SAMoVAR: Structural Aligned Mixture of VAR for Time-Series Forecasting

Jiecheng Lu Yan Sun Shihao Yang

Georgia Institute of Technology

Introduction & Motivation

- **Autoregressive Transformers** excel in sequence modelling but vanilla softmax-attention is $O(N^2)$. *Linear attention* reduces this to $O(N)$ and outperforms vanilla on some TSF tasks.
- **Classic VAR** captures lag dependencies with interpretable weights but assumes linearity/fixed lag.
- **Problem:** Deeper Transformers drift from the autoregressive generative process, obscuring an underlying *VAR structure* latent in linear attention.
- **Key Insight:** A single linear attention layer *is* a dynamic VAR model. We realign multi-layer Transformers to preserve and exploit this.

Key Contributions

1. **Linear Attention → VAR**: proof that one-layer linear attention equals a dynamic VAR.
2. **Diagnose Misalignments**: identify residual, loss, and token-drift issues when stacking layers.
3. **Structural Alignment**: reorder layers so multi-layer linear attention maintains an explicit VAR with *temporal influence paths*.
4. **SAMoVAR**: new Transformer variant combining a mixture of VARs for accurate, interpretable, efficient TSF.

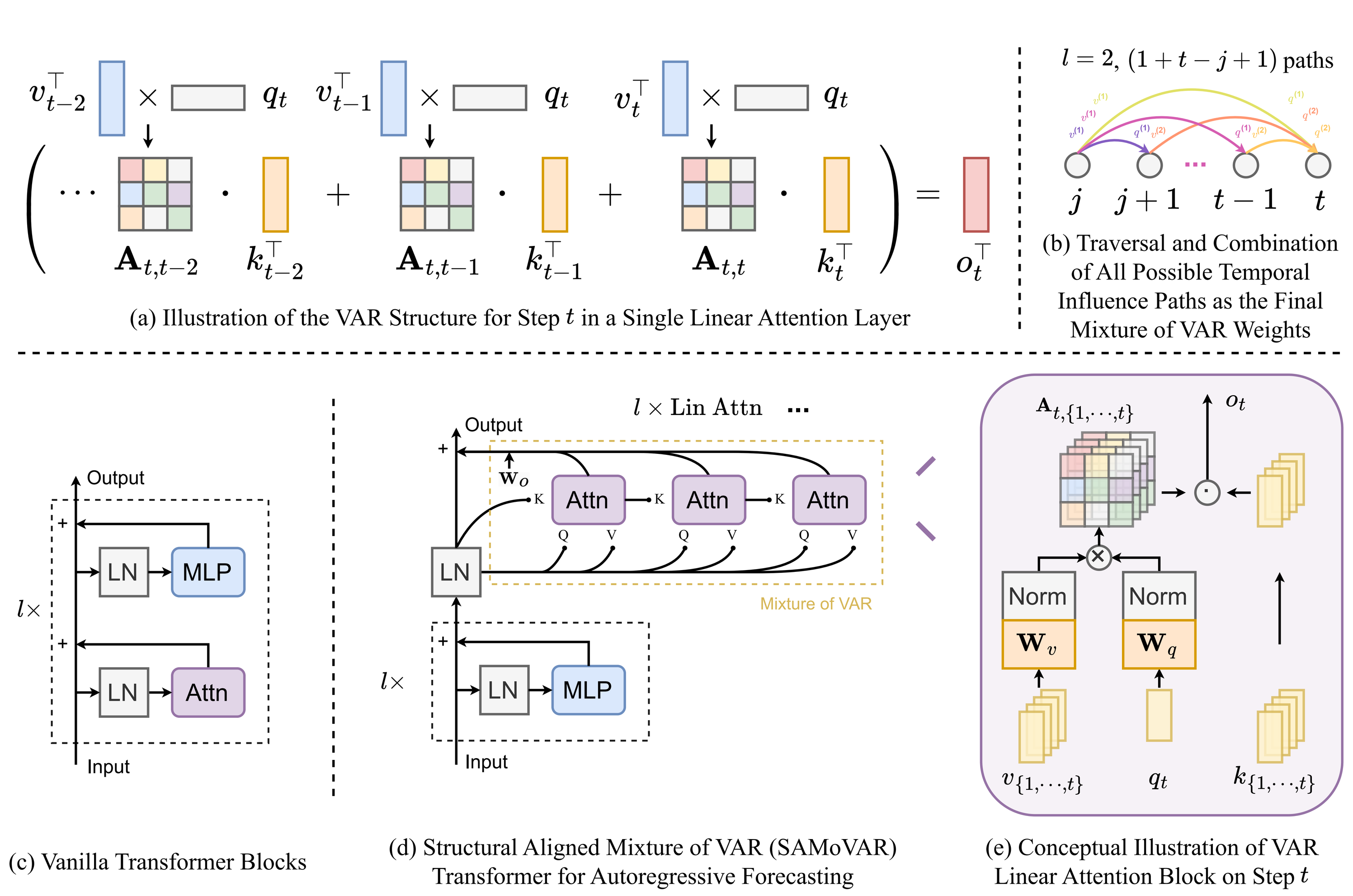

Figure 1. Visualization of Key Concepts in SAMoVAR. The subfigures highlight different structural and conceptual elements of the model.

Method Overview: SAMoVAR Architecture

- **Pipeline**: MLP preprocessing → stack of l *linear-attention* layers (no MLPs between) → mixture of VAR outputs.
- **Dynamic VAR Weights**: weight matrices $A_{t,i} = \mathbf{v}_i^\top \mathbf{q}_t$; stacking allows higher-rank & multi-step *influence paths*.
- **Stability**: identity keys, RMSNorm on $\mathbf{q}, \mathbf{v}$, passive path pruning.
- **Patch-wise ARX Tokenization**: alternate exogenous–target tokens; preserves univariate dynamics while enabling cross-series effects.

$$\mathbf{o}_t^{(1)\top} = \sum_{j=1}^{t} \mathbf{B}_{t,j}^{(1)} \mathbf{k}_j^\top, \quad \mathbf{B}_{t,j}^{(1)} = \mathbf{A}_{t,j}^{(1)} = \mathbf{v}_j^{(1)\top} \mathbf{q}_t^{(1)}.$$

$$\mathbf{o}_t^{(2)\top} = \sum_{j=1}^{t} \mathbf{B}_{t,j}^{(2)} \mathbf{k}_j^\top, \mathbf{B}_{t,j}^{(2)} = \sum_{i=j}^{t} \underbrace{\left(\mathbf{v}_i^{(2)\top} \mathbf{q}_t^{(2)}\right)}_{\mathbf{A}_{t,i}^{(2)}} \mathbf{W}_k^{(2)\top} \mathbf{B}_{i,j}^{(1)}$$

$$\mathbf{B}_{t,j}^{(l)} = \sum_{i=j}^{t} \underbrace{\left(\mathbf{v}_i^{(l)\top} \mathbf{q}_t^{(l)}\right)}_{\mathbf{A}_{t,i}^{(l)}} \mathbf{W}_k^{(l)\top} \mathbf{B}_{i,j}^{(l-1)}.$$

Temporal Influence Paths

Each past step j influences t via paths of at most $l-1$ intermediates. Number of paths: $\binom{(t-j)+l-1}{l-1}$. Paths provide interpretability: rank-1 contributions reveal lag structure and cross-series effects.

$$\mathbf{P}_{t,j,\{i_1,\cdots,i_{l-1}\}}^{(l)} = \mathbf{A}_{t,i_1}^{(l)} \mathbf{A}_{i_1,i_2}^{(l-1)} \cdots \mathbf{A}_{i_{l-1},j}^{(1)} = \mathbf{v}_{i_1}^{(l)\top} \mathbf{q}_t^{(l)} \mathbf{v}_{i_2}^{(l-1)\top} \mathbf{q}_{i_1}^{(l-1)} \cdots \mathbf{v}_j^{(1)\top} \mathbf{q}_{i_{l-1}}^{(1)}$$

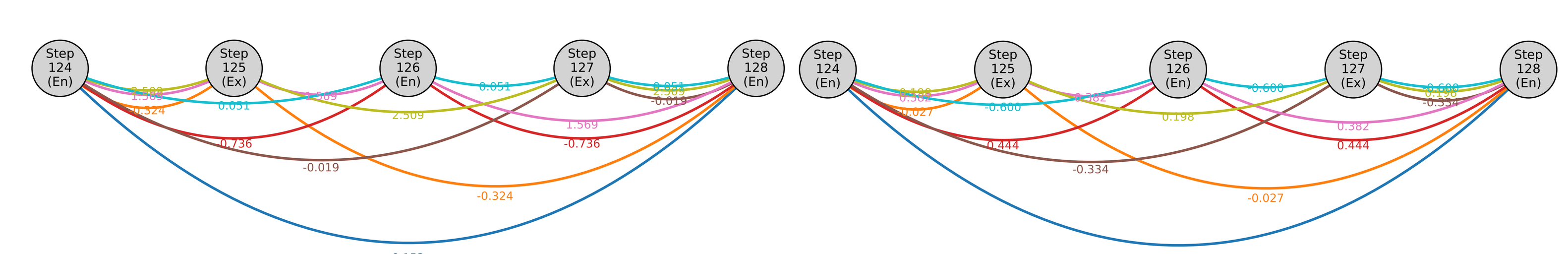

Figure 2. Visualization of the 2 temporal influence paths from step 124 to step 128 for the two series in the datapoint shown in Fig. 3, where even-numbered steps represent endogenous tokens and odd-numbered steps represent exogenous tokens.

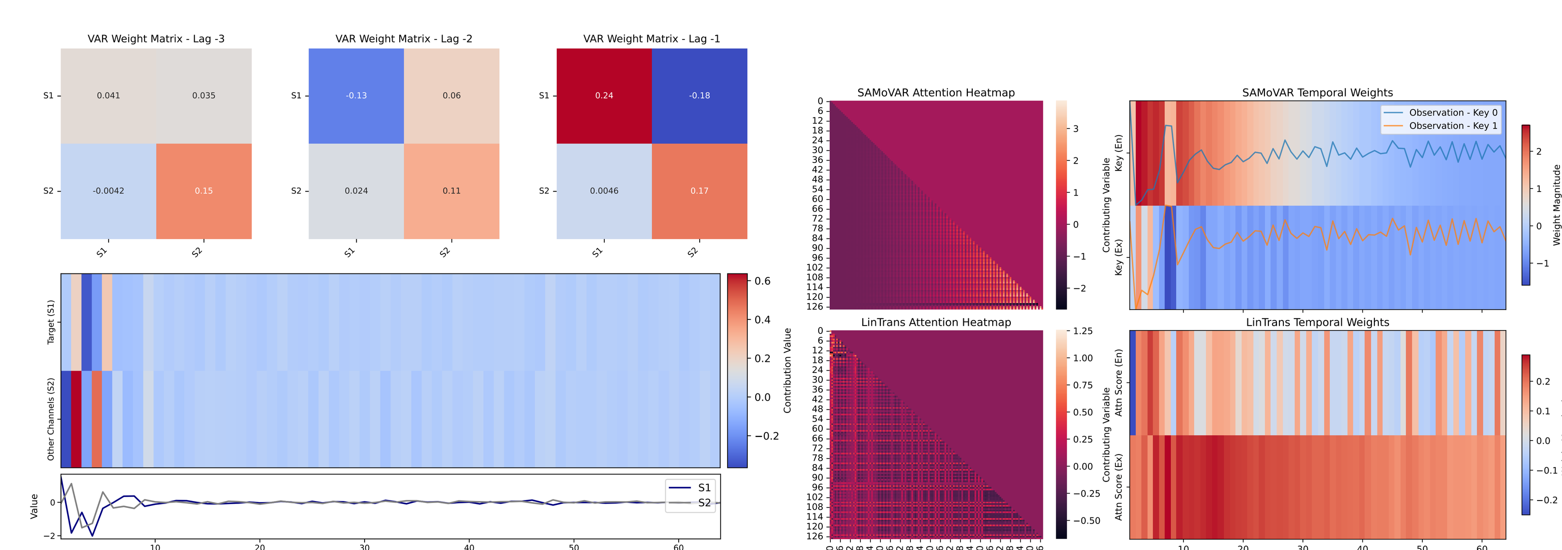

Figure 3. Visualization of the validation datapoint and model weights for the synthetic VAR task.

Experiments

- **Benchmarks**: 12 real-world multivariate datasets (ECL, Traffic, Weather, Solar, ETT, PEMS...).
- **Results**: SAMoVAR achieves best average rank (1.41) and most wins (29/33).

Table 1. Summary of Multivariate TSF Results. Averaged test set MSE are reported.

Model	SAMoVAR	LinTrans	FixedVAR	CATS	iTransformer	FITS	PatchTST	Dlinear	EncFormer
Weather	**0.214**	0.217	0.247	0.216	0.232	0.222	0.221	0.233	0.251
Solar	**0.184**	0.189	0.430	0.206	0.219	0.209	0.202	0.216	0.212
ETTh1	**0.401**	0.419	0.564	0.408	0.454	0.440	0.413	0.422	0.906
ETTh2	0.324	0.346	0.391	**0.320**	0.374	0.354	0.330	0.426	0.877
ETTm1	**0.339**	0.346	0.519	0.345	0.373	0.354	0.346	0.347	0.735
ETTm2	**0.240**	0.243	0.278	0.243	0.265	0.247	0.247	0.252	0.576
ECL	**0.151**	0.166	0.345	**0.151**	0.170	0.167	0.159	0.165	0.664
Traffic	0.391	0.438	0.717	**0.385**	0.414	0.418	0.391	0.431	0.824
PEMS03	**0.150**	0.188	0.375	0.225	0.212	0.234	0.230	0.254	0.443
PEMS04	**0.102**	0.136	0.404	0.184	0.171	0.256	0.222	0.246	0.377
PEMS08	**0.234**	0.261	0.674	0.359	0.271	0.296	0.290	0.357	0.681
AvgRank	**1.41**	3.41	8.16	2.86	5.43	5.20	4.00	5.82	8.30
#Top1	**29**	3	0	9	1	0	2	1	0

Synthetic VAR test: SAMoVAR recovers true lag weights & extrapolates beyond training lag order.

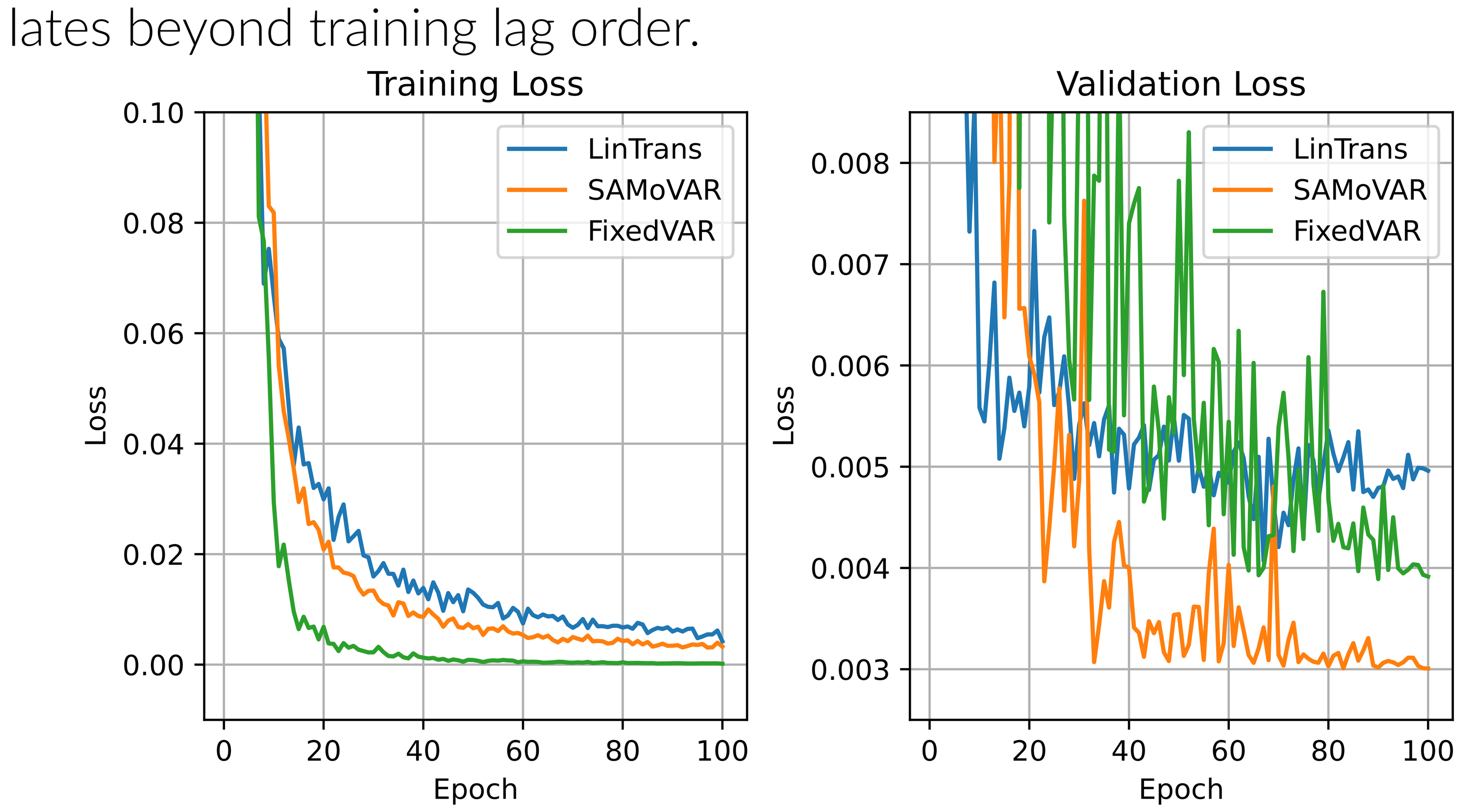

Figure 4. Visualization of the loss curves for synthetic VAR tasks.

Conclusion, Future Work

- SAMoVAR bridges classical VAR interpretability with deep attention power.
- Delivers SOTA accuracy, transparent lag coefficients, linear complexity.
- Future: scale to foundation TSF models; apply dynamic-VAR alignment to other sequence tasks.

jlu414@gatech.edu | shihao.yang@isye.gatech.edu