Summary of Model comparison results:

Overall Breakdown:

* ARIMA performs best in most scenarios

A screenshot of a computer screen

Description automatically generated

Breakdown by Type of data:

* Fine-tuned Lag-Llama performs best on commodities
* On exchange rates, it is unclear if ARIMA or fine-tuned TimeGPT is best
* Fine-tuned Lag-Llama performs best on indices

A screenshot of a computer

Description automatically generated



Type of data = Commodities, frequency = weekly, Raw metrics:

A screenshot of a computer screen

Description automatically generated

Type of data = Commodities, frequency = daily, Raw metrics (skewed by the mean) (this is because ft\_lag\_llama did really bad on data 2022-01-01-2024-01-01 for some reason):

A screenshot of a computer screen

Description automatically generated

Type of data = Index, frequency = monthly, Raw metrics:

A screenshot of a computer screen

Description automatically generated

Type of data = Index, frequency = weekly, Raw metrics:

A screenshot of a computer screen

Description automatically generated

Breakdown by training size:

* Fine-tuned Lag-Llama comes on top in train size of 32 scenario.

A screenshot of a computer

Description automatically generated



Breakdown by frequency:

* Fine-tuned Lag-Llama is the best model for monthly frequency
* Fine-tuned Lag-Llama is the best model for weekly frequency

A screenshot of a computer

Description automatically generated



Type of data = cc, frequency = weekly, Raw metrics:

A screenshot of a computer screen

Description automatically generated

Type of data = fx, frequency = weekly, Raw metrics:

A screenshot of a computer screen

Description automatically generated

Breakdown by frequency and type of data:

* With indices, fine-tuned Lag-Llama only performs best with weekly and monthly frequencies
* With commodities, fine-tuned Lag-Llama performs best with daily or weekly frequencies
* Fine-tuned timeGPT performs best on weekly cc data
* With fx, fine-tuned timeGPT only performs best on weekly data

Breakdown by stationarity of time-series:

* On stationary time-series, fine-tuned Lag-Llama performs best

A screenshot of a computer screen

Description automatically generated

Breakdown by trend of the time-series

* On time-series with no trend, fine-tuned Lag-Llama performs best

A screenshot of a computer screen

Description automatically generated

Breakdown by seasonality of time-series:

* On data with no cyclical patterns, fine-tuned Lag-Llama performs best

A screenshot of a computer screen

Description automatically generated

Summary of lag-Llama fine-tuning experimentation

Batch size:

* Batch size doesn’t seem to have significant influence on model performance, however, lower batch size seems to do better.

A screenshot of a computer

Description automatically generated

Max epochs:

* Max epochs appears to have no significant effect on performance of the model, however, lower max\_epochs seems to do better

A screenshot of a computer

Description automatically generated

Fine-tune length:

* FTL does seem to have a significant effect on performance of the model. Low FTL doesn’t seem to perform well, however, there is some evidence showing performance drops (according to certain metrics) with thigher FTL.

A screenshot of a computer

Description automatically generated