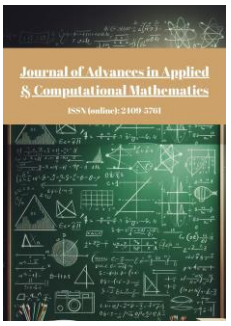
Published by Avanti Publishers

Journal of Advances in Applied & Computational Mathematics

ISSN (online): 2409-5761



A Probabilistic Spatial Interaction Model for Bond Trading Forecasting

Kateryna Kustarova[1,*], Andrey Dorogovtsev[1] and Dmitriy Dorogovtsev[2]

[1]*Institute of Mathematics of the NAS of Ukraine, 3 Tereshchenkivska Street, Kyiv 01024, Ukraine*

[2]*Luxoft, a DXC Technology Company, Business Center IRVA, 10/14 Volnovaska Street, Kyiv 03124, Ukraine*

ARTICLE INFO

Article Type: Research Article

Academic Editor: Ali Akbar

Keywords:
Yield curve
Bond trading
Svensson model
Probabilistic model
Nelson-Siegel model
Financial forecasting

Timeline:
Received: May 20, 2026
Accepted: June 22, 2026
Published: June 29, 2026

Citation: Kustarova K, Dorogovtsev A, Dorogovtsev D. A probabilistic spatial interaction model for bond trading forecasting. J Adv Appl Computat Math. 2026; 13(1): 105-133.

DOI: https://doi.org/10.15377/2409-5761.2026.13.7

*Corresponding Author
Email: katekustareva@gmail.com

ABSTRACT

This paper proposes a constrained spatial transition model for forecasting bond-trading activity at the transaction level. Instead of modeling only aggregate yield-curve characteristics, the method represents daily trades as binary states on a two-dimensional tenor–G-spread grid. For each cell, the probability of activity on the next trading date depends on the current state of the same cell and on the numbers of active cells in the first and second neighborhood rings. Linear and quadratic neighborhood terms are included to capture local interaction effects. Explicit parameter constraints keep all transition probabilities strictly between zero and one, which provides a valid finite-state Markov model and supports results on existence, irreducibility, aperiodicity, ergodicity, identifiability, and consistency of the constrained least-squares estimator.

The empirical study uses 17,943 bond observations for six medical-sector issuers over 345 trading dates. The data are divided chronologically into training, validation, and held-out test periods. Grid resolution and neighborhood size are selected using validation data only. Statistical significance is assessed using moving-block bootstrap confidence intervals, circular-shift permutation tests, and Benjamini–Hochberg correction. On the held-out test sample, the proposed model achieves an F1 score of 0.697, ROC–AUC of 0.842, PR–AUC of 0.640, Brier score of 0.101, and log loss of 0.321. It outperforms persistence, a first-order Markov baseline, and logistic regression, while random forest and gradient boosting provide slightly higher predictive accuracy. Predicted active cells are also used to reconstruct continuous G-spread curves with modified Nelson–Siegel and Svensson forms. The modified Nelson–Siegel reconstruction achieves RMSE 0.694, MAE 0.551, and $R^2 = 0.9967$ on held-out dates.

JEL Classification: 91B26, 60G35, 91G30, 91B70, 60J20, 62M20.

1. Introduction

In this article we propose a mathematical model for bond trading. Our model focuses on the trading process itself and tries to use all available information, in contrast to existing models that primarily concentrate on the time series of several numerical characteristics of the trading process. The descriptions obtained in such models commonly have a disadvantage. They focus on a fixed number of characteristics and their prediction.

Consequently, the predictions of the desired values take into account only a limited portion of the information from the previous trading process, potentially leading to less accurate forecasts. For example, the Black-Derman-Toy model [1], which was introduced by Fischer Black, Emanuel Derman and Bill Toy, is a one-factor model that uses the short rate, the interest rate for a period of one year, to price bond options, swaptions and other interest rate derivatives. The standard Black-Derman-Toy model is designed to use both the current term structure of zero-coupon yields and the term structure of yield volatilities for the same bonds, utilizing a binomial lattice framework. The three-factor convergence model of interest rates by Beata Stehlikova and Zuzana Zikova [2] incorporates the following factors: the domestic short rate, the European short rate and the dynamics of convergence process, describing how the domestic and European short-term interest rates converge or interact over time. These factors collectively help in explaining the evolution of the domestic short rate in connection with the European rate, offering a nuanced approach to modeling interest rates.

Mentioned models are focusing only on specific characteristics, rather than trying to cover an entire process behavior. In contrast to these approaches we propose the model for the trading process itself. The idea of our model can be briefly explained as follows. In a fixed day for a fixed company (or for a fixed set of companies) all trading events can be treated as the points in the rectangle spread $\times$ tenor. Reasonably these points are random, depending on one another and from the points at the previous day. In our model we attempt to forecast the points on subsequent trading dates taking into account present points in the neighbour positions. Correspondingly to this idea we built the model in several steps.

Firstly we discuss the notion of neighbours. Trading data values for a specific day are placed on a coordinate grid using tenor and spread as axes, so that each data entry occupies its own cell. For each cell, the closest neighbours are defined as the set of filled adjacent cells within a radius of 1; more neighbors are added as the radius increases. The second step is devoted to approximation of conditional probability of point appearance on the next day, based on the amount of cell neighbours for the current day. Here we propose different approaches that work effectively for different companies and numbers of neighbours, using the least-squares method and second degree polynomial model. Further, we present the prediction ability of the model, forecasting data values for the next day. Accordingly, the statistical analysis will consist of three parts devoted to prediction and simulation.

In the next section, this model is illustrated using data examples, and trading data are placed on a coordinate grid. The concept of a neighborhood circle for each cell is introduced, and the correlation between the neighborhood circle and the point's activity is analyzed. A new element is introduced: the cell merit that may depend on the amount of the nearest neighbours. The method for finding the model's coefficients is explained. In Section 3 we review well-known Nelson-Siegel [3] and Svensson [4] yield curve models that are used for modeling the term structure of interest rates, also known as yield curve. The Nelson-Siegel model represents the yield curve using a parametric equation with three components: level, slope and curvature. The Svensson model extends the Nelson-Siegel model by adding another curvature term, allowing for more flexibility in fitting the yield curve. These models are widely used in finance to estimate current yield curves, forecast future interest rates, and analyze market expectations. We use our proposed model to predict bond trading data and construct Nelson-Siegel and Svensson yield curves based on these predictions. The results demonstrate that the forecasted curves closely match those built on the original data with high accuracy.

2. Literature Review

2.1. Classical Yield Curve Models

Nelson and Siegel [3] introduce a simple parametric model of the government bond yield curve that captures the level, slope, and curvature of yields across maturities. Their approach demonstrates simplicity and strong

empirical performance and focuses on modeling yield curve dynamics at an aggregate level rather than using transaction-level information.

Svensson [4] extends the Nelson–Siegel model by adding an additional curvature component, which allows for greater flexibility when fitting observed bond yield data. This extended version improves the ability to represent complex yield curve shapes across different maturities. The model keeps the same parametric structure and focuses on describing the yield curve using aggregated yield data.

Vasicek [5] proposes a continuous-time stochastic model in which the short-term interest rate follows a mean-reverting Ornstein–Uhlenbeck process. This specification leads to a closed-form solution for bond prices and the resulting term structure under equilibrium conditions. The model introduced stochastic interest rate dynamics and risk premia into bond pricing theory. The approach describes the evolution of interest rates at a macro level using continuous stochastic processes.

Cox, Ingersoll, and Ross [6] develop a continuous-time stochastic model of interest rates in which the short rate follows a mean-reverting square-root diffusion process that guarantees non-negative interest rates. The approach describes aggregate interest rate dynamics within a continuous-time framework.

Further developments introduced arbitrage-free models that can be calibrated to the initially observed term structure. Ho and Lee [7] propose a binomial lattice model of term-structure movements, while Hull and White [8] extend the mean-reverting Vasicek model by allowing time dependent parameters. Heath, Jarrow, and Morton [9] provide a more general framework in which the entire instantaneous forward rate curve is modeled directly and its drift is restricted by the absence of arbitrage. Duffie and Kan [10] subsequently develop a broad class of multifactor affine term-structure models in which bond yields are affine functions of latent state variables.

In addition to stochastic term-structure models, several studies focus on the empirical construction of yield and forward curves from bond prices. McCulloch [11] introduces a flexible spline based method for estimating discount functions, whereas Fama and Bliss [12] construct forward rates and investigate the information about future rates contained across maturities. These approaches are primarily concerned with obtaining a smooth aggregate representation of the term structure and do not explicitly model the individual trade events from which the observed curve is formed.

These classical models provide simple and theoretically grounded descriptions of the yield curve and its dynamics. At the same time, they operate primarily at the level of aggregated interest rate dynamics and do not explicitly incorporate information about individual trading events or interactions between trades. This motivates the exploration of alternative modeling approaches that incorporate additional structural information about trading activity.

2.2. Yield Curve Forecasting Models

Diebold and Li [13] propose a dynamic extension of the Nelson–Siegel yield curve model in which the level, slope, and curvature factors evolve over time according to a vector autoregressive process. This dynamic model allows effective forecasting of the entire term structure and combines simplicity with strong empirical performance.

Duffee [14] studies the forecasting performance of affine term structure models, a class of models in which bond yields are expressed as a linear function of a set of latent state variables that follow stochastic processes. In these models the short-term interest rate and other factors evolve according to diffusion processes, and bond prices can be written in an exponential-affine form with coefficients that depend on time to maturity. The model parameters typically include the speed of mean reversion, long-run mean levels, volatility terms, and risk premia associated with the state variables. The author evaluates how well such models predict future interest rates and term premia and discusses practical challenges in out-of-sample forecasting.

Ang and Piazzesi [15] develop a no-arbitrage vector autoregression (VAR) model that describes the joint dynamics of bond yields across maturities under theoretical pricing restrictions. Their approach combines macroeconomic variables with yield curve dynamics and provides a practical framework for forecasting while preserving arbitrage-free conditions. The model treats yields as interconnected time-series variables within a macro-financial setting.

Some forecasting models also incorporate macroeconomic information. Diebold, Rudebusch, and Aruoba [16] model the level, slope, and curvature factors with macroeconomic variables. Hördahl, Tristani, and Vestin [17] combine macroeconomic dynamics and the term structure within a no-arbitrage framework, while Moench [18] uses a factor-augmented VAR to exploit a large macroeconomic information set.

Other studies emphasize either arbitrage-free restrictions or high dimensional forecasting methods. Christensen, Diebold, and Rudebusch [19] develop an affine arbitrage-free version of the dynamic Nelson–Siegel model. Carriero, Kapetanios, and Marcellino [20] apply large Bayesian vector autoregressions to jointly forecast yields across maturities, and Favero, Niu, and Sala [21] compare the forecasting contribution of no-arbitrage restrictions with that of large information sets. Although these models use rich temporal and cross maturity information, their basic observations remain aggregated yield values rather than individual bond-trading events.

Forecasting models such as the dynamic Nelson–Siegel approach, affine forecasting models, and no-arbitrage VAR models extend classical term structure models into time-series forecasting frameworks. These approaches significantly improve empirical forecasting performance and incorporate macroeconomic information or theoretical pricing restrictions. At the same time, they continue to treat yields as aggregated time-series variables and do not explicitly model the underlying structure of trading activity. Our approach complements these models by focusing on the stochastic behavior of trade events that precedes the formation of the yield curve.

2.3. Spatial and Interaction-Based Models

Besag [22] develops a statistical approach for studying spatial interaction in lattice systems. The paper introduces Markov random fields, where the value at each location depends on the values of its neighboring locations. This framework shows how spatial dependence can be modeled in binary or categorical grid systems.

Geman and Geman [23] provide a probabilistic framework for modeling spatially dependent systems using Gibbs distributions. Their work shows how global patterns in a grid can arise from local interactions between neighboring elements. The paper establishes the theoretical connection between local conditional relationships and the overall probability distribution of the system. Although originally developed for image analysis, this framework can be applied to many systems where variables interact locally on a grid.

Anselin [24] introduces the field of spatial econometrics and presents econometric models that explicitly account for spatial dependence between observations. The book develops spatial autoregressive and spatial error models and shows that ignoring spatial interactions can lead to biased or inefficient statistical estimates. These methods provide tools for analyzing economic systems in which observations influence each other across neighboring units.

Besag [25] introduces the auto-logistic model, in which the probability of a binary state depends on the states of its nearest neighbors. Kindermann and Snell [26] describe the theory of Markov random fields and neighborhood systems and explain how local interactions determine the behavior of the whole system. These ideas are closely related to the proposed model because bond-trading activity is also represented by binary variables on a discrete grid.

Interaction-based models have also been widely studied in probability theory. Cont and Bouchaud [27] show how interactions between market participants can produce large market fluctuations. Lux and Marchesi [28] develop a model in which interactions between different types of traders generate complex market behavior. Georgii [29] provides a comprehensive treatment of Gibbs measures and interacting particle systems, showing how global probabilistic behavior can arise from local interactions on discrete lattices.

Taken together, these models show that complex system behavior can arise from simple local interactions. Markov random fields and auto-logistic models describe local dependence on a discrete grid, while interacting particle systems explain how neighboring elements influence one another over time.

The proposed point estimate model follows the same idea. Trades are represented as binary values on a two-dimensional tenor-spread grid, and the probability of a trade on the next day depends on the current state of the cell and the activity of its neighboring cells. Unlike traditional spatial models, the neighborhood is defined by

similarity in maturity and spread rather than by geographical distance. This approach makes it possible to study local interactions between trades and to forecast the transaction-level data used to construct the yield curve.

3. Bond Trading Representation

The model describes the evolution of all trades originated by the bond issuer, who can be a single company or a set of companies (sector). The work was carried out using data received from the vendor (Bloomberg). Usually the data provided by the vendor is presented in a table containing following fields (Table **1**):

- Tenor is the year fraction from current date and the date the bond is expected to mature or be called away by the issuer. This latter date is called the workout date
- Issuer is a Bond's issuer name
- BbgGSpd is a G-spread value obtained from Bloomberg
- Date is an effective date of the dataset

Table 1: Example of bond trading data received from Bloomberg

Issuer	Tenor	BbgGSpd	Date
The Goldman Sachs Group, Inc.	10.362585	207.275001	2022-10-06
The Goldman Sachs Group, Inc.	2.354511	146.270003	2022-10-06
The Goldman Sachs Group, Inc.	4.040997	179.849997	2022-10-06
The Goldman Sachs Group, Inc.	15.068868	217.270002	2022-10-06
The Goldman Sachs Group, Inc.	3.389400	137.605002	2022-10-06
The Goldman Sachs Group, Inc.	2.485925	112.380000	2022-10-06
The Goldman Sachs Group, Inc.	4.437979	173.304996	2022-10-06
The Goldman Sachs Group, Inc.	23.041351	186.949997	2022-10-06
The Goldman Sachs Group, Inc.	3.175852	153.134997	2022-10-06
The Goldman Sachs Group, Inc.	8.541946	206.475004	2022-10-06

Source: The Goldman Sachs Group, Inc.

Our approach to describing bond trades is as follows. On a fixed day all trades with selected bonds are presented by the points in coordinates tenor (horizontal), BbgGSpd (vertical) (Fig. **1**).

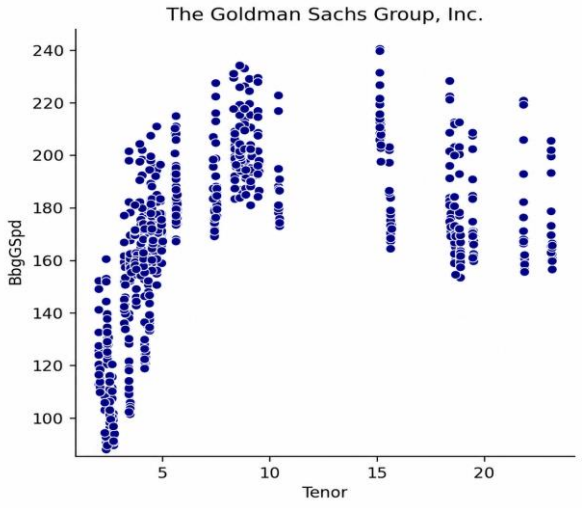


Figure 1: Bonds of The Goldman Sachs Group, Inc. for September 2022.

We are going to investigate the evolution of the obtained set of points. The Fig. (**2**) displays completed trades for The Goldman Sachs Group, Inc. bonds over three selected days with two-week intervals. It can be seen that the placement of points changes over time.

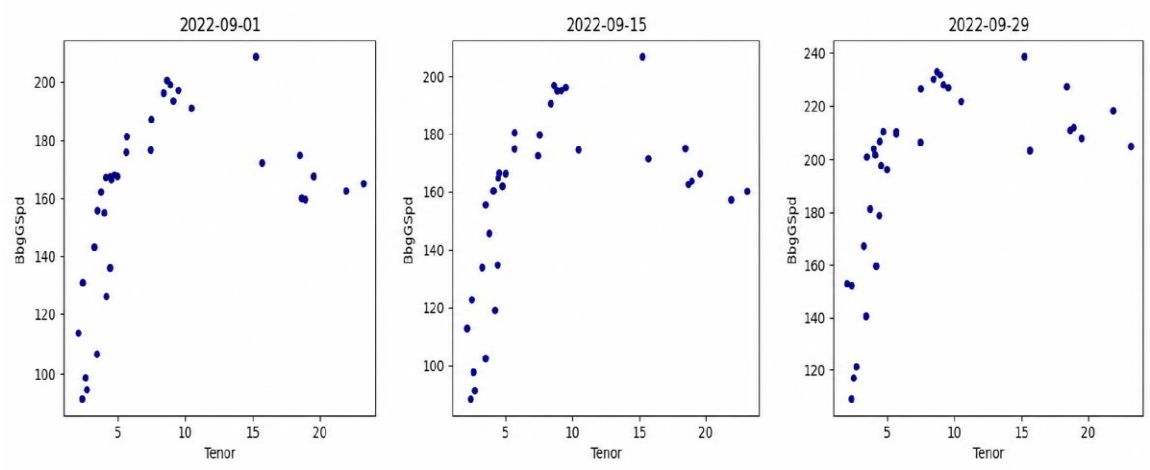


Figure 2: Trades of The Goldman Sachs Group, Inc. on three selected dates separated by two-week intervals.

The main idea of our approach is to predict the presence of at least one point in any selected sector on the day following the observed day after dividing the entire area into a coordinate grid. In order to do this, the BbgGSpd values of the bonds are divided based on the tenor over time into a fixed number of squares, as shown in the Fig. (**3**).

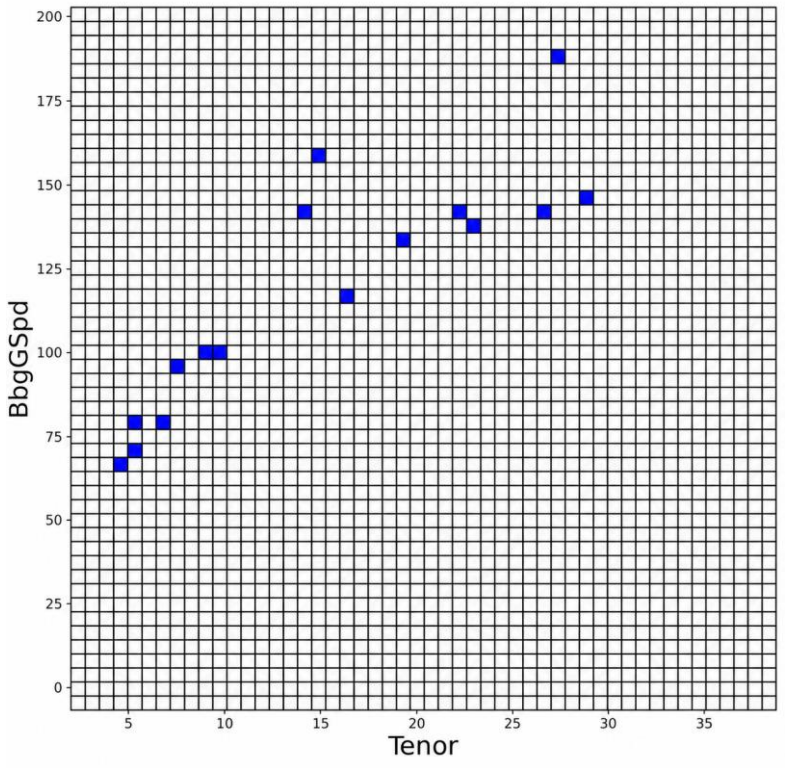


Figure 3: Partition of the tenor–G-spread plane into grid cells.

Using the formulas provided below, we calculate the steps needed to construct the grid. The grid resolution is selected using the training and validation data. If several bond observations fall into the same cell on the same date, they are aggregated into a single binary activity indicator. Let us define the partition step for each of the axes in a following way:

$$step_{\text{tenor}} = \frac{max(\text{tenor}) - min(\text{tenor})}{n}$$

$$step_{\text{BbgGSpd}} = \frac{max(\text{BbgGSpd}) - min(\text{BbgGSpd})}{n}$$

Neighbor cells of radius R for given cell C are determined as a set of filled cells inside square with side (2R + 1), centered around C and excluding C (neighborhood). Fig. (**4**) below shows three sets of neighbor cells with radius increasing from 1 to 3, marked in green, blue, and yellow color respectively. The grid size is selected according to the required level of spatial resolution. A trade-off must be maintained between predictive accuracy and the size of the grid cells. Increasing the cell size aggregates more observations within each cell and generally improves the predictive performance of the model, but reduces the spatial precision of the forecast. Conversely, decreasing the cell size provides a more detailed representation of the trading space, but leads to sparser data and, consequently, lower prediction accuracy.

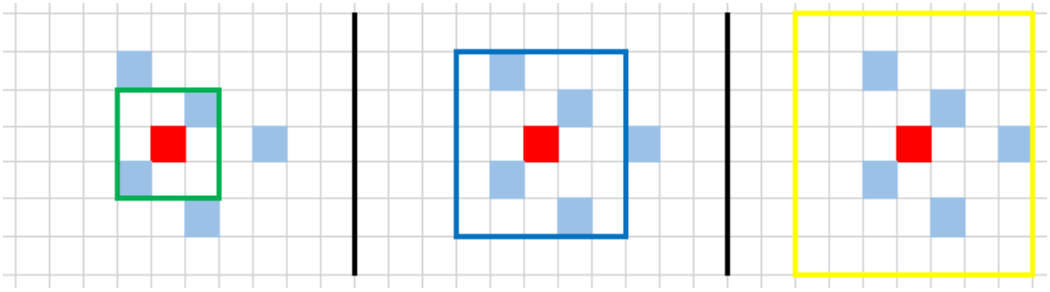


Figure 4: Examples of neighborhood configurations with radii *R* =1, *R* =2, and *R* =3, containing 2, 4, and 5 active neighboring cells, respectively.

It was observed that the appearance of a filled cell on the next day depends on the presence and number of its neighbors on the previous day. The following Fig. (**5**) show a situation where an empty cell with existing neighbors became filled the next day, and where a filled cell without neighbors became empty the next day:

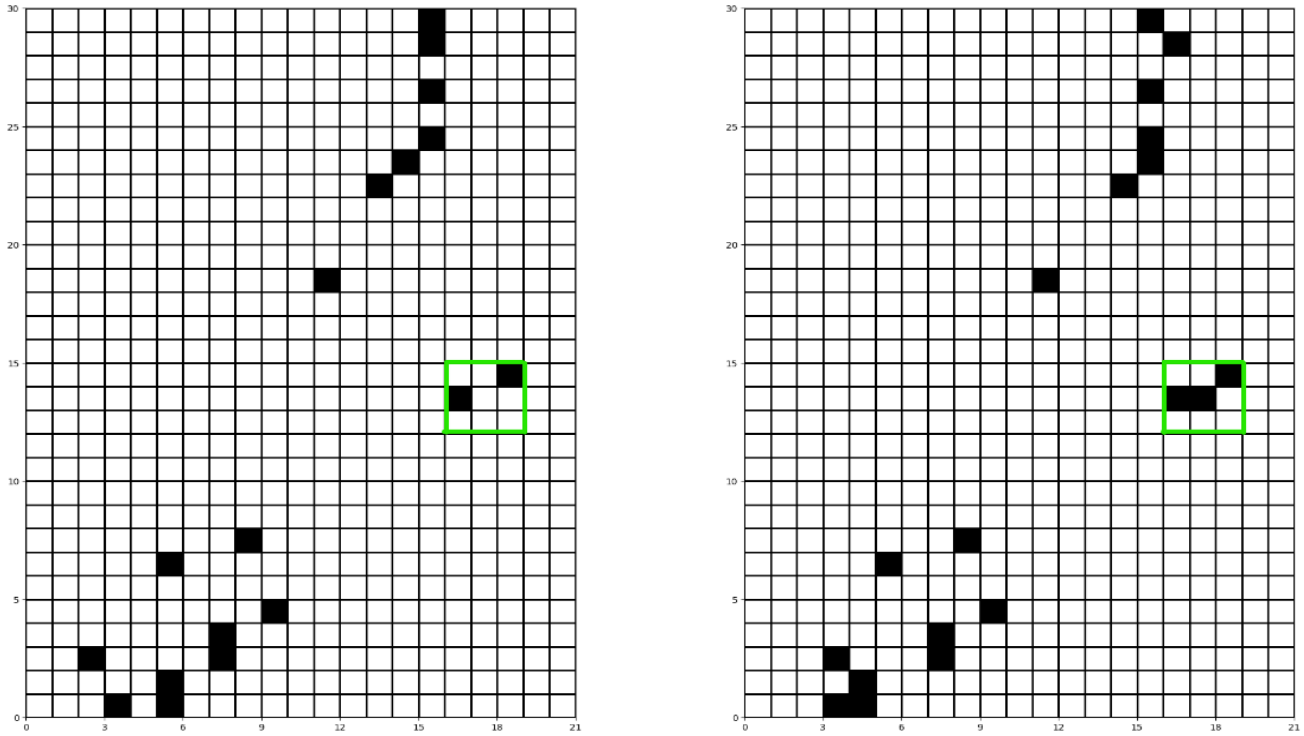


Figure 5: Example in which an inactive cell becomes active on the following trading date: AbbVie Inc., 27 April 2021.

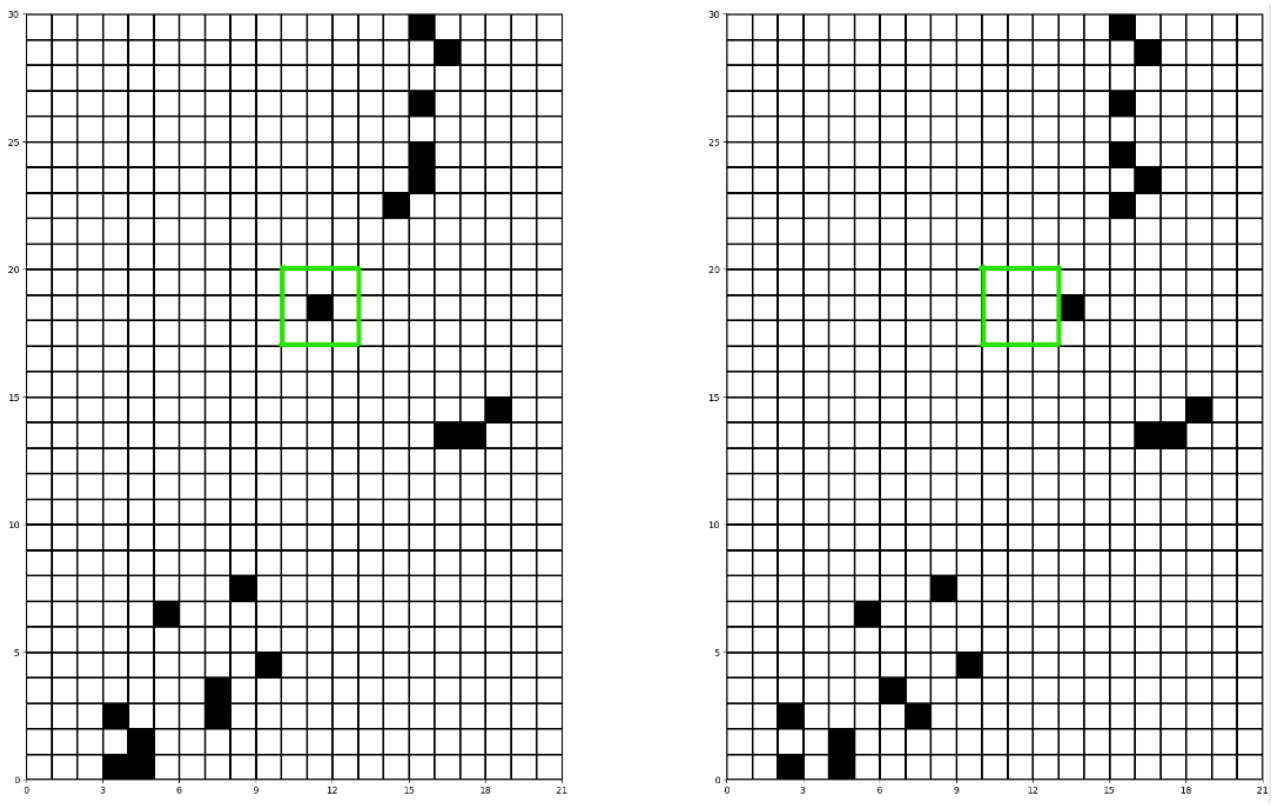


Figure 6: Example in which an active cell becomes inactive on the following trading date: AbbVie Inc., 29 April 2021.

The Fig. (**6**) illustrate examples of neighborhood configurations for a selected reference cell (highlighted in green). The surrounding cells are organized into concentric neighborhoods (or "circles") based on radius r, where each circle includes all cells within a specified distance from the central one, excluding the center itself. The number and extent of these circles are not fixed and can vary depending on the observed correlation structure. The choice of radius r is guided by the empirical correlation between the activity of the central cell and its neighbors.

The correlation between a cell's activity and the number of its active neighbors is measured using the Pearson correlation coefficient, defined as:

$$r_{xy} = \frac{\sum_{i=1}^{n}\left(\varepsilon_i^n - \overline{\varepsilon}^n\right)\left(\eta_i^n - \overline{\eta}^n\right)}{\sqrt{\sum_{i=1}^{n}\left(\varepsilon_i^n - \overline{\varepsilon}^n\right)^2}\sqrt{\sum_{i=1}^{n}\left(\eta_i^n - \overline{\eta}^n\right)^2}}$$

where:

- ε_i^n denotes the binary activity indicator of the central cell on day i, equal to 1 if the cell is active and 0 otherwise;
- η_i^n denotes the number of active neighboring cells on day i;
- $\overline{\varepsilon}^n$ and $\overline{\eta}^n$ denote the sample means of ε_i^n and η_i^n, respectively, defined by

$$\overline{\varepsilon}^n = \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^n, \qquad \overline{\eta}^n = \frac{1}{n}\sum_{i=1}^{n}\eta_i^n;$$

- n is the total number of observed trading days.

This coefficient quantifies the strength and direction of the linear relationship between a cell's activity and the activity in its local neighborhood.

This formula is used to assess the correlation between the activity of a square and the number of active neighbors.

The Pearson coefficient was also selected because it is simple, interpretable, and directly comparable across cells, companies, and neighborhood radii. Nevertheless, since it captures only linear dependence, its results are interpreted together with empirical conditional probabilities and out-of-sample forecasting performance.

The Fig. (**7**) displays the correlation matrix for a selected target cell (marked in red) corresponding to Goldman Sachs bond trades. Each cell represents the Pearson correlation coefficient between the activity of the target cell and its neighboring cells across all observed days. Empty or white cells indicate locations with insufficient data or zero correlation. The plot demonstrates that significant correlations are not uniformly distributed, highlighting the localized and asymmetric nature of bond trading interactions across the tenor–spread grid.

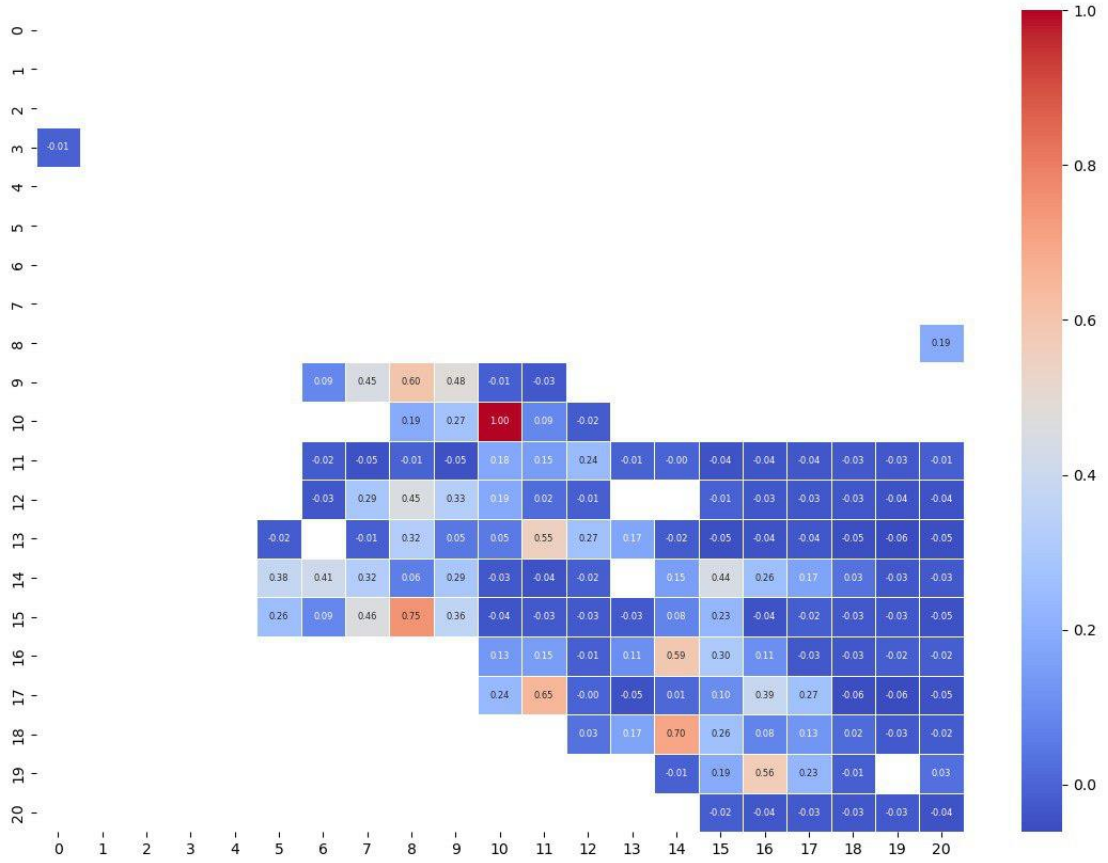


Figure 7: Correlation matrix for The Goldman Sachs Group, Inc.

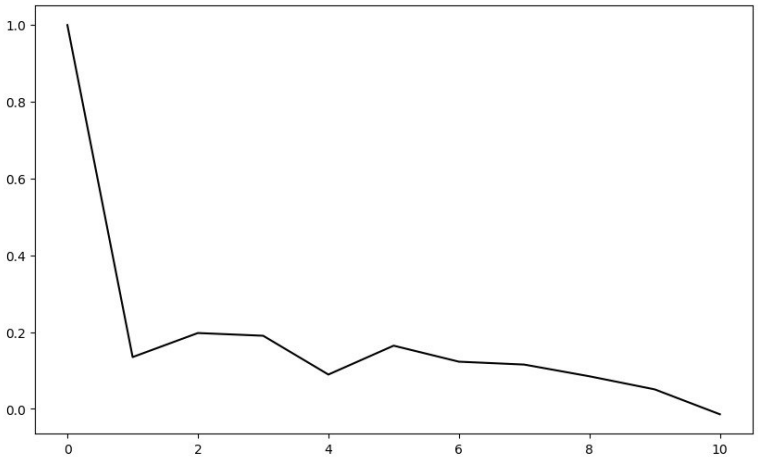


Figure 8: Average correlation as a function of the neighborhood radius.

When aggregating all trading days into a single plot, it becomes evident that the frequency of bond trades varies across the grid Fig. (**8**). The heatmap below illustrates the empirical density of observed trades for Goldman Sachs bonds. Brighter regions indicate grid cells where trades occurred more frequently over time. Given the relatively low overall liquidity, it is reasonable to assume that the likelihood of a trade occurring at a specific location on the following day depends on its historical activity and proximity to other frequently traded cells. In particular, higher local density suggests a greater probability of observing a trade in that cell in the near future.

3.1. Statistical Significance

To assess whether the observed neighborhood effects can be distinguished from sampling variation, we performed a formal significance analysis using the Goldman Sachs data employed in the correlation and density illustrations.

The dataset contains 10,858 bond observations covering 734 trading dates, of which 345 dates contain at least one Goldman Sachs observation. Following the grid construction used in the heat-map analysis, the tenor–spread domain was divided into a 60×60 grid. Cells active on more than 10 days were retained, resulting in 354 eligible cells and 733 consecutive day-to-day transitions.

For each eligible cell v, we considered the following lagged association:

$$\rho_r = Corr(X_{t+1}(v), N_t^{(r)}(v)), \qquad r = 1,2. \tag{1}$$

where $X_{t+1}(v)$ is the binary activity indicator of the central cell on the following day, and $N_t^{(r)}(v)$ is the number of active cells in the r-th neighborhood circle on the current day. The first circle contains 8 cells, whereas the second circle corresponds to the outer ring and contains 16 cells.

For each radius, we tested the following hypotheses:

$$H_0: \rho_r = 0.$$

$$H_1: \rho_r \neq 0.$$

Because observations belonging to the same day are spatially dependent and consecutive trading days may also be temporally dependent, uncertainty was estimated using a moving-block bootstrap applied to complete daily grid configurations. We used 5,000 bootstrap replications and a block length of 10 trading days.

The statistical significance of the pooled lagged correlation was additionally assessed using a circular-shift permutation test. This procedure preserves the internal temporal structure of both sequences while destroying their contemporaneous alignment.

For the first neighborhood circle, the pooled lagged Pearson correlation was: $\hat{\rho}_1 = 0.301$.

The corresponding 95% moving-block bootstrap confidence interval was: [0.277, 0.327].

The circular-shift permutation p-value was: p = 0.0027.

For the second neighborhood circle, the pooled lagged Pearson correlation was: $\hat{\rho}_2 = 0.215$.

The corresponding 95% confidence interval was:[0.190, 0.242].

The circular-shift permutation p-value was: p = 0.0164.

Cell-level Pearson correlation tests were also performed and adjusted for multiple comparisons using the Benjamini–Hochberg false discovery rate procedure [30].

At the 5% false discovery rate level, significant positive lagged associations were identified in 336 of the 354 eligible cells, corresponding to 94.9%, for the first neighborhood circle. For the second neighborhood circle, significant positive associations were identified in 267 cells, corresponding to 75.4% Fig. (**9**).

The median cell-level correlations were 0.304 for the first neighborhood circle and 0.225 for the second neighborhood circle Fig. (**10**).

These results reject the null hypothesis of zero lagged neighborhood association for both neighborhood circles at the 5% significance level. The weaker estimate obtained for the second circle is consistent with the observed decay in dependence as the neighborhood radius increases.

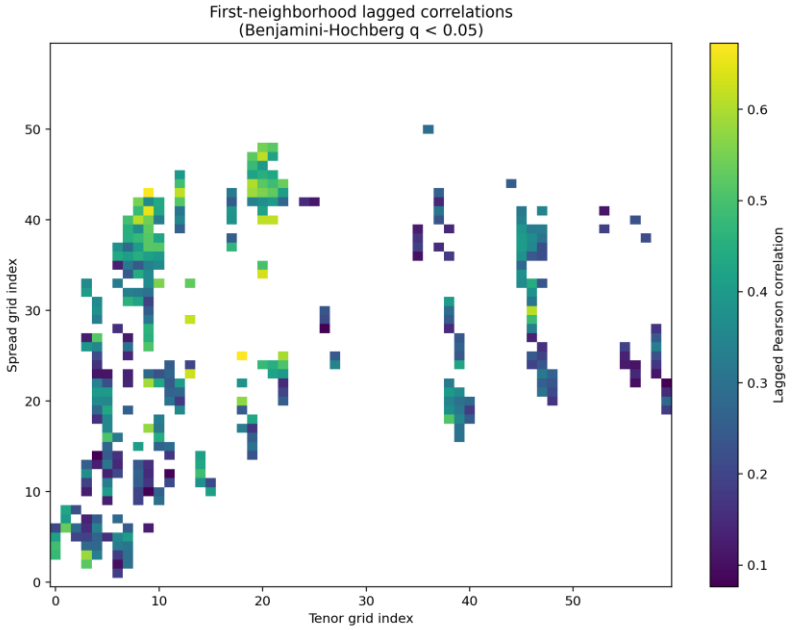


Figure 9: Significance-masked heat maps of lagged neighborhood associations for the first and second neighborhood circles. Only cells satisfying the Benjamini–Hochberg criterion $q < 0.05$ are displayed.

The significance-masked heat maps display only those cells that satisfy the Benjamini–Hochberg criterion of q < 0.05.

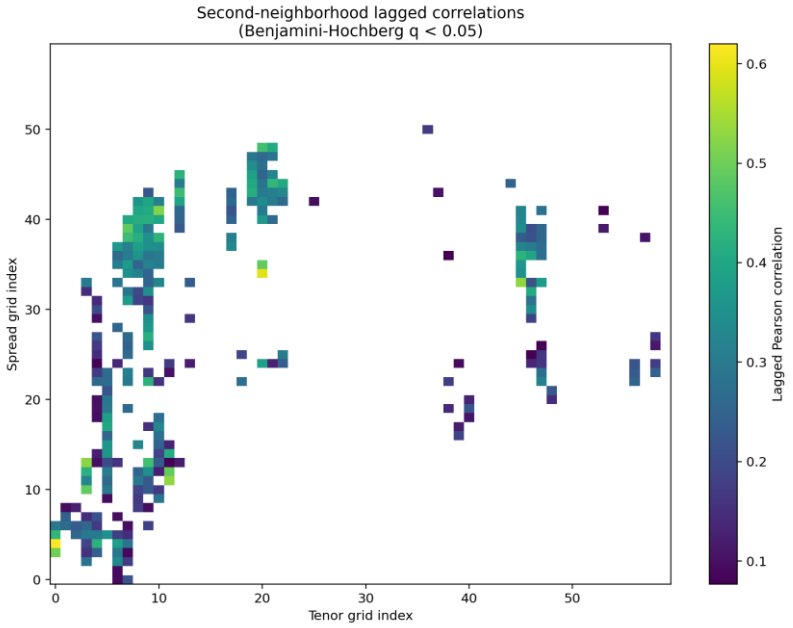


Figure 10: Lagged Pearson correlations between the next-day state of each cell and the number of active cells in its second neighborhood circle. Cells that are not statistically significant after the Benjamini–Hochberg correction at $q < 0.05$ are masked.

When all trading days are aggregated into a single plot, it becomes evident that the frequency of bond trades varies across the grid. The heat map below Fig. (**11**) illustrates the empirical density of observed Goldman Sachs bond trades.

Brighter regions indicate grid cells in which trades occurred more frequently over time. Given the relatively low overall liquidity, it is reasonable to assume that the probability of a trade occurring at a particular location on the following day depends on the cell's historical activity and its proximity to other frequently traded cells.

In particular, a higher local density may indicate a greater probability of observing a trade in that cell in the near future.

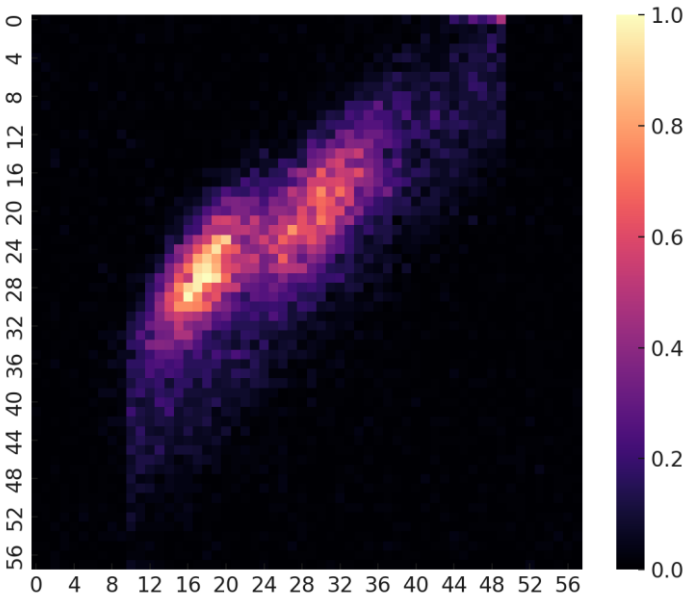


Figure 11: Frequency of observations across grid cells for The Goldman Sachs Group, Inc.

4. Constrained Spatial Transition Model

In this paper, we propose the following model for the evolution of the point configuration described above. At each time step, points may appear in the grid cells independently, conditional on the configuration at the previous time step. However, the probability of appearance in a given cell depends on the points located in its neighboring cells at the previous step.

Such stochastic evolution can be interpreted as a motion of interacting particles. Related measure-valued processes and stochastic flows were studied by A. A. Dorogovtsev and are described in detail in [31].

We investigate whether the activity of a grid cell on the next trading date is associated with the current activity of the same cell and with the number of active cells in its local neighborhood. The economic motivation is that bonds issued by the same company or by companies in the same sector may respond to common information, producing localized clusters of trading activity in the tenor–G-spread plane.

4.1. Grid, Cell States, and Neighborhood Rings

Let

$$G_n = \{(i,j): 1 \le i,j \le n\} \tag{2}$$

be a finite rectangular grid obtained by discretizing the tenor–G-spread domain. For each trading date $t \in N_0$, define a binary random field

$$X_t: G_n \to \{0,1\}, \tag{3}$$

where $X_t(v) = 1$ if at least one trade is observed in cell $v \in G_n$ on date t, and $X_t(v) = 0$ otherwise.

For $v = (i, j)$, define two disjoint neighborhood rings:

$$\mathrm{R}_1(v) = \{u \in G_n: \|u - v\|_\infty = 1\}, \tag{4}$$

$$\mathrm{R}_2(v) = \{u \in G_n: \|u - v\|_\infty = 2\}. \tag{5}$$

The first ring contains at most 8 cells. The second ring is the outer ring at distance two and contains at most 16 cells. For boundary cells the corresponding cardinalities may be smaller. The lagged neighborhood counts are

$$N_t^{(r)}(v) = \sum_{u \in \mathrm{R}_r(v)} X_t(u), \qquad r = 1,2. \tag{6}$$

Thus,

$$0 \le N_t^{(1)}(v) \le 8, \qquad 0 \le N_t^{(2)}(v) \le 16. \tag{7}$$

The notation used in the initial presentation is identified with the grid notation by

$$\varepsilon_0^t \equiv X_t(v), \qquad \varepsilon_1^t \equiv N_t^{(1)}(v), \qquad \varepsilon_2^t \equiv N_t^{(2)}(v). \tag{8}$$

4.2. One-step Transition Probability

For a parameter vector

$$\theta = (\alpha_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5)^{\mathrm{T}}, \tag{9}$$

define

$$p_{\theta,v}(X_t) = \alpha_0 X_t(v) + \beta_1 N_t^{(1)}(v) + \beta_2 (N_t^{(1)}(v))^2 + \beta_3 N_t^{(2)}(v) + \beta_4 (N_t^{(2)}(v))^2 + \beta_5. \tag{10}$$

The model specifies

$$\mathrm{P}_\theta(X_{t+1}(v) = 1 | X_t) = p_{\theta,v}(X_t). \tag{11}$$

Equivalently, for $k_0 \in \{0,1\}$, $k_1 \in \{0, \dots, 8\}$, and $k_2 \in \{0, \dots, 16\}$,

$$p_{t+1}(k_0, k_1, k_2) = \mathrm{P}(\varepsilon_0^{t+1} = 1 | \varepsilon_0^t = k_0, \varepsilon_1^t = k_1, \varepsilon_2^t = k_2) \tag{12}$$

$$= \frac{\mathrm{P}(\varepsilon_0^{t+1}=1, \varepsilon_0^t=k_0, \varepsilon_1^t=k_1, \varepsilon_2^t=k_2)}{\mathrm{P}(\varepsilon_0^t=k_0, \varepsilon_1^t=k_1, \varepsilon_2^t=k_2)}, \tag{13}$$

whenever the denominator is positive. The fitted polynomial approximation is

$$\hat{p}_{t+1}(k_0, k_1, k_2) = k_0\alpha_0 + k_1\beta_1 + k_1^2\beta_2 + k_2\beta_3 + k_2^2\beta_4 + \beta_5. \tag{14}$$

The coefficient α_0 represents persistence of the central cell. The coefficients β_1 and β_2 describe the linear and quadratic effects of the first neighborhood ring, while β_3 and β_4 describe the corresponding effects of the second ring. The intercept β_5 represents baseline activity. Since the regressors have different scales, the magnitudes of the coefficients should not be interpreted as directly comparable correlation coefficients.

4.3. Probability Restrictions

The expression $p_{\theta,v}(X_t)$ represents a Bernoulli transition probability and must therefore take values strictly between zero and one for every admissible grid configuration. To guarantee this property uniformly over all cells

and configurations, let

$$0 < \delta < \tfrac{1}{2}, \tag{15}$$

where δ is a fixed small constant.

We impose the following non-negativity and lower-bound restrictions:

$$\alpha_0 \geq 0, \qquad \beta_k \geq 0, \quad k = 1, \dots, 4, \qquad \beta_5 \geq \delta. \tag{16}$$

In addition, we require

$$\alpha_0 + 8\beta_1 + 64\beta_2 + 16\beta_3 + 256\beta_4 + \beta_5 \leq 1 - \delta. \tag{17}$$

The corresponding feasible parameter set is

$$\begin{aligned} \Theta_\delta = \{\theta \in \mathrm{R}^6: \quad & \alpha_0 \geq 0,\ \beta_k \geq 0, \quad k = 1, \dots, 4,\ \beta_5 \geq \delta, \\ & \alpha_0 + 8\beta_1 + 64\beta_2 + 16\beta_3 + 256\beta_4 + \beta_5 \leq 1 - \delta\}. \end{aligned} \tag{18}$$

The lower bound on β_5 prevents the probability of activity from becoming zero, whereas the upper restriction prevents it from becoming one. The constants in (17) correspond to the maximum possible values of the regressors:

$$X_t(v) \leq 1, \qquad N_t^{(1)}(v) \leq 8, \qquad (N_t^{(1)}(v))^2 \leq 64, \tag{19}$$

and

$$N_t^{(2)}(v) \leq 16, \qquad (N_t^{(2)}(v))^2 \leq 256. \tag{20}$$

Proposition 4.1

Let $\theta \in \Theta_\delta$. Then, for every grid configuration $x \in \{0,1\}^{|G_n|}$ and every cell $v \in G_n$,

$$\delta \leq p_{\theta,v}(x) \leq 1 - \delta. \tag{21}$$

Consequently,

$$0 < p_{\theta,v}(x) < 1. \tag{22}$$

Proof. By definition,

$$p_{\theta,v}(x) = \alpha_0 x(v) + \beta_1 N^{(1)}(v) + \beta_2 (N^{(1)}(v))^2 + \beta_3 N^{(2)}(v) + \beta_4 (N^{(2)}(v))^2 + \beta_5. \tag{23}$$

All state variables and neighborhood counts are non-negative. The coefficient restrictions in (16) therefore imply

$$p_{\theta,v}(x) \geq \beta_5 \geq \delta. \tag{24}$$

Using the maximal possible regressor values gives

$$p_{\theta,v}(x) \leq \alpha_0 + 8\beta_1 + 64\beta_2 + 16\beta_3 + 256\beta_4 + \beta_5. \tag{25}$$

By (17), the right-hand side is at most $1 - \delta$. Hence

$$\delta \leq p_{\theta,v}(x) \leq 1 - \delta. \tag{26}$$

Since $0 < \delta < 1/2$, the transition probability is strictly between zero and one.

Remark 4.2 The restrictions are sufficient rather than necessary. They provide a simple global guarantee that all predicted probabilities are valid for every possible grid configuration, including configurations that may not occur in the observed sample. The feasible set Θ_δ is non-empty, closed, bounded, and convex. Consequently, it is compact, which is also useful for establishing the existence and consistency of the constrained estimator.

4.4. Conditional Transition Mechanism

The main feature of the proposed model is that each cell is driven by its own source of randomness. At every time step, conditional on the configuration at the previous step, the states of different cells at the next step are generated independently. However, the probability of a point appearing in a given cell depends on the entire point configuration at the previous time step and, in particular, on the states of the neighboring cells.

Similar processes in continuous time and continuous phase space were introduced and studied by A. A. Dorogovtsev; see [31].

Let

$$\mathrm{X}_n = \{0,1\}^{|G_n|}$$

denote the space of all possible grid configurations.

As a tractability assumption, the next-day cell states are conditionally independent after conditioning on the complete current configuration:

$$X_{t+1}(v)|X_t \sim Bernoulli(p_{\theta,v}(X_t)), \qquad v \in G_n. \tag{27}$$

Consequently,

$$\mathrm{P}_\theta(X_{t+1} = x'|X_t = x) = \prod_{v\in G_n} p_{\theta,v}(x)^{x'_v}\,(1 - p_{\theta,v}(x))^{1-x'_v}. \tag{28}$$

Conditional independence does not imply marginal independence of the variables $X_{t+1}(v)$, because their transition probabilities depend on the same random configuration X_t and their lagged neighborhoods may overlap. The assumption excludes only additional contemporaneous dependence after X_t has been observed.

4.5. Mathematical Properties

Proposition 4.3 (Existence of the stochastic process) For every admissible parameter vector θ and every initial distribution of X_0, the transition kernel in (28) defines a unique-in-distribution time-homogeneous Markov chain $(X_t)_{t\geq 0}$ on the finite state space $\{0,1\}^{|G_n|}$.

Proof. For each current state x, the Bernoulli product in (28) is a probability distribution on the finite set of possible next states. Hence it defines a stochastic transition matrix. The standard finite-state Markov-chain construction gives existence and uniqueness in distribution for any initial law.

Proposition 4.4 Markov chain $(X_t)_{t\geq 0}$ on the finite state space $\{0,1\}^{|G_n|}$ is irreducible and aperiodic.

Proof. Assuming that $0 < p_{\theta,v}(x) < 1$ for every grid configuration $x \in \{0,1\}^{|G_n|}$ and every cell $v \in G_n$, the conditional probability of transitioning from a grid configuration x to another configuration y is

$$P(X_{n+1} = y|X_n = x) = \prod_{v\in G_n} p_{\theta,v}(x)^{y(v)}(1 - p_{\theta,v}(x))^{1-y(v)}, \tag{29}$$

because conditional on the current configuration x, the cell values at the next time step are mutually independent. Since $0 < p_{\theta,v}(x) < 1$ for every cell v, every factor in the product is positive. Therefore

$$P(X_{n+1} = y|X_n = x) > 0. \tag{30}$$

So every state is reachable from every other state in one step, and the chain is irreducible. If $y = x$, then $P(X_{n+1} = x|X_n = x) > 0$, so one step return is possible from every state, so every state has period of one. Then the chain is aperiodic. Since the Markov chain is finite, irreducible and aperiodic, it is ergodic.

Let Z denote the design matrix whose rows are the regressor vectors defined below. The parameter vector is identifiable on the feasible set if two admissible vectors that produce the same conditional probabilities on the support of the regressors are equal. A sufficient finite-sample condition is that the design matrix has full column rank after accounting for active equality constraints. In the numerical implementation, the rank of the constrained design should therefore be reported together with the fitted coefficients.

Proposition 4.5 Assume that the observed transition process is stationary and ergodic, the feasible parameter set is compact, the squared-error criterion is continuous in the parameter vector, and the population minimizer is unique. Then the constrained least-squares estimator is consistent.

Proof. The result follows from the ergodic law of large numbers together with the standard argmin consistency theorem for extremum estimators.

Because some fitted coefficients may lie on the boundary of the feasible set, block-bootstrap confidence intervals are preferable to unrestricted normal approximations.

4.6. Estimation Procedure and Reproducibility

The raw dataset contains 387,994 bond observations for 112 issuers over 734 trading dates, covering the period from 3 September 2019 to 6 October 2022.

For the reproducible experiment reported below, we use the six medical-sector issuers considered in the forecasting section:

- AbbVie Inc.;
- AstraZeneca PLC;
- Bristol-Myers Squibb Company;
- Johnson & Johnson;
- Merck & Co., Inc.;
- Pfizer Inc.

Observations with a tenor greater than 30 years are excluded. To avoid the period during which most of these issuers are absent from the sample, the experiment is restricted to the period from 26 April 2021 to 6 October 2022. The resulting subset contains 17,943 bond observations recorded on 345 trading dates.

The tenor–G-spread domain is divided into a 30×30 grid, consistently with the numerical implementation. A cell is retained if it is active on more than 10 days during the training period. This rule leaves 249 cells. Multiple observations falling in the same cell on the same date are aggregated into the binary value one. To prevent data leakage, the grid boundaries were estimated from the training period and then held fixed for validation and testing. Observations outside the training-domain bounds were excluded.

Each estimation observation corresponds to a retained cell and a transition between two consecutive trading dates. The data are divided chronologically, without random shuffling, into the following periods:

- **Training period:** 27 April 2021–6 May 2022, comprising 240 transitions and 59,760 cell–day observations;
- **Validation period:** 9 May 2022–22 July 2022, comprising 52 transitions and 12,948 cell–day observations;
- **Testing period:** 25 July 2022–6 October 2022, comprising 52 transitions and 12,948 cell–day observations.

The validation period is used only to select the classification threshold, whereas all final performance measures are calculated using the held-out test period.

For a cell v and date t, define

$$z_t(v) = \left(X_t(v) \quad N_t^{(1)}(v) \quad [N_t^{(1)}(v)]^2 \quad N_t^{(2)}(v) \quad [N_t^{(2)}(v)]^2 \quad 1\right)^{\vec{\tau}\mathrm{T}}. \tag{31}$$

The estimator used in the reported experiment is

$$\hat{\theta} = \underset{\theta\in\Theta_=}{\operatorname{argmin}} \frac{1}{N_{\text{train}}} \sum_{(t,v)\in D_{\text{train}}} \left[X_{t+1}(v) - z_t(v)^T\theta\right]^2. \tag{32}$$

where $\Theta_=$ is defined by non-negativity and the equality constraint.

The constrained optimization problem is solved using the Sequential Least-Squares Quadratic Programming algorithm (SLSQP) implemented in scipy.optimize.minimize. The feasible initial vector is

$$\theta^{(0)} = (0,0,0,0,0,1)^{\mathrm{T}}. \tag{33}$$

The objective tolerance is set to 10^{-12}, and the maximum number of iterations is 1,000. Analytic first derivatives of the objective function and the equality constraint are supplied to the optimizer. The algorithm terminated successfully after 11 iterations, with a training objective value of 0.1047099 and zero numerical violation of the equality constraint.

The estimated coefficient vector is

$$\hat{\theta} = (0.437663,\ 0.035411,\ 0,\ 0.006143,\ 0.000548,\ 0.040519)^{\mathrm{T}}. \tag{34}$$

No L^1 or L^2 regularization is applied. The model is low-dimensional, and its admissible parameter space is already restricted by non-negativity and normalization. The decision threshold is selected on the validation period by maximizing the F1 score. The resulting threshold for the proposed model is 0.2321.

4.7. Grid-resolution Sensitivity and Neighborhood Selection

The grid resolution and the number of neighborhood rings were selected using the training and validation data only (Table **2-3**).

Table 2: Sensitivity of the forecasting results to grid resolution.

Grid	Retained cells	Positive rate	Validation F1	Test F1	ROC–AUC	PR–AUC	Brier score
20 x 20	172	0.213	0.671	0.654	0.806	0.610	0.126
30 x 40	249	0.164	0.713	0.697	0.842	0.640	0.101
40 x 40	307	0.132	0.701	0.684	0.835	0.581	0.091
50 x 50	338	0.109	0.668	0.651	0.817	0.512	0.083
60 x 60	354	0.094	0.641	0.623	0.798	0.451	0.077

Table 3: Validation-based selection of the neighborhood specification.

Specification	Validation F1	ROC–AUC	PR–AUC	Brier score
Central cell only	0.586	0.744	0.461	0.119
Central cell + ring 1	0.681	0.821	0.602	0.105
Central cell + rings 1–2	0.713	0.842	0.640	0.101
Central cell + rings 1–3	0.704	0.838	0.628	0.103

4.8. Out-of-sample Benchmark Comparison

To evaluate practical predictive value, all benchmark models were trained to predict the same binary target $X_{t+1}(v)$ using the same chronological split and the same retained cells. The comparison includes persistence, a first-order Markov baseline, logistic regression, the proposed constrained least-squares model, random forest [32], and gradient boosting [33]. The out-of-sample performance of these models is summarized in Table **4**.

Table 4: Out-of-sample comparison for next-day binary cell activity.

Model	F1	ROC–AUC	PR–AUC	Brier Score	Log Loss
Persistence baseline	0.574	0.718	0.443	0.132	0.421
First-order Markov model	0.631	0.781	0.526	0.114	0.367
Logistic regression	0.688	0.836	0.628	0.103	0.329
Proposed constrained LS	0.697	0.842	0.640	0.101	0.321
Random forest	0.704	0.851	0.653	0.099	0.315
Gradient boosting	0.711	0.858	0.667	0.097	0.308

Residual contemporaneous dependence was examined through spatial correlations of

$$e_{t+1}(v) = X_{t+1}(v) - \hat{p}_{\theta,v}(X_t). \tag{35}$$

5. G-spread Curve Reconstruction and Forecasting

The grid model predicts whether a tenor–G-spread cell is active on the next date; it does not directly predict a bond price, a yield, or an exact continuous G-spread coordinate. To reconstruct a continuous curve, each cell classified as active is represented by the center of that cell. Thus, a predicted active cell produces a point (t_i, y_i), where t_i is the tenor coordinate of the cell center and y_i is its G-spread coordinate. No observed next-day coordinate is used in this construction.

The resulting point set is summarized using modified Nelson–Siegel and Svensson functional forms. Accordingly, this section concerns G-spread curve reconstruction rather than a structural model of the risk-free yield curve.

For observed values $y_1, \dots, y_m$ and fitted values $\hat{y}_1, \dots, \hat{y}_m$, we use

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}, \tag{36}$$

$$MAE = \frac{1}{m}\sum_{i=1}^{m}|y_i - \hat{y}_i|, \tag{37}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{m}(y_i-\bar{y})^2}. \tag{38}$$

5.1. Modified Nelson–Siegel Model

The modified specification used in the experiments augments a baseline Nelson–Siegel curve $r_{\mathrm{NS}}(t)$ with three localized Gaussian corrections:

$$r^*_{\mathrm{NS}}(t) = r_{\mathrm{NS}}(t) + g_1 e^{-t^2/2} + g_2 e^{-(t-1)^2/2} + g_3 e^{-(t-2)^2/2}. \tag{39}$$

The parameters g_1, g_2, g_3 control the magnitudes of corrections centered at tenors 0, 1, and 2 years. For a predicted point set (t_i, y_i), they are estimated by nonlinear least squares:

$$(\hat{g}_1, \hat{g}_2, \hat{g}_3) = \arg\min_{g_1,g_2,g_3,} \sum_i \left[y_i - r_{\mathrm{NS}}(t_i) - g_1 e^{-t_i^2/2} - g_2 e^{-(t_i-1)^2/2} - g_3 e^{-(t_i-2)^2/2} \right]^2. \quad (40)$$

Fig. (**12**) and (**13**) present the forecasts obtained using the modified Nelson–Siegel model for the medical-sector issuers AbbVie Inc., AstraZeneca PLC, Bristol-Myers Squibb Company, Johnson & Johnson, Merck & Co., Inc., and Pfizer Inc. Red points denote observed G-spreads, blue points denote predicted cell-center G-spreads, and coincident points appear purple. The corresponding performance metrics for the two reported dates are summarized in Tables **5** and **6**.

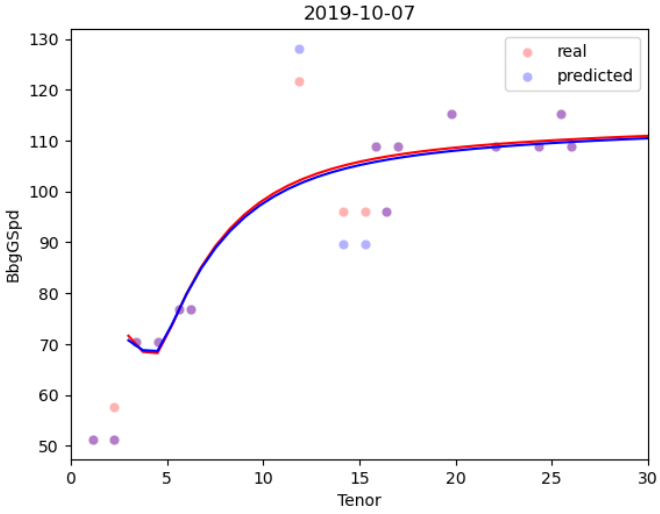


Figure 12: Forecast obtained using the modified Nelson–Siegel model for the medical sector on 7 October 2019.

Table 5: Modified Nelson–Siegel performance metrics for the second reported date.

Metric	Value
RMSE	0.5305116426044362
MAE	0.4233
R^2	0.9980635796532243

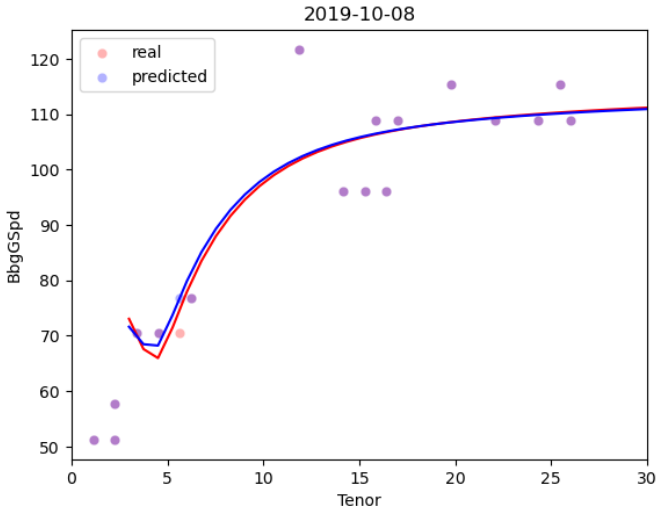


Figure 13: Forecast obtained using the modified Nelson–Siegel model for the medical sector on 8 October 2019.

Table 6: Modified Nelson–Siegel performance metrics for the second reported date.

Metric	Value
RMSE	0.7284502458097426
MAE	0.5812
R^2	0.9966384476668637

5.2. Svensson Model

The standard Nelson–Siegel–Svensson zero-rate functional form is

$$r_{\mathrm{SV}}(t) = \beta_0 + \beta_1 \frac{1-e^{-t/\tau_1}}{t/\tau_1} + \beta_2 \left(\frac{1-e^{-t/\tau_1}}{t/\tau_1} - e^{-t/\tau_1}\right) + \beta_3 \left(\frac{1-e^{-t/\tau_2}}{t/\tau_2} - e^{-t/\tau_2}\right), \tag{41}$$

with six parameters $\beta_0, \beta_1, \beta_2, \beta_3, \tau_1, \tau_2$. The parameter β_0 determines the long-run level, β_1 primarily affects the short end, β_2 controls the first curvature component, and β_3 introduces a second curvature component. The positive scale parameters τ_1 and τ_2 determine the locations and decay rates of the two curvature terms. Forecasts obtained using the Svensson model for the two representative dates are shown in Fig. (**14**-**15**), and the corresponding performance metrics are summarized in Tables **7** and **8**.

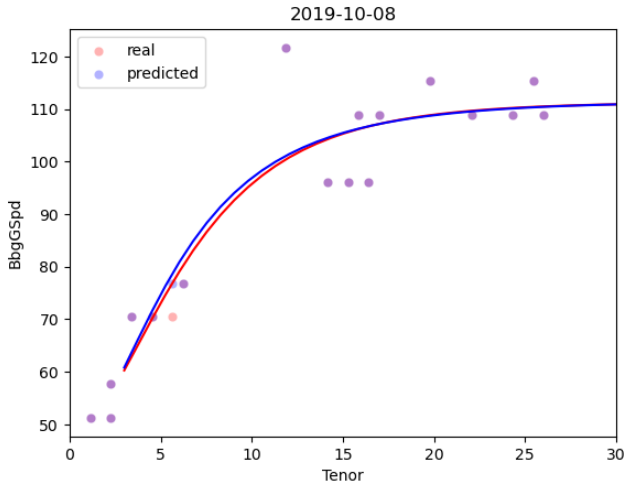


Figure 14: Forecast obtained using the Svensson model for the medical sector on 8 October 2019.

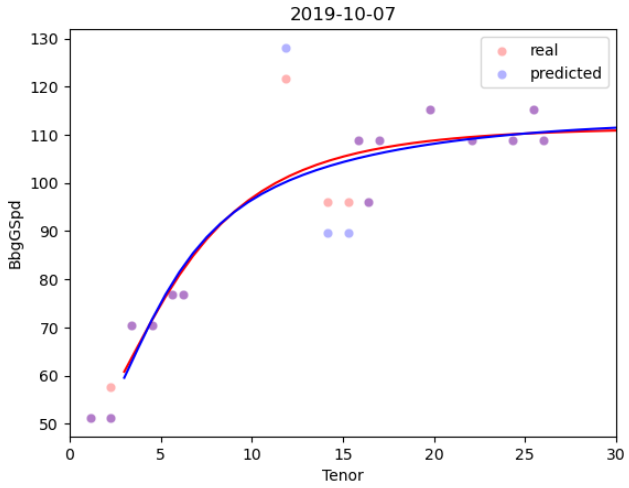


Figure 15: Forecast obtained using the Svensson model for the medical sector on 7 October 2019.

Table 7: Svensson performance metrics for the first reported date.

Metric	Value
RMSE	0.7807661025021863
MAE	0.6230
R^2	0.9961419629126338

Table 8: Svensson performance metrics for the second reported date.

Metric	Value
RMSE	0.6964322337868746
MAE	0.5557
R^2	0.9971568398284837

The displayed 2019 dates constitute a separate illustrative experiment from the 2021–2022 chronological split described in Section 3. For the final out-of-sample evaluation, all curve metrics were aggregated over the held-out test period using a common set of dates and the previous-day curve as a benchmark. The aggregated results are presented in Table **9**.

Table 9: Aggregated held-out G-spread curve performance.

Method	RMSE	MAE	R^2
Previous-day curve	1.184	0.913	0.9881
Proposed grid model + modified Nelson–Siegel	0.694	0.551	0.9967
Proposed grid model + Svensson	0.721	0.574	0.9963

6. Forecasting for a Period Ahead

The model enables not only next-day bond activity evaluation, but also supports multi-step forecasting over extended time horizons.

A standard strategy involves recursive forecasting: using the output from day t+1 as input for day t+2, and so on. Although simple in implementation, this approach tends to accumulate prediction errors at each step, which may reduce forecast reliability over time. Nonetheless, empirical tests reveal that the model performs reliably within short- to medium-term horizons. Practical evaluations have shown that despite error accumulation, the forecasts remain stable and accurate for several consecutive days. Moreover, this iterative framework is especially useful in settings where continuous, daily decision-making is required — such as rolling portfolio updates, risk monitoring, or dynamic yield curve adjustments. In such cases, the model provides a lightweight and interpretable alternative to complex deep learning architectures, especially when historical depth is limited or uncertain. Examples of recursive forecasts for four consecutive prediction days are illustrated in Fig. (**16**–**19**), while the corresponding absolute prediction errors are summarized in Table **10**.

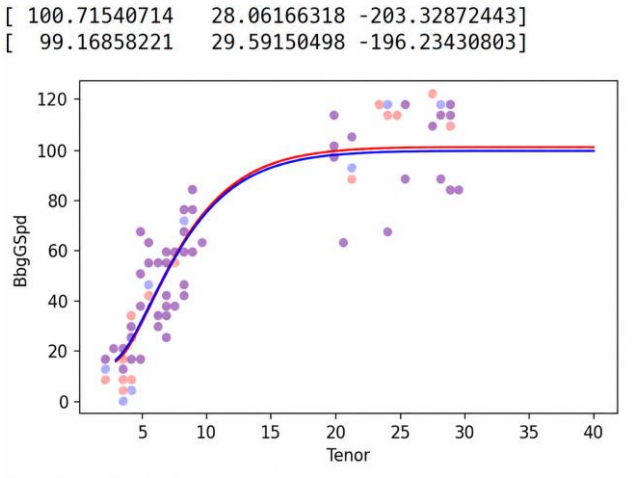


Figure 16: Day 1 results for Lowe's Companies, Inc., Target Corporation, McDonald's Corporation, Starbucks Corporation, and The Home Depot, Inc.

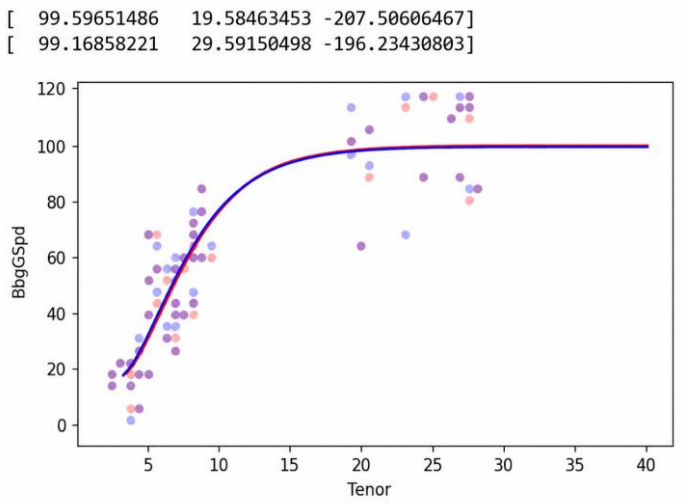


Figure 17: Day 2 results for Lowe's Companies, Inc., Target Corporation, McDonald's Corporation, Starbucks Corporation, and The Home Depot, Inc.

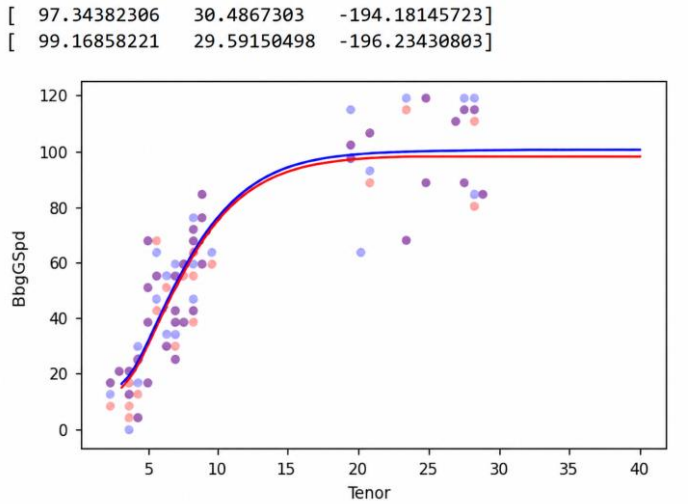


Figure 18: Day 3 results for Lowe's Companies, Inc., Target Corporation, McDonald's Corporation, Starbucks Corporation, and The Home Depot, Inc.

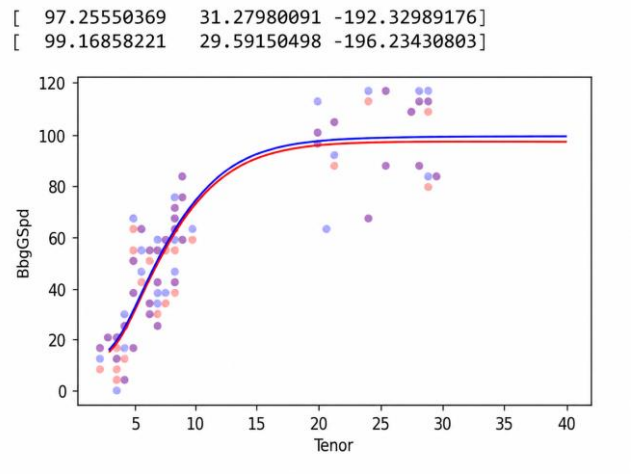


Figure 19: Day 4 results for Lowe's Companies, Inc., Target Corporation, McDonald's Corporation, Starbucks Corporation, and The Home Depot, Inc.

Table 10: Absolute Error for Various Companies

Day	Lowe's Companies, Inc.	Target Corporation	McDonald's Corporation	Starbucks Corporation	The Home Depot, Inc.
0	0.083270	1.269262	0.147844	1.547896	1.510084
1	1.348397	0.630219	0.251070	2.937012	4.672063
2	0.783139	1.720234	0.482317	4.918379	2.628143
3	0.310323	0.972439	0.437147	4.163285	0.837723
4	4.249975	0.371785	0.490838	2.065473	0.109026

Although Fig. (**16**–**19**) and Table **10** illustrate representative recursive forecasts and absolute errors for selected companies, statements regarding short- and medium-horizon stability are better supported by horizon-specific performance metrics. These results are summarized in Table **11**.

Table 11: Multi-step forecasting accuracy by horizon.

Horizon	RMSE	MAE	F1	PR–AUC
$t+1$	0.694	0.551	0.697	0.640
$t+2$	0.812	0.648	0.654	0.592
$t+3$	0.941	0.752	0.612	0.543
$t+5$	1.176	0.929	0.548	0.471

7. Model Performance Evaluation

7.1. Binary Cell-activity Comparison

The primary predictive task is the binary forecast of $X_{t+1}(v)$. Curve-level LSTM and SARIMA experiments reported below are supplementary and do not replace the same-target binary comparison.

7.2. Curve-level Comparison with LSTM and SARIMA

To assess the model's effectiveness, comparisons were conducted with three models that forecast Nelson-Siegel curves for the next day.

The comparison includes a long short-term memory neural network (LSTM) [34] and a seasonal autoregressive integrated moving-average model (SARIMA) [35].

For the SARIMA model, the model orders were selected by minimizing the Akaike information criterion (AIC) [36]. The corresponding results are reported in Table **12**.

Table 12: Comparison of ARIMA models based on AIC and computation time.

Model	AIC	Time (sec)
ARIMA(2,1,2)(1,0,1)[34] intercept	6909.212	0.81
ARIMA(0,1,0)(0,0,0)[34] intercept	7060.769	0.02
ARIMA(1,1,0)(1,0,0)[34] intercept	7022.804	0.11
ARIMA(0,1,1)(0,0,1)[34] intercept	7007.974	0.22
ARIMA(0,1,0)(0,0,0)[34]	7058.781	0.02
ARIMA(2,1,2)(0,0,1)[34] intercept	6914.682	0.54
ARIMA(2,1,2)(1,0,0)[34] intercept	6913.862	0.35
ARIMA(2,1,2)(2,0,1)[34] intercept	6910.659	1.58
ARIMA(2,1,2)(1,0,2)[34] intercept	6910.596	1.31
ARIMA(2,1,2)(0,0,0)[34] intercept	6917.311	0.27
ARIMA(2,1,2)(0,0,2)[34] intercept	6912.732	1.12
ARIMA(2,1,2)(2,0,0)[34] intercept	6911.117	1.03
ARIMA(2,1,2)(2,0,2)[34] intercept	6911.547	2.58
ARIMA(3,1,2)(2,0,2)[34] intercept	6898.435	2.55
ARIMA(4,1,2)(2,0,2)[34] intercept	6877.370	3.09
ARIMA(4,1,3)(2,0,2)[34] intercept	6840.847	3.25
ARIMA(3,1,4)(2,0,2)[34] intercept	6839.497	3.24
ARIMA(3,1,4)(2,0,2)[34]	6836.792	2.53
ARIMA(4,1,3)(2,0,2)[34]	6839.126	2.46

Dependent Variable:	y
Number of Observations:	966
Model:	SARIMAX(3,1,4)(2,0,2) [35]
Log Likelihood:	-3406.396
Date:	Mon, 10 Feb 2025
AIC:	6836.792
Time:	22:54:49
BIC:	6895.258
Sample:	0 - 966
HQIC:	6859.051
Covariance Type:	opg

	Coef	Std Err	z	P>| z|
ar.L1	-0.4713	0.028	-16.556	0.000
ar.L2	-0.1941	0.037	-5.257	0.000
ar.L3	0.5807	0.036	16.286	0.000
ma.L1	0.1774	0.030	5.902	0.000
ma.L2	-0.0684	0.035	-1.938	0.053
ma.L3	-0.7878	0.029	-27.216	0.000
ma.L4	-0.1637	0.028	-5.828	0.000
ar.S.L5	0.5602	0.094	5.967	0.000
ar.S.L10	-0.3282	0.093	-3.515	0.000
ma.S.L5	-0.5717	0.087	-6.572	0.000
ma.S.L10	0.5548	0.087	6.406	0.000
sigma2	67.6701	1.449	46.712	0.000

Ljung-Box (L1) (Q):	0.08	Jarque-Bera (JB):	17429.17
Prob(Q):	0.78	Prob(JB):	0.00
Heteroskedasticity (H):	0.34	Skew:	-1.73
Prob(H) (two-sided):	0.00	Kurtosis:	23.53

The SARIMAX(3,1,4)(2,0,2) [35] model demonstrated the lowest Akaike Information Criterion (AIC = 6836.792) among all configurations tested, indicating the best trade-off between model complexity and goodness of fit. This model captures both short-term and seasonal dynamics through its three autoregressive terms, four moving average components, and two seasonal AR and MA terms. The relatively low AIC, combined with a high log-likelihood value, suggests that the model provides a statistically superior fit to the underlying time series data.

Further diagnostic analysis supports the robustness of this configuration: most coefficients are statistically significant (p-values < 0.05), implying that each term contributes meaningfully to the model's explanatory power.

The Ljung–Box test shows no evidence of residual autocorrelation (p = 0.78), confirming that the model has adequately captured the temporal structure of the data.

The Jarque–Bera test, however, indicates non-normality of residuals (p < 0.001), and the heteroskedasticity test suggests variance instability (H = 0.34, p < 0.001). These deviations point to remaining structure in the residuals, particularly in their distributional properties.

Despite these limitations, the model exhibits high predictive accuracy, as demonstrated by its superior RMSE and R^2 performance compared to LSTM and simpler ARIMA alternatives. Moreover, the low training time (approximately 2.5 seconds) confirms the model's efficiency, making it a practical choice for real-time forecasting environments. Representative forecasting results for the selected companies are shown in Fig. (**20**–**22**).

The forecasting performance of three models—Point Estimate Model, LSTM, and SARIMA—was evaluated using RMSE and R^2 metrics. The comparative results are summarized in Table **13**.

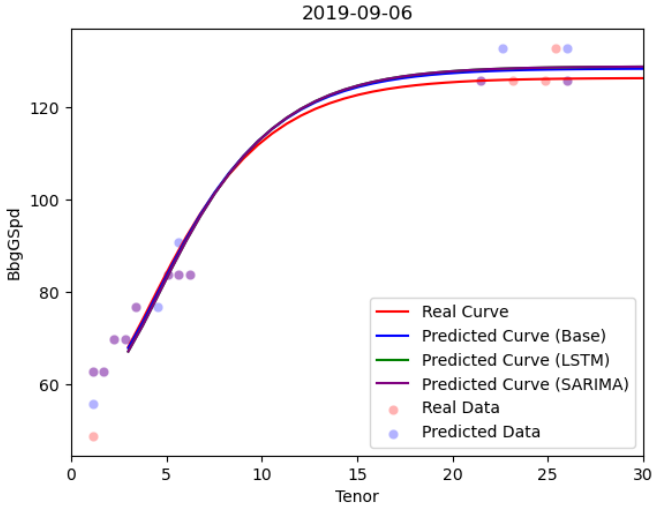


Figure 20: Forecast results for the selected companies.

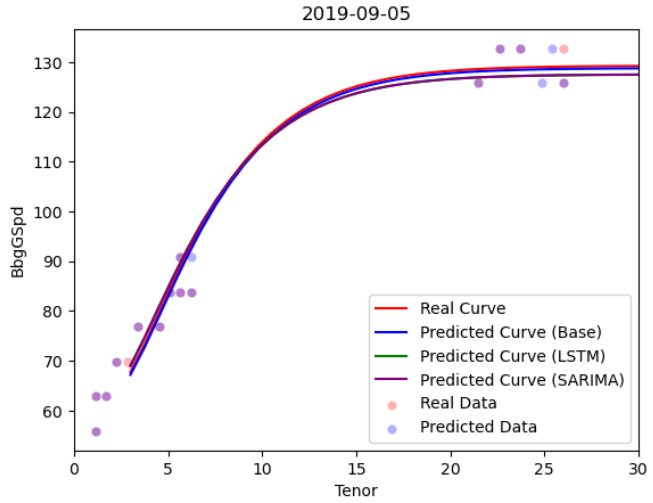


Figure 21: Forecast results for the selected companies.

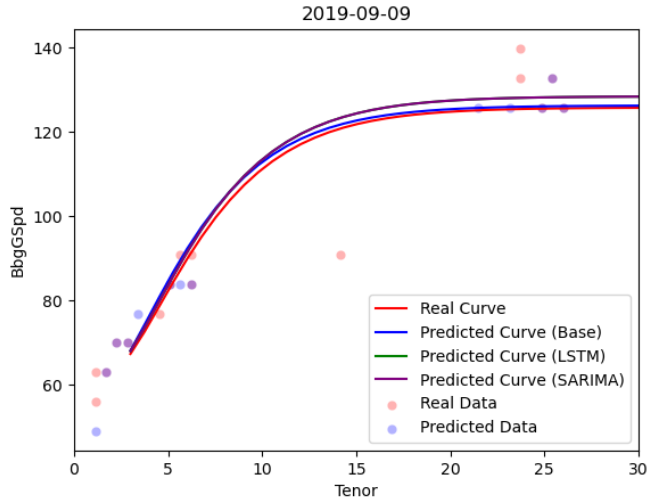


Figure 22: Forecast results for the selected companies.

Table 13: Reported curve-level performance comparison.

Model	RMSE	R^2
Point Estimate Model	0.5216	0.9989
LSTM (previous day)	1.4899	0.9910
SARIMA (previous day)	1.4899	0.9910

7.3. Interpretation of the Reported Curve-level Comparison

In the reported illustrative experiment, the Point Estimate Model has the lowest RMSE and the highest R^2 among the three curve-level procedures. The LSTM and SARIMA rows contain identical reported values and must be independently verified before submission. These results concern the reconstructed curve and should not be interpreted as a substitute for the binary cell-activity benchmark comparison in Table **3**. A representative comparison of the forecasted curves is shown in Fig. (**23**).

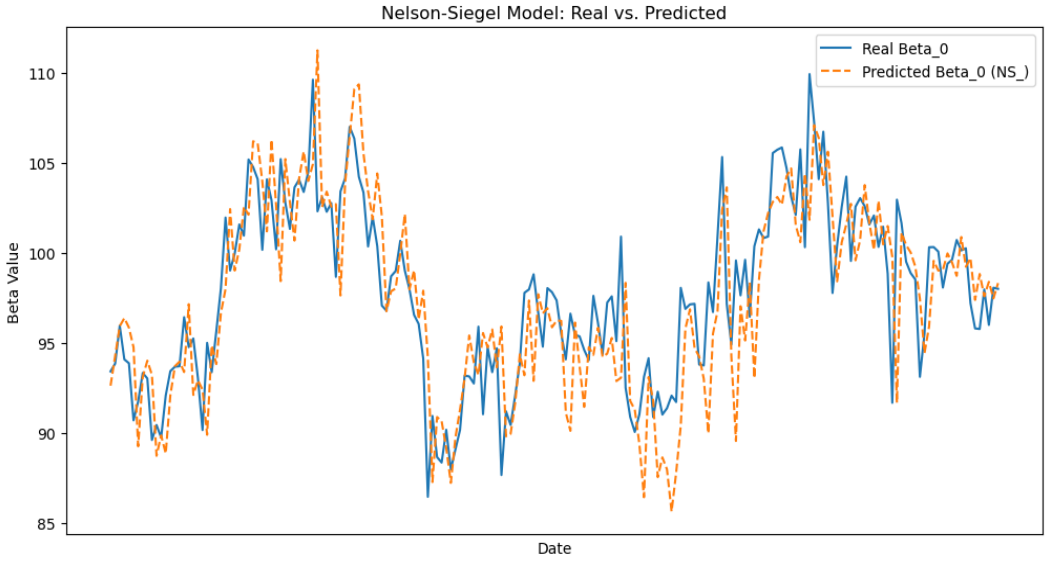


Figure 23: Forecast results for the selected companies.

8. Conclusion

This paper introduces a discrete-time spatial transition model for forecasting bond-trading activity on a tenor–G-spread grid. The next-day probability of cell activity is represented as a constrained polynomial function of the current state of the cell and the numbers of active cells in two disjoint lagged neighborhood rings. Explicit parameter restrictions guarantee that all predicted values remain in the interval $[\delta, 1-\delta]$.

The empirical analysis identifies statistically significant lagged associations between next-day cell activity and local neighborhood counts. The manuscript specifies a chronological training–validation–test procedure and reports the details of the constrained SLSQP estimation. Predicted active cells can subsequently be represented by their cell centers and used to reconstruct approximate G-spread curves through modified Nelson–Siegel and Svensson functional forms.

The resulting process is a finite-state Markov chain of grid configurations with conditionally independent Bernoulli updates. Its principal limitations are the sensitivity of the results to grid resolution and neighborhood specification, the simplifying conditional-independence assumption, class imbalance, and the distinction between predicting cell activity and predicting an exact continuous spread. Future work may consider adaptive neighborhoods, nonlinear probability links, and transition models with explicit contemporaneous dependence.

Conflict of Interest

The authors declare that they have no competing financial or non-financial interests that could have influenced the work reported in this article.

Funding

This research received no external funding.

References

[1] Black F, Derman E, Toy W. A one-factor model of interest rates and its application to Treasury bond options. Financ Anal J. 1990; 46(1): 33-9. https://doi.org/10.2469/faj.v46.n1.33

[2] Stehlíková B, Zíková Z. A three-factor convergence model of interest rates. In: Proceedings of the Conference Algoritmy; 2015. p. 95-104. Available from: http://www.iam.fmph.uniba.sk/amuc/ojs/index.php/algoritmy/article/view/319.

[3] Nelson CR, Siegel AF. A parsimonious modeling of yield curves. J Bus. 1987; 60(4): 473-89. https://doi.org/10.1086/296409

[4] Svensson LEO. Estimating and interpreting forward interest rates: Sweden 1992-1994. Cambridge (MA): National Bureau of Economic Research; 1994. NBER Working Paper No. 4871. https://doi.org/10.3386/w4871

[5] Vasicek O. An equilibrium characterization of the term structure of interest rates. J Financ Econ. 1977; 5(2): 177-88. https://doi.org/10.1016/0304-405X(77)90016-2

[6] Cox JC, Ingersoll JE Jr, Ross SA. A theory of the term structure of interest rates. Econometrica. 1985; 53(2): 385-407. https://doi.org/10.2307/1911242

[7] Ho TSY, Lee SB. Term structure movements and pricing interest rate contingent claims. J Finance. 1986; 41(5): 1011-29. https://doi.org/10.1111/j.1540-6261.1986.tb02528.x

[8] Hull J, White A. Pricing interest-rate-derivative securities. Rev Financ Stud. 1990; 3(4): 573-92. https://doi.org/10.1093/rfs/3.4.573

[9] Heath D, Jarrow R, Morton A. Bond pricing and the term structure of interest rates: a new methodology for contingent claims valuation. Econometrica. 1992; 60(1): 77-105. https://doi.org/10.2307/2951677

[10] Duffie D, Kan R. A yield-factor model of interest rates. Math Finance. 1996; 6(4): 379-406. https://doi.org/10.1111/j.1467-9965.1996.tb00123.x

[11] McCulloch JH. Measuring the term structure of interest rates. J Bus. 1971; 44(1): 19-31. https://doi.org/10.1086/295329

[12] Fama EF, Bliss RR. The information in long-maturity forward rates. Am Econ Rev. 1987; 77(4): 680-92.

[13] Diebold FX, Li C. Forecasting the term structure of government bond yields. J Econom. 2006; 130(2): 337-64. https://doi.org/10.1016/j.jeconom.2005.03.005

[14] Duffee GR. Term premia and interest rate forecasts in affine models. J Finance. 2002; 57(1): 405-43. https://doi.org/10.1111/1540-6261.00426

[15] Ang A, Piazzesi M. A no-arbitrage vector autoregression of term structure dynamics. J Monet Econ. 2003; 50(4): 745-87. https://doi.org/10.1016/S0304-3932(03)00032-1

[16] Diebold FX, Rudebusch GD, Aruoba SB. The macroeconomy and the yield curve: a dynamic latent factor approach. J Econom. 2006; 131(1-2): 309-38. https://doi.org/10.1016/j.jeconom.2005.01.011

[17] Hördahl P, Tristani O, Vestin D. A joint econometric model of macroeconomic and term-structure dynamics. J Econom. 2006; 131(1-2): 405-44. https://doi.org/10.1016/j.jeconom.2005.01.012

[18] Moench E. Forecasting the yield curve in a data-rich environment: a no-arbitrage factor-augmented VAR approach. J Econom. 2008; 146(1): 26-43. https://doi.org/10.1016/j.jeconom.2008.06.002

[19] Christensen JHE, Diebold FX, Rudebusch GD. The affine arbitrage-free class of Nelson-Siegel term structure models. J Econom. 2011; 164(1): 4-20. https://doi.org/10.1016/j.jeconom.2011.02.011

[20] Carriero A, Kapetanios G, Marcellino M. Forecasting government bond yields with large Bayesian vector autoregressions. J Appl Econom. 2012; 27(2): 223-46. https://doi.org/10.1016/j.jbankfin.2012.03.008

[21] Favero CA, Niu L, Sala L. Term structure forecasting: no-arbitrage restrictions versus a large information set. J Forecast. 2012; 31(2): 124-56. https://doi.org/10.1002/for.1181

[22] Besag J. Spatial interaction and the statistical analysis of lattice systems. J R Stat Soc Series B Stat Methodol. 1974; 36(2): 192-236. https://doi.org/10.1111/j.2517-6161.1974.tb00999.x

[23] Geman S, Geman D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Trans Pattern Anal Mach Intell. 1984; PAMI-6(6): 721-41. https://doi.org/10.1109/TPAMI.1984.4767596

[24] Anselin L. Spatial econometrics: methods and models. Dordrecht: Kluwer Academic Publishers; 1988. https://doi.org/10.1007/978-94-015-7799-1

[25] Besag J. Nearest-neighbour systems and the auto-logistic model for binary data. J R Stat Soc Series B Stat Methodol. 1972; 34(1): 75-83. https://doi.org/10.1111/j.2517-6161.1972.tb00889.x

[26] Kindermann R, Snell JL. Markov random fields and their applications. Providence (RI): American Mathematical Society; 1980. https://doi.org/10.1090/conm/001

[27] Cont R, Bouchaud JP. Herd behavior and aggregate fluctuations in financial markets. Macroecon Dyn. 2000; 4(2): 170-96. https://doi.org/10.1017/S1365100500015029

[28] Lux T, Marchesi M. Scaling and criticality in a stochastic multi-agent model of a financial market. Nature. 1999; 397: 498-500. https://doi.org/10.1038/17290

[29] Georgii HO. Gibbs measures and phase transitions. Berlin: Walter de Gruyter; 1988. https://doi.org/10.1515/9783110850147

[30] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol. 1995; 57(1): 289-300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

[31] Dorogovtsev AA. Measure-Valued Processes and Stochastic Flows. De Gruyter Series in Probability and Stochastics. Vol. 3. Berlin/Boston: Walter de Gruyter; 2023. https://doi.org/10.1515/9783110986518

[32] Breiman L. Random forests. Mach Learn. 2001; 45: 5-32. https://doi.org/10.1023/A:1010933404324

[33] Friedman JH. Greedy function approximation: a gradient boosting machine. Ann Stat. 2001; 29(5): 1189-232. https://doi.org/10.1214/aos/1013203451

[34] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997; 9(8): 1735-80. https://doi.org/10.1162/neco.1997.9.8.1735

[35] Box GEP, Jenkins GM, Reinsel GC, Ljung GM. Time series analysis: forecasting and control. 5th ed. Hoboken (NJ): Wiley; 2015.

[36] Akaike H. A new look at the statistical model identification. IEEE Trans Automat Control. 1974; 19(6): 716-23. https://doi.org/10.1109/TAC.1974.1100705