MET-C3: MIXED EXOGENOUS–TEMPORAL DEPENDENCY MODELING FOR NON-STATIONARY TIME SERIES FORECASTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Nonstationary time series forecasting is challenging, as the dependency structure among historical observations, temporal lags, and exogenous variables varies over time. Existing models capture temporal patterns implicitly, but often lack an explicit mechanism for distinguishing temporal inertia from exogenous impacts. This paper proposes MET-C3, a mixed node dependency graph framework for graph guided nonstationary forecasting. MET-C3 represents lagged target variables and exogenous variables as a unified node set, estimates dependency matrices in rolling windows, and introduces autocorrelation constrained coupling to weight dependencies among lag nodes, between lag and exogenous nodes, and among exogenous nodes. It further applies TMFG sparse topology construction and LoGo partial correlation inference to extract compact structural features, which are integrated with multiple forecasting backbones. Experiments on a proprietary Shandong Electricity Market dataset and three public datasets from electricity, mobility, and finance show that MET-C3 consistently reduces forecasting error across backbones and domains. Robustness analyses demonstrate its advantages under stronger structural nonstationarity, while runtime comparisons show comparable computational cost to the original backbones. These results indicate that explicit mixed dependency modeling offers an effective and interpretable approach to nonstationary forecasting.

Keywords: Nonstationary time series forecasting, mixed node dependency graph, exogenous impact modeling, sparse graphical inference, graph guided forecasting.

1 INTRODUCTION

Multivariate time series forecasting is a basic task in data mining and machine learning. It is used in electricity markets, power systems, financial markets, and urban mobility (Wang et al., 2024a). In many real world systems, future observations do not follow a fixed dependency structure. The value of different information sources changes over time because the system may face nonstationarity, distribution shifts, and changing operating conditions (Liu et al., 2022). Electricity price forecasting is a clear example. Prices depend on regular patterns in past market data and on exogenous variables whose patterns also change. Renewable generation may vary with the season and weather, while directly dispatched load may differ across operating cycles. Past observations may provide strong signals during stable periods. Exogenous variables may become more useful when operating conditions change (Wang et al., 2024b).

Recent forecasting studies have improved the modeling of temporal patterns and multivariate dependencies. Models such as DLinear (Zeng et al., 2023) and PatchTST (Nie et al., 2023) show that useful temporal representations can improve long horizon forecasting. Architectures based on TCN (Bai et al., 2018) remain useful for learning local temporal patterns. Graph based methods such as MTGNN (Wu et al., 2020) extend forecasting models by learning dependencies among variables. These studies show that temporal representations and variable dependencies both matter. However, most existing methods treat lagged target values and exogenous variables as common input channels. Other methods infer their relations through attention weights or adaptive adjacency matrices. These designs can learn useful associations, but they often hide how the associations change over

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

time (Qiu et al., 2025). A relation that helps in one period may weaken or disappear in another, especially after the system enters a new operating state. This creates an open representation problem. Temporal inertia and exogenous forcing both affect prediction, but their changing relations with the target are not organized clearly.

To address this problem, we use a mixed exogenous and temporal view of dependency modeling for nonstationary time series forecasting. First, lagged target states and exogenous variables should share one structural space so that their changing relations with the target can be compared. This view organizes variables by their roles in forecasting instead of treating them as common input channels. Lagged target states represent temporal inertia. Exogenous variables represent exogenous forcing. These two information sources differ in timing, meaning, and statistical role.

Second, the structural space should change over time because the target may relate to different lagged states and exogenous variables in different regimes. A distant lag or an exogenous variable may be useful in one local regime. The same relation may weaken or become misleading after the system changes. A suitable representation should make temporal inertia and exogenous forcing comparable in structure. It should also preserve their different meanings and avoid treating unstable local associations as reliable relations.

To meet these requirements, we propose Mixed Exogenous and Temporal Dependency Modeling with Correlation Constrained Coupling, called MET-C3. MET-C3 first builds a shared dependency space for the current target, lagged target states at several scales, and exogenous variables at the current time. It then estimates dependencies in rolling windows to track changing relations with the target. A coupling constraint based on the measured autocorrelation decay of the target weakens less reliable long range temporal relations before graph construction. MET-C3 then uses the Triangulated Maximally Filtered Graph, or TMFG, and local and global partial correlation inference, or LoGo, to obtain a sparse structure. This structure is converted into graph guided forecasting features and structural indicators of lag dependence, exogenous influence, and topology changes.

We evaluate MET-C3 on a high frequency Provincial Electricity Market dataset and three public datasets from power systems, urban mobility, and crypto markets. The experiments use several forecasting baselines and fixed prediction horizons. They examine forecasting accuracy, the contribution of each component, generalization across models and datasets, and structural interpretation. We also study whether the learned indicators reflect changes in exogenous influence and graph structure. Our main contributions are as follows:

- We present MET-C3, a graph based framework that organizes mixed temporal and exogenous dependencies in nonstationary forecasting. It provides a clear structural interface for studying how temporal inertia and exogenous forcing affect target predictions.
- We identify changing target dependence as a central issue in forecasting with exogenous variables. The learned topology helps the framework separate persistent temporal effects from changes driven by external variables. It also supports the interpretation of different regimes.
- We provide empirical evidence that mixed exogenous and temporal dependency modeling can improve forecasting and provide useful structural measures. The analysis also explains the role of the main components and the computational cost of the framework.

The paper is organized as follows. Section 2 reviews related work on nonstationary time series forecasting, graph based forecasting, and sparse dependency inference. The Methodology section defines the mixed exogenous and temporal forecasting problem and presents MET-C3. It covers autocorrelation constrained coupling, TMFG and LoGo structural inference, and graph guided feature construction. The next two sections report the structural analysis and forecasting results. They include component ablation, generalization across datasets, robustness tests, transition window analysis, and computational cost. The final section concludes the paper.

2 Related work

We review studies on temporal and exogenous modeling, graph based dependency modeling, and sparse structural inference for time series forecasting. We describe their main methods and explain how they differ from our focus on changing mixed dependencies.

2.1 Temporal and Exogenous Modeling in Time Series Forecasting

Recent studies have focused on improving the representation of past information (Benidis et al., 2022). Transformer based models and their variants (Zhou et al., 2021; Liu et al., 2024; Gao et al., 2024; Woo et al., 2024) use attention to learn long range temporal dependencies. Patch based methods (Zeng et al., 2023; Wang et al., 2024b) use long input windows more effectively by grouping nearby observations into local segments. Frequency based methods (Yi et al., 2023; Wang et al., 2025) learn periodic patterns from spectral components. Decomposition based methods (Challu et al., 2023) separate past sequences into trend, seasonal, and residual parts. These studies improve the encoding of past observations. They also raise the question of whether increasingly complex temporal models are always needed.

Simple architectures have therefore become an important topic in forecasting research. Linear models such as DLinear and NLinear (Zeng et al., 2023) show that strong performance does not always require complex attention models. Light frequency domain and hierarchical interpolation models such as FITS (Xu et al., 2024) and N-HiTS (Challu et al., 2023) also remain competitive in long horizon forecasting. These findings have led researchers to reconsider the relation among model complexity, the stability of temporal patterns, and forecasting accuracy. However, these methods mainly focus on representations of past data and do not clearly organize the role of exogenous information.

Recent research has also studied how to add exogenous variables to forecasting models. In electricity markets and power systems, variables such as load and renewable generation provide information beyond past target values (Olivares et al., 2023). Most studies add exogenous variables as input channels or supporting embeddings. Some methods use interaction modules such as attention to combine external variables with the target history (Lim et al., 2021). These studies show that exogenous information can improve forecasts when changing external conditions affect the target (Lago et al., 2021).

Although these studies show the value of past observations and exogenous information, they usually combine both sources in the same predictive representation. Research on nonstationary forecasting has considered distribution shifts and regime changes. However, it remains difficult to understand how the target shifts between temporal inertia and exogenous forcing. More study is needed to determine when past states drive prediction and when external variables have a stronger effect.

2.2 Graph Based Dependency Modeling for Multivariate Forecasting

Graph based methods are widely used to model dependencies among variables in multivariate time series forecasting (Jin et al., 2024b;a). In spatial temporal forecasting, predefined graphs often describe physical or spatial relations (Jiang et al., 2023a;b). Graph WaveNet (Wu et al., 2019) combines graph convolution with dilated temporal convolution to model spatial temporal dependencies. D2STGNN (Shao et al., 2022c) separates dynamic spatial temporal patterns for traffic forecasting. STEP (Shao et al., 2022b) adds pretraining to spatial temporal graph forecasting to improve representation learning. These studies show that clear dependency structures can improve forecasts when the relations among variables are meaningful and stable (Ma et al., 2024).

However, predefined graphs are hard to specify when variable relations are hidden or change over time. Researchers have developed adaptive graph learning methods to address this problem. Some methods learn graph structures from node representations or adaptive adjacency matrices. For example, AGCRN (Bai et al., 2020) learns adaptive node parameters and hidden dependencies from data. MTGNN (Wu et al., 2020) uses a graph learning module to find relations among variables and combines the learned structure with temporal convolution. Other methods treat graph construction as a structure learning problem. GTS (Shang et al., 2021), for example, treats graph structure learning as a discrete optimization problem. These methods reduce the need for known graph structures and support data driven dependency modeling.

Recent studies have also considered dynamic and different types of dependencies. DGCRN (Li et al., 2023) models changing graph structures through recurrent graph learning. DFDGCN (Li et al., 2024) studies dynamic graph convolution in the frequency domain. Other methods such as STID (Shao et al., 2022a) ask whether graph convolution is always needed and explore simpler ways to handle spatial or cross variable effects. These works show that dependency modeling requires more

than adding a graph neural network. It also requires a clear view of when and how variable relations should be represented.

Despite this progress, most graph based forecasting methods define nodes as observed variables or individual time series. Their learned edges usually describe relations among variables at the same meaning level. This design works well for sensors, regions, or observed series. However, lagged target states and exogenous variables differ in timing, meaning, and statistical role. Existing graph based forecasting methods are therefore less suited to mixed exogenous and temporal dependency modeling.

2.3 Sparse Dependency Inference and Structural Interpretation

Sparse dependency inference has been widely studied in graphical modeling and structure learning. Early sparse graph models such as graphical lasso (Friedman et al., 2008) estimate sparse inverse covariance matrices to find conditional dependencies among variables. Newer structure learning methods infer directed relations from data. NOTEARS treats DAG structure learning as a continuous optimization problem. DAGMA (Bello et al., 2022) supports differentiable DAG learning through a log determinant measure of cycles. DECI (Geffner et al., 2024) combines causal discovery and causal inference with a flow based nonlinear additive noise model. CUTS (Cheng et al., 2023) and CUTS+ (Cheng et al., 2024) extend neural causal discovery to irregular and high dimensional time series.

Network filtering provides another way to build sparse dependency structures. MST (Mantegna, 1999) and PMFG (Tumminello et al., 2005) keep important edges from dense similarity matrices under different topology rules. TMFG (Massara et al., 2015) builds a sparse triangulated topology while keeping strong dependency information. LoGo (Barfuss et al., 2016; Liu et al., 2025) then estimates local and global partial correlations by computing inverse covariance matrices on small subgraphs.

Recent time series studies have also explored structural interpretation. Autoformer (Wu et al., 2021) uses series decomposition to separate trend and seasonal parts. FEDformer (Zhou et al., 2022) models the frequency domain to learn important temporal patterns. TimesNet (Wu et al., 2023) represents temporal changes through two dimensional forms and offers another way to analyze periodic structures. In graph based forecasting, researchers sometimes inspect learned adjacency matrices or hidden graph structures to study variable relations.

Overall, existing studies improve interpretation through sparse structural modeling and clear forecasting representations. However, sparse inference methods are often used for static structural analysis. Forecasting models with clear explanations usually describe learned temporal or variable representations. Few studies connect sparse topology inference with forecasting feature construction and changing structural indicators. The use of sparse structural inference for nonstationary forecasting therefore needs more study.

3 Methodology

This section presents MET-C3 for mixed exogenous and temporal dependency modeling. We first define the forecasting problem and the mixed node set, and then introduce the main modules for rolling window dependency estimation, autocorrelation constrained coupling, sparse structural inference, graph guided feature generation, and forecasting integration. The overall workflow is illustrated in Fig. 1.

3.1 Problem Definition

The task predicts y_{t+h} from the available target history and exogenous variables. MET-C3 represents the current target, selected lagged target values, and current exogenous variables as one mixed node set. It learns a sparse dependency graph and graph guided features over this set. Equations (1)-(4) in Appendix B give the full notation.

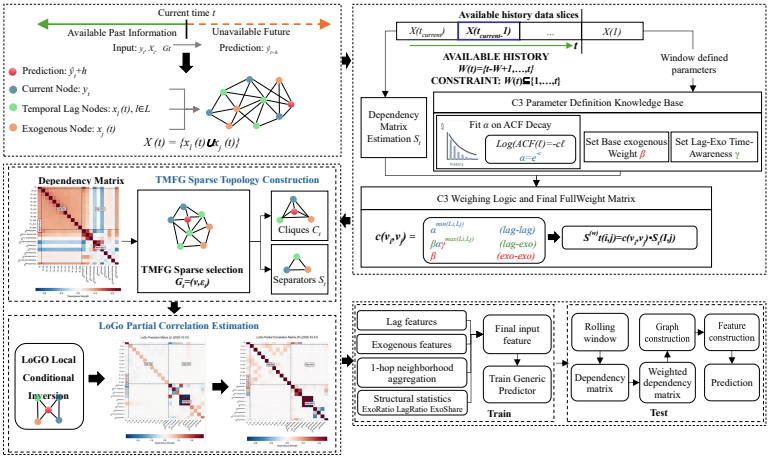

Figure 1: Overall framework of the proposed MET-C3 method.

3.2 Rolling Window Dependency Estimation

MET-C3 forms the mixed node matrix in each rolling window and estimates its local dependency matrix. Pearson correlation is used by default, while Spearman correlation and distance correlation are also supported. Full definitions are given in Equations (5)-(8) of Appendix B.

3.3 Autocorrelation Constrained Coupling

MET-C3 estimates an autocorrelation decay factor α from the target series. It uses α to reduce less reliable long range temporal relations, while β controls the base weight of exogenous edges and γ controls temporal and exogenous coupling. The constrained dependency matrix is the elementwise product of the local dependency matrix and this structural weight matrix. Equations (9)-(16) in Appendix B give the autocorrelation fitting procedure and the full coupling equations.

3.4 Sparse Topology Construction

MET-C3 uses TMFG to retain a sparse set of strong edges. It starts from the four nodes with the largest strengths and inserts each remaining node into the face with the largest total connection weight. The full construction is given in Equations (17)-(23) of Appendix B.

3.5 Partial Correlation Inference

MET-C3 combines the inverse matrices of TMFG cliques and separators to obtain a sparse inverse dependency matrix. It then converts this matrix into partial correlation weights. Equations (24)-(29) in Appendix B give the full equations.

3.6 Graph Guided Feature Generation

MET-C3 uses the partial correlations linked to the target to form exogenous impact, lag feedback, graph complexity, and graph hubness features. The resulting four dimensional vector is passed to the forecasting model. Equations (30)-(37) in Appendix B give all definitions.

3.7 Structural Indicators

The structural indicators measure lag dependence, exogenous influence, their relative balance, and changes in the target neighborhood. Their full definitions are provided in Equations (38)-(44) of Appendix B.

3.8 Forecasting Integration

The graph guided feature vector is joined with the mixed node vector and passed to the selected forecasting backbone. Training minimizes the mean absolute prediction error. Equations (45)-(47) in Appendix B give the full equations.

4 Structural Analysis of Mixed Temporal-Exogenous Dependencies

This section studies how mixed temporal and exogenous dependencies change across analysis windows. We analyze the relative dependence of different nodes and changes in the target neighborhood.

4.1 Dataset and Structural Analysis Setting

This section uses a proprietary dataset collected from the Shandong electricity market as the main dataset, which contains a target series and multiple exogenous variables that reflect changing system conditions. Variations in these variables can be reflected in the target series, which provides a practical basis for analyzing whether temporal and external dependencies remain stable over time.

For structural analysis, we construct mixed dependency graphs over rolling analysis windows. The resulting graph structures are used to examine the balance between temporal and exogenous dependence, as well as the reorganization of the target neighborhood over time.

4.2 Dynamic Shifts between Temporal and Exogenous Dependencies

We first examine how the target depends on temporal lag nodes and exogenous nodes. We use two views: the composition of the target neighborhood and dependency strength based on partial correlation. The first view shows the types of nodes in the local target neighborhood. The second measures the strength of associations related to the target after sparse partial correlation inference. Figure 2(a) (Appendix C) shows the composition of the target neighborhood across analysis windows. Its size and its mix of temporal and exogenous nodes both change over time. Lag neighbors are more common in some windows, while exogenous neighbors are more common in others. The main source of local target dependence therefore shifts between temporal inertia and exogenous forcing.

Figure 2(b) (Appendix C) shows that dependency strength around the target also changes. Lag dependence and exogenous influence often rise or fall together, but their relative roles are not fixed. The exogenous ratio rises when external variables account for a larger share of dependency strength related to the target. It falls when lag dependence becomes stronger. The shift therefore appears in both neighborhood selection and dependency strength.

Together, these findings show a changing mixed dependency pattern around the target. A dynamic structural representation is therefore needed to capture the changing roles of temporal inertia and exogenous forcing.

4.3 Target-Neighborhood Replacement and Stability

We also examine whether adjacent analysis windows keep or replace the local neighbors of the target. Figure 3 (Appendix C) shows that the change ratio S_t^{chg} remains moderate in some transitions. This pattern means that part of the target neighborhood is preserved. In other transitions, S_t^{chg} rises above the mean change baseline, which means that more local neighbors are replaced. The target therefore changes both dependency strength and local topology over time.

Table 1 (Appendix C) reports bootstrap stability for six representative windows. It tests whether the observed target neighbor relations are reliable within each window. The number of stable target neighbors ranges from 10 to 15, and the mean stable probability ranges from 0.883 to 0.950. These results show that random sampling variation does not control the local relations centered on the target. However, the number and type of stable neighbors differ across windows, so the stable dependency topology changes over time. Overall, the dependency topology centered on the target is not fixed across analysis windows. Stable local relations appear within individual windows, but nonstationary conditions reorganize the target neighborhood.

5 Forecasting Performance and Efficiency Evaluation

This section evaluates the forecasting effectiveness and computational efficiency of MET-C3. Experiments are conducted on the Shandong electricity market dataset and three public datasets from electricity, mobility, and financial domains. We compare the original forecasting backbones with their MET-C3 enhanced versions, and further examine component contributions, robustness under repeated runs, transition period performance, and runtime overhead.

5.1 Dataset and Forecasting Setting

The forecasting experiments are conducted on four datasets: Shandong Electricity Market, ENTSO-E Spain, UCI Bike Sharing Dataset, and Binance BTC/USDT. The Shandong dataset is used as the main dataset. The three public datasets cover electricity, mobility, and finance, and provide target series with exogenous variables aligned in time. Table 2 (Appendix C) summarizes the dataset statistics and forecasting settings.

The forecasting horizon is five days for Shandong Electricity Market, ENTSO-E Spain, and Binance BTC/USDT. For UCI Bike Sharing Dataset, we use a three day horizon because the dataset is shorter and shows strong daily demand cycles.

All datasets are split chronologically for model training and testing to avoid future information leakage. The evaluated forecasting backbones include LightGBM, DLinear, LSTM, TCN, MTGNN, and PatchTST. Forecasting accuracy is evaluated by MAE and RMSE. Their standard definitions are given in Appendix B. Execution time is recorded for each experimental run. All experiments are conducted on a high performance computing platform with one NVIDIA GeForce RTX 4090 GPU with 24 GB memory. The software environment includes NVIDIA driver 535.146.02, PyTorch 2.6.0, and CUDA 12.4.

5.2 Component Ablation and Mechanism Validation on the Main Dataset

The Shandong Electricity Market dataset is the main dataset because it represents the primary real world forecasting setting. We use it to examine whether the learned mixed dependencies improve forecasting accuracy. LightGBM is the ablation backbone because its errors clearly show the effects of changes in the graph guided features. Table 3 (Appendix C) reports the ablation results of the MET-C3 variants on the main dataset. These variants are designed to separate the effects of graph guided information, single source dependency modeling, mixed dependency modeling, and the complete MET-C3 structure. The comparison is used to examine the contribution of each structural component and the additional improvement provided by the complete design over partial variants.

Table 3 also indicates that graph guided information generally improves forecasting accuracy, as most variants have a lower MAE than M0. The temporal variants provide limited and unstable gains, while the exogenous variant achieves a clearer reduction in both MAE and RMSE. In particular, M3 shows a slight negative improvement, suggesting that temporal decay alone may limit useful local temporal associations when exogenous information is not incorporated. The mixed variants further improve performance over most single source variants, indicating that temporal lag nodes and exogenous nodes provide useful information together. M6 obtains the lowest MAE of 83.2810, which is 2.04% lower than M0. Although M5 achieves the lowest RMSE, M6 still reduces RMSE compared with M0. This result suggests that the full MET-C3 design mainly improves average forecasting accuracy. Figure 4 (Appendix C) further compares the weekly MAE of different variants

during the test period. The weekly results show that the improvement of M6 is not driven by a single isolated test segment. Instead, M6 maintains lower or competitive MAE values across multiple weeks, especially when forecasting performance fluctuates more strongly. This indicates that the learned mixed dependency features provide more adaptive predictive information than static or single source dependency variants.

5.3 Robustness, Sensitivity, and Transition Analysis on the Main Dataset

This part evaluates the stability of MET-C3 on the main dataset. After comparing the components, we study repeated run robustness, performance during market transitions, and sensitivity to the exogenous weight. Table 4 (Appendix C) summarizes the robustness results across forecasting backbones. For all five backbones, M6 achieves a lower mean MAE than M0, with mean MAE improvements ranging from 2.92% to 5.34%. PatchTST obtains the largest mean improvement, while DLinear still achieves a 2.92% reduction despite its simpler linear architecture. The standard deviations show that the gains are relatively stable for most backbones, whereas MTGNN exhibits larger variability, implying a stronger dependence on the integration between graph guided signals and the graph neural forecasting architecture. The successful runs are fixed at 30 for each backbone, but the required attempts vary from 40 to 73, leading to success ratios from 41.1% to 75.0%. This indicates that the learned structural information is generally useful, while stable integration still depends on the interface between graph guided features and the forecasting architecture.

Figure 5 (Appendix C) compares robustness across the backbones. The results agree with Table 4. The gap between M0 and M6 is clear for LSTM, PatchTST, and MTGNN, which shows that several neural forecasting models can use the structural features. The gap is smaller for DLinear and TCN, in line with their lower success ratios. Robustness appears in both the mean MAE reduction and the pattern of variation across repeated runs. Figure 6 (Appendix C) examines robustness during market transitions. Shaded regions mark periods in which M6 has a lower MAE than M0. Black crosses mark regime shift failure points where the MAE of M0 is more than 1.2 standard deviations above its global mean. At many of these points, M6 better limits the increase in error across the five backbones, although it does not beat M0 in every case. MET-C3 can therefore reduce forecasting loss during sudden market transitions, but its benefit varies across backbones and transition stages. Figure 7 (Appendix C) analyzes the effect of the exogenous weight factor β. The MAE curve reaches its lowest value at $\beta = 1.0$, where M6 achieves an MAE of 83.281. When β is smaller or larger than this value, the forecasting error increases, indicating that either underweighting or overweighting exogenous information may weaken the forecasting effect.

Taken together, these results show that MET-C3 provides consistent gains across repeated runs and retains predictive advantages in multiple transition periods. They also indicate that its realized effect depends on both backbone compatibility and the weighting balance between temporal and exogenous dependencies.

5.4 Cross-dataset Generalization and Backbone Transferability

To evaluate the generalization of MET-C3, we select public datasets that include target series and exogenous variables aligned in time. This is necessary because MET-C3 studies the relation between temporal inertia and external effects. We do not use common benchmarks such as ETTh1 and ETTm1 because their variables are mainly homogeneous sensor measurements. They do not clearly separate external driving factors from temporal dependence. The selected public datasets allow us to test the mixed dependency design in different domains. Figure 8 (Appendix C) compares structural changes in the selected public datasets. For ENTSO-E Spain, the exogenous influence ratio changes often, with several local increases in structural reorganization. For UCI Bike Sharing Dataset, the exogenous influence ratio stays within a smaller range, while the reorganization becomes stronger later. For Binance BTC/USDT, both curves show more irregular changes and a less stable dependency regime. Structural nonstationarity is therefore present in every selected domain, but its pattern and strength differ greatly.

Table 5 (Appendix C) reports the robustness results on these datasets across six forecasting backbones. Overall, M6 achieves a lower mean MAE than M0 in all 18 dataset and backbone combinations, indicating that the proposed graph guided design is not restricted to the main electricity market dataset. The improvements are relatively moderate on ENTSO-E Spain and UCI Bike Shar-

ing, ranging from 0.88% to 4.07% and from 0.59% to 5.77%, respectively. In contrast, the gains on Binance BTC/USDT are more pronounced, ranging from 2.17% to 32.89%. The larger standard deviations on several Binance BTC/USDT backbones further indicate that stronger nonstationarity brings higher gains but also greater variability. This suggests that MET-C3 can provide stronger benefits in highly nonstationary environments, where historical temporal patterns are more likely to become unstable. However, the success ratios also show that such gains are not equally easy to obtain across all settings, indicating that a suitable adaptation interface is still needed to effectively connect the learned structural information with different forecasting backbones.

The results across datasets further show that the effectiveness of MET-C3 depends on both the data domain and the forecasting architecture. DLinear obtains smaller but consistent gains, while nonlinear backbones such as LSTM, PatchTST, and MTGNN achieve larger improvements on Binance BTC/USDT. This indicates that models use graph guided structural information with different levels of success. Since MET-C3 learns the structural prior from the mixed dependency graph instead of using a manually specified prior, it also gives a clear view of the temporal and exogenous dependencies used by the forecasting model.

These results confirm that MET-C3 works across domains and is not limited to one dataset. Its gains are clearer when structural nonstationarity is stronger. However, stable use of the learned structural prior still depends on how well it is connected to the forecasting backbone.

5.5 Computational Efficiency and Scalability

Computational efficiency is important for practical nonstationary forecasting. We therefore measure the cost of adding structural modeling and graph guided features to standard forecasting backbones. Figure 9 (Appendix C) shows that M6 uses 1.14$\times$ the runtime of M0 with LightGBM on the main dataset. For the other five backbones, the ratios range from 0.98$\times$ to 1.07$\times$. Figure 10 (Appendix C) shows a similar pattern on the three public datasets. Most runtime ratios remain close to 1.0, and several settings have equal or lower runtime for M6. Table 6 (Appendix C) confirms that M6 remains close to M0 in absolute runtime for most dataset and backbone combinations. longer runtime mainly come from neural backbones such as LSTM and TCN. The graph guided features add little cost. Thus, the forecasting architecture determines most of the runtime, and MET-C3 improves dependency awareness without a large added cost.

6 Conclusion

This paper studied nonstationary time series forecasting through mixed temporal and exogenous dependencies. The analysis shows that the dependency topology around the target changes across windows. The roles of lag nodes and exogenous nodes change, and the target neighborhood is reorganized. Thus, nonstationarity appears not only in data distributions and prediction errors, but also in the changing structure between temporal inertia and exogenous forcing.

MET-C3 addresses this issue by placing lagged target states and exogenous variables in one mixed node dependency graph. It combines rolling window estimation, autocorrelation constrained coupling, TMFG topology construction, and LoGo partial correlation inference. Tests on the main dataset and three public datasets show better forecasts across datasets and backbones. The ablation, repeated run, and transition results confirm the value and stability of mixed dependency modeling. Runtime comparisons also show that the additional graph derived features do not substantially increase the computational scale of the original forecasting backbones.

Future work can develop a more flexible link between structural information and different backbones. Online graph updates may reduce repeated estimation and support real time forecasting. MET-C3 can also be tested in other nonstationary systems, including industrial processes, health monitoring, and supply chain demand.

REFERENCES

Lei Bai, Lina Yao, Can Li, Xianzhi Wang, and Can Wang. Adaptive graph convolutional recurrent network for traffic forecasting. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20, Red Hook, NY, USA, 2020. Curran Associates Inc. ISBN 9781713829546.

Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling, 2018. URL https://arxiv.org/abs/1803.01271.

Wolfram Barfuss, Guido Previde Massara, T. Di Matteo, and Tomaso Aste. Parsimonious modeling with information filtering networks. *Physical Review E*, 94(6), December 2016. ISSN 2470-0053. doi: 10.1103/physreve.94.062306. URL http://dx.doi.org/10.1103/PhysRevE.94.062306.

Kevin Bello, Bryon Aragam, and Pradeep Ravikumar. Dagma: learning dags via m-matrices and a log-determinant acyclicity characterization. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS '22, Red Hook, NY, USA, 2022. Curran Associates Inc. ISBN 9781713871088.

Konstantinos Benidis, Syama Sundar Rangapuram, Valentin Flunkert, Yuyang Wang, Danielle Maddix, Caner Turkmen, Jan Gasthaus, Michael Bohlke-Schneider, David Salinas, Lorenzo Stella, François-Xavier Aubet, Laurent Callot, and Tim Januschowski. Deep learning for time series forecasting: Tutorial and literature survey. *ACM Comput. Surv.*, 55(6), December 2022. ISSN 0360-0300. doi: 10.1145/3533382. URL https://doi.org/10.1145/3533382.

Cristian Challu, Kin G. Olivares, Boris N. Oreshkin, Federico Garza Ramirez, Max Mergenthaler-Canseco, and Artur Dubrawski. Nhits: neural hierarchical interpolation for time series forecasting. In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence*, AAAI'23/IAAI'23/EAAI'23. AAAI Press, 2023. ISBN 978-1-57735-880-0. doi: 10.1609/aaai.v37i6.25854. URL https://doi.org/10.1609/aaai.v37i6.25854.

Yuxiao Cheng, Runzhao Yang, Tingxiong Xiao, Zongren Li, Jinli Suo, Kunlun He, and Qionghai Dai. Cuts: Neural causal discovery from irregular time-series data, 2023. URL https://arxiv.org/abs/2302.07458.

Yuxiao Cheng, Lianglong Li, Tingxiong Xiao, Zongren Li, Jinli Suo, Kunlun He, and Qionghai Dai. Cuts+: high-dimensional causal discovery from irregular time-series. In *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence and Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence and Fourteenth Symposium on Educational Advances in Artificial Intelligence*, AAAI'24/IAAI'24/EAAI'24. AAAI Press, 2024. ISBN 978-1-57735-887-9. doi: 10.1609/aaai.v38i10.29034. URL https://doi.org/10.1609/aaai.v38i10.29034.

Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008.

Shanghua Gao, Teddy Koker, Owen Queen, Thomas Hartvigsen, Theodoros Tsiligkaridis, and Marinka Zitnik. Units: a unified multi-task time series model. In *Proceedings of the 38th International Conference on Neural Information Processing Systems*, NIPS '24, Red Hook, NY, USA, 2024. Curran Associates Inc. ISBN 9798331314385.

Tomas Geffner, Javier Antoran, Adam Foster, Wenbo Gong, Chao Ma, Emre Kiciman, Amit Sharma, Angus Lamb, Martin Kukla, Nick Pawlowski, Agrin Hilmkil, Joel Jennings, Meyer Scetbon, Miltiadis Allamanis, and Cheng Zhang. Deep end-to-end causal inference. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. URL https://openreview.net/forum?id=e6sqttxEGX.

Jiawei Jiang, Chengkai Han, Wayne Xin Zhao, and Jingyuan Wang. Pdformer: propagation delay-aware dynamic long-range transformer for traffic flow prediction. In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence*, AAAI'23/IAAI'23/EAAI'23. AAAI Press, 2023a. ISBN 978-1-57735-880-0. doi: 10.1609/aaai.v37i4.25556. URL https://doi.org/10.1609/aaai.v37i4.25556.

Renhe Jiang, Zhaonan Wang, Jiawei Yong, Puneet Jeph, Quanjun Chen, Yasumasa Kobayashi, Xuan Song, Shintaro Fukushima, and Toyotaro Suzumura. Spatio-temporal meta-graph learning for traffic forecasting. In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence*, AAAI'23/IAAI'23/EAAI'23. AAAI Press, 2023b. ISBN 978-1-57735-880-0. doi: 10.1609/aaai.v37i7.25976. URL https://doi.org/10.1609/aaai.v37i7.25976.

Guangyin Jin, Yuxuan Liang, Yuchen Fang, Zezhi Shao, Jincai Huang, Junbo Zhang, and Yu Zheng. Spatio-temporal graph neural networks for predictive learning in urban computing: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 36(10):5388–5408, 2024a. doi: 10.1109/TKDE.2023.3333824.

Ming Jin, Huan Yee Koh, Qingsong Wen, Daniele Zambon, Cesare Alippi, Geoffrey I. Webb, Irwin King, and Shirui Pan. A survey on graph neural networks for time series: Forecasting, classification, imputation, and anomaly detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12):10466–10485, 2024b. doi: 10.1109/TPAMI.2024.3443141.

Jesus Lago, Grzegorz Marcjasz, Bart De Schutter, and Rafał Weron. Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark. *Applied Energy*, 293:116983, 2021. ISSN 0306-2619. doi: 10.1016/j.apenergy.2021.116983. URL http://dx.doi.org/10.1016/j.apenergy.2021.116983.

Fuxian Li, Jie Feng, Huan Yan, Guangyin Jin, Fan Yang, Funing Sun, Depeng Jin, and Yong Li. Dynamic graph convolutional recurrent network for traffic prediction: Benchmark and solution. *ACM Trans. Knowl. Discov. Data*, 17(1), February 2023. ISSN 1556-4681. doi: 10.1145/3532611. URL https://doi.org/10.1145/3532611.

Yujie Li, Zezhi Shao, Yongjun Xu, Qiang Qiu, Zhaogang Cao, and Fei Wang. Dynamic frequency domain graph convolutional network for traffic forecasting. In *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5245–5249, 2024. doi: 10.1109/ICASSP48485.2024.10446144.

Bryan Lim, Sercan Ö Arık, Nicolas Loeff, and Tomas Pfister. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International journal of forecasting*, 37(4): 1748–1764, 2021.

Qingyang Liu, Hongjiu Liu, Ying Zhang, and Jijing Cai. Nonlinear parsimonious models for iot consumer electronics: Leveraging normal distribution transformations in information filtering networks. *IEEE Transactions on Consumer Electronics*, 71(4):11545–11555, 2025. doi: 10.1109/TCE.2025.3602881.

Yong Liu, Haixu Wu, Jianmin Wang, and Mingsheng Long. Non-stationary transformers: exploring the stationarity in time series forecasting. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS '22, Red Hook, NY, USA, 2022. Curran Associates Inc. ISBN 9781713871088.

Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. itransformer: Inverted transformers are effective for time series forecasting. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=JePfAI8fah.

Minbo Ma, Jilin Hu, Christian S. Jensen, Fei Teng, Peng Han, Zhiqiang Xu, and Tianrui Li. Learning time-aware graph structures for spatially correlated time series forecasting. In *2024 IEEE 40th*

International Conference on Data Engineering (ICDE), pp. 4435–4448, 2024. doi: 10.1109/ICDE60146.2024.00338.

R.N. Mantegna. Hierarchical structure in financial markets. *The European Physical Journal B*, 11 (1):193–197, 1999. ISSN 1434-6028. doi: 10.1007/s100510050929. URL http://dx.doi.org/10.1007/s100510050929.

Guido Previde Massara, Tiziana di Matteo, and Tomaso Aste. Network filtering for big data: Triangulated maximally filtered graph. *J. Complex Networks*, 5:161–178, 2015. URL https://api.semanticscholar.org/CorpusID:10394642.

Yuqi Nie, Nam H Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=Jbdc0vTOcol.

Kin G Olivares, Cristian Challu, Grzegorz Marcjasz, Rafał Weron, and Artur Dubrawski. Neural basis expansion analysis with exogenous variables: Forecasting electricity prices with nbeatsx. *International Journal of Forecasting*, 39(2):884–900, 2023.

Xiangfei Qiu, Xingjian Wu, Yan Lin, Chenjuan Guo, Jilin Hu, and Bin Yang. Duet: Dual clustering enhanced multivariate time series forecasting. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.1*, KDD '25, pp. 1185–1196, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400712456. doi: 10.1145/3690624.3709325. URL https://doi.org/10.1145/3690624.3709325.

Chao Shang, Jie Chen, and Jinbo Bi. Discrete graph structure learning for forecasting multiple time series, 2021. URL https://arxiv.org/abs/2101.06861.

Zezhi Shao, Zhao Zhang, Fei Wang, Wei Wei, and Yongjun Xu. Spatial-temporal identity: A simple yet effective baseline for multivariate time series forecasting. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, CIKM '22, pp. 4454–4458, New York, NY, USA, 2022a. Association for Computing Machinery. ISBN 9781450392365. doi: 10.1145/3511808.3557702. URL https://doi.org/10.1145/3511808.3557702.

Zezhi Shao, Zhao Zhang, Fei Wang, and Yongjun Xu. Pre-training enhanced spatial-temporal graph neural network for multivariate time series forecasting. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '22, pp. 1567–1577, New York, NY, USA, 2022b. Association for Computing Machinery. ISBN 9781450393850. doi: 10.1145/3534678.3539396. URL https://doi.org/10.1145/3534678.3539396.

Zezhi Shao, Zhao Zhang, Wei Wei, Fei Wang, Yongjun Xu, Xin Cao, and Christian S. Jensen. Decoupled dynamic spatial-temporal graph neural network for traffic forecasting. *Proc. VLDB Endow.*, 15(11):2733–2746, July 2022c. ISSN 2150-8097. doi: 10.14778/3551793.3551827. URL https://doi.org/10.14778/3551793.3551827.

M. Tumminello, T. Aste, T. Di Matteo, and R. N. Mantegna. A tool for filtering information in complex systems. *Proceedings of the National Academy of Sciences*, 102(30):10421–10426, 2005. ISSN 1091-6490. doi: 10.1073/pnas.0500298102. URL http://dx.doi.org/10.1073/pnas.0500298102.

Shiyu Wang, Haixu Wu, Xiaoming Shi, Tengge Hu, Huakun Luo, Lintao Ma, James Y. Zhang, and JUN ZHOU. Timemixer: Decomposable multiscale mixing for time series forecasting. In *The Twelfth International Conference on Learning Representations*, 2024a. URL https://openreview.net/forum?id=7oLshfEIC2.

Yihang Wang, Yuying Qiu, Peng Chen, Kai Zhao, Yang Shu, Zhongwen Rao, Lujia Pan, Bin Yang, and Chenjuan Guo. ROSE: Register-assisted general time series forecasting with decomposed frequency learning, 2025. URL https://openreview.net/forum?id=tdttNKCtyB.

Yuxuan Wang, Haixu Wu, Jiaxiang Dong, Guo Qin, Haoran Zhang, Yong Liu, Yunzhong Qiu, Jianmin Wang, and Mingsheng Long. Timexer: empowering transformers for time series forecasting

with exogenous variables. In *Proceedings of the 38th International Conference on Neural Information Processing Systems*, NIPS '24, Red Hook, NY, USA, 2024b. Curran Associates Inc. ISBN 9798331314385.

Gerald Woo, Chenghao Liu, Akshat Kumar, Caiming Xiong, Silvio Savarese, and Doyen Sahoo. Unified training of universal time series forecasting transformers. In *Proceedings of the 41st International Conference on Machine Learning*, ICML'24. JMLR.org, 2024.

Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: decomposition transformers with auto-correlation for long-term series forecasting. In *Proceedings of the 35th International Conference on Neural Information Processing Systems*, NIPS '21, Red Hook, NY, USA, 2021. Curran Associates Inc. ISBN 9781713845393.

Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. Timesnet: Temporal 2d-variation modeling for general time series analysis. In *International Conference on Learning Representations*, 2023.

Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, and Chengqi Zhang. Graph wavenet for deep spatial-temporal graph modeling. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, IJCAI'19, pp. 1907–1913. AAAI Press, 2019. ISBN 9780999241141.

Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, Xiaojun Chang, and Chengqi Zhang. Connecting the dots: Multivariate time series forecasting with graph neural networks. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '20, pp. 753–763, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450379984. doi: 10.1145/3394486.3403118. URL https://doi.org/10.1145/3394486.3403118.

Zhijian Xu, Ailing Zeng, and Qiang Xu. FITS: Modeling time series with \$10k\$ parameters. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=bWcnvZ3qMb.

Kun Yi, Qi Zhang, Wei Fan, Shoujin Wang, Pengyang Wang, Hui He, Ning An, Defu Lian, Longbing Cao, and Zhendong Niu. Frequency-domain MLPs are more effective learners in time series forecasting. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL https://openreview.net/forum?id=iif9mGCTfy.

Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are transformers effective for time series forecasting? In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence*, AAAI'23/IAAI'23/EAAI'23. AAAI Press, 2023. ISBN 978-1-57735-880-0. doi: 10.1609/aaai.v37i9.26317. URL https://doi.org/10.1609/aaai.v37i9.26317.

Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pp. 11106–11115, 2021.

Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun, and Rong Jin. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *International conference on machine learning*, pp. 27268–27286. PMLR, 2022.

A AI USAGE STATEMENT

In the process of developing this work, we utilized AI for paper polishing and for writing parts of the program code.

B DETAILED METHOD EQUATIONS

Let $y_t \in \mathbb{R}$ denote the target value at time t, and let $x_j(t)$ denote the j-th exogenous variable available at time t, where $j \in \mathcal{J}$. Given a forecasting horizon h, the task is to predict the future target value $\hat{y}_{t+h}$ using available target history and exogenous information. We define a lag set $\mathcal{L} = \{\ell_1, \ell_2, \ldots, \ell_K\}$. For each $\ell \in \mathcal{L}$, the temporal lag node is defined as:

$$x_\ell(t) = y_{t-\ell}. \tag{1}$$

At each time t, the mixed node set is defined as

$$\mathcal{X}_t = \{y_t\} \cup \{x_\ell(t) \mid \ell \in \mathcal{L}\} \cup \{x_j(t) \mid j \in \mathcal{J}\}. \tag{2}$$

MET-C3 learns a time varying dependency graph over the mixed node set as

$$\mathcal{G}_t = (\mathcal{X}_t, \mathcal{E}_t, \mathbf{P}_t), \tag{3}$$

where $\mathcal{E}_t$ denotes the sparse edge set and $\mathbf{P}_t$ denotes the partial correlation weight matrix. Based on $\mathcal{G}_t$, MET-C3 generates graph guided forecasting features $\mathbf{g}_t$. The forecasting function is written as

$$\hat{y}_{t+h} = F(\mathcal{X}_t, \mathbf{g}_t; \Theta), \tag{4}$$

where $F(\cdot)$ is a forecasting model and Θ denotes learnable parameters.

Given the mixed node set $\mathcal{X}_t$, MET-C3 estimates dependencies within a local rolling window. For each graph update time t, the rolling window with length w is defined as

$$\mathcal{W}_t = \{t - w + 1, t - w + 2, \ldots, t\}. \tag{5}$$

For each time point $s \in \mathcal{W}_t$, the mixed node values are arranged as

$$\mathbf{z}_s = \left[y_s,\, x_{\ell_1}(s), \ldots, x_{\ell_K}(s),\, x_{j_1}(s), \ldots, x_{j_d}(s)\right]^\top, \tag{6}$$

where $K = |\mathcal{L}|$ and $d = |\mathcal{J}|$.

Stacking the mixed node vectors in $\mathcal{W}_t$ gives

$$\mathbf{Z}_t = \left[\mathbf{z}_{t-w+1}, \mathbf{z}_{t-w+2}, \ldots, \mathbf{z}_t\right]^\top \in \mathbb{R}^{w \times m}, \tag{7}$$

where $m = 1 + K + d$.

The local dependency matrix is estimated as

$$\mathbf{R}_t = \mathrm{Dep}(\mathbf{Z}_t) \in \mathbb{R}^{m \times m}, \tag{8}$$

where $\mathrm{Dep}(\cdot)$ denotes the dependency estimator. In our implementation, $\mathrm{Dep}(\cdot)$ can be instantiated by Pearson correlation, Spearman correlation, or distance correlation, with Pearson correlation used as the default setting.

Given the local dependency matrix $\mathbf{R}_t$, MET-C3 applies autocorrelation constrained coupling before sparse graph construction. For the target series, the empirical autocorrelation at lag q is computed as

$$\rho(q) = \mathrm{Corr}(y_t, y_{t-q}). \tag{9}$$

The decay factor α is estimated by fitting a linear model to the logarithm of the absolute autocorrelation values,

$$\log(|\rho(q)| + \epsilon) = a - cq, c > 0 \tag{10}$$

where ϵ is a small constant for numerical stability, a is the intercept, and c is the fitted slope coefficient. The decay factor is then computed as

$$\alpha = \mathrm{clip}\left(\exp(-c), \alpha_{\min}, \alpha_{\max}\right), \tag{11}$$

where $\alpha_{\min} = 0.8$ and $\alpha_{\max} = 0.999$.

Let

$$\mathcal{X}_t^{temp} = \{y_t\} \cup \{x_\ell(t) \mid \ell \in \mathcal{L}\} \tag{12}$$

denote the temporal node set, and let

$$\mathcal{X}_t^{exo} = \{x_j(t) \mid j \in \mathcal{J}\} \tag{13}$$

denote the exogenous node set. For compact notation, we write

$$\mathcal{T}_t = \mathcal{X}_t^{temp}, \quad \mathcal{Q}_t = \mathcal{X}_t^{exo}. \tag{14}$$

For each node u, let λ_u denote its lag value. For y_t and $x_j(t)$, we set $\lambda_u = 0$. For $x_\ell(t)$, we set $\lambda_u = \ell$.

The structural weight matrix $\mathbf{C} \in \mathbb{R}^{m \times m}$ is defined as

$$C_{uv} = \begin{cases} 1, & u = v, \\ \alpha^{\min(\lambda_u, \lambda_v)}, & u, v \in \mathcal{T}_t, \\ \beta, & u, v \in \mathcal{Q}_t, \\ \beta\alpha^{\gamma \max(\lambda_u, \lambda_v)}, & \begin{array}{l} u \in \mathcal{T}_t,\ v \in \mathcal{Q}_t \\ \text{or} u \in \mathcal{Q}_t,\ v \in \mathcal{T}_t, \end{array} \end{cases} \tag{15}$$

where β controls the basic weight of exogenous related edges, and γ controls the decay strength for temporal exogenous coupling.

The constrained dependency matrix is obtained by

$$\widetilde{\mathbf{R}}_t = \mathbf{R}_t \odot \mathbf{C}, \tag{16}$$

where $\odot$ denotes the Hadamard product.

Given the constrained dependency matrix $\widetilde{\mathbf{R}}_t$, MET-C3 defines the nonnegative edge weight matrix as

$$\mathbf{W}_t = |\widetilde{\mathbf{R}}_t|, \tag{17}$$

where the diagonal elements are set to zero.

For each node $u \in \mathcal{X}_t$, the node strength is computed as

$$s_u = \sum_{v \in \mathcal{X}_t,\, v \neq u} W_{uv,t}. \tag{18}$$

MET-C3 constructs the sparse topology using the Triangulated Maximally Filtered Graph. The four nodes with the largest strengths initialize the first tetrahedron. For each remaining node u, the insertion face is selected by

$$f^*(u) = \arg \max_{f = \{a,b,c\}} \Omega(u, f), \tag{19}$$

where

$$\Omega(u, f) = W_{ua,t} + W_{ub,t} + W_{uc,t}. \tag{20}$$

After selecting $f^*(u) = \{a, b, c\}$, the node u is connected to a, b, and c. The selected face is stored as a separator, and the new four-node clique is written as

$$c_u = \{u, a, b, c\}. \tag{21}$$

The edge set, clique set, and separator set are obtained as

$$(\mathcal{E}_t, \mathcal{C}_t, \mathcal{S}_t) = \mathrm{TMFG}(\mathbf{W}_t). \tag{22}$$

The sparse graph topology is then

$$\mathcal{G}_t^{topo} = (\mathcal{X}_t, \mathcal{E}_t). \tag{23}$$

Given the constrained dependency matrix $\widetilde{\mathbf{R}}_t$, the clique set $\mathcal{C}_t$, and the separator set $\mathcal{S}_t$, MET-C3 estimates a sparse inverse dependency matrix by local and global partial correlation inference.

For each clique $c \in \mathcal{C}_t$, the local submatrix is

$$\widetilde{\mathbf{R}}_{t,c} = \widetilde{\mathbf{R}}_t[c, c]. \tag{24}$$

For each separator $s \in \mathcal{S}_t$, the separator submatrix is

$$\widetilde{\mathbf{R}}_{t,s} = \widetilde{\mathbf{R}}_t[s, s]. \tag{25}$$

The sparse inverse dependency matrix is computed as

$$\mathbf{J}_t = \sum_{c \in \mathcal{C}_t} \left[\widetilde{\mathbf{R}}_{t,c}^{-1}\right]^0 - \sum_{s \in \mathcal{S}_t} \left[\widetilde{\mathbf{R}}_{t,s}^{-1}\right]^0, \tag{26}$$

where $[\cdot]^0$ denotes the zero padded embedding of a local inverse matrix into the full $m \times m$ matrix. If a local submatrix is singular, the Moore Penrose pseudoinverse is used.

The partial correlation matrix is obtained from $\mathbf{J}_t$ as

$$P_{uv,t} = -\frac{J_{uv,t}}{\sqrt{J_{uu,t} J_{vv,t}} + \varepsilon}, \quad u \neq v, \tag{27}$$

and the diagonal elements are set as

$$P_{uu,t} = 1. \tag{28}$$

The resulting matrix $\mathbf{P}_t \in \mathbb{R}^{m \times m}$ is used as the partial correlation weight matrix in the time varying dependency graph

$$\mathcal{G}_t = (\mathcal{X}_t, \mathcal{E}_t, \mathbf{P}_t). \tag{29}$$

Given the partial correlation matrix $\mathbf{P}_t$, MET-C3 extracts dependency weights related to the target from the row for y_t. For an exogenous node $x_j(t)$, the graph weight is defined as

$$w_{j,t}^{exo} = P_{y,j,t}, \quad j \in \mathcal{J}. \tag{30}$$

For a temporal lag node $x_\ell(t)$, the graph weight is defined as

$$w_{\ell,t}^{lag} = P_{y,\ell,t}, \quad \ell \in \mathcal{L}. \tag{31}$$

Based on these weights, MET-C3 constructs two graph guided modulation signals. The exogenous impact signal is

$$g_t^{exo} = \sum_{j \in \mathcal{J}} w_{j,t}^{exo} x_j(t), \tag{32}$$

and the lag feedback signal is

$$g_t^{lag} = \sum_{\ell \in \mathcal{L}} w_{\ell,t}^{lag} x_\ell(t). \tag{33}$$

MET-C3 also extracts two graph summary features from partial correlations related to the target. Let

$$\mathcal{N}_t = \mathcal{X}_t \setminus \{y_t\}. \tag{34}$$

The graph complexity feature is defined as

$$g_t^{comp} = \sum_{u \in \mathcal{N}_t} \mathbb{I}(|P_{y,u,t}| > \theta), \tag{35}$$

where $\theta = 0.05$. The graph hubness feature is defined as

$$g_t^{hub} = \max_{u \in \mathcal{N}_t} |P_{y,u,t}|. \tag{36}$$

The graph guided feature vector is then written as

$$\mathbf{g}_t = \left[g_t^{exo}, g_t^{lag}, g_t^{comp}, g_t^{hub}\right]^\top. \tag{37}$$

When graph structures are updated at a lower frequency than the forecasting samples, the latest available $\mathbf{g}_t$ is used for subsequent timestamps until a new graph is generated.

Let r denote the index of the target node y_t. The partial correlation score related to the target is defined as

$$\pi_{u,t} = P_{ru,t}, \quad u \in \mathcal{X}_t \setminus \{y_t\}. \tag{38}$$

The lag dependence indicator measures the total partial correlation strength between the target and its lagged states:

$$S_t^{lag} = \sum_{u \in \mathcal{X}_t^{temp} \setminus \{y_t\}} |\pi_{u,t}|. \tag{39}$$

The exogenous influence indicator measures the total partial correlation strength between the target and exogenous variables:

$$S_t^{exo} = \sum_{u \in \mathcal{X}_t^{exo}} |\pi_{u,t}|. \tag{40}$$

The relative exogenous ratio is defined as

$$S_t^{ratio} = \frac{S_t^{exo}}{S_t^{lag} + S_t^{exo} + \varepsilon}. \tag{41}$$

For topology change, we use the target neighborhood

$$\mathcal{N}_t(y) = \{u \mid \{u, r\} \in \mathcal{E}_t\}. \tag{42}$$

The target neighborhood change ratio is computed as

$$S_t^{chg} = 1 - \frac{|\mathcal{N}_t(y) \cap \mathcal{N}_{t-1}(y)|}{|\mathcal{N}_t(y) \cup \mathcal{N}_{t-1}(y)| + \varepsilon}. \tag{43}$$

The structural indicator vector is written as

$$\mathbf{s}_t = \left[S_t^{lag},\, S_t^{exo},\, S_t^{ratio},\, S_t^{chg} \right]^\top. \tag{44}$$

Given the mixed node vector $\mathbf{z}_t$ and the graph guided feature vector $\mathbf{g}_t$, the enhanced forecasting input is defined as

$$\mathbf{u}_t = \left[\mathbf{z}_t^\top,\, \mathbf{g}_t^\top \right]^\top. \tag{45}$$

The forecasting model predicts the target value at horizon h by

$$\hat{y}_{t+h} = F(\mathbf{u}_t; \Theta), \tag{46}$$

where $F(\cdot)$ denotes the forecasting backbone and Θ denotes learnable parameters. The training objective is

$$\mathcal{L}_{pred} = \frac{1}{N} \sum_{t=1}^{N} \left| y_{t+h} - \hat{y}_{t+h} \right|. \tag{47}$$

This formulation allows the graph guided features to be integrated with different forecasting backbones without changing their internal architectures.

$$\mathrm{MAE} = \frac{1}{N} \sum_{t=1}^{N} \left| y_{t+h} - \hat{y}_{t+h} \right|. \tag{48}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N} \sum_{t=1}^{N} (y_{t+h} - \hat{y}_{t+h})^2}. \tag{49}$$

C SUPPLEMENTARY FIGURES AND TABLES

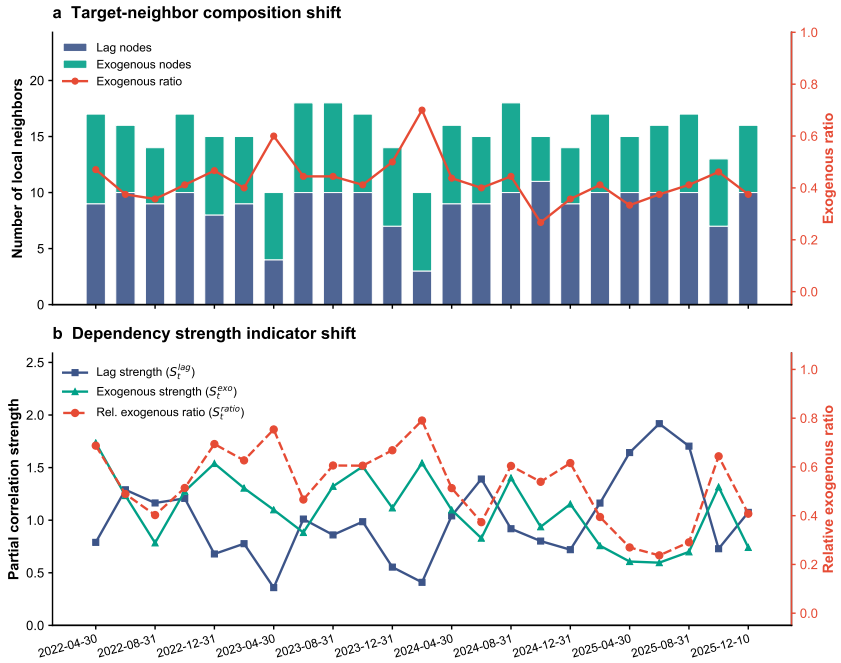

Figure 2: Dynamic shifts between temporal and exogenous dependencies across analysis windows.

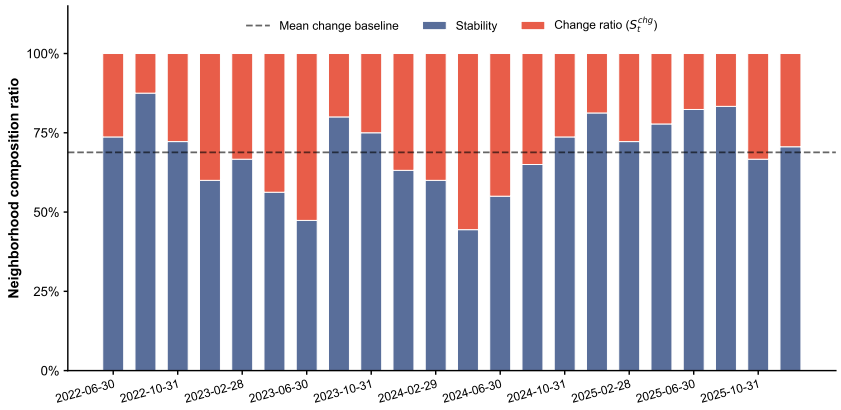

Figure 3: Reorganization of the target neighborhood across analysis windows.

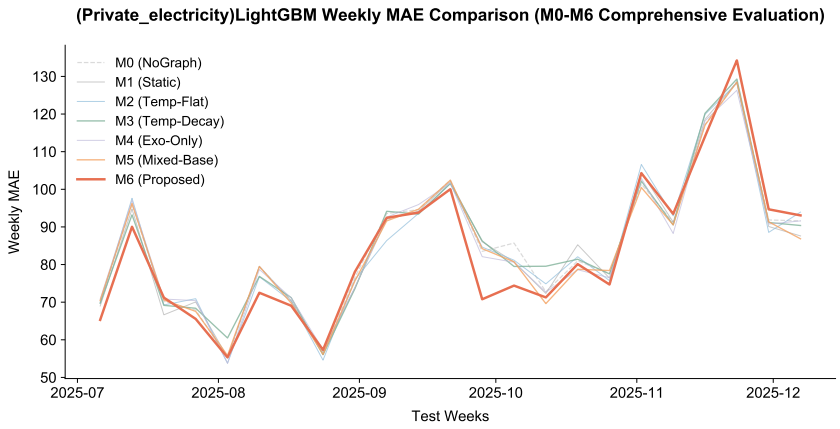

Figure 4: Weekly MAE comparison of MET-C3 ablation variants on the main dataset.

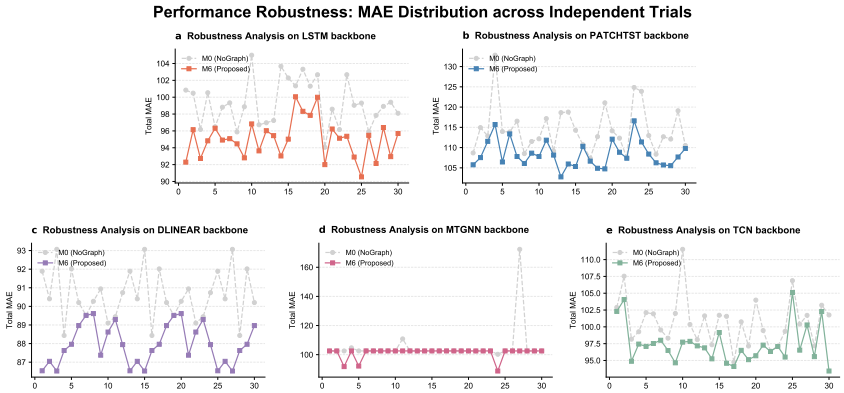

Figure 5: Performance robustness across forecasting backbones on the main dataset.

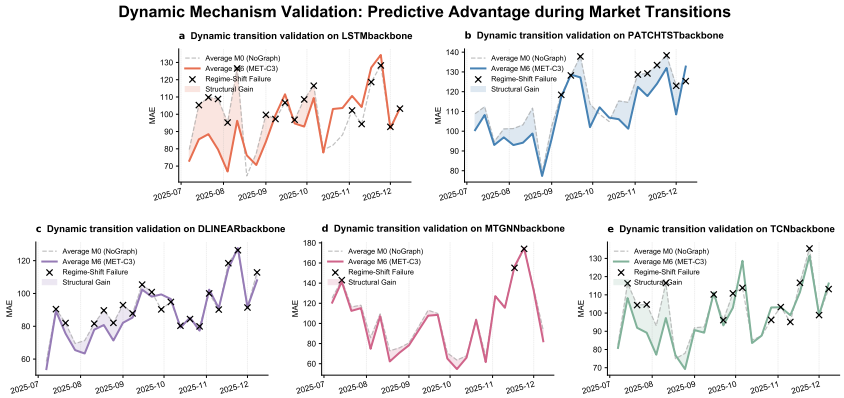

Figure 6: Dynamic transition validation across forecasting backbones on the main dataset.

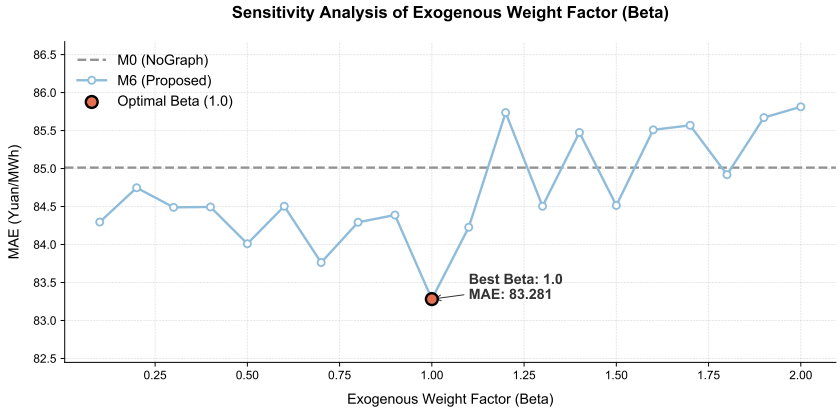

Figure 7: Sensitivity analysis of the exogenous weight factor β on the main dataset.

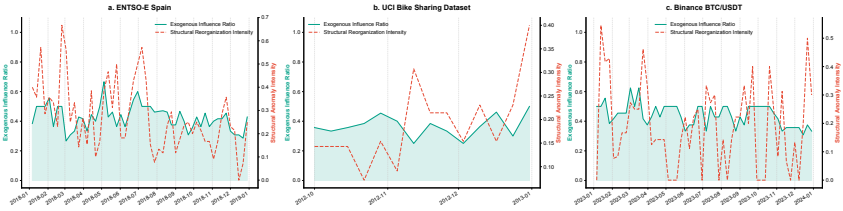

Figure 8: Cross dataset structural evolution of exogenous influence and topological reorganization. Panels (a) to (c) correspond to ENTSO-E Spain, UCI Bike Sharing Dataset, and Binance BTC/USDT, respectively. The solid curve denotes the exogenous influence ratio, while the dashed curve denotes the structural reorganization intensity.

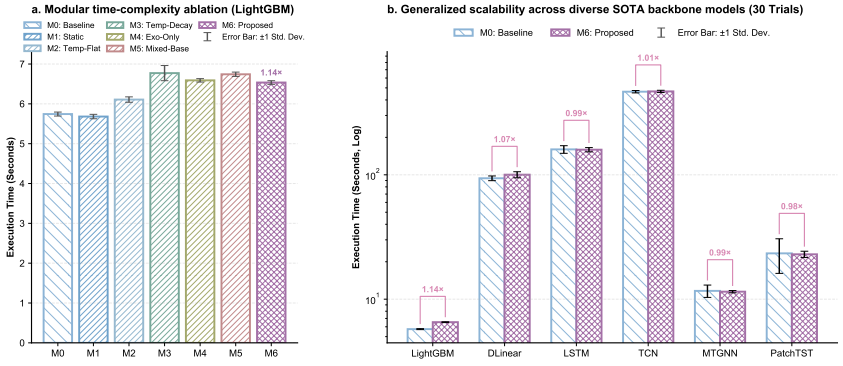

Figure 9: Computational efficiency analysis on the Shandong Electricity Market dataset. (a) Modular time complexity ablation under the LightGBM backbone; (b) runtime comparison between M0 and M6 across different forecasting backbones.



Figure 10: Computational scalability analysis across public datasets. Subfigures (a) to (c) compare the runtime of M0 and M6 on ENTSO-E Spain, UCI Bike Sharing Dataset, and Binance BTC/USDT, respectively.

Table 1. Bootstrap stability of target neighbor structures.

Window	Total	Stable	Lag	Exo	Mean Probability
2024-04-30	20	13	8	5	0.950
2024-06-30	20	14	9	5	0.914
2024-08-31	21	15	10	5	0.883
2024-10-31	20	10	6	4	0.900
2024-12-31	20	10	5	5	0.895
2025-02-28	19	12	8	4	0.925

Table 2. Dataset statistics and forecasting settings.

Dataset	Domain	Frequency	Samples	Target	Exogenous variables	Forecasting horizon
Shandong Electricity Market	Electricity	15 min	135,360	Provincial nodal price	10	480 steps / 5 days
ENTSO-E Spain	Electricity	1 h	34,344	Actual price	10	120 steps / 5 days
UCI Bike Sharing Dataset	Mobility	1 h	16,824	Rental demand	8	72 steps / 3 days
Binance BTC/USDT	Finance	1 h	34,344	Close price	6	120 steps / 5 days

Table 3. MET-C3 ablation results on the main dataset.

Model	Description	MAE	RMSE	MAE change vs. M0
M0	Original forecasting backbone without graph guided features	85.0118	127.6813	0.00%
M1	Static graph features constructed without updates over time	84.6679	127.7651	↓ 0.40%
M2	Temporal graph features with flat temporal weighting	84.7175	127.7128	↓ 0.35%
M3	Temporal graph features with autocorrelation based temporal decay	85.0631	127.8732	↑ 0.06%
M4	Exogenous graph features constructed from external variables	84.4602	126.9342	↓ 0.65%
M5	Mixed temporal and exogenous graph features without the full MET-C3 design	84.2609	126.7360	↓ 0.88%
M6	Full MET-C3 with changing mixed dependencies and graph guided features	83.2810	127.2179	↓ 2.04%

Table 4. Robustness analysis across forecasting backbones on the main dataset. (The symbol Δ denotes the standard deviation.)

Backbone	Model type	Mean MAE of M0	Mean MAE of M6	Improvement	Robustness
DLinear	Linear model	$90.60 \pm \Delta 1.36$	$87.96 \pm \Delta 1.08$	$2.92\% \pm \Delta 2.34\%$	47.6%
LSTM	Recurrent neural network	$99.26 \pm \Delta 2.71$	$95.02 \pm \Delta 2.28$	$4.28\% \pm \Delta 2.72\%$	75.0%
MTGNN	Graph neural network	$105.16 \pm \Delta 12.79$	$101.38 \pm \Delta 3.58$	$3.59\% \pm \Delta 7.92\%$	63.8%
PatchTST	Transformer based model	$114.47 \pm \Delta 5.71$	$108.36 \pm \Delta 3.27$	$5.34\% \pm \Delta 4.11\%$	58.8%
TCN	Temporal convolutional network	$100.80 \pm \Delta 3.54$	$97.40 \pm \Delta 2.85$	$3.38\% \pm \Delta 2.82\%$	41.1%

Table 5. Robustness comparison of M0 and M6 across datasets and forecasting backbones.

Dataset	Backbone	Mean MAE of M0	Mean MAE of M6	Improvement	Robustness
ENTSO-E Spain	LightGBM	$5.23 \pm \Delta\, 0.00$	$5.18 \pm \Delta\, 0.00$	$0.88\% \pm \Delta\, 0.00\%$	100.00%
	LSTM	$5.45 \pm \Delta\, 0.07$	$5.37 \pm \Delta\, 0.06$	$1.46\% \pm \Delta\, 1.14\%$	35.70%
	PatchTST	$8.14 \pm \Delta\, 0.23$	$7.81 \pm \Delta\, 0.28$	$4.07\% \pm \Delta\, 3.00\%$	51.70%
	DLinear	$5.75 \pm \Delta\, 0.11$	$5.62 \pm \Delta\, 0.10$	$2.29\% \pm \Delta\, 1.89\%$	50.00%
	MTGNN	$8.24 \pm \Delta\, 0.26$	$7.92 \pm \Delta\, 0.23$	$3.94\% \pm \Delta\, 2.76\%$	48.40%
	TCN	$6.10 \pm \Delta\, 0.18$	$5.90 \pm \Delta\, 0.17$	$3.23\% \pm \Delta\, 2.89\%$	31.90%
UCI Bike Sharing Dataset	LightGBM	$62.49 \pm \Delta\, 0.00$	$61.64 \pm \Delta\, 0.00$	$1.36\% \pm \Delta\, 0.00\%$	100.00%
	LSTM	$68.81 \pm \Delta\, 0.37$	$68.41 \pm \Delta\, 0.26$	$0.59\% \pm \Delta\, 0.51\%$	41.70%
	PatchTST	$145.47 \pm \Delta\, 5.78$	$137.08 \pm \Delta\, 5.80$	$5.77\% \pm \Delta\, 3.96\%$	44.10%
	DLinear	$69.58 \pm \Delta\, 5.21$	$67.60 \pm \Delta\, 0.94$	$2.84\% \pm \Delta\, 5.08\%$	55.60%
	MTGNN	$159.15 \pm \Delta\, 1.25$	$157.57 \pm \Delta\, 0.94$	$0.99\% \pm \Delta\, 0.87\%$	65.20%
	TCN	$79.66 \pm \Delta\, 2.25$	$76.70 \pm \Delta\, 1.80$	$3.72\% \pm \Delta\, 2.45\%$	50.00%
Binance BTC/USDT	LightGBM	$2072.98 \pm \Delta\, 0.00$	$1828.78 \pm \Delta\, 0.00$	$11.78\% \pm \Delta\, 0.00\%$	100.00%
	LSTM	$1606.12 \pm \Delta\, 195.60$	$1298.35 \pm \Delta\, 79.66$	$19.16\% \pm \Delta\, 9.42\%$	85.70%
	PatchTST	$3690.07 \pm \Delta\, 1197.17$	$2476.41 \pm \Delta\, 755.66$	$32.89\% \pm \Delta\, 16.38\%$	49.20%
	DLinear	$1139.62 \pm \Delta\, 26.96$	$1114.85 \pm \Delta\, 12.57$	$2.17\% \pm \Delta\, 1.91\%$	40.00%
	MTGNN	$3133.06 \pm \Delta\, 503.26$	$2594.97 \pm \Delta\, 351.11$	$17.17\% \pm \Delta\, 10.63\%$	37.00%
	TCN	$594.07 \pm \Delta\, 42.39$	$561.38 \pm \Delta\, 26.59$	$5.50\% \pm \Delta\, 5.37\%$	19.50%

Table 6. Average runtime comparison between M0 and M6 across datasets and forecasting backbones.

Dataset	Backbone	M0 (s)	M6 (s)	Ratio
Shandong Electricity Market	LightGBM	$5.74 \pm \Delta\, 0.05$	$6.54 \pm \Delta\, 0.05$	$1.14\times$
	DLinear	$93.80 \pm \Delta\, 4.09$	$100.25 \pm \Delta\, 5.61$	$1.07\times$
	LSTM	$160.16 \pm \Delta\, 11.28$	$159.25 \pm \Delta\, 6.31$	$0.99\times$
	TCN	$465.92 \pm \Delta\, 10.27$	$468.44 \pm \Delta\, 9.76$	$1.01\times$
	MTGNN	$11.66 \pm \Delta\, 1.33$	$11.49 \pm \Delta\, 0.24$	$0.99\times$
	PatchTST	$23.36 \pm \Delta\, 7.24$	$22.96 \pm \Delta\, 1.34$	$0.98\times$
ENTSO-E Spain	LightGBM	$3.25 \pm \Delta\, 0.01$	$3.59 \pm \Delta\, 0.03$	$1.10\times$
	DLinear	$77.31 \pm \Delta\, 2.61$	$77.25 \pm \Delta\, 4.01$	$1.00\times$
	LSTM	$37.12 \pm \Delta\, 12.09$	$35.36 \pm \Delta\, 0.76$	$0.95\times$
	TCN	$6.65 \pm \Delta\, 2.47$	$6.67 \pm \Delta\, 0.19$	$1.00\times$
	MTGNN	$7.18 \pm \Delta\, 0.17$	$7.66 \pm \Delta\, 0.12$	$1.07\times$
	PatchTST	$5.47 \pm \Delta\, 6.74$	$4.66 \pm \Delta\, 0.75$	$0.85\times$
UCI Bike Sharing Dataset	LightGBM	$8.86 \pm \Delta\, 0.07$	$7.94 \pm \Delta\, 0.17$	$0.90\times$
	DLinear	$31.37 \pm \Delta\, 1.52$	$31.91 \pm \Delta\, 1.72$	$1.02\times$
	LSTM	$20.45 \pm \Delta\, 5.92$	$19.17 \pm \Delta\, 0.41$	$0.94\times$
	TCN	$3.59 \pm \Delta\, 1.63$	$3.61 \pm \Delta\, 0.18$	$1.00\times$
	MTGNN	$8.63 \pm \Delta\, 0.39$	$8.82 \pm \Delta\, 0.28$	$1.02\times$
	PatchTST	$3.63 \pm \Delta\, 7.41$	$2.67 \pm \Delta\, 0.82$	$0.74\times$
Binance BTC/USDT	LightGBM	$0.98 \pm \Delta\, 0.03$	$1.16 \pm \Delta\, 0.05$	$1.18\times$
	DLinear	$32.58 \pm \Delta\, 0.68$	$33.36 \pm \Delta\, 1.00$	$1.02\times$
	LSTM	$193.20 \pm \Delta\, 7.88$	$223.25 \pm \Delta\, 11.91$	$1.16\times$
	TCN	$266.11 \pm \Delta\, 4.49$	$266.75 \pm \Delta\, 2.43$	$1.00\times$
	MTGNN	$4.69 \pm \Delta\, 1.83$	$3.39 \pm \Delta\, 0.89$	$0.72\times$
	PatchTST	$44.66 \pm \Delta\, 6.09$	$44.12 \pm \Delta\, 0.73$	$0.99\times$