Submitted to *Management Science*
manuscript MS-0001-1922.65

Swap Spread Term Structure Modeling and Forecasting by Manifold Learning

Chi Zhang
School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, czhang611@gatech.edu
Supervised by Prof. Xiaoming Huo

This paper proposes a nonparametric approach for the modeling and forecasting of weekly interest rate spread curves by using nonlinear dimension reduction, such as the locally linear embedding (LLE). We mainly focus on the swap spread (LIBOR substract Treasury). Benchmarking on its linear dimension reduction counterparty – principle component analysis (PCA) – we show the LLE-based framework yields a higher out-of-sample forecast accuracy for specific underlying tenors as well as a better profit and loss (PnL) profile in backtesting various systematic term structure trading strategies.

Key words: swap spread curve, short-rate model, manifold learning, systematic trading
History: This paper was first submitted on July 28, 2021 and has been with the authors for 1 years.

1. Introduction

In the fixed-income trading industry, there is an increasing demand for accurately forecasting the short-term movement of spread term structure in a neat but efficient way as the low interest rate environment persists to squeeze the spread magnitude among multiple key interest rates curves. Besides, The Alternative Reference Rates Committee (ARRC) has identified the Secured Overnight Financing Rate (SOFR) as the successor rate of LIBOR, and continues supporting the launch of SOFR-based financial products in coming years. Thus, during the process of LIBOR's fallback, market participants are imperative to measure the swap spread when USD LIBOR-based activity gradually decreases until completely unusable.

In general, modeling the dynamics of a high-dimensional curve analytically is daunting. Therefore practitioners are motivated to apply a variety of dimension reduction and predictive models in this area. The widely recognized use of such methods is limited to principal component and latent factor analysis, because more complex models may not be suitable for structural analysis and parameter interpretation. However, as a forecasting problem instead of an identification problem, the main goal of modeling the time variation of interest rate spread curves should be pursuing a stronger out-of-sample forecast power. As PCA is mainly designed for extracting linear factors of a data set, it is not suitable to capture the most intriguing part of the curves. Hence, we explore the manifold learning methods (e.g. LLE), which are designed to analyze intrinsic nonlinear structures of high-dimensional curves. The LLE starts with finding the K nearest neighbors for each sample curve, then outputs the optimal local convex combinations of the K nearest neighbors to represent each original sample. We believe the large adaptivity introduced by the novel idea of "local convex representations" could improve the forecast power of swap spread curves.

Specifically, we propose the following multistage approach for the short-term prediction of swap spread changes following the work by Chen, Deng and Huo (2008) (1): (a) dimension reduction (b) forecasting in the reduced dimension (c) mapping back to the original space. Forecasting time variation requires a careful approximation of an unknown encoder that maps information from high-dimensional (long spectrum of underlying maturities) to low-dimensional representations (latent drivers of the entire term structure). After obtaining such low-dimensional representation of spread curves, forecasts could be made by first predicting each new coordinate of the manifold using the ARMA model (or nonlinear time series models) and then map them back to the high-dimensional space utilizing the corresponding reconstruction method.

The performance evaluation of modeling the objective curve includes both statistical accuracy and the profitability of term structure relative-value trading strategies. We found that benchmarked against the well-known PCA-based framework, the LLE-based framework delivers a higher out-of-sample forecast accuracy for the spread with short to mid tenors, where most fixed-income market participants put their main focus on in their asset pricing and risk management. Besides, the profit and loss (PnL) of our framework also outperforms the benchmark in all three relative-value trading strategies.

2. Background

2.1. Term Structure Modeling

Regarding the debate that how much predictability of curve movements one could capture using simply historical panel data, we confirm the superior out-of-sample forecasting performance of short-term interest rate dynamics if we model the spread curve as a whole.

Inspired by the idea of summarizing term structures by a small set of linear combinations of yields, Diebold and Li (2006) (9) use the AR family models to obtain encouraging results for long-horizon ex-ante forecasts by reformulating the Nelson and Siegel (1987) model. Yet considering the substantial information about future curve dynamics (specifically the long-end tenors) embedded in the macroeconomic variables, researchers have also tried to extract macroeconomic information as a set of latent factors, then add these exogenous variables (e.g. real activity, inflation, and fed funds rate) into the term structure modeling framework, see Ang and Piazzesi (2003) (10), Cooper and Priestley (2008) (11), and Ludvigson and Ng (2009) (12). Though a large part of the term structure model specification has been deployed, a unified conclusion concerning the factor selection has not been achieved yet.

2.2. Swap Spreads

A large body of research has explored the potential factors that could explain the variation in swap spreads. There are three candidate sources: (a) the creditworthiness of the counter-parties (b) the risk premium (c) the liquidity premium.

First, the creditworthiness of counter-parties has driven the swap spreads in the early stage of the swap market, see Bollier and Sorensen (1994) (13). This may not hold due to the collateral agreement introduced in the current industry practice. Under this agreement, both parties net out their swap positions and impose collateral against each other based on the daily net mark-to-market values. That is to say, an investor receiving fixed in a swap agreement is equivalent to hold a long position in a default-free government bond while financing by paying LIBOR.

After ruling out the impact of counter-party worthiness, we focus on the implicit risk premium associated with the floating leg in a swap contract. Considering an auxiliary investment strategy: buying a government bond with the same maturity as the swap while financing the purchase via a repurchase agreement (repo hereafter). Because the creditworthiness of the banks involving in determining LIBOR varies significantly, the LIBOR is usually higher than the repo rate. Consequently, it is expected that fixed-leg receivers be awarded a swap rate that is higher than the treasury rate. Since the Secured Overnight Financing Rate (SOFR) is a proxy for the repo rate, we select the spread of 3-month LIBOR over 3-month SOFR as our risk premium in this paper.

Finally, many fixed-income investors are mandated to invest most of their portfolios in government bonds, but they are prohibited from using swaps as an investment pool under certain circumstances. Thus, the liquidity premium comes into our scope due to the demand difference between the two markets. According to Grinblatt (2001) (14) and Yu (2014) (16), the liquidity premium directly influences the short-term interest rates from which swaps and government bonds are priced and consequently determines the swap spread. Among several proxies for the liquidy premium, we choose the spread of the fed funds rate (FFR) over the inflation rate. As one of the most influential interest rates in the US, the FFR is the rate at which a depository institution lends funds to another overnight. The higher the difference, the less the liquid in the swap market.

2.3. Dimension Reduction

Focused on the context of dimension reduction techniques, Principle Component Analysis (PCA) has been applied in the field of term structure modeling for decades, such as Litterman and Scheinkman (1991) (15). Practitioners usually interpret the principal components as level, slope, and curvature effects. However, there is some unavoidable limit associated with the PCA approach, for example, encoding a great amount of information associated with curve shapes into a covariance matrix. This will miss a great amount of predictive information since the behavior of interest rates has been shown to depend on the absolute level of rates. Thus, to better capture the nonlinear relationship among neighbor tenors, Kondratyev (2018) (4) proposes a neural networks based term structure algorithm for brent oil forward price and USD swap curve.

While the above-mentioned term structure modeling has been well developed in the Treasury markets, to the best of our current knowledge, advances in modeling the spread term structure using nonlinear dimension reduction methods are comparatively small. Inspired by the work of Kondratyev, we believe the heterogeneity of temporal evolution will be better captured by manifold learning combined with adaptive time series forecasting.

The rest of the paper is organized as follows. Section 3 starts from the economic meaning of interest rate variables, then provides a structured formulation on building the required curves and implementing nonlinear dimension reduction, finally outlines the times series forecasting architectures and trading strategies design. Section 4 first describes the methodology of simulating swap spread curve datasets using a residual-based short-rate model, then shows the results of the empirical analysis. Section 5 concludes and discusses future work.

3. Problem Formulation

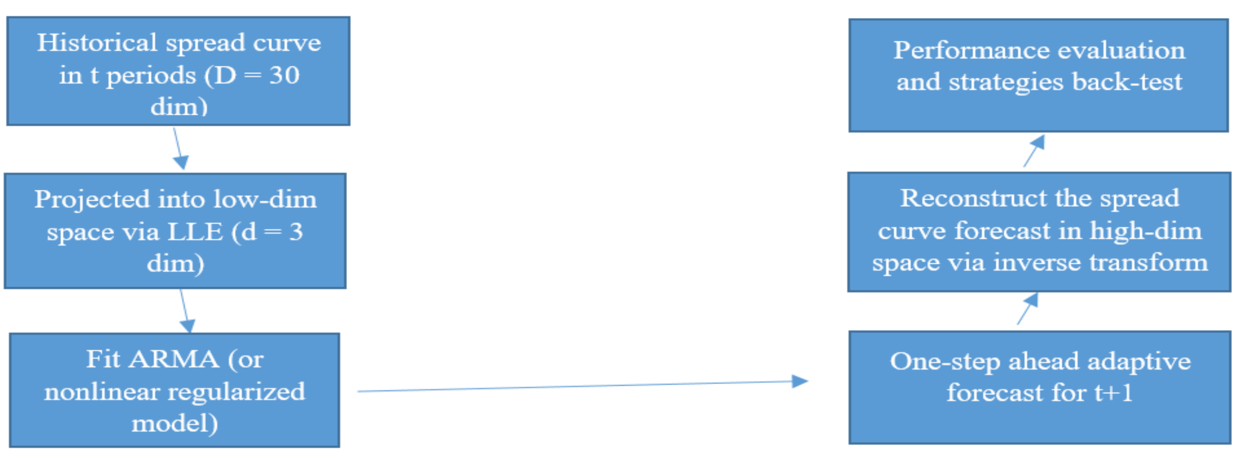

Figure 1 **Conceptual flowchart of the framework**

The panel dataset of LIBOR rate is in the form of $L_1, L_2, \dots, L_n$, where n denotes the number of time periods and each $L_i \in R^D$. Similarly, given the panel dataset of Treasury rate T_1, $T_2, \dots, T_n$, we first calculate the required spread dataset: swap spread $SS_i = L_1 - T_1, L_2 - T_2, \dots, L_n - T_n$. Then, by using a pre-specified dimension reduction technique, we obtain their low-dimensional representations $SSlow_i$, where each sample is in the R^d space for i from 1 to n. For the objective spread, we perform a one-step ahead prediction using univariate time series modeling to get $SSlow_{n+1}$. Finally, map them back to the original R^D space using the corresponding inverse transform to get $\hat{SS}_{n+1}$, then evaluate the forecasting accuracy and design trading strategies according to the comparison with realized value SS_{n+1}.

3.1. Curves Building

Considering the purpose of understanding the treasury curve's fundamental determinants, we employ a parametric yield curve specification described in Svensson (1994). As shown below, this specification could largely rule out variation resulting from a few specific securities at a given maturity.

From the website of Fed Reserve Bank of New York, we get the functional parameters of the Nelson-Siegel-Svensson structural models, assuming that instantaneous forward rates n years ahead are characterized by a continuous function with four parameters: $\beta_0 + \beta_1$ measures the initial level at horizon zero, while the asymptote level will be β_0. β_2 and β_3 determine the convexity of two humps located in between the entire maturity spectrum.

Empirically, using the calibrated parameters as shown in equation (1) and (2), we could construct the zero (forward) rate curve by plugging the maturity index into the interest rate parametric expression.

Equation (1):

$$y_t(n) = \beta_0 + \beta_1 \frac{1 - exp(\frac{-n}{\tau_1})}{\frac{n}{\tau_1}} + \beta_2 (\frac{1 - exp(\frac{-n}{\tau_1})}{\frac{n}{\tau_1}} - exp(\frac{-n}{\tau_1})) + \beta_3 (\frac{1 - exp(\frac{-n}{\tau_2})}{\frac{n}{\tau_2}} - exp(\frac{-n}{\tau_2}))$$

Equation (2):

$$f_t(n, 0) = \beta_0 + \beta_1 exp(\frac{-n}{\tau_1}) + \beta_2 \frac{n}{\tau_1} exp(\frac{-n}{\tau_1}) + \beta_3 \frac{n}{\tau_2} exp(\frac{-n}{\tau_2})$$

The data format should be in daily frequency with a continuous maturity spectrum from overnight to 30y (360 dimensions). To make it consistent with the other swap curves, we downsample it to get a weekly dataset with a discrete maturity spectrum (30 dimensions).

For the LIBOR, SOFR, and OIS curves, we directly download them from the Bloomberg terminal by manually changing the "As of Date" variable. Specifically, the OIS curve is used as the risk-free interest rate when discounting future payoff in our systematic trading.

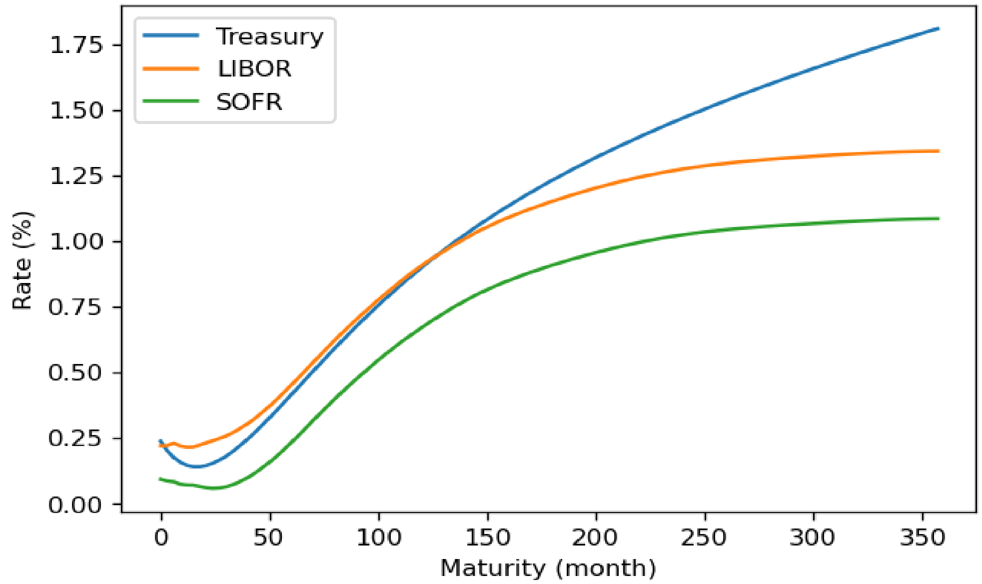

Figure 2 Original interest rate curves on 2020-10-30

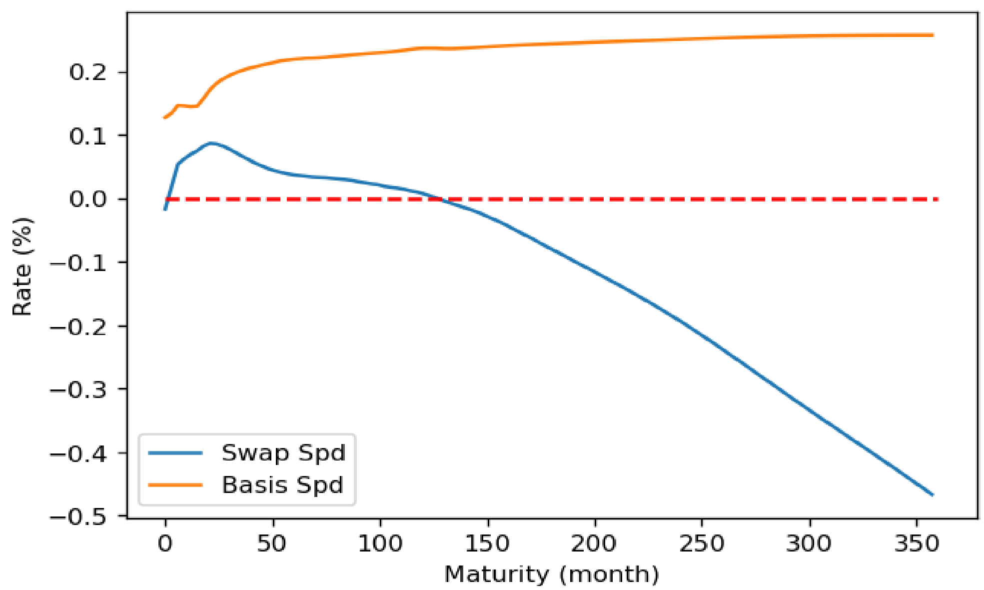

Figure 3 Interest rate spread curves on 2020-10-30

As shown in Fig 2, on 2020-10-30, all of the three curves are upward sloping. The LIBOR and SOFR curve are quite similar in this cross-sectional view, while the Treasury curve starts below the LIBOR curve at short-end and then crosses over the LIBOR curve around mid to long tenors.

As shown in Fig 3, on 2020-10-30, the basis spread curve (LIBOR substract SOFR) is quite flat and constantly positive, while the swap spread shows a hump shape at short end and sharply goes down into the negative territory as tenor increases.

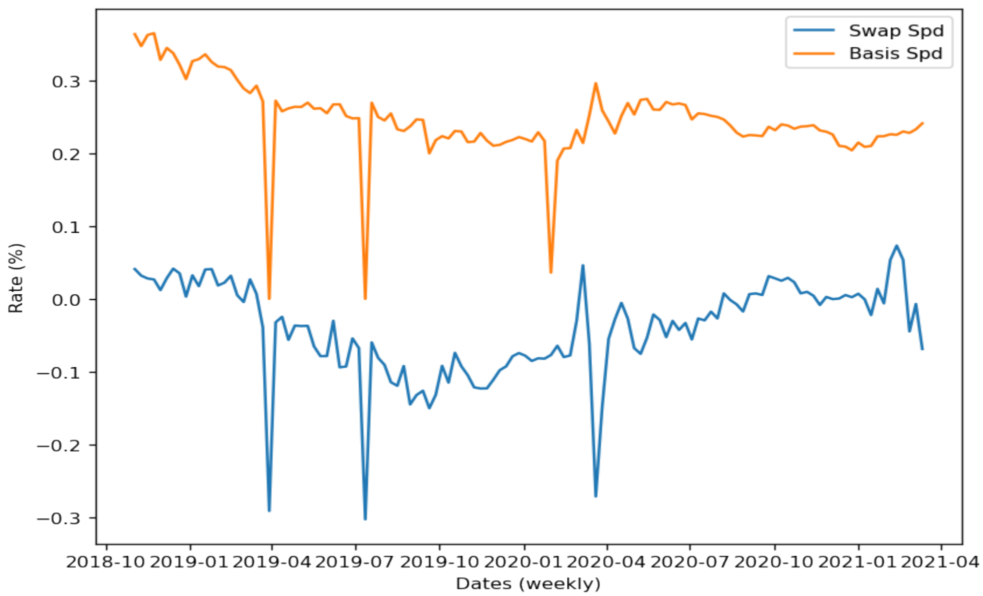

Figure 4 **Interest rate spread curves time series (10year tenor)**

As shown in Fig 4, for the 10yr tenor, the two spread time series demonstrate a similar pattern in general. Within the observation window, there are three downward jumps due to the impact associated with LIBOR around Apr 2019, Jul 2019, and Feb 2020.

3.2. Dimension Reduction

The benchmark method, PCA, is a popular technique for explaining curve dynamics. However, two strict underlying assumptions may limit the predictive power for datasets with strong nonlinearity: (a) The new orthogonal basis, which is a linear combination of the original basis, usually is not able to capture the most interesting part of nonlinear behavior; (b) Mean and variance are sufficient statistics. Therefore, if the probability distribution is not Gaussian, all other necessary high-order statistics will be lost.

The second approach is Locally Linear Embedding (LLE), which is a classical manifold searching method, see Huo, Ni, and Smith (2007) (2). Given a set of D-dimensional data points $x_1, x_2, \ldots, x_n$, we try to find the embedded d-dimensional feature vectors $y_1, y_2, \ldots, y_n$. The main steps are as follows.

(1) Identify nearest neighbors based on a pre-specified distance metric for each data point x_i, where N_i denote the set of indices of the nearest neighbors for this data point.

(2) Find the optimal local convex combination of the nearest neighbors to represent each data point. That is, we are optimizing function (3) to compute the weights.
Function (3):

$$E(w) = \sum_{i=1}^{N} \left\| x_i - \sum_{j \in N_i} W_{ij} x_j \right\|^2, s.t. \sum_{j \in N_i} W_{ij} = 1$$

(3) Find the low-dimensional feature vectors y_i, which have the optimal local convex representations with the given W_{ij}. That is, we try to compute y_i by minimizing the following cost function (4).
Function (4):

$$\Phi(y) = \sum_{i=1}^{N} \left\| y_i - \sum_{j \in N_i} W_{ij} y_j \right\|^2$$

As we try to model nonlinear relationship among different tenors, the low-dim representations may not be that interpretable as the parametric framework or the baseline PCA approach. This is the cost of improving prediction power.

3.3. Time Series Forecasting and Inverse Transform

Our objective is to predict a curve transformation given its observed shape at a particular moment in time. Our method converts original spread curves into several main drivers in low-dimensional space by manifold learning. After conversion, we employ either the linear models (e.g. ARMA) or nonlinear hessian regularized models as the time series forecasting tool.
Because manifold learning is not an injective function, we may not come up with a general approach to reconstruct our predicted term structure in high-dimensional space. Three potential solutions are: (a) simple LLE reconstruction, by Chen, Deng, and Huo (2004) (1). (b) Local Tangent Space Alignment (LTSA) (a nonparametric regression approach), by Z. Zhang and H. Zha (2004) (5). (c) inverse manifold learning (encoding and decoding), by Li S, Lin H, Zang Z, et al. (2020) (6).
By employing the simple LLE reconstruction, suppose low-dimensional feature vectors $y_1, y_2, \ldots, y_n$ have been obtained in the previous subsection and we have a new prediction y_{n+1}, we could reconstruct $\hat{x_{n+1}}$ using the following steps.

(1) Identify nearest neighbors based on the same distance metric for each data point y_{n+1}, where N_{n+1} denote the set of indices of the nearest neighbors for this data point.

(2) The weights of the local optimal convex combination w_j are obtained by optimizing function (5).
Function (5):

$$E(w) = \left\| y_{n+1} - \sum_{j \in N_{n+1}} W_j y_j \right\|^2, s.t. \sum_{j \in N_{n+1}} W_j = 1$$

(3) The reconstructed data point in the high-dimensional space is represented as

$$\hat{x_{n+1}} = \sum_{j \in N_{n+1}} W_j x_j$$

3.4. Performance Evaluation and Systematic Trading Strategies

We define a statistical measure and an economic measure to evaluate forecasting performance. For the statistical measure, regarding the relative difference of spread values between two adjacent time steps could be classified into positive and negative categories, we calculate the accuracy rate. Denote the total number of time points in the back-test procedure as N, the actual and predicted one-step changes at time t are da_t and db_t respectively, then the accuracy rate is defined as:

Function (6):

$$\frac{\sum_{t=1}^{N} I(da_t db_t \geq 0)}{N}$$

, where I is an indicator function.

For the economic measure, we back-test absolute-value strategies using each single asset and three systematic relative-value strategies (level, fly, butterfly) using multiple assets based on forecasting. Specifically, the level trading signal is defined as: $\frac{\sum_{t=1}^{M} R_i}{M}$, where M is the number of tenors used in the construction of this signal, R_i is the objective spread rate with tenor i. If we predict the signal is going to increase at the next period (average level increases), then float leg payment will increase, we will build a float leg receiver position. The PnL from this trade will be $\frac{\sum_{t=1}^{M}(P_{i,t}-P_{i,t-1})}{M}$, where $P_{i,t}$ is the synthetic swap price for tenor i and time t.

The fly trading signal will be: $R_j - R_i$, where i denotes the shorter tenor (e.g. 2yr) and j denotes the longer tenor (e.g. 10yr). If we predict the signal is going to increase at next period (curve slope increases), then float leg payment of 10yr will increase relative to the 2yr, we will build a float leg receiver position of 10yr and a float leg payer position of 2yr. The PnL from this trade will be $(P_{j,t} - P_{j,t-1}) - (P_{i,t} - P_{i,t-1})$.

The butterfly trading signal will be: $0.5 * R_k + 0.5 * R_j - R_i$, where k denotes the shortest tenor (e.g. 2yr), j denotes the longest tenor (e.g. 10yr), and i denotes the mid tenor (e.g. 5yr). If we predict the signal is going to increase at next period (curve curvature increases), then float leg payment of 10yr and 2yr will increase relative to the 5yr, we will build a float leg receiver position of 10yr and 2yr and a float leg payer position of 5yr. The PnL from this trade will be $0.5 * (P_{k,t} - P_{k,t-1}) + 0.5 * (P_{j,t} - P_{j,t-1}) - (P_{i,t} - P_{i,t-1})$.

Intuitively, if we decide to build a short position, then all we need is to add a negative sign to the above PnL expression of the long position.

By holding the corresponding synthetic swap portfolio for a week suggested by the trading signal, we will get three cumulative PnL plots for each given model specification. This will help us identify the difference among various dimension reduction methods. The notional value of the synthetic swap portfolio is 100 million USD. The initial time to maturity is 5 year. The coupon payment frequency is semiannually. The swap rate of the fixed leg is 1% (in exchange for the floating swap spread).

4. Numerical Experiments

In this section, we first extract the residual process from swap spread using linear regression, then fit a short-rate model and generate a simulated dataset accordingly. Secondly, we explore the selection of the number of neighbors in the LLE framework and determine the optimal value. Finally, for the swap spread, we report the accuracy and trading profitability for single tenor forecasting using both PCA and LLE-based framework, then repeat the similar pipeline for systematic trading including multiple tenors.

4.1. Simulation via a Residual-based Vasicek Model

Swap spreads tend to be random walk processes, which invalidates the use of a pure autoregression model. Hence, we first regress the spread on risk premium and liquidity premium to get the residual process for each tenor respectively $RS_1, RS_2, \ldots, RS_n$, where n denotes the number of time periods and each $RS_i \in R^1$. Then, we treat it as a mean-reverting Ornstein-Uhlenbeck process to estimate the parameters of the residual processes for each tenor (equation (7) and (8)). The residuals for all tenors in a term structure are driven by the same random shock, while the shocks are independent across samples along the time horizon.

As shown in the Table 1, the residual processes with different tenors are indeed stationary based on the result of the augmented Dickey-Fuller (ADF) test. The unit root hypotheses are rejected at the 5% level for all maturities, except for 10 year tenor (rejected at the 10% level).

Tenor (month)	12	24	36	60	84	120	180	240	300	348
P-value	0.002	0.013	0.005	0.042	0.013	0.076	0.009	0.003	0.004	0.003

Table 1 Result of ADF Test

As shown in Fig 5, the mean-reverting speed λ varies significantly across tenors. It keeps moving downward till the 3yr, then gradually trends upward in the mid-end till 10-15yr, finally decreases in the long-end. The volatility σ demonstrates a similar downward trend in the short-end, then monotonously increases after the 3yr point. Since the mean of the residual process is zero after the previous regression step, we do not focus too much on the mean-reverting level μ.

With the calibrated process, we simulate a large number of realizations (e.g. 1,000). For each simulated scenario, the dataset will be in the shape of N by M, where N is the number of weeks along the calendar dates and M is the number of tenors.

Equation (7):

$$dRS_t = \lambda(\mu - RS_t)dt + \sigma dW_t$$

Equation (8):

$$RS_{t+\delta} = RS_t e^{-\lambda\delta} + \mu(1 - e^{-\lambda\delta}) + \sigma\sqrt{\frac{1 - e^{-\lambda\delta}}{2\lambda}} N(0,1)$$

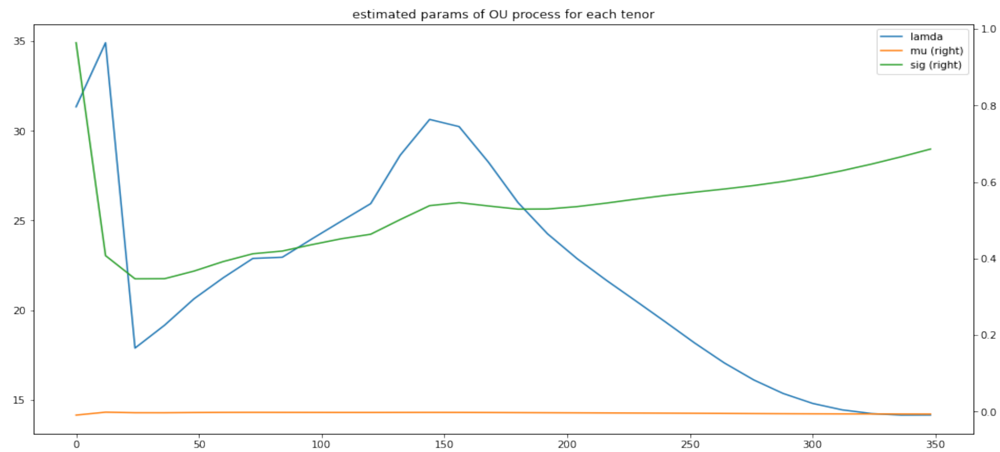

Figure 5 Estimated Parameters of O-U Processes

4.2. Choice of Hyper-parameter

The number of neighbors K in the LLE-based framework is an important hyper-parameter we need tune. Given a fixed low dimensionality (e.g. 3), we choose the K which will lead to minimum reconstruction error (RE). The RE is defined as:

$$RE = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{3} \mid \frac{\hat{x_{ij}} - x_{ij}}{x_{ij}} \mid$$

, where x_{ij} is the jth-dim embedding of the ith sample.

As shown in Fig 6, taking the average of 100 simulated scenarios, the RE trends towards zero as K increases and arrives at its minimum when K=37. We keep this selected parameter value in all the following experiments.

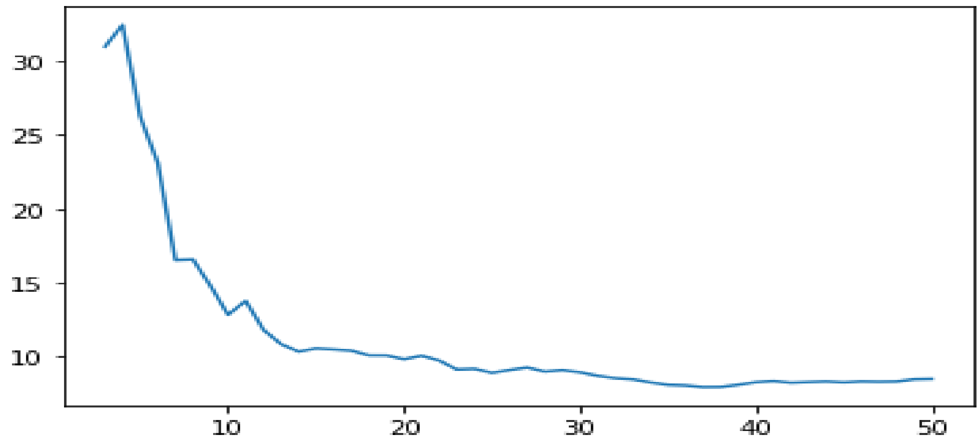

Figure 6 **Reconstruction Error v.s. Number of Neighbors**

4.3. Results: Single Asset Forecasting and Trading

For the swap spread curve, the PCA-based framework outperforms the LLE in forecasting long tenors (15yr and beyond), while the LLE is more suitable for short to mid tenors. Given a fixed model, the forecasting accuracy for tenors around 10yr are the top performers in each group, as shown in Table 2 and Fig 7. One of the potential reasons is associated with the larger liquidity of these tenors in the Treasury markets.

Tenor (month)	12	24	36	60	84	120	180	240	300	348
PCA+ARMA	56.8%	60.3%	61.0%	61.3%	61.2%	62.5% **	62.6% ***	61.4% *	60.2%	59.9%
LLE+ARMA	57.3%	60.5%	61.3%	61.7% *	61.7% *	62.8% ***	62.5% **	61.2%	60.1%	59.8%

Table 2 **Accuracy of Single Tenor Forecasting (swap spread curve); The star denotes the top accuracy among all tenors, ***: 1st, **: 2nd, *: 3rd. Red colored cell is the outperformer among the two models.**

Given a 60-week rolling window, we dynamically update our position each week according to the newest forecast, then compute the cumulative PnL from 2019-12-27 to 2021-03-12. As shown in Table 3 and Fig 8, under both the PCA and LLE-based framework, the short and mid tenors yield a positive cumulative PnL while longer tenors yield significantly poor performance. In terms of the relative difference between two dimension reduction approaches, LLE outperforms PCA when forecasting short to mid tenors while underperforms in the long end (20yr and beyond).

4.4. Results: Multiple Asset Forecasting and Trading

Following the strategies design in section 3.4, we back-tested the weekly-rebalanced “2s5s10s level”, “2s10s fly”, and “2s5s10s butterfly” trading based on the one-step ahead prediction using both two dimension reduction frameworks.

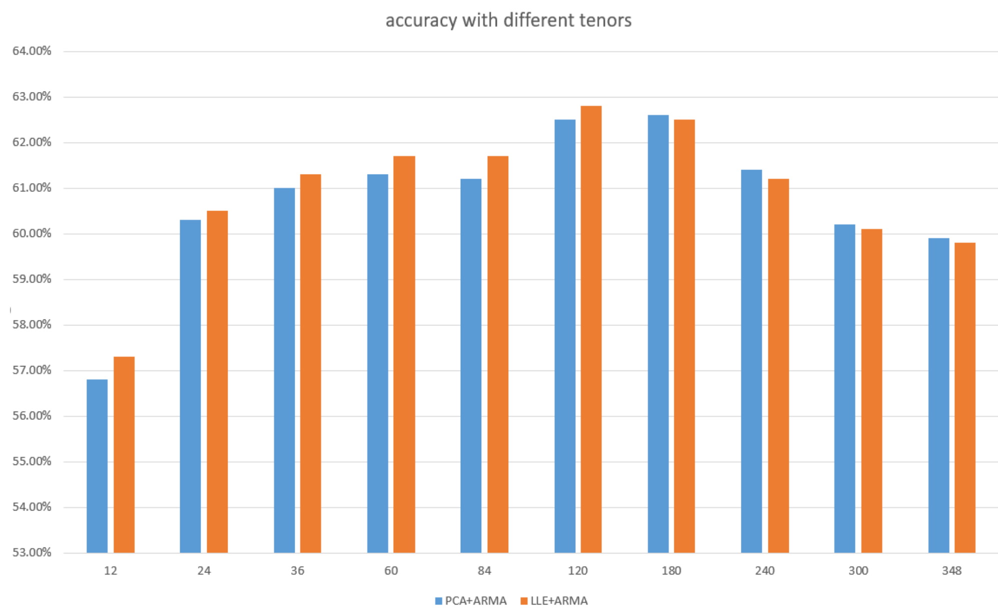

Figure 7 **Accuracy from 2019-12-27 to 2021-03-12**

Tenor (month)	12	24	36	60	84	120	180	240	300	348
PCA+ARMA	2.96 ***	2.09 **	0.15	0.56	0.61 *	-2.47	-7.32	-6.94	-6.01	-5.73
LLE+ARMA	3.95 ***	2.20 **	0.71	1.11 *	1.03	-2.17	-7.30	-7.35	-6.56	-6.21

Table 3 **Cumulative PnL of Single Tenor Forecasting (million USD); The star denotes the top accuracy among all tenors, ***: 1st, **: 2nd, *: 3rd. Red colored cell is the outperformer among the two models.**

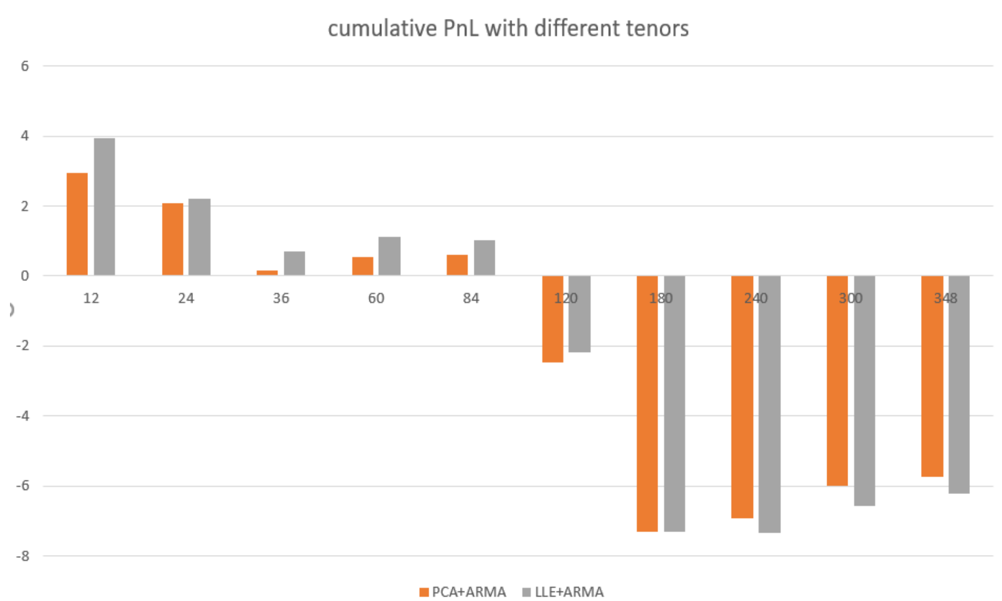

Figure 8 **Cumulative PnL from 2019-12-27 to 2021-03-12**

As shown in Table 4 and Fig 9, the forecasting result of LLE outperforms PCA in the back-test of all three term structure relative-value strategies. The average statistical accuracy of "level" trading is the highest (PCA: 61.7%, LLE: 62.0%), while the "fly" and "butterfly" strategies are the second (PCA: 57.8%, LLE: 58.5%) and third (PCA: 46.3%, LLE: 47.5%) respectively. For the "butterfly" trading, the accuracy from both frameworks is less than 50%. This further shows the higher the order we pursue in curve trading, the larger the difficulty will be.

Tenor (month)	level	fly	butterfly
PCA+ARMA	61.7%	57.8%	46.3%
LLE+ARMA	62.0%	58.5%	47.5%

Table 4 Accuracy of Multiple Tenor Forecasting (swap spread curve). Red colored cell is the outperformer among the two models.

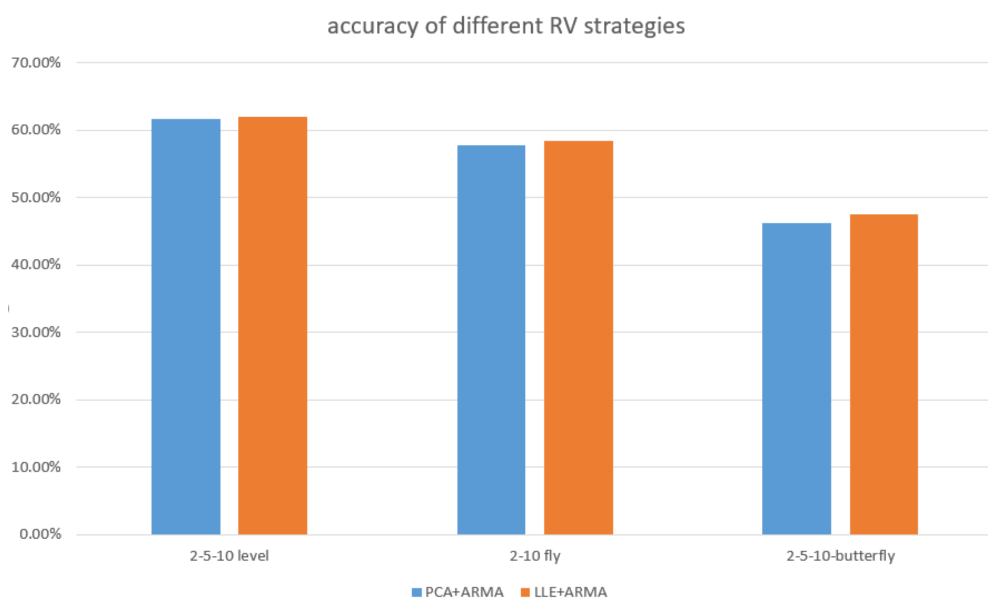

Figure 9 Accuracy from 2019-12-27 to 2021-03-12

As shown in Table 5 and Fig 10, in terms of the cumulative PnL (averaging from all simulated scenarios), the LLE outperforms the PCA in all relative-value trading, where the outperformance is the most significant in "level" trading (PCA: 0.264 million, LLE: 0.385 million). For both models, the PnL comparison among different relative-value strategies is consistent with the statistical accuracy: the higher out-of-sample forecast accuracy a given strategy could achieve, the better the PnL profile will be.

Tenor (month)	level	fly	butterfly
PCA+ARMA	0.264	0.082	-0.002
LLE+ARMA	0.385	0.085	-0.000

Table 5 Cumulative PnL of Multiple Tenor Forecasting (million USD). Red colored cell is the outperformer among the two models.

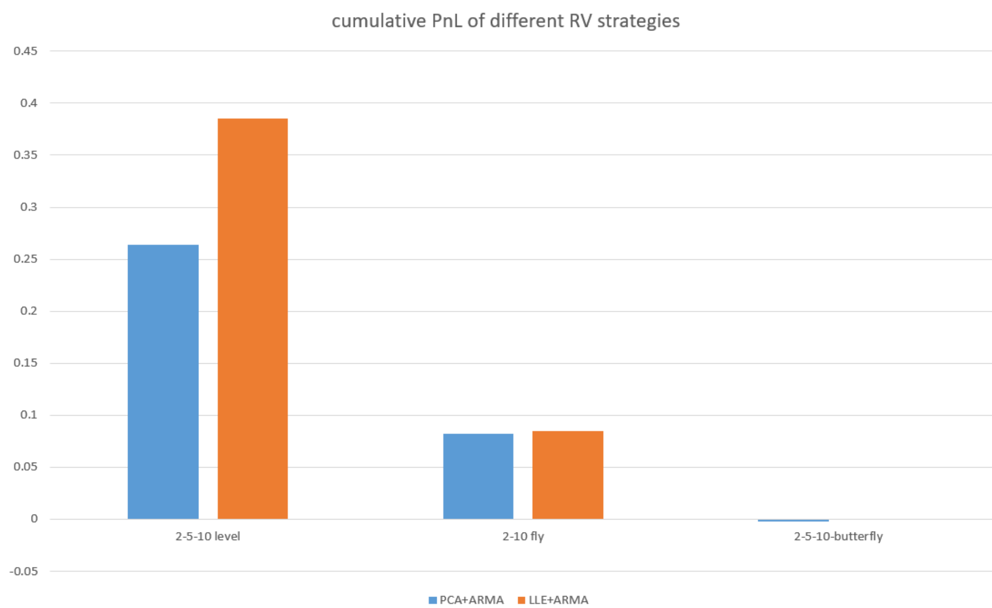

Figure 10 Cumulative PnL from 2019-12-27 to 2021-03-12

5. Conclusion and Discussion

In this paper, we implement a multi-stage framework to model the spread curve dynamics using manifold learning. We find the LLE-based approach performs well in predicting nonlinear patterns through various systematic trading strategies, while the benchmark PCA-based approach could only capture the low-order movement. The numerical experiment is built on a residual-based simulation procedure via the Vasicek model. Besides, we extensively deployed the hyper-parameters tuning of the dimension reduction method and the hypothesis test of time series forecasting.

In single asset forecasting, the LLE outperforms the benchmark in short to mid tenors. Meanwhile, given a fixed dimension reduction method, the highest forecasting accuracy appears around 5 to 10 year area due to the high liquidity. In multiple asset forecasting, the LLE outperforms the benchmark in all relative-value trading strategies. This demonstrates the superiority of LLE in modeling curve dynamics and developing complex trading strategies accordingly.

There are two directions of future work we may further explore: (a) replace the ARMA model with the Hessian regularized nonlinear time series model (HRM); (b) implement other dimension reduction methods to find the most suitable one for this task.

In the stage of time series forecasting, the basic idea of the HRM is to fit a model via penalization, where the penalty term is an unbiased estimator of the integrated Hessian of the underlying function (3). This will serve as an advanced time series forecasting approach benchmarked against the ARMA model in our work.

There are several more potential dimension reduction techniques, such as Multidimensional Scaling (MDS) and t-distributed Stochastic Neighbor Embedding (t-SNE). Compared to the direct eigen-analysis of the N data points in PCA, MDS selects influential dimensions by the eigen-analysis of the N^2 data points of a pairwise distance matrix. The goal is to preserve the pairwise distances as best as possible after mapping to the low-dimensional space. The t-SNE converts similarities between data points to joint probabilities and tries to minimize the K-L divergence between the joint probabilities of the low-dimensional embedding and the high-dimensional data.

References

[1] Chen J, Deng S J, Huo X. *Electricity price curve modeling and forecasting by manifold learning*. IEEE Transactions on Power Systems, 2008, 23(3): 877-888.

[2] Huo X, Ni X S, Smith A K. *A survey of manifold-based learning methods*. Recent advances in data mining of enterprise data, 2007: 691-745.

[3] Chen J, Huo X. *A hessian regularized nonlinear time series model*. Journal of Computational and Graphical Statistics, 2009, 18(3): 694-716.

[4] Kondratyev A. *Learning curve dynamics with artificial neural networks*. Available at SSRN 3041232, 2018.

[5] Zhang Z, Zha H. *Principal manifolds and nonlinear dimensionality reduction via tangent space alignment*. SIAM journal on scientific computing, 2004, 26(1): 313-338.

[6] Li S, Lin H, Zang Z, et al. *Invertible Manifold Learning for Dimension Reduction*. arXiv preprint arXiv:2010.04012, 2020.

[7] Blaskowitz O J. *A forecast evaluation of PCA-based adaptive forecasting schemes for the EURIBOR swap term structure*. Christian-Albrechts Universität Kiel, 2009.

[8] Sack B P, Wright J, Gürkaynak R. *The US Treasury yield curve: 1961 to the present*. Board of Governors of the Federal Reserve System (US), 2006.

[9] Diebold F X, Li C. *Forecasting the term structure of government bond yields*. Journal of econometrics, 2006, 130(2): 337-364.

[10] Ang A, Piazzesi M. *A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables*. Journal of Monetary economics, 2003, 50(4): 745-787.

[11] Cooper I, Priestley R. *Time-varying risk premiums and the output gap*. The Review of Financial Studies, 2009, 22(7): 2801-2833.

[12] Ludvigson S C, Ng S. *Macro factors in bond risk premia*. The Review of Financial Studies, 2009, 22(12): 5027-5067.

[13] Sorensen E H, Bollier T F. *Pricing swap default risk*. Financial Analysts Journal, 1994, 50(3): 23-33.

[14] Grinblatt M. *An analytic solution for interest rate swap spreads*. International Review of Finance, 2001, 2(3): 113-149.

[15] Litterman R, Scheinkman J. *Common factors affecting bond returns*. Journal of fixed income, 1991, 1(1): 54-61.

[16] Yu J. *Forecasting Interest Rate Swap Spreads: A Residual-based Approach*. Unpublished doctoral dissertation, Georgia State University, 2014.

Acknowledgments

The authors gratefully acknowledge the existence of the Journal of Irreproducible Results and the support of the Society for the Preservation of Inane Research.