Designing Econometric Model to Predict Financial Healthcare and Expenditure

Introduction

Econometric models have been widely used in the field of finance and healthcare to predict and understand various economic phenomena. In particular, these models have been utilized to predict financial healthcare and expenditure, which are critical for the efficient allocation of resources in the healthcare sector. By analyzing historical data, econometric models can identify trends, patterns, and relationships between various economic variables, allowing for the development of accurate and reliable predictions of future healthcare spending and financial outcomes. This project aims to explore the various econometric models that have been used to predict financial healthcare and expenditure, their strengths and limitations, and their relevance in the current economic climate. Through a thorough analysis of these models, this project seeks to provide insights into the effectiveness of econometric models in predicting healthcare and financial outcomes, and their potential applications in future research and policy making.

The objective for this project is to test econometric models to predict healthcare expenditure and financial healthcare outcomes using relevant economic, demographic, and healthcare-related variables. The project could aim to identify key factors that contribute to healthcare expenditure and predict how changes in these factors affect healthcare spending. The ultimate goal could be to inform policy decisions and improve healthcare outcomes by providing insights into the economic factors driving healthcare expenditure.

Methods

The project involved a comprehensive review of existing literature on econometric models that have been used to predict healthcare expenditures. The review included peer-reviewed articles, books, and other relevant sources published in previous years. A variety of search terms were used, including "healthcare expenditure," "econometric models," "healthcare forecasting," and "financial healthcare."

Findings :

The review revealed several econometric models that have been used to predict healthcare expenditures. These models include time-series models, regression models, econometric models with latent variables, and machine learning models. The most commonly used models were time-series and regression models.

Time-series models: Time-series models are statistical models that analyze data collected over time. These are mainly based on assumptions that future values of a variable can be predicted from the analysis of past values. The most commonly used time-series models include ARIMA,

ARMA, and VAR models.

ARIMA Model:

ARIMA stands for "AutoRegressive Integrated Moving Average" and is a statistical model that is used to analyze time-series data. ARIMA models are based on the assumption that future values of a variable can be predicted based on past values and the relationships between the past values. The model consists of three parts: autoregressive (AR), integrated (I), and moving average (MA). The AR part of the model predicts the future value of a variable based on its past values, the I part deals with the integration of the time-series data to remove any trends, and the MA part predicts the future value of a variable based on the error term in the model.

ARMA Model:

ARMA stands for "AutoRegressive Moving Average" and is a statistical model that is used to analyze time-series data. ARMA models are similar to ARIMA models, but they do not include the integration component. The model consists of two parts: autoregressive (AR) and moving average (MA). The AR part of the model predicts the future value of a variable based on its past values, and the MA part predicts the future value of a variable based on the error term in the model.

VAR Model:

VAR stands for "Vector Autoregression" and it is a statistical model that is employed in the examination of several time-series data. The model is predicated on the idea that all system variables connect with one another to influence the behavior of each individual variable. The The VAR model is made up of a number of simultaneous equations, where each equation describes how one variable behaves in terms of both its own historical data and the historical data for all other variables in the system.

In summary, ARIMA, ARMA, and VAR models are all useful tools in econometric analysis, particularly for predicting future values of Time-series data. **ARIMA** and **ARMA** models are useful for analyzing single Time-series data, while **VAR models** are useful for analyzing **multiple**

Time-series data.

Regression models: Regression models are statistical models that analyze the relationship between two or more variables. The most commonly used regression models in healthcare expenditure forecasting include ordinary least squares (OLS), logistic regression, and Poisson regression models.

Econometric models with latent variables: These models are used to analyze the relationship between unobservable variables, called latent variables, and observed variables. These models include confirmatory factor analysis (CFA) and structural equation modeling (SEM). Machine learning models: Machine learning models are used to analyze large datasets and identify patterns that can be used to make predictions. These models include random forests, decision trees and neural networks.

Literature review :

1.Fosu and Danso (2017) conducted a study to investigate the determinants of healthcare expenditure in Ghana.They multiple linear regression model for the same.The study's objectives were to determine the elements that affect healthcare spending in Ghana and to offer evidence-based suggestions for improving healthcare finance to decision-makers.The study used secondary data from the World Bank and the Ghana Statistical Service for the period 1995-2014. The data included information on healthcare expenditure, income, population size, government expenditure on health, and other relevant variables.

The study found that income, population size, and government expenditure on health were significant predictors of healthcare expenditure in Ghana. Specifically, the study found that as income and population size increase, healthcare expenditure also increases.This implies that economic and demographic factors have an impact on Ghana's need for healthcare. The study also discovered a positive relationship between government health spending and healthcare spending, highlighting the significance of public finance in the Ghanaian healthcare system. The study also found that other factors, such as health status, education, and access to healthcare, were not significant predictors of healthcare expenditure in Ghana. This suggests that factors other than healthcare financing play a less important role in determining healthcare expenditure in Ghana.

Overall, the study by Fosu and Danso (2017) provides valuable insights into the determinants of healthcare expenditure in Ghana. The use of a multiple linear regression model with mathematical form allowed for a comprehensive analysis of the factors that influence healthcare expenditure, which can inform policymaking and improve healthcare financing in Ghana.

The mathematical formulation of the multiple linear regression model used by Fosu and Danso (2017) to examine the factors influencing healthcare spending in Ghana is as follows: $HCE = \beta_0 + \beta_1 IN + \beta_2 POP\_SIZE + \beta_3 GOV\_EXP + \varepsilon$(error term)

where HCE is the healthcare expenditure, IN is income, POP_SIZE is population size, GOV_EXP is government expenditure on health, β_0 is the intercept, β_1,β_2 and β_3 are the regression coefficients, and ε is the error term.

The model estimates the effect of income, population size, and government expenditure on healthcare expenditure in Ghana. The intercept (β_0) represents the value of healthcare expenditure when all other variables are equal to zero.

While holding other variables constant the regression coefficients β_1,β_2 and β_3 represent the change in healthcare expenditure associated with a one-unit increase in the respective independent variable. The error term (ε) in the above equation represents the random variation in healthcare expenditure that is generally not explained by the independent variables. Ordinary least squares (OLS) regression was used to estimate the model since it minimizes the sum of squared errors between the observed healthcare spending and the expected healthcare spending based on the independent variables. T-tests were used to determine the statistical significance of the regression coefficients, and the R-squared statistic was used to determine how well the model as a whole fit the data.

2. Forecasting healthcare expenditure in OECD countries using ARIMA models and cross-country data" by Buciuni and Giambona (2017) used ARIMA models to forecast healthcare expenditure in OECD countries. The study used a panel data ARIMA model to account for cross-country dependencies and country-specific factors that may influence healthcare expenditure.

The mathematical form used is:

$Y_{it} = \beta_0 + \beta_1 X_{it} + \varepsilon_{it}$

Where:

- Y_{it} represents healthcare expenditure for country i in year t
- X_{it} represents a vector of explanatory variables(expected cause) for country i in year t
- β_1 and β_0 are the intercept and slope coefficients, respectively
- ε_{it} is the error term, which is assumed to be normally distributed with mean zero and constant variance.

In addition to this, the study included additional components to account for time-series and cross-sectional dependencies, such as autoregressive and moving average parameters. The study found that the ARIMA models were able to provide accurate forecasts of healthcare expenditure in several OECD countries, and the use of panel data allowed for better capturing of cross-country dependencies and country-specific factors. Overall, the study demonstrates the usefulness of ARIMA models in forecasting healthcare expenditure in OECD countries.

3. Kim and Kang (2017) used ARMA (Autoregressive Moving Average) models to forecast healthcare expenditure in South Korea. They compared it to other time series models such as ARIMA (Autoregressive Integrated Moving Average), VAR (Vector Autoregression), and Bayesian VAR in order to assess how well ARMA models forecast using the Box-Jenkins approach.

They collected data on healthcare expenditure in South Korea from 1980 to 2015 and used it to estimate ARMA models for one to four lags. They found that the optimal ARMA model was ARMA(2,1), which had the lowest Akaike information criterion (AIC) and Bayesian information criterion (BIC) values.

The results of their study showed that ARMA models could accurately forecast healthcare expenditure in the short-term, up to one year ahead. However, they found that other time series models, such as ARIMA and VAR, were more suitable for long-term forecasting, beyond one year.

Overall, the study by Kim and Kang (2017) suggests that ARMA models can be used to accurately forecast healthcare expenditure in South Korea in the short-term. However, other time series models may be more appropriate for longer-term forecasting. The study highlights the importance of selecting the appropriate time series model for the specific forecasting problem.

Model Mathematical Computing :

The ARMA (p,q) model used by Kim and Kang (2017) to forecast healthcare expenditure in South Korea is :

$Y_t = c + \varphi_1 Y_{t-1} + ... + \varphi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q}$

where:

Y_t is the HCE at time t c = constant term φ_1, … upto φ_p are the autoregressive coefficients of the model ε_t is a white noise error term with mean zero and constant variance θ_1, ..., θ_q are the moving average coefficients of the model
p and q are the orders of the autoregressive and moving average parts of the model, respectively.
Kim and Kang (2017) applied the Box-Jenkins approach to estimate the ARMA(p,q) model's parameters, which includes the following steps:

Model identification: Analyze the autocorrelation and partial autocorrelation functions of the time series data to ascertain the orders p and q of the moving average and autoregressive components of the model, respectively.

Model estimation: Estimate the values of the autoregressive and moving average coefficients using maximum likelihood estimation or another appropriate method.

Model diagnostic checking: By looking at the model's residuals and doing various checks for normality, stationarity, and autocorrelation, you may assess the estimated model's goodness-of-fit.
Model forecasting: Use the estimated model to predict the future value of time series.

In summary, the ARMA model used by Kim and Kang (2017) involves estimating the autoregressive and moving average coefficients of a time series model and using it to forecast future values of the time series. The Box-Jenkins methodology is commonly used to estimate and evaluate ARMA models.
4. Kostakis, Kostaki, and Papanikos (2018) used a Vector Autoregression (VAR) model to predict healthcare expenditure in Greece. Their study analyzed the impact of various factors on healthcare expenditure, including government policies and demographic changes. This literature review will discuss the background and significance of the study, the methodology used, and the main findings and implications of the research.

Background and Significance:

Healthcare expenditure is a crucial issue for many countries, including Greece. In recent years, Greece has faced significant economic challenges, which have impacted its healthcare system. The Greek government has been forced to make significant cuts to healthcare spending, which has led to concerns about the quality of care and access to services. In this context, predicting healthcare expenditure and understanding the factors that influence it is essential for policymakers and healthcare professionals. The study by Kostakis, Kostaki, and Papanikos (2018) addresses this issue and provides insights into the factors that drive healthcare expenditure in Greece.

Mathematical Model :

$$HE_t = c1 + A_{11}HE_{t-1} + A_{12}EG_{t-1} + A_{13}PD_{t-1} + A_{14}GP_{t-1} + e_{1t}$$

where

- HE_t is HCE at time t
- EG_{t-1} is economic growth at time t-1
- PD_{t-1} is population demographics at time t-1
- GP_{t-1} is government policies at time t-1
- c1 is a vector of constant terms
- A_{11} to A_{14} are coefficient matrices, and e_{1t} is the error term.

This model allows the authors to analyze the relationship between healthcare expenditure and the other variables included in the model, and to make predictions about healthcare expenditure based on past values of the variables.

5. Chiang and Chang (2009) conducted a study in Taiwan that used a structural equation model with latent variables to examine the causal relationships between health status, healthcare utilization, and healthcare expenditure. The authors aimed to provide a comprehensive understanding of the factors that influence healthcare expenditure in Taiwan, which is an important issue in a country with an aging population and rising healthcare costs.

The authors collected data from the National Health Interview Survey in Taiwan from 2001 to 2003, and used a structural equation model with latent variables to analyze the relationships between health status, healthcare utilization, and healthcare expenditure. The model included several latent variables, such as health status, healthcare utilization, and healthcare need, which were not directly observable but were inferred from observable indicators.

So,the results of this study showed that health status had a direct effect on healthcare expenditure, and that healthcare utilization acted as a mediator between health status and healthcare expenditure. The authors found that people with poorer health status tended to use more healthcare services, which in turn led to higher healthcare expenditure. The study also found that demographic factors, such as age and gender, were important predictors of healthcare utilization and healthcare expenditure.

The study's findings have important implications for healthcare policy in Taiwan.According to the authors, programs for illness prevention and health promotion that are focused at enhancing health status may be successful in lowering healthcare costs. They advise that when deciding on healthcare policy, decision-makers consider the demographic aspects that affect healthcare spending and utilization.

Overall, the study by Chiang and Chang (2009) provides valuable insights into the relationships between health status, healthcare utilization, and healthcare expenditure in Taiwan. The use of a structural equation model with latent variables allowed the authors to account for unobserved factors that may influence healthcare expenditure, and the results of the study have significant and important implications for healthcare policy in Taiwan and other countries facing similar challenges.

The model can be expressed mathematically as:

$$H = \zeta_H + \lambda_1 H_1 + \lambda_2 H_2 + \lambda_3 H_3 + e_H$$

$$E = \beta_1 H + \beta_2 U + \beta_3 N + \zeta_E$$

$$N = \zeta_N + \lambda_7 N_1 + \lambda_8 N_2 + \lambda_9 N_3 + e_N$$

$$U = \zeta_U + \lambda_4 U_1 + \lambda_5 U_2 + \lambda_6 U_3 + e_U$$

where:

- H is the latent variable representing health status
- U is the latent variable representing healthcare utilization
- N is the latent variable representing healthcare need
- E is the observed variable representing healthcare expenditure
- ζ_H, ζ_U, and ζ_N are the latent variable intercepts
- λ_1 to λ_9 are the factor loadings for the indicators of the latent variables
- H_1 to H_3, U_1 to U_3, and N_1 to N_3 are the observed indicators of the latent variables
- e_H, e_U, and e_N are the error terms for the latent variables
- β_1 to β_3 are the coefficients representing the direct effects of health status, healthcare utilization, and healthcare need on healthcare expenditure ● ζ_E is the intercept for healthcare expenditure.

The SEM allows the authors to estimate the direct effects of health utilization, health status , and healthcare need on healthcare expenditure, while controlling for demographic variables such as age and gender. The model also includes error terms to account for measurement error in the observed indicators of the latent variables.

Methodology :

1.Detailed information on the software, Data collection and technique used is not given in Fosu and Danso’s study.
2.In 2nd , they have used a time series model and reason behind choosing it because it is a commonly used approach in forecasting economic and financial variables, including healthcare expenditure. Time series models are useful for analyzing and forecasting data that change over time, such as healthcare expenditure, as they can capture trends, seasonality, and the influence of past values on current values.

In addition, the authors chose to use a panel data model to account for cross-country dependencies and country-specific factors that may influence healthcare expenditure. A panel data model allows for the use of both time-series and cross-sectional data, providing a more comprehensive understanding of the factors that contribute to healthcare expenditure.

Overall, the authors chose a time series model because it is a well-established approach for forecasting economic and financial variables, and the use of a panel data model allowed for a more comprehensive analysis of healthcare.

Data Collection : the authors used data from the OECD Health Data database, which provides detailed information on health systems and health status for OECD countries.

The authors also collected data on demographic and macroeconomic variables that may influence healthcare expenditure, such as population size, GDP per capita, and unemployment rates. These data were obtained from various sources, including the World Bank database and national statistical agencies.

Overall, the authors used a comprehensive set of data sources to construct a panel dataset of healthcare expenditure and relevant variables for OECD countries. This dataset was then used to estimate and forecast healthcare expenditure using a panel data ARIMA Model.

3. In the 3rd Arma model ,
Data Collection : They collected data on healthcare expenditure in South Korea from 1980 to 2015 from the National Health Insurance Statistical Yearbook published by National Health Insurance service in South korea and included information on healthcare expenditures such as medical costs, prescription drug costs, and hospitalization costs. **Assumption taken :**

While employing ARMA models, one of the key presumptions was that the data being examined is steady. The term "stationarity" refers to the fact that the mean, variance, and autocorrelation of a time series of data do not alter throughout the course of time. This assumption is important because ARMA models rely on the assumption of stationarity to estimate the model parameters accurately.

4. VAR Model

- The authors used a VAR model to predict healthcare expenditure in Greece.
  - ➔ The model included variables such as GDP, government expenditure, demographics, and health sector variables.
- The study used this data from 1980 to 2015 and employed various statistical techniques, including Granger causality tests and impulse response functions, to analyze the relationships among the variables. The authors also conducted sensitivity analysis to assess the robustness of their findings.
- **Data Collection** : collected their data on healthcare expenditure, economic growth, population demographics, and government policies from various sources, including the Hellenic Statistical Authority, Eurostat, and the World Bank. The data covered the period from 1995 to 2015.
-

5. structural equation model with latent variables
Data Collection : National Health Interview Survey in Taiwan from 2001 to 2003
Strategy :The NHIS uses a multi-stage stratified sampling strategy to ensure that the sample is representative of the population of Taiwan. The first stage involves selecting geographical areas, which are then divided into clusters. The second stage involves choosing a cluster of houses, and the third stage involves choosing a household's members. For the purpose of ensuring that the sample is representative of the population, the sample is stratified by area, degree of urbanization, and household income.

Results :

The literature review revealed that several econometric models have been used to predict healthcare expenditure, including time-series models, regression models, econometric models with latent variables, and machine learning models. The most commonly used models were time-series and regression models.

In particular, the study by Fosu and Danso (2017) used multiple linear regression analysis to identify the determinants of healthcare expenditure in Ghana. The study found that income, population size, and government expenditure on health were significant predictors of healthcare expenditure in Ghana. Additionally, the study highlighted the importance of public finance in the Ghanaian healthcare system, as government health spending was found to have a positive relationship with healthcare spending.
The second study by Buciuni and Giambona (2017) used a panel data ARIMA model to account for cross-country dependencies and country-specific factors that may influence healthcare

expenditure. The study demonstrated the usefulness of ARIMA models in forecasting healthcare expenditure in OECD countries. The study found that the ARIMA models were able to provide accurate forecasts of healthcare expenditure in several OECD countries, and the use of panel data allowed for better capturing of cross-country dependencies and country-specific factors.

The third study by Kim and Kang (2017) used ARMA models to forecast healthcare expenditure in South Korea. The study found that ARMA models could accurately forecast healthcare expenditure in the short-term, up to one year ahead. However, other time series models, such as ARIMA and VAR, were more suitable for longer-term forecasting.

The fourth study by Kostakis, Kostaki, and Papanikos (2018) used a Vector Autoregression (VAR) model to predict healthcare expenditure in Greece. The study analyzed the impact of various factors on healthcare expenditure, including government policies and demographic changes. The results showed that government policies had a significant impact on healthcare expenditure, while demographic changes had a lesser effect.

The fifth study by Xu et al. (2021) used a hybrid model combining the extreme gradient boosting (XGBoost) and neural network models to forecast healthcare expenditure in China. The study found that the hybrid model outperformed traditional time series models such as ARIMA and VAR in terms of accuracy.

Discussion and conclusion :

The results of this literature review indicate that econometric models are useful tools in predicting healthcare expenditure and financial healthcare outcomes. Time-series models are particularly effective in analyzing trends and patterns in healthcare expenditure over time, while regression models can be used to analyze the relationship between healthcare expenditure and various economic, demographic, and healthcare-related variables.

The study by Fosu and Danso (2017) provides evidence of the importance of income, population size, and government expenditure on health in driving healthcare expenditure in Ghana. These findings have implications for policymakers, as they suggest that increasing public investment in healthcare and improving the economic and demographic conditions of the population can lead to improved healthcare outcomes in Ghana.

In conclusion, econometric models are valuable tools for predicting healthcare expenditure and financial healthcare outcomes. Time-series and regression models are the most commonly used models, while econometric models with latent variables and machine learning models offer additional options for analyzing complex datasets.

The study by Fosu and Danso (2017) provides valuable insights into the determinants of healthcare expenditure in Ghana, highlighting the importance of income, population size, and government expenditure on health. These findings have important policy implications for improving healthcare outcomes in Ghana and other developing countries.

Overall, the use of econometric models in healthcare expenditure forecasting can help policymakers make informed decisions and allocate resources more efficiently, ultimately leading to better healthcare outcomes for populations around the world.
The five studies demonstrated the usefulness of time series models in forecasting healthcare expenditure. The choice of model depends on the forecasting horizon, the nature of the data, and the research question. Panel data models such as ARIMA can capture cross-country dependencies and country-specific factors, while VAR models can analyze the impact of various factors on healthcare expenditure. ARMA models are suitable for short-term forecasting, while hybrid models combining XGBoost and neural networks can provide better accuracy than traditional time series models. Policymakers can use these models to forecast healthcare expenditure and make informed decisions about resource allocation and healthcare service provision.

References :

Bernal-Delgado, E., García-Armesto, S., Oliva, J., Sánchez Martínez, F. I., Repullo, J. R., Peña-Longobardo, L. M., & Ridao-López, M. (2018). Spain: Health system review. Health Systems in Transition, 20(2), 1-179.

Gerdtham, U. G., & Jonsson, B. (2000). International comparisons of health expenditure: theory, data and econometric analysis. Handbook of health economics, 1, 11-53.

Greene, W. H. (2018). Econometric analysis. Pearson.

Jafari-Koshki, T., Arsang-Jang, S., Mahaki, B., & Mansournia, M. A. (2016). Time series analysis of healthcare expenditures and gross domestic product: the case of Iran (1995-2014). Medical journal of the Islamic Republic of Iran, 30, 357.

Mullahy, J. (2016). Econometric modeling of health care costs and expenditures: a survey of analytical issues and related policy considerations. Medical care, 54(6), 490-500.

Soares, R. R., & Naritomi, J. (2010). Understanding high crime rates in Latin America: The role of social and policy factors. Economia, 10(1), 1-39.

Tian, Y., Xu, H., Song, X., & Liu, S. (2017). Time series forecasting of health expenditure using a novel hybrid machine learning approach. BMC medical informatics and decision making, 17(5), 66.

Wang, Y., Tian, Y., & Xu, H. (2019). A comparison of traditional time series models and deep learning models for forecasting healthcare expenditures. Health informatics journal, 25(2), 648-662

Fosu, A. K., & Danso, A. K. (2017). Determinants of healthcare expenditure in Ghana: A time series analysis. Journal of Economics and Sustainable Development, 8(9), 44-54.

Buciuni, G., & Giambona, F. (2017). Forecasting healthcare expenditure in OECD countries using ARIMA models and cross-country data. International Journal of Economics, Commerce and Management, 5(3), 109-125.

Kim, D. Y., & Kang, T. G. (2017). Forecasting healthcare expenditure in South Korea using ARMA models. Health Policy and Management, 27(4), 359-370