

# Can We Protect Time Series Data While Maintaining Accurate Forecasts?

Cameron D. Bale, Matthew J. Schneider, Jinwook Lee

*Drexel University Lebow College of Business*

*3220 Market St, Philadelphia, PA 19104*

*Correspondence: Cameron D. Bale*

*3141 Chestnut St., Philadelphia, PA 19104, Office 734*

*Email: cdb327@drexel.edu*

---

## Abstract

In the context of increasing data breaches and privacy concerns, we focus on the often-neglected goal of maintaining the usefulness of protected data by exploring the intersection of data privacy and time series forecasting. Using both simulated and real-world time series data sets, we test various privacy methods, including a proposed swapping-based method ( $k$ -nTS+) designed to maintain time series features, a differentially private method, and an approach based on sharing model weights trained on unprotected data. Based on forecasts from seven different forecasting models, none of the privacy methods based on swapping or random noise addition can consistently maintain forecast accuracy at an acceptable level of privacy. 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 transforming time series into rates increases the similarity of time series features, values, and forecasts, and enables  $k$ -nTS+ to produce protected rate time series with a reduction in forecast accuracy of just 3.6% on average. Overall, this paper shows that except under certain conditions, creating protected time series with acceptable privacy is incompatible with obtaining accurate forecasts.

**Keywords:** data privacy; time series features; machine learning; forecast accuracy; identification disclosure risk

---