Kronos: A Foundation Model for the Language of Financial Markets

Yu Shi[1,*], Zongliang Fu[2,*], Shuo Chen[1], Bohan Zhao[1], Wei Xu[1], Changshui Zhang[2], Jian Li[1]

[1] Institute for Interdisciplinary Information Sciences [2] Department of Automation, Tsinghua University

NEURAL INFORMATION PROCESSING SYSTEMS

Problem Definition

Learn a **universal, high-fidelity representation for financial K-line** (candlestick, OHLCVA) **time series foundation model** that supports both predictive and generative downstream tasks—principally multistep price/return forecasting, realized volatility forecasting, synthetic K-line generation, and trading simulation.

K-line Seqs: unique statistical properties	*Mainstream TSFM: underserved financial tasks*
✓ low signal-to-noise ratios	✓ Low proportion in pre-training dataset
✓ strong non-stationarities	✓ Limited supports in downstream tasks
✓ high-order dependencies	✓ underperforming than full-shot models

We Need a Foundation Model for Financial K-lines!

Methods

A hierarchical quantization–autoregressive architecture for K-line time-series modeling.

Tokenization Pipeline

- ✓ **Instance-based tokenization:** converts **continuous** K-line vectors into **discrete** tokens.
  1 token = 1 K-line instance
- ✓ **BSQ: binary spherical quantization**
  strict error bounds and strong robustness to anomalies;
  scalability to huge codebook sizes, such as 2^20

Pretraining Pipeline

- ✓ **Hierarchical** discrete representation
  Model learns a **codebook** $\mathcal{C}$, each x_t is quantized into **a coarse token** and **a fine token**:
  $$x_t \to (b_t^c,\ b_t^f)$$
- ✓ **Coarse-to-Fine Factorization Decoder** factorizes the likelihood
  $$p(b_t|b_{<t}) = p(b_t^c|b_{<t}) * p(b_t^f|b_{<t}, b_t^c)$$

Experiments

Datasets

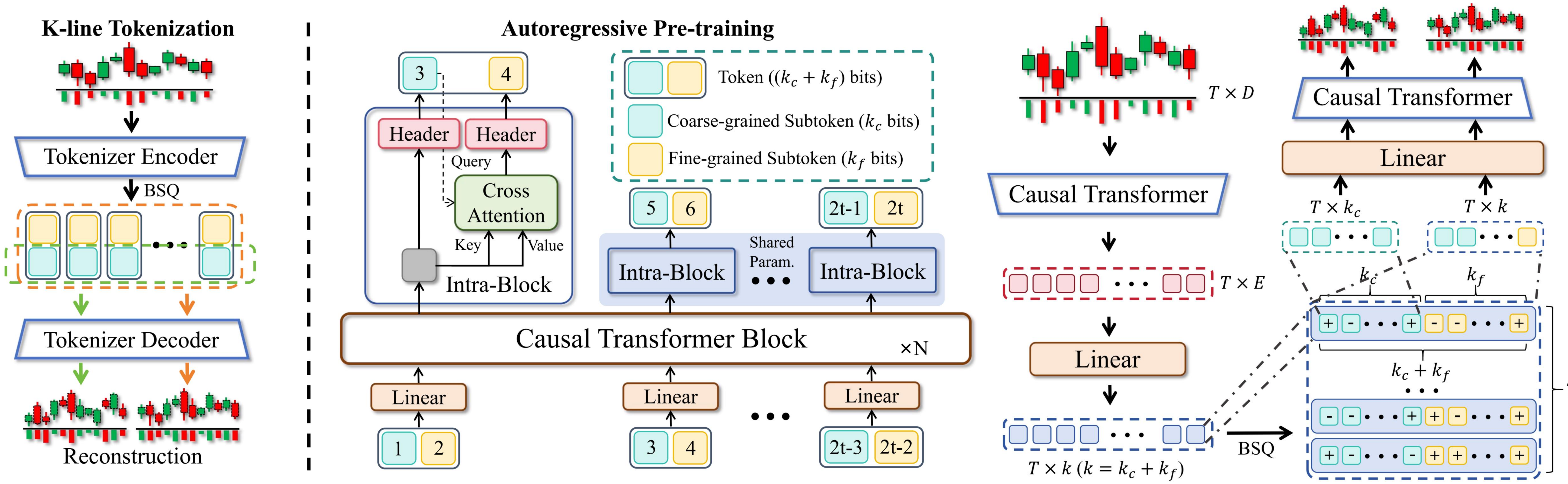
Coverage:
45 global exchanges,
12.1B K-lines
Asset Cate.: stocks, futures, forex, options, and crypto;
Time Granularity: (7)
1,5, 15, 30, 60m, d ,w, 2w
Time Range:
Train/Val: pre ~ Jun. 2024
Test: Jul. 2024 ~ Jun. 2025
Market Range:
ID: NASDAQ, China, etc
OOD: Indonesia, Brazil, etc

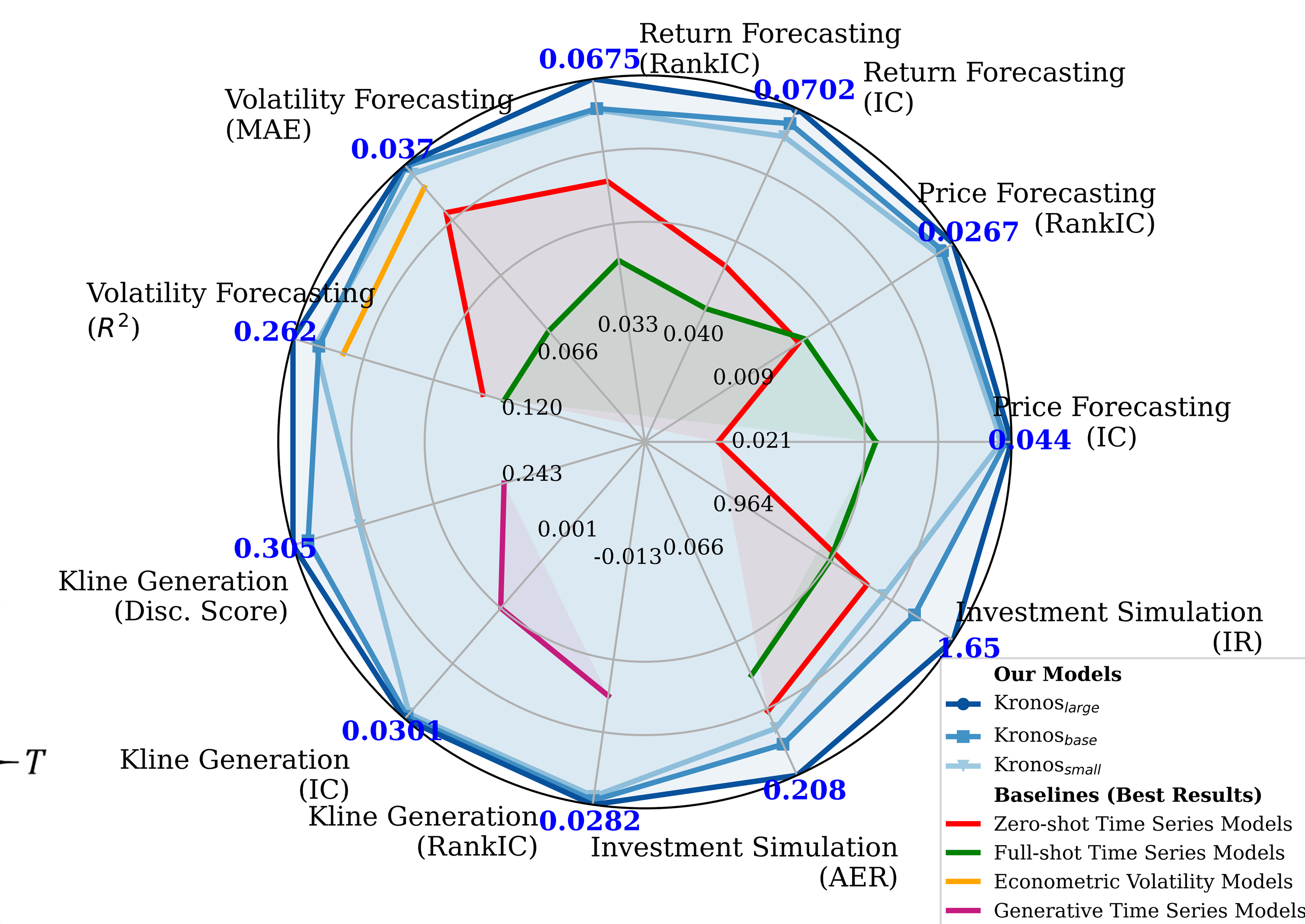
Baselines

Full-shot Time Series Models: (8)
TimeMixer, iTransformer, Dlinear, TimeXer, etc
Zero-shot Time Series Models: (5)
TimeMoE, Moirai, etc.
Econometric Volatility Models: (2)
ARCH, GARCH
Generative Time Series Models: (3)
TimeGAN, DiffuTS, etc

Evaluations

Price/Rtn Forecasting:
IC and RankIC are calculated between the predicted and true series
Realized Volatility Forecasting:
MAE and R2
Synthetic K-line Generation:
Discriminative Score
Investment Simulation:
Cumulative return and Cumulative Excess return





Results

Task	Baseline Type	Best Baseline Model	Best Baseline Perf.	Ours Model	Ours Perf.	Comparison
Price Forecasting	Zero shot	$TimeMoE_B$	0.0138	$Kronos_L$	0.0267	↑93%
	Full shot	$TimesNet$	0.0143			↑87%
Return Forecasting	Zero shot	Moirai	0.0533	$Kronos_L$	0.0675	↑27%
	Full shot	$DLinear$	0.0423			↑60%
Volatility Forecasting	Zero shot	$Moment_B$	0.0444	$Kronos_L$	0.0370	↑17%
	Full shot	$NSTrans.$	0.0627			↑41%
K-line Generation	/	$DiffuTS$	0.2504	$Kronos_L$	0.3408	↓36%
Investment Simulation	Zero shot	$Moment_L$	0.1655	$Kronos_L$	0.2193	↑33%
	Full shot	$TimesNet$	0.1441			↑52%

Insights

Discrete

***Discrete** tokenization improves sample efficiency and generalization, as well as reducing noise.*

BSQ

***B**inary **S**pherical **Q**uantization (BSQ) + hierarchical subtokens yield robustness and bounded error.*

Factorization

***Coarse→fine factorization** reduces modeling complexity while preserving expressiveness.*

Paper, Models & Code

Full paper, finetuning code, model weight, live demo can be found:

Arxiv

GitHub

Authors: shi-y23 or fzl22@mails.tsinghua.edu.cn