Datai 2026
5th Scientific Conference

Forecasting in the Era of Foundation Models: Model Review and Usage Guidelines

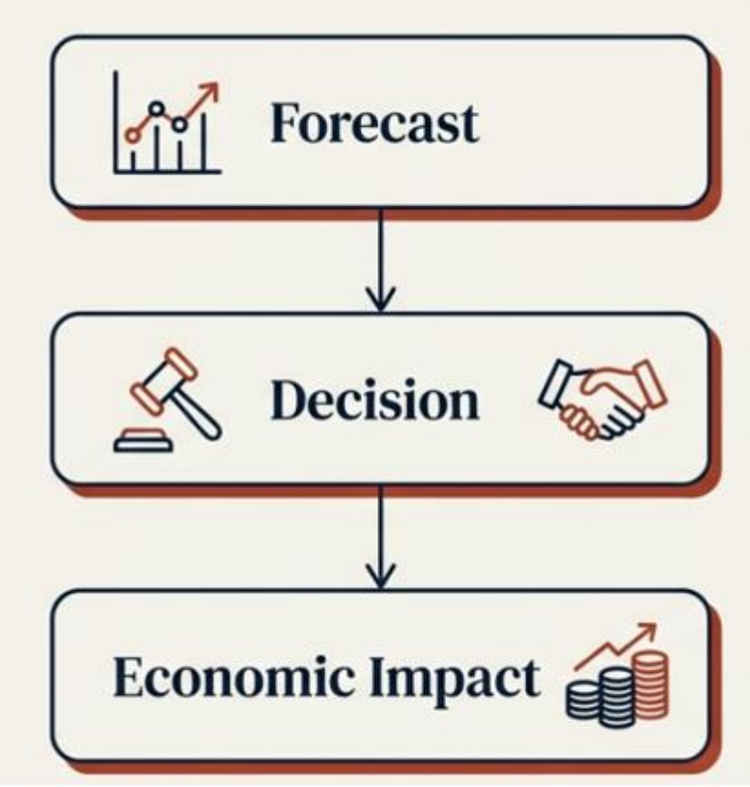
Pedro Gómez

Department of Electrical and Electronic Engineering, Tecnun

Introduction

Tecnun Universidad de Navarra

Forecasting is the Engine of Economic Decision Making



Critical Utility: Forecasting drives capital allocation, inventory levels, labor hiring and financial asset pricing.

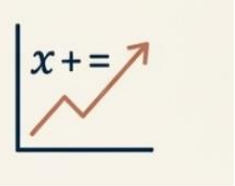
The Multiplier Effect: Small accuracy gains yield massive economic returns. Historically accuracy was capped at 50-60% due to computational limits.

The Financial Challenge: Markets are characterized by non-stationary, noisy and driven by complex exogenous factors.

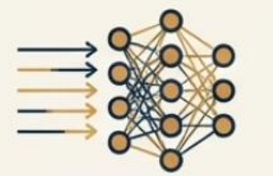
Motivation 1

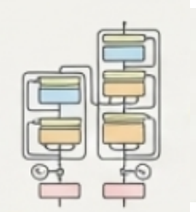
Tecnun Universidad de Navarra

A Century of Progress: The Evolution of Forecasting Tools:



Assumption: History repeats locally





Assumption: Universal temporal dynamics

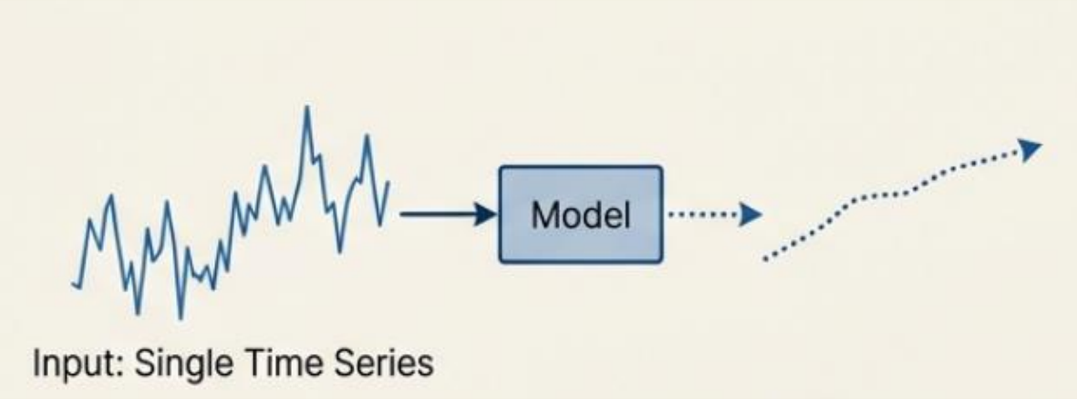
Common limitations of the Past: **Narrow Induction**: Learning solely from the history of a single time series.

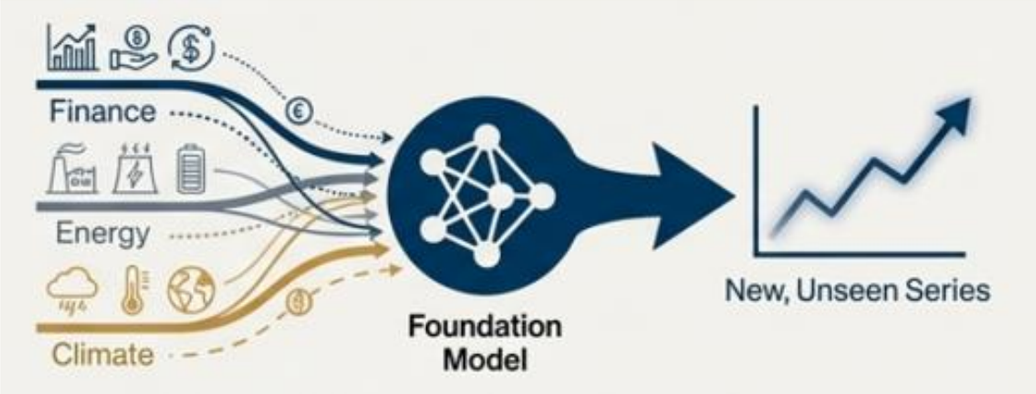
Motivation 2

Tecnun Universidad de Navarra

A Paradigm Shift: Towards Universal Forecasting

What if a single model could learn the fundamentals of time series data from millions of examples, much like LLMs learn human language?





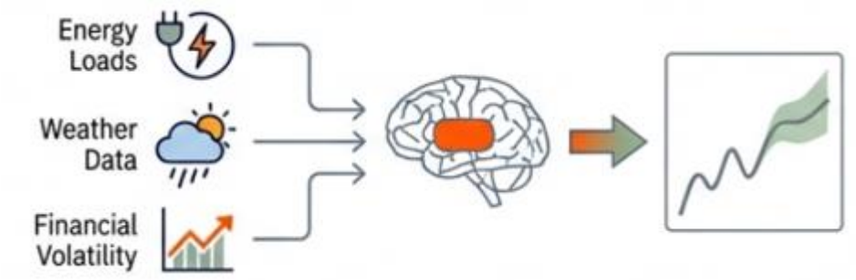
Old paradigm (narrow induction): Future behaviour is inferred from the historical observations of a time series. Assumes stationarity and is limited to a specific domain

A Foundation Model (broad induction) is pre-trained on a vast, heterogeneous collection of time series. It learns transferable temporal regularities –seasonal, trends, etc- that generalize across domains

Research Questions / Hypothesis

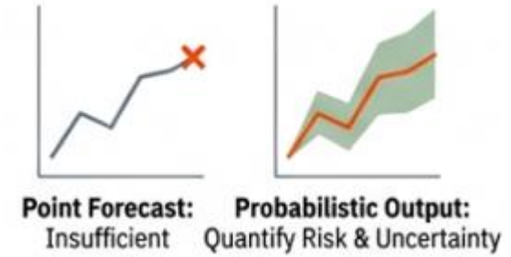
Tecnun Universidad de Navarra

01: Can TSFMs generalize across domains? A single pre-trained model can learn patterns from energy loads or weather data to effectively predict financial volatility.



02: Do they work in zero-shot settings? TSFMs can outperform specialized baselines without any fine-tuning on the target asset.

03: Do they provide reliable uncertainty? Point forecasts are insufficient for finance.



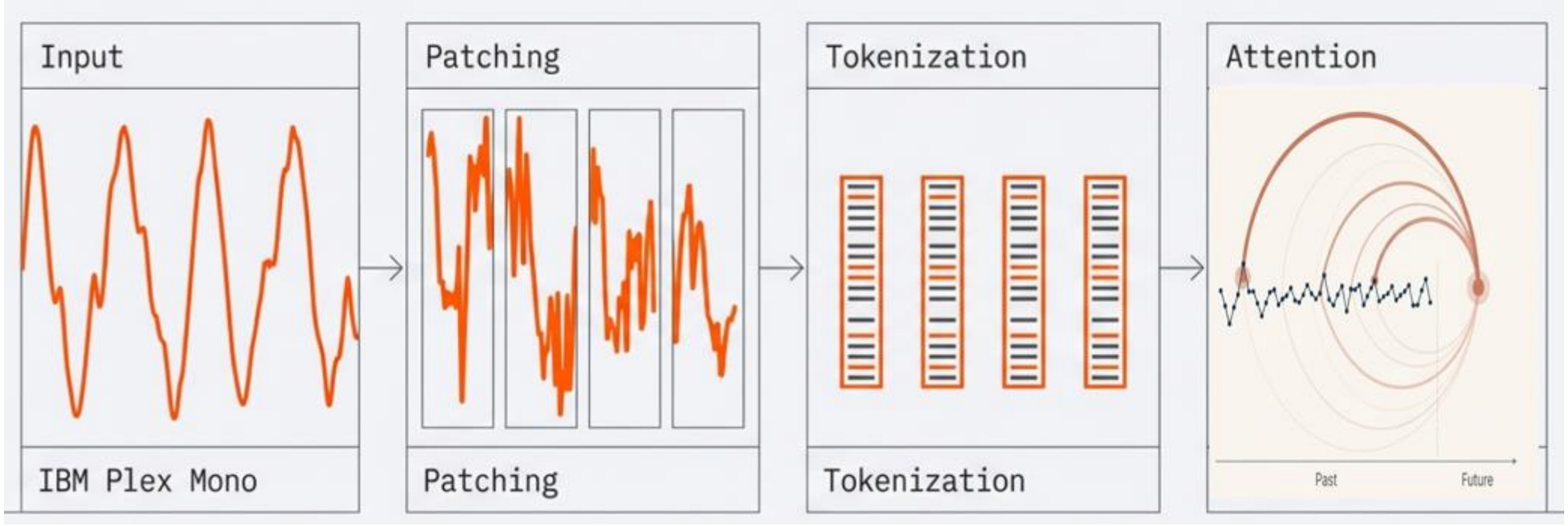
04: Is temporal data alone sufficient?

Methodology I: How TSFMs Work

Tecnun Universidad de Navarra

Inside the Engine: Tokenizing Time

Just as LLMs break sentences into words (tokens), TSFMs break time series into “patches.” This reduces computational complexity and allows the model to capture local semantic information, treating time segments like words in a sentence.



Methodology II: Architectural Trade-offs

Tecnun Universidad de Navarra

Encoder Only

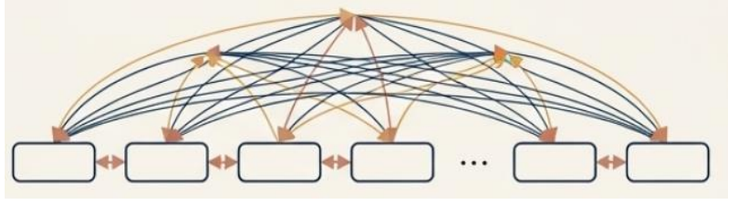
Mechanism: Processes the *entire input sequence at once* using **bi-directional attention**.
Training objective: Masked reconstruction of missing patches (only a subset of tokens is predicted).
Strengths:

- Strong representation learning
- Full-context understanding
- Naturally supports future-known covariates

Limitations:

- Inefficient pretraining (sparse loss signal)

Moirai
A 'universal forecaster' trained on 27 billion observations across 9 domains, capable of handling missing values and external variables.



Decoder Only

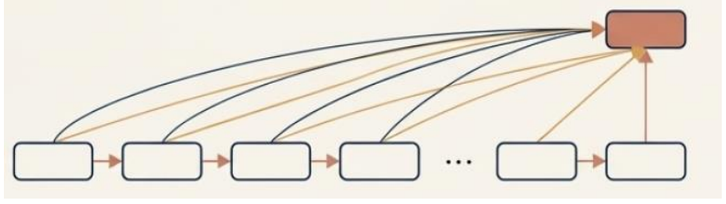
Mechanism: Autoregressive generation, predicting the next token/patch based on all previous ones.
Training objective: Next-token / next-patch prediction with **dense loss over the full sequence**.
Strengths:

- Highly efficient large-scale pretraining
- Well suited for long contexts and scaling

Limitations:

- Cannot natively condition on future-known inputs

TimesFM
A powerful decoder-only model pre-trained on 100 billion real-world time series points.



Encoder-Decoder

Mechanism: Maps an input sequence to a representation (encoder), then generates an output sequence (decoder).

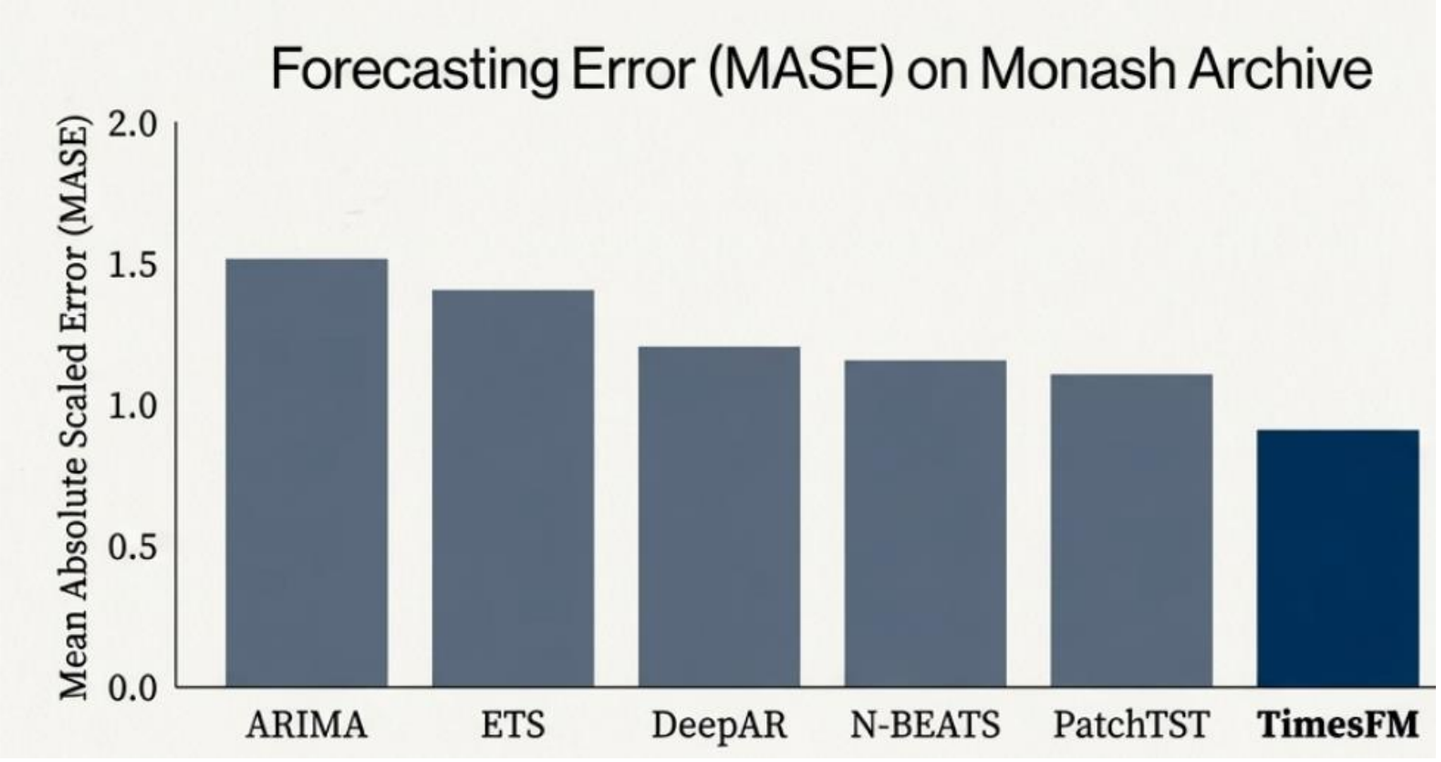
Limitations:

- Expensive pretraining as all context/horizon lengths must be presented

Moment (CMU)
T5-based (Encoder-Decoder); pre-trained via masked reconstruction for multiple tasks (forecasting, imputation, classification).

WHAT WORKS

On standardized benchmarks such as the Monash Forecasting Archive, TSFMs consistently outperform traditional statistical methods and earlier deep learning models.



Lower error scores indicate higher accuracy.

TSFMs *like TimesFM set a new performance standard.*

Strengths of TSFMs

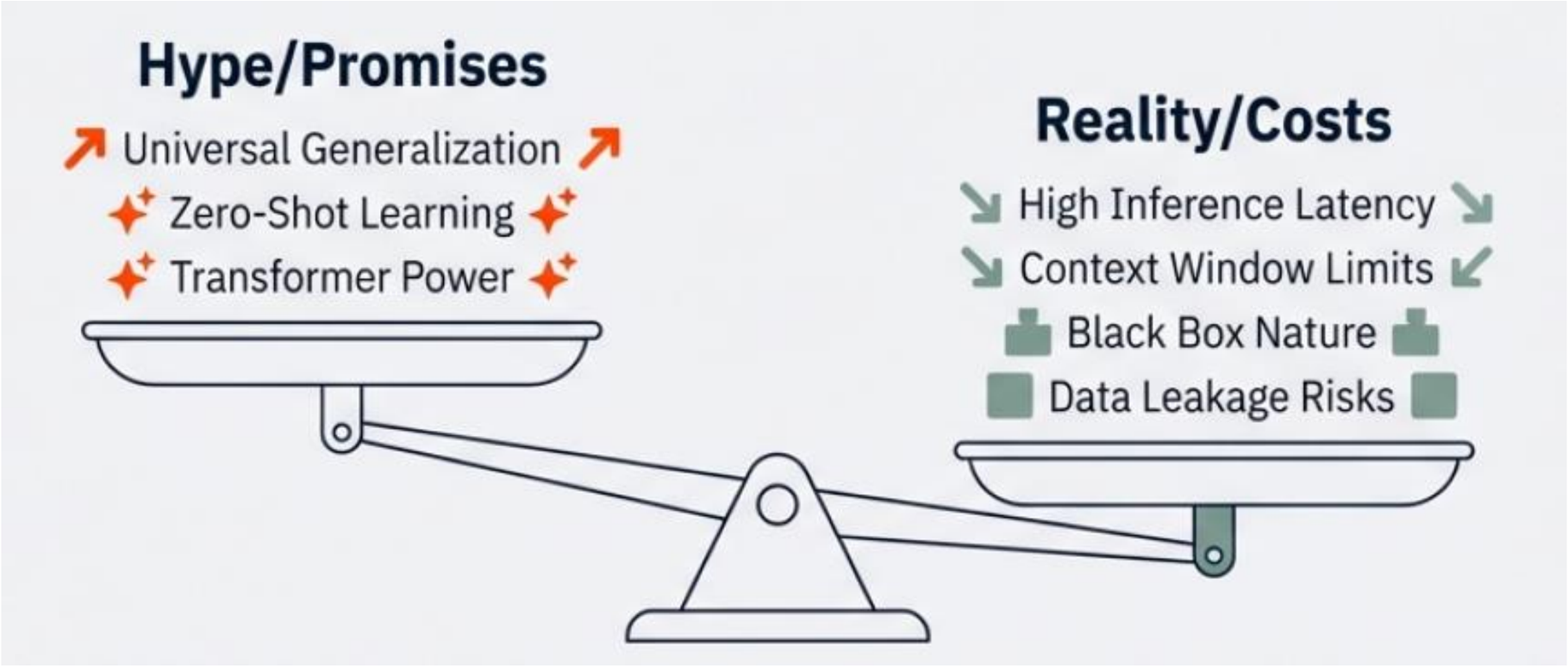
- Strong benchmark performance
- Zero-shot capability
- Cross-domain generalization

👉 **New baseline for forecasting**

Limitations

Tecnun Universidad de Navarra

Engineering Reality: The Skeptic's View



Limitations

Tecnun Universidad de Navarra

REALITY CHECK

Pretrained with a single loss (e.g., MSE, quantile loss)
Cannot be modified after training
☞ **Limited flexibility across tasks**
Possible mitigation:

- Increase context length
- Fine-tune on target dataset

Critical hyperparameter
☞ Performance highly depends on tuning

Favors certain frequencies over others
☞ Strong in structure and scale
☞ Weaker in very high-frequency, short-term tasks

No consistent dominance
☞ Performance varies across regimes
☞ Sensitive to distribution shifts outside training data

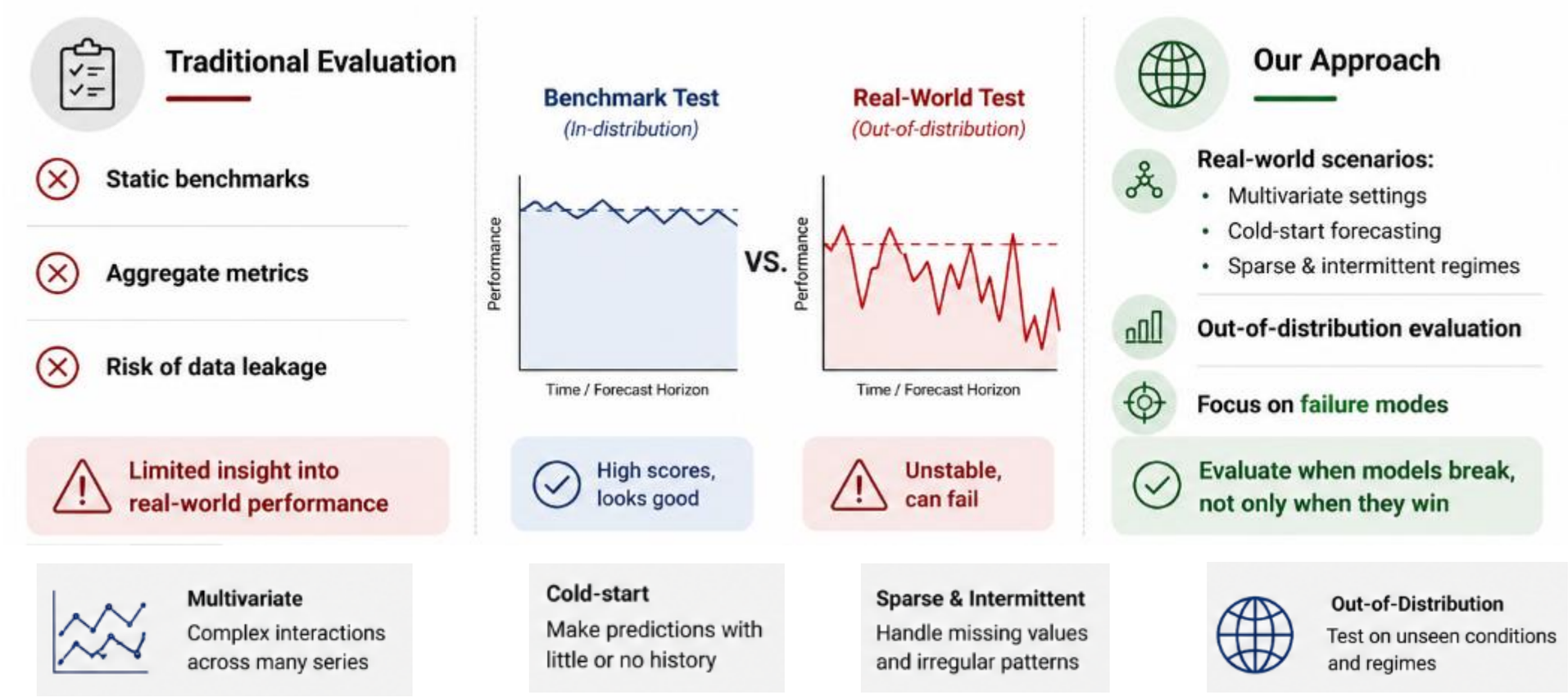
TSFMs are not plug-and-play models

Evaluation approach: Stress Testing

Tecnun Universidad de Navarra

Beyond Benchmarks: Stress Testing (zero-shot mode)

Our contribution is a stress-testing framework to evaluate TSFMs beyond standard benchmarks

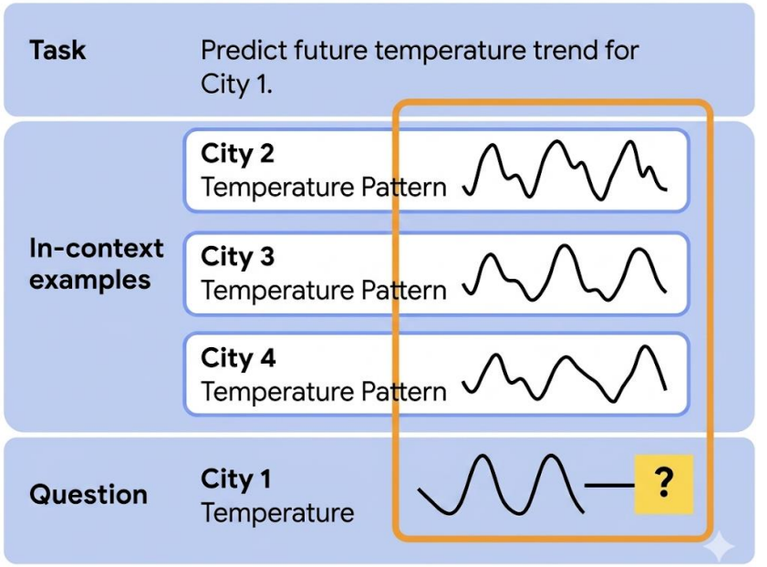


Multivariate Forecasting

Tecnun Universidad de Navarra

Multivariate TSFMs remain underexplored compared to univariate settings

Model	Univariate Forecasting	Multivariate Forecasting	Past-Only Covariates	Known Covariates	Categorical Covariates	Cross Learning	Memory Scaling
Chronos-2	✓	✓	✓	✓	✓	✓	$\mathcal{O}(V)$
Toto-1.0	✓	✓	✓	✗	✗	✗	$\mathcal{O}(V)$
TabPFN-TS	✓	✗	✗	✓	✓	✗	$\mathcal{O}(V)$
COSMIC	✓	✗	✓	✓	✗	✗	$\mathcal{O}(V^2)$
Moirai-1.0	✓	✓	✓	✓	✗	✗	$\mathcal{O}(V^2)$
Chronos-Bolt	✓	✗	✗	✗	✗	✗	-
Moirai-2.0	✓	✗	✗	✗	✗	✗	-
Sundial	✓	✗	✗	✗	✗	✗	-
TimesFM-2.5	✓	✗	✗	✗	✗	✗	-
TiRex	✓	✗	✗	✗	✗	✗	-



In context learning is another promising path.

Multivariate Forecasting

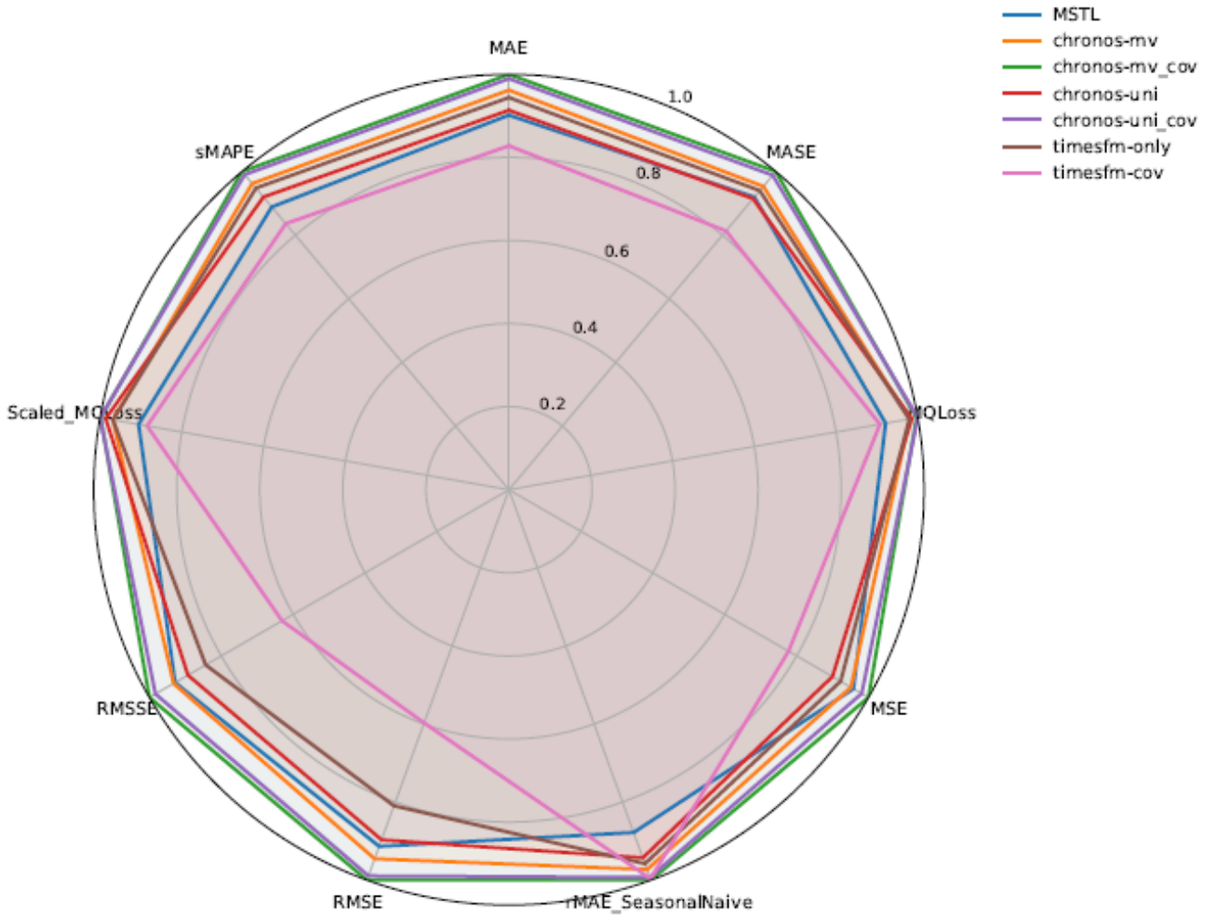
Tecnun Universidad de Navarra

Experimental Setup

Dataset: Kaggle Tabular Playground (Sep 2022)
48 related time series
(4 products × 2 stores × 6 countries)
Daily data (2017–2021)
Forecast horizon: 24 days
Models:
- Statistical baselines
- TSFMs: Chronos, TimesFM…

Evaluation:
- Rolling windows (×10)
- Multiple context lengths
- Calendar covariates

🔍 **Key Findings:**
- TSFMs improve with:
  - Longer context
  - Multivariate structure
  - External covariates



👉 **Performance gains come from context, not architecture**

Performance differences are small — context-aware models slightly dominate

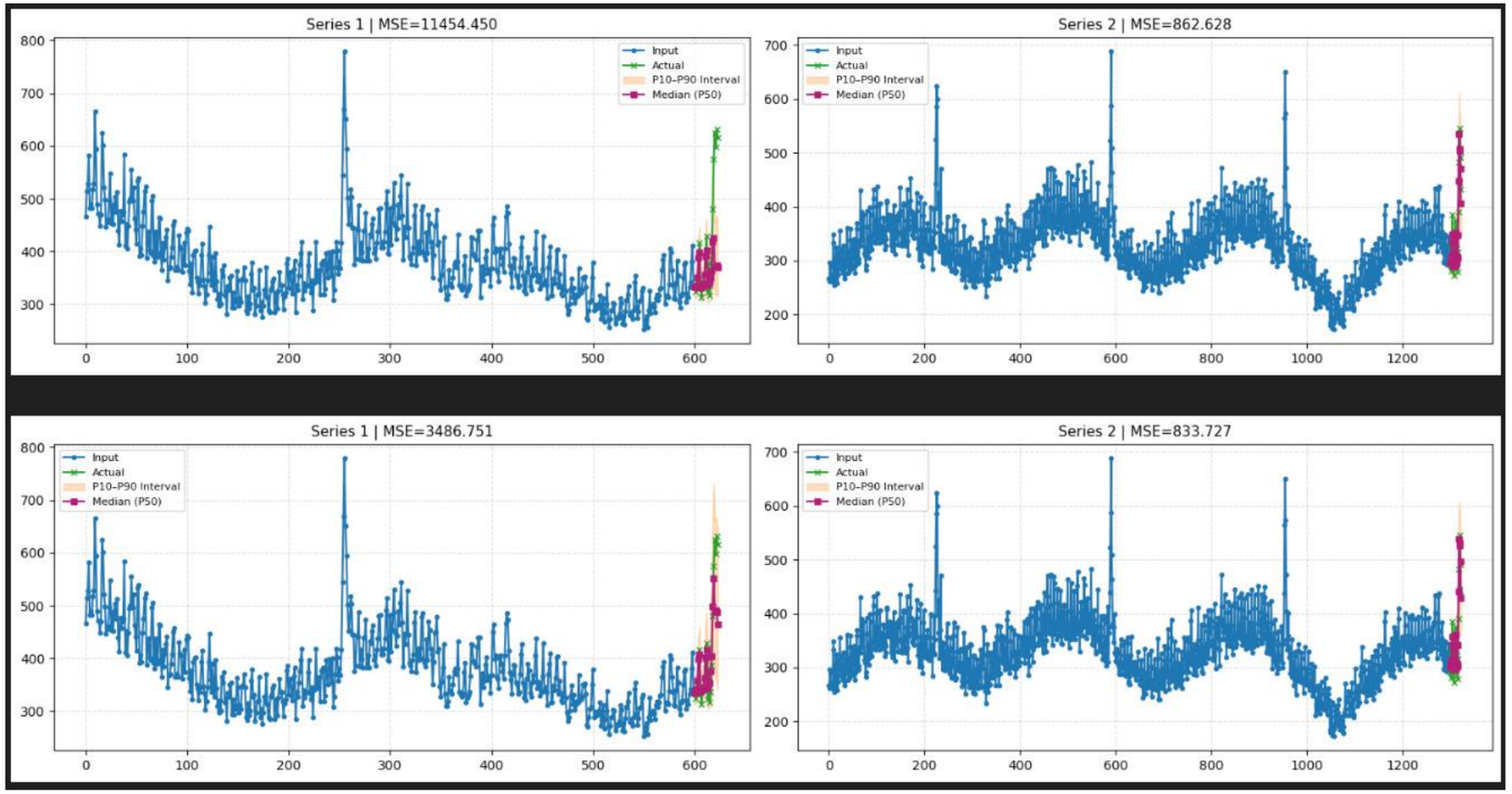
Cold Start Problem

Tecnun Universidad de Navarra

📊 Experimental Setup

Simulated new product scenario
Target series: Limited history (~600 observations)
Related series: Full history (~1300 observations)
Forecast horizon: 24 days
Models:
- Statistical baselines
- TSFMs: Chronos, TimesFM…

Evaluation:
- Point + probabilistic metrics
- Rolling Windows



🔍 **Key Findings:**
- Univariate models struggle:
  - → High error due to limited history
- Multivariate TSFMs improve significantly

→ Leverage information from related series

👉 **TSFMs transfer knowledge across series, Generalization replaces data availability**

Sparse Data

Tecnun Universidad de Navarra

📊 Experimental Setup

📉 Problem Setting

Event-driven time series
High sparsity (many zeros)
Irregular dynamics & heavy noise

Dataset: BOOM (observability data)
Highly sparse & intermittent signals
Models:
- Statistical baselines
- TSFMs: Chronos, TimesFM…

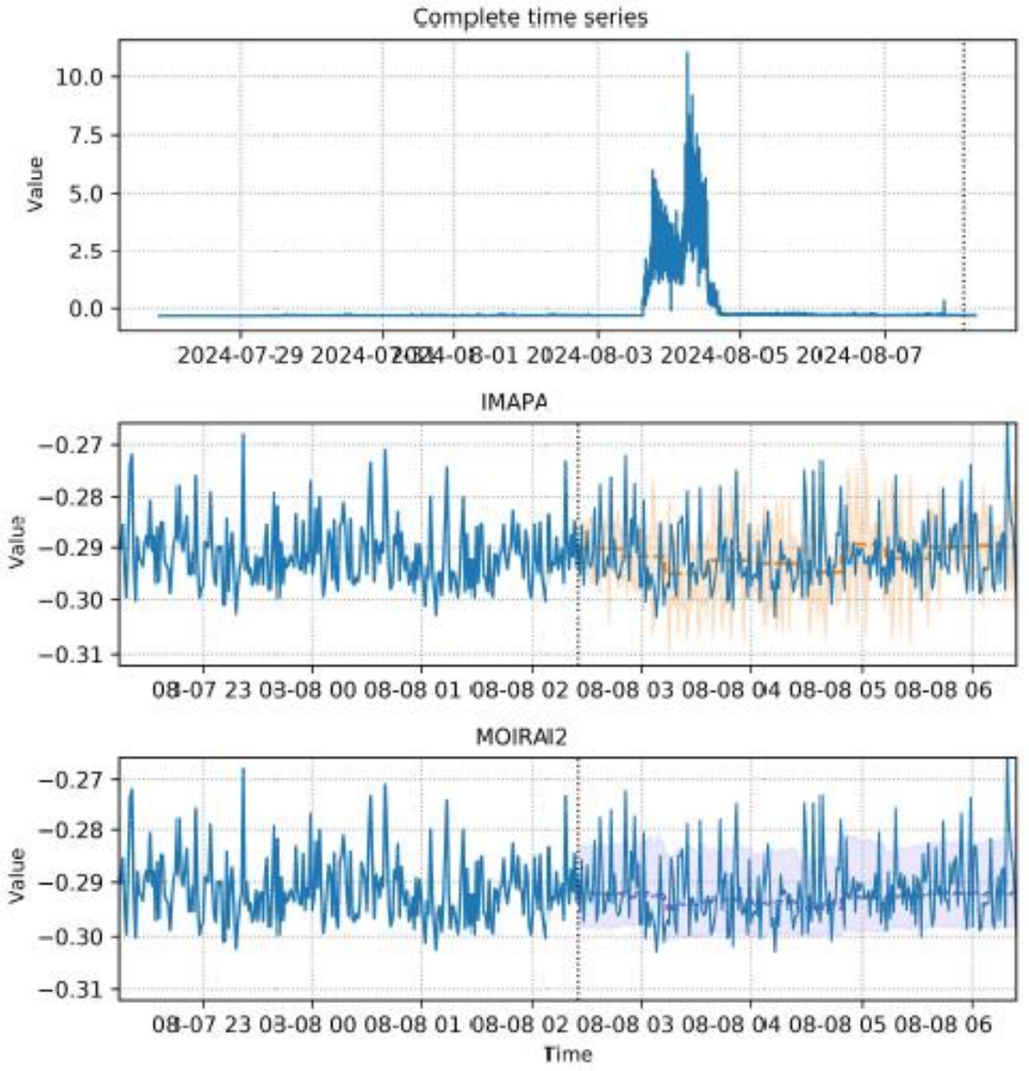
Evaluation:
- Rolling windows
- RMSE + probabilistic metrics

🔍 **Key Findings:**

Statistical models remain strong:

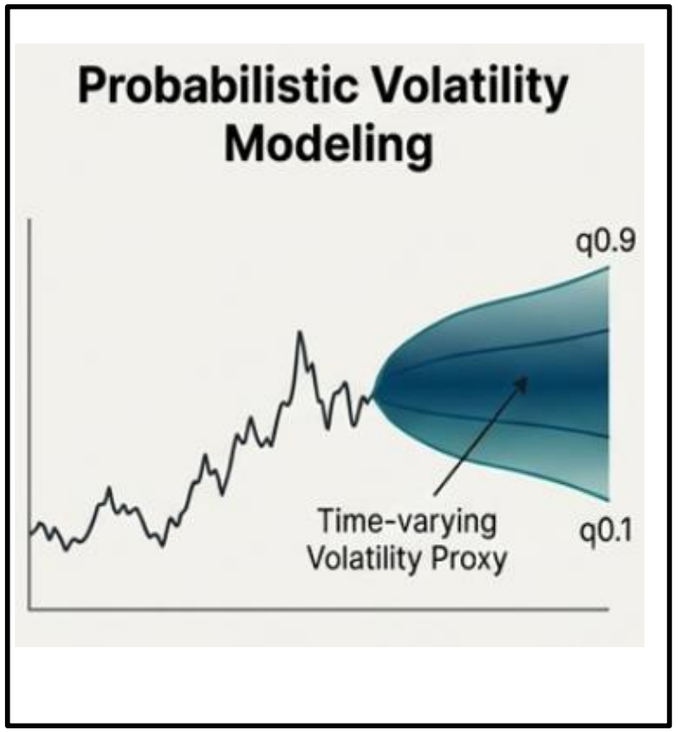
→ Especially in extreme sparsity

TSFMs provide better uncertainty estimates

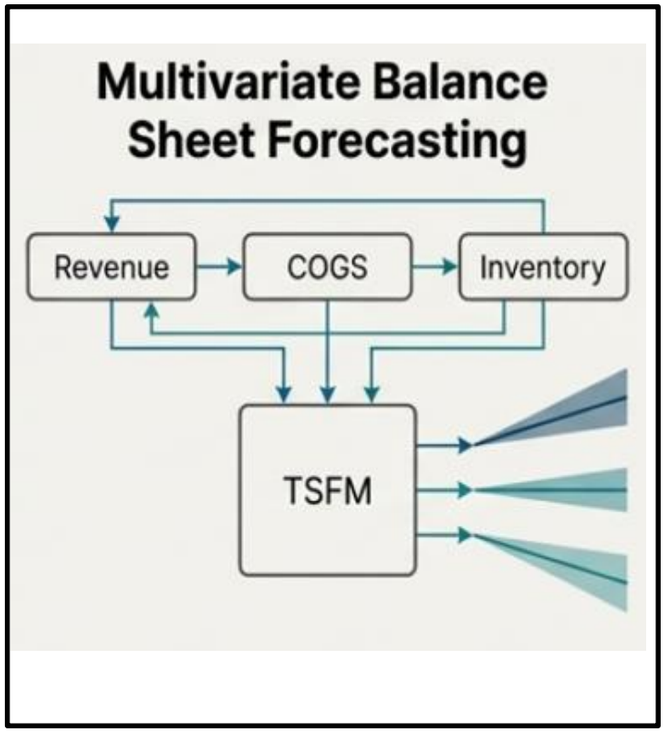


☞ **In sparse and event-driven regimes, TSFMs do not always outperform in accuracy, but they provide much better uncertainty estimates**

Financial Applications



Concept: TSFMs output predictive quantiles (e.g., q0.1, q0.9). The inter-quantile range (q0.9 − q0.1) serves as a time-varying proxy for market volatility.
Application: Dynamic risk management, Value-at-Risk (VaR) estimation, options pricing.



Concept: Jointly model interdependent accounts (revenue, COGS, inventory) using cross-series attention.
Application: Generate more coherent and accurate financial projections than running independent univariate models.

Concept: For new assets or IPOs with limited history, zero-shot TSFMs transfer patterns learned from thousands of other assets.
Application: Provide a reasonable baseline forecast where traditional models, which require long histories, would fail.

Practical Guide: When to Use TSFMs

Tecnun Universidad de Navarra

Use TSFMs when:

Complex data structure
→ multivariate + covariates
Limited or sparse history
→ cold-start scenarios
Cross-series information matters
→ shared patterns across datasets

Use Classical Models when:

Single stable time series
Large, clean datasets
Low complexity / high signal-to-noise

Practical Considerations:
Context length is data dependent
Compute & latency constraints matter
Fine-tuning can improve performance
Avoid data leakage (strict temporal splits)

Main Takeaway:

Model selection is context-dependent

No single model is optimal across all regimes

☞ **Choose the model based on the problem, not the benchmark**

Conclusions & Future Work

Tecnun Universidad de Navarra

From Fitting Models to Learning Dynamics

We are moving away from building bespoke models for individual problems toward **learning universal temporal dynamics** that can adapt across domains.

Universal Prior: While not a silver bullet, Foundation Models establish a **new baseline for financial intelligence**, enabling:

- **Zero-shot prediction** on unseen assets
- **Risk quantification at scale** through probabilistic forecasting
- **Scalable and transferable forecasting systems**

Conclusions & Future Work

Tecnun Universidad de Navarra

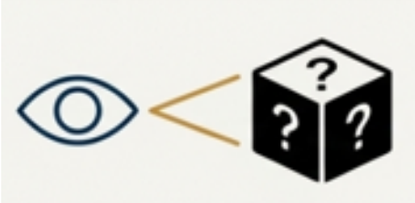
The Road Ahead: Open Challenges and New Frontiers

While powerful, TSFMs represent significant challenges –opportunities to explore.

Computational Efficiency

Challenge: Self-attention scales quadratically ($O(n^2)$) with sequence length, creating a bottleneck for very high-frequency data.
Future Work: Exploring linear-time architectures like **Mamba** (State Space Models) and sparse networks such as **Mixture-of-Experts (MoE)**.

Explainability



Challenge: The “black box” nature of large models can be a barrier in high-stakes financial applications where explainability is critical.
Future Work: Applying perturbation-based methods (SHAP, LIME) and developing inherently explainable architectures

Multimodal Forecasting

Challenge: Financial markets are influenced by more than just price history (e.g., news text).
Future Work: Integrating diverse data types (tabular, textual, visual) into a single, unified forecasting system.

Closing

Tecnun Universidad de Navarra

Final Takeaways

TSFMs are a major shift

From task-specific models to general-purpose forecasters.

Strong in complex scenarios

They excel with multiple series, limited history, and rich contexts.

Not universally optimal

Performance depends on the data regime and the problem.

→ **Best performance comes from:**

Combining models

Leverage the strengths of different approaches.

Using context

Incorporate exogenous information and domain knowledge.

Understanding the data

Tailor models to the structure, regime, and limitations of the data.

Closing

Tecnun Universidad de Navarra

Takeaway: Forecasting is no longer about one model per series or domain, but about reusable temporal intelligence.

Acknowledgements: Juan Carlos Martínez Ovando (BBVA) and Idoia Ochoa (Tecnun), PhD co-supervisors.
Alfonso Churruca (BBVA), industry supervisor.