

Discussion of the paper “From Univariate to Universal Forecasting” by Pablo Guerron et al.

Francesco Ravazzolo

BI Norwegian Business School and University of Bozen

March 24, 2026

- There is a significant **lack of high-quality, real-world pretraining data** that includes both multivariate dependencies and informative covariates. Big issue for multivariate models.
- To overcome this, **Chronos-2** applies **group attention** so it can share information within groups across the batch axis, allowing it to scale gracefully with the number of variates.
- Moreover, **Chronos-2** relies on **synthetic data** to enable its multivariate and covariate-informed capabilities.
- While effective, the model's complex reasoning is only as good as the diversity of the "fake" structures it was shown during training.

- The model cannot yet incorporate **multimodal inputs**, such as text (e.g., news headlines affecting stock prices) or images, which are often critical for high-stakes real-world forecasting. Simple solution build an index based on text and use it. More challenging solution extend the model structure.
- On strictly multivariate tasks (multiple targets but no external covariates), the **performance gains** from using "in-context learning" (ICL) were **relatively modest**. Strong univariate models can often capture the same dynamics as multivariate ones, making the added complexity of multivariate modeling less impactful in some scenarios.
- The model uses **different encoding strategies depending on the task**. For univariate targets, it uses target encoding, but for multivariate targets, it must fall back to ordinal encoding (assigning unique integers to categories), which may be less effective at capturing complex relationships.
- The model is restricted to handling **numeric and categorical data**.

- While efficient enough for mid-range GPUs, the addition of the group attention layer and alternating transformer blocks increases the **architectural complexity** compared to simple univariate channel-independent models.
- Because there is **no natural ordering** for different time series within a group, the group attention layer must omit positional embeddings, potentially missing nuances in how certain variables relate to others spatially or logically.
- Capturing **long-term seasonality** requires a two-stage training process where the context length is extended from 2,048 to 8,192 steps, adding computational overhead to the pretraining phase. Possibility to find a trade-off.