

Evaluating the Relationship Between Temporal Aggregation and Forecast Bias in Statistical Time Series Models

Claire Cook, Colton Phillips, Connor Edwards

1 Introduction

The practice of temporal aggregation represents a fundamental preprocessing step in time series analysis, routinely employed to reduce noise, manage computational requirements, and align data with decision-making timeframes. Despite its widespread application across domains including economics, environmental science, and operations management, the systematic effects of aggregation choices on forecast accuracy remain inadequately characterized. Current literature predominantly focuses on aggregation’s impact on forecast variance and computational efficiency, while largely overlooking its nuanced relationship with forecast bias. This research gap is particularly consequential given that biased forecasts can lead to suboptimal decisions with significant real-world implications, from inventory mismanagement in supply chains to inefficient resource allocation in energy systems.

Our investigation addresses this gap by examining how different levels of temporal aggregation systematically influence forecast bias across multiple statistical modeling approaches. We challenge the conventional assumption that bias remains relatively constant across aggregation levels and instead propose that aggregation induces non-linear transformations in the data generating process that fundamentally alter forecast distributions. The novelty of our approach lies in developing a comprehensive framework for bias decomposition that separates aggregation effects from inherent model limitations, enabling practitioners to make more informed decisions about temporal resolution selection.

This research is motivated by three primary questions that have received limited attention in existing literature: First, does a systematic relationship exist between temporal aggregation level and forecast bias across different statistical models? Second, can we identify optimal aggregation ranges that minimize bias for specific data characteristics? Third, how does aggregation interact with different components of time series (trend, seasonality, noise) to produce biased forecasts? By addressing these questions through rigorous empirical analysis, we contribute to both theoretical understanding and practical application of time series forecasting.

2 Methodology

Our methodological approach employs a multi-scale analytical framework designed to systematically evaluate the relationship between temporal aggregation and forecast bias. We collected fourteen diverse time series datasets spanning economic indicators, environmental measurements, energy consumption patterns, and web traffic metrics. Each dataset contained high-frequency observations (minute or hour level) covering at least three years, enabling aggregation to twelve distinct temporal levels: hourly, 3-hourly, 6-hourly, 12-hourly, daily, 2-day, weekly, 2-week, monthly, quarterly, semi-annual, and annual.

For each aggregation level, we implemented six statistical forecasting models representing different methodological approaches: ARIMA (AutoRegressive Integrated Moving Average), ETS (Error, Trend, Seasonal), TBATS (Trigonometric seasonality, Box-Cox transformation, ARMA errors, Trend, and Seasonal components), Theta method, Simple Exponential Smoothing, and Naive forecasting. This selection ensures coverage of both sophisticated and baseline approaches commonly used in practice. Each model was trained on the first 80

A key innovation in our methodology is the development of a bias decomposition metric that separates total forecast error into three components: aggregation-induced bias, model-specific bias, and irreducible error. The aggregation-induced bias is calculated by comparing forecasts from the same model across different aggregation levels, while controlling for the underlying data characteristics. We introduce a novel normalization procedure that accounts for scale differences across aggregation levels, enabling direct comparison of bias metrics.

Our analysis employs multiple bias measures including Mean Error (ME), Mean Percentage Error (MPE), and a newly developed Aggregation Bias Index (ABI) that quantifies the proportion of total error attributable to aggregation choices. Statistical significance of bias differences across aggregation levels is assessed using repeated measures ANOVA and post-hoc pairwise comparisons with Bonferroni correction. Additionally, we examine how data characteristics including seasonality strength, trend magnitude, noise-to-signal ratio, and structural breaks moderate the relationship between aggregation and bias.

3 Results

The empirical analysis reveals several significant findings that challenge conventional practices in temporal aggregation for forecasting. First, we identify a consistent U-shaped relationship between aggregation level and forecast bias across most datasets and models. Both extremely fine (hourly) and extremely coarse (annual) aggregations produce significantly higher bias compared to intermediate levels (weekly to quarterly), with the exact optimal range varying by data characteristics. For economic time series with strong seasonality, the minimal bias occurred at monthly aggregation, while for environmental data with smoother patterns, weekly aggregation produced the least biased forecasts.

Second, our bias decomposition analysis demonstrates that aggregation-induced bias accounts for 18-42

Third, we find that the interaction between data characteristics and aggregation level follows predictable patterns. Time series with strong seasonal patterns experience minimal bias at aggregation levels that preserve at least two complete seasonal cycles, while series with structural breaks show increased sensitivity to aggregation, with finer resolutions generally performing better in capturing breakpoints. The noise-to-signal ratio emerges as a critical moderating variable: noisier series benefit more from moderate aggregation (reducing noise-induced bias), while cleaner series tolerate a wider range of aggregation levels.

Fourth, our analysis reveals that conventional forecast accuracy measures like MAPE (Mean Absolute Percentage Error) and RMSE (Root Mean Square Error) often mask aggregation-induced bias, as they conflate bias and variance components. When bias is examined separately, the ranking of optimal aggregation levels changes substantially in 67

4 Conclusion

This research provides compelling evidence that temporal aggregation choices significantly influence forecast bias in ways that have been largely overlooked in both academic literature and practical applications. The identification of a systematic U-shaped relationship between aggregation level and bias represents a fundamental contribution to time series forecasting theory, suggesting that both excessive disaggregation and over-aggregation can introduce substantial systematic errors. Our findings challenge the common practice of selecting aggregation levels based solely on data availability or computational convenience, instead advocating for deliberate consideration of bias implications.

The practical implications of this research are substantial across multiple domains. In supply chain management, our results suggest that weekly or bi-weekly aggregation may provide more unbiased forecasts than the commonly used daily or monthly levels. In energy forecasting, the optimal aggregation for load prediction depends critically on the balance between seasonal patterns and random fluctuations. Financial applications requiring volatility forecasting should consider that high-frequency aggregations may systematically overestimate risk due to aggregation-induced bias.

Several limitations of the current study suggest directions for future research. Our analysis focused on univariate time series, while many practical applications involve multivariate relationships that may interact differently with aggregation. The examination of more complex forecasting approaches including machine learning methods would extend the generalizability of our findings. Additionally, theoretical work developing formal mathematical explanations for the observed U-shaped bias pattern would strengthen the conceptual foundations uncovered in this empirical investigation.

In conclusion, this research establishes temporal aggregation as a critical de-

terminant of forecast bias that warrants careful consideration in both research and practice. By providing empirical evidence of systematic aggregation effects and developing methodologies for bias decomposition, we contribute to more nuanced understanding of forecasting performance and more informed decision-making in temporal resolution selection. The relationship between aggregation and bias represents a fertile area for future investigation with significant potential to improve forecasting accuracy across diverse applications.

References

- Armstrong, J. S. (2001). Principles of forecasting: A handbook for researchers and practitioners. Kluwer Academic Publishers.
- Athanasopoulos, G., Hyndman, R. J., Song, H., Wu, D. C. (2011). The tourism forecasting competition. *International Journal of Forecasting*, 27(3), 822-844.
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C. (2015). Time series analysis: Forecasting and control. John Wiley Sons.
- De Gooijer, J. G., Hyndman, R. J. (2006). 25 years of time series forecasting. *International Journal of Forecasting*, 22(3), 443-473.
- Hyndman, R. J., Athanasopoulos, G. (2018). Forecasting: principles and practice. OTexts.
- Makridakis, S., Spiliotis, E., Assimakopoulos, V. (2018). The M4 Competition: Results, findings, conclusion and way forward. *International Journal of Forecasting*, 34(4), 802-808.
- Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M. Z., Barrow, D. K., Ben Taieb, S. (2022). Forecasting: theory and practice. *International Journal of Forecasting*, 38(3), 705-871.
- Silvestrini, A., Veredas, D. (2008). Temporal aggregation of univariate and multivariate time series models: A survey. *Journal of Economic Surveys*, 22(3), 458-497.
- Tiao, G. C. (1972). Asymptotic behaviour of temporal aggregates of time series. *Biometrika*, 59(3), 525-531.
- Wei, W. W. S. (2006). Time series analysis: Univariate and multivariate methods. Pearson Education India.