Chronos-2

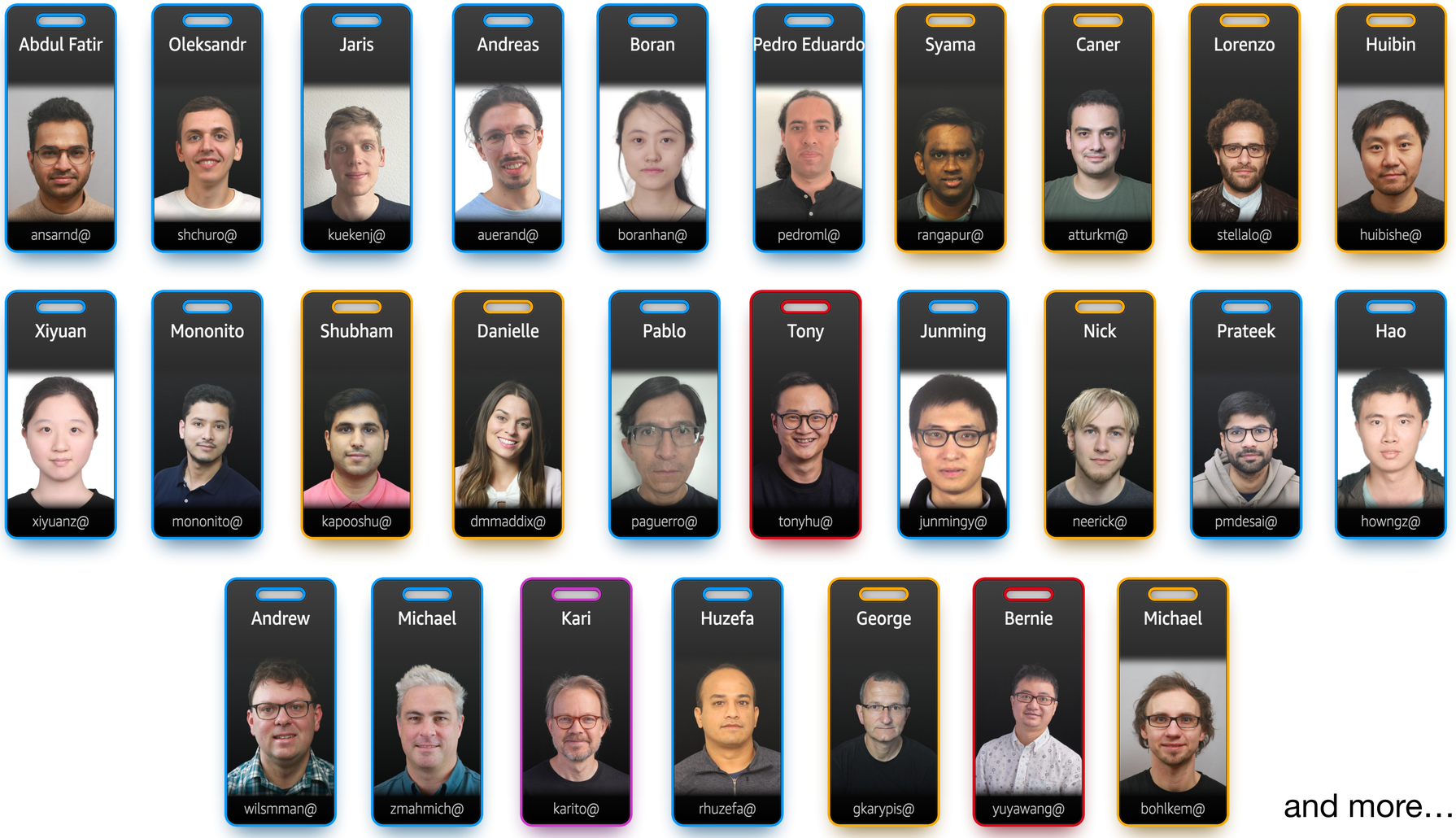
From Univariate to Universal Forecasting

Pablo A Guerron
Boston College
Amazon Scholar

DS3 DEEP SCIENCE FOR SYSTEMS AND SERVICES

aws

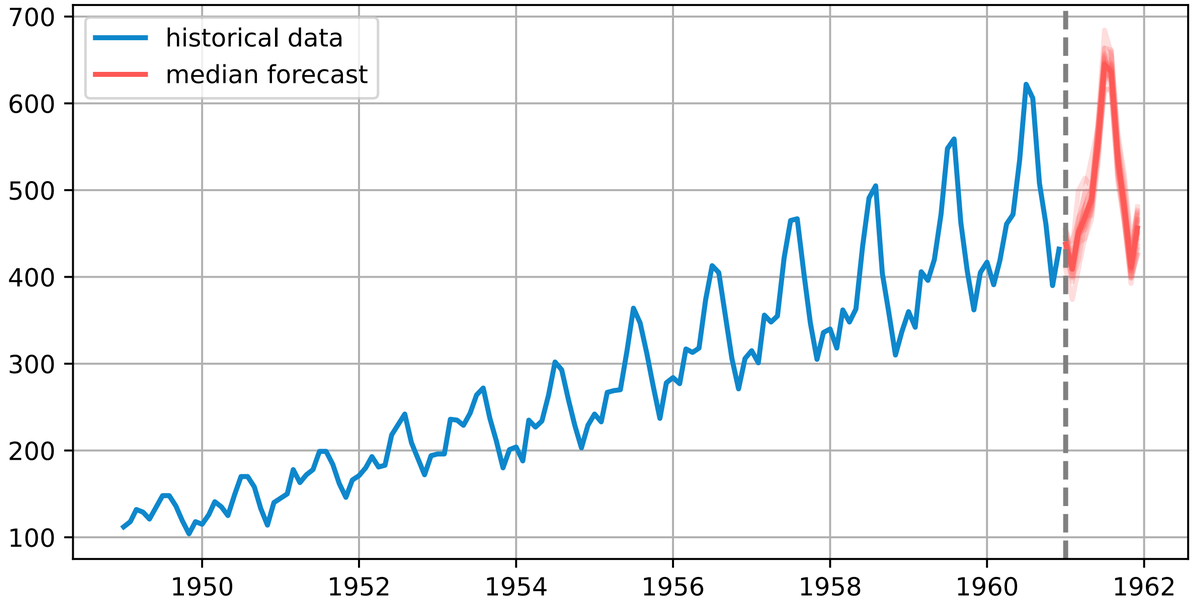
The Chronos Team



and more…

Time Series Forecasting

Predict the future behavior of a time series given its past



Monthly US Airline Passengers from 1949 to 1960

Energy

Finance

Weather

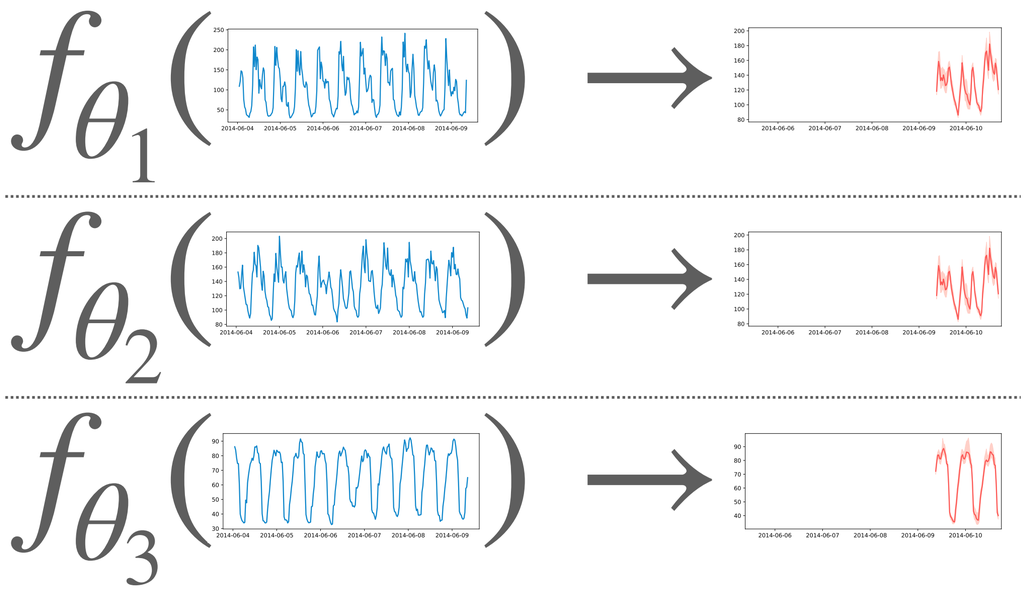
Retail

Healthcare

Traffic

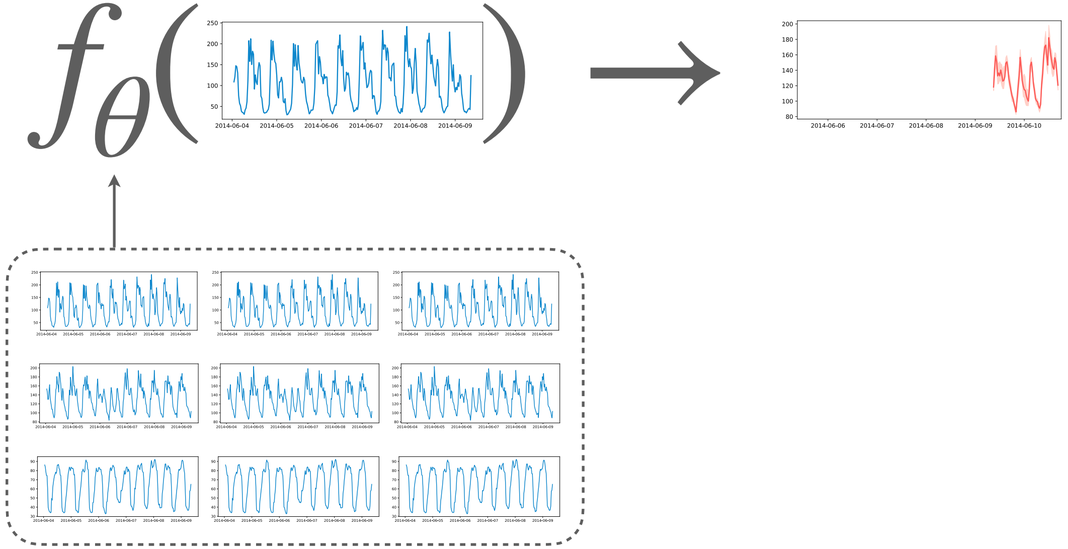
$$p(x_{T+1}, x_{T+2}, \ldots, x_{T+h} \mid x_1, x_2, \ldots, x_T)$$

Local vs. Global Models



Fit a **separate** model for each **individual** time series

(e.g., ETS, ARIMA, Prophet)



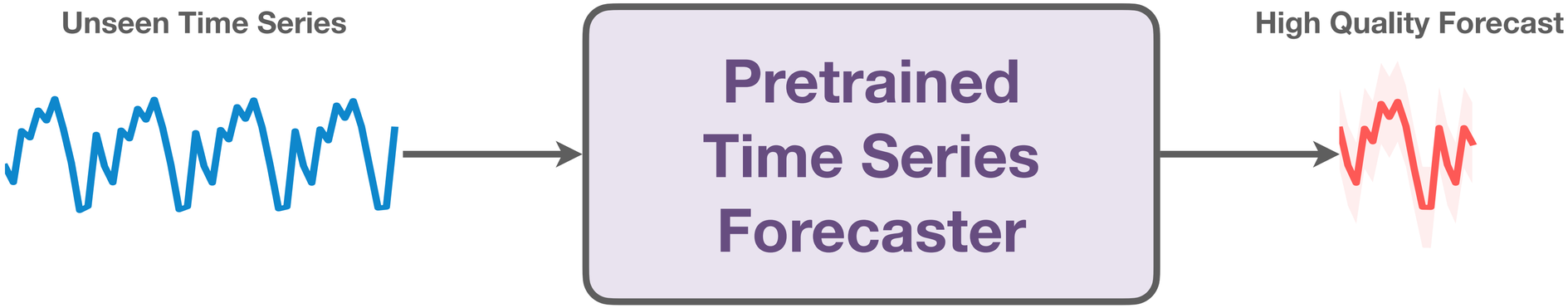
Fit a **single** model for each **dataset** or task

(e.g., DeepAR, TFT)

Time Series Foundation Models

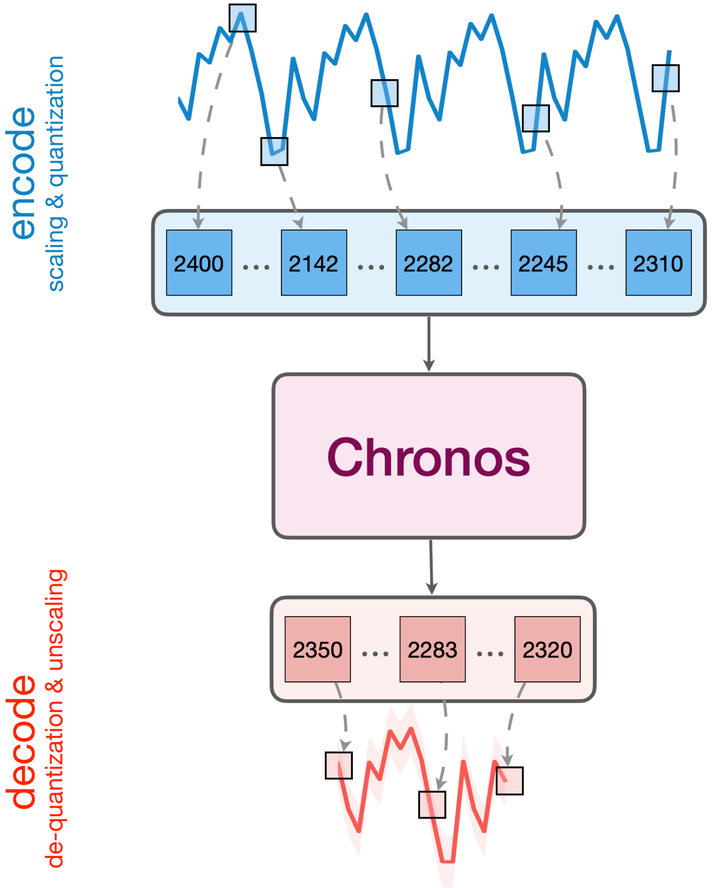
Time Series tasks have diverse **patterns**, **frequencies**, **history lengths**, **prediction horizons**, **missing values**, …

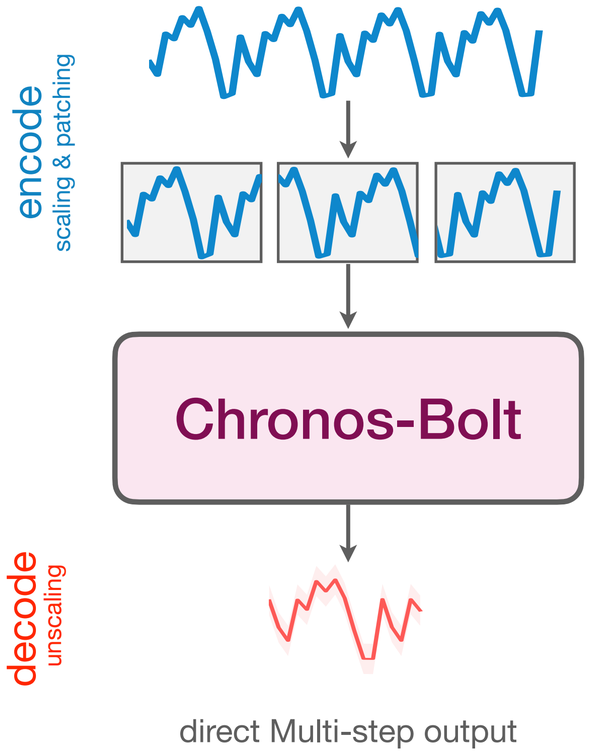
→ *long trial-and-error development cycles*



trained on large-scale data and perform well on unseen time series datasets

Chronos & Chronos-Bolt





CHRONOS

Customer

Cool! …but how can I adapt it further to *my problem*?

*my problem **requires** **covariates** (exogenous inputs)*

Electrical Load

Weather
(Temperature, Humidity, Pressure)

Energy Price

Electrical Load

Generation

Product Sales

Promotions

Holidays

Economy

CHRONOS

Customer

Cool! …but how can I adapt it further to *my problem*?

*my time series are **multivariate** (shared dynamics)*

Cloud Infra Metrics

CPU Util

Disk I/O

Network

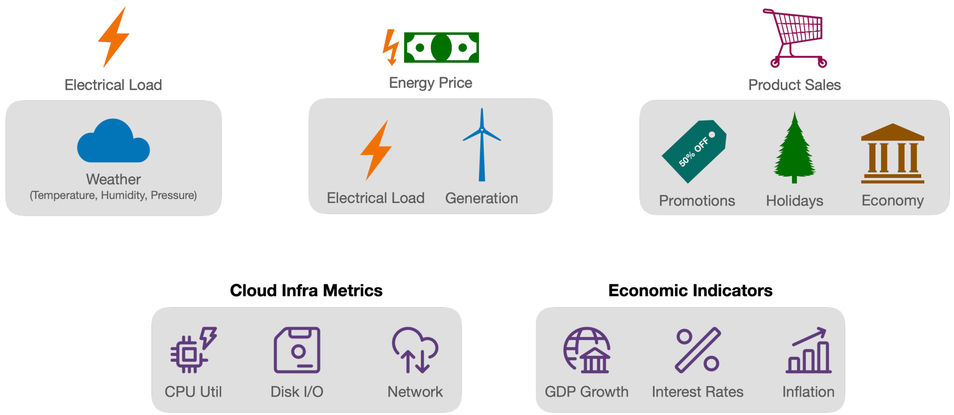
Economic Indicators

GDP Growth

Interest Rates

Inflation

The Challenge



the space of forecasting problems

the number and types of (co)variates, and their interactions,
are **not known** a priori

Introducing Chronos-2

A **unified** model supporting…

- exogenous variables

  infers how target time series depend on external factors
- multivariate forecasting

  infers dependencies across different related series
- cross learning

  infers better using multiple items in the batch

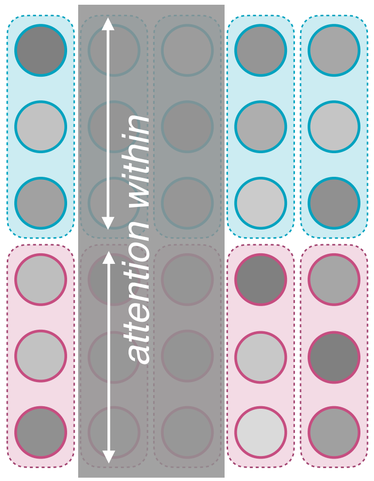
…via in-context learning (ICL).

Chronos-2 vs. Prior Models

	Chronos-2	Chronos-Bolt	Chronos
Input Type	Patches	Patches	Individual Observations
Output Type	Multi-step Quantile Forecast	Multi-step Quantile Forecast	Autoregressive Sampling
Loss Function	Quantile Loss	Quantile Loss	Cross Entropy Loss
Univariate Forecasting	✅	✅	✅
Cross Learning	✅	❌	❌
Multivariate Forecasting	✅	❌	❌
Past-only Covariates	✅	❌	❌
Known-future Covariates	✅	❌	❌
Max. Context Length	8192	2048	512
Max. Prediction Length	1024+	64+	64+
Num. Parameters	120M	9M to 205M (4 sizes)	8M to 710M (5 sizes)

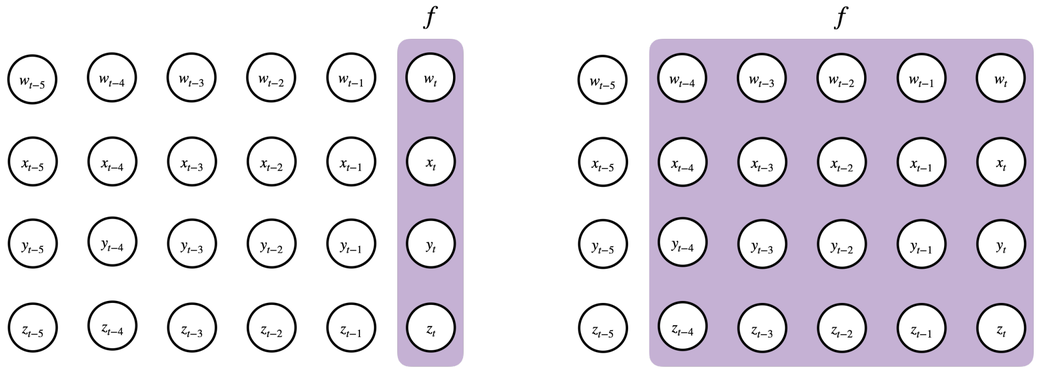
How to develop a universal model?

Model must understand interactions from the historical context



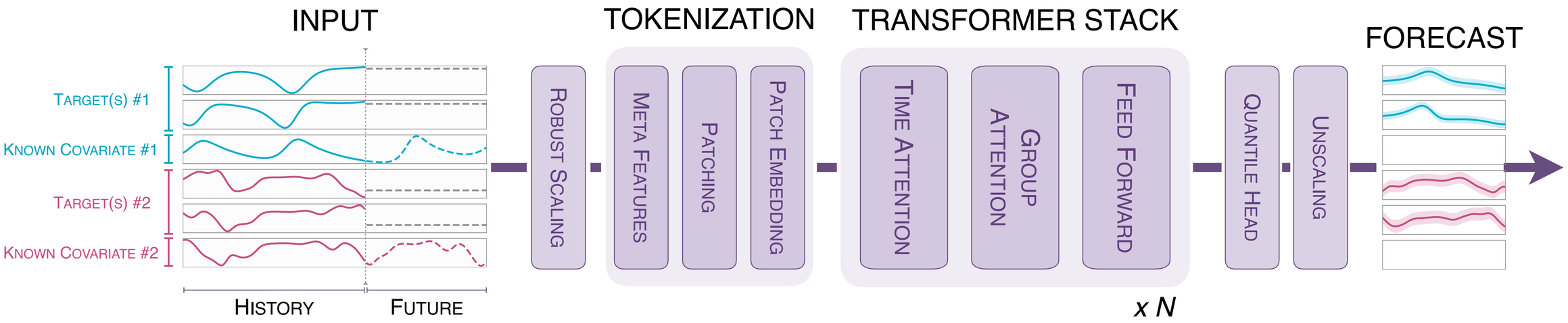
Group Attention

High-quality training **data** with multivariate dependencies



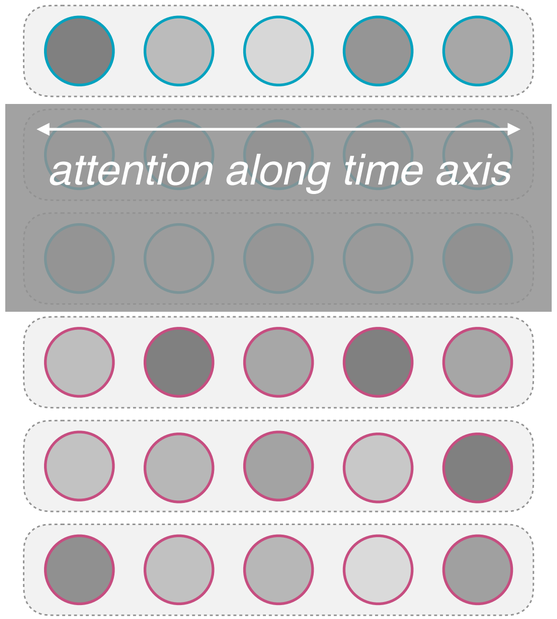
Synthetic Data

Model Architecture

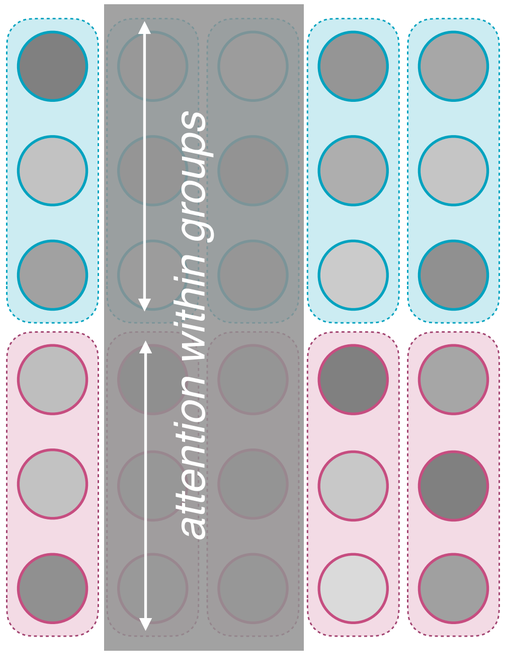


Transformer Stack

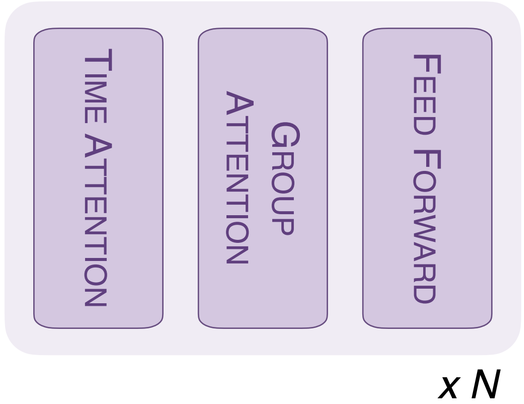
TIME ATTENTION



GROUP ATTENTION

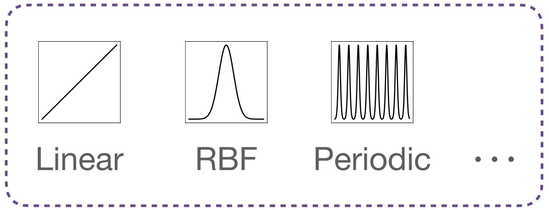


TRANSFORMER STACK



Univariate Generators

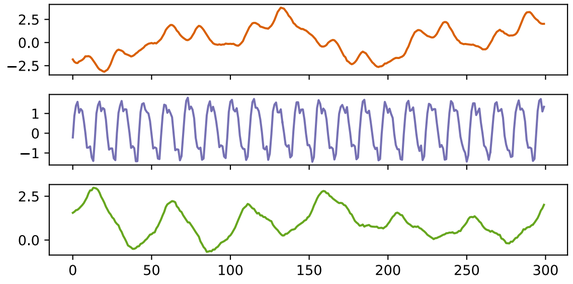
KernelSynth



Gaussian Process Kernel Bank

sample

Combine Kernels

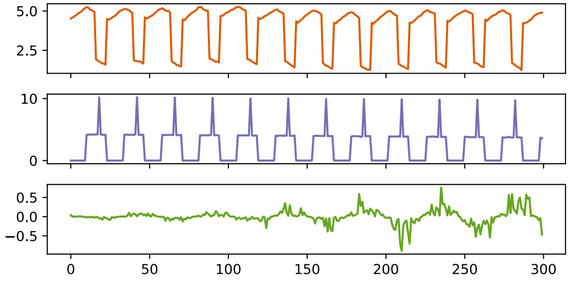


Trend, Seasonality and Irregularity (TSI)

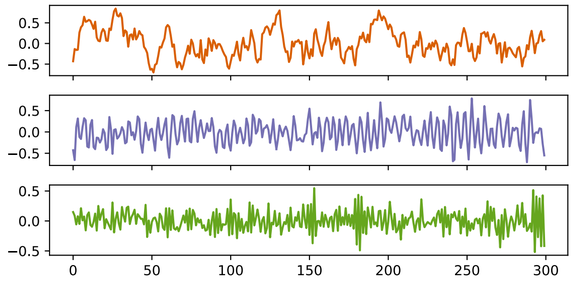
Seasonalities **Trends** **Irregularities**

sample sample sample

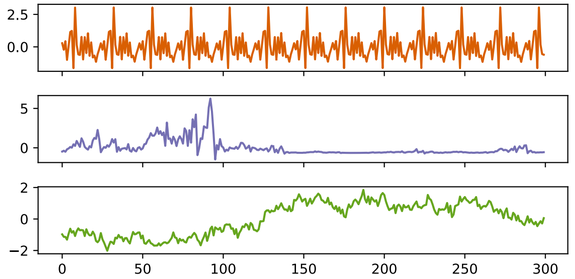
Combine Components and Post-process



AutoReg (AR)

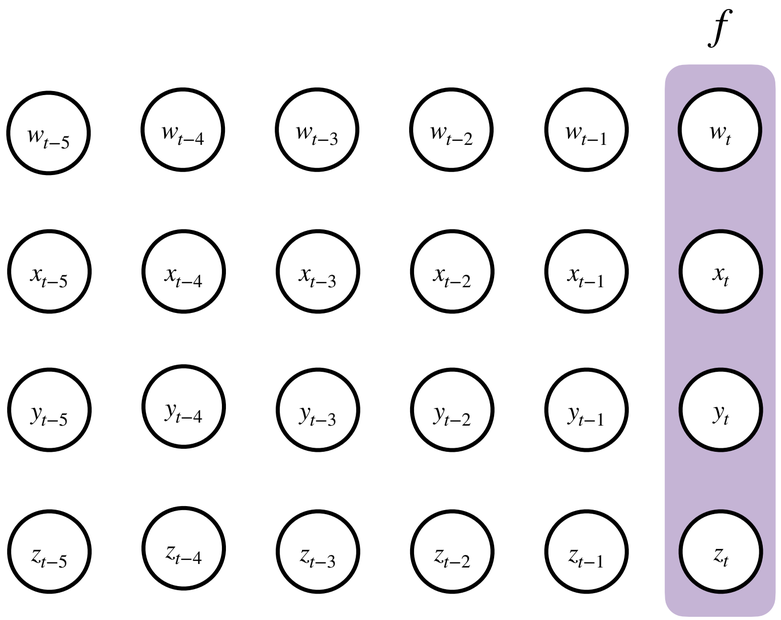


ExponenTial Smoothing

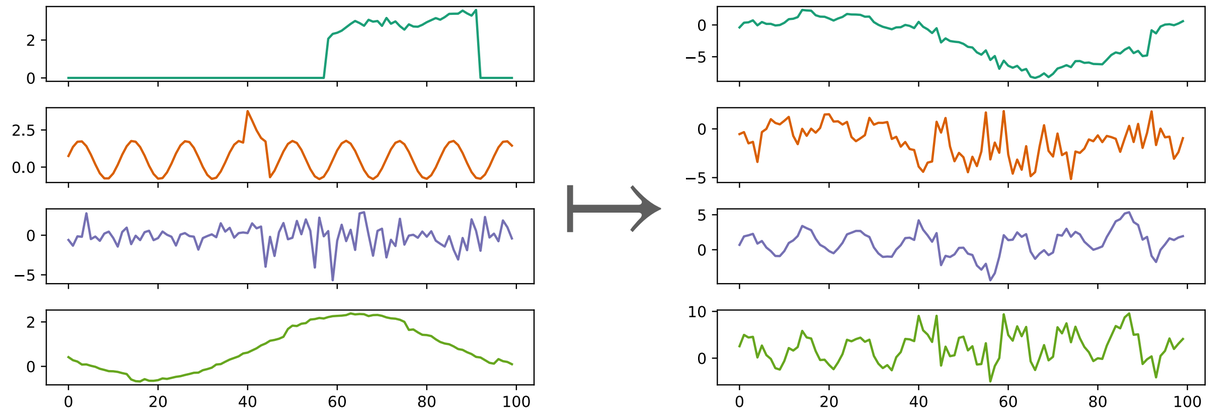


and more…

Cotemporaneous Multivariatizers

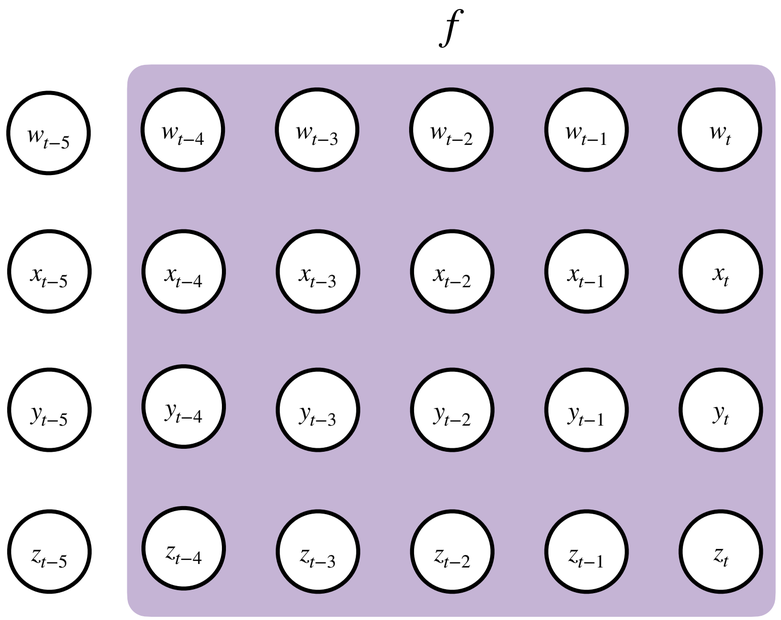


$$f: \begin{bmatrix} w_t \\ x_t \\ y_t \\ z_t \end{bmatrix} \mapsto \begin{bmatrix} \tilde{w}_t \\ \tilde{x}_t \\ \tilde{y}_t \\ \tilde{z}_t \\ \vdots \end{bmatrix}$$

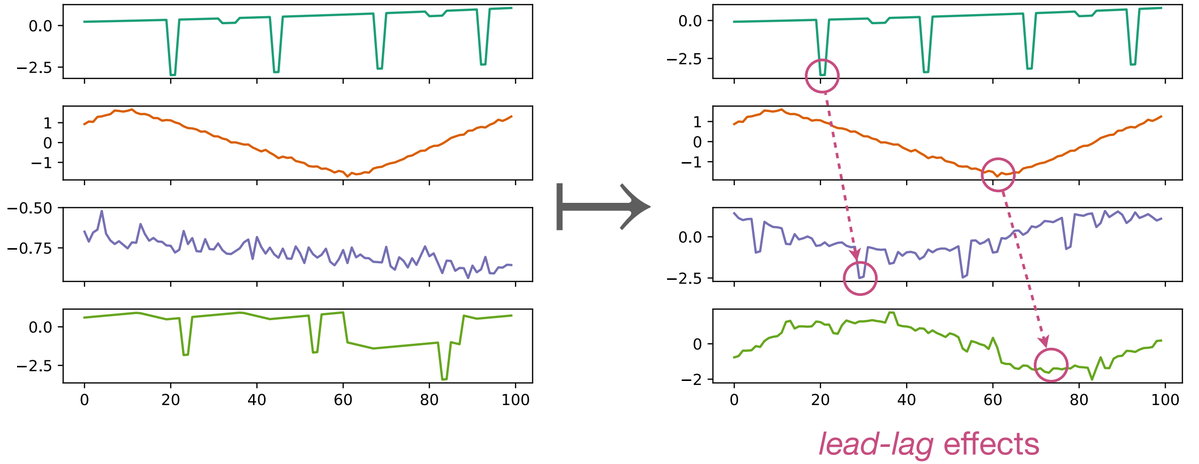


Instantaneous effects at the same time step

Sequential Multivariatizers



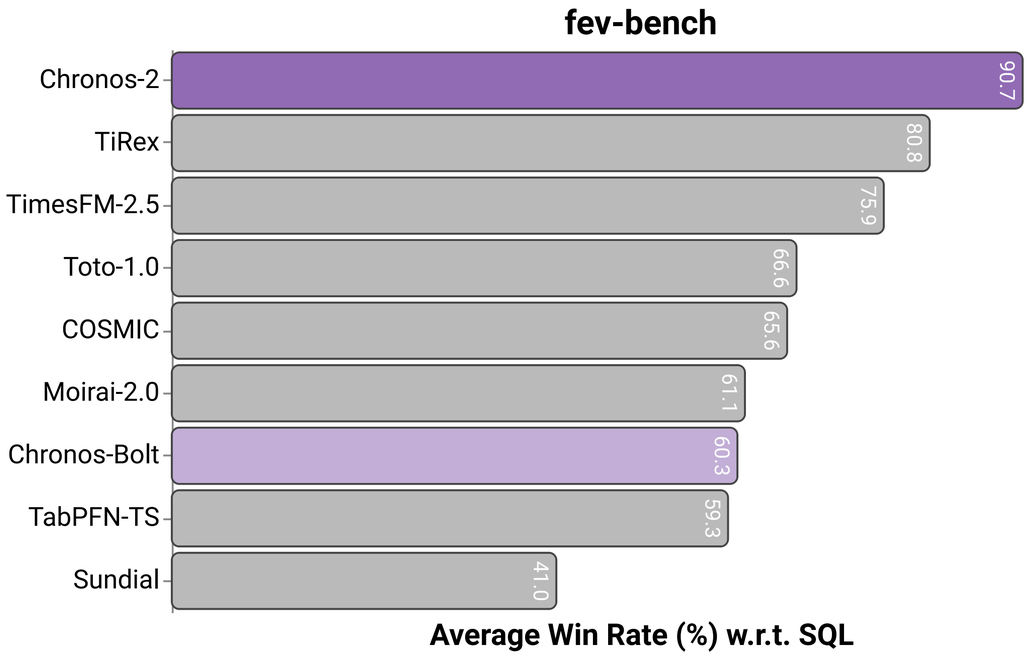
$$f : \begin{bmatrix} w_{t-k} & \cdots & w_t \\ x_{t-k} & \cdots & x_t \\ y_{t-k} & \cdots & y_t \\ z_{t-k} & \cdots & z_t \end{bmatrix} \mapsto \begin{bmatrix} \tilde{w}_t \\ \tilde{x}_t \\ \tilde{y}_t \\ \tilde{z}_t \end{bmatrix}$$

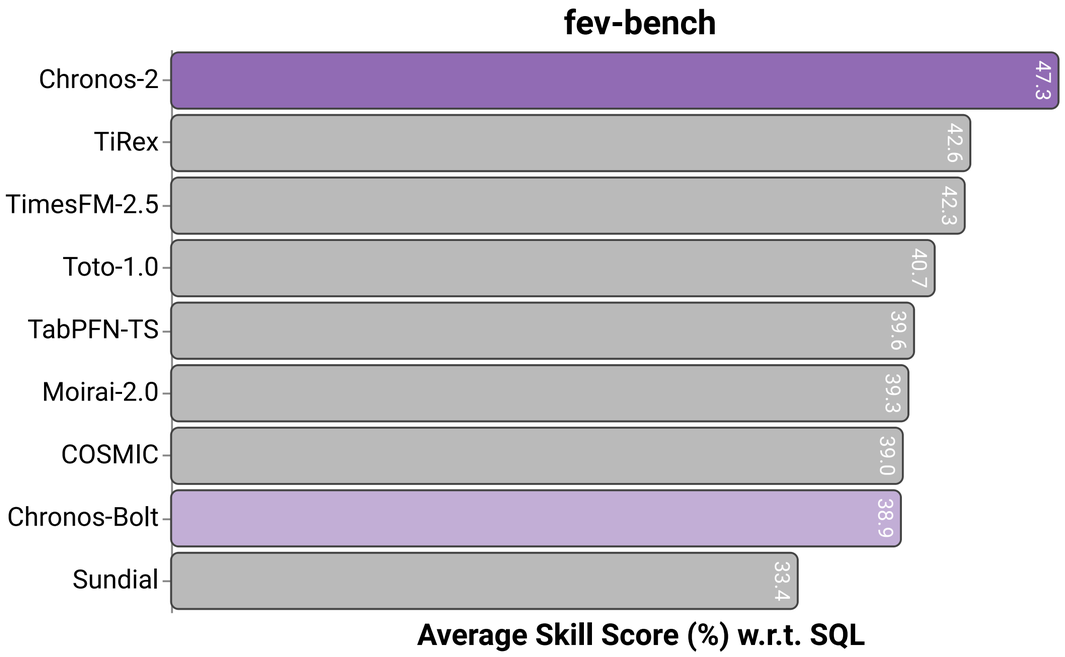


Effects across time

fev-bench Results

fev-bench: Comprehensive benchmark with **100 tasks** including univariate, multivariate and covariate-informed forecasting

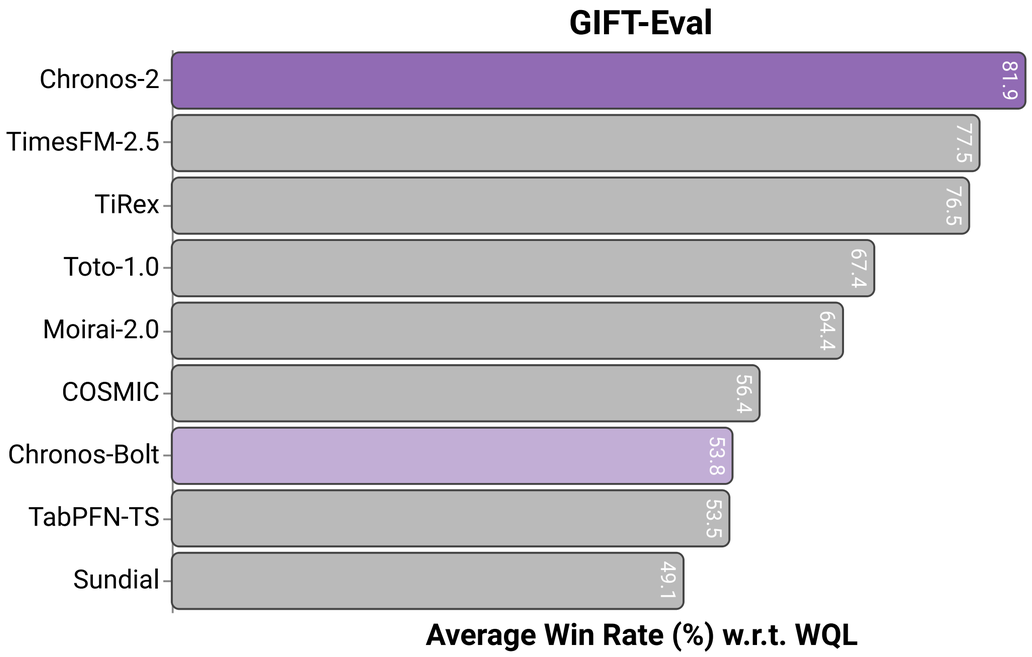


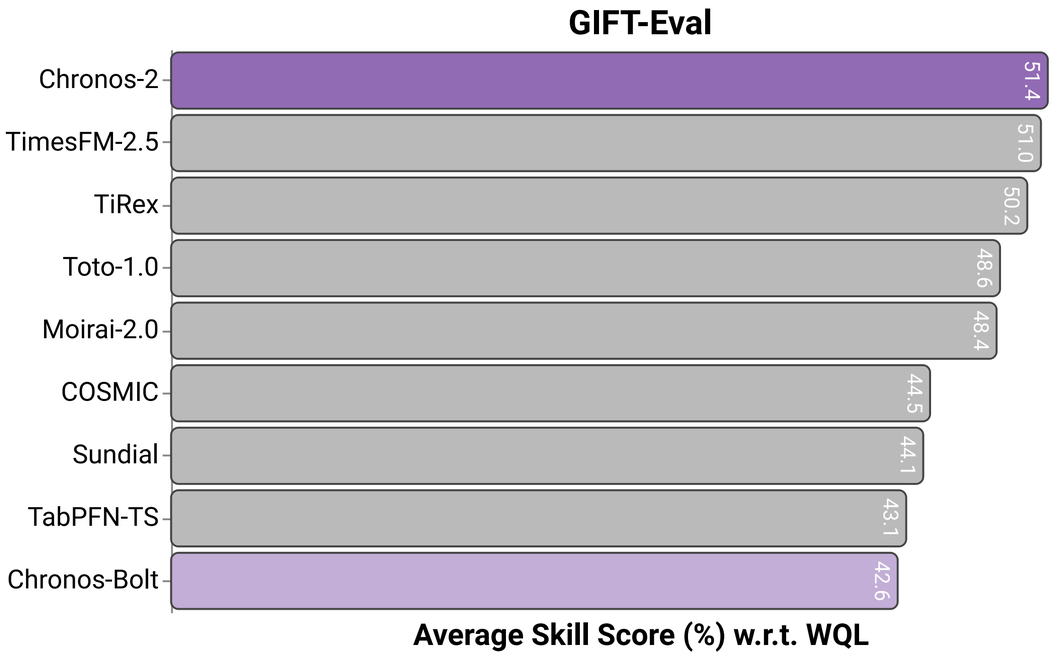


Results with respect to the scaled quantile loss (SQL) which evaluates probabilistic forecasting performance

GIFT-Eval Results

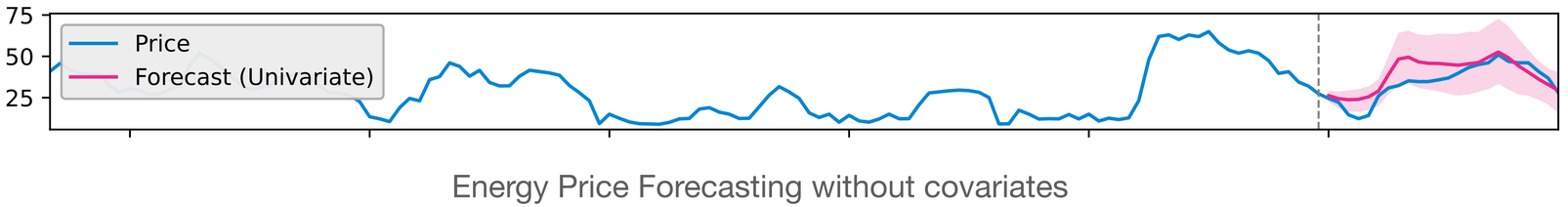
GIFT-Eval: Benchmark with **97 tasks** focusing on long-horizon forecasting tasks



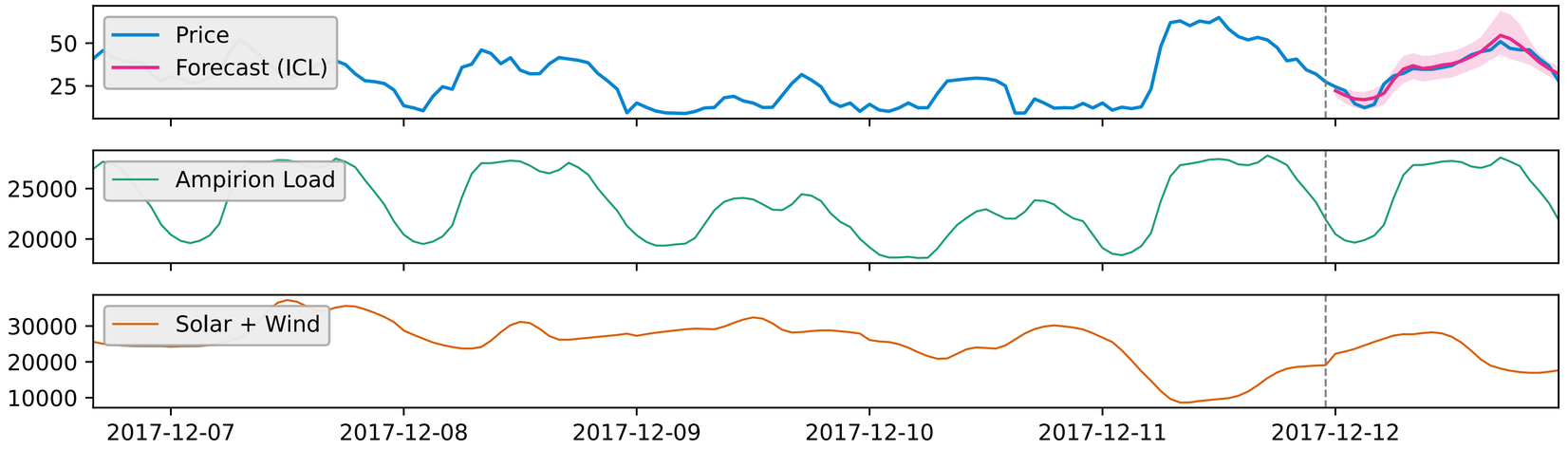


Results with respect to the weighted quantile loss (WQL) which evaluates probabilistic forecasting performance

Qualitative Results: Energy

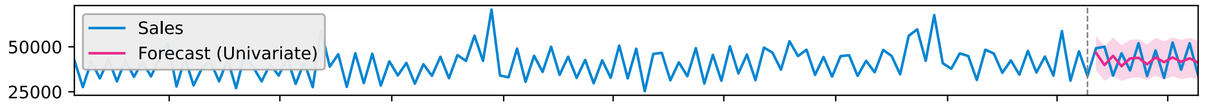


Energy Price Forecasting without covariates

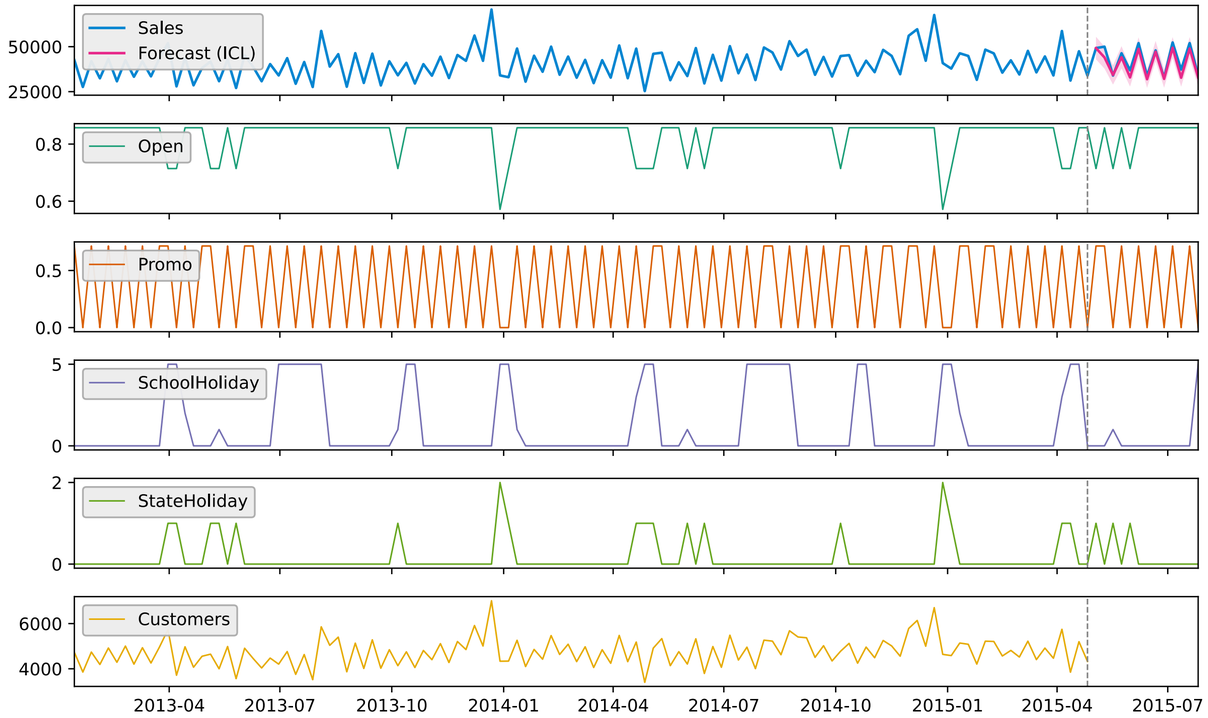


Energy Price Forecasting with covariates

Qualitative Results: Retail

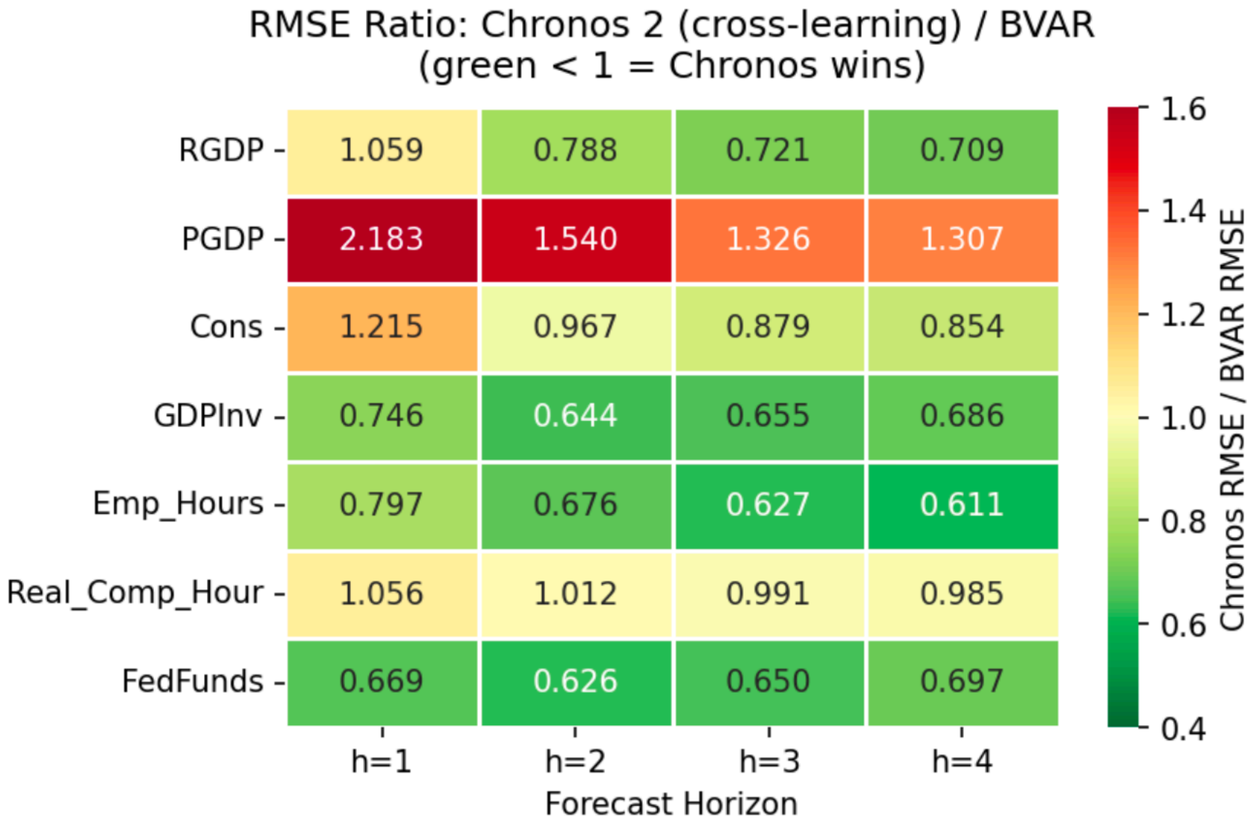


Sales Forecasting without covariates



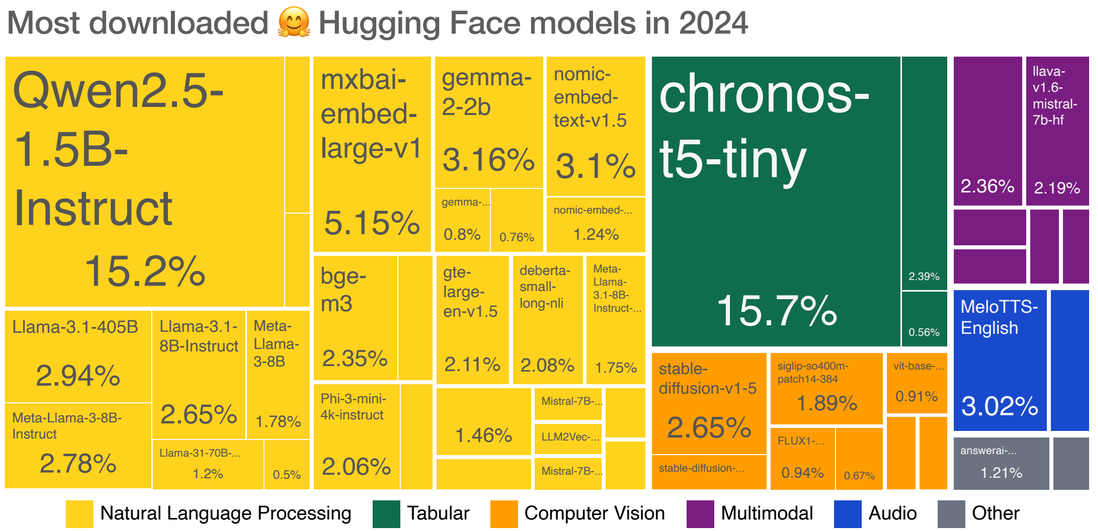
Sales Forecasting with covariates

Macro Application: BVAR vs Chronos-2



Chronos in Open Source

- Inference code available on GitHub
- Model weights available on Hugging Face
- Deploy Chronos-2 on AWS using SageMaker JumpStart
- Run Chronos with 1 line of code using AutoGluon (Chronos-2 coming soon!)



Chronos models downloaded 600M+ times

Getting Started with Chronos-2

pip install chronos-forecasting

```python
import pandas as pd
from chronos import BaseChronosPipeline

pipeline = BaseChronosPipeline.from_pretrained("amazon/chronos-2", device_map="cuda")

# Load historical data
context_df = pd.read_csv("AirPassengers.csv")

# Generate predictions
pred_df = pipeline.predict_df(
    context_df,
    prediction_length=36,  # Number of steps to forecast
    quantile_levels=[0.1, 0.5, 0.9],  # Quantiles for probabilistic forecast
    id_column="item_id",  # Column identifying different time series
    timestamp_column="Month",  # Column with datetime information
    target="#Passengers",  # Column(s) with time series values to predict
)

```

Shameless promotion: AI related work

1. *Double Descent in Time Series* (with J. Fernandez-Villaverde)

2. *Bounded Rationality as Limited Optimization: Stochastic Gradient Descent Agents*

3. *The Macro implications of Gen-AI* (with T. Mikami and J. Nosal)

4. *fev-bench: A Realistic Benchmark for Time Series Forecasting* (multiple co-authors)