A Proof of Theorem

Theorem 1. *Fix measurable functions g and g^*. Let $I \coloneqq \{(i,j) : i = 1,2,...,M, j = 1,2,...,K\}$, and let $J(\mathcal{X}) \coloneqq \{(i,j) \in I : \hat{R}_{ij}(g) < \hat{R}_{ij}(g^*) - \epsilon\}$. If the following two conditions hold:*

(a) $\forall (i,j),(k,l) \in I$ such that $(i,j) \neq (k,l)$, $\hat{R}_{ij}(g) - \hat{R}_{ij}(g^*) \perp \hat{R}_{kl}(g)$

(b) $\hat{R}_{ij}(g^*) < R_{ij}(g) + \epsilon$ for all $(i,j) \in J(\mathcal{X})$ almost surely,

then MSE$(\hat{R}(g)) \geq$ MSE$(\hat{R}^{wb}(g))$. Given the condition (a), if we have $0 < \alpha$ such that $\alpha < R_{ij}(g) - \hat{R}_{ij}(g^*) + \epsilon$ for all $(i,j) \in J(\mathcal{X})$ almost surely, then

$$\textit{MSE}(\hat{R}(g)) - \textit{MSE}(\hat{R}^{wb}(g)) \geq 4\alpha^2 \sum_{(i,j)\in I} \Pr[\alpha < \hat{R}_{ij}(g^*) - \hat{R}_{ij}(g) - \epsilon]. \tag{9}$$

Proof. We first show that if (a) and (b) hold, then MSE$(\hat{R}(g)) \geq$ MSE$(\hat{R}^{wb}(g))$.

$$\begin{aligned}
&\mathrm{MSE}(\hat{R}(g)) - \mathrm{MSE}(\hat{R}^{wb}(g)) \\
&\quad = \mathbb{E}[(\hat{R}(g) - R(g))^2] - \mathbb{E}[(\hat{R}^{wb}(g) - R(g))^2] \\
&\quad = \mathbb{E}[(\hat{R}(g) - \hat{R}^{wb}(g))(\hat{R}(g) + \hat{R}^{wb}(g) - 2R(g))] \\
&\quad = 4\int_{\mathcal{X}} \left(\sum_{(i,j)\in I} \mathbf{1}_{J(\mathcal{X})}(i,j) \left(\hat{R}_{ij}(g) - \hat{R}_{ij}(g^*) + \epsilon\right)\right) \left(\sum_{(i,j)\in I} \hat{R}_{ij}(g) - R_{ij}(g) \right. \\
&\qquad\qquad \left. + 1_{J(\mathcal{X})}(i,j) \left(\hat{R}_{ij}(g^*) - \hat{R}_{ij}(g) - \epsilon\right)\right) \mathrm{d}F(\mathcal{X}) \\
&\quad = 4 \sum_{\substack{(i,j),(k,l)\in I \\ \text{s.t.}(i,j)\neq(k,l)}} \int_{\mathcal{X}} \mathbf{1}_{J(\mathcal{X})}(i,j) \left(\hat{R}_{ij}(g) - \hat{R}_{ij}(g^*) + \epsilon\right) \left(\hat{R}_{kl}(g) - R_{kl}(g)\right) \mathrm{d}F(\mathcal{X}) \\
&\qquad + 4\int_{\mathcal{X}} \sum_{(i,j)\in J(\mathcal{X})} \left(\hat{R}_{ij}(g) - \hat{R}_{ij}(g^*) + \epsilon\right) \left(\hat{R}_{ij}(g^*) - R_{ij}(g) - \epsilon\right) \mathrm{d}F(\mathcal{X}).
\end{aligned} \tag{10}$$

From the independence between $\hat{R}_{ij}(g) - \hat{R}_{ij}(g^*)$ and $\hat{R}_{kl}(g)$.

$$\begin{aligned}
&4 \sum_{\substack{(i,j),(k,l)\in I \\ \text{s.t.}(i,j)\neq(k,l)}} \int_{\mathcal{X}} \mathbf{1}_{J(\mathcal{X})}(i,j) \left(\hat{R}_{ij}(g) - \hat{R}_{ij}(g^*) + \epsilon\right) \left(\hat{R}_{kl}(g) - R_{kl}(g)\right) \mathrm{d}F(\mathcal{X}) \\
&= 4 \sum_{\substack{(i,j),(k,l)\in I \\ \text{s.t.}(i,j)\neq(k,l)}} \int_{\mathcal{X}} \mathbf{1}_{J(\mathcal{X})}(i,j) \left(\hat{R}_{ij}(g) - \hat{R}_{ij}(g^*) + \epsilon\right) \mathrm{d}F(\mathcal{X}) \int_{\mathcal{X}} \hat{R}_{kl}(g) - R_{kl}(g) \mathrm{d}F(\mathcal{X}) \\
&= 4 \sum_{\substack{(i,j),(k,l)\in I \\ \text{s.t.}(i,j)\neq(k,l)}} \left(\int_{\mathcal{X}} \mathbf{1}_{J(\mathcal{X})}(i,j) \left(\hat{R}_{ij}(g) - \hat{R}_{ij}(g^*) + \epsilon\right) \mathrm{d}F(\mathcal{X}) \right) \cdot 0 \\
&= 0.
\end{aligned} \tag{11}$$

From (b) and by the definition of $J(\mathcal{X})$, $\hat{R}_{ij}(g) - \hat{R}_{ij}(g^*) + \epsilon \leq 0$ and $\hat{R}_{ij}(g^*) - R_{ij}(g) - \epsilon \leq 0$ for $(i,j) \in J(\mathcal{X})$ with probability 1. Therefore,

$$\begin{aligned}
&\mathrm{MSE}(\hat{R}(g)) - \mathrm{MSE}(\hat{R}^{wb}(g)) \\
&\quad = 4\int_{\mathcal{X}} \sum_{(i,j)\in J(\mathcal{X})} \left(\hat{R}_{ij}(g) - \hat{R}_{ij}(g^*) + \epsilon\right) \left(\hat{R}_{ij}(g^*) - R_{ij}(g) - \epsilon\right) \mathrm{d}F(\mathcal{X}) \\
&\quad \geq 0.
\end{aligned} \tag{12}$$

Assume that we have $\alpha > 0$ such that $\alpha < R_{ij}(g) - \hat{R}_{ij}(g^*) + \epsilon$ for all $(i,j) \in J(\mathcal{X})$ with probability 1. Let A be the event that $\alpha < R_{ij}(g) - \hat{R}_{ij}(g^*) + \epsilon$ for all $(i,j) \in J(\mathcal{X})$. Then,

$$
\begin{aligned}
&\mathrm{MSE}(\hat{R}(g)) - \mathrm{MSE}(\hat{R}^{wb}(g)) \\
&= 4 \sum_{(i,j)\in I} \int_{\mathcal{X}|A} 1_{J(\mathcal{X})}(i,j) \left(\hat{R}_{ij}(g) - \hat{R}_{ij}(g^*) + \epsilon\right) \left(\hat{R}_{ij}(g^*) - R_{ij}(g) - \epsilon\right) \mathrm{d}F(\mathcal{X}) \\
&= 4 \sum_{(i,j)\in I} \Pr[\alpha < \hat{R}_{ij}(g^*) - \hat{R}_{ij}(g) - \epsilon] \\
&\cdot \int_{\mathcal{X}|A, \alpha < \hat{R}_{ij}(g^*) - \hat{R}_{ij}(g) - \epsilon} 1_{J(\mathcal{X})}(i,j) \left(\hat{R}_{ij}(g) - \hat{R}_{ij}(g^*) + \epsilon\right) \left(\hat{R}_{ij}(g^*) - R_{ij}(g) - \epsilon\right) \mathrm{d}F(\mathcal{X}) \\
&\geq 4\alpha^2 \sum_{(i,j)\in I} \Pr[\alpha < \hat{R}_{ij}(g^*) - \hat{R}_{ij}(g) - \epsilon].
\end{aligned}
\tag{13}
$$

□

B Pseudocode of WaveBound

In Section 3, we describe how wave empirical risk can be obtained in mini-batched optimization process. In practice, WaveBound can be easily implemented by introducing an additional target network and slightly modifying the training objective in the optimization process as shown in Algorithm 1.

Algorithm 1 *WaveBound*

Input: source network g_θ, target network g_τ, training dataset $\mathcal{X}$, hyperparameters ϵ and τ
Output: target network g_τ

for $t = 1$ to T **do**
    $\mathcal{B} \leftarrow \{(x_i, y_i) \sim \mathcal{X}\}_{i=1}^{B}$ ▷ Sample a batch of size B
    for $j = 1$ to M **do**
        for $k = 1$ to K **do**
            for $(x_i, y_i) \in \mathcal{B}$ **do**
                $l_{ijk} \leftarrow \ell((g_\theta)_{jk}(x_i), (y_i)_{jk})$ ▷ Compute the loss for g_θ
                $r_{ijk} \leftarrow \ell((g_\tau)_{jk}(x_i), (y_i)_{jk})$ ▷ Compute the loss for g_τ
            end for
        end for
    end for
    $\delta\theta \leftarrow \frac{1}{MK} \sum_{j,k} \partial_\theta \left| \frac{1}{B} \sum_i l_{ijk} - \frac{1}{B} \sum_i r_{ijk} + \epsilon \right|$ ▷ Compute the loss gradient w.r.t. θ
    $\theta \leftarrow$ optimizer$(\theta, \delta\theta)$ ▷ update the source network parameter
    $\tau \leftarrow \alpha\tau + (1-\alpha)\theta$ ▷ update the target network parameter
end for

C Detailed Experimental Settings

In the experiments, WaveBound is applied as the regularization technique in Autoformer [5], Pyraformer [6], Informer [7], LSTNet [14], and TCN [16] for the multivariate settings. For the univariate settings, N-BEATS [15] is additionally included. As described in Section 4, we use six real-world benchmarks: ETT, Electricity, Exchange, Traffic, Weather, and ILI. In both the multivariate and univariate settings, we train baselines with the batch size of 32 and Adam optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The learning rate is searched in $\{0.00001, 0.00003, 0.0001, 0.0003, 0.001\}$ for each model. In all experiments, we perform early stopping on the validation set. For WaveBound, the target network g_τ is updated with exponential moving average (EMA) of the source network g_θ with a decay rate $\alpha = 0.99$. When applying WaveBound to each baseline, the hyperparameter ϵ is searched in $\{0.01, 0.001\}$. We use a single TITAN RTX GPU in all experiments.

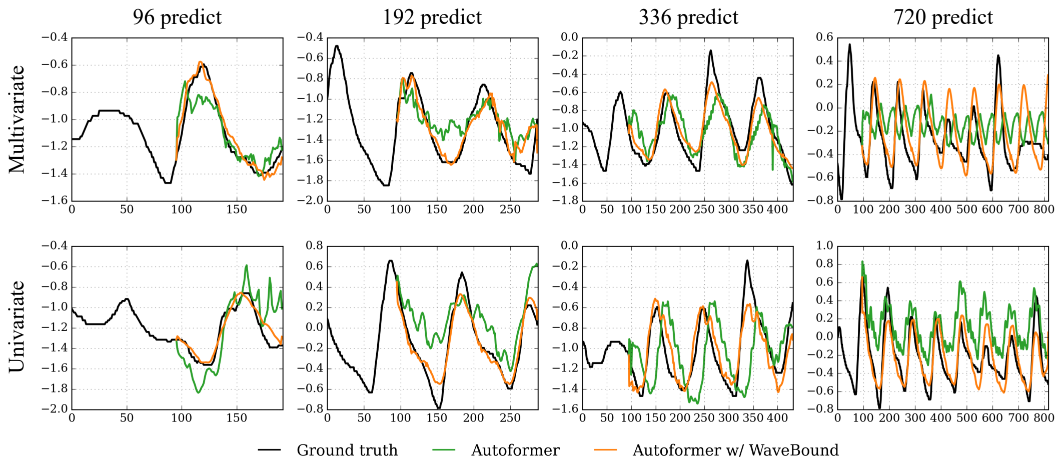

Figure 6: Qualitative results of Autoformer with and without WaveBound on the ETTm2 dataset. When using WaveBound, the results are relatively smoother and have less lag.

D Additional Experimental Results

Quantitative Results. The additional experimental results of our proposed WaveBound method for full benchamrks and baselines are reported in Table 14, 15, and 16.

Qualitative Results. To illustrate the effects of WaveBound, we depict the predictions of different models on ETTm2 dataset (96 input). For the multivariate setting, we plot the predictions for the last feature. As shown in Figure 6, the models trained with WaveBound consistently predict accurately regardless of forecast step.

E Effects of ϵ in WaveBound

Table 4 shows the results for different hyperparameter values of ϵ which are searched in $\{0.01, 0.001\}$. The performance can be improved by selecting proper ϵ in each setting.

Table 4: Performance comparison under difference choices of hyperparameter $\epsilon \in \{0.01, 0.001\}$. We evaluate the performance of Autoformer on two different datasets ETT and Electricity.

		ETTm2				Electricity			
		96	192	336	720	96	192	336	720
$\epsilon = 0.01$	MSE	0.206	0.268	0.323	0.414	0.185	0.205	0.217	0.260
	MAE	0.290	0.327	0.361	0.412	0.300	0.317	0.327	0.359
$\epsilon = 0.001$	MSE	0.204	0.265	0.320	0.413	0.176	0.233	0.228	0.282
	MAE	0.285	0.322	0.356	0.408	0.288	0.333	0.331	0.379

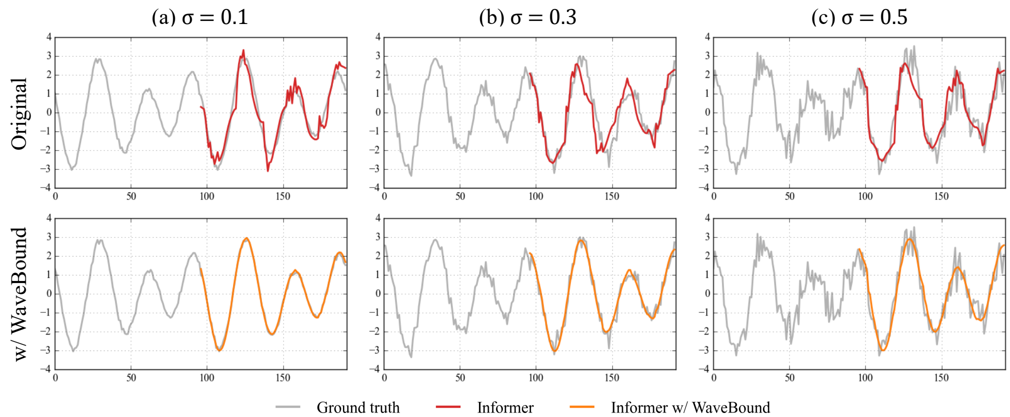

Figure 7: Qualitative results of Informer trained with and without WaveBound on the synthetic dataset. As the noise level σ increases, Informer (top) is highly affected by the noise and produces the unstable predictions. In contrast, the model with WaveBound (bottom) shows the relatively stable predictions and correctly estimate the trend of the signal regardless of the noise level in the dataset.

F Performance Analysis on Synthetic Dataset

Synthetic dataset. To analyze how WaveBound behaves in different noise levels and different dataset sizes, we synthesize a dataset composed of sine wave with the Gaussian noise. Formally, for a given noise level σ and a time step t, the synthetic dataset is formed as: $f(t) = 2\sin(2\pi t/32) + \sin(2\pi t/48) + \sigma z$, where z follows the normal distribution.

Analysis on different noise levels and dataset sizes. In time series forecasting, the overfitting issue becomes significant when the train data is small and the data contains a high level of noise. Using the synthetic dataset, we examine the effectiveness of WaveBound in different levels of overfitting. Specifically, we examine Informer [7] trained with and without WaveBound by varying the noise level σ and the datset size $|\mathcal{D}|$ of the synthetic dataset. Table 5 and Table 6 shows the performance of Informer with and without WaveBound in different noise levels and dataset sizes, respectively. As the noise level increases and dataset size decreases, Informer struggles with the unpredictable noise and produces noisy predictions, resulting in poor performance while WaveBound improves performance in all noise levels and dataset sizes. Figure 7 qualitatively shows that WaveBound allows the model to make stable predictions even in the high level of noise.

Table 5: Performance comparison under different noise levels $\sigma \in \{0.1, 0.3, 0.5\}$ with the fixed dataset size $|\mathcal{D}| = 2000$. All results are averaged over 3 trials. Subscripts denote standard deviations.

		$\sigma = 0.1$	$\sigma = 0.3$	$\sigma = 0.5$
Informer	MSE	$0.225_{\pm 0.024}$	$0.316_{\pm 0.022}$	$0.605_{\pm 0.061}$
	MAE	$0.363_{\pm 0.017}$	$0.441_{\pm 0.013}$	$0.613_{\pm 0.029}$
Informer w/ WaveBound	MSE	$\mathbf{0.015}_{\pm \mathbf{0.001}}$	$\mathbf{0.102}_{\pm \mathbf{0.001}}$	$\mathbf{0.281}_{\pm \mathbf{0.001}}$
	MAE	$\mathbf{0.096}_{\pm \mathbf{0.003}}$	$\mathbf{0.252}_{\pm \mathbf{0.002}}$	$\mathbf{0.419}_{\pm \mathbf{0.000}}$

Table 6: Performance comparison under different dataset sizes $|\mathcal{D}| \in \{4000, 3000, 2000\}$, with the fixed noise level $\sigma = 0.5$. All results are averaged over 3 trials. Subscripts denote standard deviations.

		$|\mathcal{D}| = 4000$	$|\mathcal{D}| = 3000$	$|\mathcal{D}| = 2000$
Informer	MSE	$0.311_{\pm 0.004}$	$0.399_{\pm 0.003}$	$0.605_{\pm 0.061}$
	MAE	$0.443_{\pm 0.003}$	$0.505_{\pm 0.002}$	$0.613_{\pm 0.029}$
Informer w/ WaveBound	MSE	$\mathbf{0.246}_{\pm \mathbf{0.001}}$	$\mathbf{0.290}_{\pm \mathbf{0.002}}$	$\mathbf{0.281}_{\pm \mathbf{0.001}}$
	MAE	$\mathbf{0.394}_{\pm \mathbf{0.000}}$	$\mathbf{0.435}_{\pm \mathbf{0.001}}$	$\mathbf{0.419}_{\pm \mathbf{0.000}}$

G Impact of WaveBound on Individual Features and Time Steps

To properly address an overfitting issue, WaveBound estimates adequate error bounds for each feature and time step individually. To justify the assumptions made by WaveBound, we depict the training curve of each feature and time step of the prediction made by Autoformer on the ETTm2 dataset. Each row and column in Figure 8 denote one feature and one time step, respectively. As we assumed, we can observe that different features show different training dynamics, which indicates that the difficulty of prediction varies through the features. Moreover, the magnitude of error increases as the forecast step increases. Those observations highlight that different regularization strategies are required for different features and time steps.

We also observe that existing regularization methods such as early-stopping cannot properly handle the overfitting issue, as it considers only the aggregated value rather than considering each error individually. Red dotted lines in Figure 8 denote the iteration of the best checkpoint found by early-stopping. Interestingly, even before the best checkpoint is decided by the early-stopping (red dotted lines), the model seems to be overfitted for several features; the train error (green line) is decreasing, but the test error (black line) is increasing. We conjecture that this is because of the property of the early-stopping, which considers only the mean value of error rather than individual errors of each feature and time step. In contrast, WaveBound retains the test loss (red lines) at a low level by preventing the train error (blue lines) from falling below a certain level for each feature and time step. In practice, this results in an overall performance improvement in time series forecasting, as shown in our extensive experiments.

H WaveBound with RevIN

Reversible instance normalization (RevIN) [41] is recently proposed to mitigate the distribution shift problem in time series forecasting. In this section, we conduct the additional experiments by applying both WaveBound and RevIN in training. As shown in Table 7, the performance of Autoformer [5] and Informer [7] is improved with a significant margin when both WaveBound and RevIN are applied.

Table 7: The results of WaveBound with RevIN. All results are averaged over 3 trials. Subscripts denote standard deviations.

Models		Autoformer						Informer					
		Origin		RevIN		**RevIN w/ WaveBound**		Origin		RevIN		**RevIN w/ WaveBound**	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTm2	96	0.262 ±0.037	0.326 ±0.014	0.230 ±0.007	0.303 ±0.006	**0.208 ±0.002**	**0.282 ±0.002**	0.376 ±0.056	0.477 ±0.047	0.270 ±0.041	0.326 ±0.025	**0.206 ±0.000**	**0.286 ±0.001**
	192	0.284 ±0.003	0.342 ±0.002	0.291 ±0.005	0.338 ±0.003	**0.268 ±0.001**	**0.319 ±0.001**	0.751 ±0.020	0.672 ±0.004	0.475 ±0.078	0.439 ±0.033	**0.293 ±0.011**	**0.341 ±0.008**
	336	0.338 ±0.007	0.374 ±0.005	0.346 ±0.004	0.369 ±0.004	**0.329 ±0.000**	**0.357 ±0.000**	1.440 ±0.180	0.917 ±0.067	0.517 ±0.089	0.467 ±0.041	**0.367 ±0.005**	**0.385 ±0.003**
	720	0.446 ±0.011	0.435 ±0.006	0.435 ±0.010	0.419 ±0.007	**0.417 ±0.003**	**0.407 ±0.001**	3.897 ±0.562	1.498 ±0.128	0.646 ±0.075	0.531 ±0.034	**0.476 ±0.015**	**0.442 ±0.007**
ECL	96	0.202 ±0.003	0.317 ±0.003	0.178 ±0.001	0.285 ±0.001	**0.173 ±0.008**	**0.276 ±0.006**	0.335 ±0.008	0.417 ±0.004	0.196 ±0.001	0.303 ±0.000	**0.166 ±0.001**	**0.269 ±0.001**
	192	0.235 ±0.011	0.340 ±0.009	0.218 ±0.010	0.317 ±0.009	**0.212 ±0.017**	**0.308 ±0.015**	0.341 ±0.013	0.426 ±0.011	0.217 ±0.004	0.322 ±0.004	**0.181 ±0.003**	**0.283 ±0.003**
	336	0.247 ±0.011	0.351 ±0.009	0.241 ±0.016	0.335 ±0.013	**0.223 ±0.019**	**0.318 ±0.014**	0.369 ±0.011	0.448 ±0.009	0.236 ±0.004	0.339 ±0.003	**0.194 ±0.002**	**0.297 ±0.003**
	720	0.270 ±0.006	0.371 ±0.003	0.259 ±0.006	0.349 ±0.003	**0.251 ±0.021**	**0.342 ±0.017**	0.396 ±0.009	0.457 ±0.005	0.267 ±0.001	0.363 ±0.001	**0.217 ±0.001**	**0.317 ±0.000**

I WaveBound in Short-Term Series Forecasting

In this section, we examine WaveBound in short-term series forecasting. Specifically, we use Autoformer [5] as the baseline and train the model to forecast 3, 6, and 12 steps on ECL, exchange

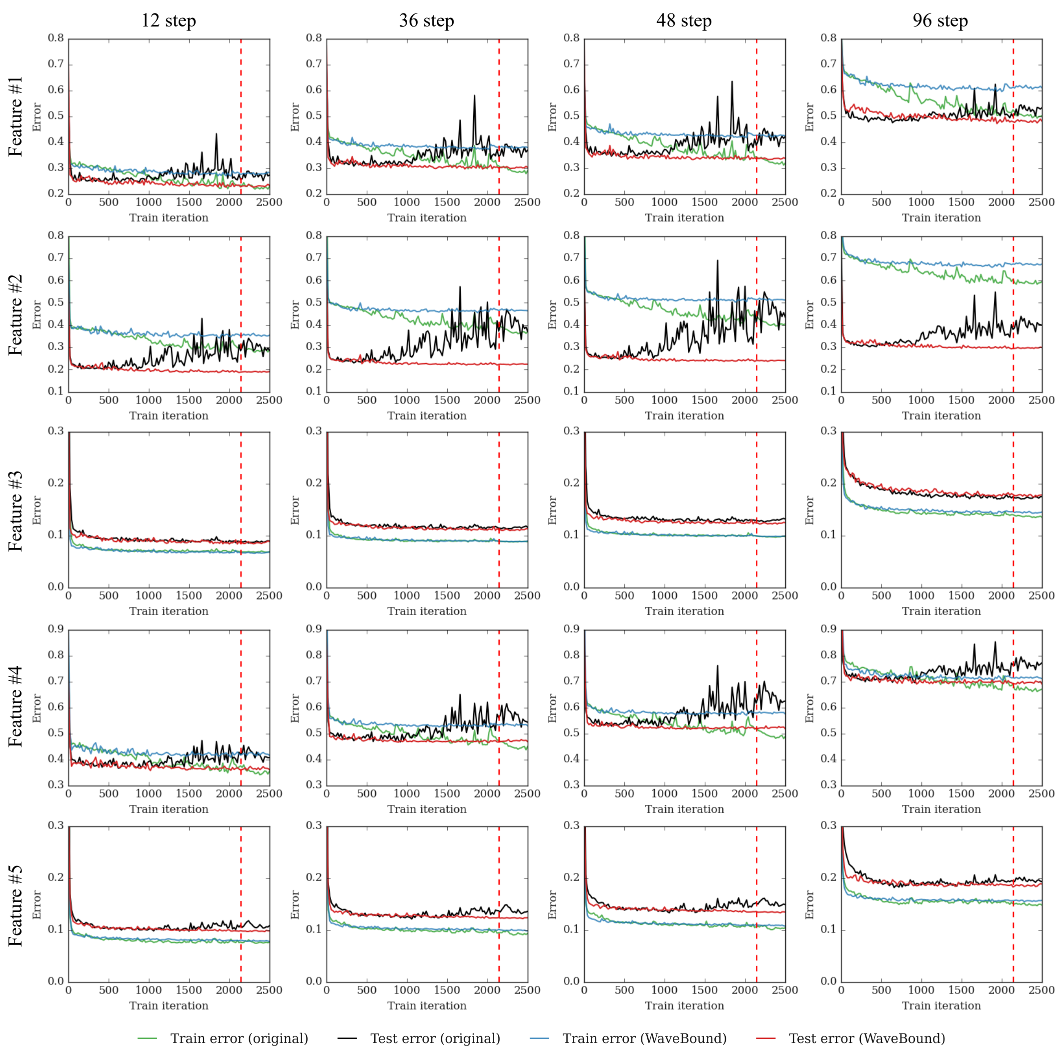

Figure 8: Training curve for each feature and time step of Autoformer on the ETTm2 dataset. Each row and column denote the feature and time step of the prediction, respectively. The red dotted line denotes the iteration of best checkpoint found by early-stopping. As we hypothesized, the output variables for different features and time steps show different trends for their training curves, which indicates that the consideration for each value is essential to properly handle the overfitting issue in time series forecasting.

rate, and traffic dataset. Table 8 shows that WaveBound still achieves the performance improvements even in short-term series forecasting.

Table 8: Short-term forecasting results of WaveBound in the multivariate setting. We use ECL, exchange rate and traffic dataset with three prediction lengths $L \in \{3, 6, 12\}$. All results are averaged over 3 trials. Subscripts denote standard deviations.

Models		Autoformer			
		Origin		**w/ Ours**	
Metric		MSE	MAE	MSE	MAE
ECL	3	$0.148_{\pm 0.001}$	$0.274_{\pm 0.001}$	$\mathbf{0.133_{\pm 0.002}}$	$\mathbf{0.256_{\pm 0.002}}$
	6	$0.152_{\pm 0.001}$	$0.277_{\pm 0.000}$	$\mathbf{0.140_{\pm 0.001}}$	$\mathbf{0.263_{\pm 0.001}}$
	12	$0.157_{\pm 0.001}$	$0.282_{\pm 0.002}$	$\mathbf{0.144_{\pm 0.001}}$	$\mathbf{0.265_{\pm 0.001}}$
Exchange	3	$0.034_{\pm 0.003}$	$0.132_{\pm 0.005}$	$\mathbf{0.029_{\pm 0.009}}$	$\mathbf{0.121_{\pm 0.020}}$
	6	$\mathbf{0.030_{\pm 0.003}}$	$0.126_{\pm 0.007}$	$\mathbf{0.030_{\pm 0.007}}$	$\mathbf{0.124_{\pm 0.016}}$
	12	$0.039_{\pm 0.004}$	$0.143_{\pm 0.007}$	$\mathbf{0.032_{\pm 0.005}}$	$\mathbf{0.129_{\pm 0.008}}$
Traffic	3	$0.555_{\pm 0.005}$	$0.385_{\pm 0.007}$	$\mathbf{0.520_{\pm 0.002}}$	$\mathbf{0.352_{\pm 0.002}}$
	6	$0.555_{\pm 0.001}$	$0.377_{\pm 0.003}$	$\mathbf{0.531_{\pm 0.001}}$	$\mathbf{0.351_{\pm 0.003}}$
	12	$0.557_{\pm 0.004}$	$0.375_{\pm 0.004}$	$\mathbf{0.530_{\pm 0.006}}$	$\mathbf{0.343_{\pm 0.004}}$

J WaveBound in Spatial-Temporal Time Series Forecasting

To examine the applicability of our WaveBound method in spatial-temporal time series forecasting, we examine Graph WaveNet [42] trained with and without WaveBound on a widely used traffic dataset, METR-LA [1]. As in Wu *et al.* [42], Graph WaveNet is trained with the mean absolute error and attempt to predict a future traffic speed of 1 hour given a historical data of 1 hour along with a graph structure. Table 9 shows that WaveBound improves the performance of Graph WaveNet in terms of all metrics, mean absolute error (MAE), root mean squared error (RMSE), and mean absolute precentage error (MAPE).

Table 9: Spatial-temporal forecasting results of WaveBound adopted to Graph WaveNet. All results are averaged over 3 trials.

Models	15min			30min			60min		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
Graph WaveNet [42]	2.70	5.17	6.93%	3.10	6.24	8.37%	3.58	7.44	10.10%
+ WaveBound	**2.67**	**5.11**	**6.91%**	**3.03**	**6.10**	**8.32%**	**3.45**	**7.15**	**9.90%**

K Ablation Study on EMA Methods

Ablation study on the effect of EMA methods is shown in the Table 10. In the experiments, models only with EMA update (Without Bound) shows the slight performance improvements compared to the original model. However, EMA with dynamic error bound (WaveBound (Indiv.)) outperforms other baselines with a large margin, which concretely shows the effectiveness of individual error bounding in WaveBound.

L WaveBound with Small-Scale Model

We also evaluate the performance of WaveBound with the small-scale model. Specifically, we train a multi layer perceptron (MLP) model consisting of three linear layers with and without WaveBound in the univariate setting. Table 11 shows that WaveBound improves the performance of this MLP model.

Table 10: Results of EMA methods on the ECL dataset. All results are averaged over 3 trials. Subscripts denote standard deviations.

Models		Origin		Without Bound		WaveBound (Avg.)		**WaveBound (Indiv.)**	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Autoformer	96	0.202$\pm$0.003	0.317$\pm$0.003	0.193$\pm$0.001	0.308$\pm$0.001	0.194$\pm$0.001	0.309$\pm$0.001	**0.176$\pm$0.003**	**0.288$\pm$0.003**
	336	0.247$\pm$0.011	0.351$\pm$0.009	0.224$\pm$0.006	0.334$\pm$0.003	0.221$\pm$0.009	0.331$\pm$0.006	**0.217$\pm$0.006**	**0.327$\pm$0.005**
Pyraformer	96	0.256$\pm$0.002	0.360$\pm$0.001	0.247$\pm$0.002	0.351$\pm$0.002	0.248$\pm$0.002	0.352$\pm$0.002	**0.241$\pm$0.001**	**0.345$\pm$0.001**
	336	0.278$\pm$0.007	0.383$\pm$0.006	0.287$\pm$0.009	0.388$\pm$0.010	0.288$\pm$0.011	0.388$\pm$0.010	**0.269$\pm$0.005**	**0.371$\pm$0.005**
Informer	96	0.335$\pm$0.008	0.417$\pm$0.004	0.305$\pm$0.005	0.391$\pm$0.006	0.302$\pm$0.004	0.388$\pm$0.003	**0.289$\pm$0.003**	**0.378$\pm$0.002**
	336	0.369$\pm$0.011	0.448$\pm$0.009	0.326$\pm$0.017	0.411$\pm$0.013	0.322$\pm$0.010	0.407$\pm$0.008	**0.305$\pm$0.008**	**0.394$\pm$0.008**
LSTNet	96	0.268$\pm$0.004	0.366$\pm$0.003	0.201$\pm$0.002	0.311$\pm$0.003	0.208$\pm$0.010	0.314$\pm$0.010	**0.185$\pm$0.003**	**0.291$\pm$0.003**
	336	0.284$\pm$0.001	0.382$\pm$0.002	0.241$\pm$0.004	0.352$\pm$0.004	0.246$\pm$0.009	0.356$\pm$0.009	**0.217$\pm$0.004**	**0.326$\pm$0.005**

Table 11: Results of the MLP model consisting of three linear layers with and without WaveBound in the univariate setting. All results are averaged over 3 trials. Subscripts denote standard deviations.

Models		Origin		**w/ Ours**	
Metric		MSE	MAE	MSE	MAE
ETTm2	96	0.071$\pm$0.001	0.195$\pm$0.002	**0.068$\pm$0.000**	**0.191$\pm$0.001**
	192	0.105$\pm$0.004	0.243$\pm$0.005	**0.104$\pm$0.000**	**0.241$\pm$0.001**
	336	0.141$\pm$0.009	0.289$\pm$0.010	**0.136$\pm$0.001**	**0.284$\pm$0.001**
	720	0.207$\pm$0.010	0.357$\pm$0.009	**0.194$\pm$0.002**	**0.343$\pm$0.002**
ECL	96	0.325$\pm$0.003	0.401$\pm$0.001	**0.317$\pm$0.003**	**0.392$\pm$0.002**
	192	0.334$\pm$0.001	0.408$\pm$0.001	**0.330$\pm$0.004**	**0.401$\pm$0.001**
	336	0.389$\pm$0.007	0.446$\pm$0.007	**0.380$\pm$0.002**	**0.436$\pm$0.001**
	720	0.441$\pm$0.005	0.486$\pm$0.002	**0.438$\pm$0.002**	**0.481$\pm$0.001**

M WaveBound with Generative Model

To show the applicability of WaveBound when training a model with a loss function other than MSE we mainly considered in Section 3, we conduct an experiment with a generative model DeepAR [43] which uses a negative log likelihood (NLL) as a training objective. Table 12 shows that WaveBound improves the performance of DeepAR in the univariate setting.

Table 12: Results of DeepAR with and without WaveBound. All results are averaged over 3 trials. Subscripts denote standard deviations.

Models		DeepAR			
		Origin		**w/ Ours**	
Metric		NLL	MSE	MLL	MSE
ECL	8	1.059$\pm$0.152	0.447$\pm$0.168	**0.935$\pm$0.045**	**0.306$\pm$0.001**
	16	1.168$\pm$0.101	0.500$\pm$0.208	**1.148$\pm$0.052**	**0.371$\pm$0.036**
	24	1.256$\pm$0.073	0.617$\pm$0.119	**1.180$\pm$0.062**	**0.394$\pm$0.043**

N Computational Cost Analysis

WaveBound introduces additional computation and memory costs in the training phase as it employs the EMA model during training. In this section, we analyze the computation and memory cost of WaveBound on the ETTm2 dataset. The input and output lengths are fixed to 96 and a single TITAN RTX GPU is used for all experiments. Table 13 shows the computation time per epoch and maximum GPU memory occupied in the training phase. We also remark that WaveBound does not introduce the additional computation and memory costs in the inference phase.

Table 13: Computation cost analysis of WaveBound.

	Training time (Sec / Epoch)		GPU memory (GB)	
Model	Origin	WaveBound	Origin	WaveBound
Autoformer	67.722	84.075	2.13	2.20
Pyraformer	27.024	39.083	0.47	0.50
Informer	75.251	97.154	1.80	1.86
LSTNet	20.380	23.347	0.02	0.02

O Limitations

While we observe the success of WaveBound with various baselines on real-world benchmarks, our method inevitably introduces the additional cost in the training procedure since our method utilizes the exponential moving average network for estimating the difficulty of predictions. Note that our method performs as the regularization technique in training and does not introduce any additional cost during inference.

P Potential Societal Impact

Our proposed method does not introduce any foreseeable societal consequence. The real-world benchmarks used in this work are widely used for research purposes.

Table 14: Results of WaveBound in the multivariate setting. All results are averaged over 3 trials. Subscripts denote standard deviations.

Models		Autoformer [5]				Pyraformer [6]				Informer [7]				LSTNet [14]				TCN [16]			
		Origin		**w/ Ours**		Origin		**w/ Ours**		Origin		**w/ Ours**		Origin		**w/ Ours**		Origin		**w/ Ours**	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.443 ±0.019	0.453 ±0.012	**0.395 ±0.011**	**0.423 ±0.007**	0.637 ±0.017	**0.592 ±0.007**	**0.630 ±0.011**	0.598 ±0.013	0.886 ±0.020	0.730 ±0.014	**0.841 ±0.013**	**0.708 ±0.010**	0.662 ±0.022	0.588 ±0.009	**0.660 ±0.007**	**0.586 ±0.002**	0.704 ±0.095	0.642 ±0.056	**0.692 ±0.009**	**0.638 ±0.005**
	192	0.509 ±0.032	0.486 ±0.020	**0.433 ±0.002**	**0.444 ±0.001**	0.814 ±0.033	**0.708 ±0.020**	**0.807 ±0.050**	**0.708 ±0.034**	0.988 ±0.014	0.781 ±0.009	**0.959 ±0.015**	**0.755 ±0.010**	0.771 ±0.008	0.652 ±0.008	**0.716 ±0.004**	**0.617 ±0.001**	0.831 ±0.027	0.722 ±0.019	**0.830 ±0.003**	**0.711 ±0.002**
	336	0.519 ±0.024	0.502 ±0.017	**0.470 ±0.005**	**0.476 ±0.004**	0.903 ±0.024	0.749 ±0.014	**0.859 ±0.026**	**0.728 ±0.016**	1.159 ±0.073	0.861 ±0.027	**1.097 ±0.040**	**0.826 ±0.019**	0.882 ±0.004	0.712 ±0.002	**0.782 ±0.004**	**0.664 ±0.007**	**0.894 ±0.016**	**0.739 ±0.007**	0.935 ±0.004	0.758 ±0.003
	720	0.516 ±0.013	0.517 ±0.009	**0.503 ±0.017**	**0.506 ±0.010**	0.932 ±0.007	0.767 ±0.004	**0.919 ±0.011**	**0.763 ±0.006**	1.178 ±0.020	**0.861 ±0.006**	**1.164 ±0.002**	0.867 ±0.003	1.030 ±0.034	0.804 ±0.013	**0.856 ±0.017**	**0.721 ±0.010**	1.139 ±0.030	0.824 ±0.013	**1.067 ±0.012**	**0.793 ±0.004**
ETTh2	96	0.392 ±0.033	0.417 ±0.016	**0.338 ±0.002**	**0.379 ±0.002**	1.453 ±0.254	0.938 ±0.083	**1.269 ±0.015**	**0.912 ±0.007**	2.967 ±0.296	1.349 ±0.068	**2.261 ±0.357**	**1.178 ±0.098**	1.189 ±0.017	0.876 ±0.018	**1.011 ±0.107**	**0.767 ±0.057**	3.038 ±0.217	1.399 ±0.059	**2.921 ±0.203**	**1.370 ±0.040**
	192	0.467 ±0.023	0.459 ±0.013	**0.420 ±0.005**	**0.428 ±0.003**	5.147 ±0.708	1.822 ±0.137	**2.072 ±0.075**	**1.172 ±0.029**	5.822 ±0.644	1.999 ±0.121	**5.361 ±0.646**	**1.929 ±0.115**	2.227 ±0.436	1.193 ±0.160	**1.612 ±0.075**	**0.991 ±0.014**	6.275 ±0.909	2.096 ±0.170	**5.060 ±0.762**	**1.865 ±0.171**
	336	0.472 ±0.004	0.475 ±0.003	**0.452 ±0.003**	**0.459 ±0.002**	4.381 ±0.188	1.736 ±0.048	**2.299 ±0.094**	**1.273 ±0.041**	4.926 ±0.260	**1.864 ±0.051**	**4.700 ±0.181**	1.878 ±0.046	2.274 ±0.111	1.235 ±0.030	**1.697 ±0.207**	**1.041 ±0.094**	6.887 ±0.936	2.245 ±0.149	**4.263 ±0.291**	**1.747 ±0.073**
	720	0.485 ±0.010	0.496 ±0.009	**0.454 ±0.009**	**0.468 ±0.003**	4.304 ±0.530	1.764 ±0.110	**2.810 ±0.098**	**1.458 ±0.033**	3.885 ±0.351	1.667 ±0.089	**3.791 ±0.168**	**1.642 ±0.057**	2.447 ±0.079	1.286 ±0.020	**1.861 ±0.062**	**1.114 ±0.027**	8.993 ±2.254	2.355 ±0.321	**3.299 ±0.073**	**1.504 ±0.010**
ETTm1	96	0.491 ±0.028	0.467 ±0.013	**0.385 ±0.002**	**0.416 ±0.003**	**0.480 ±0.015**	**0.490 ±0.010**	0.495 ±0.017	0.494 ±0.009	0.654 ±0.044	0.576 ±0.016	**0.590 ±0.008**	**0.543 ±0.002**	0.607 ±0.010	0.541 ±0.009	**0.508 ±0.021**	**0.484 ±0.009**	0.603 ±0.031	0.558 ±0.013	**0.556 ±0.008**	**0.531 ±0.006**
	192	0.571 ±0.013	0.504 ±0.008	**0.500 ±0.002**	**0.472 ±0.001**	0.572 ±0.021	0.544 ±0.021	**0.525 ±0.013**	**0.523 ±0.008**	0.783 ±0.050	0.658 ±0.023	**0.672 ±0.012**	**0.595 ±0.011**	0.615 ±0.019	0.559 ±0.007	**0.563 ±0.007**	**0.525 ±0.002**	0.702 ±0.014	0.610 ±0.002	**0.632 ±0.009**	**0.577 ±0.006**
	336	0.611 ±0.029	0.526 ±0.008	**0.497 ±0.047**	**0.475 ±0.015**	0.750 ±0.081	0.650 ±0.031	**0.629 ±0.023**	**0.592 ±0.012**	1.097 ±0.046	0.810 ±0.024	**0.889 ±0.016**	**0.724 ±0.009**	0.657 ±0.009	0.593 ±0.006	**0.652 ±0.053**	**0.581 ±0.015**	0.881 ±0.034	0.705 ±0.012	**0.749 ±0.005**	**0.643 ±0.005**
	720	0.605 ±0.031	0.527 ±0.010	**0.507 ±0.044**	**0.488 ±0.017**	0.949 ±0.022	0.749 ±0.005	**0.806 ±0.052**	**0.691 ±0.028**	1.277 ±0.036	0.855 ±0.010	**1.095 ±0.022**	**0.797 ±0.004**	0.785 ±0.050	0.667 ±0.022	**0.720 ±0.013**	**0.628 ±0.015**	1.047 ±0.023	0.767 ±0.010	**0.910 ±0.008**	**0.714 ±0.003**
ETTm2	96	0.262 ±0.037	0.326 ±0.014	**0.204 ±0.001**	**0.285 ±0.001**	0.363 ±0.089	0.451 ±0.064	**0.281 ±0.026**	**0.386 ±0.021**	0.376 ±0.056	0.477 ±0.047	**0.334 ±0.025**	**0.429 ±0.026**	0.455 ±0.096	0.511 ±0.067	**0.268 ±0.006**	**0.368 ±0.006**	0.537 ±0.128	0.555 ±0.080	**0.480 ±0.039**	**0.524 ±0.020**
	192	0.284 ±0.003	0.342 ±0.002	**0.265 ±0.002**	**0.322 ±0.002**	0.708 ±0.044	0.648 ±0.023	**0.624 ±0.011**	**0.599 ±0.006**	0.751 ±0.020	0.672 ±0.004	**0.698 ±0.056**	**0.631 ±0.013**	0.706 ±0.102	0.660 ±0.064	**0.464 ±0.027**	**0.508 ±0.020**	0.928 ±0.110	0.745 ±0.054	**0.812 ±0.030**	**0.708 ±0.013**
	336	0.338 ±0.007	0.374 ±0.005	**0.320 ±0.001**	**0.356 ±0.001**	1.130 ±0.086	0.846 ±0.026	**1.072 ±0.016**	**0.829 ±0.008**	1.440 ±0.180	0.917 ±0.067	**1.087 ±0.064**	**0.845 ±0.026**	1.161 ±0.066	0.868 ±0.049	**0.781 ±0.012**	**0.695 ±0.009**	1.118 ±0.097	0.816 ±0.037	**1.015 ±0.104**	**0.775 ±0.057**
	720	0.446 ±0.011	0.435 ±0.006	**0.413 ±0.001**	**0.408 ±0.000**	2.995 ±0.095	1.386 ±0.046	**1.917 ±0.074**	**1.119 ±0.028**	3.897 ±0.562	1.498 ±0.128	**2.984 ±0.127**	**1.411 ±0.042**	3.288 ±0.102	1.494 ±0.048	**2.312 ±0.358**	**1.239 ±0.105**	2.687 ±0.302	1.231 ±0.078	**2.120 ±0.042**	**1.092 ±0.011**

Table 15: Results of WaveBound in the multivariate setting. All results are averaged over 3 trials. Subscripts denote standard deviations.

Models		Autoformer [5]				Pyraformer [6]				Informer [7]				LSTNet [14]				TCN [16]			
		Origin		**w/ Ours**		Origin		**w/ Ours**		Origin		**w/ Ours**		Origin		**w/ Ours**		Origin		**w/ Ours**	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Electricity	96	0.202 ±0.003	0.317 ±0.003	**0.176 ±0.003**	**0.288 ±0.003**	0.256 ±0.002	0.360 ±0.001	**0.241 ±0.001**	**0.345 ±0.001**	0.335 ±0.008	0.417 ±0.004	**0.289 ±0.003**	**0.378 ±0.002**	0.268 ±0.004	0.366 ±0.003	**0.185 ±0.003**	**0.291 ±0.003**	0.268 ±0.003	0.363 ±0.003	**0.260 ±0.002**	**0.353 ±0.002**
	192	0.235 ±0.011	0.340 ±0.009	**0.205 ±0.003**	**0.317 ±0.004**	0.272 ±0.003	0.378 ±0.003	**0.256 ±0.002**	**0.360 ±0.002**	0.341 ±0.013	0.426 ±0.011	**0.298 ±0.002**	**0.388 ±0.002**	0.277 ±0.002	0.375 ±0.001	**0.197 ±0.003**	**0.304 ±0.003**	0.295 ±0.006	0.385 ±0.005	**0.272 ±0.003**	**0.358 ±0.002**
	336	0.247 ±0.011	0.351 ±0.009	**0.217 ±0.006**	**0.327 ±0.005**	0.278 ±0.007	0.383 ±0.006	**0.269 ±0.005**	**0.371 ±0.005**	0.369 ±0.011	0.448 ±0.009	**0.305 ±0.008**	**0.394 ±0.008**	0.284 ±0.001	0.382 ±0.002	**0.217 ±0.004**	**0.326 ±0.005**	0.309 ±0.013	0.394 ±0.009	**0.284 ±0.003**	**0.365 ±0.002**
	720	0.270 ±0.006	0.371 ±0.003	**0.260 ±0.024**	**0.359 ±0.018**	0.291 ±0.002	0.385 ±0.001	**0.283 ±0.003**	**0.377 ±0.003**	0.396 ±0.009	0.457 ±0.005	**0.311 ±0.008**	**0.398 ±0.008**	0.316 ±0.002	0.404 ±0.001	**0.250 ±0.001**	**0.350 ±0.002**	0.310 ±0.002	0.388 ±0.002	**0.290 ±0.003**	**0.369 ±0.001**
Exchange	96	0.153 ±0.009	0.285 ±0.009	**0.146 ±0.005**	**0.274 ±0.007**	**0.604 ±0.037**	**0.624 ±0.016**	0.615 ±0.037	0.627 ±0.015	0.979 ±0.045	0.791 ±0.021	**0.878 ±0.053**	**0.765 ±0.032**	0.483 ±0.030	0.518 ±0.017	**0.357 ±0.046**	**0.432 ±0.028**	0.805 ±0.116	0.707 ±0.060	**0.720 ±0.023**	**0.693 ±0.014**
	192	0.297 ±0.030	0.397 ±0.019	**0.262 ±0.001**	**0.373 ±0.001**	0.982 ±0.093	0.806 ±0.029	**0.953 ±0.076**	**0.797 ±0.031**	1.147 ±0.110	**0.854 ±0.042**	**1.136 ±0.021**	0.859 ±0.014	0.706 ±0.042	0.646 ±0.027	**0.621 ±0.210**	**0.593 ±0.111**	0.945 ±0.079	0.795 ±0.042	**0.912 ±0.035**	**0.781 ±0.020**
	336	0.438 ±0.018	0.490 ±0.011	**0.425 ±0.001**	**0.483 ±0.001**	1.264 ±0.131	**0.934 ±0.052**	**1.263 ±0.043**	0.944 ±0.014	1.592 ±0.067	1.014 ±0.016	**1.461 ±0.051**	**0.992 ±0.014**	1.055 ±0.124	0.800 ±0.049	**0.837 ±0.055**	**0.691 ±0.027**	1.350 ±0.105	0.962 ±0.035	**1.135 ±0.045**	**0.889 ±0.023**
	720	1.207 ±0.080	0.860 ±0.034	**1.088 ±0.004**	**0.810 ±0.002**	1.663 ±0.083	1.051 ±0.019	**1.562 ±0.097**	**1.016 ±0.012**	2.540 ±0.089	1.306 ±0.027	**2.496 ±0.253**	**1.294 ±0.071**	2.198 ±0.656	1.127 ±0.189	**1.374 ±0.139**	**0.894 ±0.040**	2.061 ±0.226	1.152 ±0.076	**1.689 ±0.162**	**1.067 ±0.045**
Traffic	96	0.645 ±0.011	0.399 ±0.012	**0.596 ±0.027**	**0.352 ±0.013**	0.635 ±0.006	0.364 ±0.006	**0.622 ±0.003**	**0.341 ±0.003**	0.731 ±0.010	0.412 ±0.001	**0.671 ±0.007**	**0.364 ±0.004**	0.735 ±0.008	0.446 ±0.003	**0.587 ±0.004**	**0.356 ±0.003**	0.645 ±0.007	0.361 ±0.005	**0.633 ±0.004**	**0.333 ±0.003**
	192	0.644 ±0.027	0.407 ±0.021	**0.607 ±0.013**	**0.370 ±0.014**	0.658 ±0.006	0.376 ±0.005	**0.646 ±0.001**	**0.355 ±0.000**	0.751 ±0.007	0.422 ±0.007	**0.666 ±0.005**	**0.360 ±0.005**	0.750 ±0.001	0.446 ±0.000	**0.595 ±0.005**	**0.365 ±0.004**	0.646 ±0.001	0.354 ±0.001	**0.625 ±0.002**	**0.320 ±0.001**
	336	0.625 ±0.021	0.390 ±0.019	**0.603 ±0.019**	**0.361 ±0.017**	0.668 ±0.003	0.377 ±0.001	**0.653 ±0.002**	**0.355 ±0.002**	0.822 ±0.003	0.465 ±0.003	**0.709 ±0.014**	**0.387 ±0.009**	0.778 ±0.002	0.455 ±0.000	**0.623 ±0.019**	**0.378 ±0.014**	0.661 ±0.005	0.363 ±0.003	**0.630 ±0.001**	**0.320 ±0.001**
	720	0.650 ±0.008	0.398 ±0.004	**0.642 ±0.011**	**0.383 ±0.007**	0.698 ±0.002	0.390 ±0.001	**0.672 ±0.001**	**0.364 ±0.002**	0.957 ±0.038	0.539 ±0.024	**0.764 ±0.025**	**0.421 ±0.014**	0.815 ±0.002	0.470 ±0.002	**0.648 ±0.008**	**0.383 ±0.009**	0.655 ±0.001	0.358 ±0.001	**0.634 ±0.000**	**0.322 ±0.000**
Weather	96	0.294 ±0.014	0.355 ±0.011	**0.227 ±0.008**	**0.296 ±0.010**	0.235 ±0.010	0.321 ±0.010	**0.193 ±0.007**	**0.272 ±0.005**	0.378 ±0.029	0.428 ±0.016	**0.355 ±0.014**	**0.415 ±0.009**	0.237 ±0.004	0.310 ±0.002	**0.202 ±0.007**	**0.275 ±0.004**	0.447 ±0.008	0.481 ±0.008	**0.412 ±0.013**	**0.457 ±0.010**
	192	0.308 ±0.004	0.368 ±0.005	**0.283 ±0.005**	**0.340 ±0.005**	0.340 ±0.041	0.415 ±0.032	**0.306 ±0.015**	**0.372 ±0.006**	0.462 ±0.012	0.467 ±0.002	**0.424 ±0.003**	**0.448 ±0.007**	0.277 ±0.003	0.343 ±0.003	**0.254 ±0.000**	**0.316 ±0.002**	**0.604 ±0.023**	**0.558 ±0.012**	0.618 ±0.006	**0.558 ±0.008**
	336	0.364 ±0.007	0.396 ±0.008	**0.335 ±0.013**	**0.370 ±0.010**	0.453 ±0.038	0.484 ±0.028	**0.403 ±0.055**	**0.441 ±0.037**	0.575 ±0.010	0.535 ±0.005	**0.506 ±0.017**	**0.484 ±0.009**	0.326 ±0.007	0.378 ±0.004	**0.309 ±0.015**	**0.358 ±0.012**	0.753 ±0.017	0.632 ±0.007	**0.689 ±0.049**	**0.587 ±0.024**
	720	0.426 ±0.014	0.433 ±0.013	**0.401 ±0.001**	**0.411 ±0.002**	0.599 ±0.045	0.563 ±0.025	**0.535 ±0.011**	**0.519 ±0.007**	1.024 ±0.067	0.751 ±0.025	**0.972 ±0.065**	**0.712 ±0.030**	0.412 ±0.007	0.431 ±0.007	**0.398 ±0.006**	**0.415 ±0.004**	0.947 ±0.050	0.722 ±0.019	**0.876 ±0.075**	**0.661 ±0.033**
ILI	24	3.468 ±0.252	1.299 ±0.058	**3.118 ±0.036**	**1.200 ±0.019**	4.822 ±0.081	1.489 ±0.015	**4.679 ±0.087**	**1.459 ±0.024**	5.356 ±0.062	1.590 ±0.016	**4.947 ±0.053**	**1.494 ±0.017**	7.934 ±0.304	2.091 ±0.031	**6.331 ±0.231**	**1.816 ±0.029**	4.561 ±0.065	1.535 ±0.015	**4.093 ±0.060**	**1.378 ±0.030**
	36	3.441 ±0.139	1.273 ±0.034	**3.310 ±0.192**	**1.240 ±0.049**	4.831 ±0.196	**1.479 ±0.041**	**4.763 ±0.198**	1.483 ±0.044	5.131 ±0.046	1.569 ±0.006	**5.027 ±0.081**	**1.537 ±0.022**	8.793 ±0.428	2.214 ±0.047	**6.560 ±0.629**	**1.848 ±0.088**	4.376 ±0.249	1.463 ±0.077	**4.213 ±0.361**	**1.425 ±0.107**
	48	3.086 ±0.176	1.184 ±0.036	**2.927 ±0.017**	**1.128 ±0.010**	4.789 ±0.153	1.465 ±0.036	**4.524 ±0.132**	**1.439 ±0.029**	5.150 ±0.037	1.564 ±0.010	**4.920 ±0.047**	**1.514 ±0.009**	7.968 ±0.255	2.068 ±0.026	**6.154 ±0.110**	**1.779 ±0.019**	4.264 ±0.085	1.418 ±0.022	**3.963 ±0.221**	**1.340 ±0.062**
	60	2.843 ±0.025	1.136 ±0.010	**2.785 ±0.021**	**1.116 ±0.003**	4.876 ±0.027	1.495 ±0.006	**4.573 ±0.158**	**1.465 ±0.026**	5.407 ±0.033	1.604 ±0.011	**5.013 ±0.059**	**1.528 ±0.016**	7.387 ±0.269	1.984 ±0.036	**6.119 ±0.186**	**1.758 ±0.031**	4.288 ±0.105	1.401 ±0.033	**3.970 ±0.108**	**1.335 ±0.032**

Table 16: Results of WaveBound in the univariate setting. All results are averaged over 3 trials. Subscripts denote standard deviations.

Models		Autoformer [5]				Pyraformer [6]				Informer [7]				N-BEATS [15]				LSTNet [14]				TCN [16]			
		Origin		**w/ Ours**		Origin		**w/ Ours**		Origin		**w/ Ours**		Origin		**w/ Ours**		Origin		**w/ Ours**		Origin		**w/ Ours**	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTm2	96	0.098 ±0.016	0.239 ±0.020	**0.085 ±0.004**	**0.221 ±0.005**	0.078 ±0.007	0.209 ±0.011	**0.070 ±0.002**	**0.197 ±0.003**	0.085 ±0.008	0.224 ±0.012	**0.081 ±0.003**	**0.218 ±0.004**	0.073 ±0.004	0.198 ±0.006	**0.067 ±0.001**	**0.188 ±0.002**	0.091 ±0.008	0.228 ±0.011	**0.067 ±0.001**	**0.188 ±0.001**	0.082 ±0.011	0.218 ±0.017	**0.073 ±0.001**	**0.207 ±0.001**
	192	0.130 ±0.012	0.277 ±0.012	**0.116 ±0.008**	**0.262 ±0.008**	0.114 ±0.003	0.257 ±0.004	**0.110 ±0.004**	**0.256 ±0.004**	0.122 ±0.007	0.273 ±0.009	**0.118 ±0.001**	**0.270 ±0.002**	0.107 ±0.007	0.246 ±0.010	**0.103 ±0.001**	**0.241 ±0.001**	0.125 ±0.005	0.272 ±0.006	**0.108 ±0.001**	**0.248 ±0.002**	0.142 ±0.016	0.295 ±0.015	**0.113 ±0.002**	**0.265 ±0.003**
	336	0.162 ±0.005	0.311 ±0.003	**0.143 ±0.008**	**0.293 ±0.008**	0.178 ±0.007	0.325 ±0.003	**0.153 ±0.006**	**0.306 ±0.005**	0.153 ±0.004	**0.304 ±0.003**	**0.148 ±0.001**	0.305 ±0.002	0.163 ±0.028	0.310 ±0.027	**0.135 ±0.001**	**0.284 ±0.001**	0.171 ±0.002	0.322 ±0.002	**0.159 ±0.004**	**0.308 ±0.004**	0.157 ±0.022	**0.314 ±0.026**	**0.154 ±0.002**	**0.314 ±0.003**
	720	0.194 ±0.001	0.344 ±0.002	**0.188 ±0.004**	**0.338 ±0.002**	0.198 ±0.024	0.351 ±0.016	**0.169 ±0.005**	**0.329 ±0.005**	0.196 ±0.005	0.351 ±0.006	**0.189 ±0.003**	**0.349 ±0.003**	0.263 ±0.033	0.402 ±0.026	**0.188 ±0.006**	**0.340 ±0.006**	0.253 ±0.003	0.395 ±0.003	**0.242 ±0.003**	**0.387 ±0.002**	0.248 ±0.038	0.404 ±0.034	**0.193 ±0.002**	**0.360 ±0.001**
Electricity	96	0.462 ±0.023	0.502 ±0.014	**0.447 ±0.029**	**0.496 ±0.022**	0.240 ±0.002	0.351 ±0.001	**0.229 ±0.002**	**0.347 ±0.002**	0.266 ±0.005	0.371 ±0.003	**0.261 ±0.006**	**0.369 ±0.005**	0.304 ±0.001	0.382 ±0.001	**0.298 ±0.005**	**0.378 ±0.002**	0.452 ±0.011	0.505 ±0.006	**0.340 ±0.022**	**0.419 ±0.013**	0.235 ±0.003	0.357 ±0.001	**0.226 ±0.005**	**0.346 ±0.004**
	192	0.557 ±0.016	0.565 ±0.013	**0.515 ±0.031**	**0.538 ±0.017**	0.262 ±0.005	0.367 ±0.001	**0.253 ±0.001**	**0.365 ±0.001**	0.283 ±0.003	0.385 ±0.003	**0.281 ±0.006**	**0.383 ±0.003**	0.323 ±0.002	0.396 ±0.001	**0.322 ±0.004**	**0.395 ±0.002**	0.419 ±0.007	0.477 ±0.003	**0.327 ±0.005**	**0.402 ±0.001**	0.284 ±0.008	0.389 ±0.007	**0.272 ±0.004**	**0.382 ±0.003**
	336	0.613 ±0.026	0.593 ±0.015	**0.531 ±0.044**	**0.543 ±0.028**	0.285 ±0.002	**0.386 ±0.002**	**0.283 ±0.002**	**0.386 ±0.002**	0.338 ±0.005	0.428 ±0.003	**0.332 ±0.005**	**0.426 ±0.004**	0.385 ±0.004	0.430 ±0.001	**0.369 ±0.001**	**0.422 ±0.000**	0.439 ±0.011	0.488 ±0.006	**0.352 ±0.003**	**0.420 ±0.001**	0.378 ±0.015	0.451 ±0.009	**0.327 ±0.012**	**0.419 ±0.007**
	720	0.691 ±0.045	0.632 ±0.019	**0.604 ±0.009**	**0.591 ±0.004**	0.309 ±0.002	**0.411 ±0.002**	**0.307 ±0.016**	0.415 ±0.016	0.631 ±0.008	0.612 ±0.004	**0.378 ±0.005**	**0.463 ±0.004**	0.462 ±0.009	0.487 ±0.004	**0.433 ±0.003**	**0.473 ±0.001**	0.492 ±0.004	0.528 ±0.003	**0.428 ±0.032**	**0.478 ±0.016**	0.435 ±0.044	0.490 ±0.026	**0.399 ±0.060**	**0.471 ±0.035**