A Foundational Model for Macroeconomic Times Series Forecasting and Nowcasting

B Koyunku, B Kwon, M Lombardi, **F Perez-Cruz*** and H S Shin

Bank for International Settlements *Prof at ETH Zurich

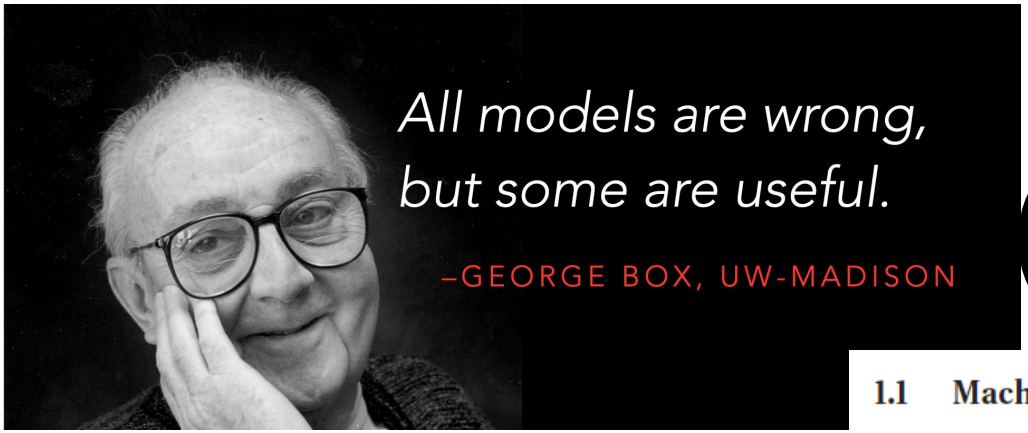
Why ML is not Statistics?

$$\min_{w} \frac{1}{n} \sum_{i=1}^{n} L\left(y_i, f_w(\boldsymbol{x}_i)\right)$$

$$\{(\boldsymbol{x}_i, y_i)\}_{i=1\ldots n} \underset{iid}{\sim} p_{X,Y}(\boldsymbol{x}, y)$$

Machine Learners and Statisticians solved the **same** problem

Why ML is not Statistics?



$$\left(y_i, f_w(\boldsymbol{x}_i)\right)$$

$$\{(\boldsymbol{x}_i, y_i)\}_{i=1}$$

Machine Learners anc

1.1 Machine learning: what and why?

We are drowning in information and starving for knowledge. — John Naisbitt.

We are entering the era of **big data**. For example, there are about 1 trillion web pages[1]; one hour of video is uploaded to YouTube every second, amounting to 10 years of content every day[2]; the genomes of 1000s of people, each of which has a length of 3.8×10^9 base pairs, have been sequenced by various labs; Walmart handles more than 1M transactions per hour and has databases containing more than 2.5 petabytes (2.5×10^{15}) of information (Cukier 2010); and so on.

This deluge of data calls for automated methods of data analysis, which is what **machine learning** provides. In particular, we define machine learning as a set of methods that can automatically detect patterns in data, and then use the uncovered patterns to predict future data, or to perform other kinds of decision making under uncertainty (such as planning how to collect more data!).

Do we need ML for time series analysis?

- We have many models to analyse time series:
  - AR
  - GARCH
  - VAR or BVAR
  - Kalman
  - Hidden Markov Models
  - Dynamical Switching Models
  - …

Do we need ML for time series analysis?

- We have many models to analyse time series:
  - AR
  - GARCH
  - VAR or BVAR
  - Kalman
  - Hidden Markov Models
  - Dynamical Switching Models
  - …
- Whatever the time series problem is, one of those methods should be good enough, right?

Do we need ML for time series analysis?

- We have many models to analyse time series:
  - AR
  - GARCH
  - VAR or BVAR
  - Kalman
  - Hidden Markov Models
  - Dynamical Switching Models
  - …
- Whatever the time series problem is, one of those methods should be good enough, right?
- The answer is not in the models but in the data:
  - ML helps with unknown nonlinear interactions or universality (maybe both).

The Game Changer: Zero-Shot Learning

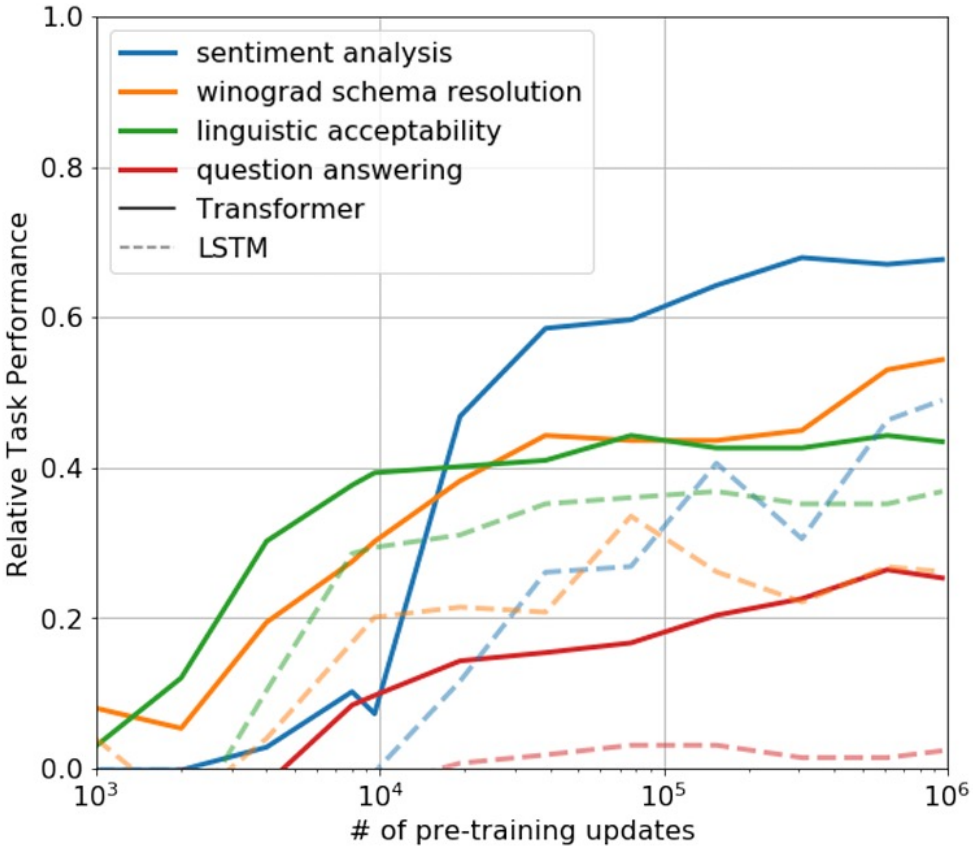
Zero-shot Behaviors We'd like to better understand why language model pre-training of transformers is effective. A hypothesis is that the underlying generative model learns to perform many of the tasks we evaluate on in order to improve its language modeling capability and that the more structured attentional memory of the transformer assists in transfer compared to LSTMs. We designed a series of heuristic solutions that use the underlying generative model to perform tasks without supervised finetuning. We visualize the effectiveness of these heuristic solutions over the course of generative pre-training in Fig 2(right). We observe the performance of these heuristics is stable and steadily increases over training suggesting that generative pretraining supports the learning of a wide variety of task relevant functionality. We also observe the LSTM exhibits higher variance in its zero-shot performance suggesting that the inductive bias of the Transformer architecture assists in transfer.

Improving Language Understanding by Generative Pre-Training

The Game Changer: Zero-Shot Learning

Zero-shot Behaviors We'd like to better understa
ers is effective. A hypothesis is that the underlying
tasks we evaluate on in order to improve its language

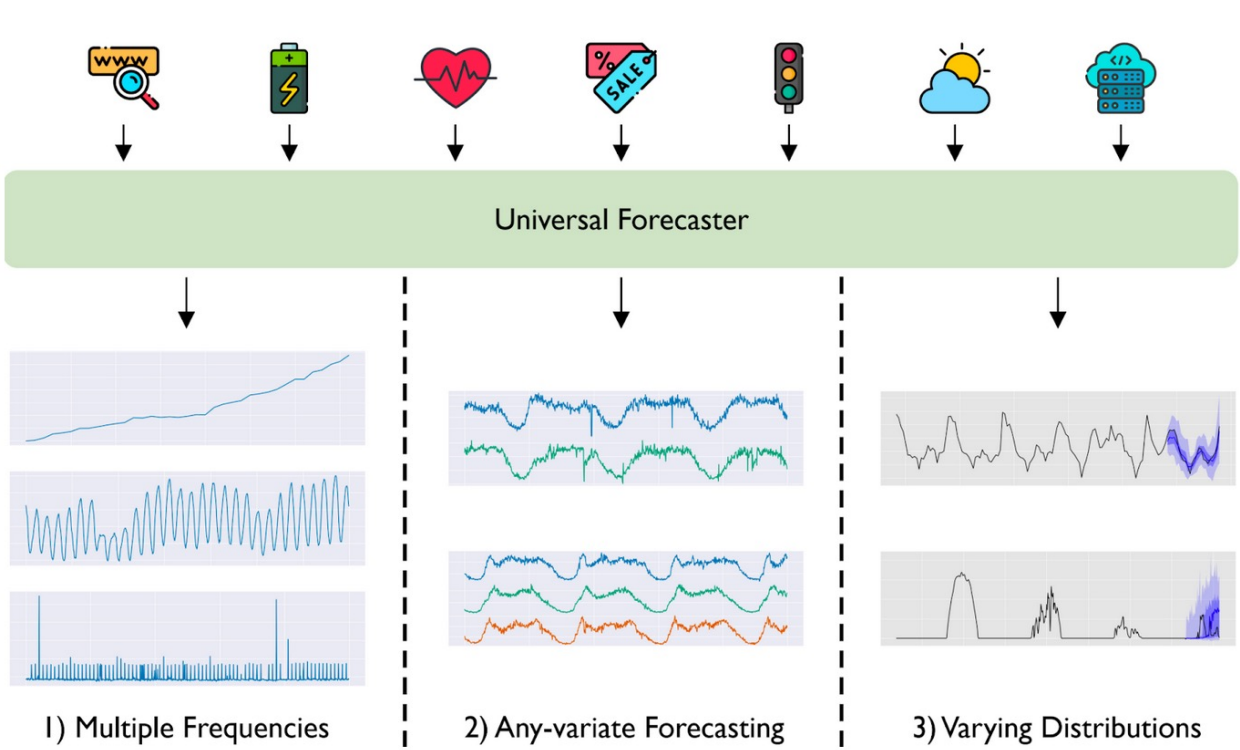
attentional memory of the transformer assists in tra
of heuristic solutions that use the underlying gener
finetuning. We visualize the effectiveness of these
pre-training in Fig 2(right). We observe the perfo
increases over training suggesting that generative p
of task relevant functionality. We also observe the
performance suggesting that the inductive bias of th



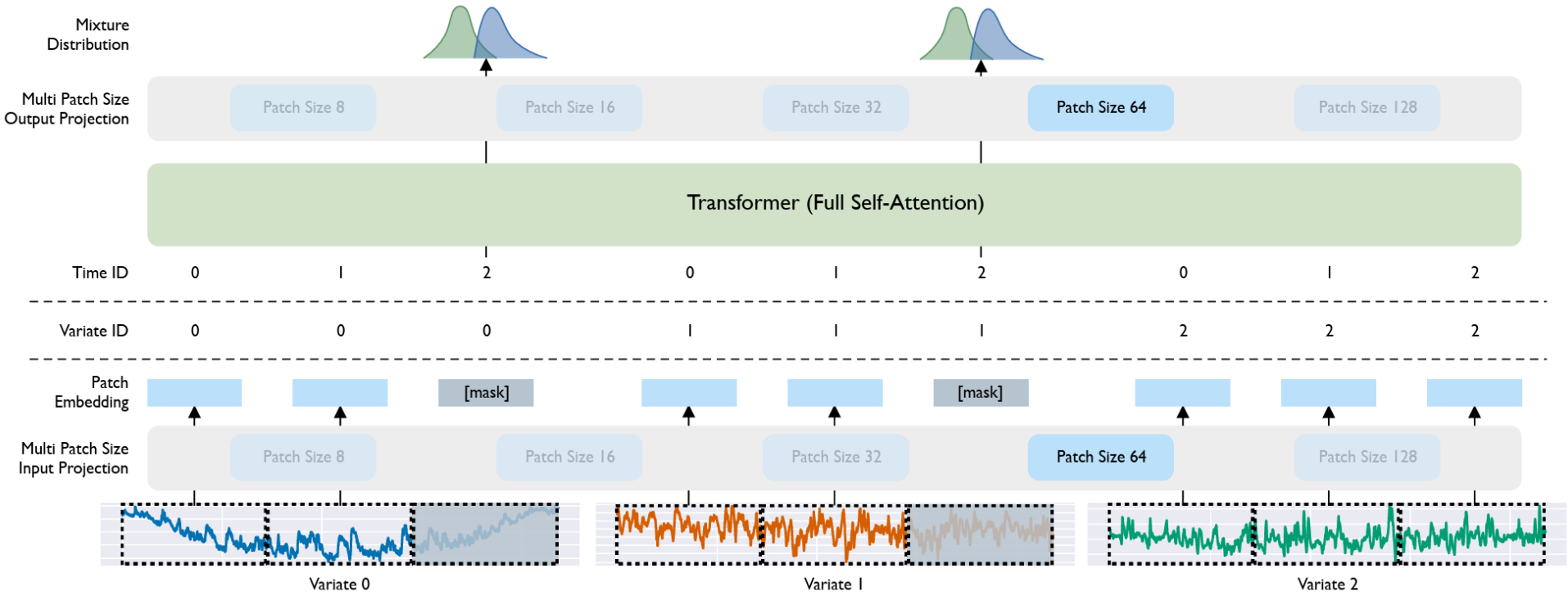
Improving Language Understanding by Generative Pre-Training

Transformers for time series: MOIRAI

- BERT based
- RoPE
- Only predict the last tokens
- Multi frequency
- Probabilistic output
- Data:
  - 60% data is about energy
  - 18% Transport
  - …
  - 0.1% Econ/Finance

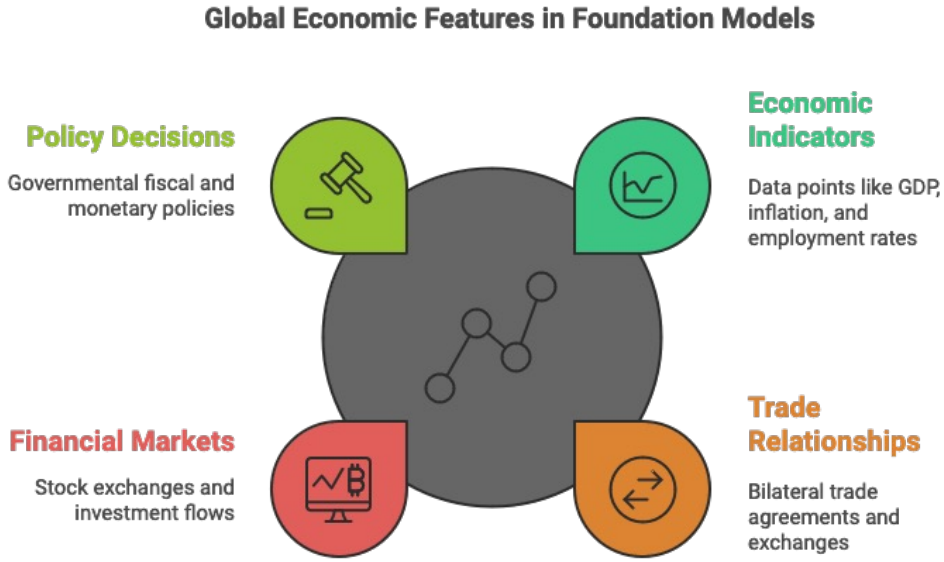


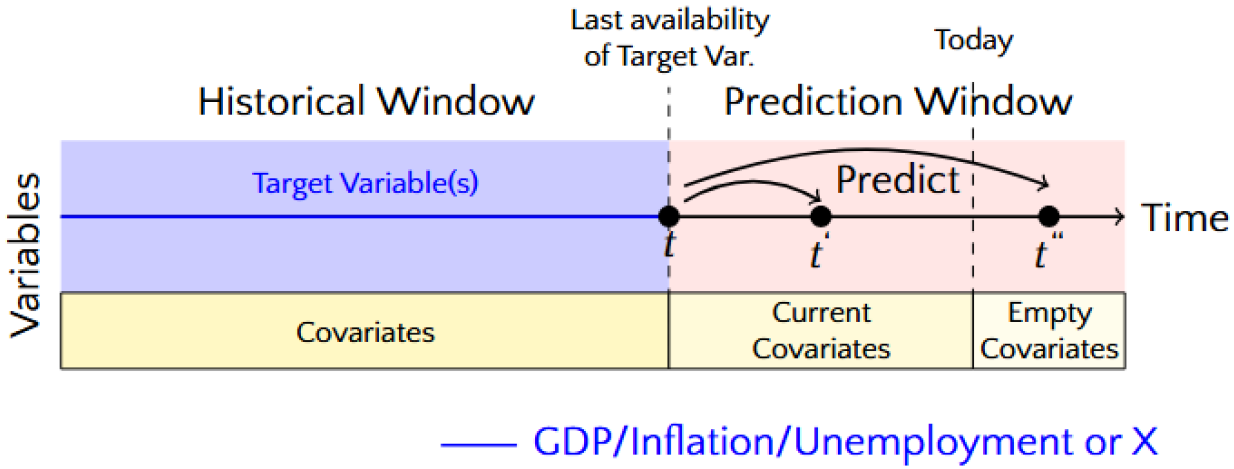
MOIRAI Architecture



MOIRAI for Macroeconomics

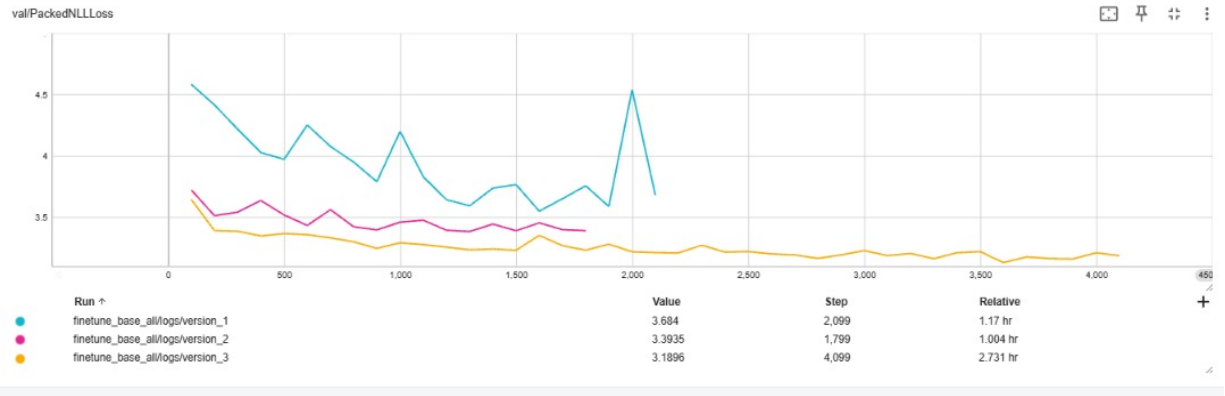
- Data:
  - BIS_MACRO data (80 countries)
  - Emphasis: GPD/CIP/Unemployment
- Granularity:
  - Daily
- Token:
  - Length 32
- Good for:
  - Forecasting
  - Nowcasting
  - Scenario planning

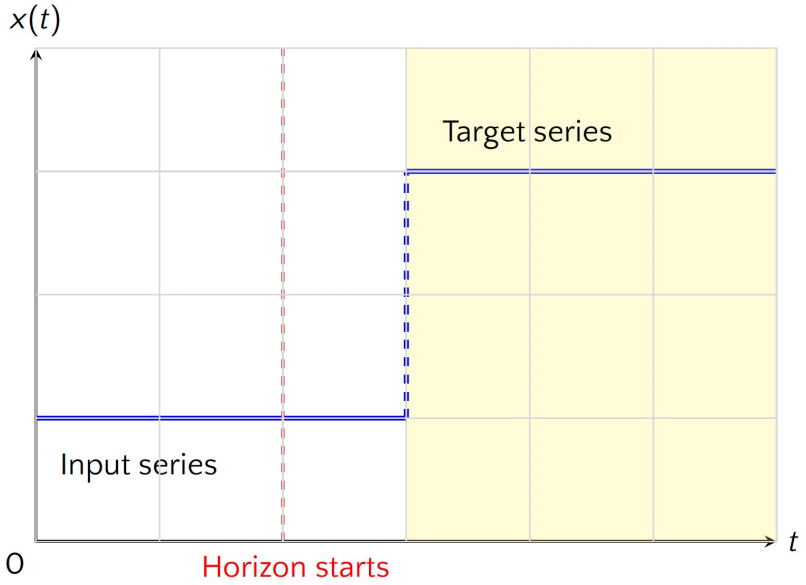




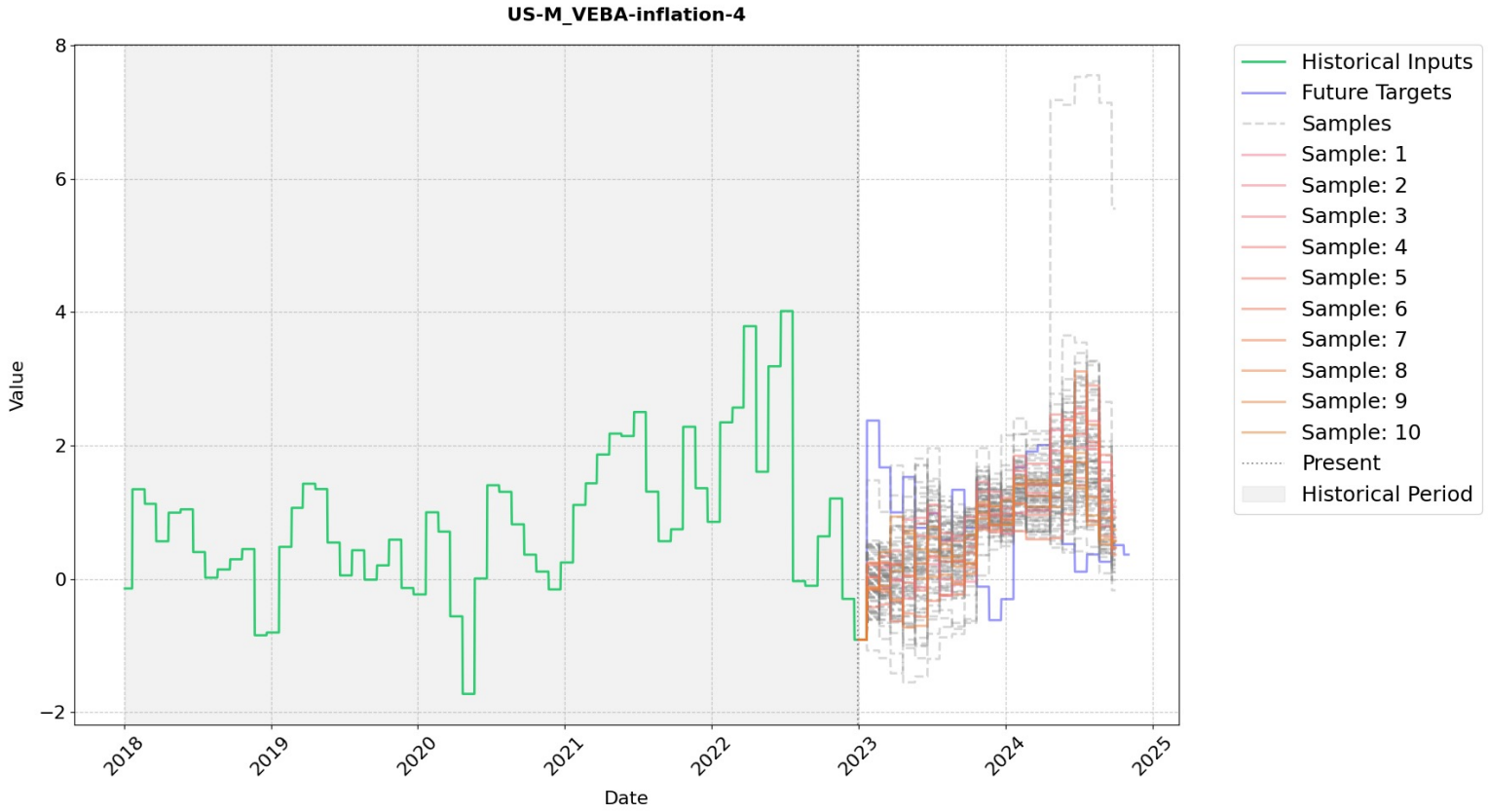
Implementation details

- Features:
  - Original data and first difference
  - Log data and first difference
- Input Vector:
  - GPD, CIP and Unemployment
  - 10 other variables
- Length: 48-60 patches (4-5 years)
- Training period [1984 to 2024]\ Test years
- Test years: 1995, 2005, 2015 and 2023+
- Masking 1 to 5 patches
- Loss over “jump mask”
- MOIRAI Base: 91m parameters

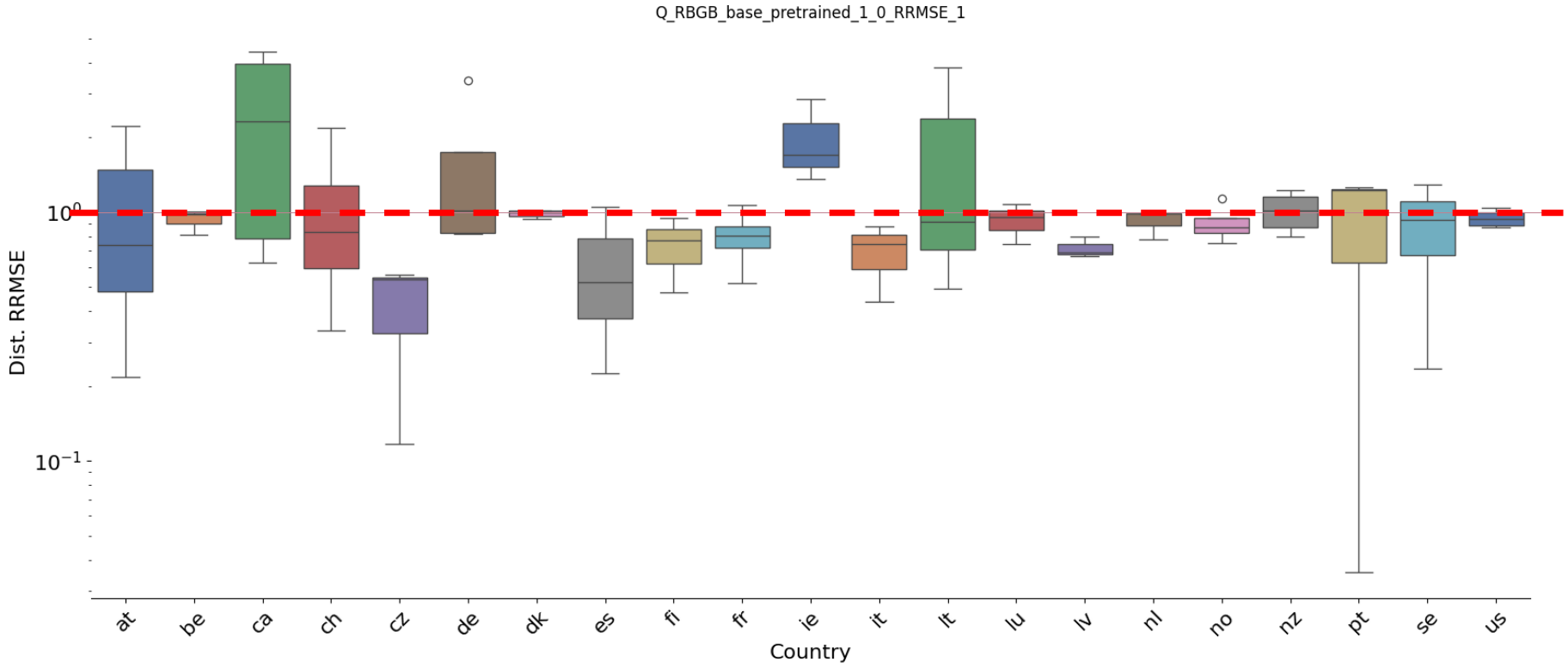




Future Looking US Inflation



Univariate Prediction of GDP Compared to AR(1)



What's next

- Complete training the model
- Univariate estimation
- Estimation with Covariates
- Nowcasting
- Scenario planning
- BIS internal validation
- Code Available Summer 2025

Reach out for testing interesting cases