DYNAMIC CROSS-MARKET SYSTEMATIC ATTENTION FOR STOCK PREDICTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Stock price prediction is challenging due to the extremely low signal-to-noise ratio in financial markets, where stock returns reflect both market-wide systematic effects and stock-specific dynamics. Recent machine learning approaches have increasingly explored cross-stock relationships for prediction, but often overlook the underlying market-wide systematic effects and their potential interactions across different markets. To address these limitations, we propose a model-agnostic dynamic cross-market systematic attention method that improves existing stock prediction models with dynamic latent systematic factors and cross-market factor interactions. The method learns market-specific latent factors that adapt to changing market conditions and models their interactions across markets through factor-level attention, providing systematic information that can be incorporated into diverse stock prediction architectures. Experiments across the China, United States, and Hong Kong stock markets demonstrate consistent improvements when the method is integrated with different prediction models. Further analysis reveals time-varying cross-market interactions, providing insights into how the method captures changing patterns of systematic dependence across markets.

1 INTRODUCTION

Stock return prediction remains a fundamental challenge in finance due to the extremely low signal-to-noise ratio of financial time series. Machine learning methods (Qin et al., 2017; Gu et al., 2020; Duan et al., 2022) have been widely explored for financial forecasting because of their ability to capture complex and nonlinear patterns in historical market data. In financial markets, however, stock returns are influenced by both market-wide systematic effects shared across assets and stock-specific idiosyncratic dynamics. Effectively modeling these distinct sources of return variation therefore requires capturing both the dependencies shared across stocks and the dynamics specific to individual stocks.

A substantial body of machine learning research has focused on modeling relationships among stocks to capture shared sources of information and improve return prediction. Many of these approaches rely on graph neural networks (GNNs) (Veličković et al., 2018; Sawhney et al., 2021; Wang et al., 2022; Qian et al., 2024; Xia et al., 2024; Song et al., 2025; Park et al., 2025; Li et al., 2025; Mei et al., 2025; Duan et al., 2025) to construct stock graphs from prior knowledge, such as industry affiliations, ownership structures, or news relations. While effective at incorporating prior relational knowledge, these approaches depend on fixed graph topologies and typically assume a static stock universe, making it difficult to naturally accommodate newly listed stocks and evolving market structures. To address these limitations, recent studies have adopted attention-based architectures (Yoo et al., 2021; Li et al., 2024; Choi et al., 2025), which can model dynamic relationships among stocks without requiring a predefined graph. Nevertheless, current methods still have several limitations.

First, existing relationship-based methods primarily focus on dependencies among individual stocks and often overlook the underlying market-wide systematic effects. Cross-stock interactions can capture shared movements implicitly, but they do not explicitly distinguish common systematic influences from stock-specific relationships. This distinction is important from a financial perspective, as asset-pricing theory suggests that stock returns reflect both systematic risk driven by market-wide and macroeconomic factors and idiosyncratic risk specific to individual firms (Sharpe, 1964;

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.

Ross, 1976; Fama & French, 1993). Market-wide influences affect many assets simultaneously and are therefore more naturally represented through shared latent factors than through an aggregation of pairwise stock relationships. Several recent studies have introduced latent factors to represent shared market-wide information more explicitly (Zhu et al., 2025; Duan et al., 2025). However, existing latent-factor approaches often make systematic information explicit without modeling how the factor representation changes with the evolving market state. In financial markets, the relevance and strength of systematic effects can change substantially across market conditions, as factor exposures and risk loadings may vary over time (Adrian & Franzoni, 2009; Fan et al., 2024). A static factor representation may therefore be insufficient to characterize the time-varying systematic component of stock returns. This motivates learning latent systematic factors that can adapt to the current market state, allowing the representation of common market effects to evolve with changing financial conditions.

Second, systematic effects are not isolated within individual markets. International equity markets exhibit substantial cross-market spillovers, with shocks and financial conditions in one market affecting others (Ren et al., 2022; Wang et al., 2023; Ouyang et al., 2025). For example, the "Liberation Day" tariff announcement on April 2, 2025 triggered sharp declines not only in the U.S. equities but also across global stock markets. Such events illustrate how market-wide effects can propagate across markets rather than remain confined to a single region. Thus, modeling systematic factors independently within each market may fail to capture how these factors interact and transmit across markets. This motivates moving beyond market-specific factor representations toward an approach that allows latent systematic factors to interact and transmit information across markets while adapting to changing market conditions.

In this paper, we propose a model-agnostic dynamic cross-market systematic attention (DCSA) method that can be integrated into diverse stock prediction architectures. Given intermediate stock representations from a host model, DCSA dynamically constructs market-specific latent factors conditioned on the current market state and models factor-level interactions across markets through cross-attention. The method is designed for straightforward integration: it independently processes the host model's temporal representations and returns a systematic representation through a gated residual connection, allowing the same method to augment different architectures without host-specific customization. Experiments across the Chinese, the U.S., and Hong Kong markets show consistent improvements when DCSA is integrated into three diverse baseline architectures, with prediction accuracy improves by 21% to 94% and Sharpe ratio gains of 29% to 63%, demonstrating its effectiveness and generality as a model-agnostic method.

2 Preliminaries and Related Work

2.1 Problem Definition

Stock price forecasting is commonly approached as a regression problem. Let us denote the set of N stocks in the market as $\mathbf{S} = \{s_1, s_2, \ldots, s_N\}$. For each stock s_i, $i \in N$, a group of predictive factors are collected at each trading day $t \in [1, T]$ to form its feature $\mathbf{x}_{i,t} \in \mathbb{R}^{1\times D}$ where D is the dimensionality of the features. The objective is to predict the future return ratio r_i for each stock based on its historical features.

Following the typical setting in existing stock prediction works (Feng et al., 2019; Sawhney et al., 2021; Wang et al., 2022; Qian et al., 2024; Xia et al., 2024; Li et al., 2024; Choi et al., 2025), we forecast return ratios instead of the absolute value of prices. Let t denote the last trading day in the observation window, let h denote the forecast horizon measured in trading days, and let l denote the length of the return interval, with $1 \leq l \leq h$. For stock s_i, we define the l-day return ending at $t+h$ as

$$r_i = \frac{p_{i,t+h} - p_{i,t+h-l}}{p_{i,t+h-l}}, \tag{1}$$

where $p_{i,u}$ denotes the closing price of stock s_i on trading day u. Thus, r_i measures the relative price change from $t+h-l$ to $t+h$. In all experiments, we set $l = 1$. We set $h = 2$ for the Chinese market to account for its T+1 trading restriction, and $h = 1$ for the U.S. and Hong Kong markets.

2.2 Related Work

Early machine learning approaches to stock return prediction primarily relied on firm-level predictors and conventional supervised learning models, including Random Forest, Gradient Boosting Regression Trees, and Multilayer Perceptrons (Gu et al., 2020). With the increasing use of sequential financial data, subsequent studies adopted recurrent neural networks and attention mechanisms to capture temporal dependencies, including LSTM (Hochreiter & Schmidhuber, 1997), DA-RNN (Qin et al., 2017), and AdaRNN (Du et al., 2021). These developments established temporal representation learning as an important foundation for stock return prediction, but they primarily model each stock's historical dynamics independently and do not explicitly characterize cross-sectional relationships among stocks.

More recent studies therefore incorporate cross-stock relationships to exploit information shared across assets. GNN-based approaches construct stock graphs using predefined relations, such as industry affiliations, ownership structures, and news relations, or data-driven similarities such as feature correlations (Veličković et al., 2018; Sawhney et al., 2021; Wang et al., 2022; Qian et al., 2024; Xia et al., 2024; Song et al., 2025; Park et al., 2025; Li et al., 2025; Mei et al., 2025). While these methods explicitly model cross-stock dependencies, their graph construction generally relies on a fixed stock universe and relational structure, which can limit their flexibility as the stock universe evolves. Recent attention-based approaches address this limitation by learning data-dependent relationships directly from stock representations without requiring a predefined graph (Yoo et al., 2021; Li et al., 2024; Choi et al., 2025). These methods demonstrate the value of modeling cross-stock relationships, but their cross-sectional interactions are primarily formulated at the stock level and do not explicitly distinguish shared systematic effects from stock-specific dynamics.

A related line of work introduces latent factors to capture shared information among stocks. MCI-GRU (Zhu et al., 2025) incorporates learned market information to enrich stock representations, while FactorGCL (Duan et al., 2025) models residual information through a factor-aware cascading hypergraph structure. These approaches recognize the importance of latent common factors, but they do not explicitly condition their factor representations on the evolving state of the market or model factor-level interactions across international markets. In contrast, our method dynamically constructs market-specific systematic factors and models their interactions across markets.

Another line of research incorporates multimodal information, such as financial news, reports, audio, and video, into stock prediction (Xu & Cohen, 2018; Luo et al., 2023; Yu et al., 2024; Qian et al., 2026). Recent studies have increasingly employed large language models to extract information from textual and other alternative data sources. These approaches address a complementary problem by enriching stock representations with external information, but their architectures generally require additional modalities and are therefore not directly compatible with our numerical-feature-based method. Moreover, most multimodal studies focus on individual-stock prediction or classification-oriented tasks rather than return prediction across large stock universes. We therefore do not include multimodal methods in our experimental comparisons.

3 Model Architecture

3.1 Overview

Figure 1a illustrates the overall architecture integrating our proposed dynamic cross-market systematic attention method with host models. The method is designed to augment existing stock prediction models without modifying their architectures. All stock prediction models begin with a common pipeline. Given the historical feature sequence of stock s_i over a lookback window of length τ,

$$\mathbf{X}_i = [x_{i,t-\tau+1}; \ldots; \mathbf{x}_{i,t}] \in \mathbb{R}^{\tau \times D}, \tag{2}$$

the aggregation of features of N stocks in the market becomes the raw input $\mathbf{X} = \{\mathbf{X}_i\}_{i=1}^{N} \in \mathbb{R}^{N \times \tau \times D}$ for the model. $\mathbf{X}$ is encoded by host models' *Temporal Encoder*, normally a GRU (Cho et al., 2014) or Transformer (Vaswani et al., 2017) layer, into intermediate representations $\mathbf{H} \in \mathbb{R}^{N \times \tau \times d}$, which capture temporal and feature interactions. The encoded representation is then processed in parallel by the *Host Model* and *DCSA*. The *Host Model* captures stock-specific and within-market dependencies and produces a idiosyncratic representation $\mathbf{L} \in \mathbb{R}^{N \times \tau \times d}$. In parallel,

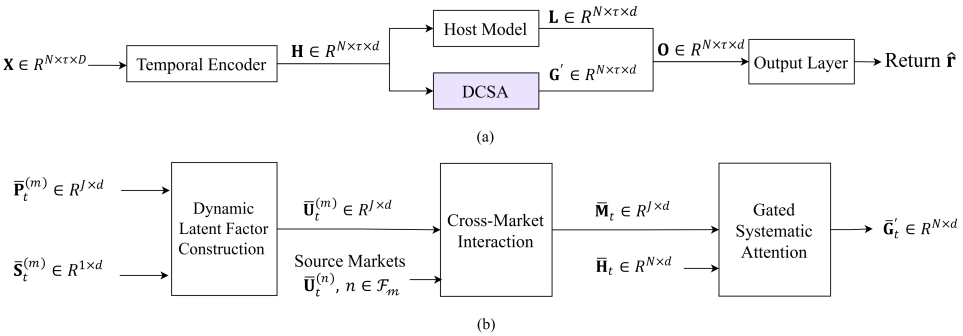

Figure 1: Overview of DCSA. (a) Pipeline for integrating DCSA into a host model. The DCSA in color represents the method we propose. The DCSA output is combined with the host model's output to get the final prediction $\mathbf{r}$. (b) Internal structure of the DCSA method.

DCSA constructs dynamic latent factors from the target market's latent factor bases and market state, models their interactions with factors from other markets, and applies gated systematic attention to generate the systematic representation $\mathbf{G}' \in \mathbb{R}^{N\times\tau\times d}$. The idiosyncratic representation $\mathbf{L}$ and the systematic representation $\mathbf{G}'$ are then fused to form $\mathbf{O} \in \mathbb{R}^{N\times\tau\times d}$, which is passed to the *Output Layer* for predicting the return $\mathbf{r}$. Note that in Figure 1a, all the blocks besides *DCSA*, including *Temporal Encoder*, *Host Model* and *Output Layer* are from the original method's architecture and do not belong to our method. The detailed structure of *DCSA* is described in the following subsections.

3.2 Dynamic Cross-market Systematic Attention (DCSA)

DCSA is a model-agnostic method that extracts systematic information from the intermediate stock representations produced by a host prediction model. As shown in Figure 1b, DCSA consists of three stages: **Dynamic Latent Factor Construction**, **Cross-Market Interaction**, and **Gated Systematic Attention**.

3.2.1 Dynamic Latent Factor Construction

We introduce a small set of J learnable latent market factors to capture shared market dynamics. For each target market m, we denote its latent factors by

$$\mathbf{P}^{(m)} = \{\mathbf{P}_j^{(m)}\}_{j=1}^{J} \in \mathbb{R}^{J\times\tau\times d}, \qquad J \ll N, \tag{3}$$

where each $\mathbf{P}_j^{(m)} \in \mathbb{R}^{\tau\times d}$ represents a latent source of systematic variation in market m and is randomly initialized with a uniform distribution. Unlike observable market indices, these factors are learned jointly with the prediction objective and are therefore not restricted to a predefined representation of market dynamics. However, systematic effects can change as market conditions evolve. We therefore dynamically adapt the latent factors according to the current state of each market. Let $\mathbf{S}_t^{(m)} \in \mathbb{R}^{1\times d}$ denote the market state of market m at time t, obtained by embedding the target market's index-level price and volume features. We first transform the market state into factor-specific updates:

$$\bar{\mathbf{\Delta}}_t^{(m)} = f_\Delta^{(m)}(\mathbf{S}_t^{(m)}) \in \mathbb{R}^{J\times d}, \tag{4}$$

where $f_\Delta^{(m)} : \mathbb{R}^{1\times d} \to \mathbb{R}^{J\times d}$ is a learnable transformation that generates an update vector for each of the J latent factors. In parallel, we combine the current market state with the J factor prototypes $\bar{\mathbf{P}}_t^{(m)} \in \mathbb{R}^{J\times d}$ at time t to produce an adaptive gate:

$$\mathbf{g}_t^{(m)} = \sigma\left(f_g^{(m)}(\mathbf{S}_t^{(m)}, \bar{\mathbf{P}}_t^{(m)})\right) \in \mathbb{R}^{J\times d}, \tag{5}$$

where $f_g^{(m)}$ is a learnable transformation and $\sigma(\cdot)$ denotes the sigmoid function. The dynamic latent factor is then obtained as

$$\bar{\mathbf{U}}_t^{(m)} = \bar{\mathbf{P}}_t^{(m)} + \mathbf{g}_t^{(m)} \odot \bar{\mathbf{\Delta}}_t^{(m)}. \tag{6}$$

Here, $\bar{\mathbf{\Delta}}_t^{(m)}$ determines the state-dependent adaptation of each latent factor, while $\mathbf{g}_t^{(m)}$ determines the extent to which this adaptation is incorporated for each factor and feature dimension. By conditioning the gate on both $\mathbf{S}_t^{(m)}$ and $\bar{\mathbf{P}}_t^{(m)}$, the model can distinguish how different latent factors should respond to the same market conditions. Thus, $\bar{\mathbf{P}}_t^{(m)}$ captures the persistent market-specific structure of systematic factors, while $\bar{\mathbf{\Delta}}_t^{(m)}$ and $\mathbf{g}_t^{(m)}$ allow this structure to adapt dynamically over time in response to evolving market conditions.

3.2.2 Cross-Market Interaction

Financial markets are interconnected, and systematic movements in one market may contain information about another. We therefore allow the dynamic factors of a target market to incorporate information from other markets. Let $\mathcal{F}_m$ denote the set of source markets, i.e., all markets other than the target market m. The dynamic latent factor $\bar{\mathbf{U}}_t^{(m)} \in \mathbb{R}^{J\times d}$ is the output of the dynamic latent factor construction stage for the target market, while $\bar{\mathbf{U}}_t^{(n)} \in \mathbb{R}^{J\times d}$ denotes the corresponding dynamic latent factor constructed for source market $n \in \mathcal{F}_m$. For each source market, we compute factor-level queries, keys, and values:

$$\mathbf{Q}_t^{(m)} = \bar{\mathbf{U}}_t^{(m)}\mathbf{W}_q^{(f)}, \qquad \mathbf{K}_t^{(m,n)} = \bar{\mathbf{U}}_t^{(n)}\mathbf{W}_k^{(f)}, \qquad \mathbf{V}_t^{(m,n)} = \bar{\mathbf{U}}_t^{(n)}\mathbf{W}_v^{(f)}. \tag{7}$$

Here, $\mathbf{W}_q^{(f)}, \mathbf{W}_k^{(f)}, \mathbf{W}_v^{(f)} \in \mathbb{R}^{d\times d}$ are learnable projection matrices shared across source markets. The factor-level information transmitted from source market n to target market m is obtained by applying multi-head attention between their dynamic latent factors:

$$\bar{\mathbf{C}}_t^{(m,n)} = \mathrm{MHA}\left(\mathbf{Q}_t^{(m)}, \mathbf{K}_t^{(m,n)}, \mathbf{V}_t^{(m,n)}\right) \in \mathbb{R}^{J\times d}. \tag{8}$$

To control the overall amount of foreign information transmitted to each target factor, we introduce a factor-specific gate:

$$\mathbf{\Gamma}_t^{(m,n)} = \sigma\left(f_{\text{source}}\left([\bar{\mathbf{U}}_t^{(n)}; \bar{\mathbf{C}}_t^{(m,n)}]\right)\right) \in (0,1)^{J\times 1}. \tag{9}$$

where $[\cdot\,;\cdot]$ denotes concatenation along the feature dimension, and $f_{\text{source}} : \mathbb{R}^{J\times 2d} \to \mathbb{R}^{J\times 1}$ is a learnable factor-wise transformation. The sigmoid function is applied elementwise, producing one scalar gate for each target factor. The gated cross-market information is aggregated as:

$$\bar{\mathbf{C}}_t^{(m)} = \sum_{n\in\mathcal{F}_m} \mathbf{\Gamma}_t^{(m,n)} \odot \bar{\mathbf{C}}_t^{(m,n)}. \tag{10}$$

Finally, the local target-market factors and the gated cross-market information are fused to obtain the dynamic latent cross-market factors:

$$\bar{\mathbf{M}}_t = f_{\text{mix}}\left([\bar{\mathbf{U}}_t^{(m)}; \bar{\mathbf{C}}_t^{(m)}]\right), \tag{11}$$

where f_{mix} is a learnable transformation implemented using a MLP layer. This cross-market interaction allows each target factor to select both relevant factors within that market, while controlling the strength of foreign information transmission.

3.2.3 Gated Systematic Attention

We next model how individual stocks respond to the resulting systematic factors. Given the stock representations

$$\bar{\mathbf{H}}_t = \{\mathbf{H}_{i,t}\}_{i=1}^N \in \mathbb{R}^{N\times d}, \tag{12}$$

and the dynamic latent cross-market factor representation $\bar{\mathbf{M}}_t$, we compute the query, key and value of systematic attention as

$$\mathbf{Q}_t^{(g)} = \bar{\mathbf{H}}_t\mathbf{W}_Q^{(g)},\ \mathbf{K}_t^{(g)} = \bar{\mathbf{M}}_t\mathbf{W}_K^{(g)},\ \mathbf{V}_t^{(g)} = \bar{\mathbf{M}}_t\mathbf{W}_V^{(g)}, \tag{13}$$

where $\mathbf{W}_Q^{(g)}, \mathbf{W}_K^{(g)}, \mathbf{W}_V^{(g)} \in \mathbb{R}^{d\times d}$ are learnable projection matrices. The attention scores are computed by

$$\mathbf{A}_t^{(g)} = \frac{\mathbf{Q}_t^{(g)} \mathbf{K}_t^{(g)\top}}{\sqrt{d}} \in \mathbb{R}^{N\times J}. \tag{14}$$

The resulting attention matrix measures the relevance of each latent systematic factor to each stock. We apply an element-wise sigmoid function to the attention scores and obtain the systematic representation

$$\bar{\mathbf{G}}_t = \sigma\left(\mathbf{A}_t^{(g)}\right) \mathbf{V}_t^{(g)} \in \mathbb{R}^{N\times d}. \tag{15}$$

Unlike the softmax operation commonly used in standard attention mechanism (Vaswani et al., 2017), the sigmoid function does not impose competition among latent factors. This allows a stock to simultaneously respond to multiple systematic factors, which is desirable when several market-wide forces affect the stock at the same time.

Applying the preceding operations across the lookback window of τ produces

$$\mathbf{G} = \{\bar{\mathbf{G}}_t\}_{t=1}^{\tau} \in \mathbb{R}^{N\times\tau\times d}, \tag{16}$$

where $\mathbf{G}$ captures the time-varying systematic representation of each stock after incorporating both local market dynamics and cross-market factor interactions. To adaptively control the amount of systematic information contributed by each stock, we compute a gate from the encoded representation:

$$\mathbf{Z} = \sigma\left(\mathbf{H}\mathbf{W}_z\right), \tag{17}$$

where $\mathbf{W}_z \in \mathbb{R}^{d\times d}$ is learnable and $\sigma(\cdot)$ is applied elementwise. The gated systematic representation is then

$$\mathbf{G}' = \mathbf{Z} \odot \mathbf{G} \in \mathbb{R}^{N\times\tau\times d}. \tag{18}$$

Thus, DCSA produces $\mathbf{G}'$, which contains the stock-level systematic information after incorporating local market dynamics and cross-market factor interactions.

The resulting systematic pathway thus represents each market through a small set of dynamic latent factors that evolve with market conditions and interact across markets. These factors subsequently provide the basis for modeling stock-level exposure to systematic effects, allowing the framework to jointly capture within-market systematic dynamics and cross-market dependencies.

3.3 Fusion

As shown in Figure 1a, the host model produces the idiosyncratic representation $\mathbf{L} \in \mathbb{R}^{N\times\tau\times d}$, which we combine with the systematic representation $\mathbf{G}'$ by

$$\mathbf{O} = \mathbf{L} + \mathbf{G}'. \tag{19}$$

The host model's output layer maps $\mathbf{O}$ to the final prediction $\hat{\mathbf{r}} \in \mathbb{R}^N$.

4 Experiments

4.1 Datasets

To evaluate the robustness of our method across different markets, we construct datasets from three major markets: China, the U.S., and Hong Kong. For each market we select the 800, 500, and 800 stocks with the highest daily trading amounts, respectively, which concentrates the universe on actively traded names with substantial market participation and ensures sufficient liquidity for reliable evaluation. For all three markets we use the publicly available Alpha158 indicators (Yang et al., 2020), which primarily capture price and trading-volume dynamics over multiple horizons of 1, 5, 10, 20, 30, and 60 trading days. The prediction label is the next-day return for US500 and HK800. For CN800 we instead predict the return realized two trading days ahead, leaving a one-day gap between the last observed bar and the predicted day to reflect the T+1 settlement rule of the Chinese A-share market. The label is a point return on the target day rather than a cumulative return over the intervening days. All datasets use a chronological split, with 2015–2022 for training, 2023 for validation, and 2024–2025 for testing. Further details are given in Appendix A.

Table 1: Performance of host models with and without DCSA. Values are means across five trials. A red subscript such as $36.1\uparrow$ denotes a 36.1% relative improvement over the corresponding host model. * indicates a statistically significant improvement at the $p = 0.05$ level using a t-test. Results with standard deviations are reported in Appendix E.

Dataset	Model	IC	ICIR	RankIC	RankICIR	AR	SR
CN800	DTML	0.061	0.42	0.091	0.59	0.20	1.2
	+ DCSA	$\mathbf{0.083}^*_{36.1\uparrow}$	$\mathbf{0.59}^*_{40.5\uparrow}$	$\mathbf{0.115}^*_{26.4\uparrow}$	$\mathbf{0.67}^*_{13.6\uparrow}$	$\mathbf{0.28}^*_{40.0\uparrow}$	$\mathbf{1.6}^*_{33.3\uparrow}$
	MASTER	0.065	0.42	0.090	0.57	0.25	1.4
	+ DCSA	$\mathbf{0.080}^*_{23.1\uparrow}$	$\mathbf{0.53}^*_{26.2\uparrow}$	$\mathbf{0.109}^*_{21.1\uparrow}$	$\mathbf{0.68}^*_{19.3\uparrow}$	$\mathbf{0.31}^*_{24.0\uparrow}$	$\mathbf{1.9}^*_{35.7\uparrow}$
	FinD3	0.078	0.53	0.101	0.60	0.23	1.3
	+ DCSA	$\mathbf{0.098}^*_{25.6\uparrow}$	$\mathbf{0.75}^*_{41.5\uparrow}$	$\mathbf{0.114}^*_{12.9\uparrow}$	$\mathbf{0.81}^*_{35.0\uparrow}$	$\mathbf{0.33}^*_{43.5\uparrow}$	$\mathbf{2.1}^*_{61.5\uparrow}$
US500	DTML	0.056	0.31	0.030	0.23	0.11	1.0
	+ DCSA	$\mathbf{0.069}^*_{23.2\uparrow}$	$\mathbf{0.35}^*_{12.9\uparrow}$	$\mathbf{0.035}^*_{16.7\uparrow}$	$\mathbf{0.24}_{4.3\uparrow}$	$\mathbf{0.24}^*_{118.2\uparrow}$	$\mathbf{1.5}^*_{50.0\uparrow}$
	MASTER	0.072	0.32	0.030	0.21	0.11	1.1
	+ DCSA	$\mathbf{0.087}^*_{20.8\uparrow}$	$\mathbf{0.35}_{9.4\uparrow}$	$\mathbf{0.041}^*_{36.7\uparrow}$	$\mathbf{0.24}_{14.3\uparrow}$	$\mathbf{0.25}^*_{127.3\uparrow}$	$\mathbf{1.6}^*_{45.5\uparrow}$
	FinD3	0.035	0.23	0.029	0.21	0.15	0.8
	+ DCSA	$\mathbf{0.068}^*_{94.3\uparrow}$	$\mathbf{0.31}^*_{34.8\uparrow}$	$\mathbf{0.033}^*_{13.8\uparrow}$	$\mathbf{0.23}_{9.5\uparrow}$	$\mathbf{0.33}^*_{120.0\uparrow}$	$\mathbf{1.3}^*_{62.5\uparrow}$
HK800	DTML	0.067	0.75	0.079	0.73	0.60	1.9
	+ DCSA	$\mathbf{0.096}^*_{43.3\uparrow}$	$\mathbf{0.80}_{6.7\uparrow}$	$\mathbf{0.099}^*_{25.3\uparrow}$	$\mathbf{0.85}^*_{16.4\uparrow}$	$\mathbf{0.75}^*_{25.0\uparrow}$	$\mathbf{2.5}^*_{31.6\uparrow}$
	MASTER	0.085	0.84	0.089	0.97	0.78	2.4
	+ DCSA	$\mathbf{0.110}^*_{29.4\uparrow}$	$\mathbf{1.05}^*_{25.0\uparrow}$	$\mathbf{0.105}^*_{18.0\uparrow}$	$\mathbf{1.01}_{4.1\uparrow}$	$\mathbf{0.88}^*_{12.8\uparrow}$	$\mathbf{3.1}^*_{29.2\uparrow}$
	FinD3	0.063	0.61	0.090	0.74	0.47	1.8
	+ DCSA	$\mathbf{0.091}^*_{44.4\uparrow}$	$\mathbf{0.75}^*_{23.0\uparrow}$	$\mathbf{0.106}^*_{17.8\uparrow}$	$\mathbf{0.90}^*_{21.6\uparrow}$	$\mathbf{0.79}^*_{68.1\uparrow}$	$\mathbf{2.7}^*_{50.0\uparrow}$

4.2 Baselines for Comparison

We applied our method to three stock prediction baselines DTML (Yoo et al., 2021), MASTER (Li et al., 2024) and FinD3 (Mei et al., 2025), including both attention-based and GNN-based stock prediction method. For FinD3, since it is a GNN method that requires fixed stock pool, we made a adaption to make it trainable on our dynamically changing stock universe. Detailed descriptions of each method can be found in Appendix B.

4.3 Experiment Setup

Each host was trained to align its predictions with the realized-return targets $\mathbf{r}$ using its corresponding host-specific objective. The objectives are detailed in Appendix C. All models were trained with five independent runs using different random initializations, and performance was reported as the mean and standard deviation across runs. Training proceeded in two stages: models were first pre-trained for 40 epochs on single markets, then fine-tuned for 10 additional epochs with cross-market interaction enabled across all three markets. All models used a fixed lookback window of length $\tau = 10$. All implementations were built on PyTorch (Paszke et al., 2019) and executed on a single NVIDIA RTX 3090 24GB GPU. The model dimension d was selected from $\{128, 256, 512\}$ via grid search on the validation set, with the feed-forward network dimension set equal to d. The DCSA method used a single attention layer throughout. The number of dynamic latent factors J was treated as a tunable hyperparameter and selected based on validation performance. For baseline methods, we followed the original implementations and reported their best-performing configurations. We adopted the ADAM optimizer (Kingma & Ba, 2017) with an initial learning rate of 1×10^{-5}. More details of training can be found in Appendix C.

4.4 Evaluation

To evaluate prediction performance, we adopted four commonly used ranking-based metrics: Information Coefficient (IC), Information Ratio of IC (ICIR), Rank Information Coefficient (RankIC), and Information Ratio of RankIC (RankICIR). In addition, we evaluated the economic value of each method using two portfolio-based metrics: Annualized Return (AR) and Sharpe Ratio (SR). For all metrics, higher values indicate better performance. More details can be found in Appendix D.

4.5 MAIN RESULTS

Table 1 reported test performance on three large-scale stock universes, CN800, US500, and HK800. Because DCSA was a model-agnostic method, each block paired a host model with its DCSA-augmented counterpart, so every row pair isolated the contribution of the DCSA pathway while holding the host architecture, features, and training protocol fixed. Across the 9 groups of comparisons, adding DCSA improved the host on every metric and most improvements were statistically significant. Relative IC improvements ranged from 21% to 94%, with eight of the nine pairs falling between 21% and 44%, and annualized returns and Sharpe ratios improved in every configuration. Only 6 of the 54 metric comparisons failed to reach significance.

The pattern of improvement differed across markets. On CN800, all four predictive metrics and both portfolio metrics improved significantly for every host. HK800 showed a similar pattern, with one exception: MASTER's RankICIR increased from 0.97 to 1.01 without reaching significance. On US500, the gains were concentrated in IC, RankIC, and the portfolio metrics, where annualized return roughly doubled for all three hosts and Sharpe ratios rose accordingly. ICIR improved significantly for DTML and FinD3 but not for MASTER, while RankICIR improved for every host without reaching significance in any case. This consistent lack of significant improvement in RankICIR across all three hosts suggested a systematic limitation of DCSA on US500 rather than a host-specific effect.

Two cases further illustrated that the gains were not simply determined by the strength of the host model. FinD3 was the weakest US500 baseline at 0.035 IC, but DCSA nearly doubled its IC to 0.068, corresponding to a 94% relative gain. In contrast, FinD3 was the strongest CN800 baseline at 0.078 IC, outperforming every other model without DCSA, but the method still increased its IC by 26%. Thus, strong performance of a host model in one market did not eliminate the benefit of cross-market information in another, suggesting that DCSA provided complementary systematic information beyond that captured by the host architecture alone. The consistency of these gains across geographically and structurally distinct markets, and across both attention-based and GNN-based relationship methods, further suggested that dynamic cross-market latent factors captured systematic structure that was not specific to a particular architecture or market.

As an additional reference, we compared the DCSA-augmented DTML, MASTER, and FinD3 models with MCI-GRU (Zhu et al., 2025) and FactorGCL (Duan et al., 2025), two methods based on static latent factors in Appendix E. Since both methods are GNN-based, we applied the same adaptation described in Section 4.2 to enable training on our dynamic stock universe. In addition, because the authors of FactorGCL did not release the source code, we reproduced the method based on its description and tuned its implementation to our data. The DCSA-augmented models achieved higher mean IC than both baselines in all nine host–market configurations. They also achieved higher mean values on every other reported metric, except for two ties: FinD3+DCSA and FactorGCL obtained the same ICIR on US500, while DTML+DCSA and MCI-GRU obtained the same RankIC on HK800. Thus, DCSA consistently achieved performance at least comparable to, and generally better than, the static latent-factor baselines across both predictive and portfolio metrics. Because these methods differ in their host architectures and factor designs, this comparison is not a controlled ablation. Nevertheless, it provides additional evidence that dynamically adapting latent factors and modeling their cross-market interactions can offer broader benefits than static-factor representations. Detailed metric-by-metric comparisons are provided in Appendix E.

4.6 ABLATION STUDY

We performed an ablation study using three variants of each host model. The original host model without a systematic-attention pathway served as the baseline. **Static Systematic Attention (SSA)** added static latent factors ($\bar{\mathbf{P}}_t^{(m)}$ from Section 3.2.1) from the target market, **Dynamic Systematic Attention (DSA)** replaced the static factors with dynamic latent factors ($\bar{\mathbf{U}}_t^{(m)}$ from Section 3.2.2), and **DCSA** further incorporated dynamic cross-market information from the source markets. In the ablation tables, these variants were indicated by checkmarks in the corresponding SSA, DSA, and DCSA columns, while rows without any checkmark denoted the original prediction model. The DCSA variant used the target market together with both available source markets and corresponded to the "+DCSA" rows in Table 1. Results for DTML were reported in Table 2, while results for MASTER and FinD3 were provided in Appendix E.1.

Table 2: Ablation study of DTML. Bold values indicate the highest value within each dataset block. * indicates a statistically significant improvement (t-test by $p = 0.05$) over the preceding row.

Dataset	SSA	DSA	DCSA	Metrics					
				IC	ICIR	RankIC	RankICIR	AR	SR
CN800				0.061	0.42	0.091	0.59	0.20	1.2
	✓			0.061	0.43	0.095	0.59	0.20	1.1
		✓		0.070*	0.47*	0.103*	0.63	0.24*	1.4*
			✓	**0.083***	**0.59***	**0.115***	**0.67**	**0.28***	**1.6***
US500				0.056	0.31	0.030	0.23	0.11	1.0
	✓			0.062*	0.31	0.030	0.22	0.12	1.0
		✓		0.066	0.34	0.034*	0.23	0.20*	1.4*
			✓	**0.069**	**0.35**	**0.035**	**0.24**	**0.24***	**1.5**
HK800				0.067	0.75	0.079	0.73	0.60	1.9
	✓			0.073*	**0.80***	0.088*	**0.87***	0.65	2.0
		✓		0.084*	0.75	0.090	0.77	0.68	2.2
			✓	**0.096***	**0.80***	**0.099***	0.85*	**0.75***	**2.5***

Home-market systematic attention. Comparing DSA with the host baseline showed that dynamic target-market factors improved IC on every host–market combination. IC increased by 14.8%, 13.8%, and 10.3% on CN800 for DTML, MASTER, and FinD3, respectively; by 17.9%, 13.9%, and 62.9% on US500; and by 25.4%, 18.8%, and 12.7% on HK800. Because Host and DSA used the same input information, these gains reflected the explicit representation of systematic structure rather than the use of additional data. The results suggested that constructing dynamic market factors and allowing stocks to attend to them was more effective than leaving systematic information implicit in the host architectures. The largest improvement occurred for FinD3 on US500, where the dynamic home-market factors substantially improved the weakest baseline.

Comparing SSA with DSA isolated the contribution of time variation in the latent factors. Dynamic factors improved IC over static factors in all nine host–market comparisons, with relative improvements ranging from 4.4% to 15.1%. They also improved ICIR and RankIC in most cases. Averaged over the nine comparisons, SSA recovered approximately 49% of the IC improvement achieved by DSA over the Host baseline. These results showed that allowing latent factors to adapt dynamically to changing market conditions provided consistent predictive gains over the static factors used in SSA. More detailed comparisons can be found in Appendix E.1.

Cross-market transmission. Adding source-market information further improved the dynamic home-market variant. Relative to DSA, DCSA increased IC in all nine host–market blocks, with incremental gains ranging from 4.5% to 28.2%. The gain was smallest for DTML on US500 (4.5%) and largest for FinD3 on HK800 (28.2%). DCSA also improved both AR and SR over DSA in all nine cases and achieved the best or joint-best value on every reported metric. These results indicated that dynamic source-market factors provide complementary information beyond the target market's own systematic representation. To further examine this pattern, we checked the influence from each source market and conducted a transmission-gate analysis. As shown in Appendix E.2, DTML and MASTER used source markets more conservatively on US500, whereas FinD3 exhibited more volatile gate dynamics, consistent with its larger cross-market gain. Further details are provided in Appendix E.2.

5 Conclusion

In this paper, we presented DCSA, a dynamic cross-market systematic attention method that captures time-varying systematic information from multiple markets and integrates it into existing stock prediction models. DCSA was model-agnostic and could be added to different host architectures without changing their original structures. Experiments on three markets showed consistent improvements across different host models, demonstrating the effectiveness of incorporating dynamic cross-market systematic information. Ablation studies further showed that both dynamic within-market factors and cross-market information contributed to the improvements, with their relative importance varying across markets. These results highlighted the value of explicitly modeling systematic cross-market dependencies for stock return prediction. Future work could explore more connections from markets beyond our research.

AI USE STATEMENT

In this work, we used generative AI tools for none of the tasks with required disclosure. We have not used generative AI tools for any of the tasks with required disclosure. Additionally, we used generative AI tools for editing a research paper to improve readability. We have reviewed all AI-assisted work. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

The DCSA architecture and its mathematical formulation are described in Section 3. The experimental datasets, market selection, chronological splits, and preprocessing procedures are described in Section 4.1 and Appendix A. The training protocol, hyperparameter selection, and hardware configuration are reported in Section 4.3, while the evaluation metrics and portfolio construction procedure are provided in Appendix D. Anonymous supplementary materials provide the source code. The CRSP and CSMAR data used in this work are obtained under licensed access and therefore cannot be redistributed.

REFERENCES

Tobias Adrian and Francesco Franzoni. Learning about beta: Time-varying factor loadings, expected returns, and the conditional capm. *Journal of Empirical Finance*, 16(4):537–556, 2009.

Kyunghyun Cho, B van Merrienboer, Caglar Gulcehre, F Bougares, H Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. In *Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)*, 2014.

Kangmin Choi, Geon Shin, Jungwoo Yang, and Hyunjoon Kim. Dycor: Capturing hidden stock relationships for stock trend prediction. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management*, CIKM '25, pp. 458–467, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400720406. doi: 10.1145/3746252.3761413.

Yuntao Du, Jindong Wang, Wenjie Feng, Sinno Pan, Tao Qin, Renjun Xu, and Chongjun Wang. Adarnn: Adaptive learning and forecasting of time series. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, CIKM '21, pp. 402–411, New York, NY, USA, 2021. Association for Computing Machinery.

Yitong Duan, Lei Wang, Qizhong Zhang, and Jian Li. Factorvae: A probabilistic dynamic factor model based on variational autoencoder for predicting cross-sectional stock returns. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(4):4468–4476, Jun. 2022.

Yitong Duan, Weiran Wang, and Jian Li. Factorgcl: A hypergraph-based factor model with temporal residual contrastive learning for stock returns prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 173–181, 2025.

Eugene F. Fama and Kenneth R. French. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56, 1993. ISSN 0304-405X. doi: https://doi.org/10.1016/0304-405X(93)90023-5.

Qingliang Fan, Ruike Wu, Yanrong Yang, and Wei Zhong. Time-varying minimum variance portfolio. *Journal of Econometrics*, 239(2):105339, 2024. ISSN 0304-4076. doi: https://doi.org/10.1016/j.jeconom.2022.08.007.

Fuli Feng, Xiangnan He, Xiang Wang, Cheng Luo, Yiqun Liu, and Tat-Seng Chua. Temporal relational ranking for stock prediction. *ACM Trans. Inf. Syst.*, 37(2), mar 2019. ISSN 1046-8188.

Shihao Gu, Bryan Kelly, and Dacheng Xiu. Empirical Asset Pricing via Machine Learning. *The Review of Financial Studies*, 33(5):2223–2273, 02 2020. ISSN 0893-9454.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Comput.*, 9(8):1735–1780, nov 1997. ISSN 0899-7667.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.

Tong Li, Zhaoyang Liu, Yanyan Shen, Xue Wang, Haokun Chen, and Sen Huang. Master: Market-guided stock transformer for stock price forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(1):162–170, Mar. 2024.

Zhengze Li, Zilin Song, Tingting Yuan, and Xiaoming Fu. Cograsp: Co-occurrence graph based stock price forecasting. In James Kwok (ed.), *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI-25*, pp. 7527–7535. International Joint Conferences on Artificial Intelligence Organization, 8 2025. doi: 10.24963/ijcai.2025/837. Main Track.

Di Luo, Weiheng Liao, Shuqi Li, Xin Cheng, and Rui Yan. Causality-guided multi-memory interaction network for multivariate stock price movement prediction. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki (eds.), *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 12164–12176, Toronto, Canada, July 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.acl-long.679.

Jieyuan Mei, Jindong Tian, Ronghui Xu, Hanyue Wei, Chenjuan Guo, and Bin Yang. Find3: A dual 3d state space model with dynamic hypergraph for financial stock prediction. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management*, CIKM '25, pp. 2084–2094, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400720406. doi: 10.1145/3746252.3761239.

Zisheng Ouyang, Zhen Chen, Xuewei Zhou, and Zhongzhe Ouyang. Imported risk in global financial markets: Evidence from cross-market connectedness. *The North American Journal of Economics and Finance*, 76:102374, 2025. ISSN 1062-9408. doi: https://doi.org/10.1016/j.najef.2025.102374.

Kijeong Park, Sungchul Hong, and Jong-June Jeon. Dynamic higher-order relations and event-driven temporal modeling for stock price forecasting. In James Kwok (ed.), *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI-25*, pp. 6048–6056. International Joint Conferences on Artificial Intelligence Organization, 8 2025. doi: 10.24963/ijcai.2025/673. Main Track.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

Hao Qian, Hongting Zhou, Qian Zhao, Hao Chen, Hongxiang Yao, Jingwei Wang, Ziqi Liu, Fei Yu, Zhiqiang Zhang, and Jun Zhou. Mdgnn: Multi-relational dynamic graph neural network for comprehensive and dynamic stock investment prediction. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(13):14642–14650, Mar. 2024.

Lingfei Qian, Xueqing Peng, Hanley Smith, Yi Han, Yueru He, Haohang Li, Yupeng Cao, Yangyang Yu, Guojun Xiong, Peng Lu, Yan Wang, Vincent Jim Zhang, Huan He, Alejandro Lopez-Lira, Jimin Huang, Jian-Yun Nie, and Sophia Ananiadou. When agents trade: Live multi-market trading arena for llm agents. In *Proceedings of the ACM Web Conference 2026*, WWW '26, pp. 7833–7844, New York, NY, USA, 2026. Association for Computing Machinery. ISBN 9798400723070. doi: 10.1145/3774904.3792821.

Yao Qin, Dongjin Song, Haifeng Chen, Wei Cheng, Guofei Jiang, and Garrison W. Cottrell. A dual-stage attention-based recurrent neural network for time series prediction. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, pp. 2627–2633, 2017.

Yinghua Ren, Wanru Zhao, Wanhai You, and Huiming Zhu. Multiscale features of extreme risk spillover networks among global stock markets. *The North American Journal of Economics and Finance*, 62:101754, 2022. ISSN 1062-9408. doi: https://doi.org/10.1016/j.najef.2022.101754.

Stephen A Ross. The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3): 341–360, 1976. ISSN 0022-0531. doi: https://doi.org/10.1016/0022-0531(76)90046-6.

Ramit Sawhney, Shivam Agarwal, Arnav Wadhwa, Tyler Derr, and Rajiv Ratn Shah. Stock selection via spatiotemporal hypergraph attention network: A learning to rank approach. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(1):497–504, May 2021.

William F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk*. *The Journal of Finance*, 19(3):425–442, 1964.

Lingyun Song, Haodong Li, Siyu Chen, Xinbiao Gan, Binze Shi, Jie Ma, Yudai Pan, Xiaoqi Wang, and Xuequn Shang. Multi-scale temporal neural network for stock trend prediction enhanced by temporal hyepredge learning. In James Kwok (ed.), *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI-25*, pp. 3272–3280. International Joint Conferences on Artificial Intelligence Organization, 8 2025. doi: 10.24963/ijcai.2025/364. Main Track.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *International Conference on Learning Representations*, 2018.

Gang-Jin Wang, Li Wan, Yusen Feng, Chi Xie, Gazi Salah Uddin, and You Zhu. Interconnected multilayer networks: Quantifying connectedness among global stock and foreign exchange markets. *International Review of Financial Analysis*, 86:102518, 2023. ISSN 1057-5219. doi: https://doi.org/10.1016/j.irfa.2023.102518.

Heyuan Wang, Tengjiao Wang, Shun Li, Jiayi Zheng, Shijie Guan, and Wei Chen. Adaptive long-short pattern transformer for stock investment selection. In Lud De Raedt (ed.), *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, pp. 3970–3977. International Joint Conferences on Artificial Intelligence Organization, 7 2022. Main Track.

Hongjie Xia, Huijie Ao, Long Li, Yu Liu, Sen Liu, Guangnan Ye, and Hongfeng Chai. Ci-sthpan: Pre-trained attention network for stock selection with channel-independent spatio-temporal hypergraph. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(8):9187–9195, Mar. 2024.

Yumo Xu and Shay B. Cohen. Stock movement prediction from tweets and historical prices. In Iryna Gurevych and Yusuke Miyao (eds.), *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1970–1979, Melbourne, Australia, July 2018. Association for Computational Linguistics. doi: 10.18653/v1/P18-1183.

Xiao Yang, Weiqing Liu, Dong Zhou, Jiang Bian, and Tie-Yan Liu. Qlib: An ai-oriented quantitative investment platform, 2020.

Jaemin Yoo, Yejun Soun, Yong-chan Park, and U Kang. 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*, KDD '21, pp. 2037–2045, New York, NY, USA, 2021. Association for Computing Machinery.

Yangyang Yu, Zhiyuan Yao, Haohang Li, Zhiyang Deng, Yuechen Jiang, Yupeng Cao, Zhi Chen, Jordan W. Suchow, Zhenyu Cui, Rong Liu, Zhaozhuo Xu, Denghui Zhang, Koduvayur Subbalakshmi, Guojun Xiong, Yueru He, Jimin Huang, Dong Li, and Qianqian Xie. Fincon: A synthesized llm multi-agent system with conceptual verbal reinforcement for enhanced financial decision making. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 137010–137045. Curran Associates, Inc., 2024. doi: 10.52202/079017-4354.

Peng Zhu, Yuante Li, Yifan Hu, Sheng Xiang, Qinyuan Liu, Dawei Cheng, and Yuqi Liang. Mci-gru: Stock prediction model based on multi-head cross-attention and improved gru. *Neurocomputing*, 638:130168, 2025. ISSN 0925-2312. doi: https://doi.org/10.1016/j.neucom.2025.130168.

A DATASETS

Because the three markets trade in different time zones and observe different holiday calendars, we align them on a common date index and admit a foreign market's features for a given date only once that market's session has closed. Cross-market information therefore reaches the home market with a lag of at least one session, so no prediction uses foreign observations that were unavailable at the time the position would have been taken. Dates on which a market did not trade are excluded from that market's panel rather than forward-filled.

All three datasets used in this paper were constructed using price and volume data provided by the Center for Research in Securities Prices (CRSP) and China Stock Market & Accounting Research Database (CSMAR). These data were accessed through a licensed institutional subscription, and they are subject to copyright and usage restrictions. As such, we are not permitted to redistribute the data. We cite CRSP and CSMAR as the data providers and respect the terms of use specified under our subscription agreement. The reason we chose CRSP and CSMAR as the data sources is because they provide information for handling events like dividends and stock splits, which was necessary in preprocessing of stock data. Besides, they also include stocks that delisted from the market. This information are typically not available in publicly available datasets. As a result, we decided to use this data source instead of publicly available alternatives.

B BASELINES

We selected three typical host models that represent distinct families of relationship-based methods, so that DCSA's model-agnostic claim was tested against architectures whose native mechanisms for capturing stock dependencies differ.

- **DTML** (Yoo et al., 2021): an attention-based method that models both temporal patterns within individual stocks and dynamic correlations across stocks. DTML first summarizes each stock's multivariate historical features with an attention LSTM, producing an individual context vector. It then combines this stock-specific context with a global market context extracted from historical index data to construct a multi-level context for each stock. A data-axis multi-head self-attention module subsequently models dynamic dependencies across stocks, with different query and key projections allowing the influence between stocks to vary by direction.
- **MASTER** (Li et al., 2024): an attention-based stock forecasting method that alternates intra-stock and inter-stock aggregation to preserve temporal details while modeling dynamic cross-stock correlations. It first uses current and historical market-index price and trading-volume information to construct a market-status representation, which generates feature-wise gates for market-guided feature selection. The gated stock features then undergo intra-stock attention across time and inter-stock attention across stocks, allowing MASTER to capture cross-time stock correlations. A temporal attention layer subsequently aggregates the resulting temporal embeddings into a stock-level representation for prediction.
- **FinD3** (Mei et al., 2025): a GNN-based method that replaces attention with a selective scan over the temporal axis, combined with an explicit hypergraph over stocks to represent higher-order relations that pairwise graphs cannot express. To accommodate the changing stock universe, we adapted FinD3 by dynamically indexing the hypergraph according to the stocks available in each market-period, allowing constituents to enter or leave without assuming a fixed stock set.

We also used the DCSA-augmented host models to directly compare to the two methods that applied static latent factors: MCI-GRU (Zhu et al., 2025) and FactorGCL (Duan et al., 2025).

- **MCI-GRU**: a stock prediction model that combines an improved GRU, graph attention, and multi-head cross-attention: it replaces the GRU reset gate with an attention mechanism to better select useful historical signals, uses GAT to capture cross-sectional relationships among stocks, and learns static latent market factors that interact with both temporal and cross-sectional features.
- **FactorGCL**: a GNN-based stock prediction method. It models stocks as nodes and factors as hyperedges, first using known prior factors such as industries, then mining static latent factors from the remaining residual information, and finally modeling stock-specific alpha.

C Experiment Setup

Training proceeded in two stages. In the first stage, each model was trained for 40 epochs on its home market alone, with DCSA constructing dynamic latent factors from home-market state only and all foreign sources disabled. In the second stage, the foreign factor sources and their per-source gates were enabled and the full model was trained for a further 10 epochs on all three markets jointly, with the home-market parameters left unfrozen so that the host representation could adapt to the cross-market pathway. For each method in each market, the corresponding J latent factors selected as a hyperparameter based on the IC over validation set. The corresponding J were reported in Table 3.

Table 3: Number of dynamic latent factors J for each host model and market.

Dataset	DTML	MASTER	FinD3
CN800	5	5	10
US500	10	10	20
HK800	10	5	20

C.1 Host Model Specific Modules and Training Objectives

During the implementation , we sticked to the original modules used in the baselines. For Temporal Encoder, DTML used an attention LSTM layer, MASTER used an attention layer and FinD3 used the raw input features. For output layer, DTML used a linear layer, MASTER used an attention mechanism with a linear layer and FinD3 used a MLP layer.

For training objectives, DCSA sticked to the original host methods as much as possible. For MASTER, we implemented its original MSE loss function:

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N}(r_i - \hat{r}_i)^2. \tag{20}$$

For DTML, since its original task was set to binary classification, we used the numerical output from the method as the prediction and applied the same MSE loss. For FinD3, we inherited its objective function introduced below, as it differs from a plain return-level MSE loss.

FinD3. FinD3 (Mei et al., 2025) combines a supervised regression objective with a pairwise ranking loss and a hypergraph regularization term. Its regression loss is

$$\mathcal{L}_{\text{reg}} = \frac{1}{N}\sum_{i=1}^{N}\|\hat{\mathbf{r}}_i - \mathbf{r}_i\|_2^2\,, \tag{21}$$

where $\hat{\mathbf{r}}_i$ and $\mathbf{r}_i$ denote the predicted and realized target sequences for stock i, respectively. The returns derived from these predictions are additionally optimized using the pairwise ranking loss

$$\mathcal{L}_{\text{rank}} = \sum_{i=1}^{N}\sum_{j=1}^{N}\phi((\hat{r}_i - \hat{r}_j)(r_i - r_j))\,, \qquad \phi(x) = \max(0, -x), \tag{22}$$

which penalizes disagreement between the predicted and realized cross-sectional rankings. FinD3 also includes the hypergraph Laplacian regularizer $\mathcal{L}_G$, which encourages smooth representations among stocks connected by the learned hypergraph. Its complete objective is

$$\mathcal{L}_{\text{FinD3}} = \mathcal{L}_{\text{reg}} + \lambda_{\text{rank}}\mathcal{L}_{\text{rank}} + \lambda_G\mathcal{L}_G, \tag{23}$$

where λ_{rank} and λ_G control the contributions of the ranking and hypergraph regularization terms, respectively.

D Evaluation Metrics

To evaluate prediction accuracy, we used four metrics: Information Coefficient (IC), Information Ratio-based IC (ICIR), Rank Information Coefficient (RankIC), and Information Ratio-based RankIC (RankICIR). The Information Coefficient (IC) was defined as the Pearson correlation coefficient between predictions and actual labels, while RankIC was the Spearman correlation coefficient between predictions and labels. These metrics quantified prediction accuracy and ranking prediction accuracy, respectively. IC and RankIC were computed as the mean values across all trading days in the test set. The resulting values lay within the range of -1 to 1, with higher values indicating better model performance. Additionally, ICIR and RankICIR were derived by normalizing IC and RankIC with their respective standard deviations, providing standardized measures of performance. These metrics, widely adopted in prior work (Yang et al., 2020; Du et al., 2021; Li et al., 2024), were standard for assessing stock price prediction performance.

We also adopted two portfolio-based metrics Annualized Return (AR) and Sharpe Ratio (SR), to evaluate the profit potential of each method. AR measured the annualized return of the portfolio, directly shows the portfolio's performance. SR captured the risk-adjusted return, computed as the portfolio's excess return over risk-free rate divided by its standard deviation, reflecting both profitability and stability. To evaluate portfolio-based performance under conditions closer to real-world trading, we simulated a portfolio starting with an initial capital of 100 million. The portfolio consistently held 50 stocks, with capital equally allocated across them. Each trading day, the 5 stocks with the lowest predicted scores in the portfolio were sold and replaced with the 5 highest-scoring stocks from the remaining market. In the A-share market, we applied a rate of 0.25% of the trading value for selling and 0.15% for buying. In the U.S. stock market, since many online brokers do not charge commissions. We therefore include the SEC transaction fee of 0.00206%, as trading cost. In the Hong Kong stock market, we assumed a transaction cost of 0.01085% of the trading value based on the rate provided by Shanghai-Hong Kong Stock Connect and Shenzhen-Hong Kong Stock Connect.

E Supplementary Results

Table 4 demonstrated the results from Table 1 with standard deviation.

Comparison with static-factor baselines. Table 5 compares the DCSA-augmented DTML, MASTER, and FinD3 models with MCI-GRU and FactorGCL, two methods that incorporate static latent factor representations. The DCSA-augmented models achieved the highest IC in all nine host–market configurations, with improvements over the strongest static-factor baseline ranging from 17.6% to 44.1% on CN800, 3.0% to 31.8% on US500, and 21.3% to 46.7% on HK800. The advantage was also observed broadly across ICIR, RankIC, RankICIR, AR, and SR, although few metrics tied to baselines. These results provided complementary evidence that dynamically constructed and cross-market-aware latent factors remained effective compared with existing methods based on static factor representations.

E.1 Static versus Dynamic Latent Factors

As shown in Tables 2, 6 and 7, dynamic latent factors improved IC in all nine host–market cases, by 4.4% to 15.1%. DTML benefited most, with significant gains of 15.1% on HK800 and 14.8% on CN800; the remaining seven cases fell between 4.4% and 7.5%. The two cases that did not reach significance were MASTER on US500 ($0.078 \rightarrow 0.082$) and FinD3 on HK800 ($0.068 \rightarrow 0.071$), in

Table 4: Performance of host models with and without DCSA method. Metrics are reported as mean $\pm$ standard deviation across 5 trials. * indicates a statistically significant improvement (by a two-sided t-test at $p = 0.05$ level) over the corresponding host model without DCSA.

Dataset	Model	IC	ICIR	RankIC	RankICIR	AR	SR
CN800	DTML	0.061 ± 0.002	0.42 ± 0.01	0.091 ± 0.002	0.59 ± 0.01	0.20 ± 0.05	1.2 ± 0.2
	+ DCSA	$0.083^* \pm 0.003$	$0.59^* \pm 0.02$	$0.115^* \pm 0.002$	$0.67^* \pm 0.01$	$0.28^* \pm 0.03$	$1.6^* \pm 0.1$
	MASTER	0.065 ± 0.002	0.42 ± 0.01	0.090 ± 0.004	0.57 ± 0.01	0.25 ± 0.05	1.4 ± 0.1
	+ DCSA	$0.080^* \pm 0.001$	$0.53^* \pm 0.02$	$0.109^* \pm 0.003$	$0.68^* \pm 0.02$	$0.31^* \pm 0.04$	$1.9^* \pm 0.2$
	FinD3	0.078 ± 0.002	0.53 ± 0.02	0.101 ± 0.003	0.60 ± 0.02	0.23 ± 0.04	1.3 ± 0.1
	+ DCSA	$0.098^* \pm 0.001$	$0.75^* \pm 0.02$	$0.114^* \pm 0.003$	$0.81^* \pm 0.03$	$0.33^* \pm 0.02$	$2.1^* \pm 0.1$
US500	DTML	0.056 ± 0.003	0.31 ± 0.02	0.030 ± 0.003	0.23 ± 0.02	0.11 ± 0.02	1.0 ± 0.2
	+ DCSA	$0.069^* \pm 0.003$	$0.35^* \pm 0.02$	$0.035^* \pm 0.002$	0.24 ± 0.03	$0.24^* \pm 0.02$	$1.5^* \pm 0.1$
	MASTER	0.072 ± 0.004	0.32 ± 0.02	0.030 ± 0.004	0.21 ± 0.04	0.11 ± 0.02	1.1 ± 0.1
	+ DCSA	$0.087^* \pm 0.004$	0.35 ± 0.01	$0.041^* \pm 0.003$	0.24 ± 0.02	$0.25^* \pm 0.04$	$1.6^* \pm 0.1$
	FinD3	0.035 ± 0.002	0.23 ± 0.01	0.029 ± 0.002	0.21 ± 0.01	0.15 ± 0.03	0.8 ± 0.1
	+ DCSA	$0.068^* \pm 0.004$	$0.31^* \pm 0.02$	$0.033^* \pm 0.003$	0.23 ± 0.02	$0.33^* \pm 0.04$	$1.3^* \pm 0.2$
HK800	DTML	0.067 ± 0.002	0.75 ± 0.08	0.079 ± 0.004	0.73 ± 0.07	0.60 ± 0.03	1.9 ± 0.1
	+ DCSA	$0.096^* \pm 0.003$	0.80 ± 0.04	$0.099^* \pm 0.002$	$0.85^* \pm 0.05$	$0.75^* \pm 0.04$	$2.5^* \pm 0.2$
	MASTER	0.085 ± 0.001	0.84 ± 0.03	0.089 ± 0.002	0.97 ± 0.02	0.78 ± 0.05	2.4 ± 0.2
	+ DCSA	$0.110^* \pm 0.002$	$1.05^* \pm 0.05$	$0.105^* \pm 0.003$	1.01 ± 0.05	$0.88^* \pm 0.07$	$3.1^* \pm 0.2$
	FinD3	0.063 ± 0.001	0.61 ± 0.02	0.090 ± 0.001	0.74 ± 0.02	0.47 ± 0.02	1.8 ± 0.1
	+ DCSA	$0.091^* \pm 0.003$	$0.75^* \pm 0.04$	$0.106^* \pm 0.002$	$0.90^* \pm 0.03$	$0.79^* \pm 0.03$	$2.7^* \pm 0.2$

Table 5: Comparison with methods using static latent factors. Values are means over five independent trials. Bold values indicate the highest mean within each dataset block. * indicates a statistically significant improvement over both static-factor baselines, MCI-GRU and FactorGCL, based on a (t-test at $p = 0.05$).

Dataset	Model	IC	ICIR	RankIC	RankICIR	AR	SR
CN800	MCI-GRU	0.068	0.52	0.080	0.58	0.15	0.8
	FactorGCL	0.062	0.42	0.097	0.50	0.24	1.3
	DTML+DCSA	0.083	0.59	**0.115***	0.67	0.28	1.6
	MASTER+DCSA	0.080	0.53	0.109	0.68	0.31	1.9
	FinD3+DCSA	**0.098***	**0.75***	0.114	**0.81***	**0.33***	**2.1***
US500	MCI-GRU	0.049	0.27	0.031	0.22	0.22	1.2
	FactorGCL	0.066	0.31	0.030	0.21	0.14	1.1
	DTML+DCSA	0.069	**0.35***	0.035	**0.24**	0.24	1.5
	MASTER+DCSA	**0.087***	**0.35***	**0.041***	**0.24**	**0**.25	**1.6***
	FinD3+DCSA	0.068	0.31	0.033	0.23	**0.33***	1.3
HK800	MCI-GRU	0.066	0.60	0.099	0.83	0.69	1.8
	FactorGCL	0.075	0.72	0.089	0.84	0.72	2.3
	DTML+DCSA	0.096	0.80	0.099	0.85	0.75	2.5
	MASTER+DCSA	**0.110***	**1.05***	0.105*	**1.01***	**0.88***	**3.1***
	FinD3+DCSA	0.091	0.75	**0.106**	0.90	0.79	2.7

both cases because the trial-to-trial spread was comparable to the gap rather than because the gain was absent. Conditioning also improved ICIR in eight of nine cases and RankIC in seven.

Decomposing the home-only gain showed why adaptation to evolving market conditions was necessary rather than merely helpful. Across the nine cases, static latent factors recovered 49% of the IC improvement over the bare host on average, leaving the remainder to time variation. The split was informative about where systematic structure was stable. On US500 the static variant recovered most of the gain for every host, indicating that much of what DCSA supplied on that market was the factor-exposure representation itself. CN800 was the opposite case. For DTML static latent factors produced no IC gain at all ($0.061 \rightarrow 0.061$), even though RankIC rose from 0.091 to 0.095, so the whole of the IC improvement there came from making the factors dynamic. For MASTER and FinD3, the dynamic component accounted for 56% and 63% of the CN800 gain respectively. Overall, static factors captured persistent systematic structure, whereas dynamic adaptation was essential when the relevant market relationships varied more substantially across time, with its importance depending on both the host architecture and the target market.

Table 6: Ablation study of MASTER. Values are means over five independent trials. Bold values indicate the highest value within each dataset block. * indicates a statistically significant improvement (t-test by $p = 0.05$) over the preceding row.

Dataset	SSA	DSA	DCSA	IC	ICIR	RankIC	RankICIR	AR	SR
CN800				0.065	0.42	0.090	0.57	0.25	1.4
	✓			0.069	0.45	0.093	0.59	0.25	1.3
		✓		0.074*	0.47	0.097	0.63	0.27	1.4
			✓	**0.080***	**0.53***	**0.109***	**0.68***	**0.31***	**1.9***
US500				0.072	0.32	0.030	0.21	0.11	1.1
	✓			0.078*	0.31	0.034*	0.22	0.15*	1.2
		✓		0.082	0.34	0.037	0.22	0.21*	1.5*
			✓	**0.087***	**0.35**	**0.041***	**0.24**	**0.25***	**1.6**
HK800				0.085	0.84	0.089	0.97	0.78	2.4
	✓			0.095*	0.91	0.095	0.98	0.80	2.4
		✓		0.101*	0.98	0.099	0.98	0.83	2.8*
			✓	**0.110***	**1.05***	**0.105**	**1.01**	**0.88***	**3.1***

Table 7: Ablation study of FinD3. Values are means over five independent trials. Bold values indicate the highest value within each dataset block. * indicates a statistically significant improvement (t-test by $p = 0.05$) over the preceding row.

Dataset	SSA	DSA	DCSA	IC	ICIR	RankIC	RankICIR	AR	SR
CN800				0.078	0.53	0.101	0.60	0.23	1.3
	✓			0.081	0.53	0.107	0.61	0.24	1.3
		✓		0.086*	0.57*	0.111	0.66*	0.28*	1.5*
			✓	**0.098***	**0.75***	**0.114**	**0.81***	**0.33***	**2.1***
US500				0.035	0.23	0.029	0.21	0.15	0.8
	✓			0.053*	0.26*	0.031	0.20	0.20*	1.0*
		✓		0.057	0.28	0.030	0.20	0.28*	1.2*
			✓	**0.068***	**0.31***	**0.033**	**0.23***	**0.33***	**1.3**
HK800				0.063	0.61	0.090	0.74	0.47	1.8
	✓			0.068	0.63	0.104*	0.88*	0.58*	2.1*
		✓		0.071	0.66	0.103	0.88	0.69*	2.5*
			✓	**0.091***	**0.75***	**0.106**	**0.90**	**0.79***	**2.7**

E.2 Cross-Market Interaction Analysis

To examine the influence from each source market, we implemented DCSA with a single source market in each target market. The results are shown in Tables 8-10. DCSA(n) denotes the addition of foreign market n as a factor source, allowing the home-market factors to receive directed, factor-level information through learned source-specific gates. The rows with checkmarks from two source markets represented the full model. As shown in Tables 8-10, opening the method to foreign factor sources improved IC in all nine host–market configurations. On CN800, the best two-source configuration improved over the home-only baseline by 18.6%, 8.1%, and 14.0% for DTML, MASTER, and FinD3. On US500, the corresponding gains were 4.5%, 6.1%, and 19.3%, and on HK800, 14.3%, 8.9%, and 28.2%. Expressed as a share of the total IC gain contributed by DCSA (measured from the bare host to the full two-source configuration), cross-market transmission accounted for 44% on average: 40–60% on CN800, 36–71% on HK800, and 23–33% on US500. Among the three markets, US500 drew the least from the cross-market pathway, which was a consistent pattern all three host models. On the contrary, the two Asian markets demonstrated larger gain from cross-market interaction. On CN800 and HK800 the balance between the two mechanisms was host-dependent: for DTML and MASTER the home-only step carried at least half of the total gain on HK800, whereas for FinD3 the cross-market step carried 71% of it, the largest share anywhere in the study. That the same market could be dominated by either mechanism depending on the host indicated that the two were substitutes to a degree, with the factor pathway supplying structure that a sufficiently expressive host could partly recover on its own.

The single-source rows exposed a largely consistent directional asymmetry. For CN800 and HK800, US was the stronger foreign source in five of the six cases, reflecting the U.S. market's influence over global equity markets; the exception was FinD3 on CN800, where HK led on all four metrics (0.094 against 0.090 IC, and 0.75 against 0.67 RankICIR). The pattern reversed on US500, where

Table 8: Ablation study of single-source and dual-source dynamic systematic attention for DTML.

Dataset	DSA	DCSA(CN)	DCSA(US)	DCSA(HK)	Metrics					
					IC	ICIR	RankIC	RankICIR	AR	SR
CN800	✓				0.070 ± 0.002	0.47 ± 0.04	0.103 ± 0.002	0.63 ± 0.02	0.24 ± 0.06	1.4 ± 0.3
	✓		✓		0.079 ± 0.003	0.53 ± 0.04	0.114 ± 0.002	0.67 ± 0.02	0.27 ± 0.03	1.5 ± 0.2
	✓			✓	0.076 ± 0.002	0.48 ± 0.03	0.111 ± 0.001	0.63 ± 0.02	0.25 ± 0.03	1.4 ± 0.1
	✓		✓	✓	0.083 ± 0.003	0.59 ± 0.02	0.115 ± 0.002	0.67 ± 0.01	0.28 ± 0.03	1.6 ± 0.1
US500	✓				0.066 ± 0.002	0.34 ± 0.02	0.034 ± 0.002	0.23 ± 0.02	0.20 ± 0.02	1.4 ± 0.1
	✓	✓			0.065 ± 0.002	0.33 ± 0.02	0.032 ± 0.003	0.23 ± 0.03	0.16 ± 0.01	1.2 ± 0.2
	✓			✓	0.066 ± 0.002	0.34 ± 0.02	0.034 ± 0.002	0.23 ± 0.02	0.18 ± 0.03	1.3 ± 0.2
	✓	✓		✓	0.069 ± 0.003	0.35 ± 0.02	0.035 ± 0.002	0.24 ± 0.03	0.24 ± 0.02	1.5 ± 0.2
HK800	✓				0.084 ± 0.001	0.75 ± 0.03	0.090 ± 0.002	0.77 ± 0.02	0.68 ± 0.04	2.2 ± 0.2
	✓	✓			0.090 ± 0.004	0.75 ± 0.03	0.095 ± 0.005	0.79 ± 0.05	0.70 ± 0.03	2.3 ± 0.2
	✓		✓		0.093 ± 0.002	0.78 ± 0.03	0.095 ± 0.003	0.82 ± 0.03	0.72 ± 0.03	2.4 ± 0.1
	✓	✓	✓		0.096 ± 0.003	0.80 ± 0.04	0.099 ± 0.002	0.85 ± 0.05	0.75 ± 0.04	2.5 ± 0.2

Table 9: Ablation study of cross-market dynamic systematic attention for MASTER.

Dataset	DSA	DCSA(CN)	DCSA(US)	DCSA(HK)	Metrics					
					IC	ICIR	RankIC	RankICIR	AR	SR
CN800	✓				0.074 ± 0.002	0.47 ± 0.04	0.097 ± 0.002	0.63 ± 0.02	0.27 ± 0.06	1.4 ± 0.3
	✓		✓		0.078 ± 0.003	0.48 ± 0.02	0.105 ± 0.003	0.65 ± 0.03	0.29 ± 0.03	1.7 ± 0.2
	✓			✓	0.075 ± 0.002	0.47 ± 0.02	0.103 ± 0.003	0.65 ± 0.03	0.29 ± 0.02	1.6 ± 0.1
	✓		✓	✓	0.080 ± 0.001	0.53 ± 0.02	0.109 ± 0.003	0.68 ± 0.02	0.31 ± 0.04	1.9 ± 0.2
US500	✓				0.082 ± 0.008	0.34 ± 0.03	0.037 ± 0.006	0.22 ± 0.03	0.21 ± 0.05	1.5 ± 0.2
	✓	✓			0.081 ± 0.005	0.33 ± 0.02	0.037 ± 0.003	0.21 ± 0.02	0.20 ± 0.05	1.5 ± 0.3
	✓			✓	0.083 ± 0.006	0.34 ± 0.02	0.037 ± 0.003	0.22 ± 0.02	0.22 ± 0.06	1.5 ± 0.3
	✓	✓		✓	0.087 ± 0.004	0.35 ± 0.01	0.041 ± 0.003	0.24 ± 0.02	0.25 ± 0.04	1.6 ± 0.1
HK800	✓				0.101 ± 0.003	0.98 ± 0.04	0.099 ± 0.003	0.98 ± 0.04	0.83 ± 0.06	2.8 ± 0.2
	✓	✓			0.104 ± 0.003	0.91 ± 0.04	0.096 ± 0.002	0.96 ± 0.03	0.83 ± 0.05	2.7 ± 0.2
	✓		✓		0.107 ± 0.002	1.00 ± 0.03	0.099 ± 0.003	0.98 ± 0.04	0.85 ± 0.06	2.8 ± 0.2
	✓	✓	✓		0.110 ± 0.002	1.05 ± 0.05	0.105 ± 0.003	1.01 ± 0.05	0.88 ± 0.07	3.1 ± 0.2

the home-only configuration contributed most of the total gain. For DTML, DCSA(CN) alone decreased IC relative to home-only and DCSA(HK) alone left it unchanged. For MASTER, both single foreign sources stayed within 1.2% of home-only in either direction. FinD3 was again the exception, gaining 7.0% from CN alone and 5.3% from HK alone, though from a baseline low enough that the absolute IC remained below both other hosts' home-only results. The contrast suggests that systematic structure within the U.S. market was more informative for U.S. stock returns than lagged signals from Asian markets, except where the host captured little of that domestic structure to begin with.

While interaction with both foreign markets showed consistent gains, adding a single source did not always improve every metric. For DTML on US500, adding CN alone slightly decreased IC relative to the home-only configuration and adding HK alone left it unchanged, yet the two-source configuration achieved the highest IC of all. CN as a single source into HK800 was informative but destabilizing for two of the three hosts: it raised IC while lowering both rank metrics for MASTER (0.099 to 0.096 RankIC) and for FinD3 (0.103 to 0.098 RankIC), indicating that a weakly informative source could add signal and variance simultaneously. FinD3 on US500 showed the same effect at the home-only step, where RankICIR fell marginally below the bare baseline (0.21 to 0.20) and recovered only once both foreign sources were present. The pattern in these cases was that a marginal source shifted performance between the level and the stability of the cross-sectional correlation. Still, the full two-source configuration was best or joint-best on every metric in all nine host–market blocks, indicating that the learned gates combined complementary information rather than selecting one market.

Cross-Market Pattern from Transmission Gate To understand how DCSA leverages information from foreign markets, we examined the learned transmission gates $\mathbf{\Gamma}_t^{(m,n)}$, which control the contribution of systematic information from source market n to target market m. Figures 2, 3, and 4 show the time-varying gate weights, averaged across the J latent factors, for each foreign source during the test period across the three target markets and host architectures.

For CN800 and HK800, the source-gate trajectories exhibited a clearer directional pattern. For CN800, DTML and FinD3 assigned larger weights to the U.S. source than to the HK source for most of the test period, whereas MASTER allocated more balanced and time-varying weights to the two sources. HK800 showed a similar pattern: DTML and FinD3 generally assigned larger weights to the U.S. source than to the CN source, while MASTER's two source gates remained relatively close and alternated over time. These patterns are broadly consistent with the single-

Table 10: Ablation study of cross-market dynamic systematic attention for FinD3.

Dataset	DSA	DCSA(CN)	DCSA(US)	DCSA(HK)	Metrics					
					IC	ICIR	RankIC	RankICIR	AR	SR
CN800	✓				0.086 ± 0.003	0.57 ± 0.03	0.111 ± 0.002	0.66 ± 0.02	0.28 ± 0.03	1.5 ± 0.2
			✓		0.090 ± 0.002	0.61 ± 0.02	0.112 ± 0.003	0.67 ± 0.03	0.29 ± 0.02	1.7 ± 0.1
				✓	0.094 ± 0.002	0.67 ± 0.03	0.113 ± 0.002	0.75 ± 0.02	0.32 ± 0.03	1.5 ± 0.1
			✓	✓	0.098 ± 0.001	0.75 ± 0.02	0.114 ± 0.003	0.81 ± 0.03	0.33 ± 0.02	2.1 ± 0.1
US500	✓				0.057 ± 0.006	0.28 ± 0.02	0.030 ± 0.004	0.20 ± 0.03	0.28 ± 0.04	1.2 ± 0.1
		✓			0.061 ± 0.005	0.30 ± 0.02	0.032 ± 0.003	0.21 ± 0.02	0.31 ± 0.03	1.2 ± 0.2
				✓	0.060 ± 0.006	0.28 ± 0.03	0.030 ± 0.004	0.20 ± 0.03	0.27 ± 0.04	1.1 ± 0.1
		✓		✓	0.068 ± 0.004	0.31 ± 0.02	0.033 ± 0.003	0.23 ± 0.02	0.33 ± 0.04	1.3 ± 0.2
HK800	✓				0.071 ± 0.002	0.66 ± 0.03	0.103 ± 0.003	0.88 ± 0.04	0.69 ± 0.02	2.5 ± 0.2
		✓			0.084 ± 0.003	0.69 ± 0.04	0.098 ± 0.002	0.84 ± 0.05	0.71 ± 0.05	2.6 ± 0.2
			✓		0.089 ± 0.002	0.72 ± 0.03	0.104 ± 0.003	0.87 ± 0.03	0.71 ± 0.03	2.6 ± 0.1
		✓	✓		0.091 ± 0.003	0.75 ± 0.04	0.106 ± 0.002	0.90 ± 0.03	0.79 ± 0.03	2.7 ± 0.2

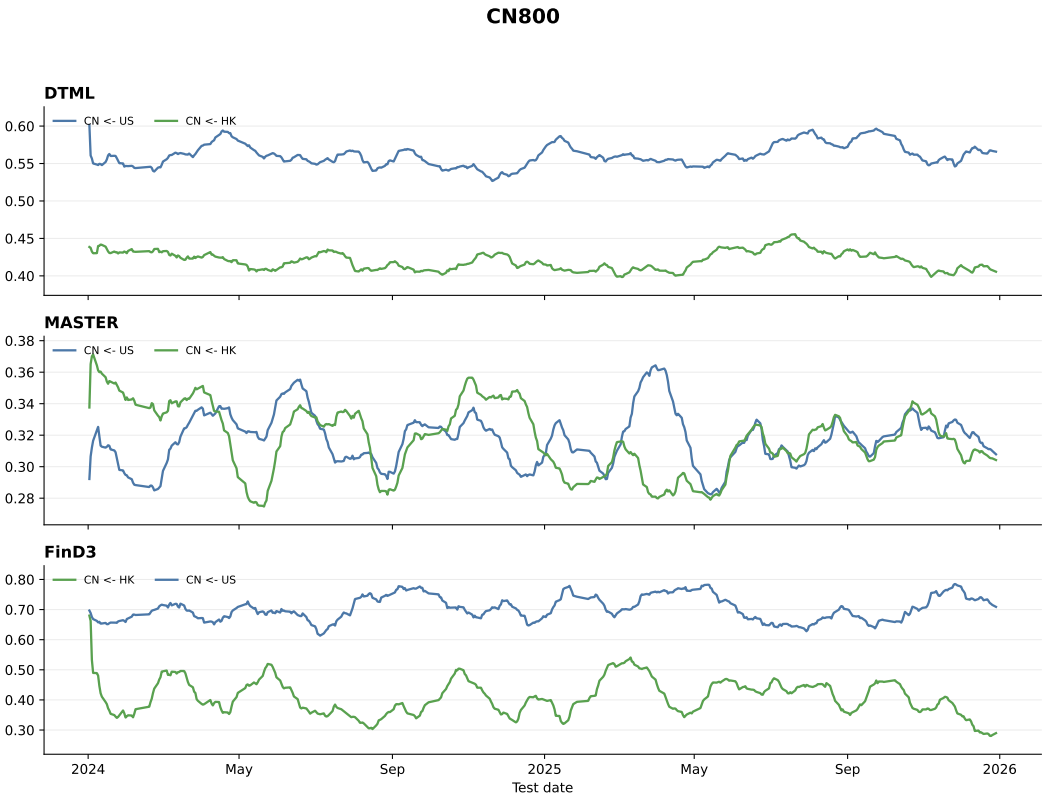

Figure 2: Transmission gate weight on CN800

source ablations: the U.S. source provided larger IC gains than the Asian counterpart for all hosts on HK800 and for DTML and MASTER on CN800. FinD3 on CN800 was an exception, obtaining a slightly larger single-source gain from HK than from the U.S. source despite assigning a higher gate weight to the latter. This difference suggests that the gate trajectories characterize the model's time-varying routing behavior rather than directly ranking the standalone predictive value of each source. Detailed transmission-gate trajectories are provided in Appendix Figures 2–4.

For US500, the transmission patterns from CN and HK differed across host architectures. For DTML and MASTER, the gate weights from CN and HK remained closely aligned and within a narrow, stable range throughout the test period, with variations of less than 0.02 across most time steps. This suggests that systematic factors from CN and HK provided limited incremental information for U.S. return prediction beyond that captured by the host models and the U.S. market's local systematic representation, consistent with the modest gains observed in the corresponding ablation results. In contrast, FinD3 exhibited substantially greater variation in its transmission gates. Although the gate weights from CN remained close to zero for extended periods, their overall variation was con-

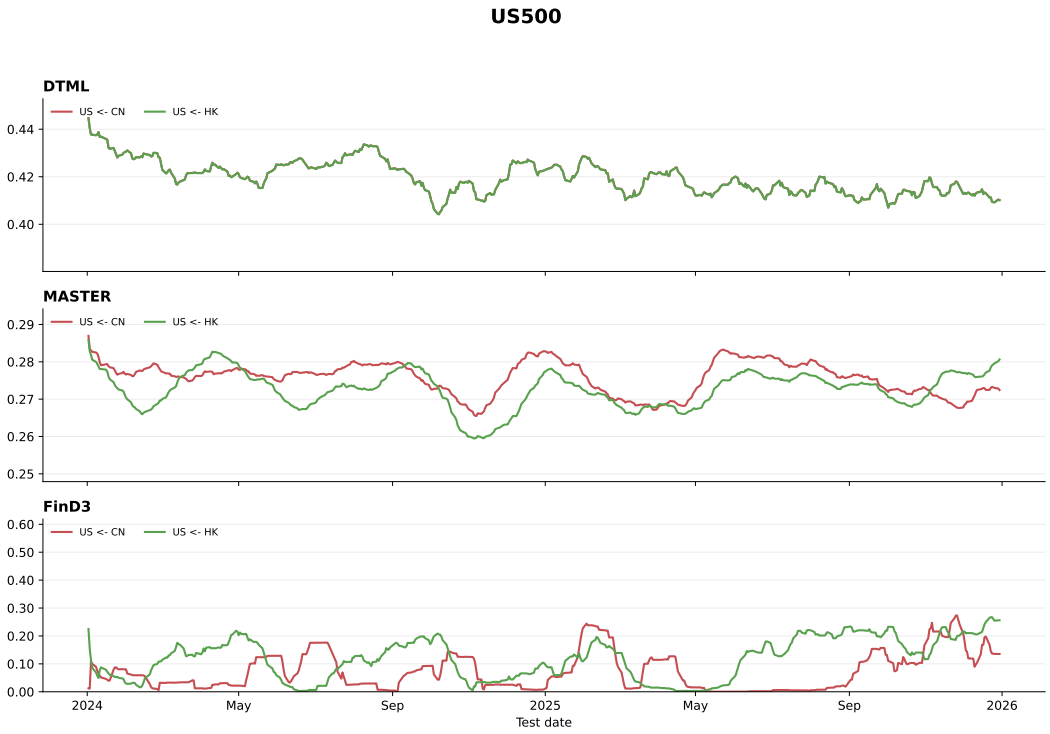

Figure 3: Transmission gate weight on US500.

siderably larger than those observed for DTML and MASTER, corresponding to a 19.3% relative improvement in IC when foreign sources were enabled. These results suggest that the contribution of cross-market systematic information depends on the representational capacity of the host architecture: more expressive host models may already capture much of the relevant cross-market information, whereas architectures with more limited baseline factor capacity can benefit more from the additional systematic signals provided by DCSA.

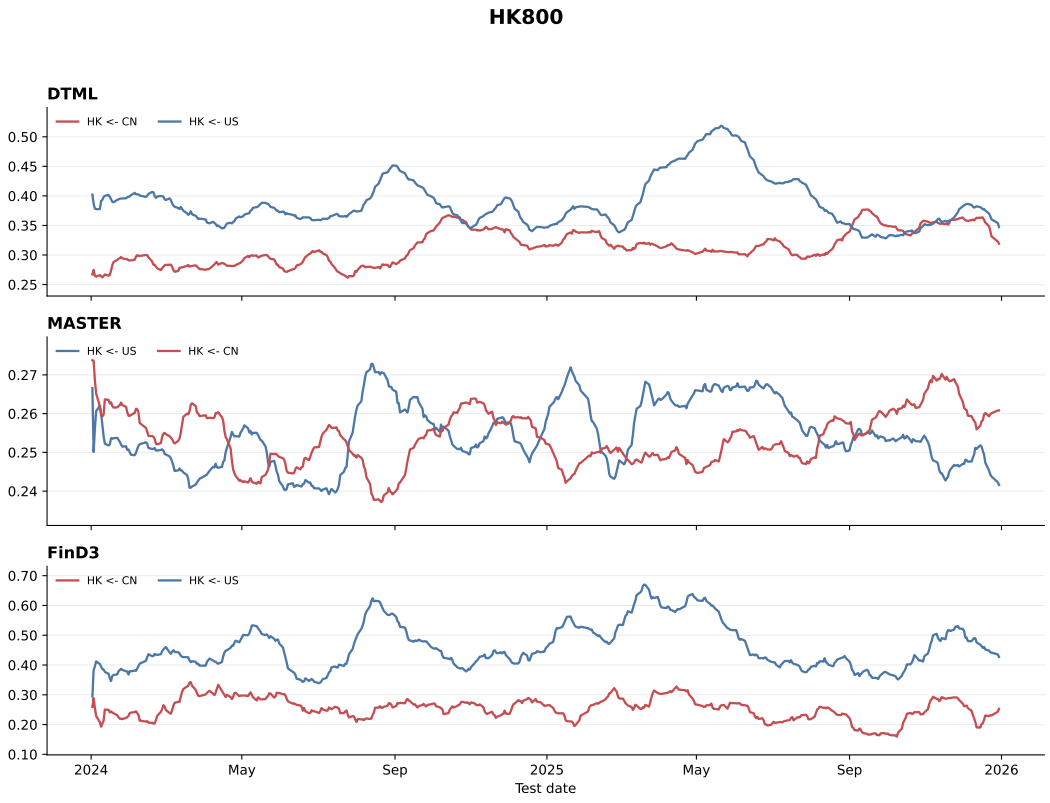

Figure 4: Transmission gate weight on HK800.