Diagnosing Data Irregularities in Financial Risk Forecasting: A Wavelet-Based Augmentation Study.

Charles B Bramble, Prof. Xianghua Xie, Dr. Gary Tam, Dr. Kevin Mclafferty
Swansea University Department of Computer Science
C.B.Bramble@Swansea.ac.uk, X.Xie@Swansea.ac.uk, K.L.Tam@Swansea.ac.uk, Kevin.Mclafferty@Swansea.ac.uk

Abstract

Financial institutions increasingly rely on predictive models to assess internal risk scores for companies, yet real-world organisational risk-scoring data remains underexplored. This task is complicated by structural inconsistencies driven by policy changes, user-entry errors, measurement noise, and sparse observations. These issues challenge models to preserve temporal coherence while handling abrupt shifts and missing values. In collaboration with a major UK commercial bank, we curate a proprietary dataset of company-level risk scores and conduct an exploratory analysis to characterise its irregularities. Building on these insights, we propose a wavelet-based diagnostic augmentation approach that perturbs discrete wavelet coefficients to introduce realistic temporal variability while maintaining structural integrity. Experiments using LSTM forecasting models show that wavelet-domain perturbations yield consistent improvements in predictive stability—reducing mean absolute error (MAE) from 3.25 to 3.22 on average—and expose systematic sensitivities to noisy inputs. Although the performance gains are modest, the results highlight the importance of robustness evaluation in financial time series and position wavelet-domain perturbation as a practical diagnostic tool for model reliability under imperfect data conditions.

1 Introduction

Financial institutions increasingly rely on data-driven models to assess risk, forecast company behaviour, and support decision-making across large and diverse portfolios. These predictive systems underpin key operational processes—from credit evaluation to portfolio monitoring—and their success depends heavily on the quality, consistency, and temporal reliability of the underlying data.

In practice, however, real-world financial datasets rarely meet these ideal conditions. They often contain structural inconsistencies caused by policy changes, user-entry errors, measurement noise, and missing values. Such irregularities disrupt temporal coherence and introduce hidden biases, posing significant challenges for models that depend on stable sequential patterns. As machine learning becomes increasingly central to financial forecasting, addressing these imperfections is critical for achieving reliable and explainable model performance in production environments.

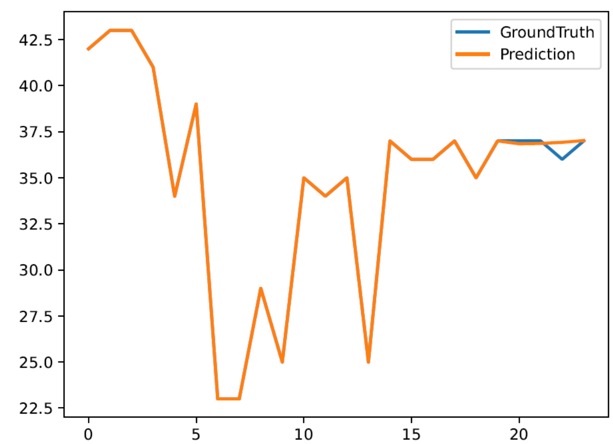

Figure 1: Example prediction of a company-level financial risk score exhibiting high volatility and measurement noise. The instability and abrupt changes illustrate the challenges faced by forecasting models trained on real-world organisational data.

Recent advances in time-series forecasting have focused on improving model robustness under noisy or irregular conditions. Augmentation techniques—ranging from simple time-domain transformations to frequency-based and generative approaches—help models generalise by introducing realistic variability while preserving temporal dependencies (Wen et al. 2021; Gao et al. 2020; Yoon, Jarrett, and van der Schaar 2019). Wavelet-based methods, in particular, have proven effective for analysing non-stationary signals by jointly capturing short-term fluctuations and long-term trends (Rhif et al. 2019; Percival and Walden 2000). Meanwhile, deep learning architectures such as LSTMs, TCNs, and Transformers now outperform traditional statistical methods for financial forecasting (Ziółkowski and Wieczyński 2025; Kabir et al. 2025). However, most of these studies rely on clean, publicly available datasets (e.g., stock or cryptocurrency prices) rather than the irregular, policy-driven organisational data examined in this work (Figure 1).

Despite extensive research in controlled settings, little attention has been paid to financial datasets derived from organisational systems—where human input, evolving policies, and internal scoring mechanisms introduce abrupt structural shifts, sparsity, and noise. These characteristics violate common modelling assumptions and hinder generalisation, yet they remain underexplored in both the financial AI and time-series literature.

In collaboration with a a large incumbent commercial banking organisation in the UK, we analyse a proprietary dataset of company-level financial risk scores to characterise these real-world irregularities. We then employ a wavelet-based diagnostic augmentation approach that perturbs discrete wavelet coefficients to simulate realistic variability, enabling a controlled evaluation of model sensitivity and robustness.

1.1 Contributions

This paper makes three primary contributions:

- **Empirical insight:** We provide the first exploratory analysis of large-scale organisational financial risk-scoring data, revealing the impact of policy and measurement irregularities on model performance.
- **Diagnostic augmentation:** We propose a wavelet-domain perturbation method that introduces realistic spectral variability, enabling controlled evaluation of model robustness under imperfect conditions.
- **Practical implications:** We discuss how diagnostic augmentation can guide data-centric improvements to forecasting pipelines in institutional finance.

Together, these contributions bridge a gap between academic time-series forecasting and the operational realities of institutional datasets, establishing a foundation for future research on learning under data irregularity.

2 Related Work

2.1 Data Augmentation in Time Series

Data augmentation is a core strategy for improving model generalisation in time-series forecasting (Wen et al. 2021; Annaki, Rahmoune, and Bourhaleb 2024; Iglesias et al. 2023), particularly under conditions of data scarcity, imbalance, or noise (Wang et al. 2021; Ramponi et al. 2018; Jiang et al. 2019; Fons et al. 2020). Unlike spatial data, temporal signals require transformations that preserve sequential dependencies, making direct adaptation of image-based augmentation techniques unsuitable (Lashgari, Liang, and Maoz 2020).

Early approaches introduced simple time-domain operations—such as jittering, scaling, time warping, and cropping—to expand training diversity while maintaining sequence semantics (Rashid and Louis 2019; Guennec, Malinowski, and Tavenard 2016; Cui, Chen, and Chen 2016). More recent work has explored frequency-domain transformations, including Fourier- and wavelet-based perturbations, to generate realistic temporal variability by modifying spectral characteristics (Gao et al. 2020; Eileen et al. 2019). These methods enhance robustness to distributional shifts and are particularly valuable in irregular domains such as finance and industry.

Generative models, including GANs and VAEs, have also been applied to synthesise new temporal sequences (Esteban, Hyland, and Rätsch 2017; Yoon, Jarrett, and van der Schaar 2019; Fu, Kirchbuchner, and Kuijper 2020). While these methods can capture complex dependencies, they often trade interpretability and computational efficiency for flexibility. Despite these advances, maintaining temporal coherence and realistic variability remains challenging, especially when dealing with short, sparse, or structurally inconsistent series—conditions typical of organisational financial data.

2.2 Wavelet Methods in Time Series

Wavelet analysis provides a powerful framework for representing non-stationary time-series data by jointly localising information in both the time and frequency domains (Rhif et al. 2019). Unlike Fourier transforms, which assume global stationarity, wavelet decompositions capture both short-term fluctuations and long-term trends (Percival and Walden 2000; Guo, Yang, and Sano 2023). This property makes wavelets particularly effective for real-world signals—such as physiological, environmental, and financial data—where structural shifts and abrupt changes are common (Addison 2017; Wen et al. 2020).

Within machine learning, wavelet transforms have been applied for feature extraction, denoising, and data augmentation. Coefficient-based representations offer compact, interpretable features that improve robustness to noise and missing values (Chen et al. 2024; Raimundo and Okamoto 2018). Denoising techniques selectively suppress high-frequency artefacts while retaining meaningful temporal dynamics, enhancing predictive performance in noisy and irregular datasets (Stefenon et al. 2024; Al Wadia and Ismail 2011; Ziółkowski and Wieczyński 2025; Yan and Ouyang 2018).

In contrast to prior work focused on feature engineering or denoising, our study employs wavelet-domain perturbation as a diagnostic augmentation tool. Rather than seeking to enlarge the dataset or denoise inputs, we use controlled perturbations of wavelet coefficients to evaluate model robustness under realistic structural irregularities—a perspective that, to our knowledge, remains largely unexplored in real world financial forecasting.

2.3 Forecasting in Financial Time Series

Financial forecasting has long been a central application of time-series modelling. Classical statistical methods such as ARIMA, VAR, and GARCH have been widely used to capture linear dependencies and volatility dynamics (Ho and Xie 1998; Bauwens, Laurent, and Rombouts 2006). With the advent of deep learning, architectures including LSTMs, GRUs, and Temporal Convolutional Networks (TCNs) have become standard for modelling nonlinear temporal relationships (Yan and Ouyang 2018; Ziółkowski and Wieczyński 2025; Kabir et al. 2025). More recently, Transformer-based approaches have been applied to multivariate financial forecasting and portfolio optimisation (Zeng et al. 2023; Olorunnimbe and Viktor 2024; Huang, Wang, and Yang 2023).

However, the vast majority of these studies rely on clean, high-frequency public datasets such as stock indices or cryptocurrency prices (Sezer, Gudelek, and Ozbayoglu 2020).

In contrast, internal financial datasets—such as company-level risk scores used in commercial banking—exhibit noise, sparsity, and policy-driven structural shifts that violate common assumptions of temporal smoothness. Few studies explicitly address how such irregularities impact model performance or explore methods for improving robustness in these settings (Pehlivanlı, Alp, and Katanalp 2024).

This work contributes to closing that gap by analysing a proprietary organisational dataset and applying wavelet-domain diagnostic augmentation to probe the stability of forecasting models trained under realistic, imperfect data conditions.

3 Dataset

3.1 Data Characteristics and Challenges

The proprietary datasetwas provided through collaboration with a large UK banking organisation. It contains monthly company-level financial risk scores used internally for monitoring client performance. For this study, only the univariate risk score variable was used, though the wider dataset includes other indicators such as profitability, loan usage, categorical ratings, and account balances. Each time series represents up to 24 months of observations, with the final four used as ground-truth forecasting targets. The data were sourced directly from the bank's internal systems and validated through consultation with data owners and analysts to ensure structural consistency. Approximately 5,000 companies were selected, each having changed categorical status at least once within 24 months to ensure dynamic behaviour. Series with over 50% missing values were excluded, and remaining sequences were padded or truncated to 24 months with masked placeholders for missing entries. This subset reflects the short, irregular, and noisy nature of operational financial data, making it a representative basis for testing model robustness and augmentation effects.

3.2 Data Pre-processing

A consistent pre-processing pipeline is applied across all experiments to ensure comparability between baseline and perturbed datasets. Each time series is padded or truncated to a uniform length of 24 time steps. Missing entries are assigned a placeholder value of –999 and masked during model training. When the first observation is missing, the first valid value is forward-filled to initialise the sequence. Series with more than 50% missing values are excluded to limit imputation bias. No normalisation or standardisation is applied, as preliminary testing indicated that such transformations reduced model interpretability and temporal consistency. These steps ensure that any observed effects on model performance can be attributed to the diagnostic perturbations rather than inconsistencies in preprocessing. Perturbations only occur on the training set, which consists of 75% of the dataset as there is a 75/25 split for training/testing.

4 Methodology

4.1 Overview

This study adopts an experimental pipeline designed to investigate the robustness of financial forecasting models under realistic data perturbations. The process involves five main stages: data preparation and preprocessing (Section 3.2), wavelet-domain transformation (Section 4.2), controlled coefficient perturbation and reconstruction (Section 4.3), diagnostic dataset expansion (Section 4.4), and model training and evaluation (Section 4.5). Rather than introducing a novel augmentation framework, the pipeline functions as a diagnostic tool to assess how small, structured variations in the data influence model stability and predictive accuracy. The approach remains model-agnostic and may be extended to other sequential datasets exhibiting irregular sampling, volatility, or missing data patterns.

4.2 Wavelet Transform

To capture both localised fluctuations and broader temporal trends, each time series is decomposed using a Discrete Wavelet Transform (DWT) with a Daubechies-1 (db1) basis—equivalent to the Haar wavelet. Wavelet transforms provide joint localisation in time and frequency, enabling analysis of non-stationary signals where abrupt changes and structural shifts occur (Rhif et al. 2019; Yan and Ouyang 2018). The db1 wavelet was chosen for its simplicity, computational efficiency, and interpretability when applied to short financial sequences. This decomposition separates the signal into low-frequency approximation coefficients (representing long-term trends) and high-frequency detail coefficients (capturing short-term variability and potential noise).

4.3 Coefficient Perturbation and Reconstruction

Let each univariate time series be denoted as $x = \{x_t\}_{t=1}^{T}$, where $T = 24$ months in this study. The series is first decomposed using the Discrete Wavelet Transform (DWT):

$$[a_0, d_1, d_2] = \mathrm{DWT}(x),$$

where a_0 represents the approximation (low-frequency) coefficients and d_i the detail (high-frequency) coefficients for bands $i = 1, 2$.

To simulate realistic variability while preserving temporal structure, both multiplicative and additive Gaussian perturbations are applied to the coefficients:

$$\begin{aligned} a_0' &= \gamma_0\, a_0 + \epsilon_0, \qquad \gamma_0 \sim \mathcal{U}(0.8, 1.2),\ \epsilon_0 \sim \mathcal{N}(0, 0.1^2), \\ d_i' &= \gamma_i\, d_i + \epsilon_i, \qquad \gamma_i \sim \mathcal{U}(0.5, 1.5),\ \epsilon_i \sim \mathcal{N}(0, 0.1^2). \end{aligned}$$

The perturbed signal $\hat{x}$ is then reconstructed via the inverse wavelet transform:

$$\hat{x} = \mathrm{IDWT}(a_0', d_1', d_2').$$

Masked or missing values remain unaltered by overwriting them with their original counterparts following reconstruction. This process yields a synthetic series that maintains the structural fidelity of the original while introducing controlled stochastic variability, enabling diagnostic evaluation of model sensitivity to realistic perturbations.

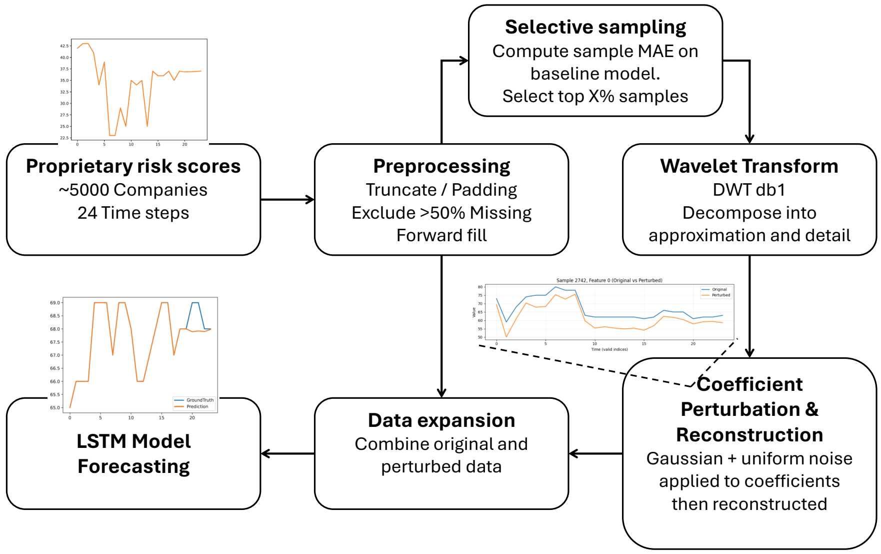

Figure 2: Overview of the experimental pipeline. The process includes data preprocessing, wavelet decomposition, coefficient perturbation and reconstruction, diagnostic dataset expansion, and LSTM forecasting. A selective sampling branch identifies high-error cases for targeted augmentation.

4.4 Diagnostic Dataset Expansion

The reconstructed signals are appended to the original dataset, effectively creating an expanded dataset $\mathcal{D}_{\text{aug}} = \mathcal{D}_{\text{orig}} \cup \mathcal{D}_{\text{pert}}$, where $\mathcal{D}_{\text{orig}}$ denotes the original time series and $\mathcal{D}_{\text{pert}}$ represents their perturbed counterparts. This expansion allows the model to encounter both original and noise-augmented series during training, enabling a diagnostic evaluation of how controlled variability influences forecasting robustness. Rather than pursuing augmentation solely for generalisation, this approach functions as a stress test—assessing whether the model learns stable representations that remain consistent under minor but realistic perturbations.

4.5 Model Architecture

A univariate Long Short-Term Memory (LSTM) network was employed as the baseline forecasting model. This architecture was selected for its ability to capture temporal dependencies within short sequences while remaining interpretable and computationally efficient for repeated experimentation. Each model received up to 24 months of company-level financial scores as input and was trained to predict the final four time steps. The LSTM framework was deliberately kept minimal to prioritise interpretability and ensure that any observed performance changes could be attributed primarily to data characteristics or augmentation effects rather than architectural complexity. No additional exogenous variables were included, aligning with the univariate focus of this initial investigation. This model therefore serves as a controlled probe into the behaviour of the dataset under varying augmentation conditions.

5 Experimental Setup and Evaluation Metrics

To evaluate the impact of wavelet-domain augmentation and assess forecasting stability under realistic data irregularities, all experiments were conducted using the proprietary commercial banking financial risk-scoring dataset described in Section 3. The forecasting task involves predicting the final four months of a 24-month univariate series, where each sequence corresponds to a distinct company. All data were preprocessed and masked following the procedures outlined in Section 3.2.

Model training employed the LSTM architecture introduced in Section 4.5. To assess sensitivity to model capacity, the hidden layer size was varied between 32 and 512 units, while other hyperparameters were kept constant to ensure comparability. Training used a masked mean squared error loss, with masking applied to both padding and invalid (zero) values. A 75/25 train–test split was used, with a batch size of 128, a maximum of 150 epochs, and early stopping (patience of 10 epochs) to mitigate overfitting.

5.1 Evaluation Metrics

Model performance was assessed using mean absolute error (MAE) and mean squared error (MSE), computed over the masked output region to ensure fair comparison between augmented and non-augmented configurations. Additionally, relative error change (ΔMAE, ΔMSE) was reported to quantify the directional influence of augmentation. These metrics collectively provide insight into both absolute forecasting accuracy and the consistency of predictions under controlled perturbations.

Three core experimental conditions were evaluated:

- **Baseline:** Model trained directly on the original dataset without augmentation.
- **Full augmentation:** Wavelet-domain perturbation applied to all training samples, doubling the effective dataset size.
- **Targeted augmentation:** Perturbation applied selectively to the 40–60% of samples exhibiting the highest predictive error, producing oversampled variants for these difficult cases.

Evaluation was based primarily on Mean Absolute Error (MAE), averaged across the full test set. To provide a naive performance reference, a simple persistence model predicting the previous observed value was also computed. In addition to quantitative metrics, qualitative visualisations were produced, including original versus perturbed samples, model prediction overlays, and error correlations against basic signal statistics (mean, standard deviation, and maximum) and individual wavelet bands. Together, these experiments allow for reproducible assessment of how data irregularity interacts with model robustness and the extent to which controlled perturbation can mitigate or reveal these issues.

5.2 Training Protocol

All models were trained under identical conditions to isolate the impact of data augmentation. The dataset was partitioned using a 75/25 train–test split, ensuring that augmented samples appeared exclusively in the training partition to prevent data leakage. Model training employed a masked mean squared error (MSE) loss function, excluding padding and placeholder values (–999 and 0) from gradient updates. The Adam optimiser was used with a learning rate of 1×10^{-3}. No dropout or weight decay was applied, as preliminary experiments indicated a reduction in performance when regularisation was introduced. Early stopping was based on validation loss measured on the test partition, with training capped at 150 epochs and a patience of 10 epochs to mitigate overfitting. Model parameters were initialised consistently across runs to ensure comparability. The primary evaluation metric was mean absolute error (MAE), computed over the unmasked prediction range.

6 Results & Discussion

6.1 Quantitative Results

For benchmarking, a naïve baseline that repeated the most recent observed value achieved a mean absolute error (MAE) of 4.424, establishing a lower bound of predictability for this dataset. In comparison, all LSTM-based models achieved substantially lower errors, confirming the presence of learnable temporal structure within the financial risk scores. The detailed performance of each configuration is reported in Table 1.

Key results:

- Naïve baseline: MAE = 4.424
- Standard LSTM (no augmentation): MAE = 3.253
- Full-sample perturbation: MAE = 3.223 (best performance)
- Targeted 60% perturbation: MAE = 3.230
- Targeted 40% perturbation: MAE = 3.265

These results indicate that wavelet-domain augmentation produces small yet consistent improvements in predictive accuracy. The best configuration, obtained through full-sample perturbation, reduced MAE by approximately 1% relative to the baseline LSTM. The marginal nature of these gains suggests that while the augmentation introduces meaningful variability, its impact is constrained by dataset size and inherent noise levels. The narrow performance range across model capacities further implies that the primary limitation lies in data quality rather than representational power.

Error decomposition across prediction horizons showed a gradual increase in MAE (2.3, 3.1, 3.5, and 3.9 for successive forecast steps), reflecting a typical decay in autoregressive accuracy. Nonetheless, the rate of degradation remained moderate given the dataset's irregular temporal structure, supporting the method's stability under sparse and noisy conditions.

These results collectively indicate that while augmentation can yield small but consistent improvements, the magnitude of change remains modest. This suggests either (1) the dataset size and variability are insufficient to fully leverage perturbation-based augmentation, or (2) that more expressive augmentation methods are required to meaningfully enhance model generalisation. The narrow range of MAE values across configurations further supports the view that the model is limited primarily by data quality and availability rather than representational capacity. Error decomposition across prediction horizons revealed a gradual increase in MAE over time: 2.3, 3.1, 3.5, and 3.9 for successive horizons. This progression demonstrates a predictable decay in accuracy typical of autoregressive forecasting, though the decline is relatively moderate given the dataset's irregularity.

6.2 Visual Analysis

To complement the quantitative results, a series of visual diagnostics were conducted to examine the behaviour of the wavelet perturbations, model predictions, and error characteristics across the dataset.

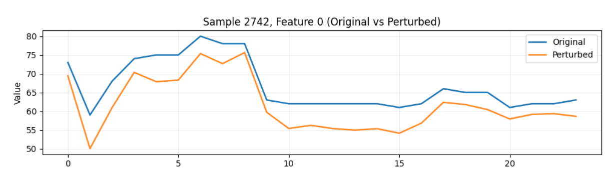

Figure 3: Example of an original and its corresponding wavelet-perturbed series. The perturbation introduces realistic temporal variability while preserving the structural integrity and overall dynamics of the original signal.

Figure 3 illustrates examples of original time series and their corresponding wavelet-perturbed versions. The perturbations introduce realistic variability without distorting the

Table 1: Performance summary of baseline and augmented models on the financial risk-scoring dataset. The best results are highlighted in bold.

Model	**Hidden Units**	**Layers**	**Augmentation Strategy**	**MAE ↓**	**RMSE ↓**	**SMAPE (%) ↓**
Naïve baseline	–	–	None	4.424	–	–
LSTM (no augmentation)	128	1	None	3.253	8.286	7.71
LSTM (Full perturbation)	128	1	All samples	**3.223**	**8.206**	**7.59**
LSTM (60% perturbation)	128	1	Top 60% error samples	3.230	8.198	7.61
LSTM (40% perturbation)	128	1	Top 40% error samples	3.265	8.273	7.82

overall temporal structure or underlying relationships between observations. In most cases, the perturbed signals retained coherent patterns consistent with genuine financial fluctuations. However, in a minority of samples where the source data contained clear anomalies as seen in Figure 5—such as abrupt drops to implausible low scores—the perturbations amplified these artifacts, occasionally producing values that deviated further from the expected range. This highlights the sensitivity of wavelet-domain perturbations to erroneous inputs and underscores the importance of pre-filtering or anomaly detection prior to augmentation.

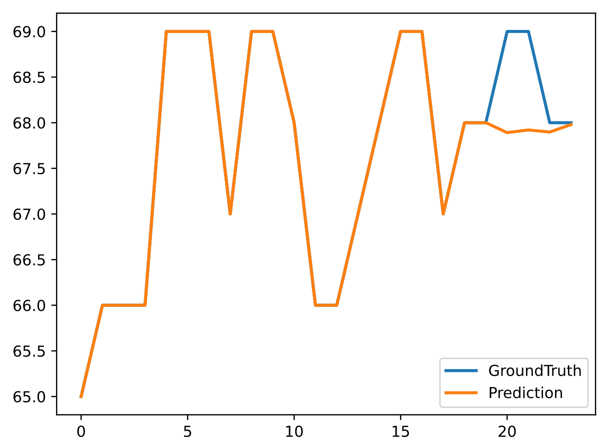

Figure 4: Illustration of a highly volatile company score series and its corresponding poor predictive fit. The prediction lag and large deviation highlight the model's difficulty in capturing abrupt shifts and irregular score behaviour.

The predicted versus actual series plots further revealed that the LSTM model captured general temporal trends more effectively for stable companies than for those that exhibit sharp or irregular changes as can be seen in Figure 4. The augmented models occasionally demonstrated slightly improved trend alignment, but these differences were not visually or quantitatively substantial. Both the baseline and perturbed models tended to lag during periods of rapid change, suggesting that the underlying challenge lies in the volatility of the data rather than in model capacity or augmentation.

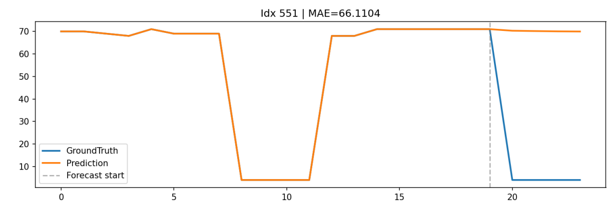

Figure 5: Example of non-zero error samples within an otherwise stable series. Localised anomalies and erratic spikes contribute disproportionately to overall forecasting error, revealing sensitivity to isolated input irregularities.

To better understand the relationship between error and signal characteristics, per-sample mean absolute error (MAE) was plotted against each sequence's maximum, mean, and standard deviation. No clear correlation was observed with the mean or maximum values, but a positive relationship emerged between standard deviation and prediction error as seen in Figure 7, indicating that volatility within a company's score history strongly influences forecast difficulty. In contrast, decomposed wavelet-band features (bands 0–2) showed no consistent relationship with error, suggesting that frequency-specific variability contributes less to overall predictability than raw volatility.

Finally, as seen in Figure 6 horizon-wise metrics showed a steady degradation in performance across successive forecast steps. MAE, RMSE, and SMAPE all increased gradually with horizon length, while MAPE displayed an anomalous dip at the final horizon—likely reflecting its instability when dealing with small denominators in percentage-based error calculations. Collectively, these patterns confirm that the dataset's structural inconsistencies and volatility, rather than the absence of augmentation, remain the dominant drivers of model error. Overall, the visual analyses reinforce the conclusion that while wavelet perturbations produce structurally coherent augmentations, they cannot compensate for the inherent complexity and noisiness of organisational financial data. More robust methods—potentially combining anomaly detection, targeted interpolation, or generative temporal modelling—are needed to meaningfully improve predictive stability in such environments.

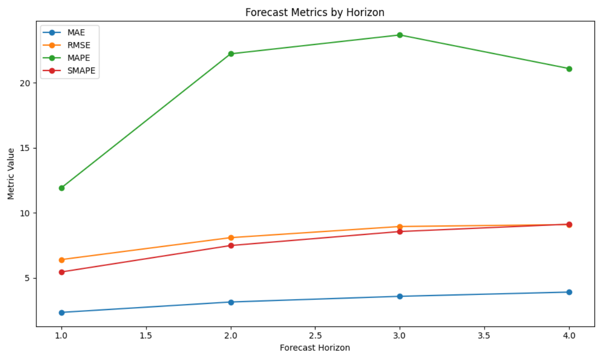

Figure 6: Forecasting error metrics across successive prediction horizons. Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Symmetric Mean Absolute Percentage Error (SMAPE) all show gradual degradation.

6.3 Limitations

This study is constrained by several factors relating to both data and methodology. First, the experiments were conducted solely on a univariate subset of approximately 5,000 companies, each represented by up to 24 monthly risk scores. While sufficient for exploratory analysis, this limited scope restricts generalisation across broader financial indicators. Multivariate extensions are expected to capture richer dependencies and may yield improved performance. Second, due to confidentiality agreements, neither the dataset nor the implementation can be publicly released, which limits external reproducibility and direct comparison with benchmark studies. Third, the wavelet-based augmentation was employed as a diagnostic tool rather than a fully optimised technique. Only the Daubechies-1 (db1) basis was used, and perturbation parameters were manually tuned within a narrow range. More extensive tuning or alternative wavelet families may better capture the complex temporal structure of the data.

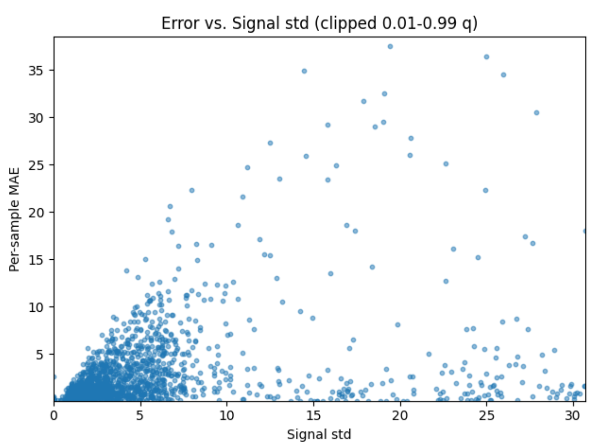

Figure 7: Relationship between per-sample prediction error and signal standard deviation.

Finally, the evaluation was restricted to an LSTM forecasting model with fixed architectural variations. Broader experimentation with more advanced sequence models, particularly under multivariate conditions, could better reveal the potential of augmentation strategies. Despite these limitations, the study provides an essential first step toward understanding the irregularities inherent in organisational financial time series.

7 Conclusion

Forecasting organisational financial risk scores remains challenging due to data sparsity, measurement noise, and policy-driven irregularities that dominate error patterns. Our experiments demonstrate that wavelet-based perturbation introduces realistic variability and marginally improves performance (MAE $\downarrow$ 3.253 $\rightarrow$ 3.223), yet the effect is limited—highlighting that data quality, rather than augmentation or model complexity, is the primary bottleneck. Nonetheless, the approach proved valuable as a diagnostic tool for probing temporal sensitivity and robustness, offering insight into how volatility and missingness shape predictive stability.

Future work will focus on richer data-centric strategies, including multivariate augmentation to capture latent dependencies, adaptive perturbation guided by uncertainty to prevent amplification of erroneous regions, and anomaly-aware preprocessing to reduce corrupted signals. We also plan to extend evaluation to modern sequence architectures such as Transformers and Temporal Convolutional Networks (TCNs), and to conduct systematic analyses linking forecast error to volatility and missingness. These directions aim to bridge the gap between controlled experimentation and real-world institutional forecasting, where imperfect and irregular data remain the norm.

A Acknowledgments

This work was supported in part by EPSRC under grant EP/S021892/1 and in part by a large incumbent commercial banking organisation in the UK.

References

Addison, P. S. 2017. *The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science, Engineering, Medicine and Finance*. CRC Press, 2 edition.

Al Wadia, M.; and Ismail, M. T. 2011. Selecting wavelet transforms model in forecasting financial time series data based on ARIMA model. *Applied Mathematical Sciences*, 5(7): 315–326.

Annaki, I.; Rahmoune, M.; and Bourhaleb, M. 2024. Overview of Data Augmentation Techniques in Time Series Analysis. *IJACSA*, 15(1).

Bauwens, L.; Laurent, S.; and Rombouts, J. V. K. 2006. Multivariate GARCH models: a survey. *Journal of Applied Econometrics*, 21(1): 79–109.

Chen, X.; Qiu, P.; Zhu, W.; Li, H.; Wang, H.; Sotiras, A.; Wang, Y.; and Razi, A. 2024. TimeMIL: Advancing Multivariate Time Series Classification via a Time-aware Multiple Instance Learning. arXiv:2405.03140.

Cui, Z.; Chen, W.; and Chen, Y. 2016. Multi-Scale Convolutional Neural Networks for Time Series Classification. *CoRR*, abs/1603.06995.

Eileen, K. M. T. L.; Kuah, K.-L.; Leo, K. H.; Sanei, S.; Chew, E.; and Zhao, L. 2019. Surrogate Rehabilitative Time Series Data for Image-based Deep Learning. *2019 27th European Signal Processing Conference (EUSIPCO)*, 1–5.

Esteban, C.; Hyland, S. L.; and Rätsch, G. 2017. Real-valued (Medical) Time Series Generation with Recurrent Conditional GANs. arXiv:1706.02633.

Fons, E.; Dawson, P.; jun Zeng, X.; Keane, J.; and Iosifidis, A. 2020. Evaluating data augmentation for financial time series classification. arXiv:2010.15111.

Fu, B.; Kirchbuchner, F.; and Kuijper, A. 2020. Data augmentation for time series: traditional vs generative models on capacitive proximity time series. In *Proceedings of the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, PETRA '20. New York, NY, USA: Association for Computing Machinery. ISBN 9781450377737.

Gao, J.; Song, X.; Wen, Q.; Wang, P.; Sun, L.; and Xu, H. 2020. RobustTAD: Robust Time Series Anomaly Detection via Decomposition and Convolutional Neural Networks. *CoRR*, abs/2002.09545.

Guennec, A. L.; Malinowski, S.; and Tavenard, R. 2016. Data Augmentation for Time Series Classification using Convolutional Neural Networks. In *ECML/PKDD Workshop on Advanced Analytics and Learning on Temporal Data*. Riva Del Garda, Italy.

Guo, P.; Yang, H.; and Sano, A. 2023. Empirical Study of Mix-based Data Augmentation Methods in Physiological Time Series Data. In *2023 IEEE 11th International Conference on Healthcare Informatics (ICHI)*, 206–213.

Ho, S.; and Xie, M. 1998. The use of ARIMA models for reliability forecasting and analysis. *Computers & Industrial Engineering*, 35(1): 213–216.

Huang, A. H.; Wang, H.; and Yang, Y. 2023. FinBERT: A Large Language Model for Extracting Information from Financial Text. *Contemporary Accounting Research*, 40(2): 806–841.

Iglesias, G.; Talavera, E.; González-Prieto, Á.; Mozo, A.; and Gómez-Canaval, S. 2023. Data Augmentation techniques in time series domain: a survey and taxonomy. *Neural Computing and Applications*, 35(14): 10123–10145.

Jiang, W.; Hong, Y.; Zhou, B.; He, X.; and Cheng, C. 2019. A GAN-Based Anomaly Detection Approach for Imbalanced Industrial Time Series. *IEEE Access*, 7: 143608–143619.

Kabir, M. R.; Bhadra, D.; Ridoy, M.; and Milanova, M. 2025. LSTM–Transformer-Based Robust Hybrid Deep Learning Model for Financial Time Series Forecasting. *Sci*, 7(1).

Lashgari, E.; Liang, D.; and Maoz, U. 2020. Data augmentation for deep-learning-based electroencephalography. *Journal of Neuroscience Methods*, 346: 108885.

Olorunnimbe, K.; and Viktor, H. 2024. Ensemble of temporal Transformers for financial time series. *Journal of Intelligent Information Systems*, 62(4): 1087–1111.

Pehlivanlı, D.; Alp, E. A.; and Katanalp, B. 2024. Introducing the overall risk scoring as an early warning system. *Expert Systems with Applications*, 246: 123232.

Percival, D. B.; and Walden, A. T. 2000. *Wavelet Methods for Time Series Analysis*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Raimundo, M. S.; and Okamoto, J. 2018. SVR-wavelet adaptive model for forecasting financial time series. In *2018 International Conference on Information and Computer Technologies (ICICT)*, 111–114.

Ramponi, G.; Protopapas, P.; Brambilla, M.; and Janssen, R. 2018. T-CGAN: Conditional Generative Adversarial Network for Data Augmentation in Noisy Time Series with Irregular Sampling. *CoRR*, abs/1811.08295.

Rashid, K. M.; and Louis, J. 2019. Times-series data augmentation and deep learning for construction equipment activity recognition. *Advanced Engineering Informatics*, 42: 100944.

Rhif, M.; Ben Abbes, A.; Farah, I. R.; Martínez, B.; and Sang, Y. 2019. Wavelet Transform Application for/in Non-Stationary Time-Series Analysis: A Review. *Applied Sciences*, 9(7).

Sezer, O. B.; Gudelek, M. U.; and Ozbayoglu, A. M. 2020. Financial time series forecasting with deep learning : A systematic literature review: 2005–2019. *Applied Soft Computing*, 90: 106181.

Stefenon, S. F.; Seman, L. O.; da Silva, E. C.; Finardi, E. C.; dos Santos Coelho, L.; and Mariani, V. C. 2024. Hypertuned wavelet convolutional neural network with long short-term memory for time series forecasting in hydroelectric power plants. *Energy*, 313: 133918.

Wang, Q.; Farahat, A. K.; Gupta, C.; and Zheng, S. 2021. Deep Time Series Models for Scarce Data. *CoRR*, abs/2103.09348.

Wen, Q.; He, K.; Sun, L.; Zhang, Y.; Ke, M.; and Xu, H. 2020. RobustPeriod: Time-Frequency Mining for Robust Multiple Periodicities Detection. *CoRR*, abs/2002.09535.

Wen, Q.; Sun, L.; Yang, F.; Song, X.; Gao, J.; Wang, X.; and Xu, H. 2021. Time Series Data Augmentation for Deep Learning: A Survey. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, IJCAI-2021, 4653–4660. International Joint Conferences on Artificial Intelligence Organization.

Yan, H.; and Ouyang, H. 2018. Financial Time Series Prediction Based on Deep Learning. *Wireless Personal Communications*, 102(2): 683–700.

Yoon, J.; Jarrett, D.; and van der Schaar, M. 2019. Time-series Generative Adversarial Networks. In Wallach, H.; Larochelle, H.; Beygelzimer, A.; d'Alché-Buc, F.; Fox, E.; and Garnett, R., eds., *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.

Zeng, Z.; Kaur, R.; Siddagangappa, S.; Rahimi, S.; Balch, T.; and Veloso, M. 2023. Financial Time Series Forecasting using CNN and Transformer. arXiv:2304.04912.

Ziółkowski, K.; and Wieczyński, P. 2025. Enhancing financial time series forecasting: a comparative study of discrete wavelet transform and LSTM models for selected global indices. *Quality & Quantity*.