# Multivariate Time Series Forecasting for ATM Cash Demand, an Adaptation to Anomalous Events

Tor Mattsson[1], Joel Ekstrand[2], Zahra Taghiyarrenani[3], Jens Lundström[4], Slawomir Nowaczyk[5] and Mikael Lindén[6]

[1]*Halmstad University, Sweden*

**Abstract**

This work investigates multivariate time series forecasting of ATM cash demand, with a focus on adapting to anomalous events such as wars, pandemics, or power outages. Using real-world transaction data from Bankomat AB, we explore modern forecasting models including Informer, FEDformer, TimesNet, TimeXer, and the foundation model MOMENT. We also explore embedding-based clustering to further improve forecasting results. Furthermore, we use fine-tuning and contrastive learning to adapt our models to real-world-based anomalous conditions. This work contributes toward reliable, interpretable, and adaptive ATM cash demand forecasting in both normal and disrupted conditions.

**Keywords**

Time Series Forecasting, Transformers, ATM Cash Demand, Anomaly Detection,

## 1. Introduction

This thesis was done in collaboration with Bankomat, the leading company in cash management through Automated Teller Machines (ATMs) in Sweden. One of Bankomats' day-to-day operations as a cash management company is to plan when to replenish ATMs. As of recently, the owners of Bankomat (the large Banks in Sweden) want Bankomat to take the role as an explicit knowledge center for cash.

Since the daily operations of Bankomat include managing ATM replenishment and general management of cash, one obvious question becomes:

> *"What will the cash demand for a single ATM, or a network of ATMs, be in the near or distant future? "*

This question is central to Bankomat's operations, as it determines how and when ATMs should be refilled. At the same time, it is also crucial from a knowledge perspective, helping to build a more general understanding of cash demand across different time horizons. It serves as the foundation of this study and defines the core forecasting challenge that the research aims to address.

Under normal circumstances, Bankomat is well-equipped to manage this challenge using its extensive domain expertise. However, external factors can significantly disrupt regular cash

*SAIS2025: Swedish AI Society Workshop 2025, 16-17 June 2025, Halmstad, Sweden.*
tormat20@student.hh.se (T. Mattsson); joeeks20@student.hh.se (J. Ekstrand); zahra.taghiyarrenani@hh.se (Z. Taghiyarrenani); jens.r.lundström@hh.se (J. Lundström); slawomir.nowaczyk@hh.se (S. Nowaczyk); mikael@mikaellindenconsulting.com (M. Lindén)

withdrawal behavior across the ATM network. Examples of such disruptions include pandemics and geopolitical crises, with the recent invasion of Ukraine serving as a prominent case.

This leads us to the research questions that guide this thesis:

1. **How can a high-performance forecasting model be developed for ATM cash demand under normal circumstances?** - To be able to create a forecasting model for anomalous events, we first create a model for normal circumstances.
2. **How can we adapt an ATM forecasting model to anomalous events with little data?** - Adapting the model to anomalous events.

Events as pandemics and wars highlight the need for forecasting models that are both reliable and adaptable. While recent advances in deep learning, particularly transformer-based models, have shown promising performance in time series forecasting, they are often regarded as black boxes. This lack of interpretability can make it difficult to understand or guide their behavior under unusual circumstances.

To address this, a promising direction explored in this report is the use of contrastive learning as an intermediate training step. Rather than relying solely on supervised learning or anomaly injection during fine-tuning, we introduce anomaly-augmented samples into the training process and leverage contrastive learning and fine-tuning to guide the model. This approach incentivizes the model to learn specific, meaningful representations in its embedding space by pulling together normal and augmented views of the same sequence while pushing apart unrelated samples.

By structuring the learning process in this way, the model is encouraged to develop embeddings that are not only robust to anomalies but also capable of distinguishing between normal and anomalous behavior. This helps ensure that the model does not merely memorize patterns from normal data but also generalizes to novel or irregular situations.

## 2. Related Works

Research on ATM cash demand forecasting has historically used statistical models and machine learning, beginning with the NN5 competition [1], where ensemble methods incorporating seasonality showed strong performance [2]. Later, Venkatesh et al. introduced clustering via weekday alignment [3] and chaotic modeling techniques [4], improving SMAPE scores. Fallahtafti et al. [5] analyzed ATM demand shifts during COVID-19, finding that simpler models (e.g., ARIMA) outperformed complex ML models during COVID-19.

More recent time-series forecasting approaches outside cash demand forecasting increasingly rely on transformers. Informer [6], FEDformer [7], TimesNet [8], and TimeXer [9] address sequence length and exogenous input challenges. MOMENT [10] and TS2Vec [11] create embeddings of time series that can be used either for forecasting, but they also provide a strong basis for modeling ATM withdrawal behaviour.

These forecasting models offer promising alternatives for our forecasting task. However, adapting models to anomalous events may benefit from approaches like contrastive learning. In TS2Vec [11], the authors introduce instance and temporal contrastive losses to guide the encoder in learning meaningful embeddings-aligning with the core idea of contrastive learning.

Building on this, [12] proposes a weighted contrastive learning framework that extends the losses in [13]. This can help the model capture anomaly-specific deviations while preserving the underlying structure in sequences.

## 3. Methodology

The methodology consists of several stages, each contributing to the development and evaluation of robust forecasting models for ATM cash demand under both normal and anomalous conditions. The full pipeline includes data exploration, clustering, model selection, anomaly synthesis, and training with and without contrastive learning. The following sections outline each component in detail.

### 3.1. Exploratory Data Analysis (EDA)

The first stage involved extensive exploratory data analysis (EDA) to understand the temporal dynamics and statistical properties of the ATM transaction data. This process revealed recurring patterns across daily, weekly, and monthly cycles, as well as seasonal variations. The insights gained from EDA informed the selection of input and output window sizes for the forecasting models.

### 3.2. Discovery of Data Subgroups and Clustering Motivation

During early model prototyping, a bug in the data loader revealed the existence of distinct geographical groups among the ATMs. Further inspection showed that while some groups exhibited highly unique behavior, many displayed similar withdrawal patterns. This observation, supported by findings from prior studies such as the NN5 forecasting competition, motivated the use of clustering to group ATMs with similar dynamics.

Rather than clustering based on geographical proximity, we opted for a data-driven approach. Specifically, we utilized embeddings generated by the MOMENT foundation model and performed clustering in this latent representation space using K-Means, with cosine similarity as the distance metric.

### 3.3. Benchmarking State-of-the-Art Forecasting Models

Before focusing on anomaly adaptation, we benchmarked several state-of-the-art forecasting models, including TimeXer, TimesNet, FedFormer, Informer, and MOMENT. Based on initial performance and time constraints, we selected TimesNet and TimeXer for further experimentation, as they showed the most promising results.

### 3.4. Anomaly Synthesis and Adaptation

To simulate the effects of rare but impactful events, we synthesized anomalies based on the cash demand shift observed during the Russian invasion of Ukraine. This was the only macro-level event in the dataset that showed a clear and lasting deviation from normal behavior.

We began by forecasting the expected cash demand during the invasion period using our trained models, then compared the predictions with the actual observed values. The resulting residuals were used to define a mathematical formulation of the anomaly.

From this formulation, we extracted key parameters describing the anomaly's magnitude, steepness, and duration. These parameters were then sampled from normal distributions to generate synthetic variants of the anomaly, which were injected into clean sequences. This allowed us to simulate a diverse range of hypothetical macro-level events and systematically evaluate model robustness and adaptability.

Two main training strategies were employed to adapt our forecasting models to anomalous conditions:

1. **Standard Training and Fine-Tuning:** Models were either trained from scratch or fine-tuned on datasets augmented with synthetic anomalies.
2. **Contrastive Learning Framework:** In this setup, the original dataset was duplicated. One copy remained unmodified, while the other received anomaly-injected samples. The model was then trained using a contrastive loss that encouraged the model to learn discriminative and robust representations in its embedding space. The objective was to pull together embeddings from clean and anomalous variants of the same sequence while pushing apart unrelated sequences.
   To achieve this, we experimented with two contrastive loss formulations inspired by TS2Vec: instance loss, which focuses on aligning whole-sequence embeddings, and temporal loss, which encourages alignment of corresponding time-step embeddings across the sequences. These losses were evaluated both individually and in combination to assess their impact on the model's ability to represent and adapt to anomalous behavior.

This contrastive approach aimed to guide the model toward a representation space where adaptation to anomalies is facilitated, while preserving the ability to perform well under normal conditions.

## 4. Experimental Results

During initial forecasting experiments, MOMENT and TimesNet got the best SMAPE scores, 29.30% and 28.73% respectively, as displayed in Appendix A, Table 2. Examples of predicted sequences for the initial experiment are displayed in Figures 2, 4, 3, 6, 5. However, since MOMENT is a foundation model, its input length is not configurable. Further evaluation was therefore done with TimesNet (best in performance) and TimeXer (third best in performance). These were evaluated with three different input/prediction lengths, the results are displayed in Appendix A, Table 3. Both TimesNet and TS2Vec showed promising results for both long and short-term forecasting.

To assess whether clustering improves forecasting accuracy, TimesNet and TimeXer were evaluated on five clusters generated using MOMENT. Performance was best in the largest clusters, with a top SMAPE of 21.72%, while smaller clusters performed significantly worse, suggesting data sparsity as a limiting factor. Results are shown in Appendix A, Table 4.

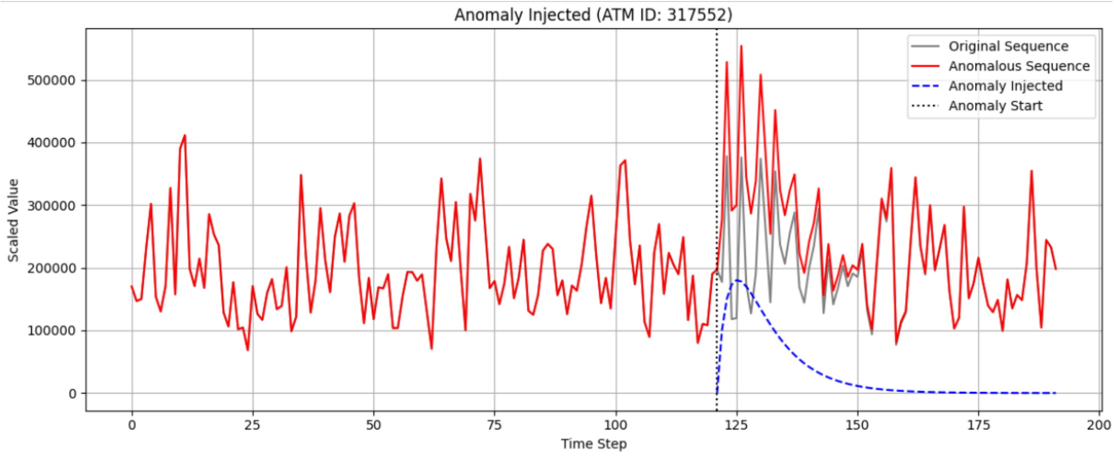


**Figure 1:** Example of an injected anomaly. This example sequence has an input length of 128 and an output length of 64. The anomaly starts at the end of the input window and continues in the output window. The red graph is the sequence with the injected anomaly (blue dashed graph), added to the grey graph (the original sequence).

After having developed the forecasting models for normal circumstances, we investigated how to adapt to anomalous circumstances.

First, to characterize the anomaly effect observed during the Russian invasion of Ukraine, we analyzed the residuals between the model's predictions and the actual cash demand during that period. A non-linear regression was then applied to fit a smooth curve to these residuals, resulting in the following parametric function used to generate synthetic anomalies:

$$AnomalyValue(n) = \frac{A * n \cdot e^{-B * n^C}}{90409} \tag{1}$$

Where AnomalyValue is the anomaly transaction amount for each time step (n) from the start of the anomalous period. A, B, and C are parameters that were sampled over a normal distribution, which was based on the A, B, and C parameter values during the invasion of Ukraine. The created anomalous sequence was added on top of the normal data. This created anomalous sequences with the same characteristics as the invasion of Ukraine, but with variation. (See figure 1)

Two main adaptation methods were used to adapt the models to the injected anomalies: fine-tuning and contrastive learning. These results are displayed in Table 1. Based on these results, a paired t-test was conducted on the SMAPE values obtained from the five evaluation runs using shared random seeds. The test revealed that contrastive learning significantly outperformed fine-tuning on normal data, with a mean SMAPE reduction and a statistically significant result ($t(4) = -25.45$, $p < 0.001$).

On anomalous data, the fine-tuning approach reached the lowest SMAPE score at 28.35%. However, the difference in SMAPE was not statistically significant ($t(4) = 1.93$, $p = 0.125$),

| Seed | Contrastive | | Fine-Tuning | |
|---|---|---|---|---|
| | Normal | Anomalous | Normal | Anomalous |
| 42 | 28.37 | 31.82 | 31.17 | 34.33 |
| 0 | 27.81 | 34.66 | 30.78 | 31.04 |
| 1 | 29.99 | 32.75 | 32.96 | 29.41 |
| 2 | 29.49 | 33.54 | 31.99 | 30.30 |
| 3 | 27.44 | 32.68 | 30.62 | 28.35 |

**Table 1**
SMAPE scores (%) for TimesNet with applied Contrastive Learning (instance loss only) and Fine-Tuning approaches across different seeds, under normal and anomalous conditions.

suggesting that the two models performed comparably under anomalous conditions across the evaluated seeds. In comparison, the TimesNet model trained on normal data only got an SMAPE score of 28.73% during normal circumstances and 37.91& when evaluated during anomalous circumstances.

In conclusion, while fine-tuning improved performance on anomalous data, it resulted in reduced accuracy under normal conditions. Contrastive learning, on the other hand, maintained strong performance during normal circumstances and achieved substantial improvements on anomalous data compared to models trained solely on normal data. We are currently conducting experiments with other contrastive learning approaches to further improve the adaptation resutls.

## Acknowledgments

We would like to express our sincere gratitude to our supervisors, Zahra Tahiyarrenani, Jens Lundstrom, and Slawomir Nowaczyk, for their continuous support, valuable guidance, and constructive feedback throughout this thesis project. We also wish to thank Bankomat AB for their support, guidance, and collaboration.

Furthermore, we acknowledge the inspiration drawn from recent advancements in time series modeling, particularly the Time Series Library [14], the Moment foundation model [15], and the transformer-based architectures—which significantly influenced the direction of this work.

## References

[1] N. Ribeiro, J. A. G. S. d. Moura, J. C. Oliveira, C. d. S. Santos, et al., The nn5 competition on forecasting of time series: An overview and results, International Journal of Forecasting 28 (2012) 397–412. doi:10.1016/j.ijforecast.2011.12.010.

[2] R. R. Andrawis, A. F. Atiya, H. El-Shishiny, Forecast combinations of computational intelligence and linear models for the nn5 time series forecasting competition, International Journal of Forecasting 27 (2011) 672–688. doi:10.1016/j.ijforecast.2010.09.005.

[3] K. Venkatesh, V. Ravi, A. Prinzie, D. Van den Poel, Cash demand forecasting in atms by clustering and neural networks, European Journal of Operational Research 232 (2014) 383–392. doi:10.1016/j.ejor.2013.07.027.

[4] V. Kamini, V. Ravi, D. N. Kumar, Chaotic time series analysis with neural networks to forecast cash demand in atms, in: 2014 IEEE International Conference on Computational Intelligence and Computing Research, 2014, pp. 1–5. doi:10.1109/CIC.2014.7019213.

[5] A. Fallahtafti, M. Aghaaminiha, S. Akbarghanadian, G. R. Weckman, Forecasting atm cash demand before and during the covid-19 pandemic using an extensive evaluation of statistical and machine learning models, SN Computer Science 3 (2022). URL: https://doi.org/10.1007/s42979-021-01000-0. doi:10.1007/s42979-021-01000-0.

[6] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, W. Zhang, Informer: Beyond efficient transformer for long sequence time-series forecasting, 2021. URL: https://arxiv.org/abs/2012.07436. arXiv:2012.07436.

[7] T. Zhou, Z. Ma, Q. Wen, X. Wang, L. Sun, R. Jin, Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting, 2022. URL: https://arxiv.org/abs/2201.12740. arXiv:2201.12740.

[8] H. Wu, T. Hu, Y. Liu, H. Zhou, J. Wang, M. Long, Timesnet: Temporal 2d-variation modeling for general time series analysis, arXiv preprint arXiv:2210.02186 (2022).

[9] Y. Wang, H. Wu, J. Dong, G. Qin, H. Zhang, Y. Liu, Y. Qiu, J. Wang, M. Long, Timexer: Empowering transformers for time series forecasting with exogenous variables, Advances in Neural Information Processing Systems 37 (2025) 469–498.

[10] M. Goswami, K. Szafer, A. Choudhry, Y. Cai, S. Li, A. Dubrawski, Moment: A family of open time-series foundation models, 2024. URL: https://arxiv.org/abs/2402.03885. arXiv:2402.03885.

[11] Z. Yue, Y. Wang, J. Duan, T. Yang, C. Huang, Y. Tong, B. Xu, TS2Vec: Towards universal representation of time series, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 36, 2022, pp. 8980–8987. URL: https://arxiv.org/abs/2106.10466.

[12] Z. Zhang, Z. Cao, Y. Zhang, B. Zong, H. Chen, W. Wang, Soft contrastive learning for unsupervised time-series anomaly detection, arXiv preprint arXiv:2305.14664 (2023).

[13] J.-Y. Franceschi, A. Dieuleveut, M. Jaggi, Unsupervised scalable representation learning for multivariate time series, in: NeurIPS, 2019.

[14] Y. Wang, H. Wu, J. Dong, Y. Liu, M. Long, J. Wang, Deep time series models: A comprehensive survey and benchmark (2024).

[15] M. Goswami, K. Szafer, A. Choudhry, Y. Cai, S. Li, A. Dubrawski, MOMENT: A family of open time-series foundation models, in: Proceedings of the 41st International Conference on Machine Learning, 2024, pp. 16115–16152. URL: https://arxiv.org/abs/2402.03885.

| Model | SMAPE (%) |
|---|---|
| ARIMA | 68.03 |
| LSTM | 80,30 |
| Informer | 31.51 |
| TimesNet | 28.73 |
| FEDformer | 31.58 |
| TimeXer | 29.99 |
| Moment | 29.30 |

**Table 2**
Forecasting results for daily aggregated data during daily operations with a multivariate time series. Trained and tested on transaction data from 2023 and 2024.

| Input len/Pred len | TimesNet | | TimeXer | |
|---|---|---|---|---|
| | SMAPE (%) | ME (%) | SMAPE (%) | ME(%) |
| 365/14 | 26.86 | 8.98 | 27.59 | 9.92 |
| 128/64 | 28.73 | 7.99 | 29.99 | 8.29 |
| 60/30 | 27.51 | 7.88 | 29.28 | 8.61 |

**Table 3**
Forecasting results for TimesNet and TimeXer with different input/prediction lengths using daily aggregated multivariate ATM transaction data. Models were trained and tested on 2023-2024 data.

| Cluster | TimesNet | | TimeXer | |
|---|---|---|---|---|
| | SMAPE (%) | ME (%) | SMAPE (%) | ME(%) |
| 0 | 21.72 | 7.92 | 21.83 | 8.03 |
| 1 | 28.86 | 8.92 | 29.09 | 9.27 |
| 2 | 46.63 | 15.45 | 44.37 | 16.64 |
| 3 | 42.13 | 15.56 | 42.05 | 13.61 |
| 4 | 28.55 | 10.25 | 28.41 | 9.83 |
| All data | 26.86 | 8.98 | 27.59 | 9.92 |

**Table 4**
Forecasting results for TimesNet and TimeXer for clustered data using multivariate daily aggregated ATM transaction data. Cluster 0 has 297 ATMs, cluster 1 has 333 ATMs, cluster 2 has 57 ATMs, cluster 3 has 75 ATMs, and cluster 4 has 492 ATMs. The models were trained and tested on 2023-2024 data.

# A. Appendix

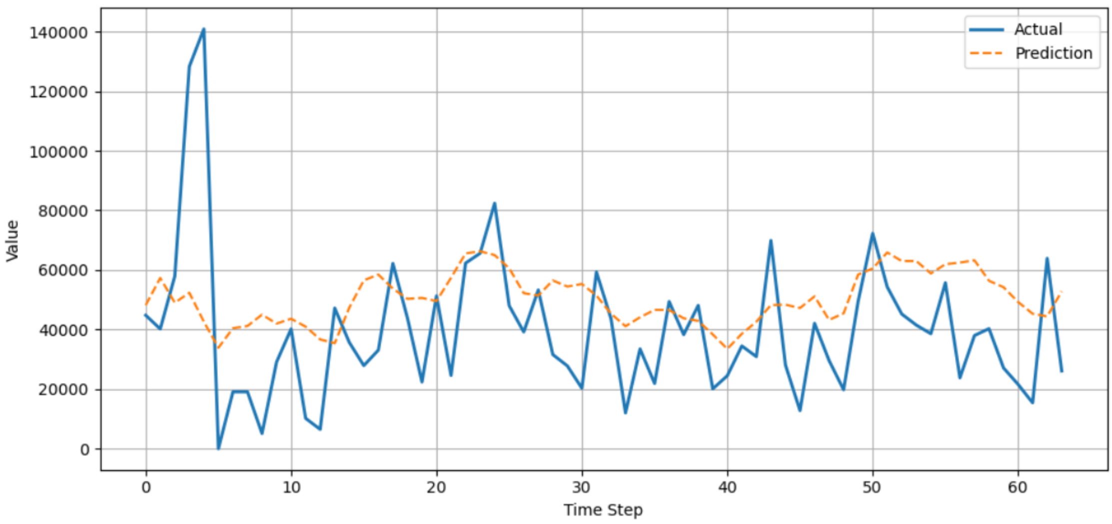


**Figure 2:** Forecasting result for one sequence with an LSTM model using 128 in sequence length and 64 in prediction length. The x-axis represents each predicted time-step, and the y-axis represents the aggregated transaction amount for one ATM over one day.

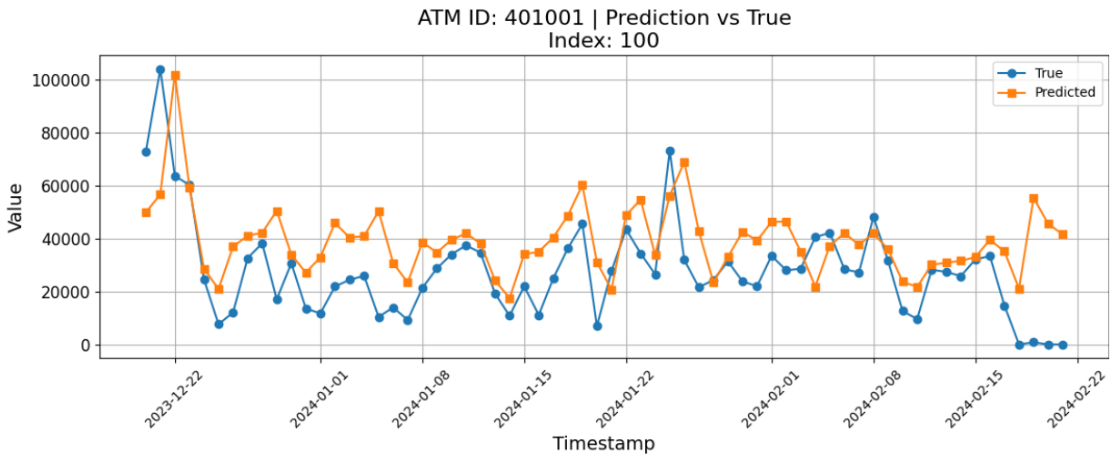


**Figure 3:** Forecasting result for one sequence with Informer with 128 in sequence length and 64 in prediction length. The x-axis represents each predicted time-step, the y-axis has not yet been transformed back to the original scale.

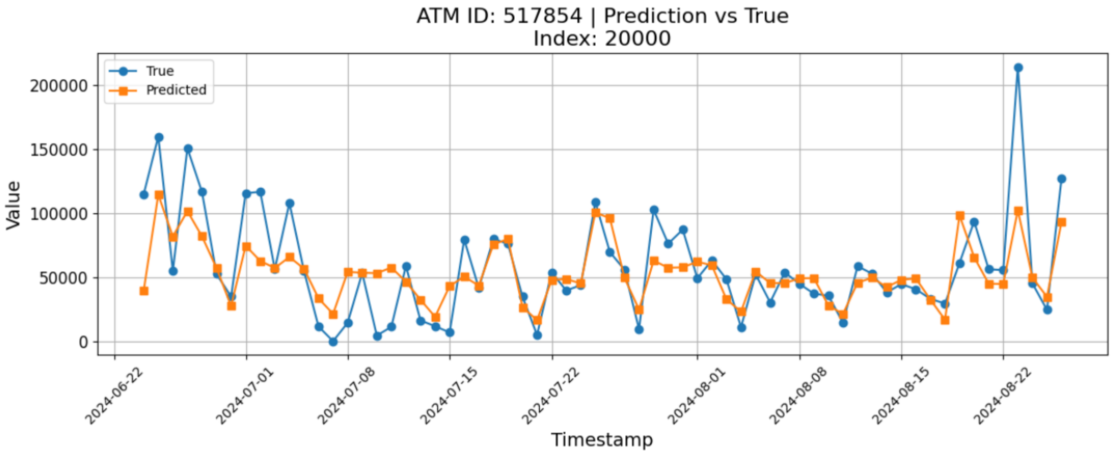


**Figure 4:** Forecasting result for one sequence with FEDformer with 128 in sequence length and 64 in prediction length. The x-axis represents each predicted time-step, the y-axis has not yet been transformed back to the original scale.

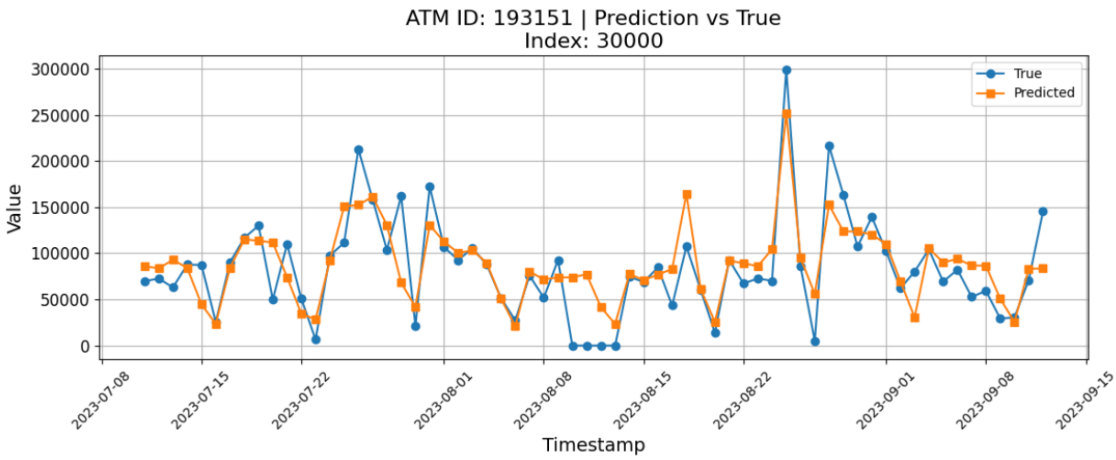


**Figure 5:** Forecasting result for one sequence with TimesNet with 128 in sequence length and 64 in prediction length. The x-axis represents each predicted time-step, the y-axis has not yet been transformed back to the original scale.

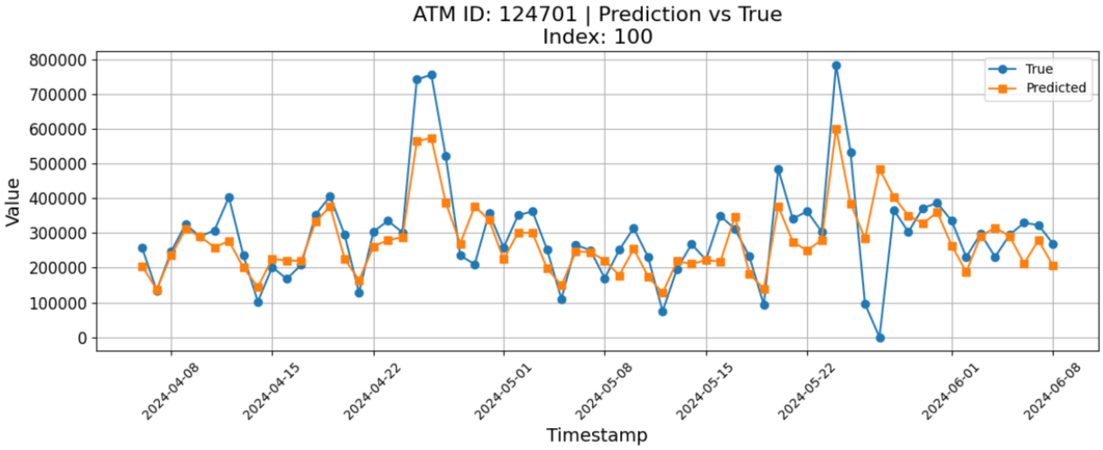


**Figure 6:** Forecasting result for one sequence with TimeXer with 128 in sequence length and 64 in prediction length. The x-axis represents each predicted time-step, the y-axis has not yet been transformed back to the original scale.