JOS

Journal of Official Statistics
2024, Vol. 40(4) 499–507
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0282423X241299913
journals.sagepub.com/home/jof
Sage

Preface to Time Series Special Issue of the Journal of Official Statistics

William R. Bell[1], Tommaso Di Fonzo[2], Dominique Ladiray[3], Tucker S. McElroy[1] and Paul A. Smith[4]

Keywords
crisis, disaggregation, high-frequency, meager data, outliers, seasonality

1. Introduction

This is a challenging era for the publication and analysis of time series data by official statistics agencies. Since the beginning of the twenty-first century and the onset of the information age, coupled with advances in data science and artificial intelligence, the demand for more timely data—with regard to release dates, as well as with regard to publication frequency—has challenged the resources of government agencies; see Ollech (2021) for a discussion of daily data. Concurrent with this global societal trend is the decline of survey response rates (Jarmin 2019), which is partially due to increasing distrust about surveys, and this degradation of response to repeated surveys impacts the resulting time series of survey estimates. Offsetting these challenges are the benefits of new technologies, both physical—allowing for new modes of data collection and curation—and methodological. Statisticians and data scientists at official statistics agencies have responded with research and development that aims to harness innovations in mathematical statistics, computer

[1]Directorate for Research and Methodology, U.S. Census Bureau, Washington, DC, USA
[2]Department of Statistical Sciences, University of Padova, Padova, Italy
[3]Institut National de la Statistique et des Études Économiques, Paris, France
[4]University of Southampton, Southampton, UK

Corresponding author:
Tucker S. McElroy, Directorate for Research and Methodology, U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20233-9100, USA.
Email: tucker.s.mcelroy@census.gov

science, and operations research in order to address the double onus of timeliness and survey degradation (Schmutte and Vilhuber 2022).

The timeliness challenge is developed through two main aspects: first, the search for higher frequency publication rates (e.g., providing data at weekly or daily rates as opposed to just quarterly or monthly); second, the publication of advance estimates (constructed from incomplete data or through nowcasting). These two aspects can be intertwined. Connected with this timeliness challenge are the related problems of temporal disaggregation, seasonal adjustment, and forecasting.

The non-response issue has prompted some to seek alternative sources for quality data, including administrative records (Jarmin and O'Hara 2016) as well as data vendors. Whereas the former requires a careful review of privacy concerns (Lagoze et al. 2013), the latter raises concerns of cost, as well as data quality and longevity (i.e., is the data vendor trustworthy, and will they stay in business); see Landefeld (2014). Another avenue of work looks to new technologies (such as tablets, or the use of smart phone apps) and mechanisms (the use of incentives, or the analysis of paradata) to remedy the response deficit; see Harris-Kojetin and Groves (2017) for an overview of these challenges. Such approaches introduce new concerns about the time series integrity and continuity. Related to this is the topic of privacy (Abowd 2016), which is in tension with data utility and granularity.

Within the current era's context, more classical questions and goals of time series analysis remain. Economists and demographers, among others, are still interested in determining turning points and trends for official time series (Trimbur and McElroy 2022). Such goals are impacted by the challenges of timeliness and non-response: the discernment of a trend's movement can be degraded by survey discontinuity (Van den Brakel et al. 2020), or can be enhanced by the usage of higher frequency indicators. The improvement of computational resources—the power of computing environments, the movement to cloud ecosystems, and the rise of the open source zeitgeist—facilitates innovative methodologies that address classical applications in the modern context.

Thrown into this exciting and tumultuous mix are global or national catastrophes, the most recent example being the Covid-19 pandemic, but also including the Great Recession. More broadly, the global linkage of the world's economies has world-wide ramifications, wherein local shocks—due to war or natural disasters—percolate throughout the entire terrestrial system. Such realities have driven deeper studies into extreme-value phenomena (Roy and McElroy 2024), and the modeling of outliers, aberrations, and discontinuities in time series data.

Broadly, we might therefore classify the time series challenges for official statistics into six overlapping categories:

1. Modeling and seasonal adjustment of higher frequency time series
2. Crisis identification and modeling
3. Computation and software
4. Detection of trends and turning points
5. Nowcasting and publication of advance estimates

Table 1. Special Issue Articles by Topic.

Article	High frequency	Crises and extremes	Computation and software	Trends and turning points	Nowcasting and advance estimates	Survey changes
Acedański (2024)	✓				✓	
Bisio (2024)	✓					
Quilis and Cuevas (2024)	✓	✓				
Smith et al. (2024)						✓
Chen et al. (2024)				✓		
Webel and Smyk (2024)	✓	✓	✓	✓		
Zalsha and Wolter (2024)						✓
Livsey and McElroy (2024)			✓			
Quartier-la-Tente (2024)		✓		✓		
Joyce and McElroy (2024)						✓
Zajac et al. (2024)			✓		✓	

6. Survey challenges, such as discontinuity of estimates and increasing non-response

These topics are of current and future interest, and diverse facets are discussed in the papers of this special issue.

2. Summary of Articles

As well as a review (Mathews 2024) on the topic of statistical learning, there are eleven research papers in this special issue, which cover the above six chief topics with a varying range of styles. Some papers are more methodological, others are focused on empirical problems, or on software, and others provide a review of the state-of-the-art. Table 1 provides a classification of the papers by the topics that they address.

First, Acedański (2024) proposes a simple but effective smoothing method for the spatio-temporal disaggregation of economic time series that can be used for countries lacking long and reliable series of regional economic indicators. As the method can be applied sequentially, the number of estimates to be revised is reduced when new aggregated data are available, which is a convenient feature when historical estimates are of interest. In addition, the smoothing procedure can be applied in nowcasting situations with scarce data. An application to quarterly disaggregation of annual real GDP for Polish regions offers interesting results compared with standard regression-based disaggregation procedures.

Next, Bisio (2024) reports on the reconstruction of the Italian time series on hours worked, a short-term statistic used in labor input estimates by the Quarterly National Accounts. The updated indicator is estimated through the Kalman filter and smoother algorithms applied to a state-space representation of a multivariate structural model. The assessment of the performance of the new indicator against the non-updated one—based on estimates of quarterly per-employee hours worked—provided smaller forecast Mean Absolute Errors in more than 60% of the series, with a positive impact for both producer and users of this valuable economic information.

Quilis and Cuevas (2024) explore and assess the impact of the COVID-19 shock on the tasks of modeling, decomposing, and seasonally adjusting the Spanish daily sales data compiled by the Tax Agency. This assessment, based on a structural time series approach, helps the understanding of how to deal with an unprecedented shock via available methods and software. The results are promising, in providing a timely perspective on the evolution of the shock (a tent-shaped customized regressor is used) and the challenges that the COVID-19 pandemic poses for the modeling and seasonal adjustment of high-frequency economic data.

Smith et al. (2024) examine the challenge of estimating and potentially adjusting for a time series discontinuity—in the form of a gradual level shift—arising due to a transition in the survey instrument. Structural time series modeling is used to discriminate between changes arising from the survey instrument and those due to real changes in the underlying series. First an assessment of the power (of the implementation plan for the new instrument) to detect discontinuities of a certain size is determined through simulation. Then a progressive analysis is undertaken as more data become available. The discontinuity and the real evolution of the series are affected by changes to travel and migration around Brexit.

Next, Chen et al. (2024) are concerned with predicting changes in the U.S. economy, focusing on business cycle fluctuations in real GDP and eleven major aggregates from the U.S. national income and product accounts. The proposed approach capitalizes on the link between cyclical models and band-pass filters, which represents an area of ongoing development in econometrics and statistics. It is shown that the adaptive model-based Butterworth and model-based balanced filters perform better than the commonly used Hodrick-Prescott and Baxter-King filters for analyzing and interpreting the properties and co-movements in business cycle components across the set of analyzed U.S. series, mainly because these filters are implicitly defined by the models and can avoid generating spurious cyclical effects by adapting the filters to the data.

Webel and Smyk (2024) focus on the software capabilities of JDemetra+ for seasonal adjustment of infra-monthly time series, such as weekly, daily, and hourly time series data that may exhibit outliers, meager values (i.e., the presence of zeroes and small values), and multiple fractional seasonal periods (e.g., weekly periodicity, where the number of weeks in a year is 52.1775). For adjustment of such series, JDemetra+ provides the STL approach along with extensions to infra-monthly series of the ARIMA model-based and X-11 approaches. This software is written in R, adopting the open source philosophy to code development. The paper provides a comprehensive introduction to this versatile and powerful computer program, with illustrations using demographic and economic time series.

Zalsha and Wolter (2024) consider modification of direct survey estimates for a population characteristic that is nondecreasing over time, such as the accumulated number of influenza cases in a flu season. The paper uses the Pool Adjacent Violators Algorithm to modify the series of standard survey estimates to be a nondecreasing series. The focus of the paper involves statistical inference for this situation via confidence intervals. Results from a simulation study compare the performance of four alternative approaches to the construction of the confidence intervals.

Next, Livsey and McElroy (2024) propose a new computational technique that expands the practical applicability of difference stationary latent component models. At the present time, estimating parameters with maximum likelihood estimation has proven numerically infeasible for even moderate dimensional problems. The paper provides methodology for applying the Expectation-Maximization algorithm to a class of latent component multivariate time series models that allow for a nuanced specification of the unobserved signal. An explicit formula for the maximization step is proposed, which facilitates computation speed while also improving the stability of the algorithm. This new result renders feasible the fitting of a six-dimensional daily time series, which previously was intractable. Although the structural time series models employed are quite simple, the new method shows promise for circumventing the challenge of parameter estimation—and model-based seasonal adjustment—for multivariate time series.

Quartier-la-Tente (2024) examines the problem of turning-point detection for trend-cycle dynamics, utilizing the methodology of moving averages constructed by local polynomial regression, and optimized for timeliness (i.e., to minimize phase delay). The method provides an extension of the Henderson and Musgrave asymmetric filters used in the X-13ARIMA-SEATS seasonal adjustment software. Also, the paper shows that local parametrization reduces the delay in detecting turning points while also reducing revisions.

Joyce and McElroy (2024) advance a methodology for survey time series wherein an agency seeks to publish the data at alternative frequencies, or even according to bespoke epochs of time. Since the temporal periods are arbitrary, the method is developed through the application of change of support ideas to a time series context, utilizing a continuous-time model for the underlying population process. The paper demonstrates how the use of continuous-time processes can be helpful for customizing flow estimates, while also highlighting the need for further research into fitting and estimation.

Next, Zajac et al. (2024) review some of the challenges facing official statistics agencies—in particular the need for timely data—and discuss the outcome of a European nowcasting competition. The motivation for the competition was to identify nowcasting methods with potential to improve official statistics, and there were therefore prizes for both accuracy and reproducibility. The paper describes the competition parameters, and also how the scoring may have had some unintended consequences. The competition did not directly identify methods for use in official statistics, but did generate several ideas for approaches for more detailed testing. Some ideas for future competitions are discussed.

3. Future Directions

The articles of this special issue identify several directions for future research and development. In addressing the challenges of high frequency data, temporal disaggregation can be accomplished through using auxiliary time series together with temporal (or spatial) modeling, as in Acedański (2024) and Bisio (2024); this is likely to be a topic of continuing development. With fractional seasonal periods (as occurs with weekly time series) there are some remaining challenges in the construction of time series models, including the determination of unit root differencing polynomials, as acknowledged and discussed in Webel and Smyk (2024). With so many seasonal frequencies to consider, models must balance parsimony against the ability to capture every seasonal periodicity.

The presence of new forms of crisis, and their impact on official time series, prompts innovation in the form of constructed outliers (as in Quilis and Cuevas (2024)). With high-frequency time series and multivariate analyses, the identification of outliers and extremes becomes more computationally challenging, and the interplay of extreme values with temporal aggregation—in the context of official statistics—is yet to be deeply developed.

While the advance of computing resources facilitates data supplementation (e.g., using administrative records to partially remedy the defect of survey nonresponse) and augmentation (e.g., using related time series to improve nowcasts or seasonal adjustment), such applications require further methodological advances. New and better algorithms are needed for combining data sources, for analyzing high-frequency time series, and for fitting high-dimensional time series models (as is emphasized in Livsey and McElroy (2024)). Here there may be a role for insights from the machine learning literature, which seeks to analyze Big Data without the luxury of carefully constructed models.

There will continue to be interest in trends, turning points, and the business cycle, as highlighted in Chen et al. (2024) and Quartier-la-Tente (2024). The future challenge is to achieve these classical goals of signal extraction analysis in the modern context of Big Data. While model-based filtering has clear advantages, facilitating fitting for millions of (potentially high frequency) time series is an ongoing challenge, particularly when manpower resources mitigate against a human review of all modeling diagnostics. A related research thread involves nowcasting and real-time turning point detection, where filters are optimized for timeliness and accuracy.

As highlighted in Zajac et al. (2024), the generation of advance estimates is a rich area for research: what are the leading indicators that can assist with nowcasting, and what statistical methodologies require development to meet that goal? A related direction of research is on indices, which compress time series data so as to furnish a data summary; it is desirable that such indices would be timely, and possibly give advance signals about the economy and society. The methodology of index design yet has open problems, such as the construction of an optimal compression.

Finally, there are significant and enduring challenges with survey time series, such as a proper quantification of survey uncertainty; one such challenge is the extension to complex sampling of the results of Zalsha and Wolter (2024) on confidence intervals for repeated survey estimates of population quantities that are nondecreasing over time. The promise of continuous-time models, discussed in Joyce and McElroy (2024), invites deeper investigation into fitting and estimation. The important problem of survey discontinuity (Smith et al. 2024) has many novel facets, including testing for concordance of "broken" time series, linking (i.e., rectification of breaks), and discriminating between fundamental level shifts and breaks due to changes in survey instrument.

One area not touched upon by the above articles is time series privacy and confidentiality, but this is a new and important topic. The key challenge is to still give privacy guarantees for published data while also preserving the utility of the time series data, and avoid degradation of serial correlation patterns that occurs from noise infusion. This might be accomplished by utilizing noise that has the same spectral character as the sensitive data; alternatively, other mechanisms such as all-pass filtering (McElroy et al. 2023) could be used.

This is an exciting time for statisticians in general, but especially so for those working in government agencies. The articles of this special issue showcase a broad spectrum of problems and solutions, which should serve to encourage and inspire future research endeavors while also improving the capabilities of those engaged in production.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Disclaimer

This report is released to inform interested parties of research and to encourage discussion. The views expressed on statistical issues are those of the authors and not those of the U.S. Census Bureau.

ORCID iDs

Tucker S. McElroy https://orcid.org/0000-0002-2991-9067
Paul A. Smith https://orcid.org/0000-0001-5337-2746

References

Abowd, J. M. 2016. "How Will Statistical Agencies Operate When All Data Are Private?" *Journal of Privacy and Confidentiality* 7 (3): 1–15. DOI: https://doi.org/10.29012/jpc.v7i3.404.

Acedański, J. 2024. "Disaggregation and Nowcasting of Regional GDP Series with a Simple Smoothing Algorithm." *Journal of Official Statistics* (in this issue). DOI: https://doi.org/10.1177/0282423X241277716.

Bisio, L. 2024. "Reconstructing a Short-Term Indicator by State-Space Models: An Application to Estimate Hours Work by Quarterly National Accounts." *Journal of Official Statistics* (in this issue). DOI: https://doi.org/10.1177/0282423X241240366.

Chen, B., K. Hood, T. McElroy, and T. Trimbur. 2024. "Cycle Fluctuations in the U.S. Real GDP and National Income and Product Accounts Aggregates." *Journal of Official Statistics* (in this issue). DOI: https://doi.org/10.1177/0282423X241263763.

Harris-Kojetin, B. A., and R. M. Groves, eds. 2017. *Innovations in Federal Statistics: Combining Data Sources While Protecting Privacy*. National Academies Press. DOI: https://doi.org/10.17226/24652.

Jarmin, R. S. 2019. "Evolving Measurement for an Evolving Economy: Thoughts on 21st Century US Economic Statistics." *Journal of Economic Perspectives* 33 (1): 165–84. DOI: https://doi.org/10.1257/jep.33.1.165.

Jarmin, R. S., and A. B. O'Hara. 2016. "Big Data and the Transformation of Public Policy Analysis." *Journal of Policy Analysis and Management* 35 (3): 715–21. https://www.jstor.org/stable/43867163.

Joyce, P., and T. McElroy. 2024. "Modeling Survey Time Series Data with Flow-Observed CARMA Processes." *Journal of Official Statistics* (in this issue). DOI: https://doi.org/10.1177/0282423X241286236.

Lagoze, C., W. C. Block, J. Williams, J. M. Abowd, and L. Vilhuber. 2013. "Data Management of Confidential Data." 8th International Digital Curation Conference, Amsterdam, January 14–16. https://hdl.handle.net/1813/30924.

Landefeld, S. 2014. "Uses of Big Data for Official Statistics: Privacy, Incentives, Statistical Challenges, and Other Issues." International Conference on Big Data for Official Statistics, Beijing, China, October 28–30.

Livsey, J., and T. McElroy. 2024. "Applying the Expectation-Maximization Algorithm to Multivariate Signal Extraction." *Journal of Official Statistics* (in this issue). DOI: https://doi.org/10.1177/0282423X241271471.

Mathews, S. 2024. "Review of 'Statistical Learning for Big, Dependent Data.'" *Journal of Official Statistics* (in this issue).

McElroy, T., A. Roy, and G. Hore. 2023. "FLIP: A Utility Preserving Privacy Mechanism for Time Series." *Journal of Machine Learning Research* 24 (111): 1–29. http://jmlr.org/papers/v24/22-0734.html.

Ollech, D. 2021. "Seasonal Adjustment of Daily Time Series." *Journal of Time Series Econometrics* 13 (2): 235–64. DOI: https://doi.org/10.1515/jtse-2020-0028.

Quartier-la-Tente, A. 2024. "Improving Real-Time Trend Estimates Using Local Parametrization of Polynomial Regression Filters." *Journal of Official Statistics* (in this issue). DOI: https://doi.org/10.1177/0282423X241283207.

Quilis, E., and Á. Cuevas. 2024. "The Covid-19 Shock Through the Lens of the Spanish Sales Daily Data: Modeling and Seasonal Adjustment Challenges." *Journal of Official Statistics* (in this issue). DOI: https://doi.org/10.1177/0282423X241285985.

Roy, A., and T. McElroy. 2024. "Modeling Extreme Events in Time Series and Their Impact on Seasonal Adjustment in the Post-Covid-19 Era." *Bayesian Analysis* 1 (1): 1–25. DOI: https://doi.org/10.1214/24-BA1424.

Schmutte, I., and L. Vilhuber. 2022. "An Interview with John M. Abowd." *International Statistical Review* 90 (1): 1–40. DOI: https://doi.org/10.1111/insr.12489.

Smith, P. A., J. van den Brakel, and G. Horsfield. 2024. "Estimation of Discontinuities from the Introduction of Tablet-Based Data Collection on the International Passenger Survey." *Journal of Official Statistics* (in this issue). DOI: https://doi.org/10.1177/0282423X241275836.

Trimbur, T., and T. McElroy. 2022. "Modeled Approximations to the Ideal Filter with Application to Time Series of Gross Domestic Product." *The Annals of Applied Statistics* 16 (2): 627–51. DOI: https://doi.org/10.1214/21-AOAS1463.

Van den Brakel, J., X. Zhang, and S.-M. Tam. 2020. "Measuring Discontinuities in Time Series Obtained with Repeated Sample Surveys." *International Statistical Review* 88 (1): 155–75. DOI: https://doi.org/10.1111/insr.12347.

Webel, K., and A. Smyk. 2024. "Seasonal Adjustment of Infra-Monthly Time Series with JDemetra+." *Journal of Official Statistics* (in this issue). DOI: https://doi.org/10.1177/0282423X241277602.

Zajac, A., M. Karlberg, and J.-M. Museux. 2024. "The European Statistics Awards for Nowcasting: A New Approach to Engage with the Scientific Community and to Stimulate Improved Nowcasting for Official Statistics." *Journal of Official Statistics* (in this issue). DOI: https://doi.org/10.1177/0282423X241274614.

Zalsha, S., and K. M. Wolter. 2024. "A Monte Carlo Investigation of Confidence Intervals for a Nondecreasing Series." *Journal of Official Statistics* (in this issue). DOI: https://doi.org/10.1177/0282423X241287663.

Received: October 22, 2024
Accepted: October 24, 2024