Dear Editors and Reviewers,

Thank you for your comments and the opportunity to revise our manuscript. This document contains our point-by-point responses (in blue) to your comments (in black). We believe the manuscript has improved significantly and hope that it addresses the review team's concerns. The major changes to the manuscript are as follows:

Major Changes to Manuscript

1. **Repositioning of the paper**. Upon expanding our analysis to include additional M3 and M4 data sets, we discovered that the performance of our proposed *k*-nTS+ method on the M3 Monthly Micro data was atypical, and that it is uncommon to create protected time series with good forecast accuracy on the original scale. On the other hand, sharing model weights trained on unprotected data enables accurate forecasting, but accurate forecasts themselves can uncover the identities of the unprotected time series. We show that normalizing time series values increases the cross-series similarity of features, values, and forecasts, enabling good forecast accuracy at acceptable levels of privacy. This is now the central premise of our paper: except under certain conditions, creating protected time series with acceptable privacy is incompatible with obtaining accurate forecasts.

2. **New datasets.** To increase the generalizability of our results, we have included all the M3 competition data in the empirical application and include accuracy and privacy results on the M4 data in the appendix.

3. **Readability of the paper.** We have rewritten the paper using LaTeX and made significant changes to every section of the paper to improve readability. We have revised all algorithms and figures within the paper. All code has been uploaded to Github for reproducibility.

4. **Contributions of the paper**. We have carefully rewritten the claimed contributions of the paper over the prior literature to include:
    a. We show that achieving accurate forecasts based on protected data is not always possible and that sharing model parameters trained on the unprotected data can enable accurate forecasts that compromise privacy. We perform a thorough evaluation of the performance of seven forecasting models (ranging from exponential smoothing to a recurrent neural network) applied to time series that are protected using four privacy methods. Using time series data sets with features that are representative of real world data, we show that time series (on their original scale) can rarely be protected with good forecast accuracy, and that data owners and forecasters will need to adapt to using scaled rate data if protected time series are desired, regardless of the forecasting model used.
    b. We show that under certain conditions such as select data sets and/or normalized time series values, we enable data owners to share a protected version of a time series data set with good forecast accuracy and reasonable privacy. We measure privacy based on two disclosure risks: the probability of reidentifying protected time series using past protected values and forecasts. While low probabilities of these risks do not guarantee the theoretical standard of differential privacy, our measures are easily interpreted and enable data owners to perform a reasonable privacy assessment as per the GDPR. Furthermore, we find that data that is differentially private in name provides weak protection against reidentification. In

comparison, our proposed k-nTS+ method consistently outperforms benchmark methods by a significant margin on the trade-off between privacy and forecast accuracy and enables highly accurate forecasts using the protected rate data.

c. We use a machine-learning based feature selection process that incorporates the usefulness of data (forecast accuracy) into the protection process. The results show that features which are statistically significant predictors of forecast accuracy are not necessarily the most useful for swapping time series values. We analyze how these features change vis-à-vis accuracy and show how various privacy methods differentially affect time series features. Overall, using this feature selection process improves the performance of our proposed swapping method anywhere from 4%-30% depending on the data set relative to swapping without the machine learning process. Furthermore, the feature selection method is not limited to data privacy applications and can be used to select an efficient set of features for predicting any categorical or continuous single valued metric.

We also want to disclose an indexing error in the code for the previous version of the paper that caused only *k*-1 nearest neighbor time series to be used for swapping instead of *k* nearest neighbor time series. This error has been corrected.

Editor Comments:

The manuscript concentrates on a timely and relevant topic. The approach is original and the point is well supported. The reviewers and I see that the manuscript potentially brings some interesting novel ideas that could be worth of publication. However, at this stage, there seems to be quite a lot of work to be done for the paper to get there.

Thank you for your helpful feedback. We hope that you find the proposed changes satisfactory.

Besides the point of the reviewers, I would like to insist on the fact the presentation of the work needs a serious upgrade. For instance, A figure like Figure 1 cannot end up in the final version of the paper. I would encourage the authors to be careful in the way to design and produce that figure.

Thank you. We have removed figure 1 from the paper and now focus on the time series with desirable/undesirable features. We have improved the presentation of all existing (and new) figures.

Similarly, in general, the paper looks like a draft, which makes it a bit tricky at stages (e.g., when I checked the validity of equations...). Using a more "profession" text editor (possibly Latex) could help a lot in improving the readability and make the work easier for the reviewers and I. Finally, for the maths, it would be good to be really thorough, e.g., one does not need to use "*" for multiplication, the flow of information around eq. (2) is quite confusing (possibly introduce relevant notation first, for diag, 1, among others), etc. Similarly, algorithms could be better presented.

We apologize for the presentation of the original submission and thank you for your feedback. Per the “Major Changes to the Manuscript”, we have repositioned the paper (Point 1) and re-written the draft using LaTeX (Point 3). We have reviewed our equations and added descriptions of relevant notation to aid the reader. To improve readability, we now present three distinct algorithms: (1)

Algorithm 1: The *k*-nTS Swapping Method, (2) Algorithm 2: Two-Stage Feature Selection, and (3) The *k*-nTS+ Swapping Method (which utilizes Algorithms 1 and 2. We have also added a detailed flowchart (Figure 2) to help the reader visualize the steps in our proposed *k*-nTS+ swapping method.

Reviewer 1 Comments:

The authors propose a matrix-based privacy method called k-nearest time series + (k-nTS+) swapping that preserves time series features to maintain forecast accuracy. The proposed privacy method has been applied to a forecasting competition data set and proven its advantages through a series of empirical studies. Overall, the paper is well-structured and written, while its contribution is clearly explained and justified.

Thank you for your constructive feedback on our manuscript. Please see the "Major Changes to the Manuscript" and our comments directly to you below. We hope that the revised version is much improved.

Below you may find some comments that could help further improve the current work.

1. Page 5: When first introducing the k-nTS+ swapping method in Figure 2, the authors should provide more details on how it works. The framework now is primitive.

Thank you for this suggestion, we have increased the amount of detail in (and improved the presentation of) Figure 2 on page 13 to better illustrate our proposed *k*-nTS+ method. We have also included the full details of the proposed privacy method using Algorithms 1, 2, and 3 on pages 10, 12, and 14.

2. Section 2.3: The literature review of time series features for forecast accuracy could profit from including the relevant works such as:

Kang Y, Cao W, Petropoulos F, et al. Forecast with forecasts: Diversity matters[J]. European Journal of Operational Research, 2022, 301(1): 180-190.

Li L, Kang Y, Petropoulos F, et al. Feature-based intermittent demand forecast combinations: accuracy and inventory implications[J]. International Journal of Production Research, 2022: 1-16.

Montero-Manso P, Athanasopoulos G, Hyndman R J, et al. FFORMA: Feature-based forecast model averaging[J]. International Journal of Forecasting, 2020, 36(1): 86-92.

We greatly appreciate these suggestions and added these references to our literature review on pages 5-6.

3. Section 4.1: Why only use the monthly micro dataset from M3 competition? I recommend using all M3 competition data and discussing the performance of the proposed *k*-nTS+ method for the data with different frequencies. More recent M4 Competition data is also a better option.

Per the "Major Changes to the Manuscript" (Point 2), we now analyze the full M3 competition data set (pages 25-38). We also analyze the accuracy and privacy results for the M4 competition data (pages 45-46 in the Appendix).

Compared to the prior version of the manuscript, we find that the majority of the original M3 data sets cannot be protected while maintaining good forecast accuracy unless model weights are shared in place of protected data. However, these model weights can be used to simulated time series and generate forecasts which are not private. On the other hand, the privacy-accuracy trade-off is very desirable on the M3 rate data sets, where our proposed method produces protected data with a reduction of 3.63% in average forecast accuracy (See Table 9 in the paper). Tables 13 and 14 in the appendix (Table 14 shown below) report the MAE (% change in MAE) for each model and each data subset, respectively.

	Original Forecasts		**Inverse Rate Forecasts**	**Rate Forecasts**	
Data	*Unprotected*	*k-nTS+ (k = 3) M = 1.5*	*k-nTS+ (k = 3)*	*Unprotected*	*k-nTS+ (k = 3)*
Monthly Demographic	112.53	444.88 (295.34%)	113.61 (0.96%)	0.0049	0.0051 (2.27%)
Monthly Finance	291.52	861.62 (195.56%)	482.21 (65.41%)	0.0073	0.0092 (26.57%)
Monthly Industry	517.46	875.63 (69.22%)	595.45 (15.07%)	0.0126	0.0134 (6.74%)
Monthly Macro	178.59	642.08 (259.53%)	206.46 (15.61%)	0.0036	0.0039 (8.50%)
Monthly Micro	687.53	789.24 (14.79%)	937.80 (36.40%)	0.0341	0.0363 (6.47%)
Monthly Other	400.28	704.90 (76.10%)	402.39 (0.53%)	0.0204	0.0122 (-40.44%)
Quarterly Finance	138.95	412.52 (196.89%)	117.95 (-15.11%)	0.0031	0.0027 (-11.19%)
Quarterly Macro	162.76	481.47 (195.82%)	193.12 (18.66%)	0.0034	0.0041 (21.54%)
Quarterly Micro	594.95	1044.80 (75.61%)	743.27 (24.93%)	0.0127	0.0142 (11.59%)
Yearly Demographic	351.97	1095.81 (211.34%)	333.12 (-5.35%)	0.0056	0.0056 (0.02%)
Yearly Finance	1678.25	3600.49 (114.54%)	982.78 (-41.44%)	0.0157	0.0139 (-11.90%)
Yearly Industry	429.82	1295.09 (201.31%)	397.04 (-7.63%)	0.0104	0.0102 (-2.19%)
Yearly Macro	184.97	358.28 (93.69%)	177.68 (-3.94%)	0.0029	0.0031 (8.09%)
Yearly Micro	949.64	1681.62 (77.08%)	864.38 (-8.98%)	0.0237	0.0198 (-16.35%)
Other Other	47.83	529.30 (1006.71%)	40.49 (-15.34%)	0.0013	0.0012 (-2.36%)

Table 16: MAE (% change in MAE) for each data subset from the *k*-nTS+ protected M3 original and rate data.

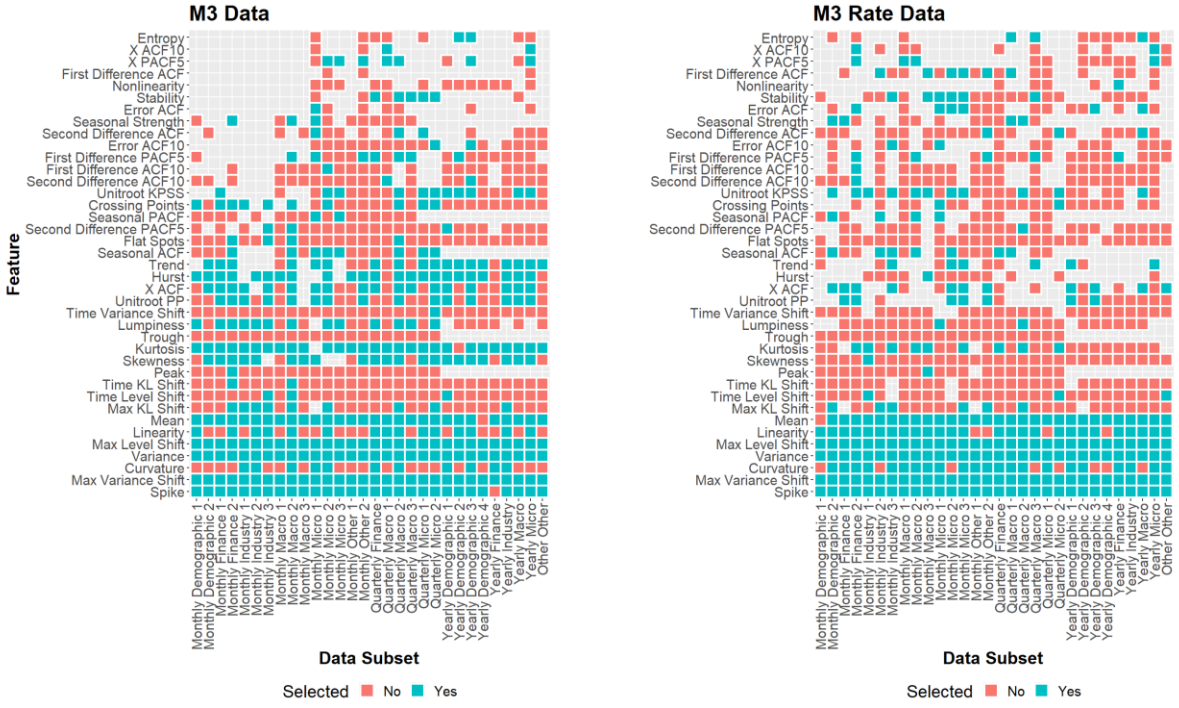
Section 4.2: The authors should clarify the details of the feature selection, e.g., why select such features for k-nTS and add new features for k-nTS+. The process seems subjective. The authors should give all the alternative features and explain the reason.

We apologize for the lack of clarity on the previous manuscript. To clarify this point, we try to avoid selection bias by including a large set of features in *k*-nTS+ which could help to maintain forecast accuracy in the swapping process. In Section 5.2 on page 25, we explain that we manually selected features for *k*-nTS swapping that were used to improve the accuracy of RNNs (Bandara et al., 2020) and/or were statistically significant predictors of forecast accuracy on the M4 competition data (Spiliotis et al., 2020).

We note that the initial feature set for *k*-nTS+ includes all features that were manually selected, as well as additional features that can be computed using the *tsfeatures* package in R. The proposed feature selection method is designed to select the features which balance the trade-off between forecast accuracy and data privacy.

Our results show that the feature selection method selects anywhere from 5 to 23 out of 39 features, depending on the subset of data, based on the minimum out of bag error of the random forest used to predict the accuracy of the forecasting model(s).

To further clarify this, we now include Figure 7 which shows the features selected for each subset of M3 data (shown below). The blue (red) squares indicate that a feature was (was not) selected for the corresponding data subset. Gray squares indicate that the feature was either not computable for that time series frequency or was eliminated in the first stage of the feature selection process. We see that some of the features that we manually selected (*e.g., Mean* and *Variance*) were commonly selected by the machine learning process. Others, such as *Spike*, *Max Variance Shift,* and *Max Level Shift* were frequently selected but have not been highlighted as statistically significant predictors of forecast accuracy. We hope this helps clarify the details of the feature selection process.



5. A brief discussion of the computational cost is useful for other researchers.

Thank you for bringing this omission to our attention. We added a paragraph to Section 3 on page 10 that describes the time complexity of the *k*-nTS swapping algorithm (Algorithm 1). In addition, Section 5.9 on page 38 includes a plot and discussion of the computation time for the full protection process as well as each step in the protection process. We find that the time to create a protected data set is reasonable, especially considering this process can be performed once for past data and continues on a rolling basis as data is generated for future time periods, and that the feature selection and swapping process can be performed in less time than it takes to train and forecast using Auto-ARIMA models.

6. Is there any reason for the error measures used in sections 4.4 and 4.5 to differ? If MSE does not provide significantly different results than MAE, personally I would prefer a consistent measure to be used for both sections.

We now use MAE to assess both the accuracy of forecasts in Section 5.4 (previously Section 4.4) and the out of bag errors for the random forest predictions of the forecast MAE in Section 5.7.

7. Figure 6: Each diagram should be numbered differently, such as A.1, A.2, A.3, B.1, B.2, B.3.

Thank you, we have adjusted Figure 8 (previously Figure 6) accordingly.

8. Figure 7: The ordinate title should not have "average"? It is unclear what "time series features for each privacy method" means exactly.

Thank you we have corrected the titles in Figure 9 (previously Figure 7).

Reviewer 2 Comments:

The authors propose a method for preserving data privacy in time series data through a swapping technique. This approach focuses on maintaining forecast accuracy by swapping the data values only if the essential features of the time series, such as mean and autocorrelation function (ACF), are likely to remain unchanged. The proposed method assumes a centralized approach, where a single data owner possesses the time series data. In this scenario, a forecaster selects a forecasting model F, and the data owner performs data swapping to prevent a decline in F accuracy.

Thank you very much for your helpful comments. We hope that the newly revised manuscript addressed your main concerns.

While the idea sounds interesting, the paper requires further clarification and enhancements to address the following points:

- Applications: The authors should provide further clarification on the potential applications of their proposed method.

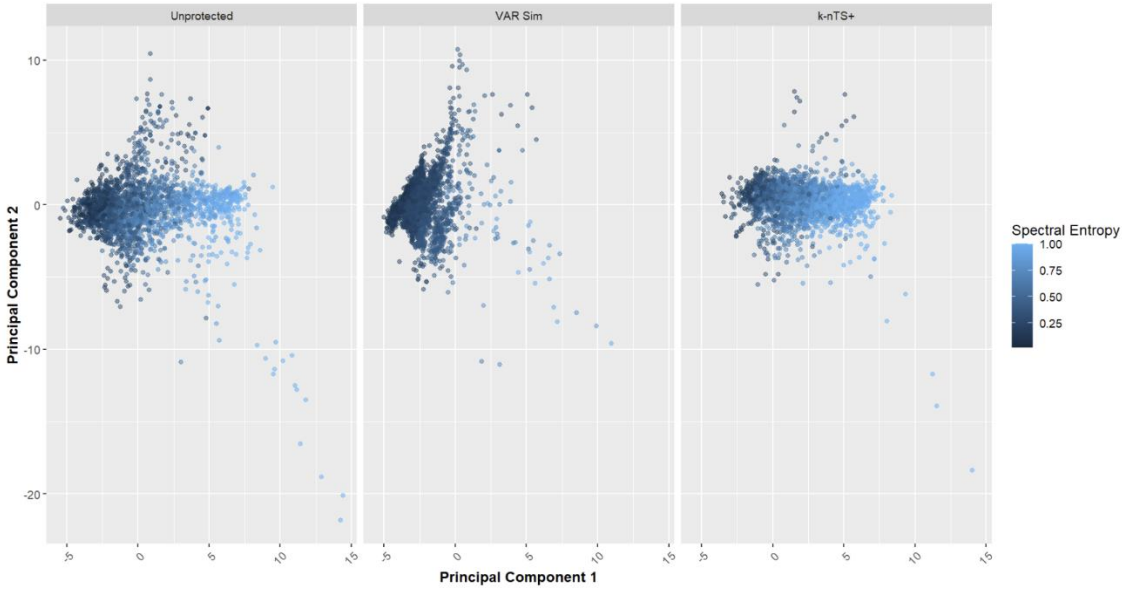
Thank you. We now mention other potential applications of our proposed swapping method, including outlier replacement, imputing missing values, and nowcasting in the last paragraph of the conclusion on page 41.

- The authors should explain why the data owner cannot provide the forecaster with the original or degraded model weights. This comparison would help illustrate the advantages and limitations of the proposed approach.

Thank you for the suggestion. To address the question on whether the data owner can provide the forecasters with the original model weights, we save the parameters of the VAR models trained on the M3 data and generated a simulated time series corresponding to each original series.

We show that the forecast accuracy of such an approach provides better forecast accuracy than all of the privacy methods we study, including *k*-nTS+, when applied to the original data. However, the forecasts from this approach compromise the privacy of the unprotected series. In short, we can rarely have accurate forecasts and privacy on the original time series scale, and time series must be normalized to increase the cross-series similarity of time series values, features, and forecasts before an acceptable trade-off between privacy and forecast accuracy can consistently be achieved. These results are shown in Section 5.5 on pages 29-30.

We also show an advantage of our methodology is that through sharing protected time series, forecasters gain access to a more representative distribution of time series values. The figure below compares the distribution of time series values for the unprotected data, time series simulated using VAR model weights, and the protected series from the *k*-nTS+ method.



Additionally, discussing whether this method could be extended or applied to cases where data are owned by multiple data owners (decentralized scenario) would be beneficial.

Thank you for the comment. We discuss this scenario in the second to last paragraph on page 41 in the conclusion.

- A crucial aspect missing is the impact of the swapping process on forecasting accuracy in a real-world setting. For instance, if a forecaster aims to perform a one-timestep ahead forecast using a VAR (Vector Autoregression) model with a lag of 1, he would require the value of X[T] to predict X[T+1]. Therefore, it is essential to evaluate whether the swapping process can change the last point of the time series without significantly affecting the forecasting accuracy.

To clarify from the previous version of the paper, our existing swapping process changed all past time series values including X[T]. T values for each time series were swapped at time T when the data owner decided to protect all its past data.

To address your comment, we also used the shared VAR model weights mentioned in the point above to forecast using shared lags protected using both additive noise and differential privacy. The forecast accuracy using the model weights from the unprotected data and protected lags was quite poor. We include these results with the results from simulating time series using the shared weights in Section 5.5. For your reference, we include a table showing the accuracy and disclosure risks $\bar{P}$ (identification disclosure risk) and $\overline{P^f}$ (forecast disclosure risk).

Privacy Method	Shared Data	Shared Model	$\bar{P}$	$\bar{P}^f$	Accuracy (% Change MAE)
Unprotected	Full Data	No	98.4%	18.36%	418.01
k-nTS+ (k = 3, M = 1.5)	Full data	No	8.00%	4.43%	848.16 (102.91%)
AN (s = 0.25)	Lags only	Yes (VAR)	*70.89%	8.59%	882.64 (111.16%)
k-nTS+ (k = 3, M = 1.5)	Lags only	Yes (VAR)	*8.00%	3.30%	2910.43 (596.27%)
VAR-simulation	Lags only	Yes (VAR)	9.70%	17.99%	508.70 (21.70%)

Table 10: Disclosure risks and VAR model accuracy for the unprotected, k-nTS+ ($k = 3, M = 1.5$), and VAR-based approaches applied to the original M3 data. The * values indicate that the disclosure risk value from the privacy analysis on the full data, *i.e.*, the average $\bar{P}$ if an adversary collect ten lagged values over time.

Figure 6 in the paper shows that the last point in the window is very close to the original one. Does this mean that a curious forecaster could reconstruct the data by running the model for some time?

Thank you for this comment. Reconstruction attacks are a valid concern with non-differentially private protection mechanisms. In our paper, we measured the reidentification risk of an intruder which is only one form of a privacy attack. We simulated an intruder matching 10 time series values in a row to the protected dataset and found that our method performs quite well (less than 9% reidentification risk). In our Conclusion section on pages 40-41, we now note the limitation that we only considered one privacy attack and that it's possible for an intruder to attack the data in another way.

However, with the theoretically private method of differential privacy, at acceptable levels of $\epsilon \leq \ln 3$, we showed that the forecasts are unusable. When increasing ϵ , the forecast accuracy improved while maintaining some bound on overall privacy risk. However, our empirical application shows that $\epsilon \geq 4.6$ results in significantly higher identification disclosure risk (40% and 44% for $\epsilon = 10$ on the original and rate M3 data sets) than our proposed method *k*-nTS+. Essentially, the differentially private data is either not useful or not protective and there is no acceptable trade-off between data privacy and forecast accuracy.

Detailed comments:
- Acronyms meaning is missing. Some examples: SES, DES, LGBM, OOB, MSE, MAE, etc.
- Equations should not be figures, e.g., (1) and (5).
- Figures 2 and 8 should be of better quality.
- Figure 4: I suggest a unique plot with different color/shape lines representing the methods.
- Notation needs to be introduced appropriately in many equations - see, e.g., (9).

Thanks for these comments. We have corrected all of them in the paper.

- I would include the detailed proposal version in the main text.

We have included detailed versions of the *k*-nTS swapping, the machine learning based feature selection, and *k*-nTS+ privacy algorithms in the main text on pages 10, 12, and 14.

- Text needs revision. Typo example: "matices", sometimes ':' is used instead of '='.

Thank you. We have carefully edited the entire paper. We hope that the revised manuscript addresses your concerns, and we sincerely appreciate your enhancements to our paper.