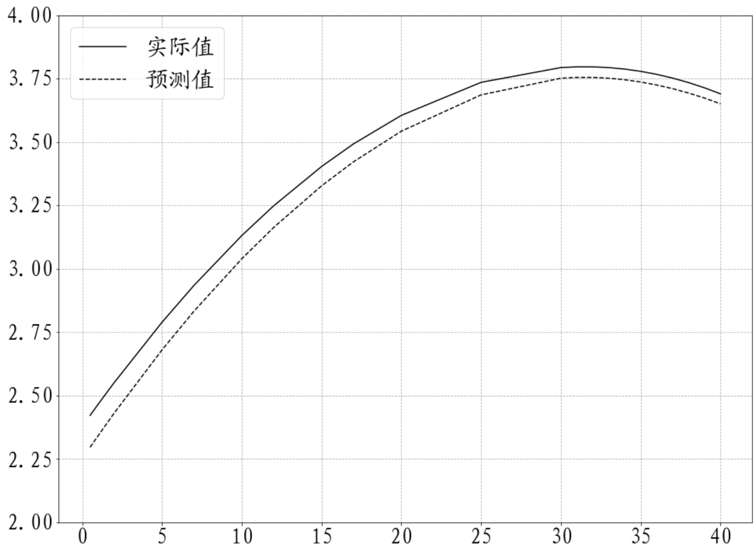


Figure 2: Performance for Yield Curve Predictions, the x-axis represents the time-to-maturity

3.3.2 Bond Portfolio Management

In China exchange-based bond markets, the bond prices are listed with clean prices and settled with dirty prices. This paper considers the performance of the prediction and portfolio management techniques in both cases. The baseline strategy is the buy and hold strategy (BH hereafter).

The bond portfolio management starts with bond prices prediction, which is done via 1) conduct no arbitrage pricing based on Vasicek short rate model; 2) prediction of short rate process based on factor models (potentially A.I. techniques); 3) plug in the predicted short rate and obtain the future yield curve estimate; 4) bond price prediction with this predicted YTM curve; 5) select the top x% bond and formulate an equally-weighted portfolio; 6) choose the portfolio weights according to the results documented in (Hong, et al., 2021).

The prediction of the short rate processes is done via Linear Regression (LR), artificial neural networks (ANN) and XGBoost (XGB) methods and we compare two portfolio construction methods:

1) the mean-variance portfolios (OptMV) solved using the numerical framework discussed above in (Hong, et al., 2021) and 2) the equally-weighted portfolio (EW). 3 prediction methodologies (LR, ANN and XGB) and 2 portfolio construction schemes (OptMV and EW) result in 6 combinations in all, written as LR-EW, ANN-EW, XGB-EW, LR-OptMV, ANN-OptMV and XGB-OptMV. For each investment strategy, we document the performance considering the clean price prediction or dirty price prediction and compare it with the baseline strategy (BH). Tables 1 and 2 present the rate of returns, volatility, Sharpe ratio, Calmar ratio and max drawdown of the considered strategies. Tables 3 to 8 report the net equity curves of the strategies based on linear regression, ANN and XGBoost.

As the accrued interest is the deterministic returns of holding a treasury bond, the prediction on the clean price change is more meaningful and its performance better measures the accuracy of our forecasting methodologies. The BH strategy earns -0.60% during the investment period. The facts prove that our methodologies enhance the returns greatly, with LR-EW 8.00%, ANN-EW slightly better. XGB-EW improves other EW strategies by significant margin. The Sharpe ratio is as high as 3.03, and the Calmar is 2.83. This means that XGBoost algorithm can better capture the instantaneous nonlinearities in bond returns and provides better extrapolation abilities.

Based on the predictions, portfolio management can effectively diversify risks and enhance portfolio returns. Based on a companion paper (Hong, et al., 2021), we dynamically adjust portfolio weights of the selected bonds at each period of time and found that the portfolio management yields 4 times the rate of returns compared to the case without portfolio choice. The advantage of nonlinear methods is even more significant in this case. ANN-OptMV yields 53.47%, XGB-OptMV achieves 128.62% in the whole out-of-sample period. From risk-adjusted return point of view, XGB-EW boasts a Sharpe ratio of 3.03. However, XGB-OptMV has highest Calmar ratio of 5.17. This further indicates a better advantage of XGBoost algorithm over ANN.

This paper further considers accrued interest, which represents the actual earnings of our strategies. After considering the accrued interests, the BH strategy yields 3.30%. All our three strategies report even higher rate of returns in this case. XGB-OptMV earns an annualized return of 19.07% and is the best among the three strategies. Two XGB strategies, XGB-EW and XGB-MVOpt, yield Calmar ratios as high as 7.70 and 7.62, which means the two strategies effectively lower the max-drawdowns meanwhile achieve high rate of returns. To conclude, the no-arbitrage and factor model-based prediction and pricing framework, applied to the field of optimal investment, is working perfectly.

Table2: Bond Portfolio Performance (Clean Price)

	TR	AR	Vol	SR	CR	MDD
B&H	-0.60%	-0.11%	0.80%	-0.17	-0.03	4.72%
LR-EW	8.00%	1.38%	0.80%	1.68	0.87	1.55%
ANN-EW	8.77%	1.51%	0.83%	1.78	0.86	1.70%
XGB-EW	22.99%	3.75%	1.21%	3.03	2.83	1.29%
LR-OptMV	46.17%	6.99%	3.55%	1.79	2.71	2.35%
ANN-OptMV	53.47%	7.92%	3.73%	1.94	2.72	2.67%
XGB-OptMV	128.62%	15.85%	10.28%	1.43	5.17	2.85%

注：TR、AR、Vol、SR、CR、MDD represent total return, annualized return, annualized volatility, Sharpe ratio, Calmar ratio and maximum drawdown, respectively.

Table2: Bond Portfolio Performance (Dirty Price)

	TR	AR	Vol	SR	CR	MDD
B&H	20.05%	3.30%	0.80%	4.03	2.78	1.15%
LR-EW	30.45%	4.84%	0.80%	5.88	6.04	0.78%
ANN-EW	31.47%	4.99%	0.89%	5.42	5.56	0.87%
XGB-EW	48.13%	7.24%	1.19%	5.85	7.70	0.90%
LR-OptMV	71.38%	10.06%	3.58%	2.56	3.47	2.65%
ANN-OptMV	76.94%	10.69%	3.54%	2.75	4.08	2.39%
XGB-OptMV	166.71%	19.07%	10.24%	1.71	7.62	2.30%

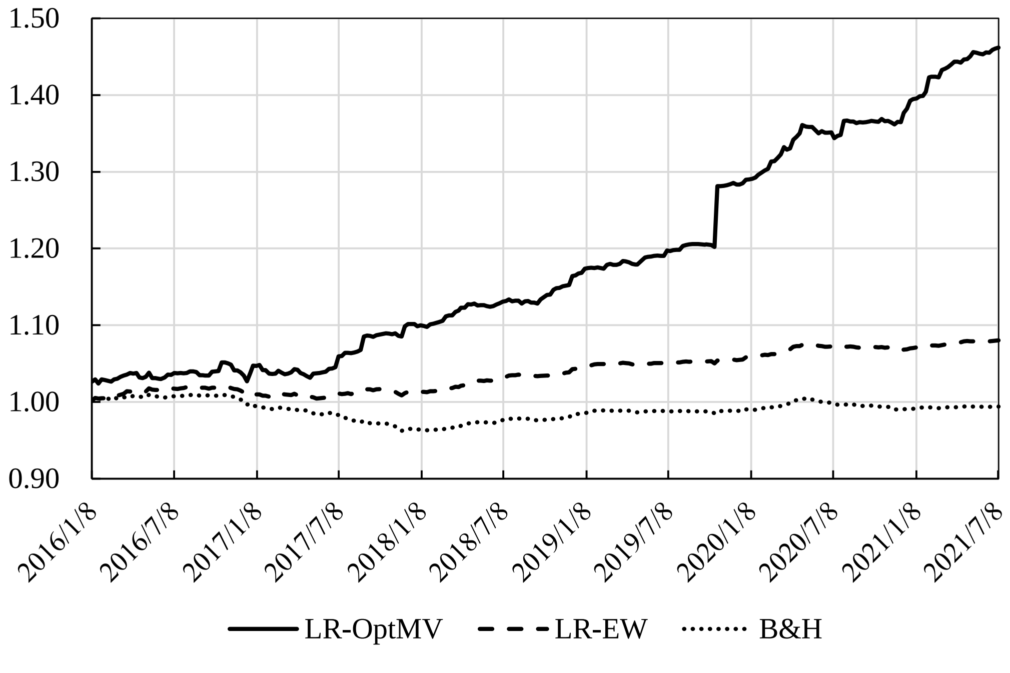


Figure 3: Net Equity Curve for LR Bond Portfolio Management (Clean Price)

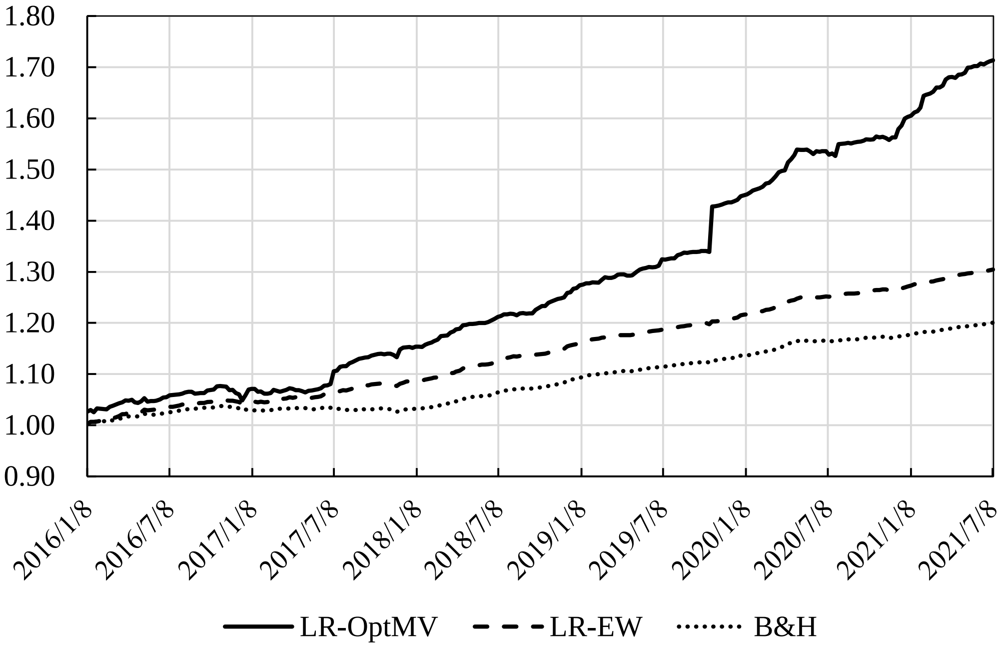


Figure 4: Net Equity Curve for LR Bond Portfolio Management (Dirty Price)

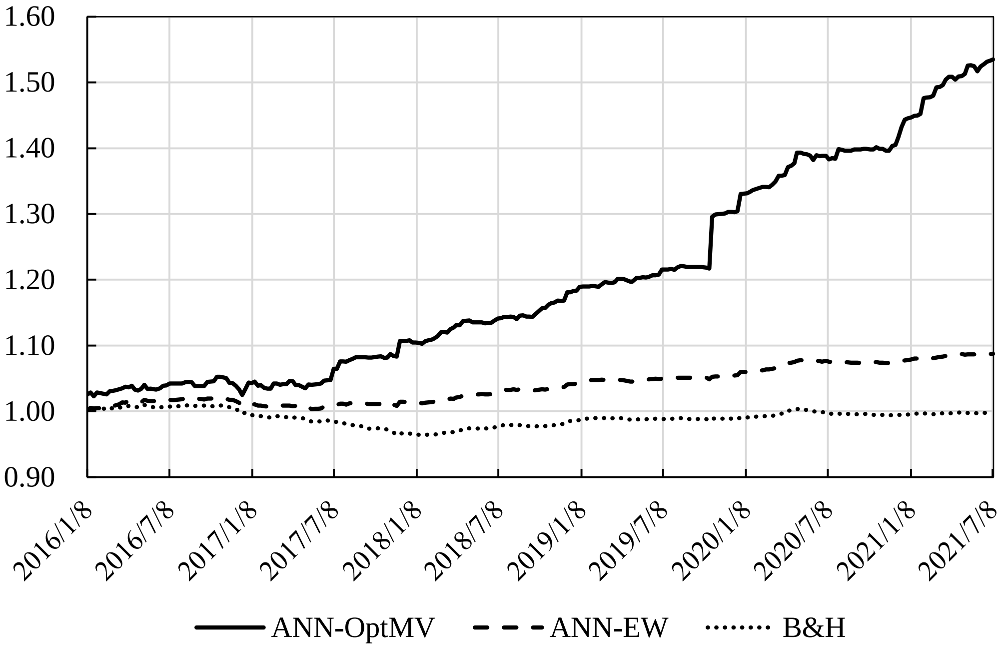


Figure 5: Net Equity Curve for ANN Bond Portfolio Management (Clean Price)

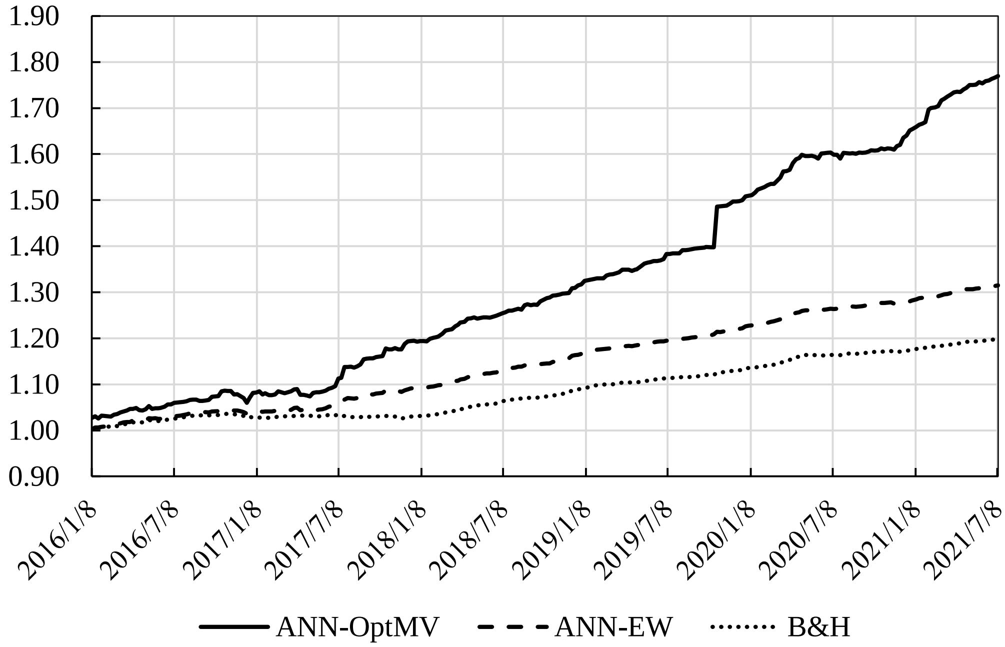


Figure 6: Net Equity Curve for ANN Bond Portfolio Management (Dirty Price)

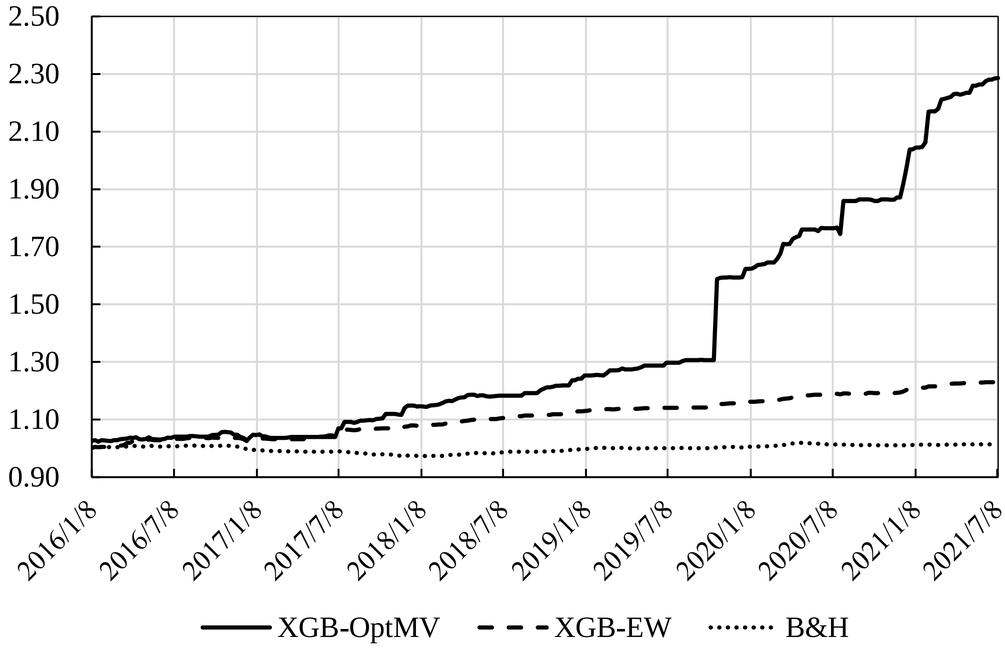


Figure 7: Net Equity Curve for XGB Bond Portfolio Management (Clean Price)

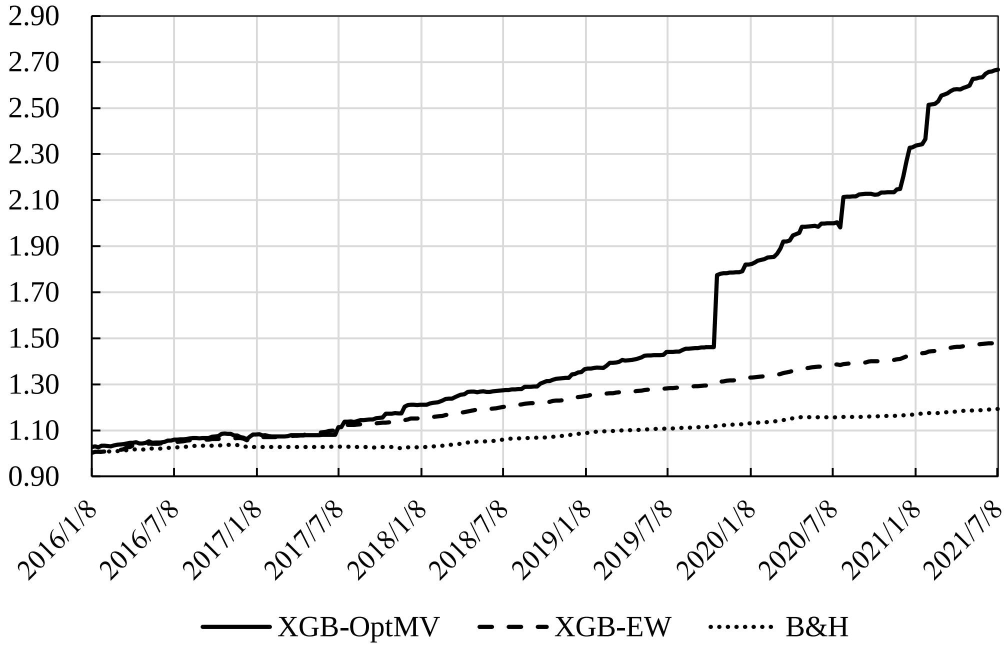


Figure 8: Net Equity Curve for XGB Bond Portfolio Management (Dirty Price)

4 Conclusion

In this paper, we document a novel framework of bond pricing and yield curve prediction methodology, unifying the traditional and theoretical no arbitrage pricing theory and machine learning factor models. The framework turns out to be effective in predicting yield curve movements in China's exchange-based treasury markets. A direct extension of the results in this paper is to consider the prediction of credit spread and therefore the corporate bond returns. In addition, the same methodology can be applied to any financial product, whose pricing relies on no arbitrage arguments and factor models.

5 References

Anderson, David and Ulrych, Urban. 2022. Accelerated American Option Pricing with Deep Neural Networks. *SSRN.* 2022.

Artzner, F. and Delbaen, F. 1988. *Term Structure of Interest Rates: The Martingale Approach.* Strasbourg : IRMA, 1988.

Azevedo, Vitor and Hoegner, Christopher. 2021. Enhancing Stock Market Anomalies with Machine Learning. *SSRN.* 2021.

Bianchi, Daniele, Buchner, Matthias and Tamoni, Andrea. 2020. Bond Risk Premia with Machine Learning. *SSRN.* 2020.

Black, Fischer and Scholes, Myron. 1973. The Pricing of Options and Corporate Liabilities. *Journal of Political Economy.* 1973, Vol. 81, 3, pp. 637--654.

Bouchard, Bruno, Lepinette, Emmanuel and Taflin, Erik. 2013. Robust no-free lunch with vanishing risk, a continuum of assets and proportional transaction costs. *ArXiv.* 2013.

Brace, A. G., Gatarek, D. and Musiela, M. 1997. The Market Model of Interest Rate Dynamics. *Mathematical Finance.* 1997, Vol. 7, 2, pp. 127--155.

Brennan, M. J. and Schwartz, E. S. 1979. A continuous time approach to the pricing of bonds. *Journal of Banking and Finance.* 1979, Vol. 3, 2, pp. 133-155.

Briys, E., Crouhy, M. and Schoebel, R. 1991. THe Pricng of Default-Free Interest Rate Cap, Floor and Collar Agreements. *The Journal of Finance.* 1991, Vol. 46, 5, pp. 1879--1892.

Chen, T. and Guestrin, C. 2016. XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco.* 2016.

Cohen, Samel, Reisinger, Christoph and Wang, Sheng. 2021. Arbitrage-Free Neural-SDE Market Models. *SSRN.* 2021.

Coqueret, Guillaume and Deguest, Romain. 2020. Predictive Regressions: A Machine Learning

Perspective. *SSRN.* 2020.

Cox, J. C., Ingersoll, J. E. and Ross, S. A. 1985. A theory of the term structure of interest rates. *Theory of Valuation.* 1985, Vol. 53, 2, pp. 385-407.

—. 1985. An intertemporal general equilibrium model of asset prices. *Econometrica.* 1985, Vol. 53, 2, pp. 363-384.

Cox, J. C., Intersoll, J. E. and Ross, S. A. 1981. The relation between forward prices and futures prices. *Journal of Financial Economics.* 1981, Vol. 9, 4, pp. 321-346.

Cuchiero, Christa, Filipovic, Damir and Teichmann, Josef. 2008. Affine Models. *arxiv.* 2008.

DeLise, Timothy. 2021. Neural Options Pricing. *ArXiv.* 2021.

Diebold, Francis and RudeBusch, Glenn D. 2013. *Yield Curve Modeling and Forecasting: The Dynamic Nelson-Siegel Approach.* s.l. : Princeton University Press, 2013.

Geertsema, Paul and Lu, Helen. 2019. Machine Valuation. *SSRN.* 2019.

Geng, Runjie and Zhang, Ally. 2021. Recursive equilibria in dynamic economies with bounded rationality. *SSRN.* 2021.

Gu, Shihao, Kelly, Bryan and Xiu, Dacheng. 2021. Empirical Asset Pricing via Machine Learning. *Review of Financial Studies.* 5 2021, Vol. 33, 5, pp. 2223--2273.

Guerrón-Quintana, Pablo and Nason, James M. 2012. Bayesian Estimation of DSGE Models. *SSRN.* 2012.

He, Xin, et al. 2021. Asset Pricing with Panel Trees under Global Split Criteria. *SSRN.* 2021.

Heath, D., Jarrow, R. and Morton, A. 1990. Bond pricing and the term structure of interest rates: A discrete time approximation. *JFQA.* 1990, Vol. 25, 4, pp. 419-440.

Ho, T. S. and Lee, S. B. 1986. Term Structure Movements and Pricing Interest Rates Contingent

Claims. *Journal of Finance.* 1986, Vol. 41, 5, pp. 1011-1029.

Hong, Zhenning, et al. 2021. Asset Allocation via Machine Learning. 2021.

Hornik, Kurt. 1991. Approximation capabilities of multilayer feedforward networks. *Neural Networks.* 1991, Vol. 4, 2.

Hornik, Kurt, stinchcombe, Mexwell and white, Halbert. 1989. Multilayer feedforward networks are universal approximators. 1989, Vol. 2, 5.

Jamshidian, F. 1997. LIBOR and Swap Market Models and Measures. *Finance and Stochastics.* 1997, Vol. 1, 4, pp. 293--330.

Kubler, Felix and Scheidegger, Simon. 2021. Uniformly Self-Justified Equilibria. *SSRN.* 2021.

Langetieg, T. C. 1980. A multivariate model of the term structure. *Journal of Finance.* 1980, Vol. 35, 1, pp. 71-97.

Lecuyer, Emy and Martins-da_Rocha, Filipe. 2021. Convex Asset Pricing. *SSRN.* 2021.

Leippold, Markus, Wang, Qian and Zhou, Wenyu. 2021. Machine-Learning in the Chinese Stock Market. *SSRN.* 2021.

Lotfinezhad, Mahdi and Lee, Brian. 2021. Unified Framework for Pricing, Return, and Risk of Assets. *SSRN.* 2021.

Miltersen, K. R., Sandmann, K. and Sondermann, D. 1997. Closed-Form Solutions for Term Structure Derivatives with Log-Normal Interest Rates. *Journal of Finance.* 1997, Vol. 52, 1, pp. 409--430.

Nard, Gianluca De, Hediger, Simon and Leippold, Markus. 2020. Subsampled Factor Models for Asset Pricing: The Rise of Vasa. *SSRN.* 2020.

Rady, S. 1994. The direct approach to debt option pricing. *Review of Futures Markets.* 1994, Vol.

2, pp. 416--515.

Rapach, David E. and Zhou, Guofu. 2022. Asset Pricing: Time-Series Predictability. *SSRN.* 2022.

Schonbucher, P. 2003. *Credit Derivatives Pricing Models: Models, Pricing and Implementation.* 2003.

Schonbucher, Philipp. *Credit Derivatives Pricing Models.* s.l. : Wiley Finance.

Vasicek, O. 1977. An Equilbrium Characterization of the Term Strcture. *Journal of Finance.* 1977, Vol. 35, 1, pp. 71-97.

Wu, L. 2003. Fast at-the-money Calibration of LIBOR Market Model Through Lagrangie Multipliers. *Journal of Computational Finance.* 2003, Vol. 6, 2, pp. 39--77.

Yang, Qing, Ye, Tingting and Zhang, Liangliang. 2019. A Unified Theory of Asset Pricing. *SSRN.* 2019.