SOCIOECONOMIC ANALYTICS

Parametric and non-parametric data-driven analytics for socioeconomic challenges in a contemporary world

Thyago C. C. Nepomuceno [1] [2]

[1] Department of Statistics, Federal University of Pernambuco, Recife, Brazil
[2] Dep. of Computer, Control, and Management Engineering Antonio Ruberti, Sapienza University of Rome, Rome, Italy.
Email: thyago.nepomuceno@ufpe.br; nepomuceno@diag.uniroma1.it

How to cite this paper: Nepomuceno, T. C. C. (2023). Data-driven Analytics for Socioeconomic Challenges in a Contemporary World, *Socioeconomic Analytics,* 1(1), 1–4. https://doi.org/10.51359/2965-4661.2023.259300

EDITORIAL

Socioeconomic Analytics
https://periodicos.ufpe.br/revistas/SECAN/
ISSN Online: 2965-4661

Submitted on: 29.07.2023.
Accepted on: 29.07.2023.
Published on: 30.07.2023.

Abstract

The rapid advancement of artificial intelligence and data-driven technologies have presented novel opportunities for addressing complex socioeconomic challenges in the contemporary world. By harnessing diverse datasets and employing sophisticated analytical techniques, researchers, policymakers, and practitioners can gain profound insights into the root causes of socioeconomic challenges and devise innovative strategies to promote inclusive development and sustainable growth. In this editorial, I share some perspectives on exploring the application of non-parametric and parametric data-driven analytics methodologies (Data Envelopment Analysis and Econometric Models) as powerful tools for understanding and resolving multifaceted issues that impact societies globally, and what I believe to be the future of these methods. Readers of *Socioeconomic Analytics* can find those interesting methodologies as empirical applications in the journal's inaugural and forthcoming issues.

Keywords

Data Analytics; Econometrics; Operations Research; Data Envelopment Analysis; Time Series Analysis; Regression models; Natural Language Processing; Sentiment Analysis.

1. Introduction

Socioeconomic challenges continue to pose critical obstacles to human development, social cohesion, and economic progress in the post-pandemic global landscape (Ciuffetelli Parke & Conversano, 2021). As scholars and analysts, committing to a comprehensive examination of the pressing challenges facing society today is imperative. The interplay of socioeconomic challenges is intricate and multifaceted. We cannot overlook the persistent issues of crime and violence, income, healthcare and education inequality, and business inefficiency which

exacerbates social disparities, limits upward mobility, and fuels societal unrest. The impact of globalization, technological advancements, and automation presents a new set of challenges, including job displacement, skill gaps, and the uneven distribution of the benefits of progress.

Multidimensional Big-data approaches are needed and in constant development to overcome socioeconomic challenges that confront us (Blazquez & Domenech, 2018). Collaboration between governments, academia, civil society, and the private sector is critical for implementing effective policies and interventions. In the following section, I present two avenues of quantitative options to cope with the most recurrent socioeconomic issues. Non-parametric frontier estimations can be used for measuring productive capacity and efficiency, while parametric econometric estimations can identify possible relationships among different variables. Readers of the journal *Socioeconomic Analytics* will be able to find some of them applied in engaging assessments over the forthcoming volumes and issues highlighting significant socioeconomic challenges in our contemporary world and emphasizing the urgent need for robust analysis and innovative solutions.

2. Selected Data-driven Methodologies

This section explores two common but distinct data-driven approaches for conducting many data analyses, offering unique insights into evaluating socioeconomic problems and trends. In addition, I offer a short discussion on what I believe to be the future of each type of analysis. These approaches provide valuable perspectives on efficiency and performance, classifications, identification of trends and exogenous components, and assessment of causality in data, forecasts and predictions, catering to different data types and research objectives.

2.1. Data Envelopment Analysis

Data Envelopment Analysis (DEA) is a powerful and widely used mathematical programming technique developed by Charnes, Cooper & Rhodes (1978) based on the concepts of technical efficiency developed by Farrell (1957). The method is designed to evaluate the relative efficiency and performance of multiple decision-making units (DMUs) based on their inputs and outputs. Decision-making units (DMUs) represent entities that convert inputs (production resources) into outputs (products), such as companies, organizations, schools, hospitals, or government agencies in all sectors of economic activities (Nepomuceno et al. 2023). Efficiency analysis using DEA involves comparing the performance of each DMU with respect to its peers, aiming to identify the most efficient units that achieve the best output from a given set of inputs or minimize inputs to achieve a given set of outputs.

According to Avenali et al. (2023), DEA is currently one of the hot topics in the field of Operations Research and Data Analysis. According to the authors, combining such methodology with simulation and scheduling techniques is a bibliometric-based suggestion for the future of operational research. Time Series Data Envelopment Analysis can be a common avenue for this purpose. Time Series Data Envelopment Analysis (TSDEA) is an invaluable tool that helps measure the relative efficiency of a single decision-making unit over time.

According to Nepomuceno & Costa (2019), TSDEA is an extension of the traditional DEA approach that incorporates the temporal dimension, enabling performance evaluation of a single unit over time, considering months instead of competitors. It leverages time series data, typically with periodic observations, to measure the efficiency of DMUs while considering their evolution and dynamic behavior and avoid comparing units that might significantly differ in their internal structure (production technology). Time series DEA offers several advantages over cross-sectional or static DEA approaches: It does not require data from competitors which

can be difficult to obtain or may not exist (in the case of natural monopolies); it can facilitate internal benchmarking of best practices, and it enables decision-makers to evaluate the efficiency of DMUs, identify trends, and uncover underlying patterns. This analysis is particularly useful in monitoring and improving performance over time, predicting future trends, and benchmarking against period peers within an industry. Together with new developments on conditional frontier estimations (Daraio & Simar, 2007), TSDEA can be the future of the Efficiency Analysis.

2.2. Econometric Tools

Unlike DEA, which is a non-parametric method, econometric approaches allow for incorporating specific functional forms and statistical assumptions. Considering the scope of estimating production functions and efficiency scores, the use of Aigner, Lovell & Schmidt (1977) Stochastic Frontier Analysis (SFA) has the benefit of accounting for random errors in the estimation process, allowing for a more comprehensive evaluation of DMU efficiency. On the other hand, considering the scope of addressing social and economic problems under the lens of causal inference and prediction, econometric models, including a wide range of Regression Models (Linear and Multiple regression, Beta Regression, Logistic Regression, Panel Data Regression for temporal and cross-sectional analysis, among others) and Time Series Analysis (Exponential Smoothing and ARIMA models), play a crucial role in understanding and addressing these problems.

Regression models are widely used in econometrics to identify and quantify causal relationships between variables. They enable researchers to analyze the impact of specific socioeconomic factors on outcomes of interest, leading to evidence-based policy recommendations. These models also help policymakers design targeted interventions by identifying the most significant determinants of socioeconomic outcomes. Regression models can isolate true causal effects, aiding policymakers in understanding the relationships between socioeconomic variables better. This, in turn, helps in the formulation of effective policies that can have a positive impact on the lives of individuals and communities.

On the other hand, time series analysis is a powerful econometric tool for understanding and predicting economic trends and behaviors over time. By analyzing historical data, policymakers can identify patterns, trends, and cycles in socioeconomic variables. Constructing Time-scenarios is also vital for studying the effectiveness of past policies and interventions (Nepomuceno et al. 2017). By examining the impact of past policies on socioeconomic indicators, policymakers can learn from successes and failures to refine their strategies and develop evidence-based policies. My perspective on this topic is that new developments in systematic modeling of social events combining both approaches for empirical lag identification that measures not only the causal effect of variables but also the time for the expected response, will be an important topic in the research agenda for the near future.

3. Conclusion

As we navigate the complexities of a contemporary world marred by socioeconomic challenges, it is imperative that we foster a collective commitment to finding sustainable solutions. In our contemporary society, addressing socioeconomic problems necessitates comprehensive and nuanced approaches. Data Envelopment Analysis is in constant expansion with countless opportunities for application in social and economic segments such as Higher Education, Banks, Healthcare, Public Administrations, among others. Econometric tools, such as different types of regression models and time series analysis, are important for understanding causal

relationships, forecasting trends, and evaluating policy interventions. From simple linear regression to more complex panel data regression, these models provide policymakers and researchers with versatile techniques to navigate the challenges and complexities of socioeconomic issues. By combining such methodologies to identify potential lags and relationships, evidence-based policy decisions can be made, fostering sustainable development and societal progress.

The journal *Socioeconomic Analytics* aims to facilitate interdisciplinary dialogue offering new theoretical developments and social-relevant empirical applications in the scope of, but not limited to those tools, encouraging rigorous research, and promoting evidence-based policies. We can address these challenges head-on by embracing innovative approaches, collaborating across sectors, and prioritizing social inclusion.

Conflict of Interest Declaration

The author has no conflicts of interest to declare and there is no financial interest to report.

References

1. Aigner, D., Lovell, C. K., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. Journal of econometrics, 6(1), 21-37.
2. Blazquez, D., & Domenech, J. (2018). Big Data sources and methods for social and economic analyses. Technological Forecasting and Social Change, 130, 99-113.
3. Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. European journal of operational research, 2(6), 429-444.
4. Ciuffetelli Parke,r D. and Conversano, P. (2021). Narratives of Systemic Barriers and Accessibility: Poverty, Equity, Diversity, Inclusion, and the Call for a Post-Pandemic New Normal Frontiers in Education, 6:704663. doi: 10.3389/feduc.2021.704663
5. Daraio, C., & Simar, L. (2007). Conditional nonparametric frontier models for convex and nonconvex technologies: a unifying approach. Journal of productivity analysis, 28, 13-32.
6. Farrell, M. J. (1957). The measurement of productive efficiency. Journal of the Royal Statistical Society Series A: Statistics in Society, 120(3), 253-281.
7. Nepomuceno, T. C. C., & Costa, A. P. C. S. (2019). Resource allocation with time series DEA applied to Brazilian federal saving banks. Economics Bulletin, 39(2), 1384-1392.
8. Nepomuceno, T. C. C., Costa, A. P. C. S., & Daraio, C. (2023). Theoretical and Empirical Advances in the Assessment of Productive Efficiency since the introduction of DEA: A Bibliometric Analysis. *International Journal of Operational Research*, 46(4), 505-549.
9. Nepomuceno, T. C. C., de Moura, J. A., e Silva, L. C., & Costa, A. P. C. S. (2017). Alcohol and violent behavior among football spectators: An empirical assessment of Brazilian's criminalization. International journal of law, crime and justice, 51, 34-44.