Editorial

Application and Scope of Econometrics in Public Health

Running title: Econometrics in Public Health

Divya Sharma[1], Himika Kaundal[2], Tanvi Kiran[3]

Department of Community Medicine and School of Public Health, PGIMER, Chandigarh, India

Corresponding author: Dr. Tanvi Kiran

Email id: **tanvikiran3@yahoo.com**

Abstract

This editorial describes the basic concept of Econometrics and its use in the diverse areas of public health. It also touches upon the various tools of econometrics that play a significant role in public health and its future scope in the field. Econometric applications in public health have great prospects. Thus, its use should be encouraged as it has the potential to contribute to better evaluation and forecasting of public health outcomes.

Keywords: Public Health, Econometrics, Prediction Modelling, Regression

Introduction

The term 'Econometrics,' conceived and crafted by Ragnar Frisch, who is also the co-winner of the First Nobel Prize in 'Economics' in 1969 [1], is generally used to elucidate the application of statistical and mathematical methods in quantifying and critically assessing hypothetical relationships using data [2]. It is formed by an amalgamation of economic theory, economic statistics, mathematical economics, and statistics [3]. The literal meaning of econometrics is 'measurement in economics' since the Greek word metrics means measurement [4]. Though the original meaning of the term suggested using statistical methods only in economics, its usage in the current times has been extensive, transcending the world of economics to branch out towards many clinical areas and public health sciences. Today, the demand and application of econometrics are spread across almost all social sciences and beyond [2]. There has been a considerable increase in the use of econometrics in the field of social sciences and health. It has mainly resulted from easy access to comprehensive datasets, the emergence of statistical software, and the boom of advanced statistical and mathematical techniques [5]. The health data has also evolved, and we now widely come across panel data encompassing cross-sectional and time-series data in health literature. Econometric models [6], based on complex regression equations between a string of dependent and explanatory variables, are now increasingly used for prediction and forecasting in health care policy decisions, health technology assessment, impact

evaluations of health interventions, policy, and so on.

Data use in Econometrics

Just as a pediatrician monitors the baby's growth by analyzing its bodily symptoms and vital signs, the health of any community/state/district/country can be evaluated by analyzing the health-related fertility, morbidity, and mortality data [7]. The data used in econometrics is usually collected from either primary, secondary or mixed sources. Primary sources contain data that is collected by individuals or organizations for the first time [8], whereas data used from any published or unpublished sources are secondary data. Secondary data sources include data from warehouses of government, quasi-government organizations like the Global Health Observatory (GHO) of the World Health Organization (WHO) [9], the Human Development Report database of the United Nations Development Programme (UNDP) [10], Health Statistics by Organization of Economic Co-operation and Development (OECD) [11], Sample Registration reports by Census of India, Ministry of Home Affairs [12], and many more. Several health impact assessment studies and health trend analyses worldwide have been conducted using data from these databases [13] [14]. Researchers across the world are increasingly using both basic and advanced econometric models for analyzing the health data pertaining to their respective fields.

Analytical techniques, models, and software commonly used in Econometrics

Regression analysis, which primarily indicates the dependence of one variable on a single or multitude of explanatory variables, is the prime tool used in econometrics [3]. For instance, a health state can be explained by a diverse set of explanatory variables like; genetics, nutrition, environmental factors, lifestyle, and alike through various types of regression analytical techniques. It is to be noted that regression does not necessarily mean causation; therefore, the regression analysis results need to be interpreted with some caution. A Health Econometrician typically proceeds with a definite methodology while analyzing the health data. The traditional methodology in econometrical applications in health sciences begins with a health-based theory or a hypothesis, followed by an econometric modelling equation described below:

$$\text{Health State} = \partial + b_1 X_1 + b_2 X_2 + \ldots\ldots + b_k X_k + e$$

X_1, X_2, and X_k are the number of explanatory variables that are either theory-driven or literature supported and are hypothesized to have a relation with the dependent variable (health state), and e refers to the random error term in the econometric model. After the equation formulation, the required data is obtained, and the model parameters are estimated through the appropriate regression technique. The choice of the regression model depends primarily upon the nature of the dependent and explanatory variables [15]. After shortlisting some appropriate regression models to be used, the following key step is to check whether or not the dataset fulfills the assumptions of the identified regression model. The choice of the appropriate econometric model and the assumption fulfillment requires a detailed discussion and therefore is beyond the scope of this article. We, however, stress that model selection criteria and fulfillment of its underlying assumption are critically important as the reliability and accuracy of the regression estimates used for predicting health outcomes and related policy depend primarily on them. Therefore, these should be considered while planning the methodology for econometrical applications in health sciences. After estimating the model parameters, the next step is to test the proposed hypothesis using either or both

the point estimation and confidence interval approach, followed by interpreting the results and drawing inferences from them. If we talk about the popularly used regression techniques and models employed for the impact assessment in health sciences, the OLS (Ordinary Least squares) [16] method is often used when the dependent variable is continuous.

On the other hand, when the dependent variable is categorical, then maximum likelihood-based estimation (MLE) models [17] such as logistic and probit regression models are used. For example, the OLS model can be used to understand which factors play a significant role in predicting disease-specific mortality rates (dengue-related deaths, COVID-19 fatality rates, etc.). In contrast, MLE can be used to predict the likelihood/odds of occurrence of a disease/medical condition, which are contingent upon several explanatory variables (odds of smoking resurgence, cancer relapse, and alike). Autoregressive Moving Average (ARMA) based models are usually employed to analyze time-series health data (overtime fertility rates, birth, death rates, health spending spanning across different years, etc.). Panel data regression models are used to predict health outcomes concerning cross-sectional and time-series data (for example, analyzing maternal health service coverage as influenced by various associated factors for several Indian states spread across 10 years). Different computing software are used to run the aforementioned complex regression models. IBM SPSS most commonly used for running regression models based on cross-sectional data, whereas Stata, EViews, and R software are often employed for running advanced econometric models mainly based on time series and panel data points.

Application of Econometrics in diverse areas of Public Health

One of the pivotal objectives of public health mainly entails monitoring community health to prevent diseases. The monitoring of health is achieved by collecting and analyzing health data. Assessing significant disease risk factors, identifying vulnerable groups, ascertaining health threats, *etc.*, are some essential outcomes from statistical findings in public health [7]. These outcomes are achieved by analyses of various types of variables in the data. Socioeconomic variables provide data on gender, education, occupation, income, housing, etc. Based on these variables, various socioeconomic scales have been devised to measure the status of a family. The commonly employed scales include Kuppuswamy, Udai Pareek, and BG Prasad [18]. Since socioeconomic status is a crucial indicator of health, these scales are widely used in public health to draw an association between socioeconomic status and health [19] using different econometric models. Regression-based techniques are widely used for assessing various health policies and interventions on a diverse set of mortality and fertility indicators. Patient-level microdata for assessing treatment costs of a particular disease and the health system costs incurred by the providers of health care services are utilized for economic evaluations of health interventions. Cost-effectiveness, cost-utility, cost-benefit, and cost-minimization [20] are the prime techniques employed mainly by health economists for economic modelling. Health care cost data serves as the critical factor by the health insurers for predicting the risk adjustment while devising various insurance plans with different sets of insurance premiums. The regression analysis tools are even utilized for supply chain planning, inventory management, operation research, and business forecasting for health managers. All the

aforementioned examples provide strong evidence for the widespread usage and application of econometric modelling in public health, spanning across the diverse fields of health promotion, epidemiology, health economics, health financing, health management, etc.

Future Scope of Econometrics in Public Health and Challenges ahead

Econometric applications in evaluating, assessing, predicting, and forecasting public health outcomes have excellent prospects, especially when complemented by machine learning (ML) and Artificial intelligence (AI). Android or iPhone Operating System IOS applications installed on our computers, mobile phones, and smartwatches to track our physical activity, heart rate, menstrual cycle, dietary habits, etc. are all embedded with an AI framework and coupled with an inbuilt regression algorithm that can help forecast the risk of heart attack, ovulation cycle, obesity, and other morbid conditions. The same can be applied in large hospital settings and can be used to forecast the probability of recurrence of particular cancer, episodes of psychological disorders, etc. In fact, time-series regression forecasting was widely employed to predict the waves of COVID-19, case fatality rates, etc., during the current pandemic. Having discussed the potentially broad scope of econometrics in public health, researchers have many challenges surrounding its employment and usage. Different econometric models are based on some underlying assumptions. Few techniques are pretty robust and can withstand the violation of some assumptions. However, reporting it while presenting the results is recommended when some critical assumptions are violated. Transforming the data as per the requirement and using non-parametric alternate modelling techniques are some solutions in case the critical assumptions are violated so that robust, unbiased, and reliable health impact assessments can be made. As a public health practitioner, using such econometric modelling involves the knowledge of different types of data, its nature, data handling, management, and skills for applying the appropriate and relevant econometric model. Lack of this knowledge leads to misreporting the results, which can lead to wrongful and sometimes dangerous policy decisions. To avoid this, it is recommended that a public health practitioner who wishes to employ regression-based econometric models must have adequate exposure to the various learning courses involving the understanding of different data types, different tools of analysis, and what to do in the event of violations of key assumptions underlying such analytical models. Econometric application in public health is an interdisciplinary process. Therefore, a public health researcher should coordinate and form a team or liaison with researchers across the related field of health economics, biostatistics, econometrics, and epidemiological modelers for reasonably accurate health care predictions.

References

1 Hansen B e. Econometrics. https://www.ssc.wisc.edu/~bhansen/econometrics/Econometrics.pdf%0A%0A

2 Christopher Dougherty. Introduction to Econometrics. Oxford 2011. https://books.google.co.in/books?hl=en&lr=&id=UXucAQAAQBAJ&oi=fnd&pg=PP2&dq=econometrics&ots=8NTFxdb2kp&sig=dXauJWvcB-OasO00HJy9pX56jc0&redir_esc=y#v=onepage&q&f=false (accessed 8 Jul 2022).

3 Gujarati DN. Basic Econometrics. 4th ed. McGraw-HiII/lrwin 2003.

4 Dimitrios Asteriou, Stephen G. Hall. Applied Econometrics . 4th ed. Red Globe Press 2021. https://books.google.co.in/books?hl=en

&lr=&id=2BlHEAAAQBAJ&oi=fnd&pg=PR7&dq=econometrics&ots=eNF5mHaDWU&sig=IXlZoQ3nkjF8KGpngHY8PZTCIco&redir_esc=y#v=onepage&q&f=false (accessed 8 Jul 2022).

5 Baltagi BH, Sadka E. Health Econometrics Contributions economic analysis. Emerald Publishing Limited Reprints 2018.

6 Newhouse JP, McClellan M. Econometrics in outcomes research: The Use of Instrumental Variables. http://dx.doi.org/101146/annurev.publhealth19117 2003;19:17–34. doi:10.1146/ANNUREV.PUBLHEALTH.19.1.17

7 S Leena. Role of Statistics in Public Health. Nti Bulltin. 2007;:41–2.

8 Kiran T, Junaid KP, Sharma D, et al. Sociodemographic Determinants of Willingness and Extent to Pay for COVID-19 Vaccine in India. Front Public Heal 2022;10. doi:10.3389/fpubh.2022.870880

9 World Health Organization (WHO). Global Health Observatory. https://www.who.int/data/gho (accessed 9 Jul 2022).

10 United Nations Development Programme (UNDP). Home | Human Development Reports. https://hdr.undp.org/ (accessed 9 Jul 2022).

11 OECD. Health Statistics 2022 - OECD. https://www.oecd.org/els/health-systems/health-data.htm (accessed 9 Jul 2022).

12 Census of India. Government of India | Office of the Registrar General & Census Commissioner, India. https://censusindia.gov.in/census.website/data/SRSSTAT (accessed 9 Jul 2022).

13 Sharma D, Kiran T, Junaid K, et al. Composite measure of Age-Specific Mortality Rate of rural females at sub-national level in India - Mendeley Data. 2022.https://data.mendeley.com/drafts/v5gskzj78x (accessed 21 Apr 2022).

14 K M, KP J, Aggarwal A, et al. A Comparative Analysis of Exports and Imports of 'Protective Garments and the like' in India. Int J Heal Syst Implement Res 2021;5:39–48.

15 Ericsson NR. Testing Linear versus Logarithmic Regression Models: A Comment. Rev Econ Stud 1982;49:477–81. doi:10.2307/2297371

16 Zdaniuk B. Ordinary Least-Squares (OLS) Model. Encycl Qual Life Well-Being Res 2014;:4515–7. doi:10.1007/978-94-007-0753-5_2008

17 Balaban J. A Gentle Introduction to Maximum Likelihood Estimation | by Jonathan Balaban | Towards Data Science. Towar. Data Sci. 2018.https://towardsdatascience.com/a-gentle-introduction-to-maximum-likelihood-estimation-9fbff27ea12f (accessed 13 Jul 2022).

18 Ramesh Masthi N, Gangaboraiah, Kulkarni P. An exploratory study on socio economic status scales in a rural and urban setting. J Fam Med Prim Care 2013;2:69. doi:10.4103/2249-4863.109952

19 National Health Profile. Socio-economic indicators. 2010. www.cbhidghs.nic.in

20 Higgins AM, Harris AH. Health economic methods: cost-minimization, cost-effectiveness, cost-utility, and cost-benefit evaluations. Crit Care Clin 2012;28:11–24. doi:10.1016/J.CCC.2011.10.002

-----*-----