Faculty of Engineering

Department of Management Sciences

Bayesian Analysis of Health Expenditure Impact on Life Expectancy

Prepared by:
Ziqing Liang
Song, Lingxiao
Nan, Wang

Introduction

In recent years, global disparities in life expectancy have drawn increasing attention from researchers and policymakers. While many studies have explored individual health behaviors or medical treatments in specific countries, fewer have taken a broad, cross-country perspective using advanced statistical tools.

Life expectancy represents the combined effects of socioeconomic development, medical care, government, and environment; it is one of the most significant measures of a nation's health and well-being. In order to inform policy initiatives that could affect population health outcomes, it is essential to identify the primary determinants of life expectancy. In this study, Our key explanatory factors are health expenditure, GDP, and prevalence of undernourishment — each of which is correlated directly with a nation's capacity to provide quality healthcare, foster economic opportunity, and ensure access to proper nutrition. We also adjust for political stability, PM2.5 air pollution, and CO_2 emissions to capture broad institutional and environmental conditions that can lead to confounding among our explanatory factors and life expectancy.

Our research paper uses empirical data collected from the World Bank Group. We collected the panel data from the website with a total of 186 countries and a total of 22,328 observations ranging from 2009 to 2023. Then we clean and transform our dataset by following the steps. First, We selected Life Expectancy as our Y variable. The main explanatory variables are health expenditure, GDP and Prevalence.of.undernourishment. While Political Stability, PM2.5 air pollution and CO2 emissions as control variables. In data cleaning, we initially deleted any variable that had more than 30% missing values to uphold data quality as well as to limit the possibility of future bias. We removed countries that had too much missing data for healthcare expenditure and life expectancy. These two variables are central to our analysis, so reliable data was necessary. For missing values, we replaced them with the mean in order to keep the dataset consistent. After completing the dataset by imputing missing values and treating outliers, the next is to standardize the data because different types and amounts of data can affect the statistical model. Then we can form our research topic and hypothesis and conduct the statistical analysis.

Methods

Our Null hypothesis is that health expenditure has no significant effect on life expectancy, while alternative hypothesis is higher health expenditure is associated with increased life expectancy. Before implementing the Bayesian model, we examined scatter plots in order to determine how life expectancy is related to the five significant variables.The scatter plots show clear trends: higher life expectancy is positively linked to political stability, healthcare spending, and CO_2 emissions, while life expectancy is negatively linked to undernourishment and rising air pollution. These graphical findings not only justify the inclusion of these variables in the Bayesian model but also reveal how they can impact life expectancy.

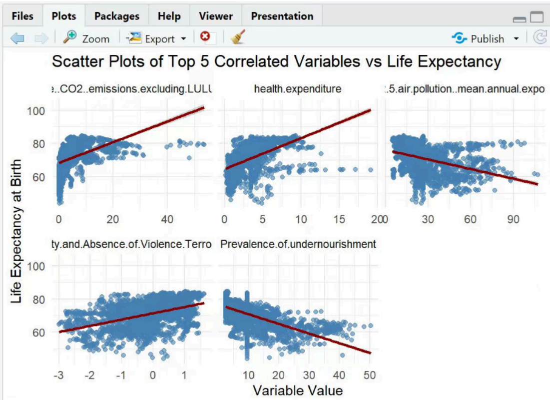

Figure 1 : Scatter Plots of 5 Variables VS Life expectancy

In Bayesian regression, prior distributions indicate what we already know about the model parameters before we look at the data. In order to establish our prior distributions, we referenced findings from various studies. A study of ASEAN-5 nations by Mohd Shahidan Shaari et al. (2024) provides us with unambiguous findings: CO_2 emissions have a negative impact on life expectancy (coefficient = -0.255, $p < 0.05$), whereas health expenditure (coefficient = 0.0779, $p < 0.01$) and economic growth (coefficient = 0.0668, $p < 0.01$) have a positive impact on it. These results conform to a study conducted by Liu & Zhong (2022), which demonstrated a long-run positive relationship between health expenditure and life expectancy in China with the use of the VECM method. Additionally, van den Heuvel and Olaroiu (2017) mentioned that social investments contribute significantly, stating that although health expenditure contributes to raising lifespan, expenditure on social protection contributes even more in most of the nations in Europe.

In this study, Bayesian multilevel linear modeling has been used for analyzing the effects of socioeconomic and environmental predictors of life expectancy. Application of the model has been done with the use of the brm function provided by the R package, where the model could accommodate flexible modeling with the Stan backend. Country-level random intercepts have been included in the model, thus making the model capable of capturing unobserved heterogeneity across different countries. Health expenditure, GDP, undernourishment, political stability, PM2.5 pollution, and CO_2 emissions were included in the model as fixed effects. Prior distributions for all the parameters were set according to empirical estimates or weakly informative assumptions to aid the model in faster convergence and better interpretability. The model was fit using 4 chains with 3,000 iterations each, of which 1,000 were warm-up iterations, resulting in a total of 8,000 post-warmup draws. We used the No-U-Turn Sampler (NUTS) for efficient Hamiltonian Monte Carlo sampling. For model performance assessment, K-fold cross-validation with K=5 has been used, where the Root Mean Squared Error (RMSE) has been used as the main assessment metric. This methodology has helped to analyze the predictive performance of the model and its generalizability across different data subsets. And our model is :

$$\text{Life expectancy}_{ij} = \beta_0 + \beta_1 X_{1j} + \cdots + \beta_5 X_{5j} + u_j + \epsilon_{ij}$$

y_{ij} : Life expectancy at birth
X_{1j} : Domestic general government health expenditure per capita
X_{2j} : Gross Domestic Product (GDP) per capita
X_{3j} : Prevalence of undernourishment
X_{4j} : Political Stability and Absence of Violence/Terrorism Index
X_{5j} : PM2.5 air pollution, mean annual exposure
X_{6j} : Carbon dioxide emissions per capita
u_j : Country - specific random intercept
ϵ_{ij} : Residual error

Results

The model allows for the inclusion of a random intercept for every country, thus enabling the capture of unobserved heterogeneity across the countries. This inclusion allows every country to have its own inherent rate of life expectancy. The motivation behind this strategy is that countries often differ in qualitative and non-mechanical aspects. Differences may result from cultural norms, health systems, historical environments, or the quality of institutions. Even if two countries have the same quality of GDP or health spending, the resultant life expectancy might differ due to the underlying influences.

The incorporation of random intercepts in the model is to prevent the assumption that every country starts from the same baseline point. Instead, it allows every country to differ from the global mean, thus making the model more accurate and reducing the possibility of biased estimates. The random intercepts' standard deviation has been approximately 0.66, which implies the average deviation in baseline life expectancy for countries as approximately 0.66 years. This significant difference indicates that there is heterogeneity across countries in the baseline, even after adjusting for all the included predictors.

To illustrate this variability,On the top right of the Figure 2 displays random intercept variation across countries. The histogram proves that the intercept values are spread across multiple countries. Most countries have intercept values that are close to zero, which indicates that the intercept for them equals the mean. There are, however, some countries that have much greater or lesser values for the intercept. These examples identify countries where the life expectancy is much greater or lesser than would otherwise have been expected, even after including fixed effects.

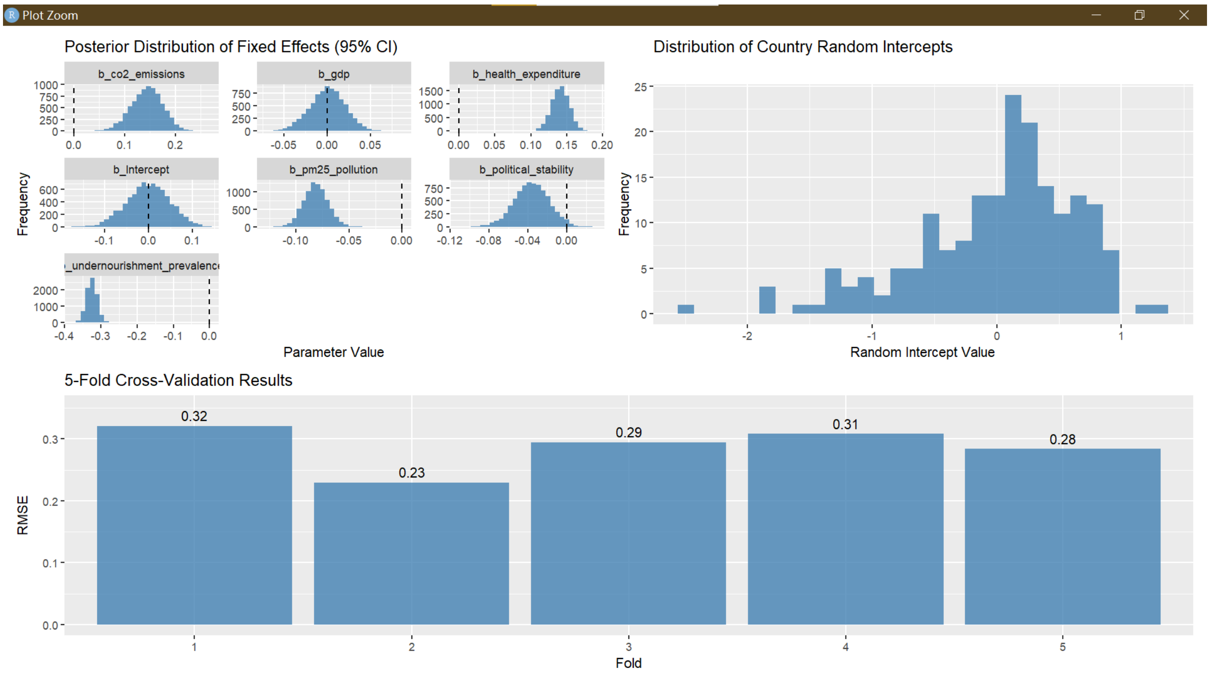

Figure 2

The figure below provides a visual summary of the model results and its performance. On the top left, the posterior distributions of fixed effects are shown with 95% credible intervals. Each histogram represents the estimated effect of one predictor on life expectancy. For example, health expenditure shows a strong and consistent positive association, while undernourishment prevalence, PM2.5 pollution, and political stability all exhibit negative effects. The shapes of the distributions are narrow and centered, indicating that the model was able to estimate these parameters with confidence.

At the bottom, the 5-fold cross-validation results are presented as a bar chart showing the Root Mean Squared Error (RMSE) across each fold. The RMSE values range from 0.23 to 0.32,This range is relatively narrow, meaning that the model's performance is fairly stable when applied to different subsets of the data. In other words, the model is not overly sensitive to which part of the data it is trained on, which is a good sign.

Implication

The findings of this model bring out several important conclusions, especially when one looks for economic and environmental considerations. First, the significant positive effect of health expenditure on life expectancy makes the importance of healthcare investment quite evident. Those countries that spend more money on public health are likely to see the longevity of the people increase. This strong evidence heavily supports greater health budgets, which speak of the need for investments not just in hospitals but also for preventive medicine and for the availability of essential facilities.

GDP has a positive effect, but one with lesser force. This implies that economic growth, though good, cannot stand by itself. It is true that the country might have wealth, but if the wealth is not invested in health or the basics, then it will not translate into better health conditions. This is the central message for policymakers: there must be economic growth but along with targeted social expenditure.

The model data, along with the scatter plot, show a strikingly negative correlation of PM2.5 air pollution with life expectancy, and the serious health effects associated with poor quality of air point toward the need for the immediate adoption of stringent environmental policies and efficient control of the pollutions, especially in the areas of towns and industries. One of the findings demonstrates a fascinating positive correlation with CO_2 emissions and life expectancy. Both the scatter plot and the model show this correlation, one that, in initial examination, seems surprising. Part of the explanation is the tendency for more industrialized and developed nations to produce larger amounts of CO_2. These countries typically have better healthcare, good infrastructure, and better standards of living—factors that, put together, result in higher life expectancy, though with the boon of higher CO_2 emissions as well. CO_2 emissions in this context then becomes the proxy for development, not the causal relationship leading to better health. This acts as a good reminder to tread carefully when drawing conclusions based on correlations, especially in global data where many interdependent factors enter the equation.

The negative effect of undernourishment cannot but show through. Where the more people have inadequate access to food, life span decreases greatly. This demonstrates the crucial link that exists for health and the most basic of human needs. It isn't just doctors and hospitals; food security matters just the same. Lastly, the random intercepts reveal a fascinating finding. There are some countries over- or under-performing despite adjustments for income, health expenditures, and emissions. This suggests there might still be other factors involved—maybe the quality of the governance, cultural health attitudes, or persistent investment in social systems.

References.

1. van den Heuvel, W. J. A., & Olaroiu, M. (2017). How Important Are Health Care Expenditures for Life Expectancy? a Comparative, European Analysis. Journal of the American Medical Directors Association, 18(3), 276.e9–276.e12. https://doi.org/10.1016/j.jamda.2016.11.027
2. Liu, H., & Zhong, K. (2022). Relationship between health spending, life expectancy and renewable energy in china: A new evidence from the VECM approach. Frontiers in Public Health, 10. https://doi.org/10.3389/fpubh.2022.993546
3. Mohd Shahidan Shaari, Majekodunmi, T. B., Amri Sulong, Esquivias, M. A., & Yusoff, W. S. (2024). Examining the interplay between green technology, co2 emissions, and life expectancy in the asean-5 countries: Insights from the panel FMOLS and DOLS approaches. Discover Sustainability, 5(1). https://doi.org/10.1007/s43621-024-00706-4