ISyE8900 Project

Chi Zhang

Georgia Institute of Technology

Abstract: This project 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 two objective spread curves: Swap Spread (LIBOR substract Treasury) and Basis Spread (LIBOR substract SOFR). 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.

I. 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 LIBOR-SOFR spread when USD LIBOR-based activity gradually decreases until completely unusable. *(https://www.newyorkfed.org/arrc/sofr-transition)*

The recent development of statistical learning greatly inspired the interest in applying dimensionality reduction and predictive models with sparse and nonlinear features in the finance industry, where the curve spread modeling is a canonical example. The use of such methods are 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.

In this paper we propose the following multistage approach for the short-term prediction of Swap/Basis spread changes following the work by Chen, Deng and Huo (2008): (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.

Benchmarking on the well-known PCA-based framework, the performance evaluation of different modeling objective curves (Swap and Basis spread) includes both statistical accuracy and the profitability of term structure relative-value trading strategies.

II. Literature Review

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) 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 (Ang and Piazzesi, 2003; Diebold, Rudebusch and Aruoba, 2006; Cooper and Priestley, 2008; Ludvigson and Ng, 2009). 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.

Specifically, focused on the context of dimension reduction techniques, PCA has been applied in the field of term structure modeling for decades, such as Steeley (1990) and Litterman and Scheinkman (1991). Usually, 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 non-linear relationship among neighbor tenors, Kondratyev (2018) proposes a neural networks based term structure algorithm for brent oil forward price and USD swap curve.

(https://github.com/DarseZ/CurveFrcst-Using-ManifoldLrn/blob/main/papers/ANNLearnCurveDynamics.pdf)

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 III starts from the economic meaning of four interest rate curves used in this project (LIBOR, SOFR, Treasury, and OIS), then provides a structured formulation of how to build the required curves and implement nonlinear dimension reduction, finally outlines the times series forecasting architectures and trading strategies design. Sections IV describe the design of the empirical analysis and the results. Section V concludes and discusses future work.

III. Problem formulation and Application

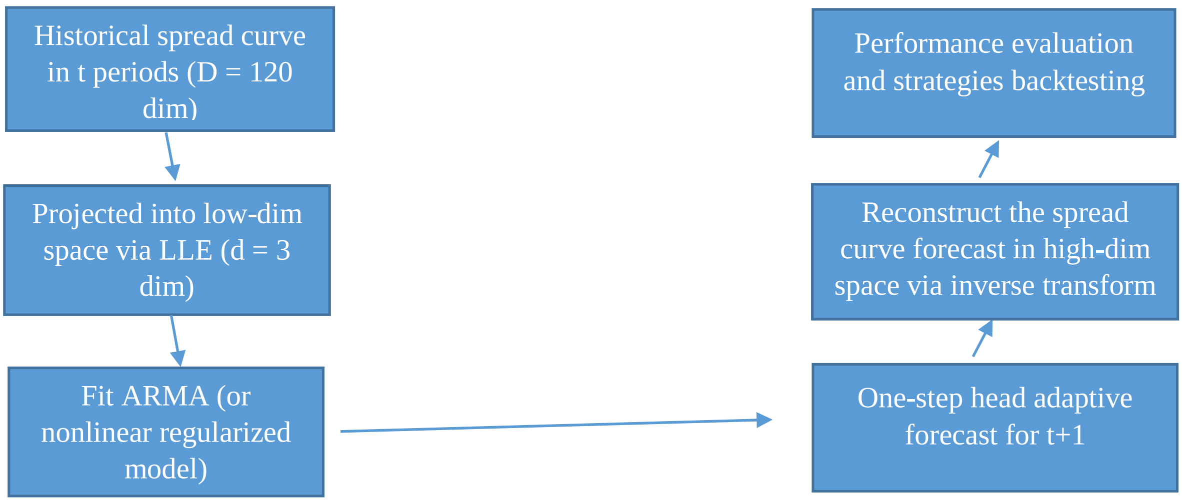

Fig. 1. Conceptual flowchart of the framework

Given the panel dataset of LIBOR rate $\{L_1, L_2, \dots, L_n\}$ where n denotes the number of time periods and each $L_i \in R^D$, and similar panel datasets of Treasury rate $\{T_1, T_2, \dots, T_n\}$ and SOFR rate $\{S_1, S_2, \dots, S_n\}$, we first calculate the two required spread datasets: swap spread $SS_i = \{L_1 - T_1, L_2 - T_2, \dots, L_n - T_n\}$ and basis spread $BS_i = \{L_1 - S_1, L_2 - S_2, \dots, L_n - S_n\}$. Then, by using a pre-specified dimension reduction technique, we obtain their low-dimensional representations SS_low_i and BS_low_i respectively, where each sample is in the R^d space for i from 1 to t. For each objective spread, we do a one-step head prediction using univariate time series modeling to get SS_low_{n+1} and BS_low_{n+1} respectively. Finally, map them back to the original R^D space using the corresponding inverse transform to get $\widehat{SS_{n+1}}$ and $\widehat{BS_{n+1}}$, then evaluate the forecasting accuracy and design trading strategies according to the comparison with realized value SS_{n+1} and BS_{n+1}.

A. Curves Building and Interest Rate Processes Simulation

Since there are a great amount of frameworks to construct interest rate curves using data from different providers, in order to eliminate some ambiguity, we first go through our data preparation process.

Considering the purpose of understanding the treasury curve's fundamental determinants, we will 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 only 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 zero rate expression. (*https://www.federalreserve.gov/data/nominal-yield-curve.htm)*

$$y_i(n) = \beta_0 + \beta_1 \frac{1-\exp(-\frac{n}{\tau_1})}{\frac{n}{\tau_1}} + \beta_2 \left[\frac{1-\exp\left(-\frac{n}{\tau_1}\right)}{\frac{n}{\tau_1}} - \exp\left(-\frac{n}{\tau_1}\right)\right] + \beta_3 \left[\frac{1-\exp\left(-\frac{n}{\tau_2}\right)}{\frac{n}{\tau_2}} - \exp\left(-\frac{n}{\tau_2}\right)\right], \qquad (1)$$

$$f_t(n,0) = \beta_0 + \beta_1 \exp(-\frac{n}{\tau_1}) + \beta_2 \frac{n}{\tau_1} \exp(-\frac{n}{\tau_1}) + \beta_3 \frac{n}{\tau_2} \exp(-\frac{n}{\tau_2}), \qquad (2)$$

The data format should be daily data with a continuous maturity spectrum from overnight to 30y (360 dimensions). To make it consistent with the other two swap curves, we re-sample it to get a weekly dataset with a discrete maturity spectrum (120 dimensions).

For the LIBOR and SOFR curves, we directly download them from the Bloomberg terminal by manually changing the "As of Date" variable. If time permits, we will replicate the whole curve bootstrapping pipeline to get these curves using market instruments prices (deposits, futures/forwards, swaps).

To generate more datasets for the performance evaluation, we will add parametric simulations of underlying interest rates to serve as a more solid numerical experiment. More details will be addressed in Section V.

The data preparation could be seen in *https://github.com/DarseZ/CurveFrcst-Using-ManifoldLrn/blob/main/CurveBuild.ipynb*

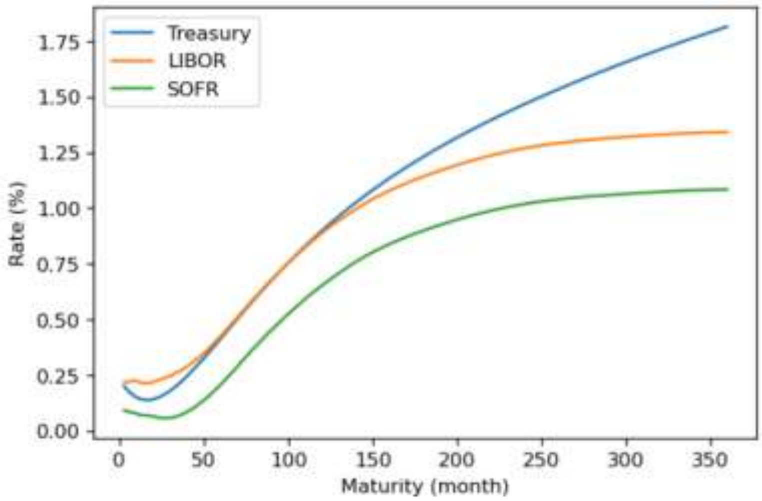

Fig. 2. Original interest rate curves snapshot on 2020-10-30

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.

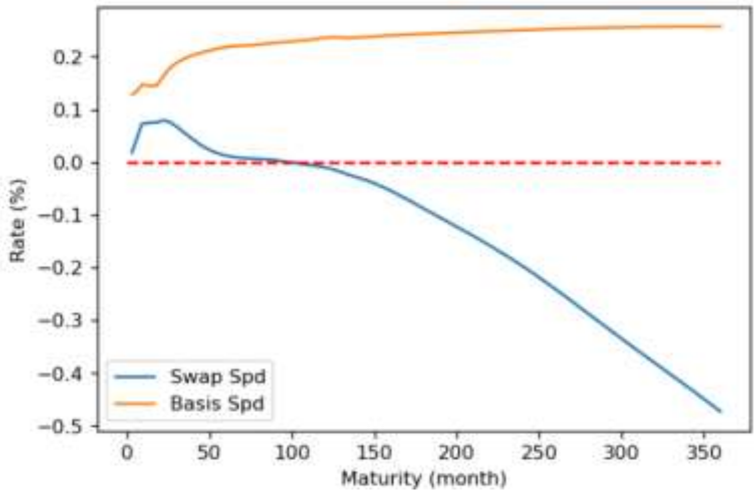

Fig. 3. Interest rate spread curves snapshot on 2020-10-30

As shown in Fig. 3, on 2020-10-30, the basis spread curve 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.

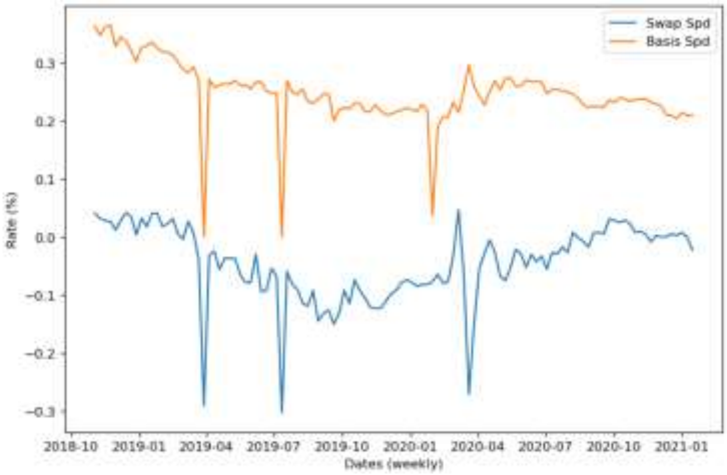

Fig. 4. Interest rate spread curves time series (10year tenor)

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

B. Nonlinear Dimension Reduction

The benchmark method, PCA, is a popular technique for explaining curve dynamics. However, there are two underlying strong assumptions that may limit the predictive power for datasets with strong nonlinearity: (1) 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; (2) 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. Given a set of D-dimensional data points $x_1, x_2, \dots, x_n$, we try to find the embedded d-dimensional feature vectors $y_1, y_2, \dots, y_n$. The main steps are as follows.

(1) Identify nearest neighbors based on some 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 (5) to compute the weights.

$$E(w) = \sum_{i=1}^{N} \left\| x_i - \sum_{j \in N_i} W_{ij} x_j \right\|^2 \text{ , s.t. } \sum_{j \in N_i} W_{ij} = 1 \qquad (3)$$

(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 (6).

$$\phi(y) = \sum_{i=1}^{N} \left\| y_i - \sum_{j \in N_i} W_{ij} y_j \right\|^2 \qquad (4)$$

The dimension reduction part could be seen in *https://github.com/DarseZ/CurveFrcst-Using-ManifoldLrn/blob/main/DmnsRdct_StateFrcst.ipynb*

As we try to model non-linear 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.

C. 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 both the linear time series models (ARMA) and nonlinear hessian regularized models as main time series forecasting tools.

Because the (inverse) manifold learning algorithm is not an injective function, we may not come up with a general approach. Three potential solutions: (1) Local Tangent Space Alignment (LTSA) (a nonparametric regression approach), by Z. Zhang and H. Zha (2004), "Principal manifolds and nonlinear dimension reduction via tangent space alignment,"; (2) inverse manifold learning (encoding and decoding) by the scholars in Xihu University (2020). (3) simple LLE reconstruction, by Jie Chen, Shijie Deng, and Xiaoming Huo (2004).

By employing the simple LLE reconstruction, suppose low-dimensional feature vectors y_1, y_2, … , y_n have been obtained in the previous subsection and we have a new prediction y_{n+1}, we could reconstruct x_{n+1} using the following steps.

(1) Identify nearest neighbors based on some 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 (7).

$$E(w) = \left\| y_{n+1} - \sum_{j \in N_{n+1}} W_j x_j \right\|^2 \text{ , s.t.} \sum_{j \in N_{n+1}} W_j = 1 \qquad (5)$$

(3) The data point in high-dimensional space should be $\widehat{x_{n+1}} = \sum_{j \in N_{n+1}} W_j X_j$.

D. 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 non-negative (positive) and negative categories, we calculate the accuracy rate.

Denote the total number of time points in the testing 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:

$$\frac{\sum_{t=1}^{N} I(da_t * db_t \geq 0)}{N}, \text{ where I is an indicator function} \qquad (6)$$

For the economic measure, we backtest absolute value strategies using a 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: $\frac{\sum_{i=1}^{M} R_i}{M}$, 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 $\frac{\sum_{i=1}^{M}(P_{i,t}-P_{i,t-1})}{M}$, 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 and j denotes the longer tenor. 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 $\left(P_{j,t} - P_{j,t-1}\right) - \left(P_{i,t} - P_{i,t-1}\right)$.

The butterfly trading signal will be: $0.5*R_k + 0.5*R_j - R_i$, where k denotes the shortest tenor, j denotes the longest tenor, and i denotes the mid tenor. 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 $0.5 * \left(P_{k,t} - P_{k,t-1}\right) + 0.5 * \left(P_{j,t} - P_{j,t-1}\right) - \left(P_{i,t} - P_{i,t-1}\right)$.

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 swap portfolio for one period (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 between different model specifications (dimension reduction and time series forecast approaches) in predicting specific patterns of curve dynamics.

IV. Results

A. Single Tenor Forecasting and Trading

For swap spread curve:

Tenor (month)	24	36	60	84	120	240	357
PCA+ARMA	63.6% **	63.6% **	59.1%	56.8%	65.9% ***	43.2%	47.7%
LLE+ARMA	56.8% **	63.6% ***	52.3%	50.0%	63.6% ***	45.5%	50.0%

Table 1. 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.

The PCE+ARMA outperforms the LLE+ARMA in forecasting short to mid tenors, while the LLE+ARMA is more suitable for long tenors (20yr and 30yr). Given a fixed model, the forecasting accuracy for 2yr, 3yr, and 10yr are the top performers in each group.

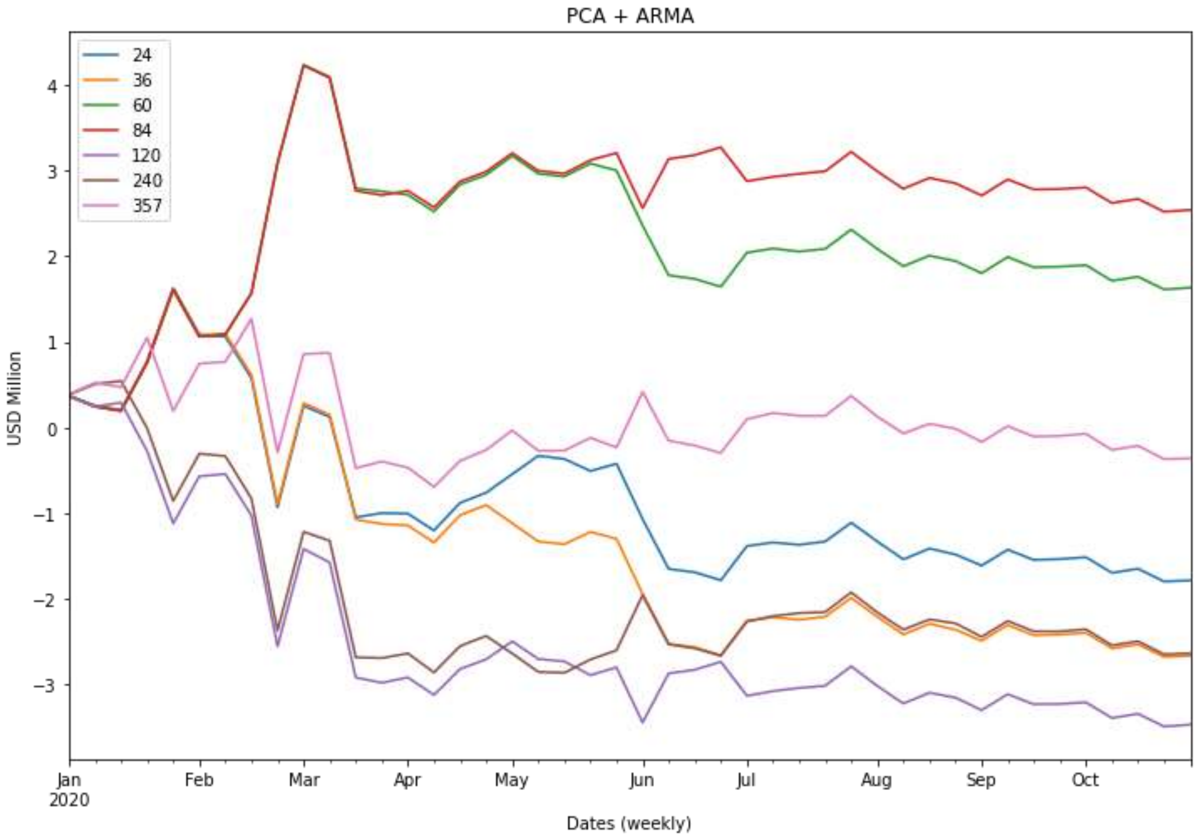

Fig. 5. PnL of single tenor trading using PCA+ARMA (swap spread curve)

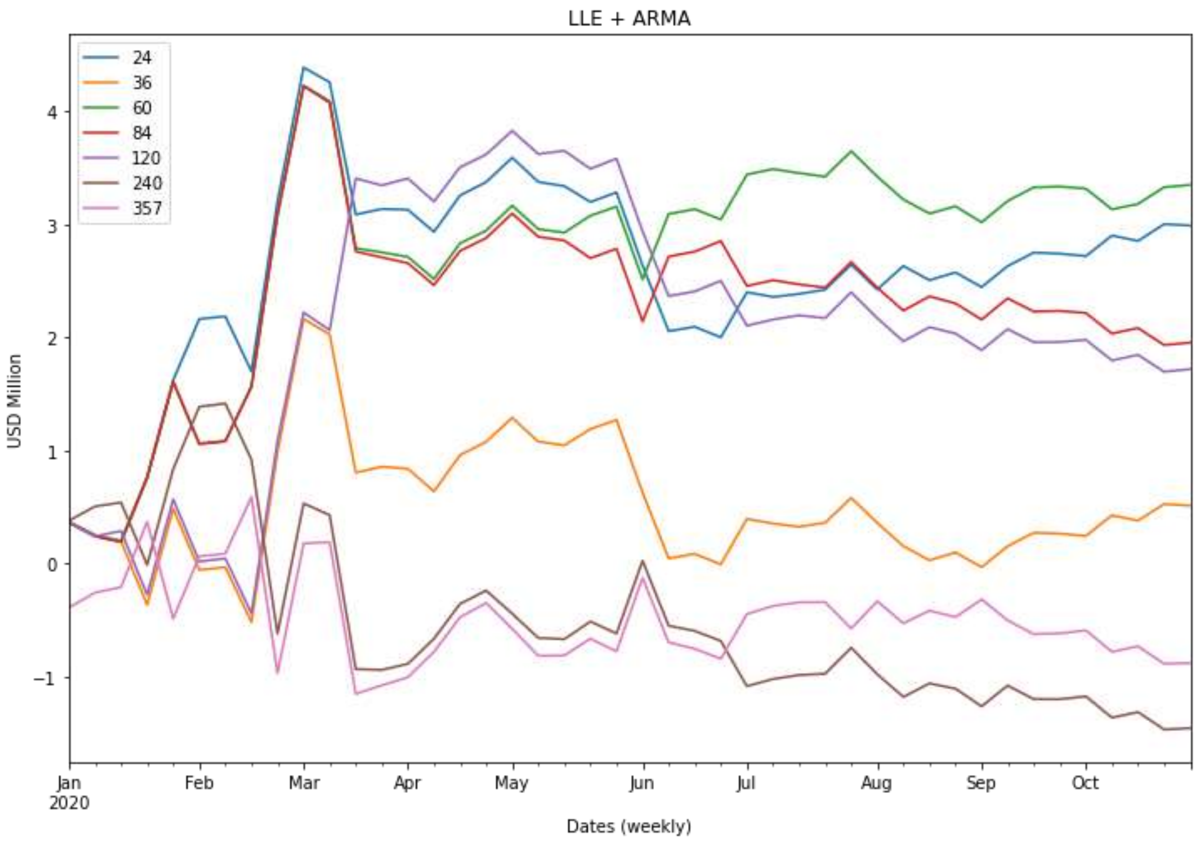

Fig. 6. PnL of single tenor trading using LLE+ARMA (swap spread curve)

For basis spread curve:

Tenor (month)	24	36	60	84	120	240	360
PCA+ARMA	50.0%	54.5%	47.7%	54.5%	61.4% **	59.1% *	68.2% ***
LLE+ARMA	59.1% *	52.3%	52.3%	56.8%	65.9% ***	63.6% **	54.5%

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

There is no significant evidence in the accuracy comparison that could support the superiority of LLE in basis spread forecasts. However, except for the 3yr and 30yr tenor, the accuracy of LLE+ARMA is indeed higher than PCA+ARMA.

Comparing the accuracy across tenors in each group, the long-end forecasting is more accurate than short–end tenors.

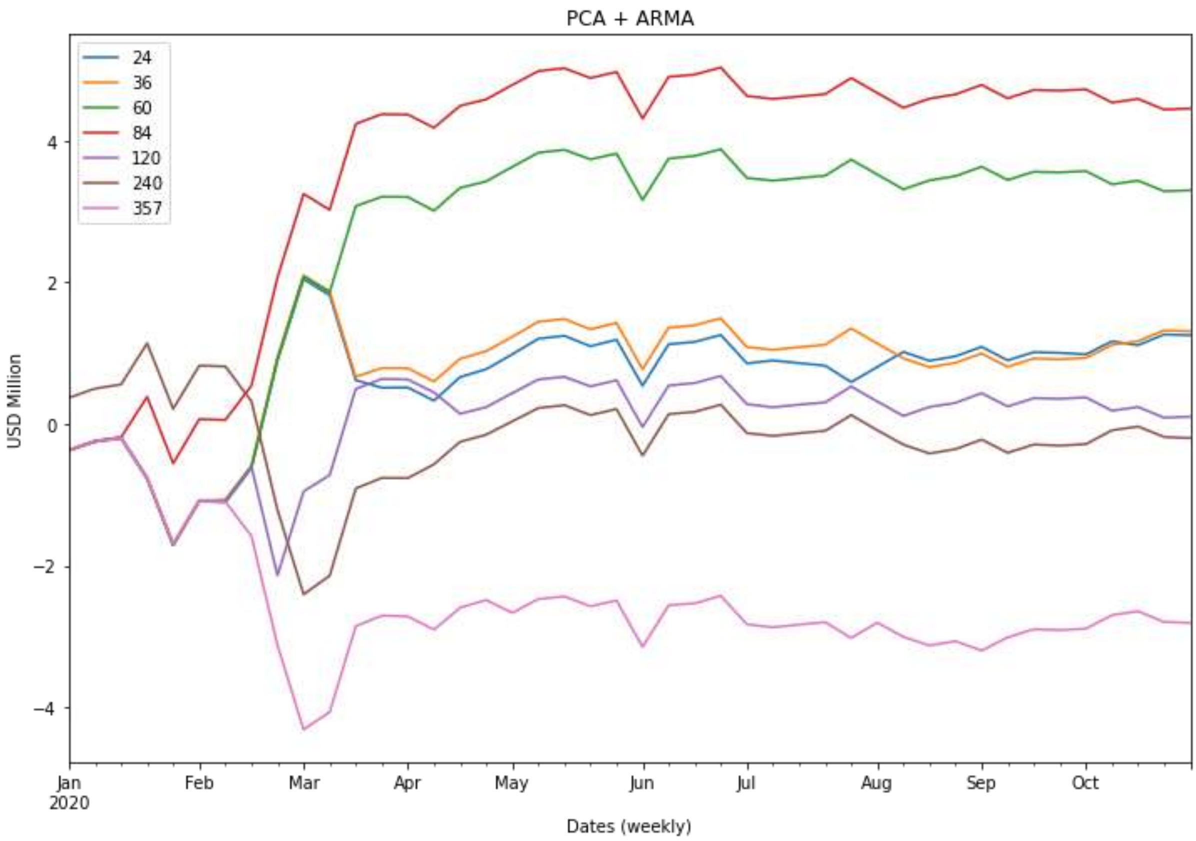

Fig. 7. PnL of single tenor trading using PCA+ARMA (basis spread curve)

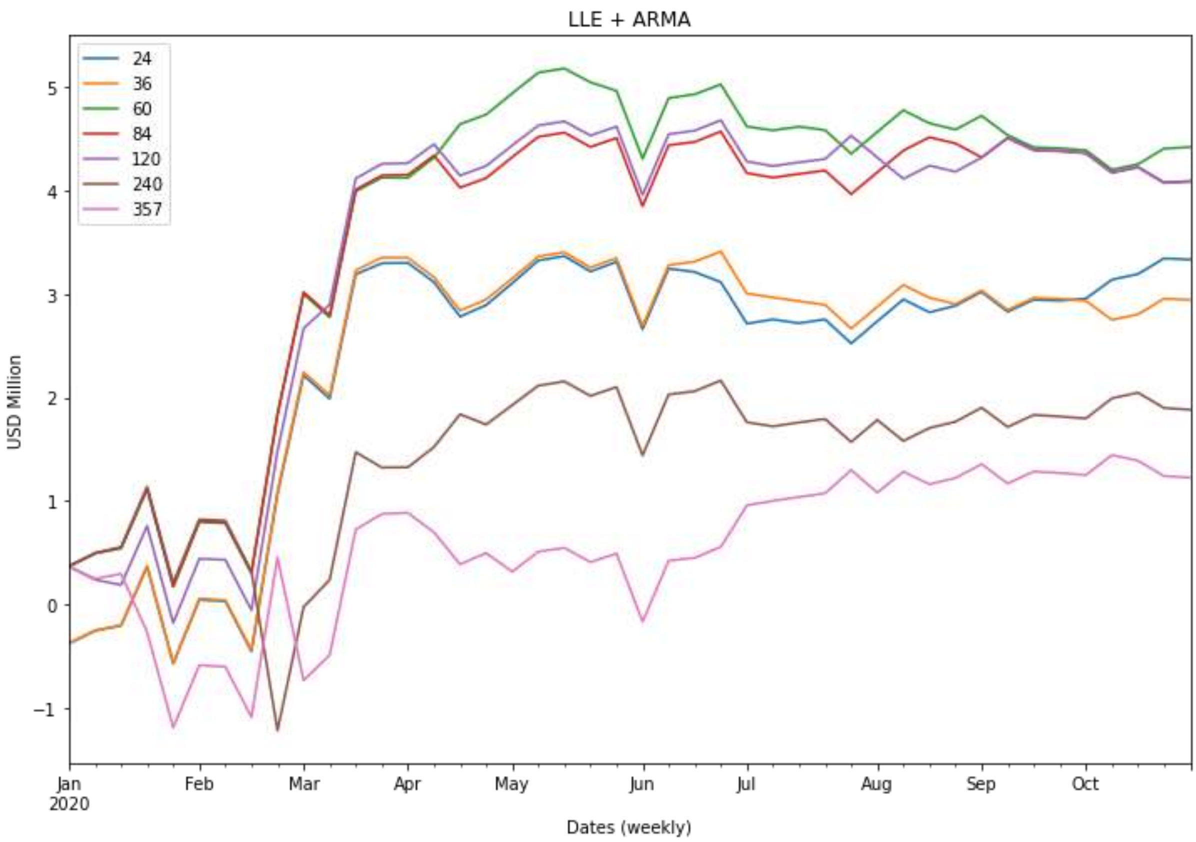

Fig. 8. PnL of single tenor trading using LLE+ARMA (basis spread curve)

B. Multiple Tenor Forecasting and Trading

For swap spread curve:

Relative-value positions	level	fly	butterfly
PCA+ARMA	56.8%	59.1%	50.0%
LLE+ARMA	54.5%	59.1%	54.5%

Table 3. Accuracy of Multiple Tenor Forecasting (swap spread curve)

As shown in Table 3, the fly and butterfly relative-value strategies using the LLE+ARMA model outperform the PCA+ARMA model, where the difference is more significant in the butterfly case. However, in the level trading strategy, the performance of LLE is not good as PCA to capture this linear signal. Specifically, given the notional amount of 1 million USD, the level trading cumulative PnL achieves 1 million using either LLE+ARMA or PCA+ARMA. Similarly, the butterfly trading cumulative PnL exceeds 0.04 million using LLE+ARMA, compared to the negative value in the benchmark.

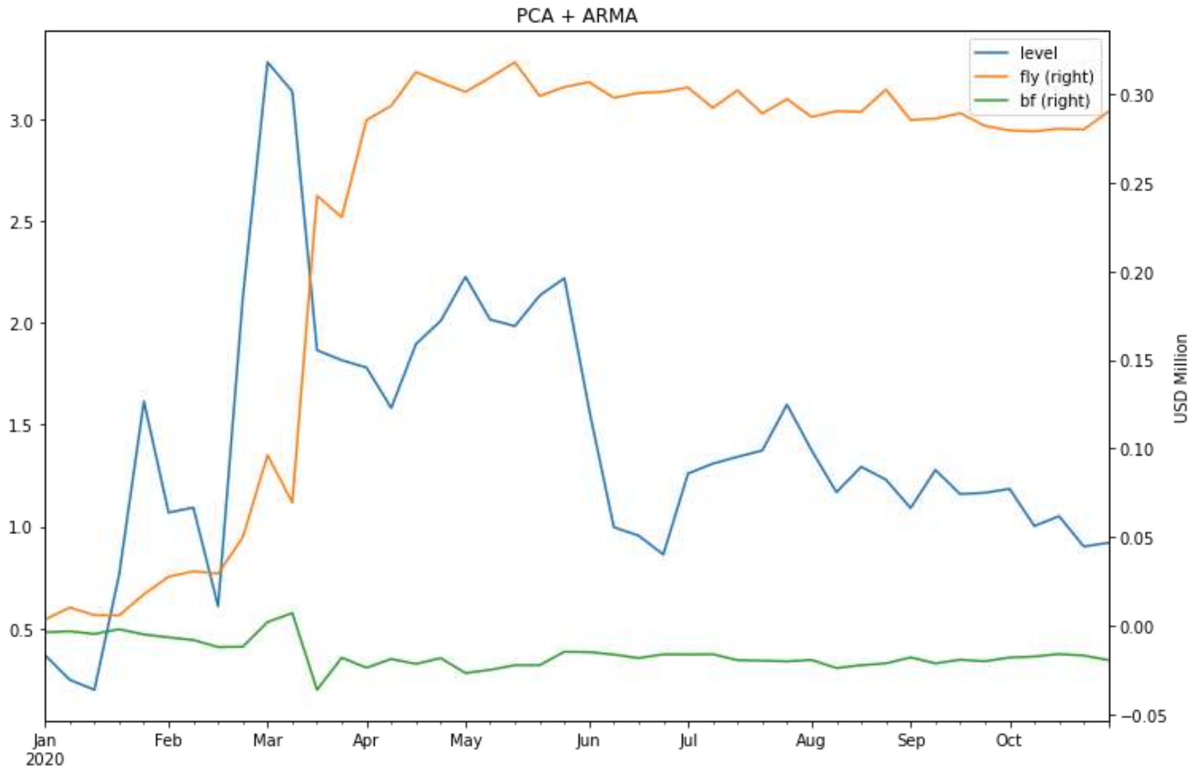

Fig. 9. PnL of multiple tenor trading using PCA+ARMA (swap spread curve)

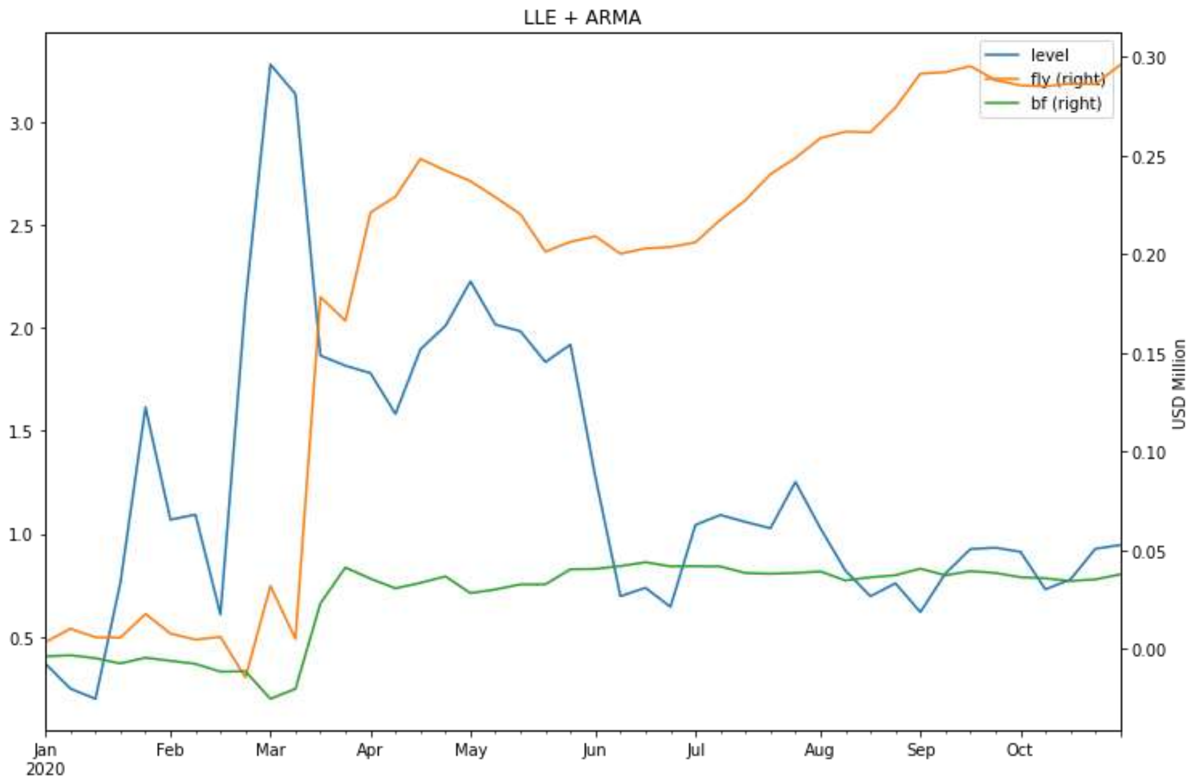

Fig. 10. PnL of multiple tenor trading using LLE+ARMA (swap spread curve)

For basis spread curve:

Relative-value positions	level	fly	butterfly
PCA+ARMA	52.3%	47.7%	56.8%
LLE+ARMA	43.2%	50.0%	54.5%

Table 4. Accuracy of Multiple Tenor Forecasting (basis spread curve)

Similar to the single tenor case, the PCA could capture most of the temporal variation in basis spread forecasting according to the accuracy comparison, especially the first-order variation (corresponding to the level signal).

However, the story from the PnL plot is much more intriguing. Given the fact that both forecasting accuracy is around 50%, the level trading cumulative PnL exceeds 4 million using LLE+ARMA, while the benchmark cumulative PnL is around 3.5 million. Our elementary conjecture for this result is the "unbalanced" movement of interest rate spread. This also emphasizes the necessity of more numerical experiments using simulated datasets. For the butterfly trading, the PnL of LLE+ARMA underperforms its benchmark, which is consistent with the statistical measure.

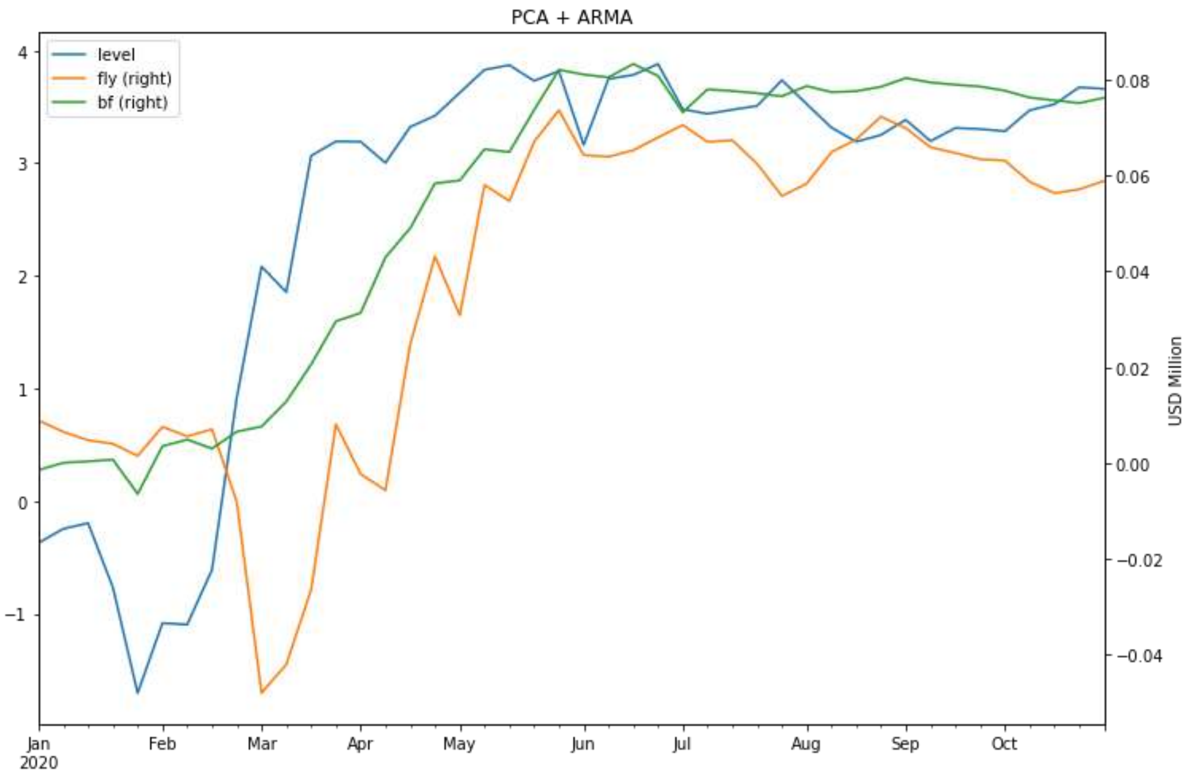

Fig. 11. PnL of multiple tenor trading using PCA+ARMA (basis spread curve)

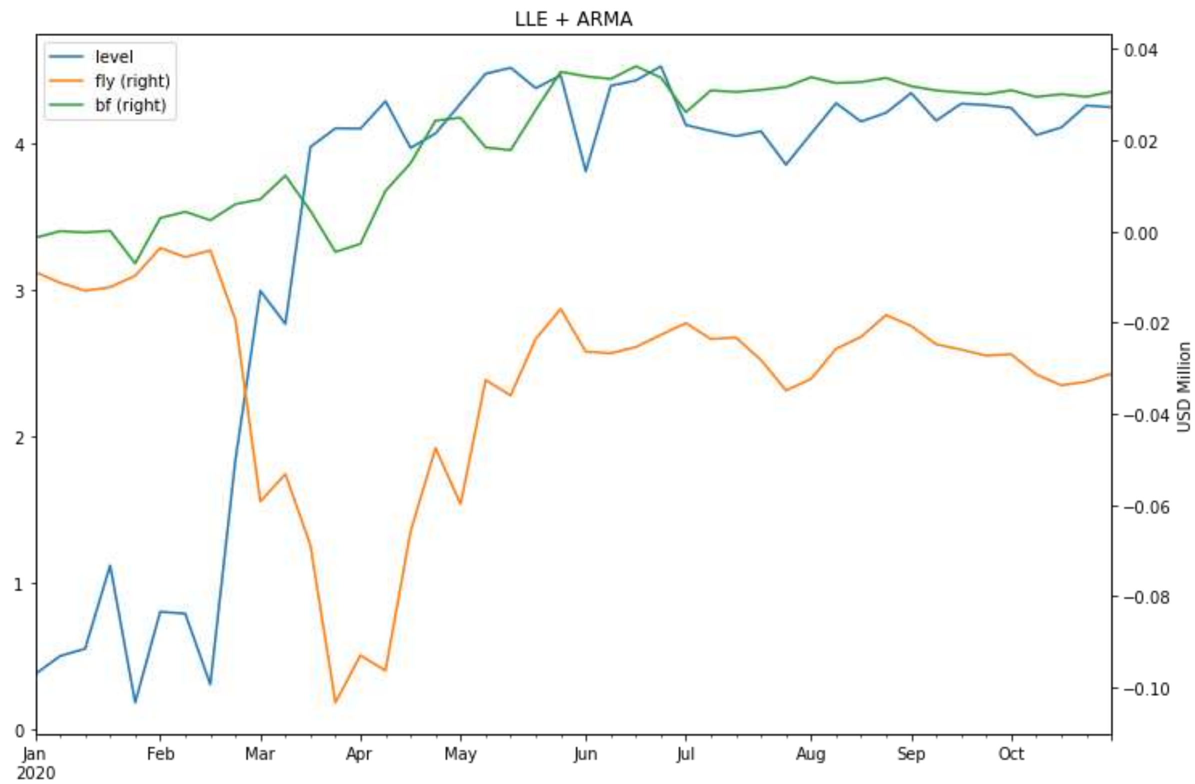

Fig. 12. PnL of multiple tenor trading using LLE+ARMA (basis spread curve)

V. Conclusion and Future Work

The first part in future work lies in the interest rates simulations. We will take a parametric model, the Ornstein-Uhlenbeck process (equation (3) and (4)), to estimate the parameters of underlying interest rates process for Treasury, LIBOR, SOFR, and OIS respectively. With the calibrated process, we simulate a large number of realizations (e.g. 10,000).

$$\mathrm{d}S_t = \lambda(\mu - S_t)dt + \sigma dW_t, \qquad (7)$$

$$S_{(t+\delta)} = S_t e^{-\lambda\delta} + \mu(1 - e^{-\lambda\delta}) + \sigma\sqrt{\frac{1-e^{-\lambda\delta}}{2\lambda}}\,\mathrm{N}(0, 1), \qquad (8)$$

For each simulated scenario, we will obtain two panel data for the swap spread and basis spread respectively. Each of them 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.

The second part will lie in applying the Hessian regularized nonlinear time series forecast model (HRM). The basic idea is to fit a model via penalization, where the penalty term is an unbiased estimator of the integrated Hessian of the underlying function. This will serve as an advanced time series forecasting approach benchmarked against the ARMA model in our work.

The third part will lie in exploring more dimension reduction techniques. We would like to implement Multidimensional Scaling (MDS) and t-distributed Stochastic Neighbor Embedding (t-SNE), which belong to the category of semi-classical methods. Compared to the direct eigen-analysis of the N data points themselves 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[J]. 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[J]. Recent advances in data mining of enterprise data, 2007: 691-745.

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

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

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

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