| | |
|---|---|
| **Course code** | ECON1313 |
| **Course name** | Basic Econometrics |
| **Lecturer** | Dr. Phuong Do Hoang |
| **Student name** | Tran Van Phuoc - s3825778 |
| **Assignment** | Assignment 2<br>Empirical Project Individual Report |
| **Group** | SGS – G07 |
| **Word Count** | 2955 – Excluding Tables, Figures, In-text Citations & Reference |

**ABBERVIATION**

**MLR.1: Assumption of Linearity**

**MLR.2: Assumption of random sampling**

**MLR.3: Assumption of no perfect multicollinearity**

**MLR.4: Assumption of zero conditional means**

**MLR.5: Assumption of no heteroskedasticity**

**MLR.6: Assumption of normality**

**H.E.: Health Expenditure**

**PART 1: Overview and Data Validation**

1. **Overview**

There have been an extensive number of debates and research on the determinants of healthcare expenditure, which is divided into three main categories: economic, demographic, and environmental determinants. Firstly, regarding economic aspects, GDP per capita is widely discovered to be the most influential factor of the changes in the level of H.E (Newhouse, 1977; Hitiris & Posnett, 1992). The fair explanation for the sturdy positive is that the higher income of the population is, the more spending on well-being on products and services will be. Other economic determinants are the trade openness and FDI inflows, as the increased FDI is poured into the marketing of health-related products and services, thereby attracting customer interest (Nagel et al. 2015). Secondly, CO2 emissions and urbanization are found to be the environmental determinants of H.E. (Atuahene et al. 2020), stating that the more toxic substances are being exhausted, the more health expenditure that authorities are willing to pay. Furthermore, urbanization also leads to increases in CO2 emission. Another explanation of the urbanization's effect on H.E is from the demographic factor, in which the people in populated cities have higher accessibility to healthcare products and services than rural citizens due to differences in city planning and income level. Nevertheless, key demographic determinant of H.E. is the population age structure, consisting of elderly and dependency ratios. Han et al. (2013) empirical study found that a high percentage of elderly in the population leads to a rise in H.E; also, a high proportion of the economically dependent population is positively linked with high H.E. The reason is that the elderly and economical-dependent population (under 15 and over 65 years of age) have their immune system to be weaker than the working-age population, leading to a higher expenditure for well-being products and services (Hou & Li 2011).

2. **Data Validation**

   a. **Data importation & validation**

| COUNTRY GROUP A YEAR 2013 | |
|---|---|
| **Variable** | **Description & Unit** |
| **PCHE** | Current health expenditure per capita (Current US$) |
| **CBR** | Birth rate, crude (per 1000 people) |
| **TOTPOP** | Total Population |
| **URBAN** | Urbanization (urban population, % of the total population) |
| **DEPRATIO** | Share of the population that is under 15 years of age or above 65 years of age as a percentage of the population |
| **GDPPC** | GDP per capita (current US$) |

| | |
|---|---|
| **NETODA** | Net official development assistance received (current US$) |

*Table 1: Summary table of assigned sub-dataset of country group A and year 2013.*

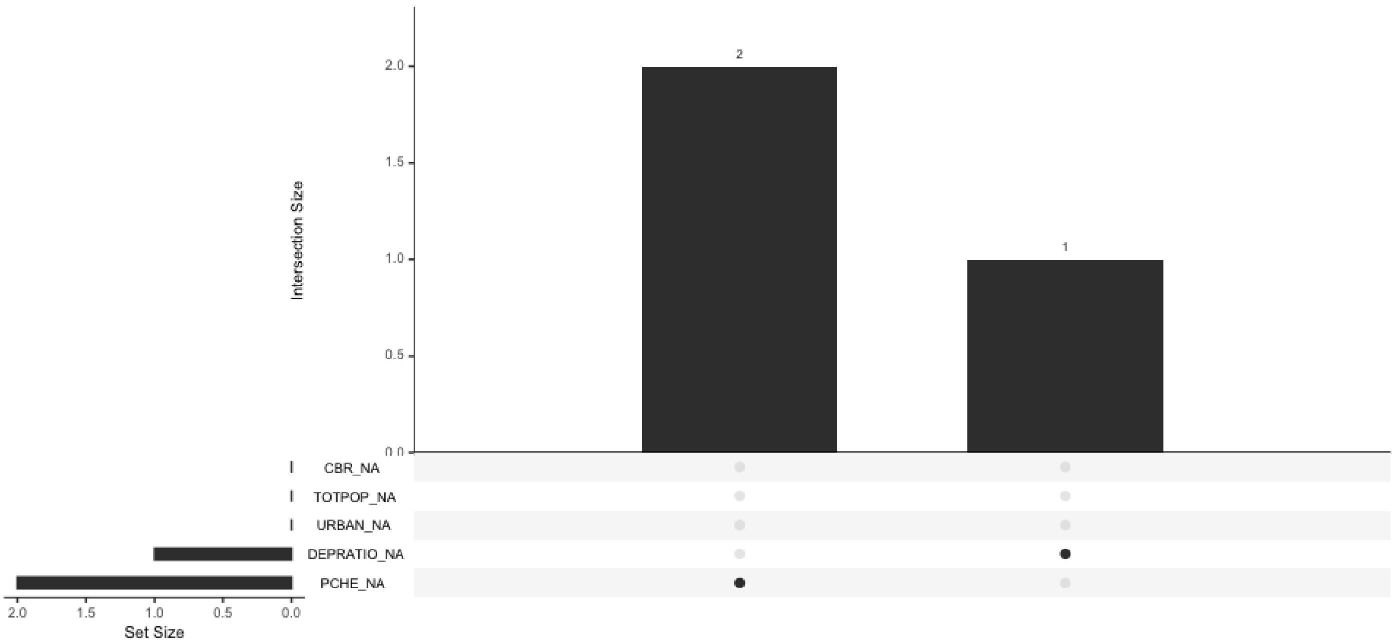



*Figure 1: Upset plot illustrates missing values within the assigned dataset*

For this study, a dataset of country group A and the year 2013, which consist of 37 observations, is imported, and validated in R. However, PCHE and DEPRATIO variables contain missing values, respectively (Figure 1). As missing data could lead to a reduction in statistical power of a study and produce biased estimates, missing data treatment is necessary. In this case, omitting rows which contain mising value is not optimal owing to insufficient sample size; hence, latest observation carried forward/backward method is selected. Specifically, in the case of longitudinal data, a missing value is imputed as that subject's previous/following observed value. Despite being easy to conduct, this method wrongly assumes that the variable remains unchanged over time, thereby reducing the variability/variance of the sample and producing bias estimates (Kang 2013). Nevertheless, this method is still considered better than omitting missing value rows, as it suggests an approximate value based on the nearest observation, thereby conserving the characteristics of the observed country and the sample size.

| **NA replacing method: Latest observation carried forward/backward** | | | | |
|---|---|---|---|---|
| **Descriptive Statistics** | **PCHE before imputation** | **PCHE after imputation** | **DEPRATIO before imputation** | **DEPRATIO after imputation** |
| **Number of observation** | 35 | 37 | 36 | 37 |
| **Number of NA value** | 2 | 0 | 1 | 0 |
| **Min** | 18.8310 | 18.8310 | 1.8700 | 1.8700 |

| | | | | |
|---|---|---|---|---|
| **Max** | 713.2413 | 713.2413 | 63.1650 | 63.1650 |
| **Range** | 694.4104 | 694.4104 | 61.2950 | 61.2950 |
| **Sum** | 7166.4366 | 7262.6480 | 1161.8970 | 1179.3550 |
| **Median** | 143.6069 | 130.4246 | 33.4095 | 31.9120 |
| **Mean** | 204.7553 | 196.2878 | 32.2749 | 31.8745 |
| **Standard Error Mean** | 31.3303 | 30.2077 | 3.1926 | 3.1308 |
| **Confidence Interval Mean 95%** | 63.6707 | 61.2640 | 6.4813 | 6.3496 |
| **Variance** | 34355.4650 | 33762.6401 | 366.9349 | 362.6758 |
| **Standard Deviation** | 185.3523 | 183.7461 | 19.1555 | 19.0440 |
| **Coefficient of Variation** | 0.9052 | 0.9361 | 0.5935 | 0.5975 |

*Table 2: Descriptive statistics of PCHE and DEPRATIO before and after missing value imputation*

As aforementioned before, the latest observation carried forward/backward imputation has caused the variance of two variable to decreased, owing to its assumption that variable remain unchanged over time.

### b. Log transformation

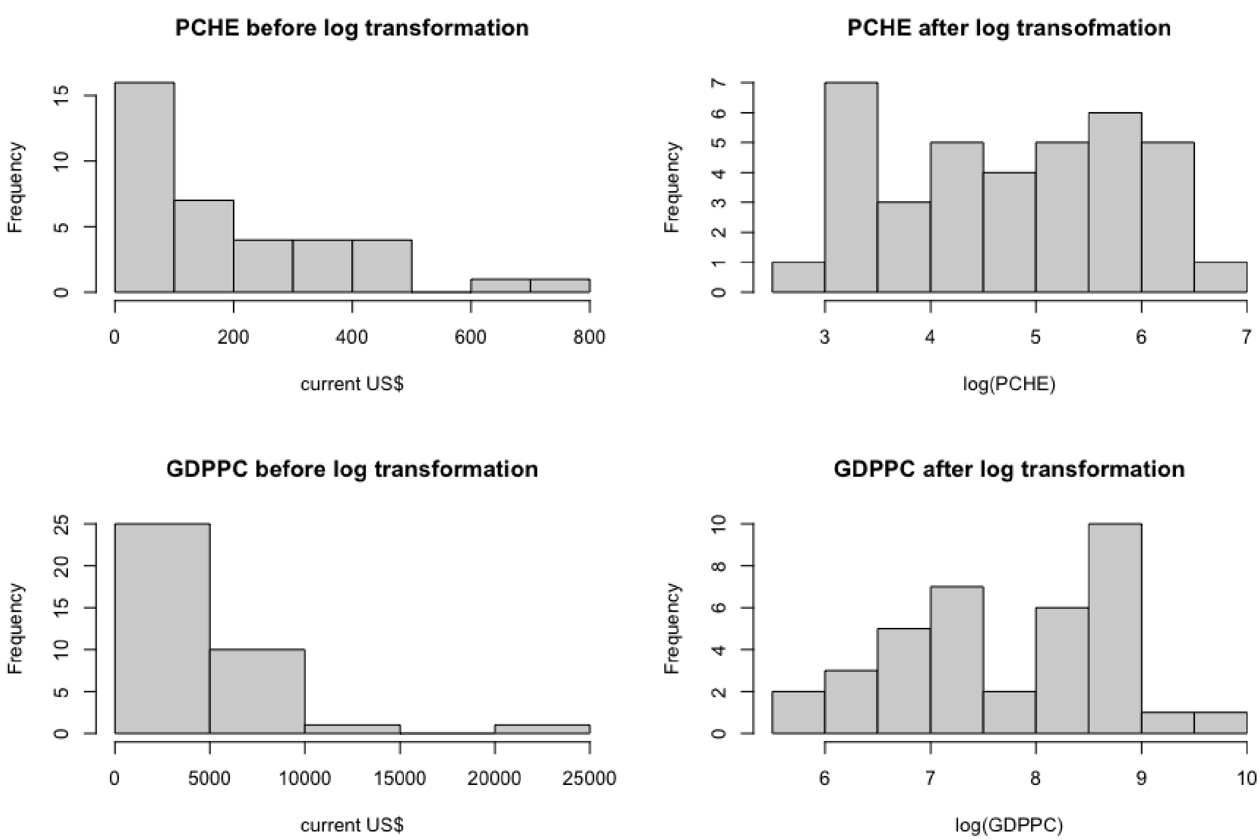



*Figure 2: Distribution of PCHE, CBR and GDPPC before and after taking the natural log*

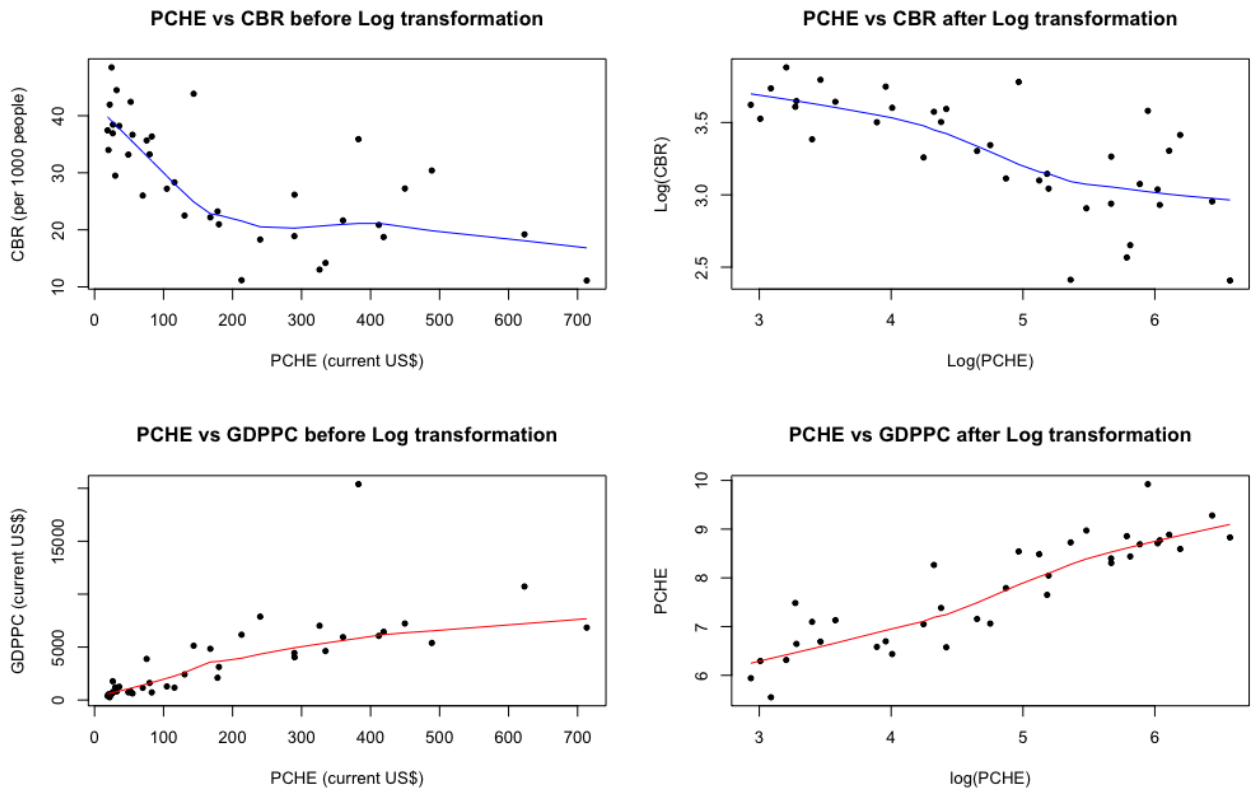



*Figure 3: Scatterplot illustrates the linear relationship between PCHE vs CBR/GDPPC before and after taking the natural log*

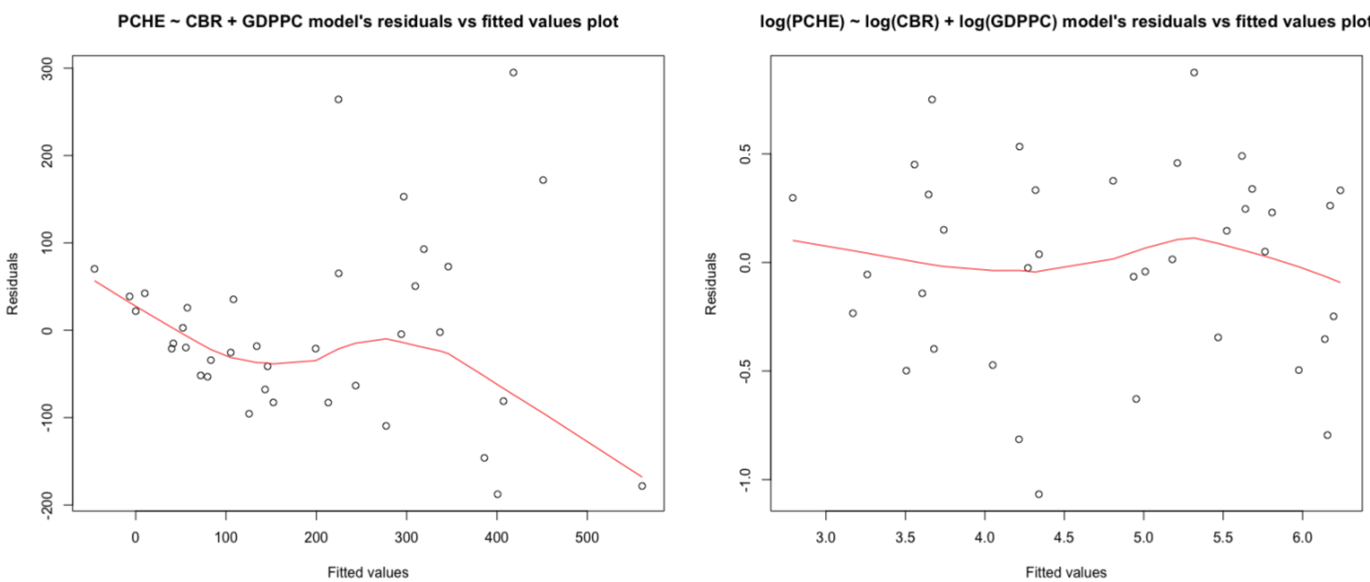



*Figure 4: Residual vs Fitted value scatterplots before and after taking double-log transformation*

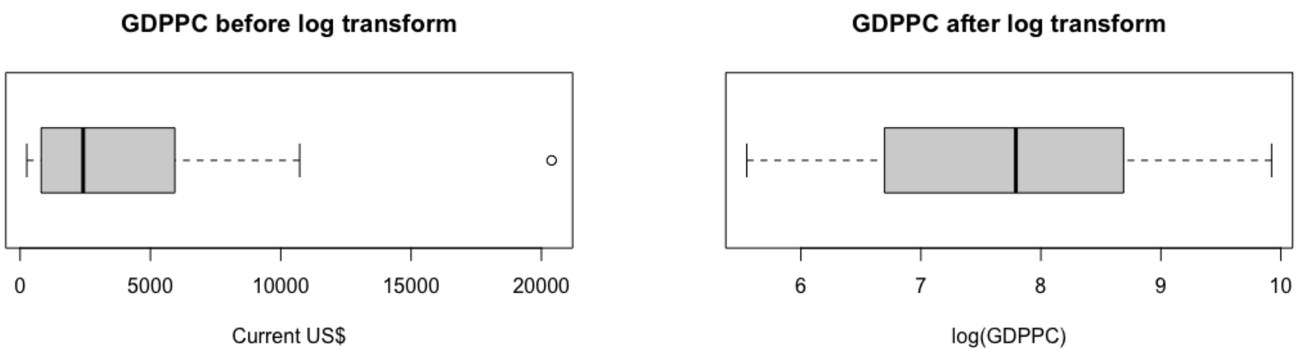



*Figure 5: Boxplots of GDPPC before and after log transformation*

Firstly, unlike MLR.6, real-life data often does not follow the symmetric bell-shaped distribution. Still, they are significantly skewed, causing the standard statistical result of these data to become invalid. Hence, log transformation is essential to conform skewed data to normality (Changyong & Houngyue 2014). Effectively, log transformation has yielded the right-skewed distributions of PCHE-GDPPC variables closer to the Gaussian distribution's bell shape (Figure 2). Secondly, log transformation may strengthen MLR.1 since it can turn a not-so-linear relationship between dependent and independent variables into a more linear relationship. Graphically proved, the PCHE-GDPPC and PCHE-CBR relationships have become more linear after log transformation (Figure 3). Thirdly, log transformation may enhance MLR.5 by mitigating the heteroskedasticity of the residuals. Specifically, by compressing the scales in which the variables are measured, log transformation can significantly minimize the variables' variance, thereby reducing the residuals' variance – the root of heteroskedasticity (Woolridge 2015). Initially, in Figure 4, the residual plots show that there is an increase in the variance of residual across fitted values, indicating the existence of heteroskedasticity. After double-log transformation, the residuals' variance is the same across the full range of fitted values, indicating that the residuals are now homoscedastic. Lastly, as log transformation decompresses the scales of the variables, it can help mitigate the outlier effect (Figure 5). Nevertheless, log transformation is inapplicable if some of the variable's values are zero or negative; hence NETODA could not be log-transformed. Consequently, only the uses of the natural logarithm of PCHE, CBR, and GDPPC are necessary.

## PART 2: Descriptive statistics and initial estimation

### 1. Descriptive Statistics

| Variable | Min | Max | Range | Median | Mean | Standard Deviation |
|---|---|---|---|---|---|---|
| **PCHE (Current US$)** | 18.83 | 713.24 | 694.41 | 130.42 | 196.29 | 183.75 |
| **CBR (per 1000 people)** | 11.11 | 48.47 | 37.36 | 28.32 | 28.86 | 10.03 |

| | | | | | | |
|---|---|---|---|---|---|---|
| **URBAN (% of the total population)** | 11.482 | 89.125 | 77.643 | 50.950 | 47.713 | 19.385 |
| **DEPRATIO (% of the population)** | 1.87 | 63.17 | 61.30 | 31.91 | 31.87 | 19.04 |
| **GDPPC (Current US$)** | 256.97 | 20390.74 | 20133.76 | 2419.72 | 3792.13 | 3905.63 |
| **NETODA (Current US$)** | -669909973 | 5152540039 | 5822450012 | 467540009 | 782401897 | 1076529948 |

*Table 3: Sample's descriptive statistic after replacing missing values with the country's latest data*

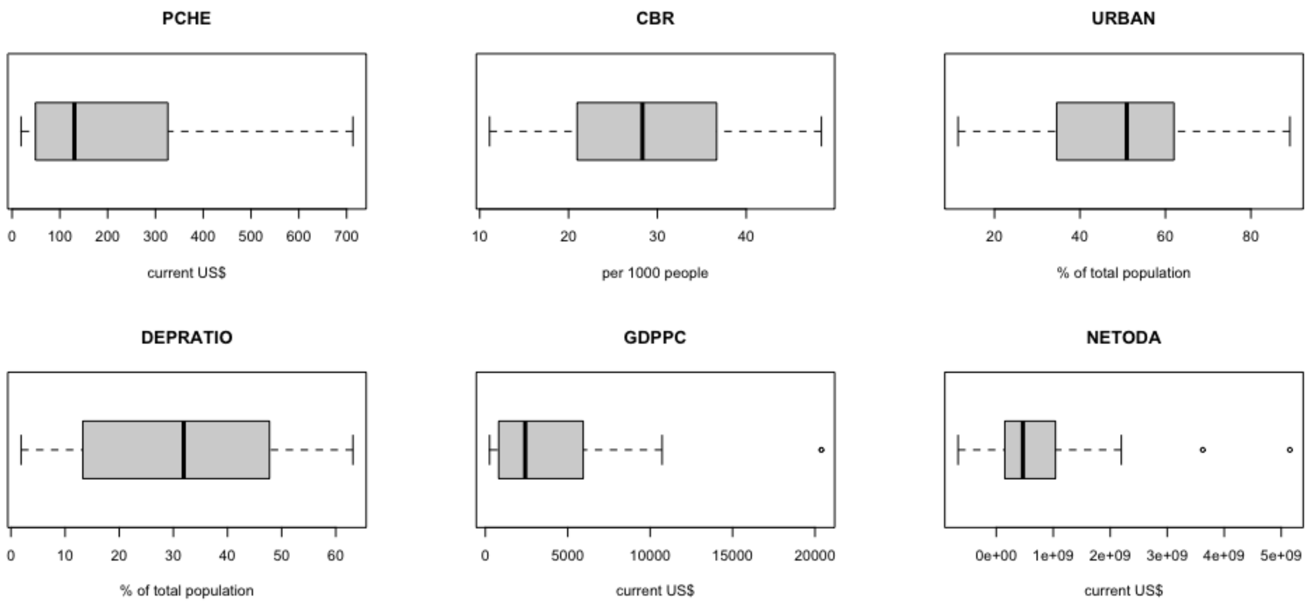



*Figure 6: Boxplots of PCHE, CBR, URBAN, DEPRATIO, GDPPC, and NEDA*

The Table 1 inform about the descriptive statistic results of 6 variables through 37 observations of country group A in the year 2013. Overall, owing to the random sampling of assigned country and the insufficient sample size, half of the variables are significantly skewed (PCHE, GDPPC and NETODA), and 4/6 variables suffer from dramatically high standard deviations. Nevertheless, only two 2/6 variables (GDPPC & NETODA) suffer from outlier effect.

Specifically, by having a relatively high standard deviation (183.75), PCHE is highly volatile. Furthermore, the variable has its mean higher than median, indicating a right skew distribution. Like PCHE, GDPPC and NETODA also have their means lower than their median, meaning their distributions are also right skewed. This might result from the imbalance between the number of income groups in the assigned sample, in which the lower income group accounts for a major proportion of the assigned sample. Contrastively, CBR, URBAN, and DEPRATIO have their mean approximate to their median, indicating near normal distributions.

2. **Model 1's multiple linear regression function:**

$$\widehat{log(PCHE)} = \beta_0 + B_1 log(CBR) + \beta_2 URBAN + \beta_3 DEPRATIO + \beta_4 log(GDPPC) + \beta_5 NETODA + \varepsilon$$

3. **Regression Result**

| Dependent Variable: Log(PCHE) | | | | |
|---|---|---|---|---|
| **Parameter** | **Coefficient** | **Standard Error** | **t-Statistic** | **p-value** |
| **Intercept** | 0.43399 | 2.1630594 | 0.200639 | 0.8423 |
| **Log(CBR)** | -0.34885 | 0.5175337 | -0.67406 | 0.5053 |
| **URBAN** | 0.009444 | 0.0068264 | 1.383464 | 0.1764 |
| **DEPRATIO** | 0.0089644 | 0.0112876 | 0.794187 | 0.4331 |
| **Log(GDPPC)** | 0.6095277 | 0.1418968 | 4.295571 | 0.0002 |
| **NETODA** | 3.17E-11 | 8.02E-11 | 0.395676 | 0.6951 |
| ***Regression Statistics*** | | | | |
| **Multiple R-squared** | 0.8502 | | | |
| **Adjusted R-squared** | 0.826 | | | |
| **Residual Standard Error** | 0.465 | | | |
| **F-statistic** | 35.18 on 5 and 31 DF | | | |
| **Observations** | 37 | | | |

*Table 4: Regression result of Model 1*

**PART 3: Regression Result Interpretation**

**Model 1 Regression Model:**

$$\widehat{log(PCHE)} = 0.43399 - 0.34885 * log(CBR) + 0.009444 * URBAN + 0.0089644 * DEPRATIO + 0.6095277 * log(GDPPC) + 3.17e^{-11} NETODA + \varepsilon$$

1. **R-squared interpretation**

The R-squared of 0.8502 indicates that 85.02% of the dependent variable's (log(PCHE)) variation is explained by the variations of the explanatory variables (Log(CBR), URBAN, DEPRATIO, log(GDPPC), NETODA). As the R-squared is relatively close to 1 (0.085), the model 1 is considered to have a good fit to the assigned data. Nevertheless, R-squared does not truly reflect the multiple regression's explanatory power, as R-squared always increases when number of explanatory variable increases. Contrastively, adjusted R-squared only increases when the new variable truly improves the model's explanatory power. Hence, adjusted R-squared provides better interpretation for multiple regression, which is 0.826 for model 1, indicating that the model is still good in explaining the variation of dependent variable.

2. **F-test for model's overall significance**

A critical element of hypothesis testing is the level of significance that controls the critical value and power of the test, thereby having a substantial impact on the test outcome and decision making. Hence, it is crucial to determine the significance level of the test. Since statistically significant findings are more difficult to detect with small sample size (Type II error) owing to large variability, Zimmerman (2011) recommends lowering the confidence level to deal with small sample size. In this case, with an insufficient sample size of only 37 observations, the hypothesis test should be conducted at 90% confidence level.

By conducting the F-test, we can check the overall fit of Model 1:

$$\begin{cases} H_0: B_1 = B_2 = B_3 = B_4 = B_5 = 0 \\ H_a: at\ least\ one\ explanatory\ variable\ \neq\ 0 \end{cases}$$

Since the critical F-value on 5 parameters and 31 degrees of freedom is 2.042 (significance level = 0.1), which is lower than the F-statistics of 35.18; hence, the null hypothesis is rejected, meaning there is at least one of the explanatory variables is statistically significant.

### 3. T-test for model's individual parameter significance

$$\begin{cases} \mathrm{H_0}: \beta_j = 0 \text{ (no linear relationship)} \\ \mathrm{H_1}: \beta_j \neq 0 \text{ (linear relationship exists)} \end{cases}$$

| Dependent Variable : Log(PCHE) | | | | | |
|---|---|---|---|---|---|
| **Independent Variable** | **Coefficients** | **p-value** | **Signficance level** | **Hypothesis Testing's result** | **Coefficient Interpretation** |
| **Log(CBR)** | -0.3489 | 0.5053 | 10% | Fail to reject H0 | Log(CBR) has no effect on PCHE |
| **URBAN** | 0.0094 | 0.1764 | | Fail to reject H0 | URBAN has no effect on PCHE |
| **DEPRATIO** | 0.0090 | 0.4331 | | Fail to reject H0 | DEPRATIO has no effect on PCHE |
| **Log(GDPPC)** | 0.6095 | 0.00016 | | Reject H0 | PCHE will increase by 0.00016% when log(GDPPC) increases by 1 percentage, ceteris paribus |
| **NETODA** | 3.17.E-11 | 0.6951 | | Fail to reject H0 | NETODA has no effect on PCHE. |

*Table 5: Result of T-test for individual parameter significance*

### 4. Model 1's parameter examination

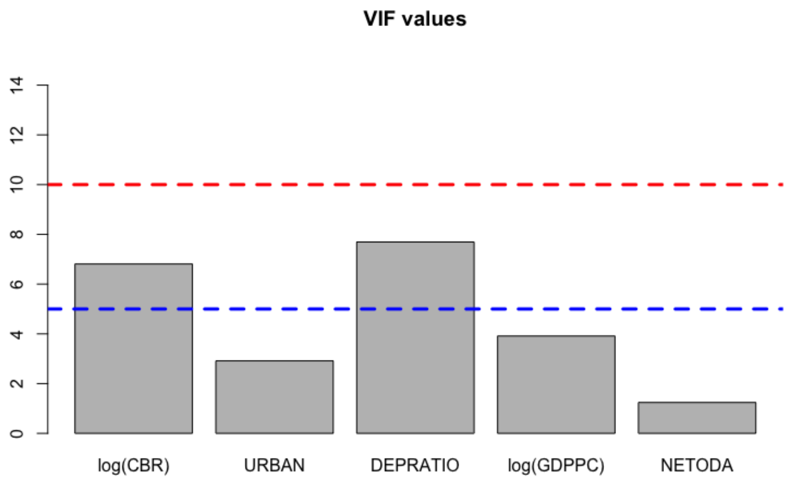
| **Predictor** | **Empirical Result** | **Description** |
|---|---|---|

| | | |
|---|---|---|
| **Log(CBR)** | Negative Coefficient<br><br>Statistically Insignificant | CBR has no impact on PCHE of the assigned country group in 2013, matching the empirical result of Adeel (2016). Moreover, the negative coefficient of the parameter is consistent with the hypothesis of Toor & Butt (2005), stated that an increase in crude birth rate will increase a country total population, thereby decreasing the healthcare expenditure per capita, assuming country's aggerate healthcare expenditure remains constant or has a lower growth rate compared to crude birth rate. |
| **URBAN** | Positive Coefficient<br><br>Statistically Insignificant | URBAN has no effect on PCHE of the assigned country group in 2013, matching the empirical finding of Kraipornsak (2017) regarding ASEAN countries' H.E. However, the positive coefficient of the parameter still matches with the theoretical explanation of a higher urbanization rate may increase healthcare expense due to overcrowded healthcare facilities or excessive demand for well-being services (Ibid.) |
| **DEPRATIO** | Positive Coefficient<br><br>Statistically Insignificant | DEPRATIO has no impact on PCHE of the assigned country group in 2013, matching the empirical finding of Okunade (2005) and Yetim et al. (2021). The positive coefficient of the parameter still matches with the hypothesis of Li (2021), stating that the family with elderly aged 65 and over had significantly higher demand for health services than family without elderly, owing to elderly's deteriorated immune system and old age. |
| **Log(GDPPC)** | Positive Coefficient<br><br>Statistically Significant | Despite being conducted under insufficient sample size, the result still shows a positive relationship between GDP and PCHE, which matches with the previous findings of New House (1997); Hitiris & Posnett (1992), thereby proving GDP is still the most reliable determinant of H.E. |
| **NETODA** | Positive Coefficient<br><br>Statistically Insignificant | NETODA has no impact on PCHE of the assigned country group in 2013, matching the empirical finding of Ali et al. (2020). Nevertheless, the positive coefficient of the result still matches with the hypothesis of Lim et al. (2022), stated that an increase in foreign aid will be channeled through government expenditure which eventually be spent on health-related public investment. |

*Table 6: Model 1's parameter examination*

## 5. Multicollinearity

### a. Variance Inflation Factor (VIF)




| Variable | VIF |
|---|---|
| Log(CBR) | 6.80 |
| URBAN | 2.91 |
| DEPRATIO | 7.69 |
| Log(GDPPC) | 3.91 |
| NETODA | 1.24 |

*Figure 7: Bar plot illustrate VIF statistic of Model 1's explanatory variables*

| | |
|---|---|
| **VIF = 1** | No multicollinearity |
| **VIF < 5** | Minor multicollinearity |
| **5 < VIF < 10** | Noticable multicollinearity |
| **VIF > 10** | Problematic multicollinearity |

*Table 7: Level of multicollinearity phenomenon according to VIF (Gareth et al. 2013)*

Multicollinearity is a phenomenon when there are two or more independent variables of a regression model are correlated. If the degree of multicollinearity is strong, it will undermine the independent variables' statistical significance and deteriorate the regression model's goodness of fit. Hence, testing for multicollinearity is important in regression modelling, which can be carried out by calculating the VIF. According to Figure, model 1's contains two variables *log(CBR)* and *DEPRATIO*, that have VIF to be over 5, indicating the existence of noticeable multicollinearity. Consequently, this high level of multicollinearity may be the key reason, besides insufficient sample size, causing most of model 1's explanatory variables to be statistically insignificant.

| Correlation | Log(CBR) | URBAN | DEPRATIO | Log(GDPPC) | NETODA |
|---|---|---|---|---|---|
| Log(CBR) | 1 | | | | |
| URBAN | -0.5585 | 1 | | | |
| DEPRATIO | -0.9216 | 0.5710 | 1 | | |
| Log(GDPPC) | -0.6361 | 0.8048 | 0.6885 | 1 | |
| NETODA | 0.2889 | -0.2884 | -0.3480 | -0.4188 | 1 |

*Table 8: Correlation matrix of model 1's explanatory variables*

| Correlation | Log(CBR) | URBAN | DEPRATIO | Log(GDPPC) | NETODA |
|---|---|---|---|---|---|
| Log(PCHE) | -0.7221 | 0.7866 | 0.7557 | 0.8933 | -0.3527 |

*Table 9: Correlation coefficients between Model 1's dependent and independent variables*

Fortunately, multicollinearity can be mitigated by omitting one of the correlated explanatory variables, increasing sample size, and redefining variable. However, omitting one of the variables which correlates with both explanatory variables and outcome variable would cause omitted variable bias, thereby breaching the MLR.4 (Buck 2015). According to Table 8 & 9, removing either Log(CBR) or DEPRATIO would introduce omitted variable bias, owing to their high correlation with explanatory and outcome variables. Consequently, only two last options are applicable.

**PART 4: Further estimation**

1. **The dummy variable HDP** is for the purpose of contrasting the effects on health expenditure capital between high level of dependent population and low dependent population (HDP = 1 if dependent population is higher than its median; otherwise HDP = 0).

**The model(2) regression equation:**

$$\widehat{log(PCHE)} = 0.138384 - 0.287996 * \log(CBR) + 0.008759 * URBAN + 0.404036 * HDP + 0.638208 * log(GDPPC) + 2.55e^{-11} NETODA + \varepsilon$$

| Dependent Variable: Log(PCHE) | | | | |
|---|---|---|---|---|
| **Parameter** | **Coefficient** | **Std. Error** | **t-Statistic** | **p-value** |
| **(Intercept)** | 0.138384 | 1.744654 | 0.079319 | 0.937289 |
| **log(CBR)** | -0.287996 | 0.365264 | -0.788459 | 0.436417 |
| **URBAN** | 0.008759 | 0.006612 | 1.324702 | 0.194953 |
| **HDP** | 0.404036 | 0.256835 | 1.573133 | 0.125840 |

| | | | | |
|---|---|---|---|---|
| **log(GDPPC)** | 0.638208 | 0.132780 | 4.806510 | 0.000037 |
| **NETODA** | 2.55.E-11 | 7.74.E-11 | 0.328717 | 0.744580 |
| ***Regression Statistics*** | | | | |
| **Multiple R-squared** | 0.8584 | | | |
| **Adjusted R-squared** | 0.8356 | | | |
| **Residual Standard Error** | 0.452 on 31 degree of freedoms | | | |
| **F-statistic** | 37.59 on 5 and 31 DF | | | |
| **Observations** | 37 | | | |

*Table 10: Regression result of Model 2*

**2. Hypothesis test for HDP binary variable of Model 2**

Similar to Model 1's hypothesis test, the null hypothesis is rejected as the p-value of *HDP* is higher than the significant level (0.12584 > 0.1). Consequently, there is no relationship between *log(PCHE)* and the level of dependent population.

**3. Interaction term**

**Model 3 regression result:**

| **Dependent Variable: Log(PCHE)** | | | | |
|---|---|---|---|---|
| **Parameter** | **Coefficient** | **Standard Error** | **t-Statistic** | **p-value** |
| **(Intercept)** | 0.1658 | 1.8026 | 0.0920 | 0.9273 |
| **log(CBR)** | -0.3013 | 0.4031 | -0.7475 | 0.4606 |
| **URBAN** | 0.0087 | 0.0067 | 1.3014 | 0.2030 |
| **HDP** | 0.5382 | 1.6040 | 0.3356 | 0.7395 |
| **log(GDPPC)** | 0.6413 | 0.1399 | 4.5857 | 0.0001 |
| **NETODA** | 0.0000 | 0.0000 | 0.2996 | 0.7666 |
| **HDP:log(GDPPC)** | -0.0176 | 0.2078 | -0.0848 | 0.9330 |
| ***Regression Statistics*** | | | | |
| **Multiple R-squared** | 0.8585 | | | |
| **Adjusted R-squared** | 0.8302 | | | |
| **Residual Standard Error** | 0.4595 on 30 degrees of freedom | | | |
| **F-statistic** | 35.18 on 6 and 30 degree of freedom | | | |
| **Observations** | 37 | | | |

*Table 11: Regression Result of Model 3*

**4. Hypothesis test for the interaction term**

By having a p-value higher than the significance level (0.9330 > 0.1), the interaction term between *log(GDPPC)* and *HDP* is statistically insignificant.

### 5. Different functional form Model 4

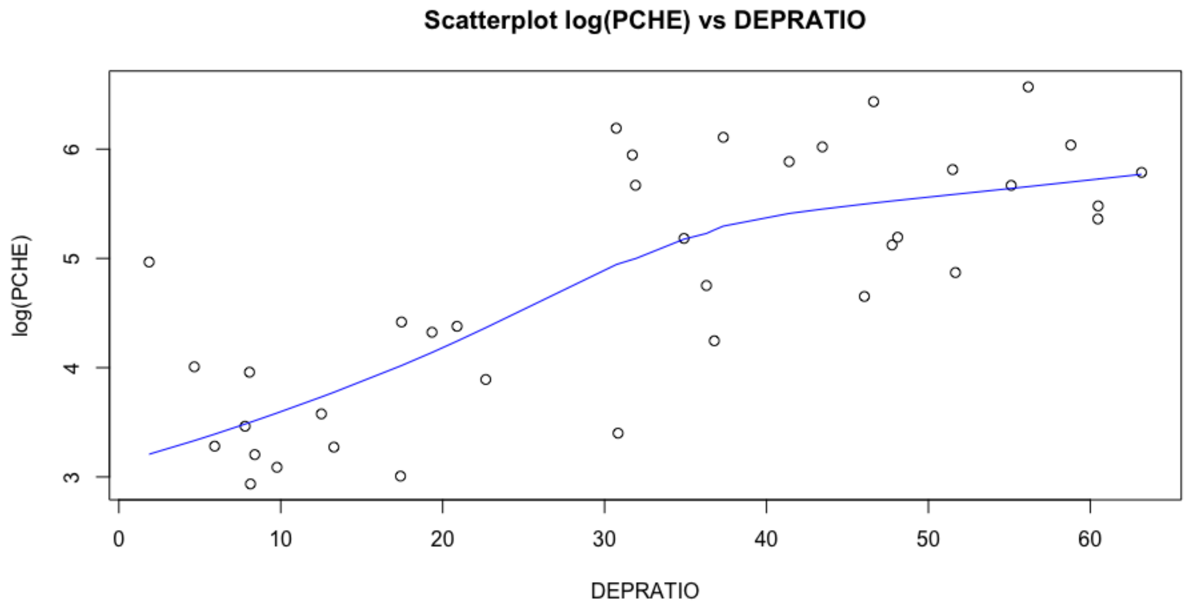



*Figure 8: Scatterplot illustrate the relationship between log(PCHE) and DEPRATIO*

In econometrics, quadratic function is effective in capturing the increasing/decreasing marginal effect and giving straight forward outcome without difficult interpretation, especially in real-life applications that aim to find the maximum or minimum point to achieve maximum yields (Woolridge 2015). Figure 8 suggests *DEPRATIO* has a diminishing effect on *Log(PCHE)* as the curve become more flatten as DEPRATIO increases.

Li et al.'s (2020) hypothesis stated that the initial increase in dependency ratio will result in increasing demand for H.E., which can be afforded by the working-age population for their children/elderly. Nevertheless, when dependency ratio surpasses a certain point, they would cause economic burden on the working class due to human capital shortage and economic slowdown (Cruz & Ahmed 2018), thereby discouraging working class's spending on well-being services for their elderly. Hence, Model 4 should imply DEPRATIO quadratic form to study its decreasing marginal effect on PCHE.

| Dependent Variable: Log(PCHE) | | | | |
|---|---|---|---|---|
| **Parameter** | **Coefficient** | **Standard Error** | **t-Statistic** | **p-value** |
| **(Intercept)** | 1.3224 | 2.2644 | 0.5840 | 0.5636 |
| **log(CBR)** | -0.6394 | 0.5653 | -1.1310 | 0.2670 |
| **URBAN** | 0.0079 | 0.0069 | 1.1468 | 0.2605 |

| | | | | |
|---|---|---|---|---|
| **log(GDPPC)** | 0.6200 | 0.1410 | 4.3967 | 0.0001 |
| **DEPRATIO** | 0.026544 | 0.0182 | 1.4600 | 0.1547 |
| **DEPRATIO-SQ** | -0.000357 | 0.0003 | -1.2273 | 0.2293 |
| **NETODA** | 1.87.E-11 | 0.0000 | 0.2323 | 0.8179 |
| ***Regression Statistics*** | | | | |
| **Multiple R-squared** | 0.8573 | | | |
| **Adjusted R-squared** | 0.8288 | | | |
| **Residual Standard Error** | 0.4612 on 30 degrees of freedom | | | |
| **F-statistic** | 30.05 on 6 and 30 DF | | | |
| **Observations** | 37 | | | |

*Table 12: Quadratic regression result of model 4*

$$log\widehat{(PCHE)} = 1.3224 - 0.6394 * log(CBR) + 0.0079 * URBAN + 0.62 * \log{(GDPPC)} + 0.026544 * DEPRATIO - 0.000357 * \mathrm{DEPRATIOSQ} + 1.87e^{-11} NETODA + \varepsilon$$

### Signficance Testing

The null hypothesis is rejected as the p-value of *DEPRATIO-SQ* is higher than the significant level (0.2293 > 0.1). Hence, with 90% confidence, there is no quadratic relationship between DEPRATIO and Log(PCHE) among assigned developing countries in 2013.

## 6. The final preferred model

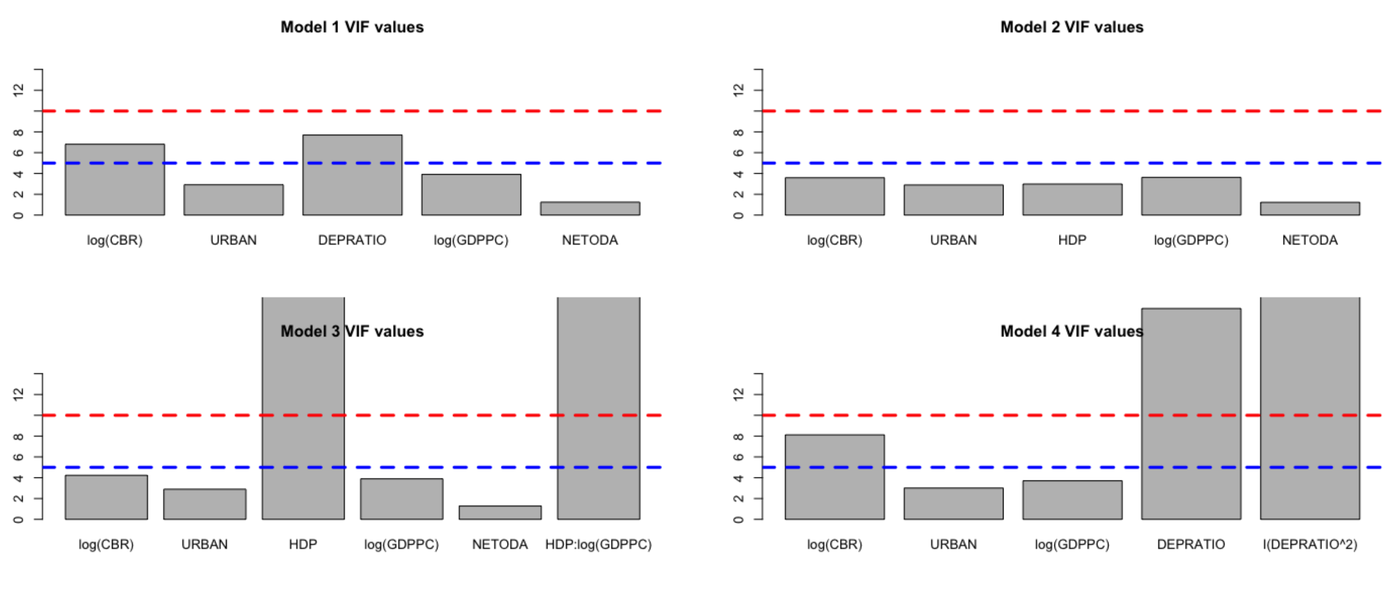



*Figure 9: VIF comparison between four models*

| **H0: Residuals are distributed with equal variance (homoscedasticity)** |
|---|

| Ha: Residuals are distributed with unequal variance (heteroskedasticity) | | | |
|---|---|---|---|
| | **Breusch-Pagan test p-value** | **Critical p-value** | **Result** |
| **Model 1** | 0.3559 | 0.05 | Residuals are homoscedastic |
| **Model 2** | 0.1015 | | |
| **Model 3** | 0.09819 | | |
| **Model 4** | 0.1049 | | |

*Table 13: Breusch-Pagan test for heteroskedasticity of four models*

| OLS Assumption | Model 1 | Model 2 | Model 3 | Model 4 |
|---|---|---|---|---|
| MLR 1 | **All satisfy** | | | |
| MLR 2 | **All satisfy** | | | |
| MLR 3 (Figure 9) | **Two parameters have VIFs > 5** | **All VIFs < 5** | **Problematic multicollinearity** | **Problematic multicollinearity** |
| MLR 4 | **All not satisfy** | | | |
| MLR 5 (Table 13) | **All satisfy** | | | |
| Adjusted R-squared | 0.826 | **0.8356** | 0.8302 | 0.8288 |

*Table 14: Summary Table for OLS assumptions and Adjusted R-squared of four models*

Since all four models are computed from the assigned country-random dataset and have similar parameters, all four models satisfy MLR.1 and 2. Also, all four models satisfy MLR.5 as their Breusch-Pagan test prove their residuals are homoscedastic (Table 13). Contrastively, MLR.4 were breached by all four models as they omit relevant factors that are proved to strongly correlate with PCHE, such as CO2 emission. Consequently, Model 2 is the most preferred model thanks to its lowest multicollinearity and highest adjusted $R^2$ among four models.

| Variable | Coefficients | Interpretation |
|---|---|---|
| **log(CBR)** | -0.2879 | If CBR increases by 1%, PCHE will decreases by -0.2879%, ceteris paribus. |
| **URBAN** | 0.0087 | If URBAN increases by 1 unit, PCHE will increase by roughly 0.87%, ceteris paribus. |

| | | |
|---|---|---|
| **HDP** | 0.4040 | The country with high DEPRATIO has its PCHE higher by 0.4040% compared to country with low DEPRATIO on average, ceteris paribus. |
| **log(GDPPC)** | 0.6382 | If GDPPC increases by 1%, PCHE will increase by 0.6382 %, ceteris paribus. |
| **NETODA** | $2.55e^{-11}$ | If NETODA increases by 1 unit, PCHE will increases by roughly $2.55e^{-9}$ %, ceteris paribus. |

Table 14: Interpretation of Model 2's coefficients.

**PART 5: Conclusion**

1. **Summary**

This empirical research is conducted to examine the determinants of healthcare expenditures per capita (PCHE) in developing and transitional countries, which primarily focus on six prominent predictors comprising Crude Birth Rate, Urbanization, Dependency Ratio, GDP per capita, and net ODA of 37 assigned countries in 2013. The statistical result of all four models indicates that GDP per capita (GDPPC) is the only determinant that have significant and positive impacts on PCHE while other factors have no effects. Resultantly, this finding supports some empirical findings of previous studies regarding the dominance of GDP per capita as the most influential factor of PCHE.

2. **Policy Recommendation**

Based on the findings of this study that GDP is a profound determinant of PCHE, this paper recommends that governments should target the policy that promote the growth of GDP per capita so that it can increase the demand for well-being services. Specifically, the policy should promote strong growths in the size of the workforce and in economic productivity, which can be achieved by improvements in educational system, technological advancements, and job opportunities.

3. **Limitation & Room For Improvements**

It is challenging to determine the factors that influence health spending in developing and transitional nations due to the small sample size of 37 observations. Also, data imputation for missing values also affects the liability of the model. Hence, to enhance the model's estimation, increasing the number of observations must be conducted to mitigate the significant variability of insufficient dataset caused by random sampling. Lastly, the study failed to introduce relevant environmental factors that highly correlates with PCHE, such as $CO_2$ emission, resulting in breaching of zero conditional mean assumption of that affects the unbiasedness of four models, which can be avoided by considering adding all relevant factors.

## REFERENCE


Adeel, U 2016, 'Impact of government expenditure on health sector of Pakistan', *Bulletin of Business and Economics (BBE)*, vol. 5, no. 4, pp. 177-192.

Ali, MEM, Ebaidalla, EM & Rizk, R 2020, 'Foreign aid and out-of-pocket health expenditure in Sub-Saharan Africa: does institutional development matter?', *International journal of sustainable economy*, vol. 12, no. 3, pp. 259-280.

Atuahene, SA, Yusheng, K & Bentum-Micah, G 2020, 'Health expenditure, CO2 emissions, and economic growth: China vs. India'.

Buck, S 2015, 'Discussion of Omitted Variable Bias versus Multicollinearity', *Berkeley Depart of Agricultural & Resource Economics*, Course material, viewed 11 December 2022, < https://are.berkeley.edu/courses/EEP118/current/handouts/OVB%20versus%20Multicollinearity_eep118_sp15.pdf>

Changyong, F, Hongyue, W, Naiji, L, Tian, C, Hua, H & Ying, L 2014, 'Log-transformation and its implications for data analysis', *Shanghai archives of psychiatry*, vol. 26, no. 2, p. 105.

Gareth, J, Daniela, W, Trevor, H & Robert, T 2013, *An introduction to statistical learning: with applications in R*, Spinger.

Han, K, Cho, M & Chun, K 2013, 'Determinants of health care expenditures and the contribution of associated factors: 16 cities and provinces in Korea, 2003-2010', *Journal of Preventive Medicine and Public Health*, vol. 46, no. 6, p. 300.

Hitiris, T & Posnett, J 1992, 'The determinants and effects of health expenditure in developed countries', *Journal of health economics*, vol. 11, no. 2, pp. 173-181.

Hou, JW & Li, K 2011, 'The aging of the Chinese population and the cost of health care', *The Social Science Journal*, vol. 48, no. 3, pp. 514-526.

Hou, JW & Li, K 2011, 'The aging of the Chinese population and the cost of health care', *The Social Science Journal*, vol. 48, no. 3, pp. 514-526.

Kang, H 2013, 'The prevention and handling of the missing data', *Korean journal of anesthesiology*, vol. 64, no. 5, pp. 402-406.

Kraipornsak, P 2017, 'Factors determining health expenditure in the Asian and the OECD countries', *Economics World*, vol. 5, no. 5, pp. 407-417.

Li, L, Du, T & Hu, Y 2020, 'The effect of population aging on healthcare expenditure from a healthcare demand perspective among different age groups: Evidence from Beijing City in the People's Republic of China', *Risk Management and Healthcare Policy*, vol. 13, p. 1403.

Lim, Y, Kim, Y & Connolly, D 2022, 'Assessing the impact of aid on public health expenditure in aid recipient countries', *DEVELOPMENT POLICY REVIEW*.

Nagel, K, Herzer, D & Nunnenkamp, P 2015, 'How does FDI affect health?', *International Economic Journal*, vol. 29, no. 4, pp. 655-679.
Nagel, K, Herzer, D & Nunnenkamp, P 2015, 'How does FDI affect health?', *International Economic Journal*, vol. 29, no. 4, pp. 655-679.
Newhouse, JP 1977, 'Medical-care Expenditure', *Journal of Human Resources*, vol. 12, no. 1, pp. 67-79.

Okunade, AA 2005, 'Analysis and implications of the determinants of healthcare expenditure in African countries', *Health Care Management Science*, vol. 8, no. 4, pp. 267-276.

Toor, IA & Butt, MS 2005, 'Determinants of health care expenditure in Pakistan', *Pakistan Economic and Social Review*, vol., pp. 133-150.

Wooldridge, JM 2015, *Introductory econometrics: A modern approach*, Cengage learning.

Yetim, B, İlgün, G, Çilhoroz, Y, Demirci, Ş & Konca, M 2021, 'The socioeconomic determinants of health expenditure in OECD: An examination on panel data', *International Journal of Healthcare Management*, vol. 14, no. 4, pp. 1265-1269.

Zimmerman, DW 2011, 'A simple and effective decision rule for choosing a significance test to protect against non-normality', *British Journal of Mathematical and Statistical Psychology*, vol. 64, no. 3, pp. 388-409.