Experiments

* Hypothesis: different frequency data has different seasonality and trend structures. I expect that models’ performance will change if we use different context lengths and how the performance changes depends on the frequency of data.
  1. How different frequencies affect results?
  2. How different context length affects results?
  3. How the combination of context length and frequency affects results?
     + Why? Because depending on what kind of data we want to use the model for, we want to know which context length to use.
     + Solution: Experimenting with different fine-tuning lengths
       - Different types of data and different frequencies might benefit from different fine tuning lengths
         * Due to different seasonality and trend structures
* Experimenting with separated fine-tuning and “train” set vs “train” set included in fine-tuning set
  1. Theoretically there isnt reason why “train” and “fine-tune” sets thould be separate
  2. Theoretically there should be benefit to “train” to be included in fine-tune because that brings the fine-tune set closer to the time of prediction
  3. Running this experiment across different context lengths (train sizes) and comparing results could answer how important the proximity of fine-tuning data to prediction is
* Repeat the TSCV across different periods of the time-series
  1. No need to try out for different number of folds and different number of times TSCV is repeated across a single time-series. I should pick a number for both and stick with that
  2. Make sure that the TSCV runs are equally spaced out throughout the time-series
* Stock data vs return data
  1. stock data results will probably be bad so put them in appendix along with other results that turn out bad
  2. Inquire about and use different time-series preprocessing methods that Ramin mentioned
     + Differencing, etc…
* When is the data from?
  1. Use data after February 2024 to avoid data leakage.
  2. Including data before February 2024 could cause data leakage but it would be interesting to compare TSFM’s performance with and without leakage
     + In order to maximize the chances of leakage, we should use index prices because they are most likely to have been included in training
     + Tie this to the concept of “machine unlearning”
       - Talk about “machine unlearning” in the Future Work section
* Consider other types of data
  1. Exchange rates
  2. Commodities
  3. Crypto
  4. Credit card data (from Lending Club)
* Repeat the whole experiment multiple times (seats)