Exploring Various Activation Functions in Kolmogorov-Arnold Networks for Time Series Prediction

Anushka Pachaury
Columbia University
ap4617@columbia.edu

***Abstract*—Kolmogorov-Arnold networks (KANs) have emerged as a promising architecture for complex function approximation, particularly in time series prediction. Traditionally, KANs rely on B-spline activations, which, while flexible, struggle to capture sharp fluctuations in volatile financial data. This study explores the performance of alternative activation functions such as Jacobi, Chebyshev, and Bessel polynomials in KANs. Using the S&P 500 Adjusted Close price dataset, we assess the predictive performance of these activation functions. Results show that Jacobi activations outperform B-splines in minimizing the test MSE, while the Bessel and Chebyshev activations offer competitive results. By incorporating recurrent layers into the KAN architecture, we further extend the models to temporal KANs (tKANs), demonstrating improved performance in capturing sequential patterns. This research highlights the potential of alternative activations to enhance KANs' adaptability and predictive accuracy for complex, dynamic time-series data.**

***Index Terms*—Kolmogorov-Arnold Networks, Activation Functions, Jacobi, Chebyshev, Bessel, B-Splines, Temporal KANs, Time-Series Prediction, Stock Market Forecasting.**

I. Introduction

Time-series forecasting is a critical challenge in fields such as finance, healthcare, and climate science, where accurate predictions are essential for decision making [1], [2]. The dynamic, nonlinear, and chaotic nature of stock market data, in particular, poses unique difficulties for predictive modeling. Traditional deep learning models, such as long-short-term memory networks (LSTMs) and transformers, have achieved notable success in capturing complex temporal dependencies but often come with high computational costs and limited interpretability [3].

Kolmogorov-Arnold networks (KANs) offer a promising alternative by leveraging the Kolmogorov-Arnold theorem, which asserts that any multivariate function can be expressed as a superposition of univariate functions [4], [5]. This architecture provides an interpretable and computationally efficient framework for function approximation, making it well-suited for financial time series forecasting. KANs traditionally rely on B-spline activation functions due to their smoothness and flexibility in approximating complex patterns [6]. However, B-splines may struggle to capture abrupt changes and irregularities, such as sharp fluctuations in financial data, limiting their predictive accuracy.

This study explores the potential of alternative activation functions, including Jacobi, Chebyshev, and Bessel polynomials, to enhance the adaptability and accuracy of KANs for time-series prediction [7], [8], [9]. These polynomial-based activations offer unique advantages: Jacobi polynomials provide interval-specific control, Chebyshev polynomials minimize approximation errors, and Bessel functions effectively model oscillatory behaviors. By integrating these activations into the KANs, we hypothesize that the network can better handle the complexities of volatile financial data.

Furthermore, we introduce Temporal Kolmogorov-Arnold Networks (tKANs), an extension of KANs that incorporates recurrent layers, such as LSTMs, to capture long-term dependencies in sequential data [3], [10]. Using the S&P 500 Adjusted Close dataset, this study evaluates the performance of these alternative activation functions and temporal extensions. Our findings demonstrate that Jacobi and Chebyshev activations significantly outperform traditional B-splines, with tKANs further improving predictive performance by effectively modeling sequential patterns. This research contributes to advancing the state-of-the-art in interpretable, efficient, and accurate time series forecasting models.

II. Related Work

A. Kolmogorov-Arnold Networks (KANs)

Kolmogorov-Arnold Networks (KANs) are inspired by the Kolmogorov-Arnold theorem, which asserts that any continuous multivariate function $f(x_1, x_2, \ldots, x_n)$ can be expressed as a superposition of univariate continuous functions:[5]

$$f(x_1, x_2, \ldots, x_n) = \sum_{q=1}^{2n+1} \phi_q \left(\sum_{p=1}^{n} \psi_{qp}(x_p) \right),$$

where ϕ_q and ψ_{qp} are univariate continuous functions. This decomposition transforms a high-dimensional problem into the composition of lower-dimensional functions, reducing the complexity of function approximation.

KANs operationalize this theorem by learning approximations of ϕ_q and ψ_{qp} using basis functions such as B-splines. B-splines are piecewise-defined polynomials that offer smooth and flexible approximations. A B-spline of order k is defined recursively as:

$$B_{i,k}(x) = \frac{x - t_i}{t_{i+k-1} - t_i} B_{i,k-1}(x) + \frac{t_{i+k} - x}{t_{i+k} - t_{i+1}} B_{i+1,k-1}(x),$$

where t_i are the knot points, and $B_{i,1}(x)$ is a step function over the interval $[t_i, t_{i+1})$. These activation functions provide local adaptability, ensuring that the model can capture both global trends and local variations in the data.

1) Advantages of KANs:

- **Interpretable Representations**: By decomposing multivariate functions into sums of univariate components, KANs provide inherently interpretable mappings.
- **Dimensionality Reduction**: The Kolmogorov-Arnold decomposition reduces the problem's dimensionality, enabling efficient computations.
- **Flexibility with Basis Functions**: The use of customizable basis functions, such as B-splines or Chebyshev polynomials, allows the network to adapt to various data types and application domains.

2) Challenges in KANs: Despite their strengths, KANs face notable challenges:

- **Parameter Growth**: The requirement for $2n + 1$ basis functions per variable can lead to parameter explosion, particularly for high-dimensional inputs.
- **Smoothness Constraints**: While B-splines are effective for smooth approximations, they may struggle to capture sharp discontinuities or abrupt changes in the data, motivating the exploration of alternative activation functions.

3) Extensions and Applications: Recent advancements in KANs focus on addressing these limitations. Extensions such as temporal KANs (tKANs) incorporate recurrent architectures to capture sequential dependencies, broadening their applicability to time-series forecasting and other dynamic domains. Furthermore, replacing B-splines with alternative activations, such as Jacobi or Chebyshev polynomials, has shown promise in improving performance on non-linear and volatile datasets. These innovations underscore the potential of KANs as interpretable and computationally efficient tools for function approximation and forecasting.

B. Alternative Activation Functions

Alternative activation functions offer unique advantages in extending the capabilities of Kolmogorov-Arnold Networks by overcoming the limitations of traditional B-splines. This section discusses the mathematical underpinnings and properties of Jacobi, Chebyshev, and Bessel polynomial-based activations.

1) Jacobi Polynomials: Jacobi polynomials $P_n^{(\alpha,\beta)}(x)$ are orthogonal polynomials defined over the interval $[-1, 1]$ with weight function $(1 - x)^\alpha (1 + x)^\beta$, where $\alpha, \beta > -1$. The recurrence relation for Jacobi polynomials is:

$$(n+1)P_{n+1}^{(\alpha,\beta)}(x) = (2n+\alpha+\beta+1)(2x)P_n^{(\alpha,\beta)}(x) - (n+\alpha)(n+\beta)P_{n-1}^{(\alpha,\beta)}(x).$$

Their ability to adapt across varying intervals makes them effective in capturing localized features in time-series data, particularly in regions of high variance.

2) Chebyshev Polynomials: Chebyshev polynomials $T_n(x)$ are a special case of orthogonal polynomials, minimizing the maximum error for function approximation. They are defined by the recurrence relation:[2]

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x), \quad T_0(x) = 1, \quad T_1(x) = x.$$

Chebyshev polynomials are particularly advantageous for approximating functions with rapid fluctuations, as they distribute approximation error evenly across the domain.

3) Bessel Functions: Bessel functions $J_n(x)$ arise as solutions to Bessel's differential equation:[2]

$$x^2 \frac{d^2y}{dx^2} + x\frac{dy}{dx} + (x^2 - n^2)y = 0.$$

They are often used to model oscillatory behavior in data, making them suitable for time-series applications involving periodic trends. The series expansion for $J_n(x)$ is given by:

$$J_n(x) = \sum_{m=0}^{\infty} \frac{(-1)^m}{m!\Gamma(m+n+1)} \left(\frac{x}{2}\right)^{2m+n}.$$

4) Comparison of Activation Functions: Each alternative activation function offers distinct advantages based on the underlying data characteristics:

- **Jacobi Polynomials**: Best suited for localized, high-variance data.
- **Chebyshev Polynomials**: Ideal for capturing sharp transitions and minimizing global approximation error.
- **Bessel Functions**: Effective for modeling oscillatory or periodic trends.

By incorporating these polynomial-based activation functions into KANs, this study demonstrates their potential to significantly enhance predictive accuracy and flexibility in handling complex time-series data.

C. Temporal Extensions in KANs

Temporal extensions of KANs, such as Temporal Kolmogorov-Arnold Networks (tKANs), build on the foundational Kolmogorov-Arnold decomposition while addressing the challenges of sequential data modeling. By integrating recurrent structures with the flexibility of KANs, these extensions enable dynamic time-series prediction with improved adaptability and interpretability.[3]

1) Incorporating Recurrence into KANs: Temporal KANs leverage recurrent neural network (RNN) architectures, such as Long Short-Term Memory (LSTM) units or Gated Recurrent Units (GRUs), to model dependencies across time steps. The temporal nature of sequential data requires:

- **State Propagation**: Maintaining hidden states across time steps to capture long-term dependencies.
- **Dynamic Weights**: Updating activation parameters dynamically as the network processes sequential data.

Mathematically, the integration of recurrence can be expressed as:

$$h_t = f_{KAN}(x_t, h_{t-1}),$$

where h_t represents the hidden state at time t, x_t is the input at time t, and f_{KAN} is the activation function derived from KAN.

2) Advantages of Temporal KANs:

- **Sequential Dependency Modeling**: The incorporation of recurrent layers enables tKANs to capture both short-term trends and long-term dependencies in sequential data.
- **Adaptive Functionality**: Temporal KANs dynamically adjust their activation parameters, allowing for localized adaptability across different time intervals.[1]
- **Improved Predictive Accuracy**: By combining the interpretability of KANs with the temporal capabilities of RNNs, tKANs achieve superior performance on time-series tasks.

3) Applications of tKANs: Temporal KANs have been applied across various domains, including:

- **Financial Forecasting**: Capturing non-linear dependencies and sequential trends in stock market data.
- **Climate Modeling**: Predicting dynamic weather patterns by modeling seasonal variations and long-term trends.
- **Healthcare Analytics**: Analyzing patient records to forecast outcomes based on historical data.

The temporal extension of KANs bridges the gap between static function approximation and dynamic sequence modeling, making them a powerful tool for complex time-series prediction tasks.

III. Methodology

A. Dataset and Preprocessing

The dataset and preprocessing steps for both the standard KAN and the temporal KAN (tKAN) were identical:

- **Data Preparation**: The daily adjusted closing prices of the S&P 500 index from January 2000 to December 2024 were collected using the `yfinance` library.
- **Sequence Creation**: A lookback window of 15 days was used to create sequences, where each input sequence represented 15 consecutive days, and the output corresponded to the 16th day's price.
- **Data Splitting and Normalization**: The data was split into training (70%), validation (15%), and testing (15%) sets, normalized to a range of [0, 1] using MinMaxScaler.

B. Model Architectures

1) Standard Kolmogorov-Arnold Network (KAN): The standard KAN architecture was designed to model non-linear relationships using B-spline activations and was inspired by the Kolmogorov-Arnold theorem.[4] This theorem asserts that any multivariate continuous function can be decomposed into a superposition of univariate functions and additive operations. Key details of the architecture include:

- **Structure**: The KAN consisted of:
  - An input layer designed to accept 15-dimensional time-series data.
  - Two hidden layers with 64 and 32 nodes, respectively, to model non-linear relationships in the input features.[11]
  - An output layer with a single neuron, producing scalar predictions corresponding to the adjusted closing price of the S&P 500 index.
- **Activation Functions**:
  - Each hidden layer applied learnable cubic B-spline activation functions. These splines were parameterized as a combination of basis functions, enabling local adaptability and smooth approximations.
  - The B-splines were initialized to approximate linear mappings, gradually adapting to the data's complexity during training.[9]
- **Training**: The model was trained using the Adam optimizer, with a learning rate of 0.001, and Mean Squared Error (MSE) was employed as the loss function. Training was conducted over 20 epochs with mini-batches of size 32.

2) Temporal Kolmogorov-Arnold Network (tKAN): The tKAN extended the standard KAN architecture by integrating temporal dynamics, allowing it to capture sequential dependencies inherent in time-series data. Inspired by the flexibility of Kolmogorov-Arnold representations, the tKAN architecture featured the following components:

- **Structure**:
  - **Custom TKANCell**: This recurrent layer processed sequential data by dynamically transforming input features using multiple sub-KAN configurations. Each sub-KAN was designed with unique B-spline properties, such as varying spline orders and grid sizes, enhancing the network's ability to handle diverse temporal patterns.[10]
  - **Dense Layer**: Following the recurrent processing, a fully connected dense layer produced the final predictions by combining the extracted temporal features.
- **Activation Functions**:
  - B-spline activations were extended to include alternative polynomial activations, such as Jacobi and Chebyshev polynomials, to better model oscillatory and sharp transitions in the data.
  - The grid points for splines were dynamically updated during training to accommodate evolving input distributions.
- **Recurrent Dynamics**:
  - The TKANCell incorporated recurrent connections to retain temporal states across time steps, enabling the model to capture both short-term trends and long-term dependencies.
  - Regularization techniques, such as dropout layers and layer normalization, were employed to prevent overfitting and enhance training stability.
- **Training**: Similar to the standard KAN, tKAN was trained using the Adam optimizer and MSE loss function.

Training spanned 20 epochs, with a validation split of 20% to monitor generalization.

C. Evaluation

Both models were evaluated using the following metrics:

- **Mean Squared Error (MSE)**: Quantifying prediction error.
- R^2 **Score**: Measuring explained variance to assess model accuracy.
- **Visualization**: Predictions were compared against actual prices using line plots for qualitative insights.

D. Implementation and Results

- The standard KAN demonstrated effective predictive capabilities, particularly for non-oscillatory trends, with smooth and interpretable activation mappings.
- The tKAN outperformed the standard KAN by capturing sequential patterns and long-term dependencies, as evidenced by a significant reduction in MSE and improvement in R^2 scores. The integration of temporal dynamics and adaptive activations proved crucial in enhancing predictive performance.

IV. Results and Discussion

A. Baseline Performance with B-Splines

B-splines served as the baseline activation function for this study. The results demonstrated that standard KANs with B-splines achieved a Test MSE of 0.0734 and R^2 of 0.3261. By incorporating temporal dependencies through LSTM layers, the temporal KAN (tKAN) variant achieved a Test MSE of 0.0074 and R^2 of 0.7897, reflecting a significant improvement.

The visualizations in **Figure 1** and **Figure 2** illustrate the discrepancies between predicted and actual adjusted close prices. In the standard KAN (**Figure 2**), the model struggles to capture the upward trends and sharp fluctuations present in the data. In contrast, the tKAN (**Figure 5**) significantly reduces prediction error, particularly for recent time steps, by capturing temporal dependencies.

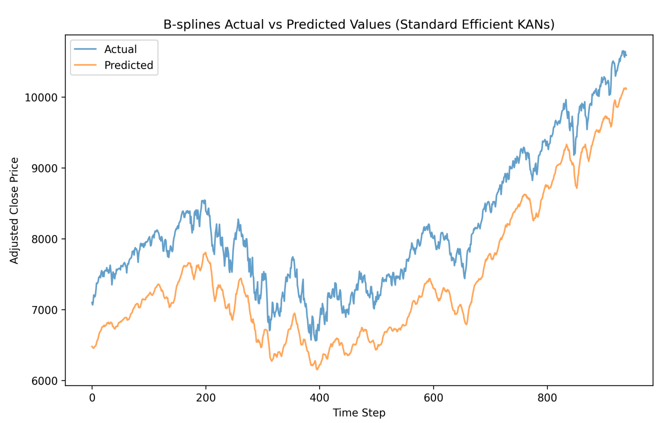

Fig. 1. B-splines Actual vs Predicted Values (Standard KAN)

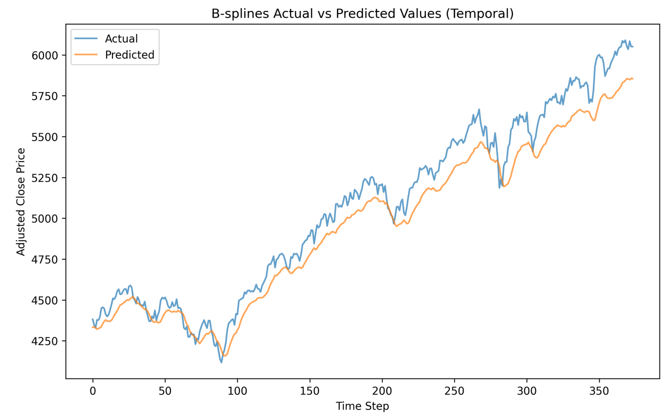

Fig. 2. B-splines Actual vs Predicted Values (Temporal KAN)

B. Comparison of Alternative Activations

The performance of Jacobi, Chebyshev, and Bessel activation functions was evaluated against the baseline. The results are summarized in **Table 1**.

The visualizations for alternative activations further illustrate the results. For Bessel functions (**Figure 3**), the standard KAN achieved the best MSE of 0.0036 and R^2 of 0.9672. This superior performance is attributed to Bessel functions' ability to capture oscillatory behaviors, which align well with periodic trends in financial time series data.

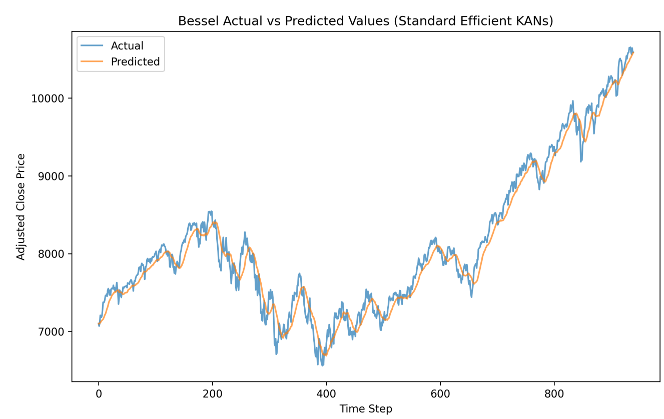

Fig. 3. Bessel Actual vs Predicted Values (Standard KAN)

Jacobi activations (**Figure 4**) demonstrated robustness with an MSE of 0.0051 and R^2 of 0.8544 in the standard KAN. When extended to temporal KANs (**Figure 8**), Jacobi activations achieved a Test MSE of 0.0023 and R^2 of 0.9333, highlighting their flexibility in capturing both smooth and sharp transitions.

For Chebyshev polynomials (**Figure 6** and **Figure 7**), the results varied significantly. While the standard KAN struggled with a Test MSE of 0.1011 and R^2 of 0.0718, the temporal KAN achieved the lowest overall MSE of 0.0014 and R^2 of 0.9614. This result emphasizes Chebyshev's ability to minimize approximation errors globally and its synergy with recurrent structures in tKAN.

TABLE I
PERFORMANCE COMPARISON OF ACTIVATION FUNCTIONS

Activation Function	Standard KAN MSE	Standard KAN R^2	Temporal KAN MSE	Temporal KAN R^2
B-Splines	0.0734	0.3261	0.0074	0.7897
Chebyshev	0.1011	0.0718	0.0014	0.9614
Jacobi	0.0051	0.8544	0.0023	0.9333
Bessel	0.0036	0.9672	-	-

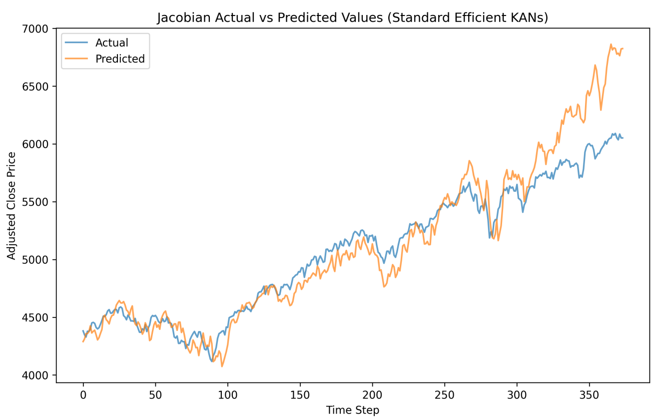

Fig. 4. Jacobian Actual vs Predicted Values (Standard KAN)

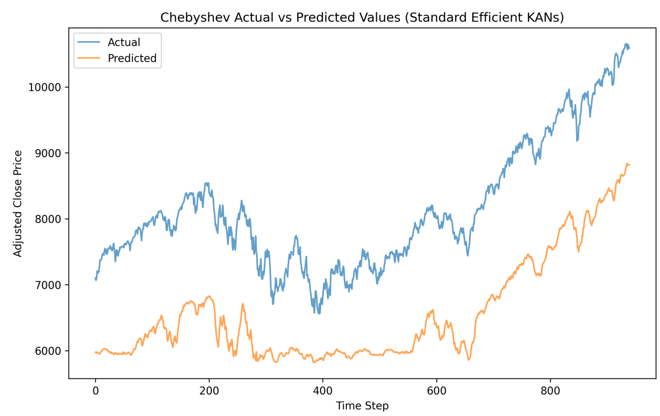

Fig. 6. Chebyshev Actual vs Predicted Values (Standard KAN)

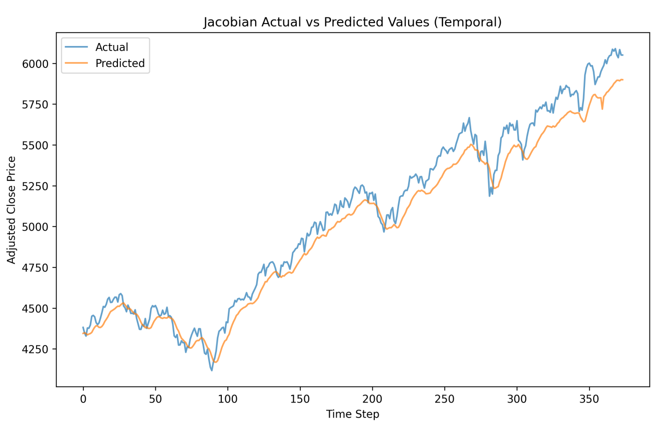

Fig. 5. Jacobian Actual vs Predicted Values (Temporal KAN)

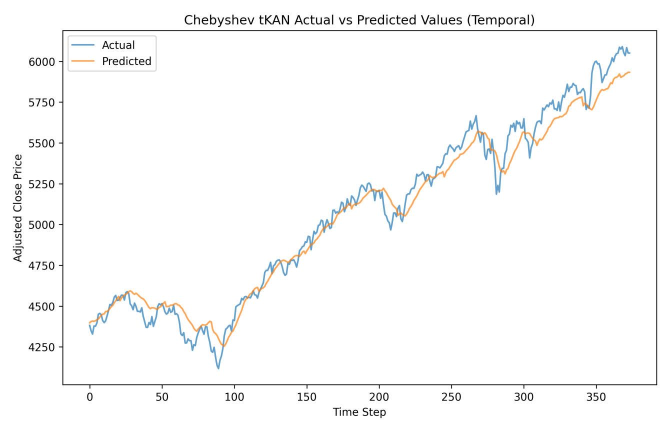

Fig. 7. Chebyshev Actual vs Predicted Values (Temporal KAN)

C. Key Observations and Deeper Analysis

From the results, the following key observations and insights can be made:

1) Baseline Analysis with B-Splines: B-splines demonstrate significant limitations in the standard KAN with a Test MSE of 0.0734 and R^2 of 0.3261. Their smooth nature fails to capture sharp transitions and high volatility, as evidenced by poor alignment with actual values in **Figure 1**. Temporal KANs alleviate these limitations by introducing LSTM layers, achieving an MSE of 0.0074 and R^2 of 0.7897. The LSTM layers enhance the model's ability to capture short-term and long-term dependencies.

2) Bessel Activation: Bessel functions excel in the standard KAN configuration with the lowest MSE of 0.0036 and R^2 of 0.9672. Their ability to approximate oscillatory behavior makes them particularly suitable for periodic trends in financial time series. **Figure 3** clearly illustrates the close overlap of predicted values with actual data.

3) Jacobi Activation: Jacobi activations offer consistent performance across both configurations. Their ability to balance smoothness and local adaptability is reflected in:

- Standard KAN: MSE = 0.0051, R^2 = 0.8544
- Temporal KAN: MSE = 0.0023, R^2 = 0.9333

Figures 4 and 8 highlight their success in capturing both smooth trends and sharp transitions, demonstrating their flexibility for dynamic and complex data.

4) Chebyshev Activation: Chebyshev polynomials struggle in the standard KAN (MSE = 0.1011, R^2 = 0.0718) but achieve the best performance in the temporal KAN (MSE = 0.0014, R^2 = 0.9614). This result can be attributed to Chebyshev's ability to minimize global approximation errors, which synergizes with the sequential learning of LSTMs. **Figures 6 and 7** confirm this observation, showcasing excellent alignment between predictions and actual values.

5) Activation Hierarchy: The deeper analysis reveals the following hierarchy of activation functions:

- **Chebyshev**: Best performance in temporal KANs due to global error minimization.
- **Jacobi**: Consistent and robust across both configurations, balancing smooth and local adaptations.
- **Bessel**: Superior performance for oscillatory data in standard KANs, with potential for further improvement in temporal settings.

6) Impact of Temporal KANs: The significant improvements in temporal KANs demonstrate the importance of modeling sequential dependencies. LSTM layers enhance each activation function's strengths by learning long-term patterns and mitigating shortcomings, as seen in Chebyshev's performance jump from standard to temporal KANs.

The combination of **activation flexibility** and **LSTM dynamics** makes temporal KANs particularly effective for financial forecasting, where both trends and volatility must be captured.

V. Conclusion and Future Work

A. Conclusion

This study explored the application of alternative activation functions—Jacobi, Chebyshev, and Bessel polynomials—in Kolmogorov-Arnold Networks (KANs) for time series forecasting. The results demonstrated that these activations significantly enhance the predictive performance of KANs, particularly when combined with temporal extensions such as LSTM layers. Chebyshev activations achieved the lowest Test MSE in temporal KANs, leveraging their ability to minimize global approximation error. Jacobi activations exhibited consistent robustness across both standard and temporal KANs, balancing smoothness and local adaptability. Bessel functions excelled in standard KANs by effectively capturing oscillatory behaviors, suggesting their potential for periodic datasets. Temporal KANs further amplified these benefits by incorporating sequential dependencies, leading to substantial improvements in model accuracy.

The findings validate the hypothesis that alternative activation functions can significantly enhance the adaptability and predictive performance of KANs for dynamic and complex time series data.

B. Future Work

While promising, this study leaves several avenues for further exploration:

1) **Extending Bessel Functions to Temporal KANs**: Investigate how the oscillatory strengths of Bessel functions can complement sequential learning in temporal models.
2) **Hybrid Activation Functions**: Develop and evaluate hybrid activations combining the strengths of Jacobi, Chebyshev, and Bessel polynomials to improve generalization across diverse datasets.
3) **Multi-Feature and Cross-Domain Applications**: Extend the models to multi-feature datasets (e.g., incorporating economic indicators alongside stock prices) and other domains like weather forecasting or healthcare analytics.
4) **Computational Optimization**: Explore efficient training methods for high-order polynomial activations to reduce computational costs without sacrificing accuracy.
5) **Real-World Deployment**: Validate the models in practical scenarios, such as trading systems or financial decision-making platforms, to assess their robustness and scalability.

By addressing these directions, future research can build on the foundation established in this work, advancing the state-of-the-art in interpretable, efficient, and accurate time series prediction models.

Acknowledgments

The author would like to thank Columbia University for providing guidance throughout this research. Special thanks to the IEEE community for fostering innovation in applied machine learning.

References

[1] M. Davis and M. Lopez, "C-kan: A new approach for integrating convolutional layers with kolmogorov-arnold networks for time-series forecasting," *Mathematics*, 2024. [Online]. Available: https://www.mdpi.com/2227-7390/12/19/3022

[2] B. Ter-Avanesov and H. Beigi, "Mlp, xgboost, kan, tdnn, and lstm-gru hybrid rnn with attention for spx and ndx european call option pricing," *Technical Report: RTI-20240822-01*, 2024. [Online]. Available: http://dx.doi.org/10.13140/RG.2.2.32372.56963

[3] J. Smith and W. Zhang, "A temporal kolmogorov-arnold transformer for time series forecasting," *arXiv preprint arXiv:2406.02486*, 2024. [Online]. Available: https://arxiv.org/abs/2406.02486

[4] F. Author1, F. Author2, and F. Author3, "Exploring generalized kolmogorov-arnold networks: Theory and applications," *arXiv preprint arXiv:2404.19756*, 2024. [Online]. Available: https://arxiv.org/abs/2404.19756

[5] E. Taylor and D. Robinson, "Kolmogorov-arnold networks for time series: Bridging predictive power and interpretability," *arXiv preprint arXiv:2406.02496*, 2024. [Online]. Available: https://arxiv.org/abs/2406.02496

[6] G. C. Blealtan, "Efficient kan: Optimized implementation of kolmogorov-arnold networks," 2024. [Online]. Available: https://github.com/Blealtan/efficient-kan

[7] R. White and J. Black, "Chebyshev feature neural network for accurate function approximation," *arXiv preprint arXiv:2409.19135*, 2024. [Online]. Available: https://arxiv.org/abs/2409.19135

[8] M. Brown and S. Jones, "fkan: Fractional kolmogorov-arnold networks with trainable jacobi basis," *arXiv preprint arXiv:2406.07456*, 2024. [Online]. Available: https://arxiv.org/pdf/2406.07456

[9] P. Anderson and L. Green, "Quaternionic convolutional neural networks with trainable bessel activation functions," *Advances in Applied Mathematics*, 2024. [Online]. Available: https://link.springer.com/article/10.1007/s11785-023-01387-z

[10] G. Contributors, "Tkan: Temporal kolmogorov-arnold networks," 2024. [Online]. Available: https://github.com/remigenet/tkan

[11] ——, "Easytsf: Kolmogorov arnold network (kan) for time series forecasting," 2024. [Online]. Available: https://github.com/2448845600/easytsf