MASTER:Market-Guided Stock Transformer for Stock Price Forecasting[1]

Group 7
110705009 陳重光、313551047 陳以瑄、313554043 戴明貴

Outline

- Introduction & Main Purpose
- Related Work - Yoo et al., KDD 2021
- Method
- Experienement & Ablation Study
- Conclusion

Intro - Dynamic Stock Correlation Modeling

Why Correlation Matters

In stock prediction, not only individual histories matter, but also:

1. Cross-stock correlation
2. Market-stock correlation

Types of Correlation Modeling

1. Static correlation modeling
   Predefined rules, cannot capture real-time changes
2. Dynamic correlation modeling
   Transformer architecture, learns time-varying correlations

Main Purpose

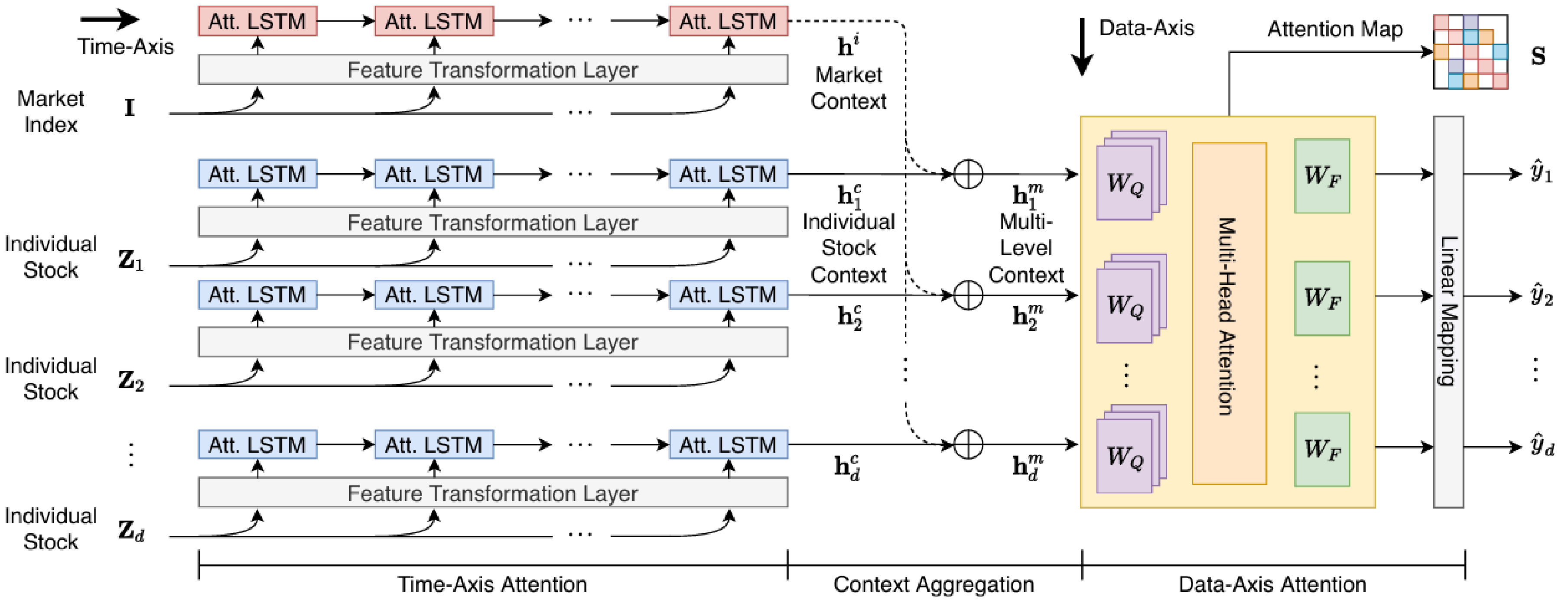
This study aims to model two phenomena:

1. Market status influence on feature relevance
2. Momentary and cross-time stock correlations

Both issues are limitations identified in Yoo et al. (KDD 2021).[2]

Related Work – Yoo et al., KDD 2021

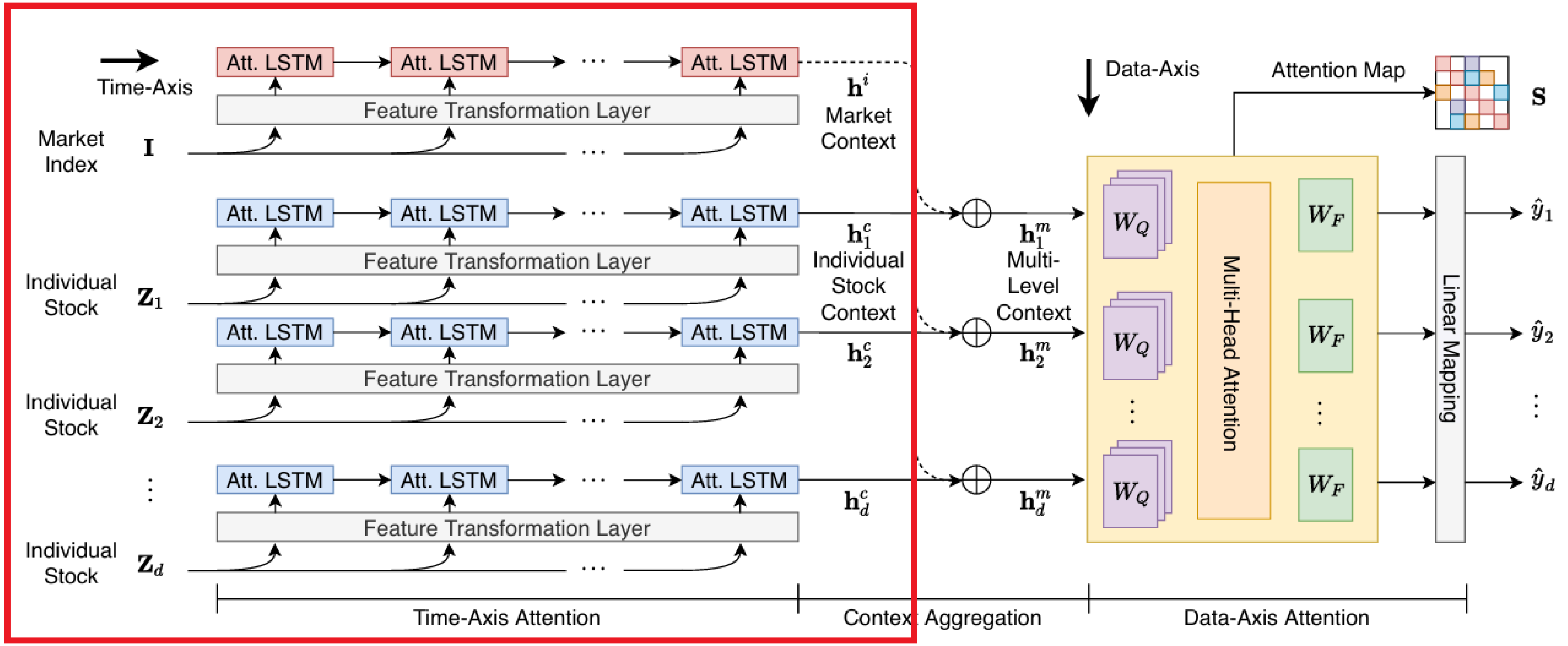
Accurate Multivariate Stock Movement Prediction via Data-Axis Transformer with Multi-Level Contexts



Related Work (Cont.)

Part 1: Time Axis Attention

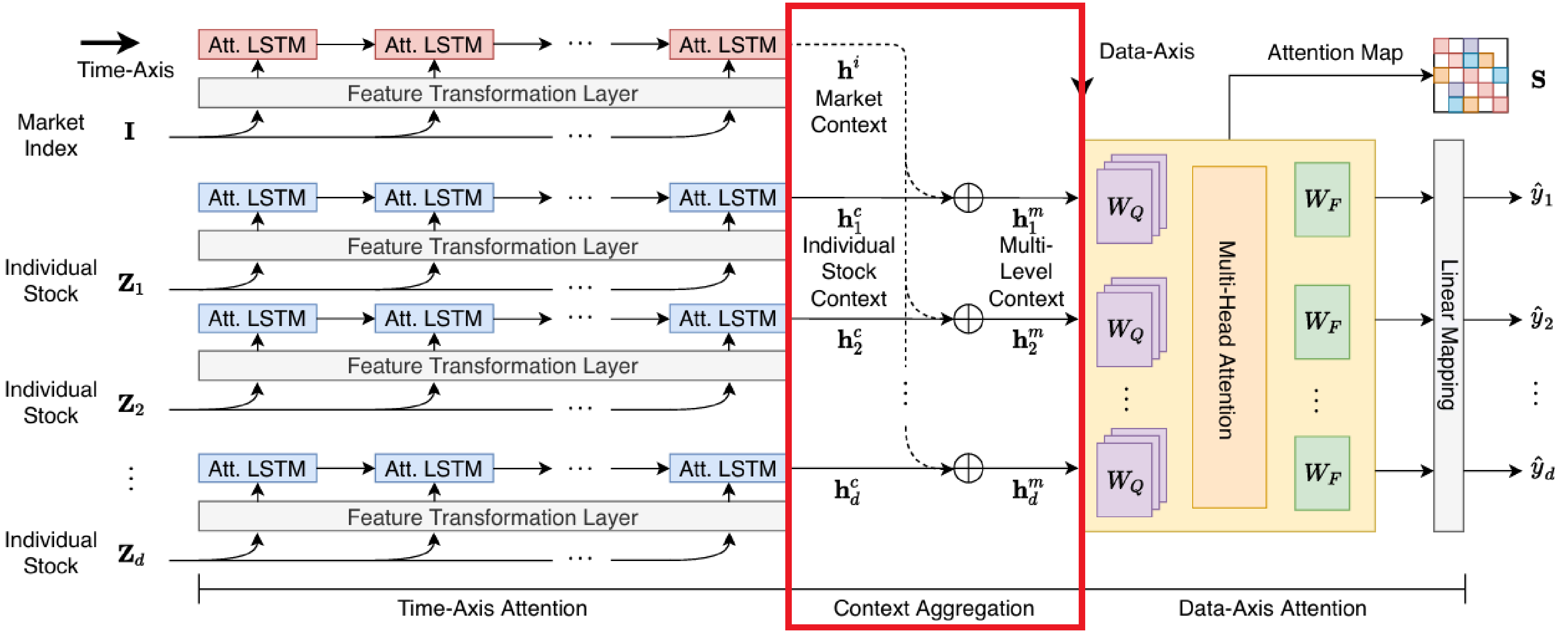
Single stock time series → Attention LSTM → Summarized embedding



Related Work (Cont.)

Part 2: Content Aggregation

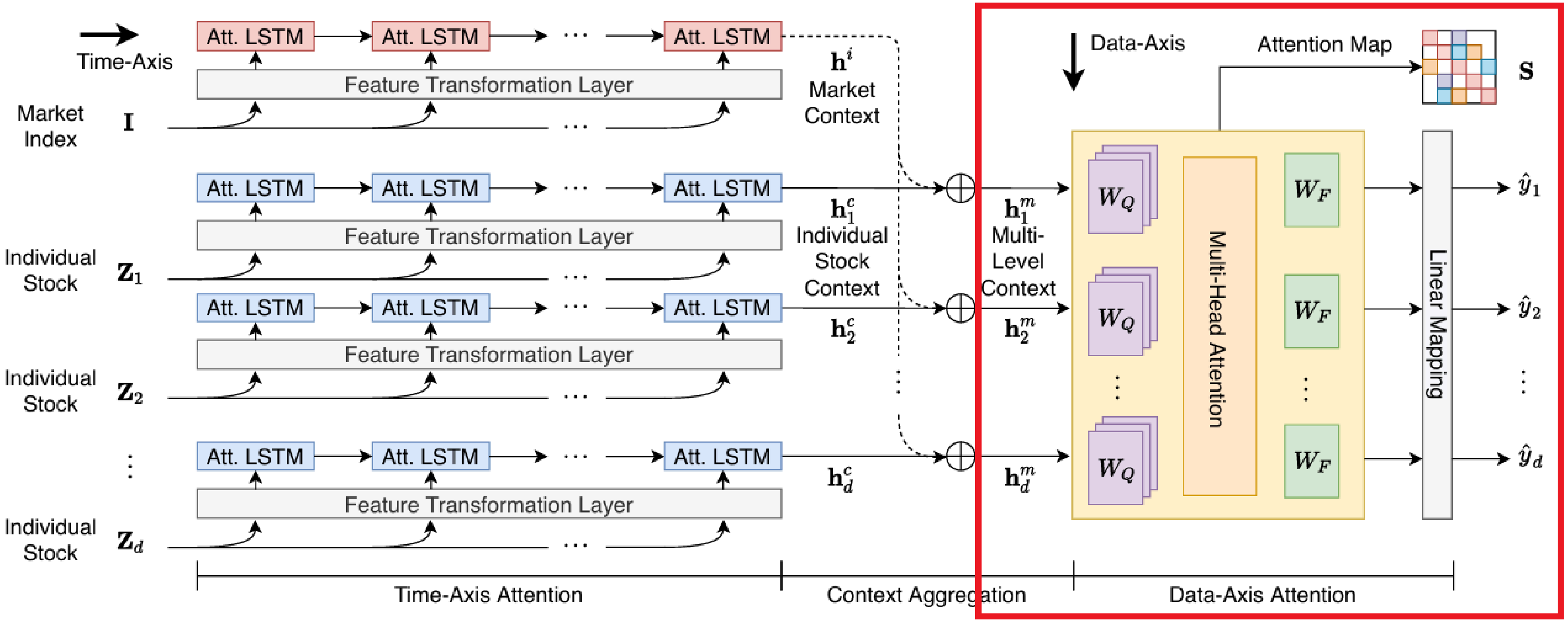
Incorporates market and stock representations: $h_u^m = h_u^c + \beta h^i$



Related Work (Cont.)

Part 3: Data Axis Attention

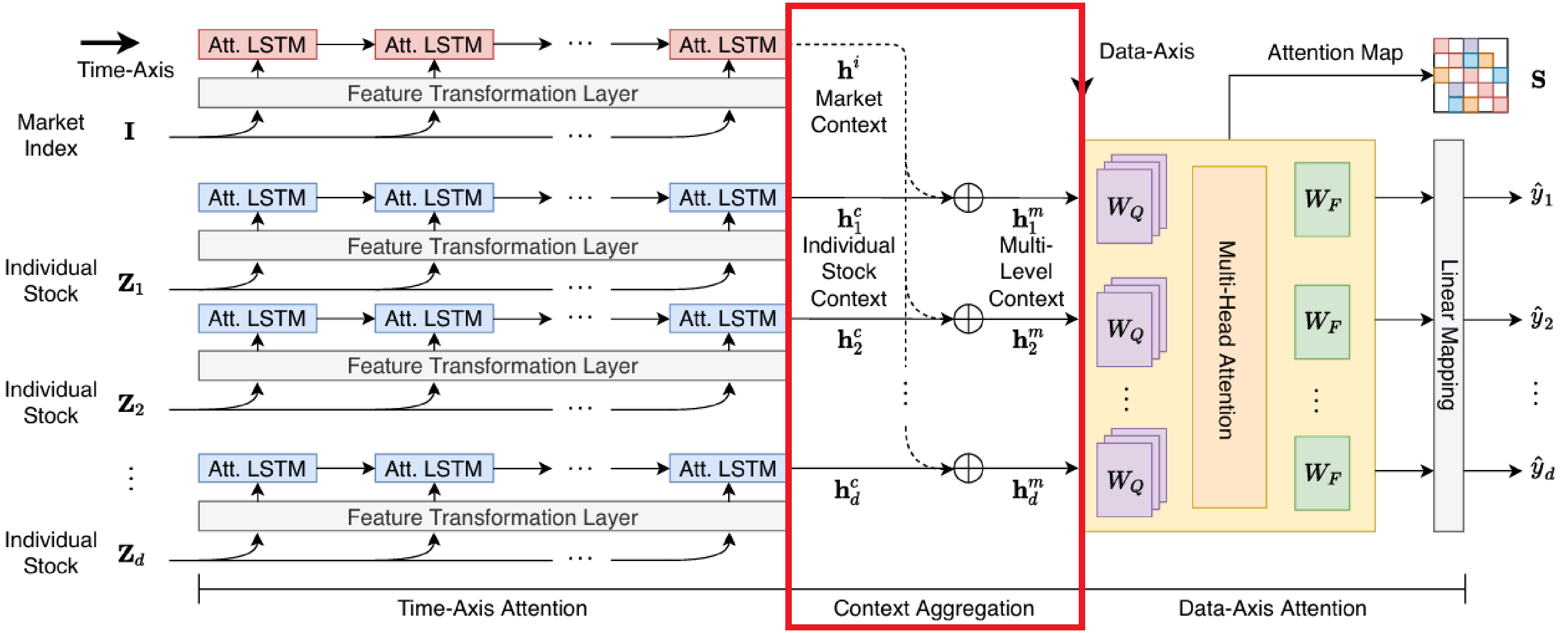
Uses Multihead attention to learn inter-stock correlation.



Limitation 1 - Market Status and Feature Selection

Problem in Prior Work (Step 2):

Integrates market information directly into stock representations.



Limitation 1 - Market Status and Feature Selection

What Is Missed:

Market status influences which features are useful for prediction.

Example:

In bullish markets, investor optimism and capital inflow lead to:

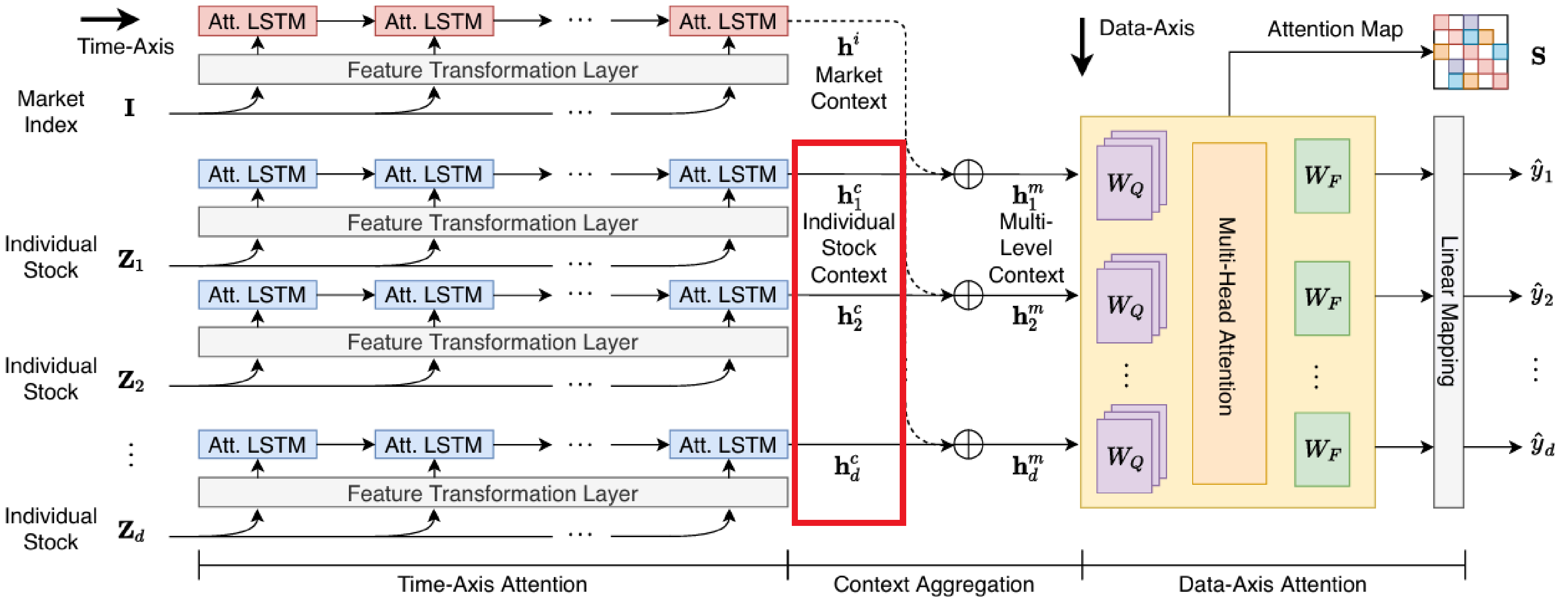
- Broader investment across sectors
- Stronger cross-stock correlations

→ Features like peer stock movements become more relevant

Limitation 2 - Cross-time Stock Correlation

Problem in Prior Work (Step 1 & 3):

Uses overall stock representation to learn cross-stock correlation.

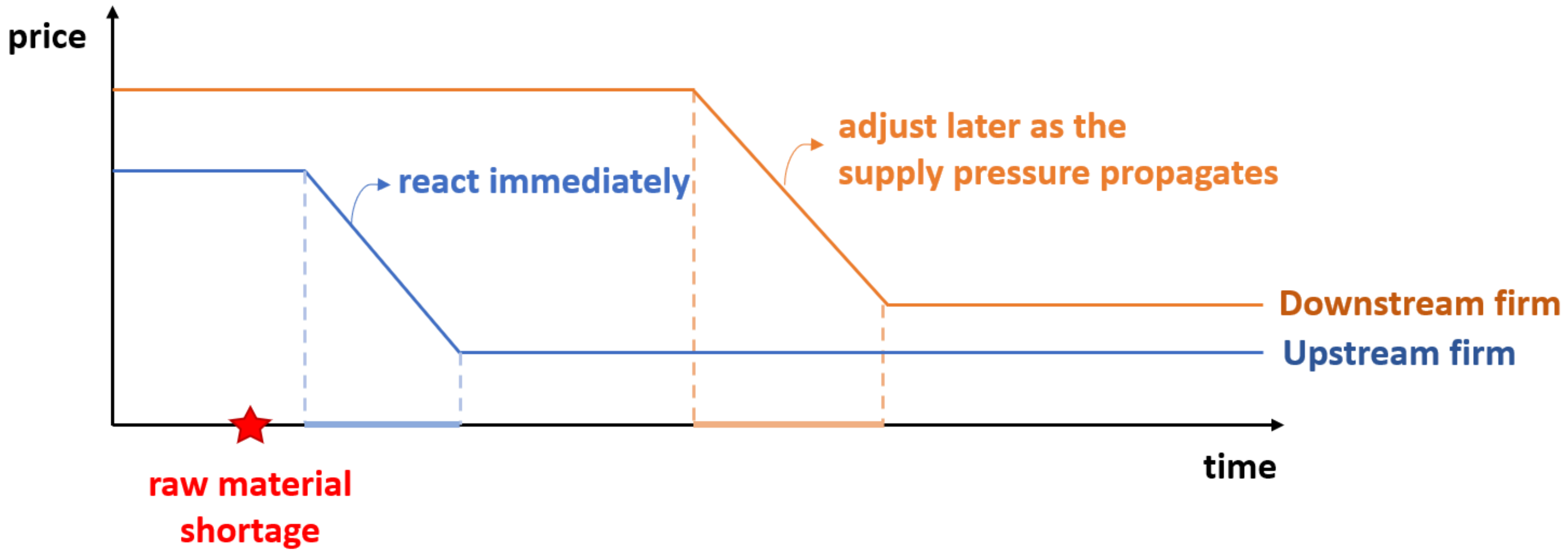


Limitation 2 - Cross-time Stock Correlation

What Is Missed:

Stocks often respond to the same external factors at different times.
→ Correlation is not aligned in time but occurs across time steps.

Example: Raw Material Shortage



Method-Framework

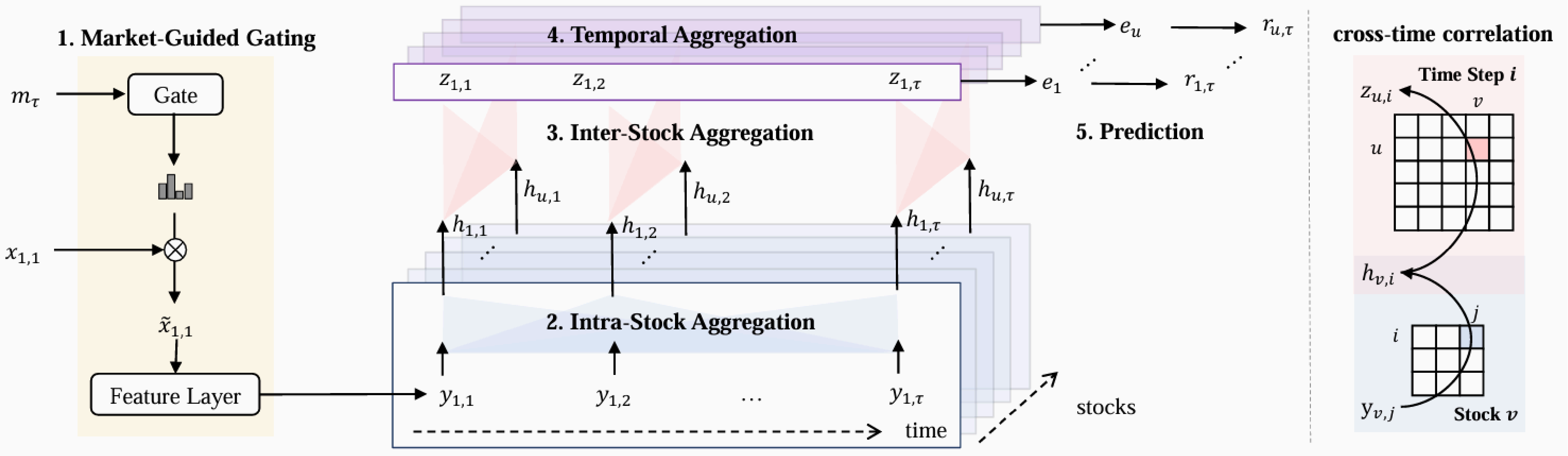


Figure 2: Overview of the MASTER framework.

Method-Framework (Cont.)

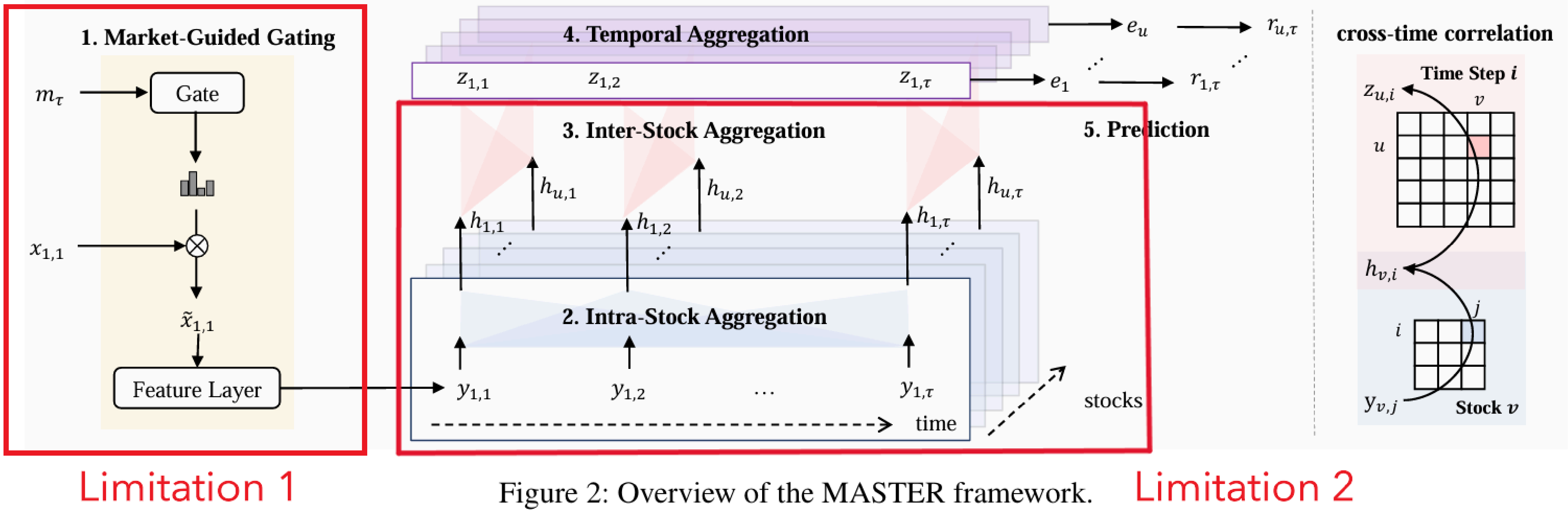


Figure 2: Overview of the MASTER framework.

Step1: Market-Guided Gating

Idea: adaptively select and scale features based on market conditions

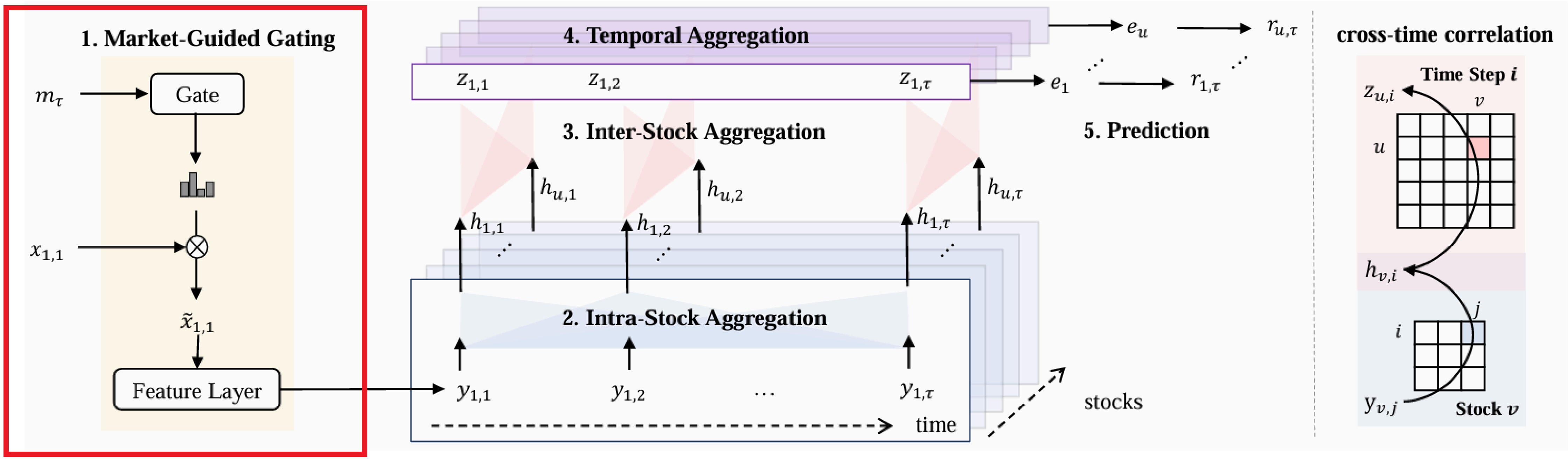


Figure 2: Overview of the MASTER framework.

Market-Guided Gating

Guiding Feature: Market status vector m_τ

- $|m_\tau| = F'$
- Market index price (current value, historical mean, std dev)
- Market index trading volume

Guided Feature: Stock Features $x_{u,t}$

- $|x_{u,t}| = F$
- Raw input features for stock u at time t

Market-Guided Gating

Gating Mechanism

1. Learn the scale factor: $\alpha(m_\tau) = F \cdot \mathrm{softmax}_\beta(W_\alpha m_\tau + b_\alpha)$
    - α: shared across all stocks $x_{u,t}$
    - β: smaller β → stronger filtering
2. Rescale each stock features: $\tilde{x}_{u,t} = \alpha(m_\tau) \circ x_{u,t}$

Addresses limitation 1:

Captures how market status influences the usefulness of features.

Step2: Intra-Stock Aggregation

Idea: Single stock behavior is continuous → a simpler distribution

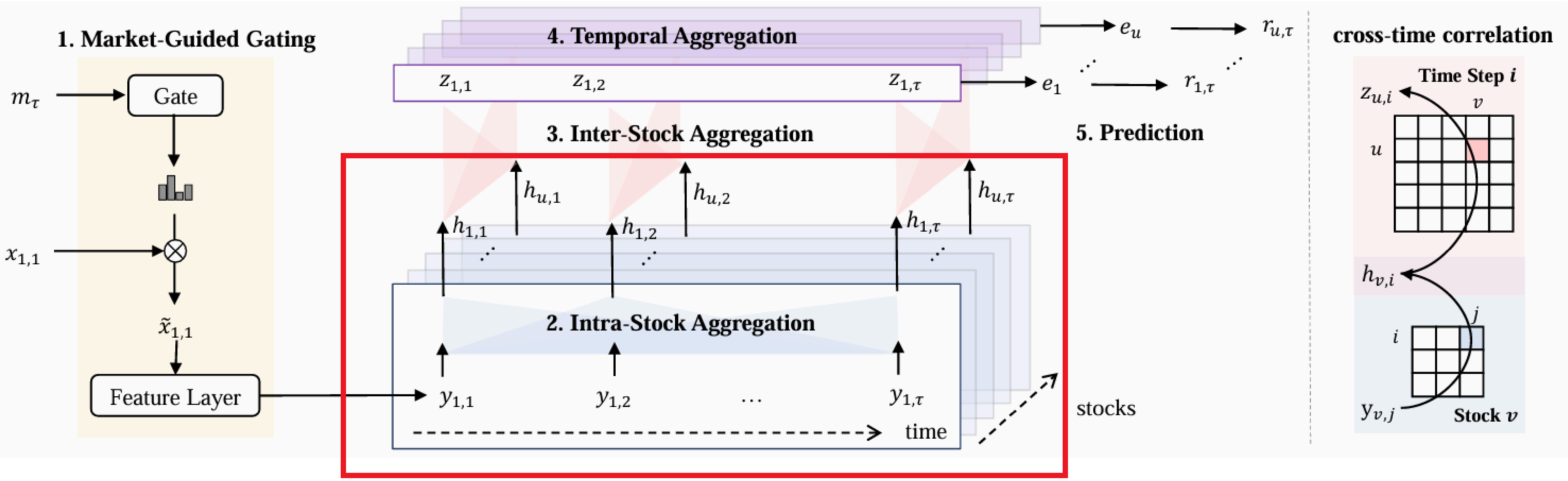


Figure 2: Overview of the MASTER framework.

Intra-Stock Aggregation

Multi-Head Attention

- Input: Rescaled features $\tilde{x}_{u,t}$
- Embedding: $y_{u,t} = f(\tilde{x}_{u,t}), |y_{u,t}| = D$
- Positional Encoding: Add sinusoidal p_t to retain temporal order
- Sequence: $Y_u = \|_{t \in [1,\tau]} \mathrm{LN}(f(\tilde{x}_{u,t}) + p_t)$

Intra-Stock Aggregation

Multi-Head Attention

- Sequence: $Y_u = \Vert_{t \in [1,\tau]} \mathrm{LN}(f(\tilde{x}_{u,t}) + p_t)$
- Attention

$$Q_u^1 = W_Q^1 Y_u, K_u^1 = W_K^1 Y_u, V_u^1 = W_V^1 Y_u$$

$$H_u^1 = \mathrm{FFN}^1(\mathrm{MHA}^1(Q_u^1, K_u^1, V_u^1) + Y_u)$$

- Output: Time-aware feature embeddings $h_{u,t}$ for each time step

Step3: Inter-Stock Aggregation

Idea: Capture the momentary stock correlation at each time step.

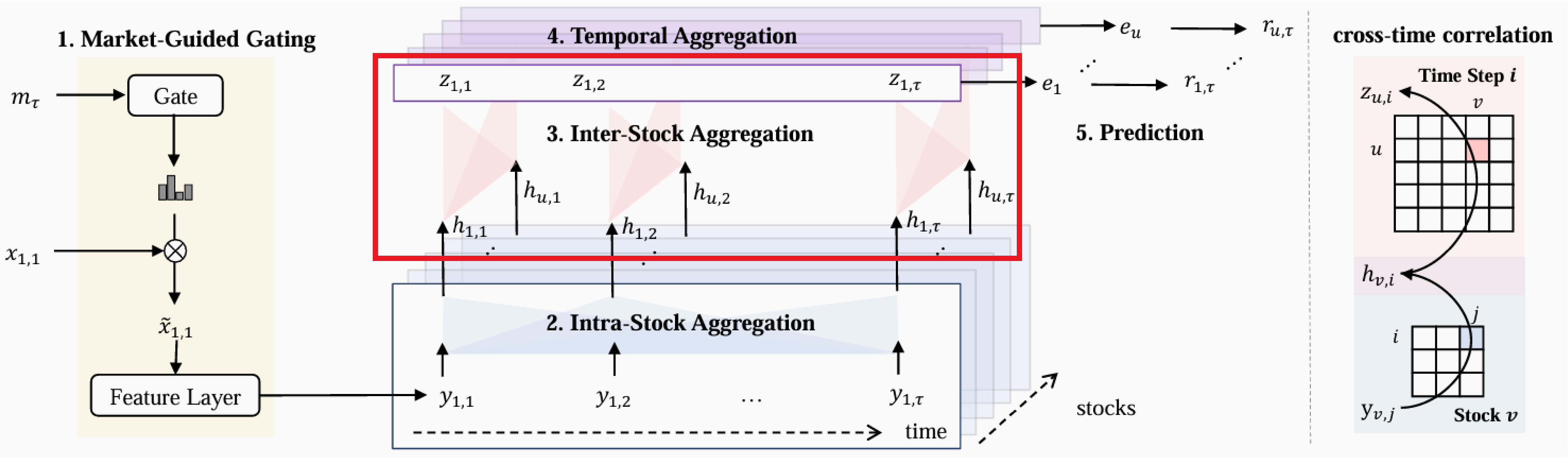


Figure 2: Overview of the MASTER framework.

Inter-Stock Aggregation

Multi-Head Attention

- Input: Gather intra-stock embeddings: $H_t^2 = \Vert_{u \in S} h_{u,t}$
- Attention:

$$Q_t^2 = W_Q^2 H_t^2, K_t^2 = W_K^2 H_t^2, V_t^2 = W_V^2 H_t^2$$

$$Z_t = \Vert_{u \in S} z_{u,t} = \mathrm{FFN}^2 \left(\mathrm{MHA}^2(Q_t^2, K_t^2, V_t^2) + H_t^2 \right)$$

- Output: Temporal embedding $z_{u,t}$ for each stock

Step 2+3: Inter-Stock & Intra-Stock Aggregations

Addresses Limitation 2:

Capture dynamic cross-stock correlations across time steps.

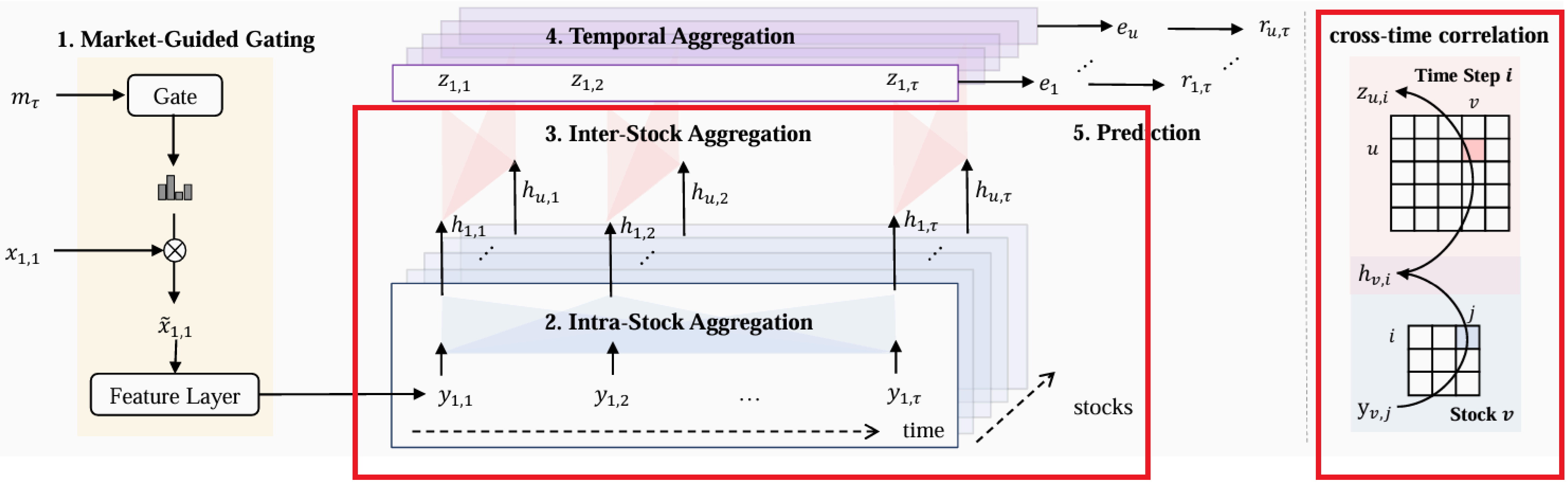


Figure 2: Overview of the MASTER framework.

Step 4: Temporal Aggregation

Idea: Summarize the obtained temporal embeddings.

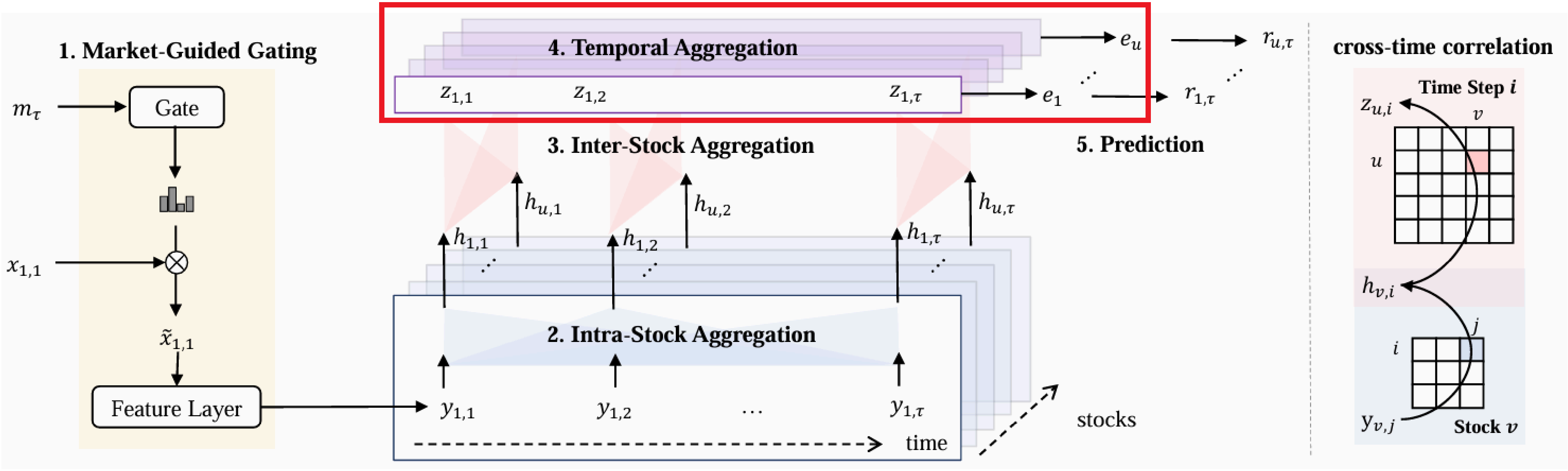


Figure 2: Overview of the MASTER framework.

Temporal Aggregation

- Input: Temporal embedding $z_{u,t}$
- Temporal attention layer:
    - Use the latest temporal embedding $z_{u,\tau}$ as the query vector
    - Compute attention scores: $\lambda_{u,t} = \frac{\exp(z_{u,t}^T W_\lambda z_{u,\tau})}{\sum_{i \in [1,\tau]} \exp(z_{u,i}^T W_\lambda z_{u,\tau})}$
- Aggregate embeddings: $e_u = \sum_{t \in [1,\tau]} \lambda_{u,t} z_{u,t}$

Step 5: Prediction

Idea: Use the stock embedding to predict the stock price.

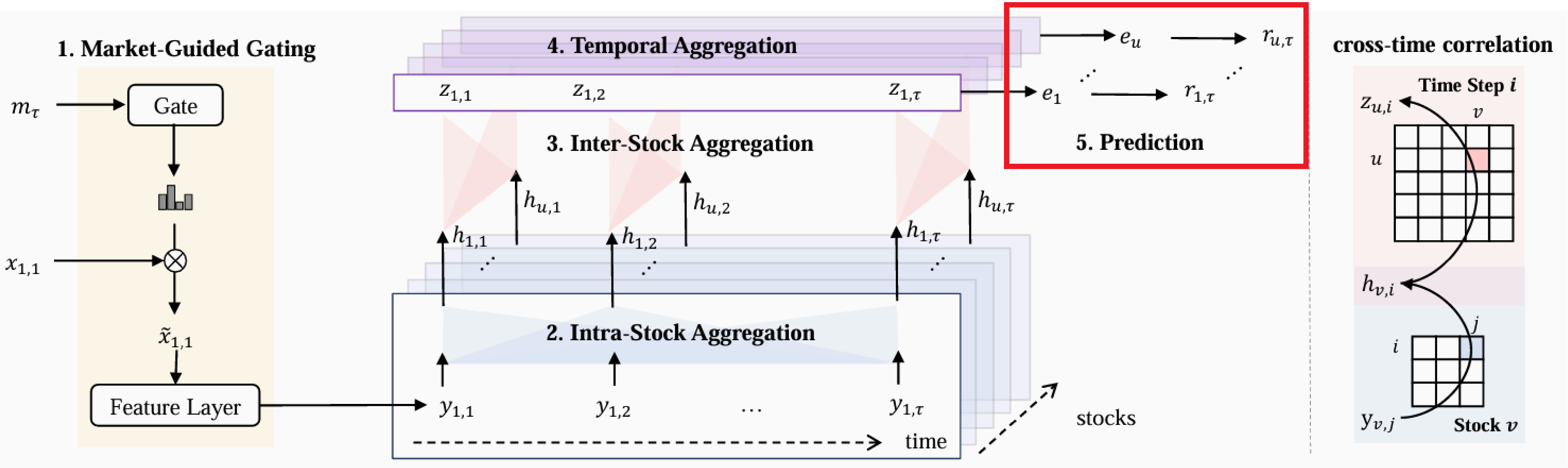


Figure 2: Overview of the MASTER framework.

Prediction

- Predicted return:
  $\hat{r}_u = g(e_u), g(.)$ is a linear layer for regression.
- Ground truth return:
  r_u is the normalized return ratio
- Loss Function: $L = \sum_{u \in S} MSE(r_u, \hat{r}_u)$

Experiments - Dataset

- **Stock sets:** Chinese stock market CSI300 and CSI800
- **Time:** 2008 to 2022

	Time
Training	Q1 2008 to Q1 2020
Validation	Q2 2020
Testing	Q3 2020 to Q4 2022

- **Feature:**
  Individual stock feature: Alpha158 indicators
  Shared market feature: CSI300, CSI500 and CSI800 index

Evaluation Metrics

Metric	Type	Description
IC	Ranking-based	Pearson correlation btw predictions and actual returns.
RankIC	Ranking-based	Spearman rank correlation btw predictions and actual returns.
ICIR	Ranking-based	Normalized IC (divided by std dev.)
RankICIR	Ranking-based	Normalized RankIC
AR	Portfolio-based	Annualized excess return from investing in top-ranked stocks
IR	Portfolio-based	Normalized AR

For all metrics, higher values indicate better performance.

Experiments - Result

Dataset	Model	IC	ICIR	RankIC	RankICIR	AR	IR
CSI300	XGBoost	0.051 ± 0.001	0.37 ± 0.01	0.050 ± 0.001	0.36 ± 0.01	0.23 ± 0.03	1.9 ± 0.3
	LSTM	0.049 ± 0.001	0.41 ± 0.01	0.051 ± 0.002	0.41 ± 0.03	0.20 ± 0.04	2.0 ± 0.4
	GRU	0.052 ± 0.004	0.35 ± 0.04	0.052 ± 0.005	0.34 ± 0.04	0.19 ± 0.04	1.5 ± 0.3
	TCN	0.050 ± 0.002	0.33 ± 0.04	0.049 ± 0.002	0.31 ± 0.04	0.18 ± 0.05	1.4 ± 0.5
	Transformer	0.047 ± 0.007	0.39 ± 0.04	0.051 ± 0.002	0.42 ± 0.04	0.22 ± 0.06	2.0 ± 0.4
	GAT	0.054 ± 0.002	0.36 ± 0.02	0.041 ± 0.002	0.25 ± 0.02	0.19 ± 0.03	1.3 ± 0.3
	DTML	0.049 ± 0.006	0.33 ± 0.04	0.052 ± 0.005	0.33 ± 0.04	0.21 ± 0.03	1.7 ± 0.3
	MASTER	$\mathbf{0.064^* \pm 0.006}$	$\mathbf{0.42 \pm 0.04}$	$\mathbf{0.076^* \pm 0.005}$	$\mathbf{0.49 \pm 0.04}$	$\mathbf{0.27 \pm 0.05}$	$\mathbf{2.4 \pm 0.4}$
CSI800	XGBoost	0.040 ± 0.000	0.37 ± 0.01	0.047 ± 0.000	0.42 ± 0.01	0.08 ± 0.02	0.6 ± 0.2
	LSTM	0.028 ± 0.002	0.32 ± 0.02	0.039 ± 0.002	0.41 ± 0.03	0.09 ± 0.02	0.9 ± 0.2
	GRU	0.039 ± 0.002	0.36 ± 0.05	0.044 ± 0.003	0.39 ± 0.07	0.07 ± 0.04	0.6 ± 0.3
	TCN	0.038 ± 0.002	0.33 ± 0.04	0.045 ± 0.002	0.38 ± 0.05	0.05 ± 0.04	0.4 ± 0.3
	Transformer	0.040 ± 0.003	$\mathbf{0.43 \pm 0.03}$	0.048 ± 0.003	$\mathbf{0.51 \pm 0.05}$	0.13 ± 0.04	1.1 ± 0.3
	GAT	0.043 ± 0.002	0.39 ± 0.02	0.042 ± 0.002	0.35 ± 0.02	0.10 ± 0.04	0.7 ± 0.3
	DTML	0.039 ± 0.004	0.29 ± 0.03	0.053 ± 0.008	0.37 ± 0.06	0.16 ± 0.03	1.3 ± 0.2
	MASTER	$\mathbf{0.052^* \pm 0.006}$	0.40 ± 0.06	$\mathbf{0.066 \pm 0.007}$	0.48 ± 0.06	$\mathbf{0.28^* \pm 0.02}$	$\mathbf{2.3^* \pm 0.3}$

Ablation Study - Beta

Temperature β

- A hyperparameter in the gating mechanism.
- Small β → stronger gating effect → filter more features.

Result: X-axis: β; Y-axis: Performance Metric

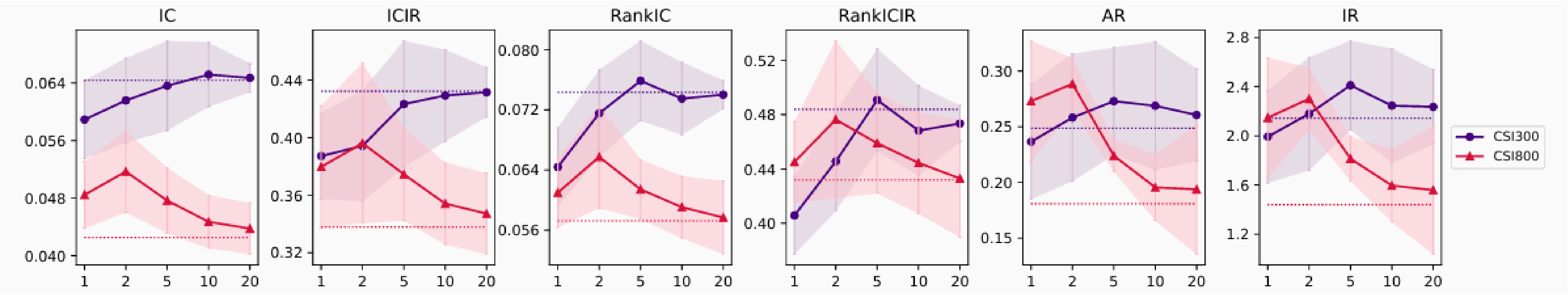


Figure 4: MASTER performance with varying β. The horizontal dash lines are performance without market-guided gating.

Ablation Study - Beta (Cont.)

Result: X-axis: β; Y-axis: Performance Metric

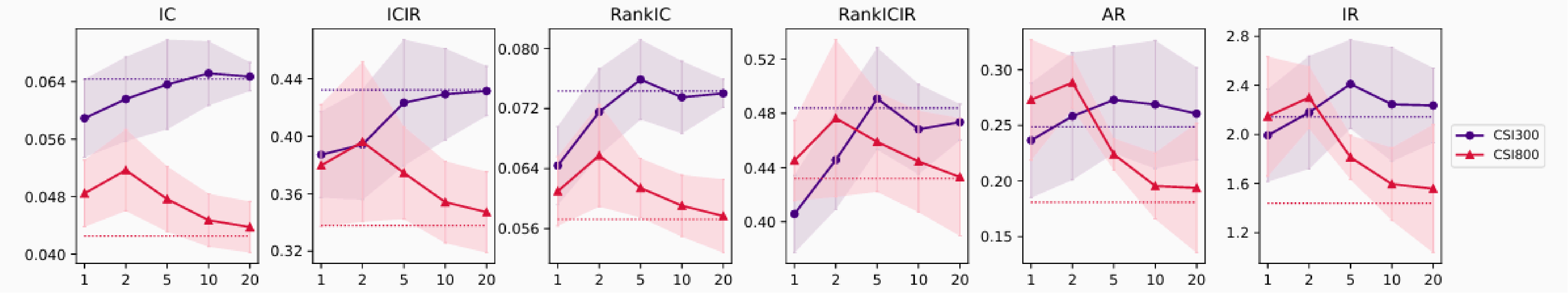
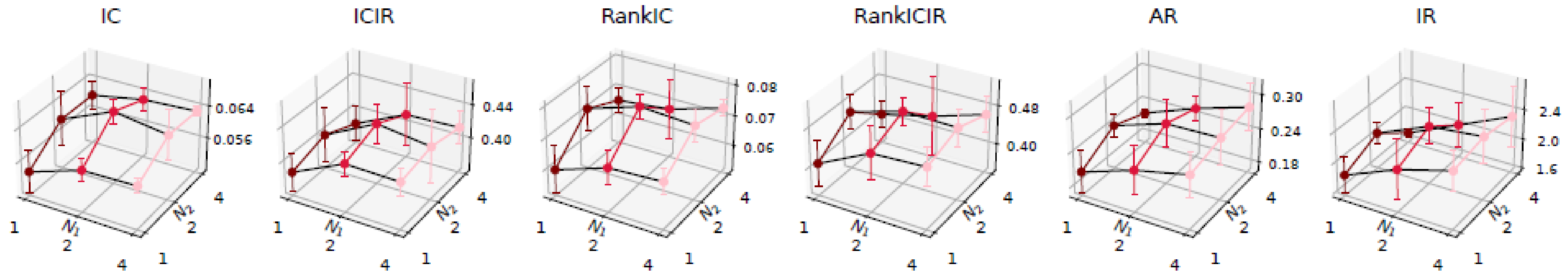


Figure 4: MASTER performance with varying β. The horizontal dash lines are performance without market-guided gating.

- CSI300: easier dataset → most features are effective → larger β
- CSI800: complex dataset → need feature selection → smaller β

Ablation Study - (N1, N2)

(N1, N2): The number of multi-heads in the attention mechanism.



Findings:

- Differences between head combinations are not significant.
- Most settings outperform the baseline → robustness

Conclusion

MASTER Key Features:

1. Models realistic stock correlations.
2. Guides feature selection with market information.

Experiments Result on the CSI300 and CSI800

- improves 13% on ranking metrics
- improves 47% on portfolio-based metrics

References

[1] Li, T., Liu, Z., Shen, Y., Wang, X., Chen, H., and Huang, S. (2024). MASTER: Market-Guided Stock Transformer for Stock Price Forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence, 38(1), 162-170.

[2]Yoo, J., Soun, Y., Park, Y.-c., and Kang, U. (2021). Accurate multivariate stock movement prediction via data-axis transformer with multi-level contexts. In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, 2037–2045.

Thank you for listening.