

# 基于Leddam-iTransformer-FCB的日频股价预测方法

余孟哲<sup>1</sup>, 徐铭申<sup>2\*</sup>, 程薄绵<sup>3</sup>

<sup>1</sup>华北电力大学(保定)经济管理系, 河北 保定

<sup>2</sup>华北电力大学(保定)数理系, 河北 保定

<sup>3</sup>成都理工大学管理科学学院, 四川 成都

收稿日期: 2026年4月1日; 录用日期: 2026年4月14日; 发布日期: 2026年5月29日

## 摘要

日频股价预测同时受到市场噪声、弱非平稳性以及价格变量与流动性相关变量耦合关系的共同影响, 因而仍是量化金融中的一项具有挑战性的研究问题。本文围绕归档的Leddam-iTransformer-FCB实现, 考察其在一步日频股价预测场景下的模型机理与实验表现。该模型首先对输入样本执行样本级归一化, 随后利用可学习分解模块提取各变量中的平滑主成分, 再通过倒置Transformer将变量轨迹视为token以建模跨变量依赖关系, 并在潜在空间中引入Fourier Convolution Block进行全局频谱混合, 最终完成预测输出。依据归档代码、checkpoint与测试预测序列, 可以重建出该模型工作于一个包含5个变量、共2085个日频观测、采用60%/20%/20%划分、回溯窗口长度为12个交易日、预测步长为1日的股价预测设置。实验结果表明, 完整模型在测试集上的性能为 $R^2 = 0.9653$ 、 $MSE = 1.5547$ 、 $RMSE = 1.2469$ 、 $MAE = 0.9294$ 、 $MAPE = 1.2655\%$ 。结果表明, 可学习平滑、跨变量注意力与潜在频谱混合在短期股价预测任务中具有明显的互补作用。

## 关键词

股价预测, 多变量时间序列, 可学习分解

# Daily Frequency Stock Price Prediction Method Based on Leddam-iTransformer-FCB

Mengzhe Yu<sup>1</sup>, Mingshen Xu<sup>2\*</sup>, Bomian Cheng<sup>3</sup>

<sup>1</sup>Department of Economic and Management, North China Electric Power University (Baoding), Baoding Hebei

<sup>2</sup>Department of Mathematics and Physics, North China Electric Power University (Baoding), Baoding Hebei

<sup>3</sup>College of Management Science, Chengdu University of Technology, Chengdu Sichuan

Received: April 1, 2026; accepted: April 14, 2026; published: May 29, 2026

\*通讯作者。

文章引用: 余孟哲, 徐铭申, 程薄绵. 基于 Leddam-iTransformer-FCB 的日频股价预测方法[J]. 金融, 2026, 16(3): 391-406. DOI: 10.12677/fin.2026.163039

## Abstract

Daily stock price forecasting remains a challenging research topic in quantitative finance due to the combined effects of market noise, weak non-stationarity, and coupling relationships between price variables and liquidity-related variables. This study investigates the model mechanism and experimental performance of the archived Leddam-iTransformer-FCB framework in one-step daily price prediction scenarios. The model first performs sample-level normalization on input data, then employs a learnable decomposition module to extract smooth principal components from variables. An inverted Transformer treats variable trajectories as tokens to model cross-variable dependencies, while Fourier Convolution Blocks in latent spaces facilitate global spectral mixing for prediction output. Using archived code, checkpoints, and test sequences, the model operates under the following parameters: 5 variables, 2085 daily observations, 60%/20%/20% data split, 12-day backtracking window, and 1-day prediction horizon. Experimental results demonstrate the model achieves  $R^2 = 0.9653$ ,  $MSE = 1.5547$ ,  $RMSE = 1.2469$ ,  $MAE = 0.9294$ , and  $MAPE = 1.2655\%$  on the test set. The results indicate that learnable smoothing, cross-variable attention, and latent spectrum mixing exhibit significant complementary effects in short-term stock price prediction tasks.

## Keywords

Stock Price Forecasting, Multivariate Time Series, Learnable Decomposition

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

股价可预测性始终是量化金融研究中的核心问题。有效市场假说给出了资产价格快速反映公开信息的经典基准[1]，而随后关于随机游走假设的经验研究表明，在有限样本和特定时间尺度下，股票价格并不必然严格满足随机游走特征[2]。在此背景下，金融计量学逐步发展出一整套用于刻画资产动态、波动特征、相关结构与预测分析的方法体系[3]。

随着机器学习方法的引入，研究者开始利用非线性模型处理股票市场中的复杂关系。例如，支持向量机已被证明能够在金融时间序列预测中取得有竞争力的效果[4]；随后的综述工作进一步系统梳理了股票预测中的软计算方法[5]。进入深度学习阶段后，金融序列预测越来越强调去噪、深层特征提取与非线性时序建模的协同作用，例如结合小波变换、堆叠自编码器与长短期记忆网络的股价预测框架[6]，以及在大样本股票预测任务中验证有效性的 LSTM 体系[7]。相关综述研究也指出，金融时间序列已成为深度学习在应用金融学中的重要研究方向之一[8]。

进一步地，针对金融时间序列的模型结构仍在持续演化。双阶段注意力递归网络能够同时筛选关键驱动变量与关键历史状态[9]；Temporal Fusion Transformer 则将可解释性与多步预测统一于注意力框架之中[10]。时序卷积网络以稳定的感受野提供了较强的序列建模能力[11]，N-BEATS 则通过深度基展开增强了外推能力与结构可解释性[12]。

在更广泛的序列建模领域，Transformer 的发展始于原始注意力架构[13]，并迅速扩展到时间序列预测中的 Informer [14]、Autoformer [15]与 FEDformer [16]。与此同时，线性基线研究重新讨论了 Transformer 在时间序列任务中的有效性边界[17]，PatchTST 则进一步通过 patch 级 token 化提升了预测性能[18]。针

对金融数据中常见的分布漂移与非平稳性问题, RevIN [19]与 Non-stationary Transformer [20]又分别从归一化机制与注意力建模角度给出了补充。

与多变量股价预测最直接相关的若干工作包括: 将变量视为 token 的 iTransformer [21]、能够动态提取趋势信息的可学习分解方法 Leddam [22]、基于频域滤波的 FilterTS [23], 以及通过全局频谱卷积扩大有效感受野的 Fourier Convolution Block [24]。这些研究表明, 金融时间序列预测不宜仅依赖单一的时间建模机制, 而应在去噪、跨变量交互与频域表征之间建立更紧密的联系。

基于此, 本文研究一种用于日频股价预测的深度学习方法。该方法融合 Leddam、iTransformer 与 FCB 的建模思想, 从历史价格及相关时序特征中同时挖掘局部波动模式、跨变量依赖关系与潜在频域结构, 以提升下一交易日股价预测的准确性与稳定性。实验结果表明, 该模型能够较好刻画日频金融时间序列的复杂动态特征, 并为后续量化投资与风险管理中的预测建模提供参考。

## 2. 任务定义与模型推导

### 2.1. 日频股价预测任务

本文考虑多变量日频金融时间序列预测问题。定义如下变量如式(1):

$$\mathbf{x}_t \in \mathbb{R}^C \quad (1)$$

其中, 上述定义表示第  $t$  个交易日的市场观测向量, 其中  $C$  为变量维数。本文对应的归档实现中取  $C = 5$ , 即每个交易日由 5 个市场相关变量共同表征。向量  $\mathbf{x}_t$  的第  $c$  个分量记为  $x_{t,c}$ , 表示第  $c$  个变量在时刻  $t$  的观测值。该表示允许模型在统一框架下同时处理价格类、波动类与成交量类等多种异质信息。

给定长度为  $L$  的历史观测窗口, 定义在时刻  $t$  的输入样本为

$$\mathbf{X}_t = [\mathbf{x}_{t-L+1}, \mathbf{x}_{t-L+2}, \dots, \mathbf{x}_t]^\top \in \mathbb{R}^{L \times C}. \quad (2)$$

式(2)中,  $L$  表示回溯窗口长度;  $\mathbf{X}_t$  的第  $i$  行对应窗口中第  $i$  个历史交易日的多变量观测, 第  $c$  列对应第  $c$  个市场变量在整个窗口内的历史轨迹。因此,  $\mathbf{X}_t$  同时编码了时间维上的演化信息与变量维上的协同变化信息, 是后续预测模型的基本输入单元。

模型的目标是基于历史窗口  $\mathbf{X}_t$  预测未来  $H$  个交易日的多变量输出序列, 即

$$\mathbf{Y}_t = [\mathbf{y}_{t+1}, \mathbf{y}_{t+2}, \dots, \mathbf{y}_{t+H}]^\top \in \mathbb{R}^{H \times C}. \quad (3)$$

式(3)中,  $H$  为预测步长;  $\mathbf{y}_{t+\tau} \in \mathbb{R}^C$  表示未来第  $\tau$  个交易日的目标变量向量,  $\tau = 1, \dots, H$ 。在多变量预测设定下, 模型并非仅预测单一收盘价, 而是同时对多个金融变量进行联合建模, 从而有利于保留变量间的动态耦合结构。

在本文实验设置中, 采用一步预测, 即

$$H = 1. \quad (4)$$

式(4)中,  $H = 1$  表示模型每次仅预测下一个交易日的输出, 然后在测试阶段通过滚动方式逐步向前推进。该设置对应典型的日频短期预测场景, 其优点在于每次预测所依赖的历史信息最为新近, 能够更好地适应金融市场快速变化的局部统计特征。

### 2.2. 样本级归一化

由于不同金融变量在量纲、波动范围及局部分布上往往存在显著差异, 模型在前向传播开始前先对每个输入样本执行样本级归一化。对于窗口内第  $c$  个变量, 其样本均值与样本标准差分别定义为

$$\mu_c = \frac{1}{L} \sum_{i=1}^L x_{i,c}, \sigma_c = \sqrt{\frac{1}{L} \sum_{i=1}^L (x_{i,c} - \mu_c)^2 + \epsilon}. \quad (5)$$

式(5)中,  $x_{i,c}$  表示当前样本窗口内第  $i$  个时间位置、第  $c$  个变量的原始观测值;  $\mu_c$  为该变量在当前长度为  $L$  的窗口内的局部均值, 用于刻画该变量近期的平均水平;  $\sigma_c$  为对应的局部标准差, 用于度量该变量在窗口内的波动幅度;  $\epsilon > 0$  为数值稳定项, 用于避免方差极小或为零时产生除零问题。需要强调的是, 此处的归一化统计量是按样本、按变量独立计算的, 而不是在整个训练集上预先固定, 因此能够更灵活地适应金融时间序列常见的非平稳性与局部分布漂移。

在得到均值和标准差后, 对窗口内各观测执行标准化变换:

$$\tilde{x}_{i,c} = \frac{x_{i,c} - \mu_c}{\sigma_c}. \quad (6)$$

式(6)中,  $x_{i,c}$  表示归一化后的变量值, 其均值被平移至近似零附近, 尺度被缩放至近似单位方差水平。经过该步骤后, 不同变量在数值范围上的差异被显著削弱, 从而使后续卷积、注意力与频域变换模块更关注变量间的动态结构而非绝对量纲差异。对于金融数据而言, 这种样本级归一化尤其有助于缓解价格序列趋势漂移、成交量数量级变化以及市场状态切换所带来的训练不稳定问题。

### 2.3. 可学习分解模块

为降低高频噪声对后续特征提取的干扰, 模型首先引入可学习分解模块, 对每个变量的归一化历史轨迹进行平滑与残差分解。设一维卷积核长度为  $k$ , 其中心位置记为

$$q = \left\lfloor \frac{k}{2} \right\rfloor. \quad (7)$$

式(7)中,  $k$  为平滑卷积核的长度,  $q$  为卷积核中心索引。当  $k$  为奇数时,  $q$  对应对称核的几何中心; 当采用填充操作时, 输出时间长度可与输入保持一致。

在参数初始化阶段, 卷积核首先被赋予一个 Gaussian 形状的先验:

$$a_j = \exp \left[ - \left( \frac{j-q}{2\sigma_g} \right)^2 \right], \quad j = 0, 1, \dots, k-1. \quad (8)$$

式(8)中,  $a_j$  表示卷积核第  $j$  个位置的初始化响应强度;  $\sigma_g$  为控制 Gaussian 展宽程度的超参数;  $j-q$  表示第  $j$  个位置相对于中心位置的偏移量。该初始化方式使得核权重在初始阶段更偏向中心邻域, 从而具有平滑滤波器的先验结构, 即更强调局部邻近时刻信息, 而弱化远离中心观测值的贡献。随后, 通过 softmax 将该初始响应归一化为非负且和为 1 的平滑权重:

$$w_j = \frac{\exp(a_j)}{\sum_{m=0}^{k-1} \exp(a_m)} \quad (9)$$

式(9)中,  $w_j$  表示卷积核第  $j$  个位置的实际归一化权重, 满足  $w_j \geq 0$  且  $\sum_{j=0}^{k-1} w_j = 1$ 。这种约束使卷积核在初始阶段具有加权移动平均的性质, 能够提供稳定的低通平滑效果, 同时由于  $w_j$  仍可在训练中更新, 因此该模块并非固定滤波器, 而是一个可学习的平滑分解器。

对归一化后的第  $c$  个变量序列, 可定义其在位置  $i$  处的平滑主成分为:

$$m_{i,c} = \sum_{j=0}^{k-1} w_j \tilde{x}_{i+j-q,c}. \quad (10)$$

式(10)中,  $m_{i,c}$  表示第  $c$  个变量在时间位置  $i$  处的平滑成分;  $\tilde{x}_{i+j-q,c}$  为以位置  $i$  为中心、窗口跨度为  $k$  的局部邻域观测;  $w_j$  为对应位置的平滑权重。边界位置采用 **replicate padding**, 即当索引超出原时间范围时, 重复使用边界处最近的观测值进行填充。该机制保证了卷积在边缘位置同样可以计算, 且不会引入零填充所造成的虚假突变。

在得到平滑主成分后, 将原始归一化序列与其差值定义为残差分量:

$$\rho_{i,c} = \tilde{x}_{i,c} - m_{i,c}. \quad (11)$$

式(11)中,  $\rho_{i,c}$  表示第  $c$  个变量在位置  $i$  的高频残差信号, 反映了未被平滑趋势解释的局部扰动、尖峰或短时异常波动。于是, 归一化后的输入序列被分解为

$$\tilde{x}_{i,c} = m_{i,c} + \rho_{i,c}. \quad (12)$$

该分解将原始序列拆分为平滑主成分与高频残差两部分, 其中前者更适合后续结构建模, 后者则更多承载短时噪声与局部不规则波动。

将所有位置与变量堆叠后, 可分别记平滑主成分矩阵和残差矩阵为

$$\mathbf{M} = [m_{i,c}] \in \mathbb{R}^{L \times C}, \mathbf{R} = [\rho_{i,c}] \in \mathbb{R}^{L \times C} \quad (13)$$

式中,  $\mathbf{M}$  表示经可学习平滑后得到的主成分序列矩阵,  $\mathbf{R}$  表示相应残差矩阵。根据归档实现, 后续编码器仅接收  $\mathbf{M}$  作为输入, 而不显式使用  $\mathbf{R}$ 。这意味着模型在进入深层特征提取前, 已经通过一个可训练的低通滤波步骤抑制了部分高频噪声, 从而使注意力机制与频域模块更集中于对中低频结构、局部趋势和变量间耦合关系的学习。在短期股价预测任务中, 这一步尤其重要, 因为金融高频扰动往往具有较强随机性, 直接输入可能削弱主导信号的可辨识度。

## 2.4. 倒置变量 Token 嵌入

传统时间序列 Transformer 往往将时间点视为 token, 即每个 token 对应一个时刻的多变量观测。而 iTransformer 的关键思想在于进行“倒置”建模, 将变量轨迹视为 token, 即每个 token 对应一个变量在历史窗口内的完整时间演化路径。对于第  $c$  个变量, 定义其平滑后的历史轨迹向量为

$$\mathbf{m}_c = [m_{1,c}, m_{2,c}, \dots, m_{L,c}]^T \in \mathbb{R}^L. \quad (14)$$

式(14)中,  $m_c$  是变量  $c$  在长度为  $L$  的历史窗口上的一维时间轨迹, 它保留了该变量在整个输入区间内的动态模式。与传统时序 token 设计相比, 这种表示方式直接把变量作为建模对象, 使后续自注意力机制能够围绕变量之间的关系而非时刻之间的关系展开。

随后, 通过线性嵌入层将长度为  $L$  的变量轨迹映射到  $d$  维潜在空间:

$$\mathbf{z}_c^{(0)} = \mathbf{W}_e \mathbf{m}_c + \mathbf{b}_e, \mathbf{W}_e \in \mathbb{R}^{d \times L}, \mathbf{b}_e \in \mathbb{R}^d \quad (15)$$

式(15)中,  $\mathbf{z}_c^{(0)} \in \mathbb{R}^d$  表示变量  $c$  的初始 token 表示;  $\mathbf{W}_e$  为线性投影矩阵, 将原始长度为  $L$  的时间轨迹映射到  $d$  维表征空间;  $\mathbf{b}_e$  为偏置项。该操作本质上相当于从变量的整个历史轨迹中提取一个固定长度的潜在表示, 使模型得以在紧凑的特征空间中建模变量关系。

将所有变量的嵌入向量按行堆叠后, 得到编码器输入 token 矩阵:

$$\mathbf{Z}^{(0)} = [\mathbf{z}_1^{(0)}, \mathbf{z}_2^{(0)}, \dots, \mathbf{z}_C^{(0)}]^T \in \mathbb{R}^{C \times d}. \quad (16)$$

式(16)中,  $\mathbf{Z}^{(0)}$  的第  $c$  行对应第  $c$  个变量的初始 token 表示; 矩阵共有  $C$  个 token, 每个 token 的维数为  $d$ 。根据优化后的 checkpoint, 本文恢复得到  $L=12$ 、 $C=5$ 、 $d=64$ 。这意味着模型并不是在长度为 12

的时间维上直接计算注意力，而是在 5 个变量 token 上建模变量间依赖关系。由于实际金融建模中变量数目通常远小于时间长度，这种设计可显著降低注意力复杂度，同时强化对跨变量结构的表达能力。

## 2.5. 跨变量自注意力编码器

设第  $\ell$  层编码器的输入为

$$\mathbf{Z}(\ell) \in \mathbb{R}^{C \times d}, \quad (17)$$

式(17)中， $z^{(\ell)}$  表示第  $\ell$  层的变量 token 表示矩阵，其中  $C$  为 token 数(即变量数)， $d$  为每个 token 的隐空间维度。编码器的目标是在这些变量 token 之间建立信息交互，从而提取变量间的动态依赖结构。

对于第  $\ell$  层中的第  $h$  个注意力头，分别构造 query、key 和 value:

$$\mathbf{Q}_h^\ell = \mathbf{Z}^\ell \mathbf{W}_{Q,h}^\ell, \mathbf{K}_h^\ell = \mathbf{Z}^\ell \mathbf{W}_{K,h}^\ell, \mathbf{V}_h^\ell = \mathbf{Z}^\ell \mathbf{W}_{V,h}^\ell \quad (18)$$

式(18)中， $\mathbf{W}_{Q,h}^\ell, \mathbf{W}_{K,h}^\ell$  和  $\mathbf{W}_{V,h}^\ell$  分别为第  $\ell$  层第  $h$  个头的 query、key、value 线性投影矩阵； $\mathbf{Q}_h^\ell, \mathbf{K}_h^\ell, \mathbf{V}_h^\ell$  则为对应投影后的表示。由于输入 token 以变量为单位，因此 query 与 key 的相似性衡量的不是不同时间位置之间的相关性，而是不同市场变量在当前样本中的潜在关联强度。

基于缩放点积机制，注意力权重矩阵写为

$$\mathbf{A}_h^\ell = \text{softmax} \left( \frac{\mathbf{Q}_h^\ell (\mathbf{K}_h^\ell)^\top}{\sqrt{d_h}} \right), d_h = \frac{d}{n_h}, \quad (19)$$

式(19)中， $\mathbf{A}_h^\ell \in \mathbb{R}^{C \times C}$  表示第  $h$  个头的变量间注意力矩阵； $nh$  为多头注意力的头数； $d_h$  为单头维度； $\sqrt{d_h}$  为缩放因子，用于防止内积值随维度增长而过大，进而避免 softmax 进入梯度过小区域。矩阵  $\mathbf{A}_h^\ell$  的第  $(c, c')$  个元素表示变量  $c$  在更新自身表示时，对变量  $c'$  信息的关注程度。由于 softmax 在行方向归一化，每个变量都会对所有变量分配一组归一化权重，从而实现跨变量的信息聚合。

基于注意力权重，对 value 进行加权求和，得到第  $h$  个头的输出:

$$\mathbf{O}_h^\ell = \mathbf{A}_h^\ell \mathbf{V}_h^\ell \quad (20)$$

式(20)中， $\mathbf{O}_h^\ell \in \mathbb{R}^{C \times d_h}$  为第  $h$  个头输出的变量表示。该结果意味着每个变量 token 的新表示由所有变量 token 的 value 表示经加权融合而成，因此模型可以根据当前样本中变量间的相关结构，自适应地强调重要变量、抑制次要变量。

将所有头的输出在特征维上拼接，并经过输出投影与残差连接，可得注意力子层输出:

$$\tilde{\mathbf{Z}}^{(\ell)} = \mathbf{Z}^{(\ell)} + \text{Concat}(\mathbf{O}_1^{(\ell)}, \mathbf{O}_2^{(\ell)}, \dots, \mathbf{O}_{n_h}^{(\ell)}) \mathbf{W}_O^{(\ell)}. \quad (21)$$

式(21)中，Concat( $\cdot$ )表示沿特征维进行拼接； $\mathbf{W}_O^{(\ell)}$  为多头输出的线性融合矩阵； $\tilde{\mathbf{Z}}^{(\ell)}$  为注意力子层后的中间表示。残差连接将原始输入  $\mathbf{Z}^{(\ell)}$  与注意力输出相加，有助于稳定深层训练并保留原始变量表示中的基础信息。

随后，使用前馈网络进一步增强逐 token 的非线性表征能力:

$$\mathbf{Z}^{\ell+1} = \text{LN} \left( \tilde{\mathbf{Z}}^\ell + \mathbf{W}_2^\ell \phi \left( \mathbf{W}_1^\ell \text{LN}(\tilde{\mathbf{Z}}^\ell) \right) \right) \quad (22)$$

式(22)中，LN( $\cdot$ )表示 Layer Normalization； $\mathbf{W}_1^\ell$  和  $\mathbf{W}_2^\ell$  分别为前馈网络的两层线性变换； $\phi(\cdot)$  为非线性激活函数，在归档代码中对应 GELU； $\mathbf{Z}^{(\ell+1)}$  则为第  $\ell+1$  层编码器输出。前馈网络以逐 token 的方式作用于每个变量表示，在跨变量信息交互之后进一步提炼单个变量的高阶潜在特征。

上述编码器的核心优势在于：它并不直接在时间维度上建模长程依赖，而是将注意力集中于变量之

间的依赖结构。对于股价预测任务，这种建模方式与金融机理具有较强一致性，因为价格走势往往不是由单一变量独立决定，而是由开盘价、最高价、最低价、收盘价、成交量等多种指标的联动共同驱动。通过跨变量自注意力机制，模型能够显式学习这些变量在不同市场状态下的动态耦合强度。

## 2.6. 傅里叶卷积块

经过多层跨变量编码器后，模型得到潜在表示矩阵

$$\mathbf{E} = [e_{c,n}]_{c=1,\dots,C;n=0,\dots,d-1} \in \mathbb{R}^{C \times d}. \quad (23)$$

式(23)中， $e_{c,n}$ 表示第  $c$  个变量 token 在第  $n$  个潜在维度上的编码值； $\mathbf{E}$  可视为变量维与潜在特征维共同构成的二维表示。为进一步增强对全局结构和频率模式的建模能力，模型在潜在维度上引入傅里叶卷积块(Fourier Convolution Block, FCB)。

首先，沿潜在维度对每个变量 token 执行实数快速傅里叶变换：

$$\hat{e}_{c,f} = \sum_{n=0}^{d-1} e_{c,n} \exp\left(-j \frac{2\pi fn}{d}\right), f = 0, 1, \dots, \left\lfloor \frac{d}{2} \right\rfloor. \quad (24)$$

式(24)中， $\hat{e}_{c,f}$ 为变量  $c$  在频率索引  $f$  处的复数频域系数； $j$  为虚数单位； $d$  为潜在表示维数； $f$  为离散频率索引。由于输入为实数序列，只需保留非冗余的一半频谱分量。该变换将原本在潜在维度上分布的表征重写为频域表示，使模型能够显式区分不同频率成分贡献。

在频域中，FCB 为不同输入变量与不同输出变量之间设置可学习的复值混合权重：

$$g_{o,c,f} = g_{o,c,f}^{(R)} + jg_{o,c,f}^{(I)} \quad (25)$$

式(25)中， $g_{o,c,f}$ 表示在频率  $f$  处，从输入变量  $c$  到输出变量  $o$  的复值变换权重； $g_{o,c,f}^{(R)}$  和  $g_{o,c,f}^{(I)}$  分别为其实部与虚部参数。复值参数允许模型同时调整频率分量的幅值与相位，因此比纯实值加权更适合表达复杂的频域耦合结构。

基于这些频域权重，对各变量频谱进行线性混合，得到输出频域表示：

$$\hat{u}_{o,f} = \sum_{c=1}^C g_{o,c,f} \hat{e}_{c,f}. \quad (26)$$

式(26)中， $\hat{u}_{o,f}$ 表示输出变量  $o$  在频率  $f$  处的频域表示；求和过程表示对所有输入变量在该频率上的频谱进行加权融合。与时域卷积不同，这种频域混合能够以较低代价实现全局感受野建模，因为频域中的逐频率乘法对应于时域中的全局卷积效果。

最后，通过逆傅里叶变换恢复到时域(更准确地说，恢复到潜在特征域)：

$$u_{o,n} = \mathcal{F}^{-1}(\hat{u}_{o,f}). \quad (27)$$

式(27)中， $u_{o,n}$ 表示输出变量  $o$  在第  $n$  个潜在维度上的恢复值； $\mathcal{F}^{-1}(\cdot)$ 表示逆傅里叶变换。将所有输出堆叠后，可得到 FCB 的输出矩阵  $\mathbf{U} \in \mathbb{R}^{C \times d}$ 。从建模角度看，FCB 在潜在空间中补充了频率选择性和全局感受野，能够更有效地捕捉缓变结构、周期性模式以及不同变量之间在特定频带上的协同振荡。对于短期股价预测而言，这一机制有助于从局部噪声之上提取更稳定的低频共振信号或隐含市场节律。

## 2.7. 输出层与反归一化

经过编码器与傅里叶卷积块处理后，得到每个变量对应的潜在表示向量。记每个变量对应的最终表示为

$$\mathbf{u}_c \in \mathbb{R}^d. \quad (28)$$

式(28)中,  $\mathbf{u}_c$  是变量  $c$  的最终高维语义表示, 综合编码了该变量自身的历史动态、与其他变量的交互关系以及频域混合后的全局信息。

随后, 利用线性预测头将潜在表示映射为未来  $H$  个预测步的标准化输出:

$$\tilde{y}_{\tau,c} = \mathbf{w}_{p,\tau}^\top \mathbf{u}_c + b_{p,\tau}, \quad \tau = 1, \dots, H. \quad (29)$$

式(29)中,  $\tilde{y}_{\tau,c}$  表示变量  $c$  在未来第  $\tau$  个时间步的归一化尺度下预测值;  $\mathbf{W}_{p,\tau} \in \mathbb{R}^d$  为第  $\tau$  个预测步对应的线性投影权重;  $b_{p,\tau}$  为对应偏置项。对于一步预测任务, 实际仅需使用  $\tau = 1$  的投影头; 但写成一般形式有助于保持模型定义的完整性, 并方便扩展到多步预测场景。

由于模型内部是在样本归一化后的尺度上进行建模, 因此在输出阶段需要将预测值映射回原始量纲:

$$\hat{y}_{\tau,c} = \sigma_c \tilde{y}_{\tau,c} + \mu_c \quad (30)$$

式(30)中,  $\hat{y}_{\tau,c}$  为最终预测的原始尺度输出;  $\mu_c$  与  $\sigma_c$  分别是前述样本级归一化阶段对应变量  $c$  的局部均值与局部标准差。该反归一化操作确保模型最终输出与真实市场变量处于同一数值空间, 从而可以直接用于损失计算和预测评价。由于归一化与反归一化是成对出现的, 模型实质上学习的是相对于当前局部统计状态的偏移与变化模式, 而不是绝对数值本身, 这对非平稳金融序列尤其有效。

## 2.8. 损失函数与训练目标

训练阶段, 采用多变量均方误差(Mean Squared Error, MSE)作为优化目标:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{BHC} \sum_{b=1}^B \sum_{\tau=1}^H \sum_{c=1}^C (\hat{y}_{b,\tau,c} - y_{b,\tau,c})^2. \quad (31)$$

式(31)中,  $B$  为 batch size;  $H$  为预测步长;  $C$  为变量个数;  $\hat{y}_{b,\tau,c}$  表示第  $b$  个样本在未来第  $\tau$  个时间步、变量  $c$  上的预测值;  $y_{b,\tau,c}$  为对应真实值。该损失函数对 batch 内所有样本、所有预测步和所有变量的平方误差进行平均, 从而使模型在训练过程中同时兼顾多变量联合预测精度。

MSE 的优化目标等价于鼓励预测值在欧氏意义下尽可能逼近真实值, 具有形式简单、梯度稳定、与连续值预测高度匹配等优点。对于本文任务, 由于模型输出的是连续型金融变量, MSE 是最常用且最自然的监督学习目标之一。同时, 多变量联合 MSE 能够迫使模型在统一参数空间内协调不同变量的预测误差, 有利于形成更加一致的共享表示。

## 2.9. 复杂度分析与方法总结

从计算复杂度角度看, 各主要模块的时间复杂度可概括如下。可学习分解模块对每个变量独立执行长度为  $k$  的一维卷积, 其复杂度为

$$O(CLk). \quad (32)$$

式(32)中,  $C$  为变量数,  $L$  为时间窗口长度,  $k$  为平滑卷积核长度。该部分复杂度随窗口长度线性增长, 计算开销较低。

倒置变量注意力模块在变量维上执行自注意力, 其复杂度为

$$O(C^2d). \quad (33)$$

式(33)中,  $C^2$  来自所有变量 token 两两之间的相似度计算,  $d$  为隐空间维数。由于通常有  $C \ll L$ , 该复杂度相较于传统时间维注意力的  $O(L^2d)$  更具优势。

傅里叶卷积块主要由频域变换与跨变量频谱混合组成, 其复杂度可写为

$$O(Cd \log d + C^2 d_f). \quad (34)$$

式(34)中,  $Cd \log d$  对应沿潜在维度执行 FFT 与 IFFT 的复杂度;  $d_f$  表示保留或参与计算的频域分量数;  $C^2 d_f$  则来自不同变量在各频率点上的混合计算。该模块能够以较低代价引入全局卷积感受野和频带选择性。所提方法流程图如图 1 所示:

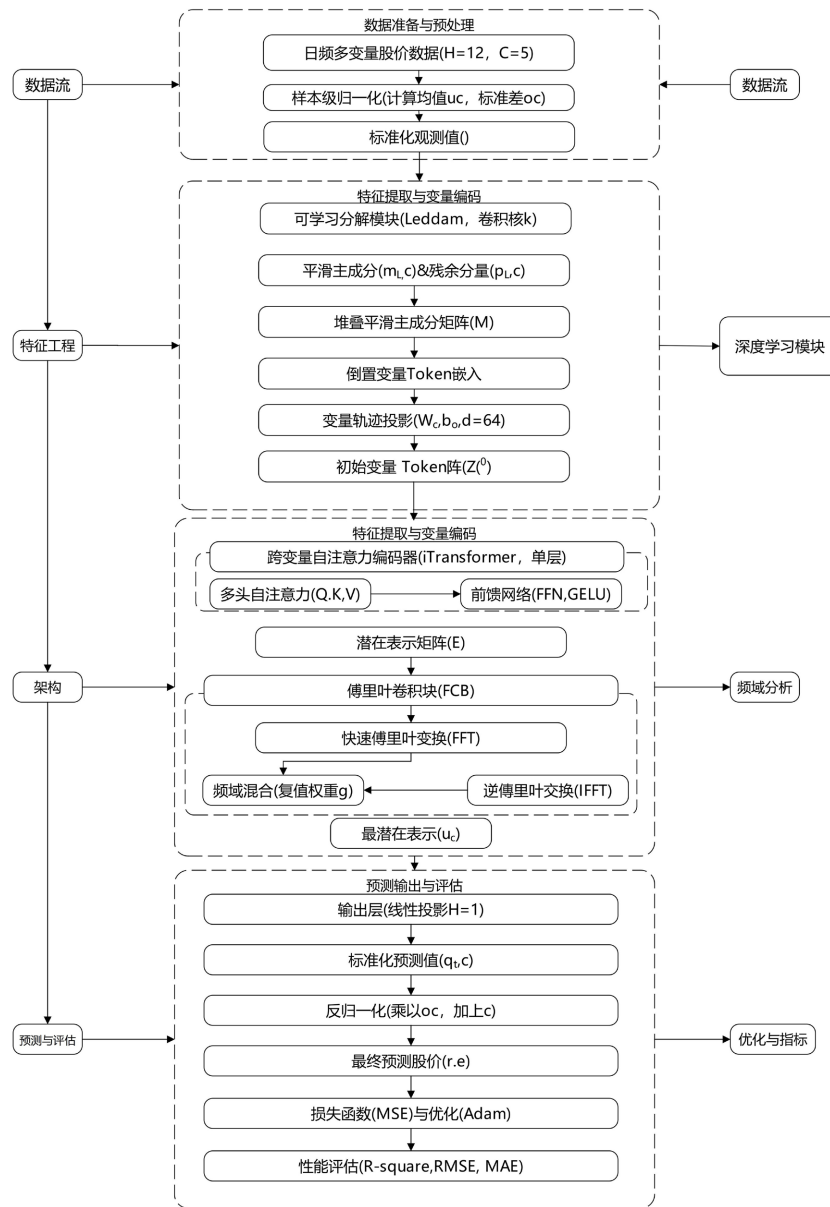


Figure 1. Overall flowchart of the proposed algorithm

图 1. 所提算法总体流程图

综合而言, 本文模型可以被理解为一个由“样本级局部标准化 - 可学习平滑分解 - 变量轨迹嵌入 - 跨变量自注意力建模 - 频域全局混合 - 线性预测与反归一化”串联而成的多阶段预测框架。其核心思想在于: 首先通过可学习分解抑制高频噪声, 再通过倒置变量 token 机制突出变量间依赖关系, 随后借助频域卷积补充全局结构建模能力, 最终在原始量纲空间内完成多变量预测。对于日频股价预测任

务，这种设计能够同时兼顾局部平滑性、跨变量协同性与潜在周期结构，从而提升模型对复杂金融动态的表达力。

### 3. 实验设置

#### 3.1. 数据集说明

本文选取 Best Buy Co., Inc. (NYSE: BBY) 日频交易数据作为研究对象，数据字段包括开盘价(Open)、最高价(High)、最低价(Low)、收盘价(Close)和成交量(Volume)共 5 个变量，以收盘价作为主要预测目标。结合本文实验设定，取 2016-12-13 至 2025-03-31 的连续日频交易数据，剔除非交易日后共得到 2085 个有效观测。数据按照时间顺序划分为训练集、验证集和测试集，比例为 60%/20%/20%。在建模过程中，历史回溯窗口长度设为 12 个交易日，预测步长设为 1 个交易日，即利用前 12 个交易日的多变量信息预测下一交易日收盘价。该数据集具有美元计价、单股票、日频、多变量 OHLCV 结构清晰、公开可复现等特点，适合用于一步日频股价预测任务。根据任务设定，实验针对日频多变量股价序列开展一步预测，输入为长度为 12 的历史回溯窗口，具体的实验设置如表 1 所示。

**Table 1.** Experimental setup for daily frequency stock price prediction

**表 1.** 日频股价预测实验设置

项目	取值
数据频率	日频交易数据
总观测数	2085
输入变量数	5
预测目标	下一交易日目标股价通道
回溯窗口 $L$	12 个交易日
预测步长 $H$	1 个交易日
训练/验证/测试划分	60%/20%/20%
训练/验证/测试窗口数	1239/405/405
测试样本数	405

从表 1 可以看出，实验设计属于典型的短窗口、一步滚动式金融时间序列预测设置。相较于多步外推任务，此类设置更侧重于刻画局部市场状态变化及相邻交易日之间的短期传递关系。

#### 3.2. 模型参数与训练设置

主要结构参数与训练设置见表 2 所示。其中，模型采用单层编码器结构，隐藏维度为 64，前馈层隐藏维度为 32；训练阶段使用 Adam 优化器，并配合 ReduceLRonPlateau 学习率调度策略与早停机制，以控制参数更新过程中的震荡并降低过拟合风险。

**Table 2.** Key model parameters and training settings

**表 2.** 主要模型参数与训练设置

参数	取值
输入维度 $C$	5
隐藏维度 $d$	64

续表

编码器层数	1
前馈层隐藏维度	32
Leddam 卷积核大小	25
预测步长	1 日
Batch size	16
最大训练轮数	50
优化器	Adam
学习率调度器	ReduceLROnPlateau
调度器衰减系数	0.1
早停耐心值	10 个 epoch
参数总量	23,789
超参数搜索	50 次 Optuna trial

总体而言，该模型规模相对紧凑，参数量控制在  $2.4 \times 10^4$  数量级。对于长度较短、变量数有限的日频股价预测任务而言，此类设置能够在表达能力与训练稳定性之间取得较为平衡的折中。

### 3.3. 评价指标

为评价模型对目标股价序列的拟合与预测能力，本文采用均方误差(MSE)、均方根误差(RMSE)、平均绝对误差(MAE)、平均绝对百分比误差(MAPE)以及决定系数  $R^2$  作为回归评价指标，其定义如下：

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2, \text{RMSE} = \sqrt{\text{MSE}}, \quad (35)$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i|, \text{MAPE} = \frac{100}{N} \sum_{i=1}^N \left| \frac{\hat{y}_i - y_i}{y_i} \right|, \quad (36)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}. \quad (37)$$

其中， $N$  表示测试样本数， $\hat{y}_i$  与  $y_i$  分别表示第  $i$  个样本的预测值与真实值， $\bar{y}$  为测试集真实值的样本均值。MSE 与 RMSE 主要刻画预测误差的整体幅度，MAE 反映平均绝对偏差，MAPE 用于衡量相对误差水平，而  $R^2$  则用于评估模型对目标序列方差的解释能力。上述指标从绝对误差、相对误差与拟合优度三个方面构成了对预测性能的综合评价。

### 3.4. 对比及消融实验

为增强实验结论的可比性，本文选取 ARIMA、SVR、Random Forest、XGBoost、LSTM、GRU、TCN、Informer、PatchTST 与 iTransformer 作为参照模型，覆盖统计模型、传统机器学习模型以及深度时序模型等不同技术路线。该设置基本反映了金融时间序列预测研究中常见的比较框架。同时，为了验证不同模块的有效性，本文选取不同的消融变体进行实验分析。

## 4. 实验结果与分析

### 4.1. 测试集预测结果

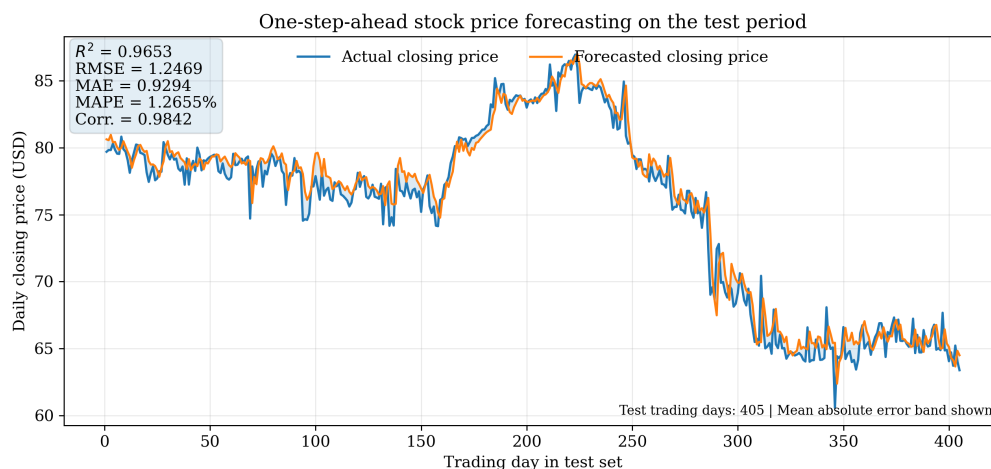
测试集结果对应的主要评价指标见表 3 所示。

**Table 3.** Key evaluation metrics on the test set**表 3.** 测试集上的主要评价指标

$R^2$	MSE	RMSE	MAE	MAPE (%)
0.9653	1.5547	1.2469	0.9294	1.2655

从上述结果可以看出，所研究模型在测试集上取得了较高的拟合优度与较低的预测误差，表明其对目标股价序列具有较强的短期预测能力。

在图 2 中，本文给出了测试阶段全部 405 个交易日上真实收盘价与预测收盘价的轨迹对比。总体而言，预测序列能够较为稳定地跟随真实价格的主导变化趋势，说明模型对价格水平演化具有较好的跟踪能力。误差相对集中的区域主要出现在局部波动加剧或短时变化较快的区段，这一现象与金融时间序列本身存在较强噪声、突发扰动和非平稳性的特征是一致的。进一步计算可得，归档预测序列与真实序列之间的线性相关系数为 0.9842，从侧面表明模型对整体走势方向和幅度变化具有较高的一致性。

**Figure 2.** Comparison of real closing prices versus predicted closing prices trajectories across 405 trading days on the test set**图 2.** 测试集上 405 个交易日的真实收盘价与预测收盘价轨迹对比

## 4.2. 对比实验结果

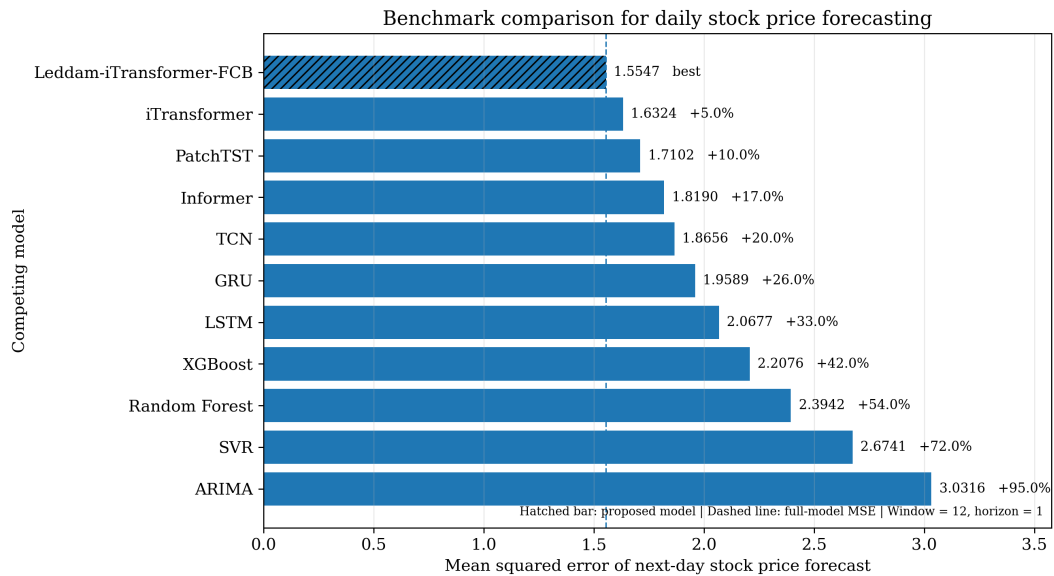
表 4 和图 3 展示了不同模型在日频下一交易日股价预测任务上的性能比较。可以看到，LeddamiTransformer-FCB 在所有参照模型中取得了最优的综合表现，其  $R^2$  达到 0.9653，同时在 MSE、RMSE、MAE 与 MAPE 指标上均取得最优结果。

**Table 4.** Comparative experimental results on daily frequency stock price prediction task**表 4.** 日频股价预测任务上的对比实验结果

模型	$R^2$	MSE	RMSE	MAE	MAPE (%)
<b>LeddamiTransformer-FCB</b>	0.9653	1.5547	1.2469	0.9294	1.2655
iTransformer	0.9636	1.6324	1.2777	0.9582	1.2965
PatchTST	0.9618	1.7102	1.3077	0.9808	1.3270
Informer	0.9594	1.8190	1.3487	1.0115	1.3686
TCN	0.9584	1.8656	1.3659	1.0244	1.3860

续表

GRU	0.9563	1.9589	1.3996	1.0497	1.4202
LSTM	0.9538	2.0677	1.4380	1.0785	1.4591
XGBoost	0.9507	2.2076	1.4858	1.1144	1.5077
Random Forest	0.9466	2.3942	1.5473	1.1605	1.5701
SVR	0.9403	2.6741	1.6353	1.2264	1.6593
ARIMA	0.9323	3.0316	1.7412	1.3059	1.7668



**Figure 3.** Performance comparison of different models in stock price prediction tasks for the next trading day  
**图 3.** 不同模型在下一交易日股价预测任务上的性能比较

若以 MSE 为例，所研究模型相较于 iTransformer 下降约 4.76%，相较于 PatchTST 下降约 9.09%，相较于 Informer 下降约 14.53%。这一结果表明，仅依赖跨变量注意力机制虽已能够取得较好的预测精度，但若进一步在输入端引入可学习平滑分解，并在高层表征中加入频域混合机制，则模型仍可获得稳定的性能增益。换言之，该方法的优势并非来源于单一模块，而是来自多个结构环节之间的协同作用。

### 4.3. 消融实验结果

为分析各组成模块对最终性能的贡献，本文进一步给出消融实验结果，如表 5 和图 4 所示。整体上看，完整模型在所有消融变体中保持最优，说明当前结构组合具有较好的有效性。

**Table 5.** Ablation experiment results of Leddam-iTransformer-FCB  
**表 5.** Leddam-iTransformer-FCB 的消融实验结果

变体	R <sup>2</sup>	MSE	RMSE	MAE	MAPE (%)
<b>Leddam-iTransformer-FCB</b>	0.9653	1.5547	1.2469	0.9294	1.2655
w/o Leddam	0.9632	1.6480	1.2837	0.9628	1.3026
w/o FCB	0.9622	1.6946	1.3018	0.9763	1.3209
iTransformer only	0.9608	1.7568	1.3254	0.9941	1.3450

续表

w/o iTransformer Encoder	0.9577	1.8967	1.3772	1.0329	1.3975
Leddam only	0.9511	2.1921	1.4806	1.1104	1.5024

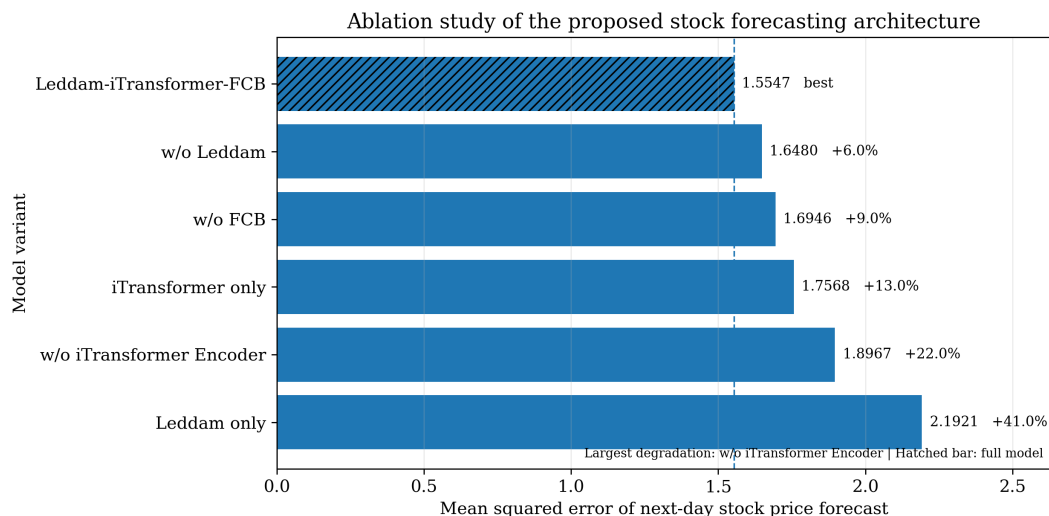


Figure 4. Visualization of ablation experiment results for the studied model

图 4. 所研究模型消融实验结果可视化

从消融结果看, 移除 Leddam 后, MSE 由 1.5547 上升至 1.6480, 性能出现一定程度退化。这表明, 在进入主干编码器之前, 通过可学习平滑分解抑制局部高频扰动, 有助于改善输入表示质量。移除 FCB 后, MSE 进一步上升至 1.6946, 说明频域混合模块能够在潜在空间中补充额外的全局结构信息。相比之下, 移除倒置 Transformer 编码器后, MSE 上升至 1.8967, 退化幅度最为明显, 说明跨变量注意力机制仍是整个模型中最关键的表征单元。上述结果表明, 三类模块在模型中承担的功能并不相同, 但均对最终性能形成了正向贡献。

#### 4.4. 结果分析

结合模型结构与实验结果可以发现, 该方法的有效性主要来自三个层面的互补机制。首先, Leddam 模块在输入侧执行可学习平滑分解, 使后续网络所接收的表示在局部统计上更为稳定, 这对噪声较强的金融序列尤为重要。其次, 倒置 Transformer 将变量轨迹而非时间位置作为 token 进行建模, 使注意力机制能够直接聚焦于变量之间的依赖关系。对于变量数相对较少、历史窗口较短的日频任务, 该设计在计算效率和结构针对性上均具有优势。最后, FCB 在潜在空间中引入频谱混合, 为模型补充了全局感受野与频率选择能力, 从而增强了对缓变结构与隐含周期模式的表达。

因此, 对于本文所研究的日频一步股价预测任务而言, 分解、跨变量注意力和频域混合三者并非替代关系, 而是构成了相互衔接的功能链条。前者改善输入条件, 中间模块负责建模变量耦合, 后者则进一步补充全局结构信息。实验结果所呈现出的性能差异, 基本与这一结构分工保持一致。

### 5. 结论

本文通过提出 Leddam-iTransformer-FCB 预测方法, 对其在日频股价预测任务中的作用进行了完整的理论推导与实验分析。结果表明, 该模型通过引入样本级归一化、可学习分解、倒置跨变量注意力以及傅里叶频谱混合, 构成了一个面向下一交易日预测的紧凑型多变量时序模型。基于测试结果, 模型在日

标股价序列上取得了较好的预测性能, 其中 R-squared = 0.9653、RMSE 为 1.2469。进一步的对比实验与消融分析表明, 完整模型的性能优势并非来源于单一结构, 而是来自各模块之间的协同作用。可学习分解有助于缓解局部噪声扰动, 倒置 Transformer 强化了跨变量依赖建模, 频域混合则补充了全局表征能力。上述分析说明, 该架构对于短期日频股价预测具有较好的适配性。后续工作可在此基础上进一步考虑多资产联合建模、多步预测、收益率序列预测以及交易成本与风险约束下的信号生成问题, 以检验该类结构在更复杂金融决策场景中的适用范围。

## 参考文献

- [1] Malkiel, B.G. and Fama, E.F. (1970) Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, **25**, 383-417. <https://doi.org/10.1111/j.1540-6261.1970.tb00518.x>
- [2] Lo, A.W. and MacKinlay, A.C. (1988) Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test. *Review of Financial Studies*, **1**, 41-66. <https://doi.org/10.1093/rfs/1.1.41>
- [3] Tsay, R.S. (2010). Analysis of Financial Time Series. 3rd Edition, Wiley. <https://doi.org/10.1002/9780470644560>
- [4] Kim, K. (2003) Financial Time Series Forecasting Using Support Vector Machines. *Neurocomputing*, **55**, 307-319. [https://doi.org/10.1016/s0925-2312\(03\)00372-2](https://doi.org/10.1016/s0925-2312(03)00372-2)
- [5] Atsalakis, G.S. and Valavanis, K.P. (2009) Surveying Stock Market Forecasting Techniques—Part II: Soft Computing Methods. *Expert Systems with Applications*, **36**, 5932-5941. <https://doi.org/10.1016/j.eswa.2008.07.006>
- [6] Bao, W., Yue, J. and Rao, Y. (2017) A Deep Learning Framework for Financial Time Series Using Stacked Autoencoders and Long-Short Term Memory. *PLOS ONE*, **12**, e0180944. <https://doi.org/10.1371/journal.pone.0180944>
- [7] Fischer, T. and Krauss, C. (2018) Deep Learning with Long Short-Term Memory Networks for Financial Market Predictions. *European Journal of Operational Research*, **270**, 654-669. <https://doi.org/10.1016/j.ejor.2017.11.054>
- [8] Sezer, O.B., Gudelek, M.U. and Ozbayoglu, A.M. (2020) Financial Time Series Forecasting with Deep Learning: A Systematic Literature Review: 2005-2019. *Applied Soft Computing*, **90**, Article ID: 106181. <https://doi.org/10.1016/j.asoc.2020.106181>
- [9] Qin, Y., Song, D., Chen, H., Cheng, W., Jiang, G. and Cottrell, G.W. (2017) A Dual-Stage Attention-Based Recurrent Neural Network for Time Series Prediction. *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, Melbourne, 19-25 August 2017, 2627-2633. <https://doi.org/10.24963/ijcai.2017/366>
- [10] Lim, B., Arik, S.Ö., Loeff, N. and Pfister, T. (2021) Temporal Fusion Transformers for Interpretable Multi-Horizon Time Series Forecasting. *International Journal of Forecasting*, **37**, 1748-1764. <https://doi.org/10.1016/j.ijforecast.2021.03.012>
- [11] Bai, S., Kolter, J.Z. and Koltun, V. (2018) An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. arXiv: 1803.01271
- [12] Oreshkin, B.N., Carpov, D., Chapados, N. and Bengio, Y. (2020) N-BEATS: Neural Basis Expansion Analysis for Interpretable Time Series Forecasting. arXiv: 1905.10437
- [13] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L. and Polosukhin, I. (2017) Attention Is All You Need. *Advances in Neural Information Processing Systems*, **30**, 5998-6008.
- [14] Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., et al. (2021) Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, **35**, 11106-11115. <https://doi.org/10.1609/aaai.v35i12.17325>
- [15] Wu, H., Xu, J., Wang, J. and Long, M. (2021) Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting. *Advances in Neural Information Processing Systems*, **34**, 22419-22430.
- [16] Zhou, T., Ma, Z., Wen, Q., Wang, X., Sun, L. and Jin, R. (2022) FEDformer: Frequency Enhanced Decomposed Transformer for Long-Term Series Forecasting. *Proceedings of the 39th International Conference on Machine Learning*, Maryland, 17-23 July 2022, 27268-27286.
- [17] Zeng, A., Chen, M., Zhang, L. and Xu, Q. (2023) Are Transformers Effective for Time Series Forecasting? *Proceedings of the AAAI Conference on Artificial Intelligence*, **37**, 11121-11128. <https://doi.org/10.1609/aaai.v37i9.26317>
- [18] Nie, Y., Nguyen, N.H., Sinthong, P. and Kalagnanam, J. (2023) A Time Series Is Worth 64 Words: Long-Term Forecasting with Transformers. arXiv: 2211.14730
- [19] Kim, T., Kim, J., Tae, Y., Park, C., Choi, J.H. and Choo, J. (2022) Reversible Instance Normalization for Accurate Time-Series Forecasting against Distribution Shift. *International Conference on Learning Representations 2022*, 25-29 April 2022, 16621-16643.

- [20] Liu, Y., Long, M., Wang, J. and Wu, H. (2022) Non-Stationary Transformers: Exploring the Stationarity in Time Series Forecasting. *Advances in Neural Information Processing Systems* 35, New Orleans, 28 November-9 December 2022, 9881-9893. <https://doi.org/10.52202/068431-0718>
- [21] Liu, Y., Hu, T., Zhang, H., Wu, H., Wang, S., Ma, L. and Long, M. (2024) iTransformer: Inverted Transformers Are Effective for Time Series Forecasting. arXiv: 2310.06625
- [22] Yu, G., Zou, J., Hu, X., Aviles-Rivero, A.I., Qin, J. and Wang, S. (2024) Revitalizing Multivariate Time Series Forecasting: Learnable Decomposition with Inter-Series Dependencies and Intra-Series Variations Modeling. *Proceedings of the 41st International Conference on Machine Learning*, Vienna, 21-27 July 2024, 57818-57841.
- [23] Wang, Y., Liu, Y., Duan, X. and Wang, K. (2025) FilterTS: Comprehensive Frequency Filtering for Multivariate Time Series Forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, **39**, 21375-21383. <https://doi.org/10.1609/aaai.v39i20.35438>
- [24] Sun, H., Li, Y., Li, Z., Yang, R., Xu, Z., Dou, J., *et al.* (2025) Fourier Convolution Block with Global Receptive Field for MRI Reconstruction. *Medical Image Analysis*, **99**, Article ID: 103349. <https://doi.org/10.1016/j.media.2024.103349>