Supplementary

I. Proof of Interval Construction

In the main paper, we use t-distribution to generate intervals. Here we show the detailed statistical inference.

Given the assumption that stock price at an individual time step is a random variable $Y_t \sim \mathcal{N}(\mu_t, \sigma_t^2)$, and the set $\tilde{\boldsymbol{Y}}_t$ with $\tilde{N} = N \times H$ samples, the sample mean $\bar{Y}_t$ and sample variance s_t^2 are calculated as:

$$\bar{Y}_t = \frac{1}{\tilde{N}} \sum_{i=1}^{\tilde{N}} \tilde{y}_{t,i}, \quad s_t^2 = \frac{1}{\tilde{N}-1} \sum_{i=1}^{\tilde{N}} (\tilde{y}_{t,i} - \bar{Y}_t)^2 \tag{1}$$

where $\tilde{y}_{t,i}$ is i-th sample in the set, $\tilde{N}-1$ of s_t^2 is for unbias correction.

For the future observation $Y_{t,\tilde{N}+1}$, we have:

$$Y_{t,\tilde{N}+1} \sim \mathcal{N}(\mu_t, \sigma_t^2), \quad \bar{Y}_t \sim \mathcal{N}\left(\mu_t, \frac{\sigma_t^2}{\tilde{N}}\right) \tag{2}$$

$$Y_{t,\tilde{N}+1} - \bar{Y}_t \sim \mathcal{N}\left(0, \sigma_t^2\left(1 + \frac{1}{\tilde{N}}\right)\right) \tag{3}$$

$$\frac{Y_{t,\tilde{N}+1} - \bar{Y}_t}{\sigma_t\sqrt{1+\frac{1}{\tilde{N}}}} \sim \mathcal{N}(0,1) \tag{4}$$

$$(\tilde{N}-1)\frac{s_t^2}{\sigma_t^2} \sim \chi^2_{(\tilde{N}-1)} \tag{5}$$

$$\frac{Y_{t,\tilde{N}+1} - \bar{Y}_t}{s_t\sqrt{1+\frac{1}{\tilde{N}}}} \sim t_{(\tilde{N}-1)} \tag{6}$$

Empirically, the inference above can also be used for the true value of close price y_t. Hence, we can use the percentile of t-distribution T to calculate a prediction interval for future observation $Y_{t,\tilde{N}+1}$, as well as for actual close price:

$$P(-T \leq \frac{Y_{t,\tilde{N}+1} - \bar{Y}_t}{s_t\sqrt{1+\frac{1}{\tilde{N}}}} \leq T) = c \tag{7}$$

The probability density function of t-distribution is symmetric. For a prediction interval with probability of $100c\%$ containing true value, there's $100\left(\frac{1-c}{2}\right)\%$ of probability that the random variable $\frac{Y_{t,\tilde{N}+1} - \bar{Y}_t}{s_t\sqrt{1+\frac{1}{\tilde{N}}}}$ is less than T and same probability that the random variable is larger than T. Therefore, the $(1-\frac{1-c}{2})$-th percentile of t-distribution is used for T [1]. The lower and upper bounds are calculated as follows:

$$\begin{aligned} Y_t^L &= \bar{Y}_t - t_{1-\frac{(1-c)}{2}}(\tilde{N}-1)s_t\sqrt{1+\frac{1}{\tilde{N}}} \\ Y_t^U &= \bar{Y}_t + t_{1-\frac{(1-c)}{2}}(\tilde{N}-1)s_t\sqrt{1+\frac{1}{\tilde{N}}} \end{aligned} \tag{8}$$

II. Details of Dataset

A. Feature Preprocessing

Table I shows calculation details on price-based and volatility-based features preprocessing.

B. Technical Indicators

Denote x_t as the stock price at time t. Besides OHLC (open, high, low, close) and volume, we deliberately choose four types of empirically technical indicators that can be derived from the price in the market and thus are proved to be useful for revealing trading patterns in industry [2]:

- **Moving Average (MA)** smooths out the prices over a specific period. SMA is calculated by taking the arithmetic average of the previous days.

$$SMA_t(n) = \frac{1}{n}\sum_{t}^{i} x_i \tag{9}$$

  EMA is calculated by the weighted average of prices, where recent prices are assigned with higher weight.

$$EMA_t(n) = \frac{2 \times x_t}{1+n} + EMA_{t-1}(n) \times (1 - \frac{2}{1+n}) \tag{10}$$

- **Bollinger Bands** indicate oversold or overbought signals in the market. Upper Bollinger Band (BOLU) and Lower Bollinger Band (BOLD) are calculated as follows:

$$BOLU = SMA(TP, n) + 2 \times \sigma_t(TP, 20) \tag{11}$$

$$BOLD = SMA(TP, n) - 2 \times \sigma_t(TP, 20) \tag{12}$$

  where TP represents the average value of high, low, and close price, $\sigma_t(TP, n)$ denotes the standard deviation of TP in the last n time steps.

- **Relative Strength Index (RSI)** is a popular momentum oscillator that provides bullish and bearish signals.

$$RSI = 100 - (\frac{100}{1+RS}) \tag{13}$$

$$RS = \frac{Gain}{Loss} \tag{14}$$

  where $Gain$ and $Loss$ represent the average percentage gain and loss calculated by comparing close prices in the past periods. An asset with RSI over 0.8 is considered overbought, and one with an RSI of less than 0.2 is believed to be oversold.

- **Moving Average Convergence Divergence(MACD)** is a collection of MACD lines, Signal lines, and Divergence series.

$$MACD = EMA(12) - EMA(26) \tag{15}$$

$$Signal = EMA(9) \tag{16}$$

  If the MACD line crosses above the Signal line, it is considered a buy signal; otherwise, it shows a sell signal.

III. Implementation Details

A. Hyperparameters

Table II shows the range of hyperparameters for tuning during the training process and the final chosen value of hyperparameters for all stock indices. Table III presents the hyperparameters in other baselines.

IV. Supplementary Experiments

A. Comparison on Predicted Intervals from Baselines and RAGIC

The main paper shows the comparison of predicted interval from four baselines only. Here we provide the comparison of all baselines with *RAGIC*., as shown in Figure 1.

B. Visualization of Predicted Intervals from RAGIC

In the main paper, we present the figure of intervals predicted from all benchmarks and *RAGIC* on SPX. Here we also visualize predicted intervals from the proposed model on other indices, as shown in Figure 2. The interval from *RAGIC* can capture the true price of the next day with precise width 95% of the time. DAX and Nikkei show more fluctuations in 2017/6-2019/12 than other indices; hence the normalized width of intervals on these two indices are larger than those of the U.S. market. The high coverage of *RAGIC* is steady over multiple markets worldwide.

References

[1] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. Springer, 2009, vol. 2.

[2] M. Wu and X. Diao, "Technical analysis of three stock oscillators testing macd, rsi and kdj rules in sh & sz stock markets," in *2015 4th International Conference on Computer Science and Network Technology (ICCSNT)*, vol. 1. IEEE, 2015, pp. 320–323.

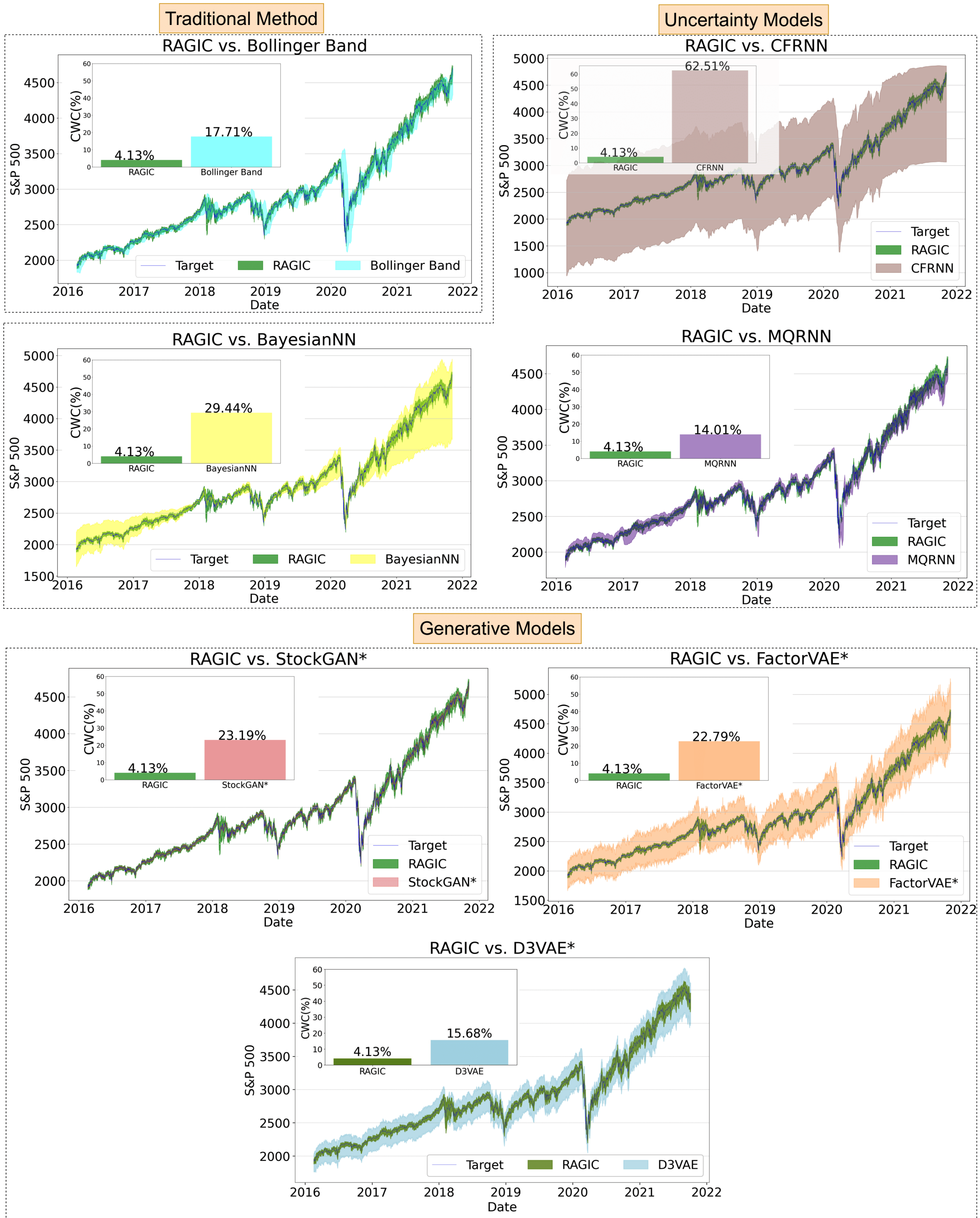

Fig. 1. Comparison of predicted intervals on SPX.

TABLE I
FEATURE PREPROCESSING. x_t DENOTES THE CORRESPONDING FEATURES BEFORE PROCESSING.

Type	Features	Calculation
Price	Open, Close, High, Low	$x_t/close_{t-1} - 1$
	Volume	$x_t/x_{t-1} - 1$
Technical indicator	RSI	-1 if $x_t < 20$, 1 if $x_t \geq 80$, 0 otherwise
	MACD DIFF	min-max scaling
	Upper Bollinger band indicator	1 if $close_t > BOLU$, 0 otherwise
	Lower Bollinger band indicator	1 if $close_t < BOLD$, 0 otherwise
	EMA: 5-day	$x_t/close_t - 1$
	SMA: 13-day, 21-day, 50-day	
Volatility	Volatility index	min-max scaling
	Return of Volatility index	$x_t/x_{t-1} - 1$

TABLE II
HYPERPARAMETERS ON *RAGIC* FOR THE DATASETS.

Parameter	Range	DJI	SPX	Nasdaq	DAX	Nikkei
Epochs	[30,300]			50		
Optimizer	Adam, RMSProp			Adam		
Learning rate of critic	[0.00001, 0.01]			0.0002		
ξ	[0.001,0.01]			0.01		
α_l	90%			90%		
α_u	99.9%			99.9%		
$\Delta\epsilon$	0.0001			0.0001		
N	50			50		
γ	[0.1,2]			0.3		
D	[2,4]			2		
δ_k for volatility return	[0.05,0.25]	0.1	0.05	0.2	0.1	0.1
δ_k for normalized volatility index	[0.05,0.25]	0.25	0.05	0.1	0.05	0.2
λ_k for volatility return	[0.5,4]	2.5	1.5	2.5	1.5	1.5
λ_k for normalized volatility index	[0.5,4]	1.5	1.5	4	2.5	2.5
Batch size	[16,512]	256	256	256	150	256
n_{critic}	[2,6]	4	4	3	6	5
Learning rate of generator	[0.00001, 0.05]	0.0001	0.0001	0.00005	0.0001	0.0001
L	[2,6]	3	2	2	3	3
p	[2,8]	8	5	5	8	8
Dimension of hidden layer	[50, 300]	150	100	100	150	150
v_l	[10,20]	10	10	10	12	10
v_u	[20,30]	22	20	20	22	25

TABLE III
HYPERPARAMETERS ON BASELINES FOR THE DATASETS.

Model	Parameter	Range	DJI	SPX	Nasdaq	DAX	Nikkei
Bollinger bands	Sequence Length	20			20		
	Number of std	2			2		
BayesianNN MQRNN	Epochs	[30,300]			50		
	Learning rate	[0.0001,0.01]			0.01		
	Batch size	[32,256]			100		
	Steps in an epoch	[10,100]			500		
	Optimizer	Adam, RMSProp			Adam		
	Embedding size	[10,100]			20		
	Sequence Length	[5,40]			30		
	Prediction Horizon	[1,10]			5		
	Dropout rate (BayesianNN)	[0,1]			0.5		
CFRNN	Epochs	[100,1500]			1000		
	Learning rate	[0.0001,0.01]			0.01		
	Batch size	[32,256]			100		
	Steps in an epoch	[10,100]			500		
	Optimizer	Adam, RMSProp			Adam		
	Embedding size	[10,100]			20		
	Sequence Length	[5,40]			30		
	Prediction Horizon	[1,10]			5		
	Correction factor	[1,5]			1		
D3VAE	Epochs	[30,50]			50		
	Learning rate	[0.001,0.01]			0.005		
	Batch size	[16,64]			16		
	Optimizer	Adam, RMSProp			Adam		
	Dimension	[32,128]			64		
	Maximum numebr of diffusion steps	[30,100]			50		
StockGAN	Epochs	[30,100]			50		
	Learning rate	[0.0001,0.01]			0.002		
	Batch size	[32,256]			64		
	Optimizer	Adam, RMSProp			Adam		
	Sequence Length	[5,40]			30		
	Layers	[1,4]			2		
	Dimension(layer1)	[10,500]	72	72	100	100	100
	Dimension(layer2)	[10,500]	10	10	10	30	30
FactorVAE	Epochs	[30,100]			50		
	Learning rate	[0.0001,0.01]			0.001		
	Batch size	[32,256]			128		
	Optimizer	Adam, RMSProp			Adam		
	Sequence Length	[5,40]			30		
	Number of head	[1,5]			5		
	Weight of KLD loss	[0,10]			0.2		
	Embedding size	[50,1000]	500	500	500	800	850
HMG-TF	Epochs	[30,100]			50		
	Batch size	[32,256]			256		
	Optimizer	Adam, RMSProp			Adam		
	Sequence Length	[5,40]			20		
	Number of block	[1,5]			3		
	Learning rate	[0.0001,0.01]	0.001	0.001	0.001	0.0005	0.0005
	Number of head	[1,5]	2	2	2	4	4
	Weight of penalty loss	[0,1]	0.05	0.05	0.05	0.02	0.02
AdaRNN	Epochs	[30,300]			200		
	Number of layer	[1,5]			2		
	Batch size	[32,512]			256		
	Optimizer	Adam, RMSProp			Adam		
	Weight of penalty	[0,1]			0.5		
	Learning rate	[0.00001,0.01]			0.00005		
	Hidden size	[30,200]			64		

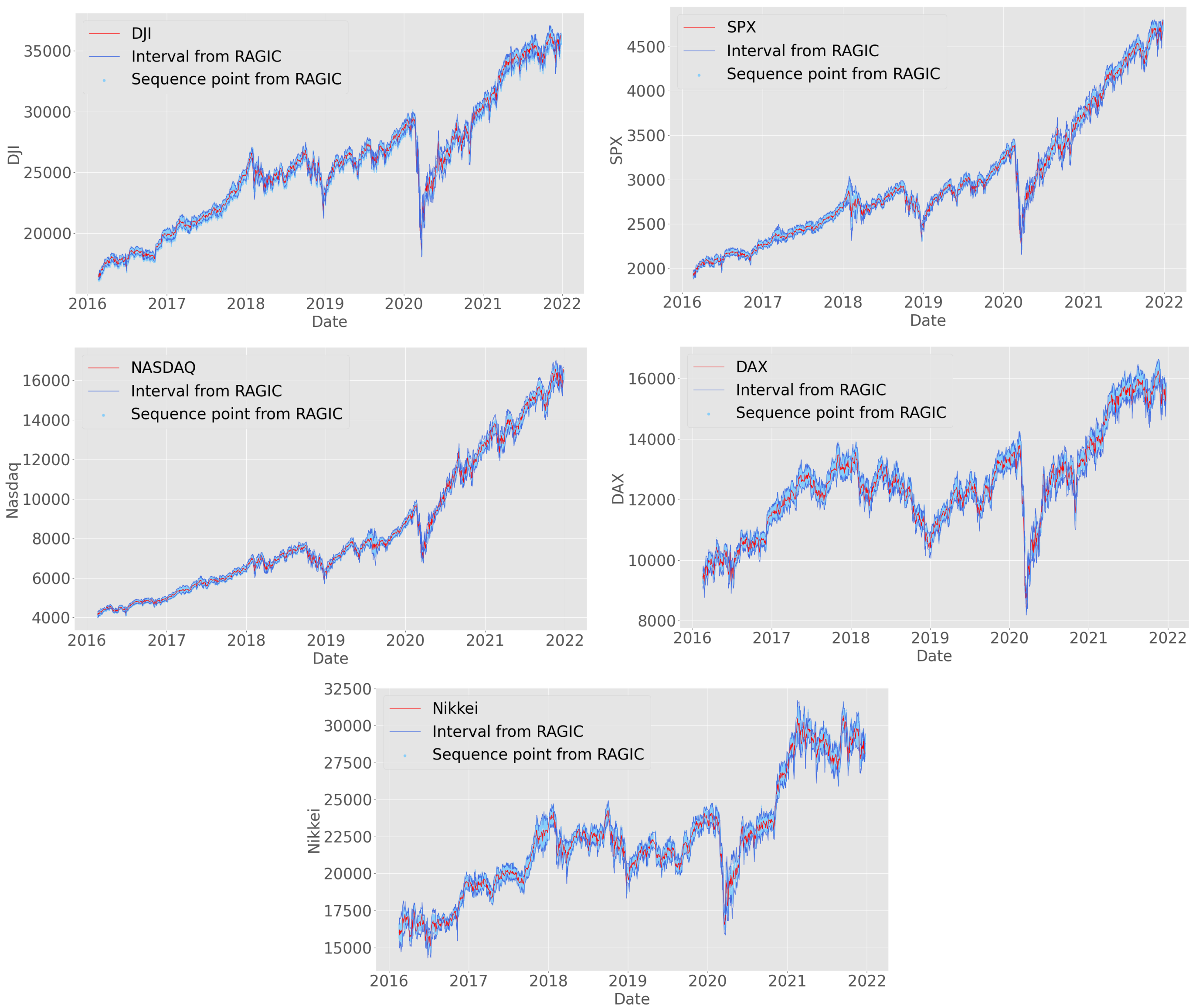

Fig. 2. Predicted intervals on stock indices.