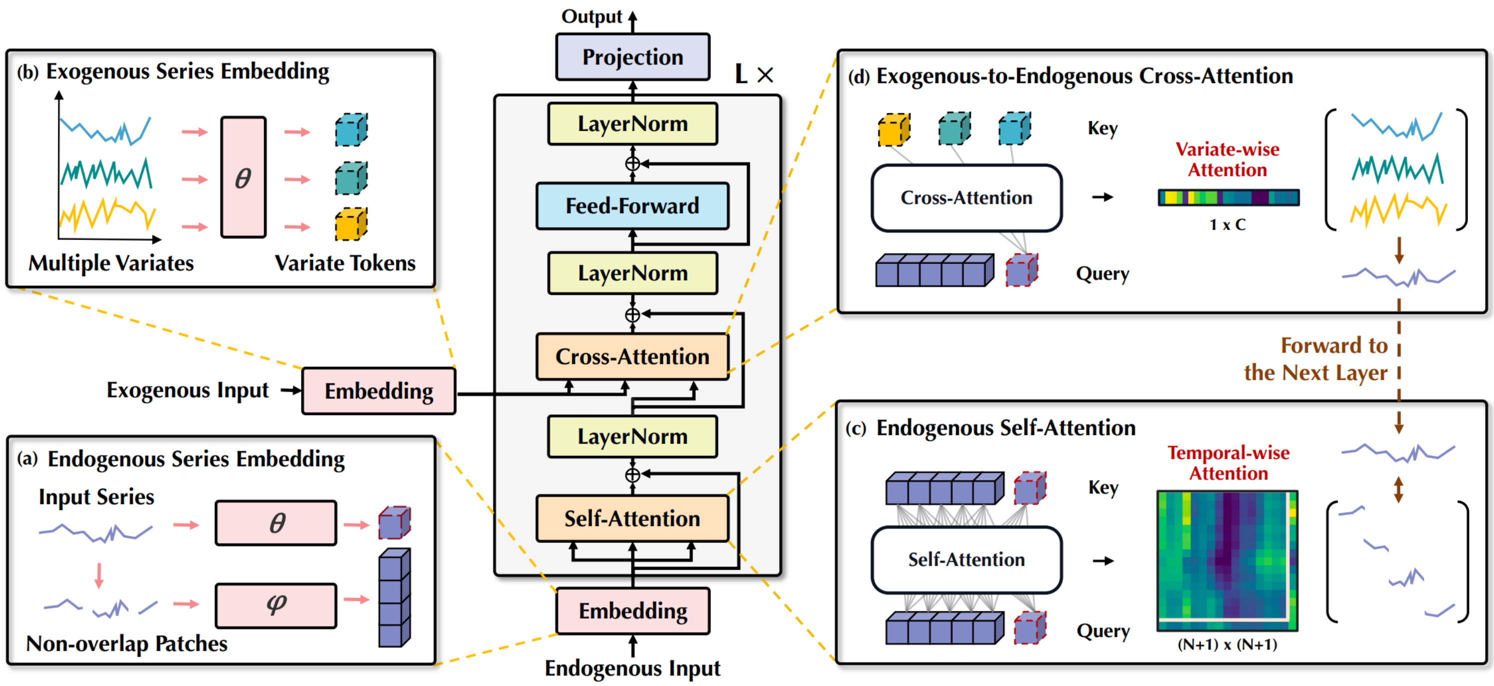
arXiv: 24.02.29

TimeXer

Empowering Transformers for Time Series Forecasting with eXogenous Variables



arXiv: 24.02.29

TimeXer

Empowering Transform**er**s for **Time** Series Forecasting with e**X**ogenous Variables

24.3.26

Presented by Yyyq

01 问题描述

- **Endogenous & Exogenous**
  - Endogenous：内生的，内因的，内源性的
  - Exogenous：外生的，外因的，外源性的
- **Endogenous variables & Exogenous variables**
  - 内生变量：只关注感兴趣的目标
  - 外生变量：为内生变量提供有价值的外部信息
- **通过外生变量引入辅助信息来促进内生变量的预测**

01 问题描述

➢ **En. & Ex. (from Datasets)**

(1)ECL：客户的每小时用电量数据。

将最后一个客户端的用电量作为内生变量，其他客户端作为外生变量。

(2)**天气**：气象站每10分钟收集的21个气象因子。

使用Wet Bulb因子作为内生变量，其他指标作为外生变量。

(3)ETT：变压器油温数据。

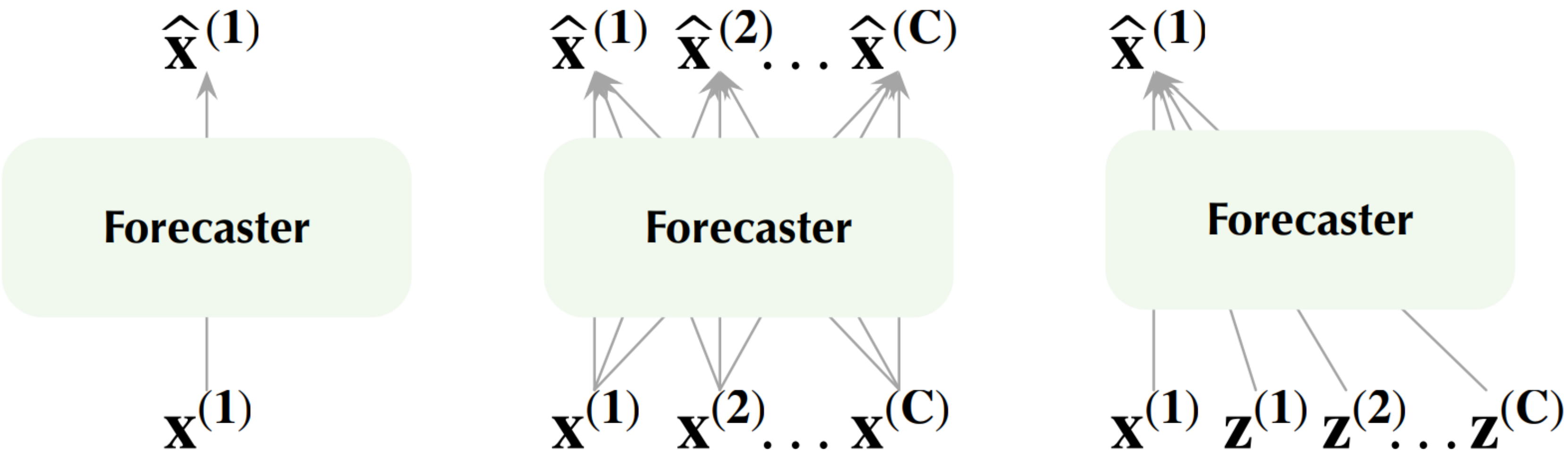
内生变量为油温，外生变量为6个电力负荷特征。

(4)**交通**：高速公路传感器测量的每小时道路占用率。

将最后一个传感器的测量作为内生变量，将其他传感器作为外生变量。

01 问题描述

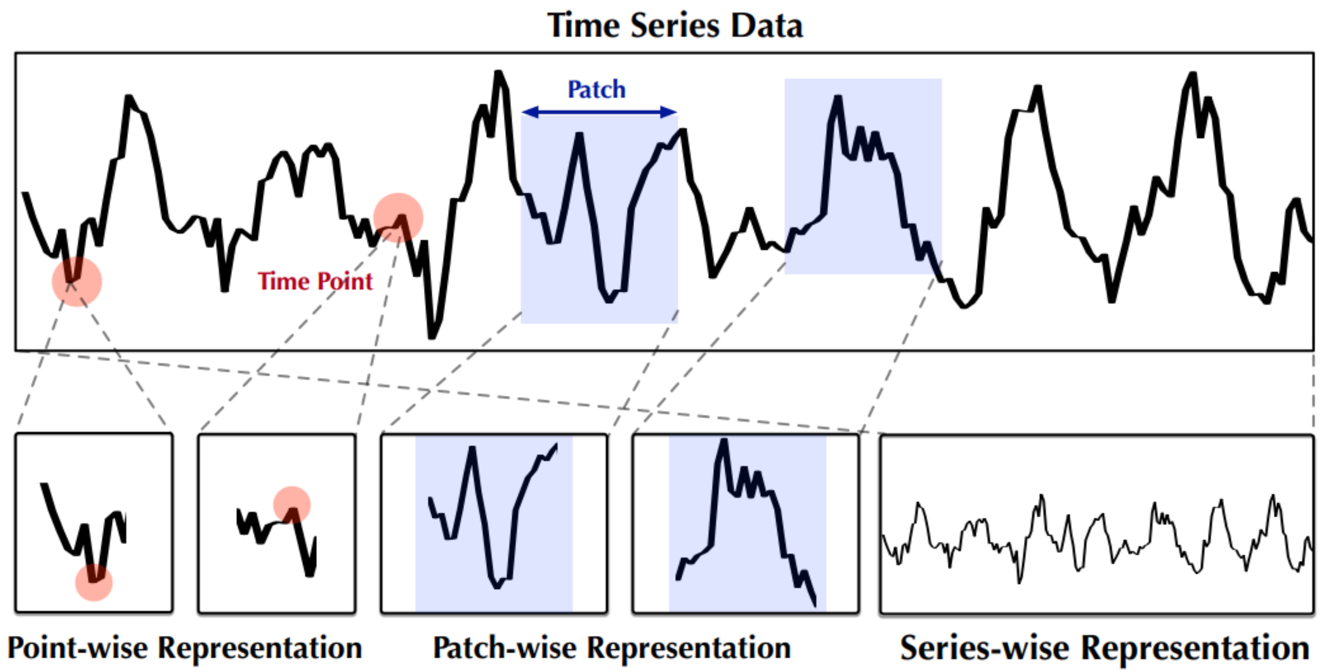
- 内生变量和外生变量之间：**需要调和差异和依赖性**
- 外部因素对内生序列的影响：**可能是连续和时滞的**
  - patchTST：只能捕获时间依赖性，而不能捕获多变量相关性
  - iTransformer：无法捕获不同子序列之间的时间变化



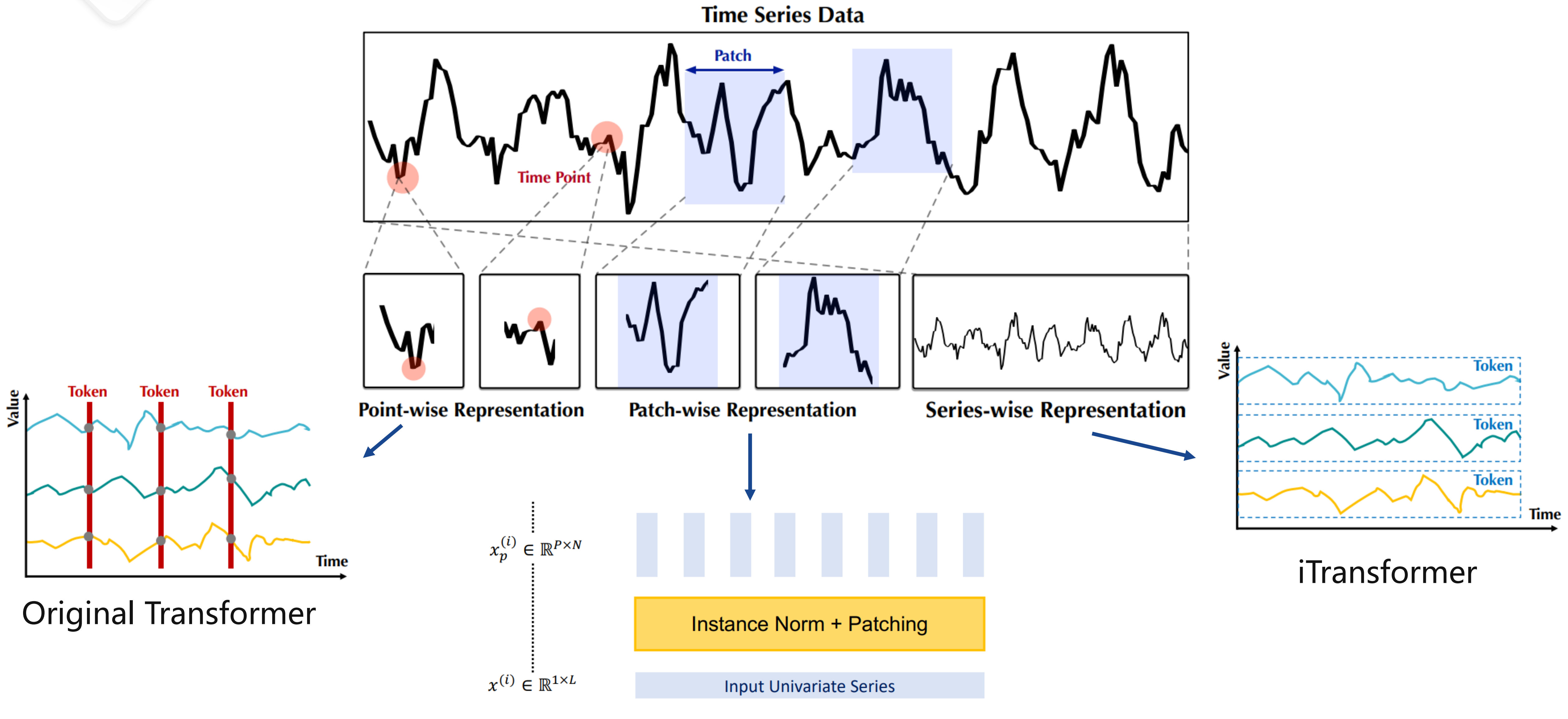
(a) Univariate (b) Multivariate (c) Ours

02 创新点

- **TimeXer**
  - 在包含外生变量的基础上，**不修改transformer架构**
  - **Self-attention**：对patch-level的内生时序token提取时间相关性
  - **Cross-attention**：对变量token提取多变量相关性



03 相关工作



PatchTST

04 算法实现：问题定义

- **只预测内生时间序列，外生变量是附加因素**
  - 内生序列：

$$\mathbf{x}_{1:L} = \{x_1, x_2, ..., x_L\} \in \mathbb{R}^{L\times 1}$$

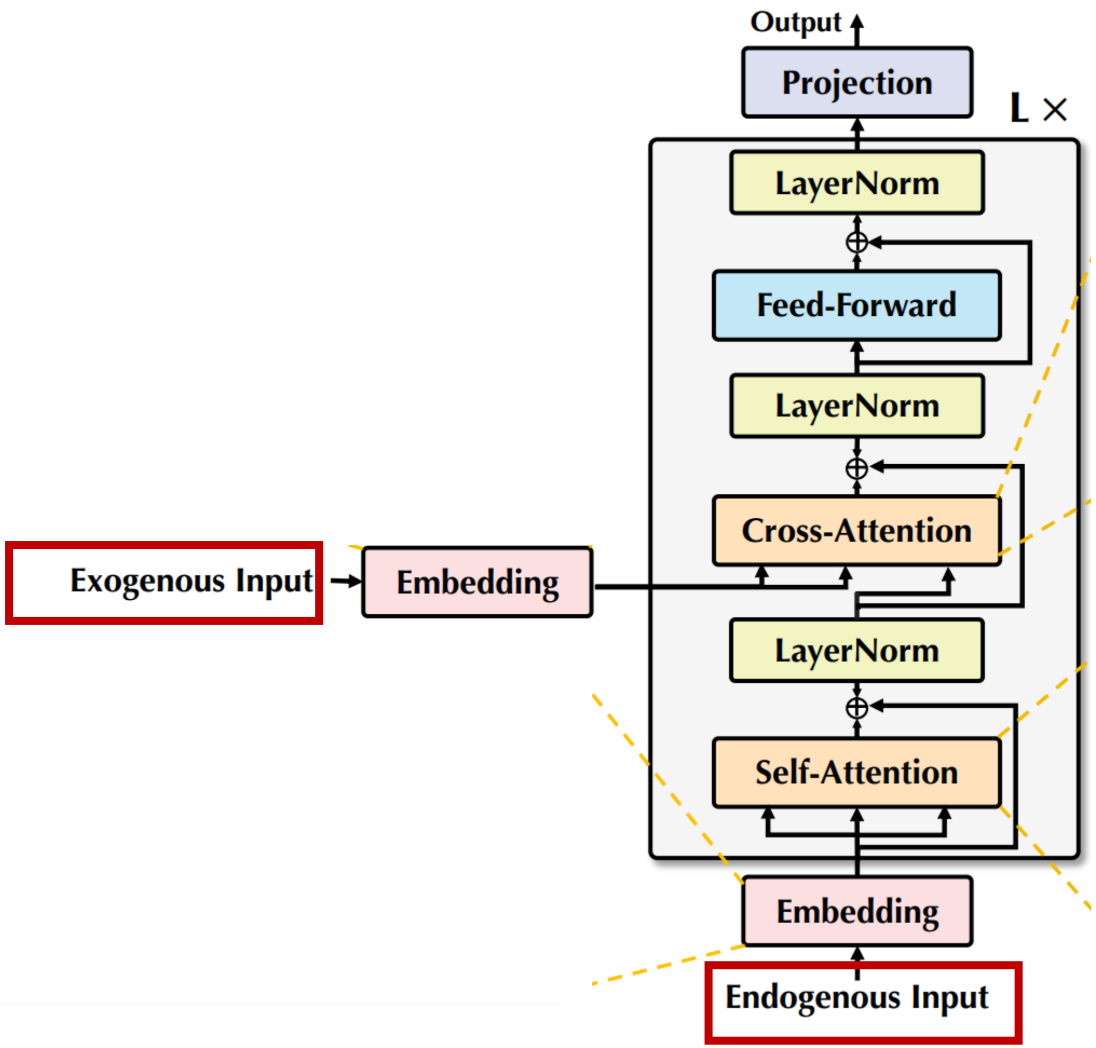
  - 外生序列：

$$\mathbf{z}_{1:L'} = \{\mathbf{z}_{1:L'}^{(1)}, \mathbf{z}_{1:L'}^{(2)}, ..., \mathbf{z}_{1:L'}^{(C)}\} \in \mathbb{R}^{L\times C}$$

  - L 和 L' 可以不相等
- **预测问题**

$$\widehat{\mathbf{x}} = f\left(\mathbf{x}_{1:L}, \mathbf{z}_{1:L'}\right).$$

04 算法实现：整体结构

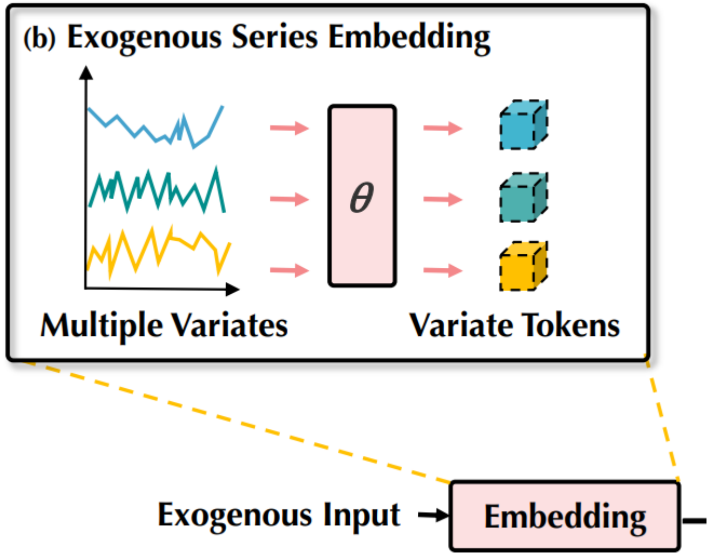

Output
Projection
L ×
LayerNorm
Feed-Forward
LayerNorm
Cross-Attention
Exogenous Input
Embedding
LayerNorm
Self-Attention
Embedding
Endogenous Input

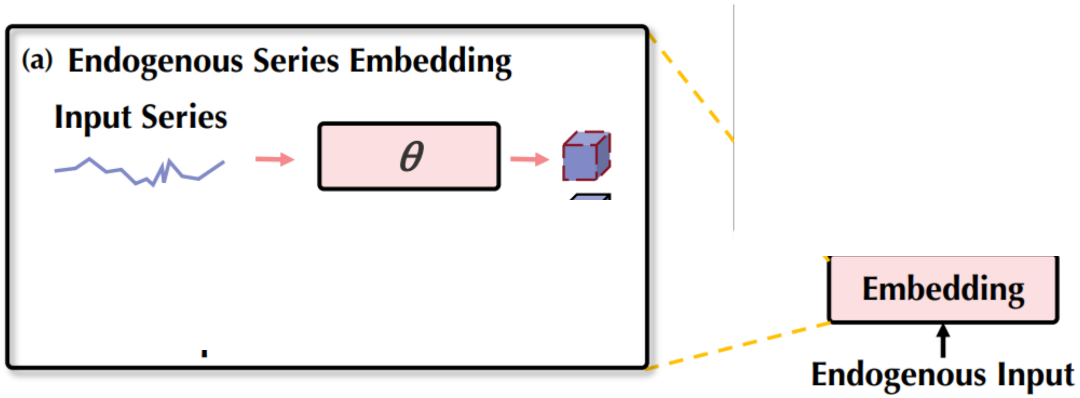
04 算法实现：Variate Embedding

- **基于itransformer的嵌入方式**
  - Series → Token
  - Series-global representation：全局标记
- **线性层：** $L \rightarrow D$，$L' \rightarrow D$

$$\mathbf{V}_{en} = \mathrm{EnVariateEmbed}\left(\mathbf{x}\right), \quad \mathbb{R}^{L} \rightarrow \mathbb{R}^{D}$$

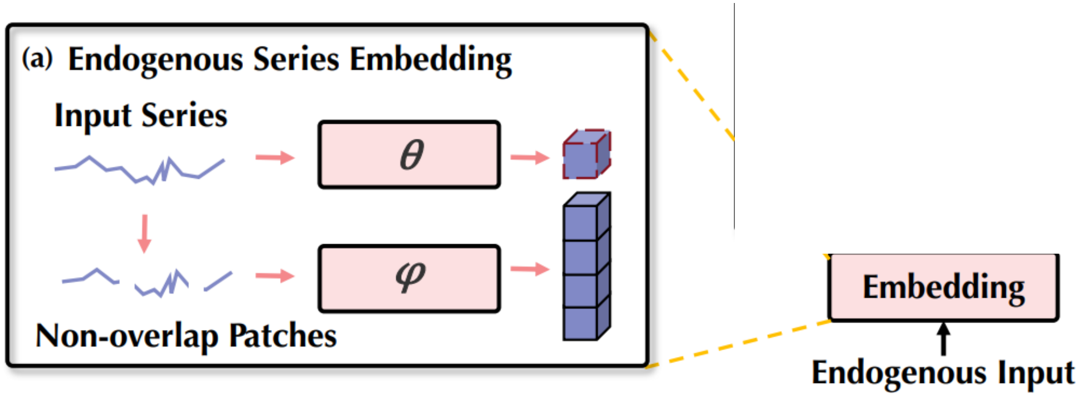
$$\mathbf{V}_{ex,i} = \mathrm{ExVariateEmbed}\left(\mathbf{z}^{(i)}\right)\ i \in \{1, \cdots, C\}. \quad \mathbb{R}^{L'} \rightarrow \mathbb{R}^{D}$$





04 算法实现：Patch Embedding

- **基于patchTST的嵌入方式**
  - Patch → Token
  - Patch不重叠
- **线性层 + PE：** $P \rightarrow D$
  - P： patch的长度，N：总Patch个数

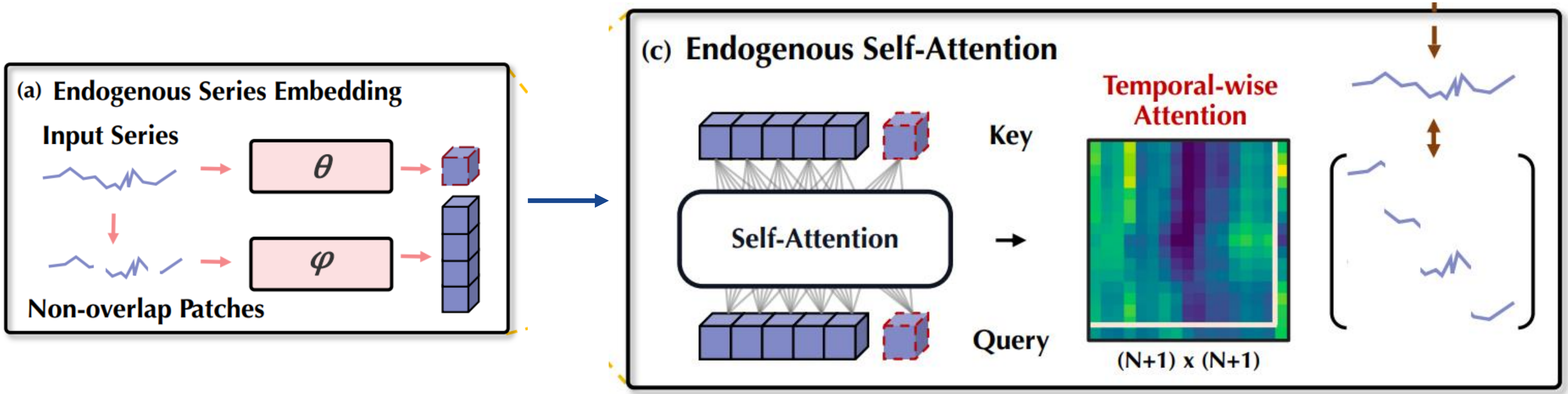


$$\{\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_N\} = \mathrm{Patchify}\left(\mathbf{x}\right),$$
$$\mathbf{P}_{en} = \mathrm{PatchEmbed}\left(\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_N\right).$$

04 算法实现：Patch-wise Self-Attention

- **内生变量的内在时间依赖性**（不预测外生变量）
  - Concat：patch-wise 和 series-wise tokens
  - Attention：$(N+1)\times(N+1)$ （N个patch，1个series）
- **Self-attention + 残差 + LN归一化层**

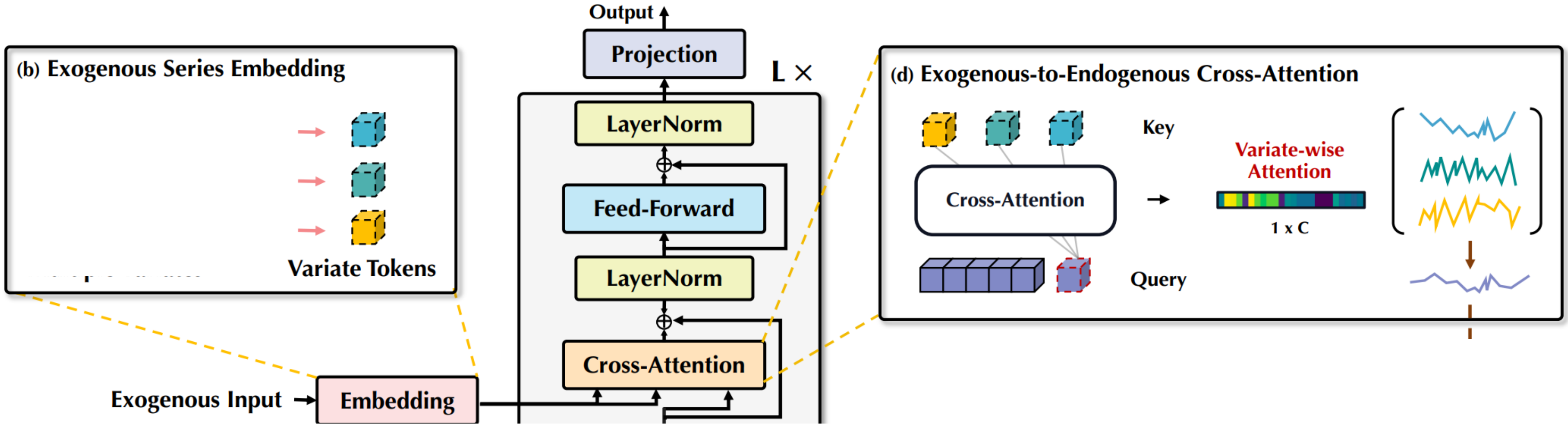
$$\widehat{\mathbf{P}}^{l}_{en},\widehat{\mathbf{V}}^{l}_{en}=\mathrm{LN}\left(\left[\mathbf{P}^{l}_{en},\mathbf{V}^{l}_{en}\right]+\text{Self-Attn}\left(\left[\mathbf{P}^{l}_{en},\mathbf{V}^{l}_{en}\right]\right)\right)$$



04 算法实现：Variate-wise Cross-Attention

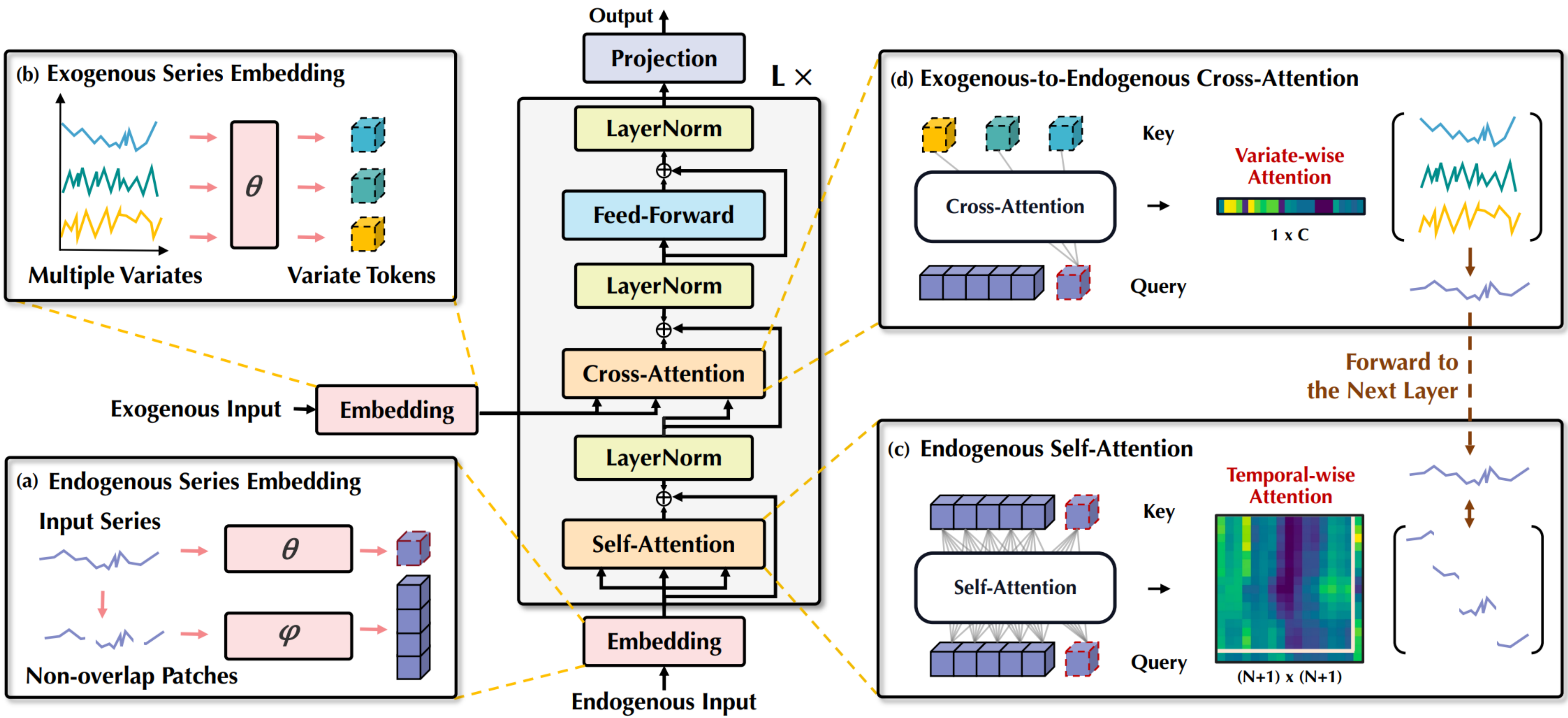
- **内生变量的内在时间依赖性**（不预测外生变量）
  - Query：内生变量series-wise的注意力结果
  - Key 和 Value：外生变量的token
- **Cross-attention + 残差 + LN归一化层**

$$\mathbf{V}_{en}^{l+1} = \mathrm{LN}\left(\widehat{\mathbf{V}}_{en}^{l} + \text{Cross-Attn}\left(\widehat{\mathbf{V}}_{en}^{l}, \mathbf{V}_{ex}, \mathbf{V}_{ex}\right)\right).$$



04

算法实现：整体结构



(b) Exogenous Series Embedding
θ
Multiple Variates
Variate Tokens
Output
Projection
L ×
LayerNorm
Feed-Forward
LayerNorm
Cross-Attention
LayerNorm
Self-Attention
Exogenous Input
Embedding
Embedding
Endogenous Input
(a) Endogenous Series Embedding
Input Series
θ
φ
Non-overlap Patches
(d) Exogenous-to-Endogenous Cross-Attention
Key
Cross-Attention
Variate-wise Attention
1 x C
Query
Forward to the Next Layer
(c) Endogenous Self-Attention
Key
Self-Attention
Temporal-wise Attention
Query
(N+1) x (N+1)

05

实验：数据集

Dataset	#Num	Ex. Descriptions	En. Descriptions	Sampling Frequency	Dataset Size
Electricity	320	Electricity Consumption	Electricity Consumption	1 Hour	(18317, 2633, 5261)
Weather	20	Climate Feature	CO2-Concentration	10 Minutes	(36792, 5271, 10540)
ETTh	6	Power Load Feature	Oil Temperature	1 Hour	(8545, 2881, 2881)
ETTm	6	Power Load Feature	Oil Temperature	15 Minutes	(34465, 11521, 11521)
Traffic	861	Road Occupancy Rates	Road Occupancy Rates	1 Hour	(12185, 1757, 3509)
NP	2	Grid Load, Wind Power	Nord Pool Electricity Price	1 Hour	(36500, 5219, 10460)
PJM	2	System Load, SyZonal COMED load	Pennsylvania-New Jersey-Maryland Electricity Price	1 Hour	(36500, 5219, 10460)
BE	2	Generation, System Load	Belgium's Electricity Price	1 Hour	(36500, 5219, 10460)
FR	2	Generation, System Load	France's Electricity Price	1 Hour	(36500, 5219, 10460)
DE	2	Wind power, Ampirion zonal load	German's Electricity Price	1 Hour	(36500, 5219, 10460)

长期预测——用电量，天气，变压器油温，交通：一个内生EN.，多个外生EX.
短期预测——5个主要市场的电价短期预测数据集：电价作为内生变量，两个在实践中有影响的外生变量

05 实验：短期预测

patch长度为24，用168预测24

Model	TimeXer (Ours)		iTrans. (2023)		RLinear (2023)		PatchTST (2022)		Cross. (2022)		TiDE (2023)		TimesNet (2023a)		DLinear (2023)		SCINet (2022a)		Stationary (2022b)		Auto. (2021)	
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
NP	**0.238**	**0.268**	0.265	0.300	0.335	0.340	0.267	0.284	0.245	0.289	0.335	0.340	0.250	0.289	0.309	0.321	0.373	0.368	0.294	0.308	0.402	0.398
PJM	**0.088**	**0.188**	0.097	0.197	0.124	0.229	0.106	0.209	0.149	0.198	0.124	0.228	0.097	0.195	0.108	0.215	0.143	0.259	0.122	0.228	0.168	0.267
BE	**0.374**	**0.241**	0.394	0.270	0.520	0.337	0.403	0.264	0.436	0.294	0.523	0.336	0.419	0.288	0.463	0.313	0.731	0.412	0.433	0.289	0.500	0.333
FR	**0.381**	**0.211**	0.439	0.233	0.507	0.290	0.411	0.220	0.440	0.216	0.510	0.290	0.431	0.234	0.429	0.260	0.855	0.384	0.466	0.242	0.519	0.295
DE	**0.440**	**0.418**	0.479	0.443	0.574	0.498	0.461	0.432	0.540	0.423	0.568	0.496	0.502	0.446	0.520	0.463	0.565	0.497	0.483	0.447	0.674	0.544
AVG	**0.304**	**0.265**	0.335	0.289	0.412	0.339	0.330	0.282	0.362	0.284	0.412	0.338	0.340	0.290	0.366	0.314	0.533	0.384	0.360	0.303	0.453	0.368

05 实验：长期预测

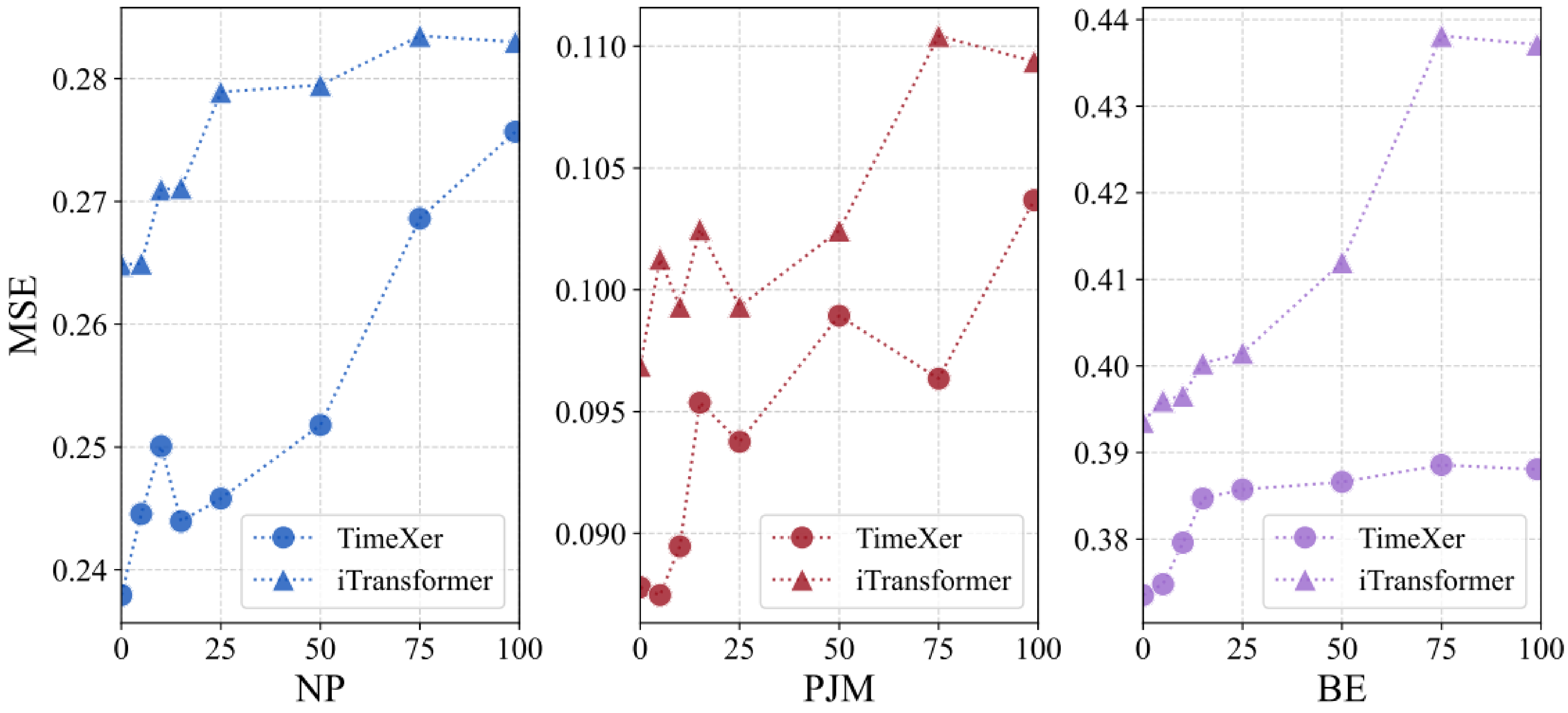
patch长度为16，历史序列长度为96，预测{96,192,336,720}

Model	TimeXer (Ours)		iTrans. (2023)		RLinear (2023)		PatchTST (2022)		Cross. (2022)		TiDE (2023)		TimesNet (2023a)		DLinear (2023)		SCINet (2022a)		Stationary (2022b)		Auto. (2021)	
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ECL	**0.336**	**0.414**	0.365	0.442	0.444	0.486	0.394	0.446	0.344	0.412	0.419	0.468	0.410	0.476	0.393	0.457	0.427	0.490	0.372	0.450	0.495	0.528
Weather	0.002	0.031	0.002	0.031	**0.002**	**0.029**	0.002	0.031	0.005	0.055	**0.002**	**0.029**	0.097	0.115	0.006	0.066	0.007	0.071	0.002	0.031	0.006	0.060
ETTh1	**0.074**	**0.210**	0.075	0.211	0.084	0.224	0.078	0.215	0.285	0.447	0.083	0.223	0.076	0.215	0.116	0.259	0.437	0.565	0.110	0.256	0.130	0.282
ETTh2	**0.183**	**0.337**	0.199	0.352	0.205	0.356	0.192	0.345	1.027	0.873	0.205	0.356	0.210	0.362	0.224	0.369	1.155	0.955	0.262	0.405	0.242	0.386
ETTm1	**0.051**	**0.169**	0.053	0.175	0.053	0.173	0.053	0.173	0.411	0.548	0.053	0.173	0.054	0.175	0.066	0.188	0.098	0.241	0.077	0.204	0.085	0.230
ETTm2	**0.116**	**0.252**	0.127	0.267	0.122	0.261	0.120	0.258	0.976	0.769	0.122	0.261	0.129	0.271	0.126	0.263	0.685	0.713	0.207	0.333	0.154	0.305
Traffic	**0.150**	**0.227**	0.161	0.246	0.324	0.412	0.173	0.253	-	-	0.324	0.411	0.171	0.264	0.323	0.404	0.447	0.500	0.361	0.361	0.302	0.353

05

实验：消融实验——屏蔽外生变量

使用完整的内生序列，对外生序列掩码从**0~99%**



05

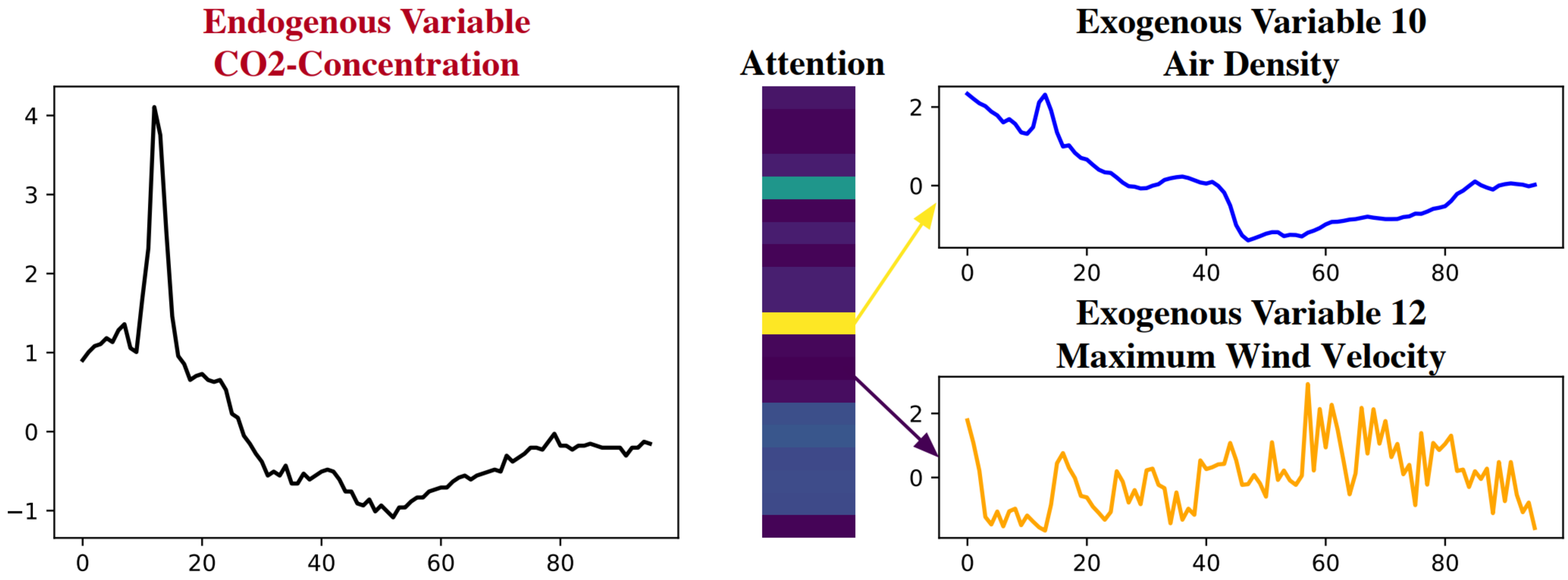
实验：消融实验——不同嵌入层的有效性

T：temporal token；V：variate token

Design	En.	Ex.	ETTh2		ETTm2		Traffic	
			MSE	MAE	MSE	MAE	MSE	MAE
Ours	T+V	V	**0.183**	**0.337**	**0.116**	**0.252**	**0.150**	**0.227**
Replace	T+V	T	0.192	0.343	0.122	0.259	0.158	0.239
w/o	w/o V	V	0.197	0.347	0.117	0.254	0.158	0.239
	w/o T	V	0.188	0.341	0.118	0.256	0.152	0.233

05 实验：模型分析——变量间的相关性

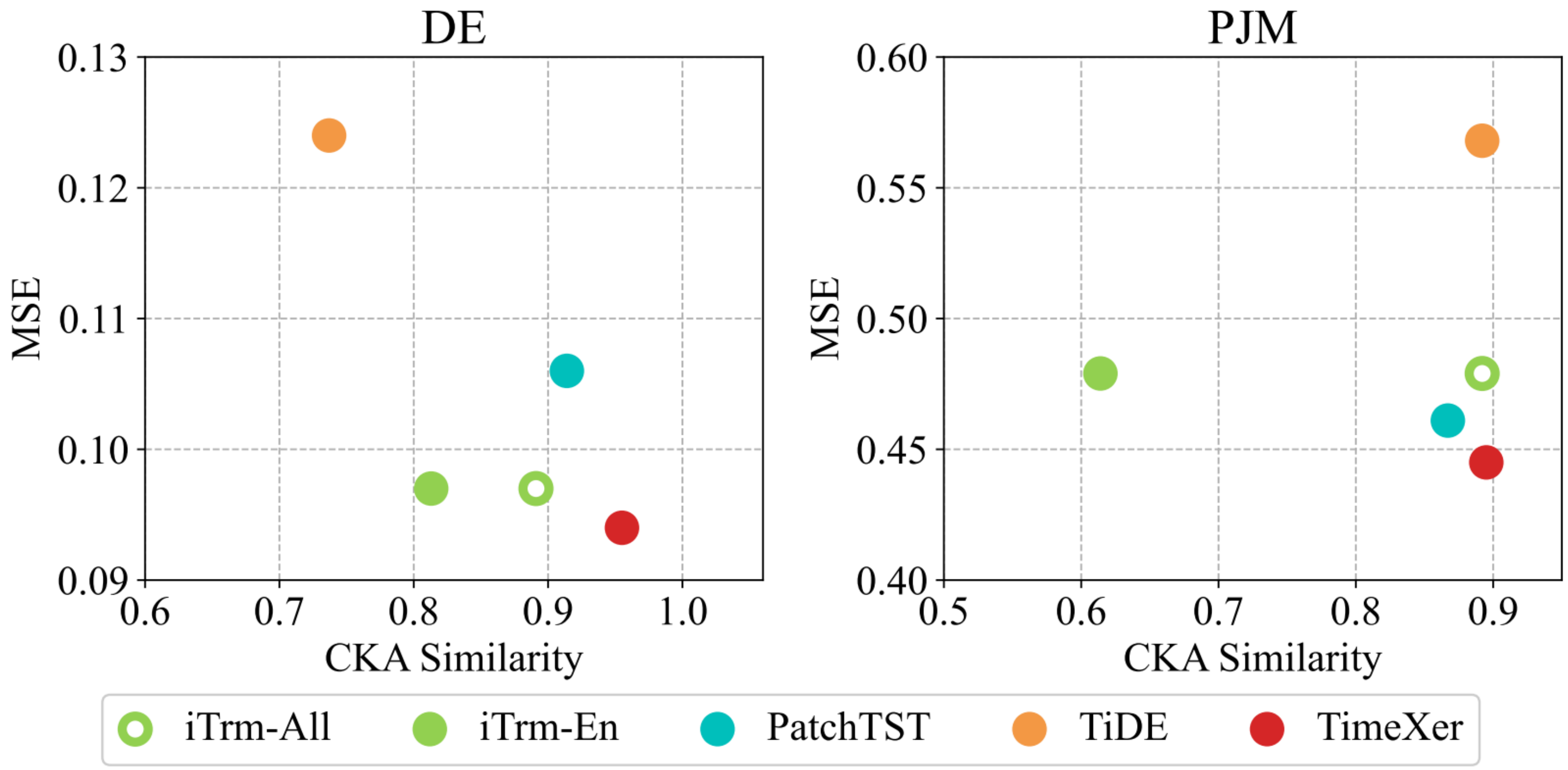
交叉注意力机制中的内生变量和外生变量之间相关性的注意力图



05

实验：模型分析——表示分析

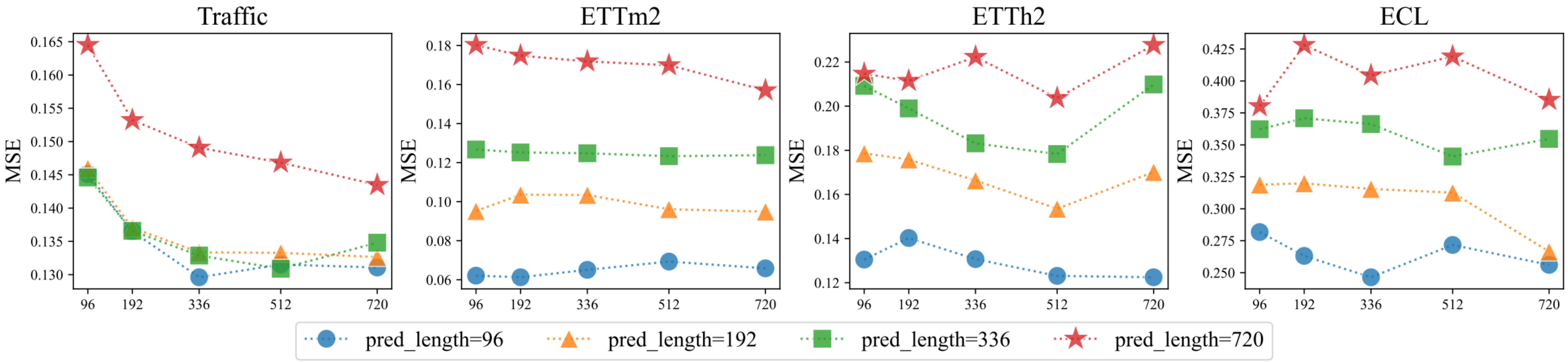
CKA相似度：比较两个特征空间的相似性



绿色空心点：itransformer用所有序列表示；
绿色点： itransformer只用内生变量的序列表示

05 实验：模型分析——不同外生变量的回望窗口

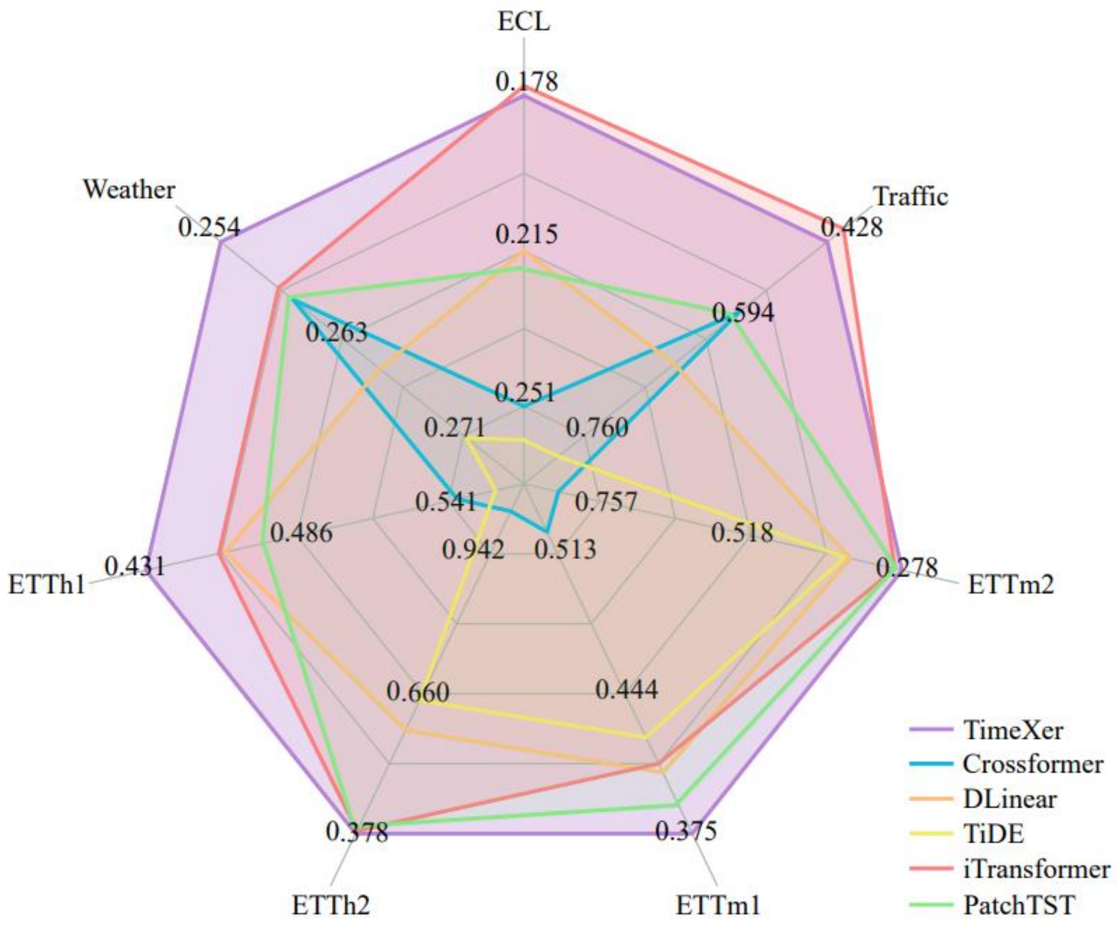
内生变量的长度固定为96，并增加外生变量的长度



05

实验：模型分析——多变量预测

数据集中的变量视为相互独立的内生变量，每个变量考虑所有其他变量作为外生变量



谢谢观看

MANY THANKS !

24.3.26