Hi Matt and Jinwook,

Here is the current draft of the paper. I was able to go in depth up to Section 5.4. Major changes are as follows:

* revising introduction and lit review

No comments for now, I can’t tell how my comments have been improved without looking at three documents, having you walk me through it for a few hours, or reviewing everything again from a basic level forgetting my prior comments (a significant time investment, can do at a more final stage). I suggest moving to MS Word for the writing/track changes or making it incredibly clear which comments you addressed and where those changes were made, or waiting to ask me how it is when you have a more complete paper (may increase the number of revisions for you). It’s fine to do intro/lit review/conclusion in MS Word and the methodology in Overleaf. Just do that…this is too messy….and then bring back when finished.

* revising k-nts section to use time series features, and implementing k-nts privacy method in R

I don’t need compare documents for the methodology section, it’s too hard to read.

2nd paragraph missing a citation or two where past IJF/other papers actually did something very similar on past time series patterns for forecasting. Then, you need to say how your method is slightly different too (swapping for privacy).

Li et al. (2022) should be written in a better way if you are building from their approach. Don’t tell them to do homework and read their paper.

Not specific enough for a researcher replication. Be more specific about all the features being calculated. What scale or range are they on? How many seasonalities are there? Are they normalized prior to computing distances (they probably should be if the scales are very different). What happens if AR1 is close to 0 for all time series, but seasonalities are high for all time series? Then, AR1 should not count for much…

Seems like an example with 4 time series (with example data) on a plot should show how a similarly patterned time series has low distance to another one, compared to a time series on the same scale with different features. And perhaps highlight the swaps that could happen.

* revising results section to have stronger storyline, currently focusing on changes in time series characteristics and model behavior

Delete your text in half and write more directly with quantitative results.

Empirical Application

I’m pretty sure the word is complex, not complicated. But you can check the literature. Complicated is a subjective word.

Results

I’m not so sure how effective the intro paragraph is. Maybe just do it, instead of wasting space writing about doing it, I find it distracting currently but maybe that is just me.

5.1 - writing needs to improve and be more direct, especially that 1st paragraph. But, I like the idea of 22 features to predict accuracy, with the citations (the writing around the citations are not well written or specified).

The writing is off and very confusing. How can you predict accuracy when you haven’t even stated a forecasting model? I would avoid talking about what you are going to do and instead write what you did and show the results as you write it. Within the paper, be a storyteller, not a future planner. There’s too many “by the ways” or “however” – not very clear what is going on.

The interesting part of this paper should be the features, but after reading it, we don’t really learn anything about the features. They are treated superficially in the writing.

Figure 3 is the first thing that provides any detail.

I’m concerned you’re using future information to make choices in k-nTS in the present. The writing seems to indicate this.

“Privacy protection tends to bias the series downward, resulting in decreases in the series means and decreasing the accuracy of privacy adjusted forecasts.” - for this data set or in general?

The writing could be cut in half and be more focused

5.2

Writing like this is not particularly desired. The statements are very general, but the evidence is not clearly specified. “For meaningful levels of ϵ ∈ {0.1, 1}, we see the destruction of time series features that are vital for accurate forecasting. The Hurst coefficient distributions show the removal of most of the persistence of the series, making them have a tendency to revert to a long-term mean which is biased downward, leading to extremely poor forecast accuracy.” You need to quantitatively specify what happened (e.g., MAPE increased by 115%). It’s all generalities here…and why doesn’t the reader understand what a Hurst coefficient is by now? For all the space it takes up, we can’t simply understand it or how it is calculated? Then why would someone want to read about it…

Not a good practice to reference results in section 5.4 while writing in section 5.2. Do these coefficients needs forecast accuracy to talk about their changes? What’s the purpose of this section? I thought it was to understand the features and how protection changes them, before moving to forecast accuracy. Very confusing as it’s written now.

5.3 - what’s the purpose of this? Why are we talking about forecasting model again? Do we really need that to see changes in characteristics? Why are we only using SES when we have so many other features other than level?

5.4

The accuracy results make sense.

For SES, how is alpha chosen? So, that figure is not really forecasting…it’s some AIC based on in-sample data (read the R documentations). In which case, the discussion around alpha for forecast accuracy, is actually not forecast accuracy…it’s in-sample analysis.

Average rank analysis , i think you’ll need something next to average rank like average error variance and average mae

Figure 9 seems appropriate

The comment “Models with lower forecast error variance will be more desireable since forecasters can have higher confidence that they have not generated forecasts for some series which are significantly worse than others” is unnecessary. Please think about why you included this because a lot of the writing is like this. This is more teaching than research…any serious forecasting researcher knows this, so it’s not an interesting statement and shows a review team that you might be new to research.

The statement “the increase in error variance for all models is at least 50% compared to the forecast error variance for auto-ARIMA on the original data” is an interesting statement because it is written in a matter-of-fact way.

After reading this, I have no understanding of why certain forecasting models did worse on protected data than others. You need to tie in common themes (time series features) and connect those to the explanation literature from the forecasting models, to explain changes in accuracy in a last subsection. And preferably, this should be done quantitatively in a logical framework instead of just writing. As of now, the results section is very confused.

My next steps:

* write out equations for models in empirical section to motivate later results
* add results for changes in all model equations/behavior
* display model rankings for point forecast accuracy and forecast error variance for each privacy method (rather than tables and tables of MAE values - put these in appendix)
* comparison of privacy adjusted and judgmentally adjusted forecasts

Thanks for your thoughts and input so far. Looking forward to seeing what you think of the current version.

I'll plan to provide another update on September 23rd.

Cameron