**A Comparative Analysis on The Impact of Household Size on Economic Prosperity In Developing Countries**

**ABSTRACT**

*This research explores the relationship between household size and economic prosperity in five developing countries: Brazil, India, Nigeria, Ukraine, and Vietnam. Using secondary data from 2010 to 2021, the study employs multiple regression analysis (MRA), tests for homoscedasticity, normality, autocorrelation, and multicollinearity to assess the impact of health, education, and income indices on average household size. The findings reveal country-specific effects, highlighting the importance of considering socio-economic and cultural differences. Ridge regression is applied to address multicollinearity issues in Nigeria and Vietnam. The research concludes that tailored policy interventions informed by these findings can contribute to more effective development strategies and reduce socio-economic disparities in developing countries.*

***Keywords:*** *HOUSEHOLD, ECONOMIC PROSPERITY, MULTIPLE REGRESSION ANALYSIS*

 **1.0 INTRODUCTION**

Access to healthcare services is a critical determinant of economic prosperity, especially in low- and middle-income countries (LMICs) where disparities in healthcare access are pronounced. Despite efforts to improve accessibility, significant portions of the population, particularly the poor, continue to face barriers in accessing quality healthcare services. These barriers encompass various dimensions, including geographic reach, availability, financial affordability, acceptability, and quality of care. Numerous studies have highlighted the disadvantages faced by the poor in LMICs across these dimensions of healthcare access. Peters et al. (2008), emphasize that while disparities persist, they are not inevitable, and various strategies can enhance access for vulnerable populations. These strategies encompass targeted or universal approaches involving governmental, nongovernmental, or commercial organizations, coupled with diverse financing and service organization methods. Key factors for success include targeted outreach, community engagement, local adaptation, and rigorous monitoring of impacts on disadvantaged populations. Despite improvements in healthcare access, many people in developing countries, particularly the poor, remain underserved and disproportionately affected by diseases. Addressing these disparities requires understanding local factors, enhancing services for the poor, and implementing innovative financing and regulatory strategies. A key challenge is ensuring vulnerable populations have a say in developing and implementing these solutions. Additionally, policymakers are recognizing the link between spatial transformation and economic development, though empirical data remains limited. To fill this gap, researchers in India have integrated poverty analysis and urban economics to map household survey data to specific locations. Findings reveal that spatial factors significantly impact economic outcomes, with productivity varying across locations, not just urban centers. This highlights the need for tailored spatial development policies to address diverse economic conditions within countries. Household size emerges as another crucial factor in understanding socio-economic conditions in developing countries. Scholars such as Adams (2010) and Bongaarts (2001), argue that household size dynamics are influenced by urbanization, industrialization, and educational opportunities. While larger households may experience modest increases in total income, augmenting the number of children often diminishes the family's quality of life, particularly evident in young families with limited resources. Given the complexities of economic development in developing countries, understanding the relationship between household size and economic prosperity is vital.

Easterlin (1967), groundbreaking work challenges conventional paradigms, prompting a paradigm shift in our understanding of the intricate relationship between population growth and economic development. Maralani (2008), Extends this exploration into the temporal dimension, emphasizing the dynamic nature of demographic transitions, with a specific focus on the Indonesian context.

Within the realm of household dynamics and economic impact, Bhattacharjer (2019), study on household size and poverty in Bangladesh becomes a pivotal touchstone. It unravels a web of findings that highlight both positive and negative correlations, emphasizing the necessity of nuanced approaches. The study underscores the importance of adjusting poverty measurements for household size and composition, illuminating potential distortions in the relationship between household size and poverty without such adjustments.

Delving into the enduring effects of early poverty on children's well-being, Krishna et al. (2015), provide an insightful perspective on intergenerational impacts. Carr (2005), Further conceptualizes households as evolving entities, dynamically shaping economic trajectories. In parallel, Hans (2001), general equilibrium model sheds light on efficiency gains achievable through collective rationality within households, challenging traditional models and underscoring the significance of understanding power and knowledge flows.

The nuanced examination of expenditure composition's influence on a nation's steady-state growth rate by Devarajan et al. (1996), and Ahmad et al. (2006), insights into the inverse relationship between household saving and demographic variables contribute to the discourse on formulating effective poverty reduction policies.

Kufenko et al. (2018), comprehensive assessment delves into the effects of changes in household size on the long-term evolution of living standards, emphasizing the need for nuanced considerations. Meanwhile, the United Nations dataset (2019), serves as a cornerstone, elucidating the intertwined nature of household size and composition with various social and economic processes, crucial for achieving sustainable development goals.

Yue and Martin (2015), Examination, challenges traditional rural-urban dichotomies by introducing a nuanced understanding of regional economic dynamics in India.

Barnes et al. (2005), focus on the urban household energy transition illustrates the shift from traditional to modern fuels and accentuates the socio-environmental repercussions. Infrastructure needs and changing household sizes are addressed by Zhu et al. (2022), study on poverty reduction policies in Vietnam and Bradbury et al. (2014), exploration of environmental implications. Notably, these studies underscore the complexity of navigating economic growth while considering the evolving dynamics of household structures.

However, despite the wealth of existing literature, notable gaps persist. A more nuanced exploration of demographic factors, regional disparities, and cultural nuances is crucial. Future research could delve into the long-term implications of changing household sizes on resource consumption, environmental sustainability, and social equity.

Adding a unique perspective, Mumford (2016), study introduces national wealth accounting as a metric for determining the underlying productive base's size. This study, conducted across Asian countries, correlates GDP growth with wealth growth but notes instances where these metrics show different signs, highlighting the importance of measuring sustainability.

In conclusion, the dynamic interplay between household sizes and economic prosperity in developing countries necessitates a multidimensional and nuanced approach. This comprehensive literature review not only contributes to the ongoing discourse but also lays a solid foundation for future investigations into the complex dynamics of household sizes and their profound impact on the economic trajectories of developing nations.

**2.0 Methodology**

The primary source of data for this study is reputable online databases, including the World Bank and the United Nations. These databases provide comprehensive demographic, economic, and social indicators for a range of developing countries, ensuring a robust dataset for analysis.

**3.2 MULTIPLE REGRESSION ANALYSIS**

Our analytical journey transitions from a single explanatory variable to the complexity of a scenario where the dependent variable *Y* intertwines with multiple independent variables a phenomenon aptly named the multiple linear regression model. In this expansive model, we embrace the possibility that the study variable is influenced by more than one explanatory variable, thereby allowing for shapes beyond linear relationships, albeit constrained within linear forms.

**3.2.1 THE MODEL**

Let Y denote the dependent variable, intricately linked to *k* independent variables X1,*X*2​,…,*Xk*​ through parameters *β*. We express this relationship as:

*Y*=*β*0​+*β*1​*X*1​+*β*2​*X*2​+…+*βk*​*Xk*​+*ϵ. 3.1*

Representing it in matrix form:

*Y*=*Xβ*+*ϵ. 3.2*

 3.3

Linear model in matrix terms:

 3.4

**3.2.2 PROPERTIES OF THE MULTIPLE LINEAR REGRESSION MODEL**

Expressing each of the *n* observations in (3.3), we can represent them individually as:

 3.5

Utilizing matrices, the above expressions can be represented as:

 3.6

This can be further simplified as:



Hence,

 3.7



Here, *X* is a matrix of order *m* × (*i* +1), with *m* >*i* +1 and *ρ*(*X*) = *i* +1. The *α* Parameters are the regression coefficients, sometimes referred to as partial regression coefficients. These coefficients indicate the change in *E*(*Y*) with a unit change in a particular *X* when other *X* variables are held constant. The influence of each *X* on *E*(*Y*) is analyzed in the presence of other *X*'s, emphasizing the importance of considering all relevant variables in the model. If one *X* is removed, the parameters will be altered, highlighting the interconnectedness of the variables in the model.

This matrix representation encapsulates the linear regression model, where *β* denotes the regression coefficients associated with *X*1​,*X*2​,…,*Xk*​, and *ϵ* symbolizes the random error component capturing the variance between the observed and the fitted linear relationship.

Note that the regression coefficient β1, for instance, signifies the anticipated change in *Y* per unit alteration in the Education Index.

**3.2.3 STATISTICAL MODEL**

*Y* = *β*0​ + *β*1​(Education Index) + *β*2​(Health Index) + *β*3​(Income Index) + *ϵ*

Estimation with Ordinary Least Squares (OLS)

Ordinary Least Squares (OLS) aims to minimize the squared differences between the observed and predicted *Y*. The OLS estimator is computed as:

 *3.8*

where:

*Y*  Is the dependent variable.

β0 Is the constant term.

β1, *β*2 ​, *β*3​ are coefficients for Education Index, Health Index, and Income Index.

X1, *X*2​ , *X*3​ are the Education Index, Health Index, and Income Index.

*ϵ* represents the error term. This model allows us to delve into the intricate dynamics between the Education Index, Health Index, and Income Index, unraveling their impact on the dependent variable *Y*.

**3.2.4 ASSUMPTIONS OF MULTIPLE LINEAR REGRESSION**

The reliability of Regression Analysis hinges upon several crucial assumptions about the variables involved. When these assumptions are unmet, the outcomes may lack trustworthiness, leading to Type I or Type II errors, or an over- or under-estimation of significance or effect size(s). To ensure the validity of estimation and inference procedures, certain conditions must be fulfilled:

* **Non-Randomness of Explanatory Variables:** The Xi variables are non-random (fixed) and non-stochastic. Often referred to as explanatory or predictor variables, they play a pivotal role in elucidating the variability in the dependent variable *Y*.
* **Independent and Normally Distributed Error Term Ui:** The error term Ui is independent and follows a normal distribution with a mean of zero and variance equal to *σ*2. This is denoted by *E* (*U*)=0 and Var (*U*)=*σ*2.
* **Non-Multicollinearity:** Assumption that there is no linear relationship between some or all independent variables. Multicollinearity can introduce challenges in interpreting the individual impact of predictors.
* **No Autocorrelation:** Absence of correlation among successive terms in the model, implying that *Cov* (*Ui*​,*Uj*​) = 0.
* **Independence of *Y* Values:** The values of *Y* are independent. This means that the selection of *Y* for a specific *X* value is not contingent on the selection of another *X* value.

**3.2.5 LINEARITY**

Standard multiple regression attains accurate estimations when the relationships between dependent and independent variables are linear. Given the prevalence of non-linear relationships in the social sciences (e.g., anxiety), it becomes imperative to scrutinize analyses for nonlinearity. Failure to address nonlinearity risks underestimating the true relationship. This carries the potential for increased Type II errors for that independent variable and, in multiple regression, an elevated risk of Type I errors for other variables sharing variance.

* **Detection Method:** Examination of residual plots, portraying standardized residuals as a function of standardized predicted values, provides a valuable means of detecting nonlinearity, readily available in most statistical software.

**3.2.6 HOMOSCEDASTICITY**

Homoscedasticity implies that the variance of errors remains consistent across all levels of the independent variable. Any divergence in error variance across different values of the independent variable signals heteroscedasticity. While slight heteroscedasticity may have minimal impact on significance tests, pronounced cases can distort findings and significantly weaken the analysis, increasing the risk of a Type I error. Visual examination of a plot of standardized residuals against the regression standardized predicted value can assess homoscedasticity. Ideally, residuals are randomly scattered around 0, providing an even distribution. Detecting heteroscedasticity involves observing irregularities in this pattern, such as a bow-tie or fan shape. More formal tests like the Goldfeld-Quandt test or Glejser tests for heteroscedasticity can be employed. The presence of skewness in independent variables may contribute to heteroscedasticity, and variable transformations can mitigate this effect.

* **STUDENTIZED BREUSCH-PAGAN TEST**

The studentized Breusch-Pagan test is a statistical test used to assess the presence of heteroscedasticity in a regression model. Heteroscedasticity refers to the situation where the variance of the errors (residuals) in a regression model is not constant across all levels of the independent variables. The Breusch-Pagan test is a specific test for detecting this issue.

The formula for the Breusch-Pagan test statistic is as follows:

***LM*=*n*∗*R*2** 3.9

Where:

* 𝐿𝑀 is the Lagrange Multiplier (LM) statistic.
* 𝑛 is the sample size.
* *R*2 is the coefficient of determination from an auxiliary regression of the squared residuals from the original regression on the independent variables from the original regression.

The LM statistic follows a chi-square distribution with degrees of freedom equal to the number of independent variables in the auxiliary regression.

To studentize the LM statistic, you divide it by its standard deviation. The studentized Breusch-Pagan test is based on this studentized LM statistic and is used to assess the significance of the heteroscedasticity in the model.

**3.2.7 NORMALITY OF RESIDUALS**

Regression assumes normally distributed residuals. Deviations from normality can be identified through visual inspection of data plots, skewness, kurtosis, Shapiro-Wilks W Test, and P-P plots. Inferential statistics on normality can be obtained using the Kolmogorov-Smirnov test. Outliers, detected through histograms or z-scores, can also impact normality and distort relationships in the analysis. Addressing these issues ensures the robustness of regression analysis results.

In regression analysis, the assumption of normality for residuals is crucial. Non-normally distributed variables, characterized by high skewness or kurtosis, may lead to uncorrelated

errors. Autocorrelation is a common issue, especially in time series data, where positive autocorrelation is frequently observed.

**CONCLUSION**: If Da is the critical value, then P(DD)=1−4D can be utilized to test whether a random variable aligns with a specific distribution function F(X). Critical values du and dl depend on the significance level (u), the number of observations, and the predictors in the regression equation. Detecting non-normality involves inspecting data plots, assessing skewness, kurtosis, conducting the Shapiro-Wilks W Test, and examining P-P plots. Addressing outliers, identified through histograms or z-scores, is essential for maintaining the integrity of relationships and significance tests.

**3.2.8 KOLMOGOROV-SMIRNOV TEST FOR NORMALITY**

Definition:

Let X1,X2,…,Xn be an ordered sample with X1≤X2≤…≤Xn, and define S(X) as follows: 3.10

Now, suppose that the sample comes from a population with a cumulative distribution function F(X) and define D as follows:

D = max ∣F(X)−S(X)∣ ≤ D 3.11

Observation:

It can be shown that D does not depend on F since S(X) depends on the sample chosen. D is a random variable. The objective is to use D as a way to estimate F(X).

**3.2.9 AUTOCORRELATION**

The assumption of independent and uncorrelated error terms is fundamental in regression analysis. Deviation from this assumption, where errors exhibit a non-zero covariance, indicates the presence of autocorrelation. The Durbin-Watson test serves as a diagnostic tool for detecting autocorrelation in regression residuals.

* **DURBIN-WATSON TEST**

The Durbin-Watson statistic is employed to identify autocorrelation, examining the relationship between values with a time lag in the residuals of regression analysis. Named after James Durbin and Geoffrey Watson, this test assesses whether errors are serially correlated.

The Durbin-Watson test evaluates autocorrelation, testing the hypothesis:

H0:*ρ*=0

H1:*ρ*>0

The test statistic (*d*) is compared to critical values (d1 and d2):

If *d* < *d*1​ , There is statistical evidence of positive autocorrelation.

If *d* > *d*2​ , There is no statistical evidence of positive autocorrelation.

If d1 ≤ d ≤ d2 , The test is inconclusive.

Positive serial correlation implies that a positive error for one observation increases the likelihood of a positive error for another observation.

**3.3 MULTICOLLINEARITY**

Multicollinearity refers to the existence of linear relationships among explanatory variables in a regression model. It can be perfect or less than perfect, leading to challenges in estimating regression coefficients.

**DETECTION OF MULTICOLLINEARITY**

* **High R² but Few Significant t Ratios:** High overall model fit (R²) coupled with few statistically significant t ratios for individual coefficients suggest multicollinearity.
* **High Pair-wise Correlations among Regressors:** Elevated zero-order correlation coefficients between two regressors may indicate multicollinearity, but caution is needed as high correlations are not necessary for collinearity.
* **Eigenvalues and Condition Index:** The condition number (*K*) and condition index (*CI*) derived from eigenvalues provide insights into multicollinearity. A moderate to strong multicollinearity is suggested if *K* is between 100 and 1000 or if *CI* is between 10 and 30.
* **Tolerance and Variance Inflation Factor (VIF):** The VIF, measuring the increase in the variance of an estimated regression coefficient due to multicollinearity, is a useful indicator. A VIF exceeding 10 indicates significant collinearity. Tolerance (*TOL*) is another measure, where values close to zero suggest greater collinearity.

Addressing multicollinearity is crucial to ensure the reliability of regression coefficient estimates and the overall model fit.

**3.4 ANALYSIS OF VARIANCE (ANOVA)**

Analysis of Variance (ANOVA) plays a crucial role in understanding the variability within a regression model and conducting tests of significance. This statistical method involves hypothesis testing for the significance of partial regression coefficients and the overall significance of the regression model.

**HYPOTHESIS TESTING FOR PARTIAL REGRESSION COEFFICIENTS**

* **Null Hypothesis (Ho):** Bi (Partial regression coefficient) is not significant for *i*=1,2,3,…,*k*.
* **Alternative Hypothesis (H1):** Bi is significant for at least one *i*.

Test Statistic

 3.12

**DECISION RULE**

If the p-value is less than the chosen level of significance (*α*), reject the null hypothesis; otherwise, do not reject the null hypothesis.

**Testing Hypothesis using the F-test**

* **Null Hypothesis (Ho):** The regression model is not significant; none of the explanatory variables has explanatory power.
* **Alternative Hypothesis (H1):** The regression model is significant; at least one explanatory variable has explanatory power.

**PARTITIONING THE SUM OF SQUARES**

SSR = SS Total−SSE

SSR=SS Total−SSE

Where:

SST=Total Sum of Squares. 3.13

SSR=Regression Sum of Squares. 3.14

SSE= Error Sum of Squares. 3.15

MSR=SSR/1-p 3.16

**ANOVA Table**



Multiple Coefficient of Determination

 3.17

The multiple coefficient of determination indicates the proportion of the total variation in the dependent variable explained by all the explanatory variables, ranging between 0 and 1.

**3.5 DATA TRANSFORMATIONS**

Data transformations involve applying a mathematical function (e.g., squaring the data) to each data point. Transformations are necessary when the data is excessively skewed positively or negatively.

* **REASONS FOR USING TRANSFORMATION:**

The most frequent reason for researchers to transform their data is to make the distribution "normal," fulfilling one of the assumptions for parametric means comparison. Other reasons include creating more informative graphs, better outlier identification, increasing the sensitivity of statistical tests, and ensuring equal spreads, linear relationships, additive relationships, etc.

* **CHOICE FOR TRANSFORMATION:**

Several types of data transformations exist, but in this context, the Natural Log Transformation is used due to its significant effect on distribution shape. The natural logarithm of a number, denoted as ln(*x*) or log*e*​(*x*), is taken with a base *e*, Euler's constant (approximately equal to 2.71828).

ln(*x*). 3.18

**3.6 RIDGE REGRESSION**

Ridge regression, also known as Tikhonov regularization, is a type of linear regression that includes a regularization term to penalize large coefficients. This regularization helps to prevent overfitting, especially in cases where there are many correlated predictors or when the number of predictors is greater than the number of observations.

The ridge regression model minimizes the following objective function:

Minimize ∥ **y** − **X** *β*∥2 + *λ* ∥*β*∥2

where:

* **y**  is the vector of observed values.
* 𝑋 is the matrix of input features.
* 𝛽is the vector of coefficients.
* 𝜆 is a regularization parameter (also called the ridge parameter or shrinkage parameter).

The term ∥𝑦−𝑋𝛽∥2 Represents the sum of squared residuals (the usual objective in ordinary least squares regression), while 𝜆∥𝛽∥2 is the penalty term that adds the sum of squared coefficients, scaled by 𝜆.

**4.0 Data Analysis & Results**

The data collected for this research work as earlier stated are secondary data on average household sizes and economic indicators (health, education, and income) from 2010 to 2021 for Brazil, India, Nigeria Ukraine, and Vietnam.

**4.1 MULTIPLE REGRESSION ANALYSIS**

**Table 1. Multiple Regression coefficients**

Consolidated Coefficients Table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Country** | **Intercept** | **Health Index** | **Education Index** | **Income Index** |
| **BRA** | 1.6199 | 0.0576 | -14.8111 | 15.3776 |
| **IND** | 14.04 | 20.74 | 10.53 | -49.71 |
| **NIG** | 7.9101 | -3.8783 | 2.3942 | -0.6156 |
| **UKR** | -25.396 | 2.639 | 36.033 | -3.559 |
| **VET** | -1.179 | 1.274 | 26.639 | -20.434 |

**Interpretation**

1. **Brazil (BRA)**
	* **Intercept (1.6199)**: The baseline average household size when all indices are zero.
	* **Health Index (0.0576)**: A slight positive effect; as the Health Index increases by 1 unit, the average household size increases by approximately 0.058.
	* **Education Index (-14.8111)**: A strong negative effect; as the Education Index increases by 1 unit, the average household size decreases by approximately 14.811.
	* **Income Index (15.3776)**: A strong positive effect; as the Income Index increases by 1 unit, the average household size increases by approximately 15.378.
2. **India (IND)**
	* **Intercept (14.04)**: The baseline average household size when all indices are zero.
	* **Health Index (20.74)**: A significant positive effect; as the Health Index increases by 1 unit, the average household size increases by approximately 20.74.
	* **Education Index (10.53)**: A positive effect; as the Education Index increases by 1 unit, the average household size increases by approximately 10.53.
	* **Income Index (-49.71)**: A significant negative effect; as the Income Index increases by 1 unit, the average household size decreases by approximately 49.71.
3. **Nigeria (NIG)**
	* **Intercept (7.9101)**: The baseline average household size when all indices are zero.
	* **Health Index (-3.8783)**: A slight negative effect; as the Health Index increases by 1 unit, the average household size decreases by approximately 3.878.
	* **Education Index (2.3942)**: A slight positive effect; as the Education Index increases by 1 unit, the average household size increases by approximately 2.394.
	* **Income Index (-0.6156)**: A very small negative effect; as the Income Index increases by 1 unit, the average household size decreases by approximately 0.616.
4. **Ukraine (UKR)**
	* **Intercept (-25.396)**: The baseline average household size when all indices are zero.
	* **Health Index (2.639)**: A slight positive effect; as the Health Index increases by 1 unit, the average household size increases by approximately 2.639.
	* **Education Index (36.033)**: A strong positive effect; as the Education Index increases by 1 unit, the average household size increases by approximately 36.033.
	* **Income Index (-3.559)**: A slight negative effect; as the Income Index increases by 1 unit, the average household size decreases by approximately 3.559.
5. **Vietnam (VET)**
	* **Intercept (-1.179)**: The baseline average household size when all indices are zero.
	* **Health Index (1.274)**: A slight positive effect; as the Health Index increases by 1 unit, the average household size increases by approximately 1.274.
	* **Education Index (26.639)**: A strong positive effect; as the Education Index increases by 1 unit, the average household size increases by approximately 26.639.
	* **Income Index (-20.434)**: A strong negative effect; as the Income Index increases by 1 unit, the average household size decreases by approximately 20.434.

**Summary**

The coefficients suggest that different factors influence average household size differently across countries. For instance, Education Index has a strong negative effect on household size in Brazil, while it has a strong positive effect in Ukraine and Vietnam. Similarly, Income Index shows a significant negative impact on household size in India and Vietnam, but a positive impact in Brazil. This indicates that socio-economic and cultural differences might play a crucial role in determining household sizes in different countries.

**4.3 LINEARITY PLOTS**

**Fig. 1. Normal Residual Q-Q Plot Plot**



**4.4 TEST FOR HOMOSCEDATICITY**

**Table 2.** **Studentized Breusch-Pagan test**

|  |  |  |  |
| --- | --- | --- | --- |
| **Country** | **BP-value** | **df** | **p-value** |
| **Brazil** | 8.8268 | 3 | 0.03168 |
| **India** | 1.4194 | 3 | 0.701 |
| **Nigeria** | 5.2008 | 3 | 0.1577 |
| **Ukraine** | 3.7705 | 3 | 0.2873 |
| **Vietnam** | 3.5319 | 3 | 0.3166 |

**Interpretation:**

* + For Brazil, the p-value (0.03168) is less than 0.05, indicating evidence against homoscedasticity, suggesting the presence of heteroscedasticity.
	+ For India, Nigeria, Ukraine, and Vietnam, the p-values are greater than 0.05, suggesting no evidence against homoscedasticity, indicating that the assumption of constant variance of residuals is not violated significantly.

**4.5 TEST FOR NORMARLITY OF RESIDUALS**

**Fig 2. Normal Residual P-P Plot**

     

**Interpretation:**

* In the normal P-P plot **Fig 2**, some points deviate from the diagonal line, particularly at the tails of the distribution, indicating the residuals likely aren't perfectly normally distributed.
* Diagnostic tests like the Kolmogorov-Smirnov test in **4.5.1** provide a more formal assessment of normality.

**4.5.1 KOLMOGOROV-SMIRNOV TEST FOR NORMALITY**

**Table 3. Kolmogorov-Smirnov Test**

|  |  |  |  |
| --- | --- | --- | --- |
| **Country** | **D-value** | **p-value** | **Alternative Hypothesis** |
| **Brazil** | 0.19607 | 0.6766 | Two-sided |
| **India** | 0.24107 | 0.4217 | Two-sided |
| **Nigeria** | 0.12939 | 0.9726 | Two-sided |
| **Ukraine** | 0.27431 | 0.2735 | Two-sided |
| **Vietnam** | 0.088079 | 0.9998 | Two-sided |

The exact one-sample Kolmogorov-Smirnov test was conducted to assess the normality of the residuals.

**Interpretation:**

* + For Brazil, India, Nigeria, and Ukraine, the p-values are all greater than conventional significance levels (0.05), indicating insufficient evidence to reject the null hypothesis. Thus, the residuals for these countries likely follow the specified theoretical distribution.
	+ For Vietnam, the p-value is extremely high (0.9998), providing strong evidence in favor of the null hypothesis. Therefore, the residuals for Vietnam most likely conform well to the specified theoretical distribution.

**4.6 TEST FOR AUTOCORRELATION**

**Table 4. Durbin Watson test**

|  |  |
| --- | --- |
| **Country** | **Durbin-Watson Statistic** |
| Brazil | 2.228326 |
| India | 1.723313 |
| Nigeria | 0.919588 |
| Ukraine | 1.035475 |
| Vietnam | 0.8338283 |

**Interpretations:**

* + For Brazil, India, and Ukraine, the Durbin-Watson statistics are all around 2, suggesting no significant autocorrelation in the residuals.
	+ For Nigeria, the Durbin-Watson statistic is below 2, indicating positive autocorrelation, which may violate the assumption of independence of residuals.
	+ For Vietnam, the Durbin-Watson statistic is significantly below 2, indicating strong positive autocorrelation in the residuals. This suggests that neighboring residuals are correlated, potentially violating the assumption of independence.

**4.7 TESTS FOR MULTICOLINEARITY**

**Table 5. Tolerance and Variance Inflation Factor (VIF)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Country** | **Health index** | **Education index** | **Income index** |
| **Brazil** | 1.459813 | 1.814887 | 1.308657 |
| **India** | 2.202729 | 6.119807 | 5.437874 |
| **Nigeria** | 72.560249 | 73.106138 | 1.152528 |
| **Ukraine** | 3.142373 | 3.019199 | 1.113740 |
| **Vietnam** | 1.469003 | 54.899732 | 55.919559 |

**Interpretation:**

* + For Brazil, all VIF values are comfortably below 10, indicating low multicollinearity among the independent variables.
	+ For India, all VIF values are comfortably below 10, indicating low multicollinearity among the independent variables.
	+ For Nigeria, both Health index and Education index have extremely high VIF values, well above 10, indicating severe multicollinearity issues.
	+ For Ukraine, all VIF values are below 10, suggesting low multicollinearity.
	+ For Vietnam, while Health index has a reasonable VIF, both Education index and Income index have extremely high VIF values, indicating severe multicollinearity issues.

In summary, Brazil, India and Ukraine seem to have low multicollinearity, Vietnam show moderate to severe multicollinearity issues, and Nigeria exhibits very severe multicollinearity problems, particularly with Health index and Education index.

**4.9 TRANSFORMATION OF DATA**

**4.9.1 USING LOG TRANSFOMATION**

Transformation of the dataset was done as stated in chapter three for corrective measures on Homoscedasticity **4.4** Brazil.

**Results:**

**Table 6.** **Studentized Breusch-Pagan test for Transformed data for Brazil**

|  |  |  |  |
| --- | --- | --- | --- |
| **Country** | **BP-value** | **df** | **p-value** |
| **Brazil** | 4.4107 | 3 | 0.2204 |

**Interpretations:**

* + After Transformation For Brazil, the p-value (0.2204) is greater than 0.05, indicating evidence of homoscedasticity and fulfilling the required assumption.

**4.10 USE OF RIDGE REGRESSION**

Due to high multicollinearity in the data of both Nigeria and Vietnam, ridge regression was applied.

**Results:**

**Table 7. Ridge Regression Co-efficient**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Country | Intercept | Health Index | Education Index | Income Index |
| Nigeria | 5.2805341 | 2.4665818 | 0.7208227 | -0.1157075 |
| Vietnam | 7.423812 | -1.666689 | -2.265419 | -1.477154 |

**Interpretation:**

1. **Nigeria (NIG)**
	* **Intercept (5.2805341)**: The baseline average household size when all indices are zero.
	* **Health Index (2.4665818)**: A positive effect; as the Health Index increases by 1 unit, the average household size decreases by approximately **2.4665818**.
	* **Education Index (0.7208227)**: A slight positive effect; as the Education Index increases by 1 unit, the average household size increases by approximately **0.7208227**.
	* **Income Index (-0.1157075)**: A very small negative effect; as the Income Index increases by 1 unit, the average household size decreases by approximately **-0.1157075**.
2. **VIETNAM (VET)**
	* **Intercept (7.423812)**: The baseline average household size when all indices are zero.
	* **Health Index (-1.666689)**: A negative effect; as the Health Index increases by 1 unit, the average household size in decreases by approximately **-1.666689**.
	* **Education Index (-2.265419)**: A negative effect; as the Education Index increases by 1 unit, the average household size decreases by approximately **-2.265419**.
	* **Income Index (-1.477154)**: A slight negative effect; as the Income Index increases by 1 unit, the average household size decreases by approximately **-1.477154**.

**4.10 COMPARATIVE ANALYSIS**

**1. Homoscedasticity (Table 2):**

* Brazil stands out with a significant p-value (0.03168), indicating heteroscedasticity, while other countries don't exhibit significant evidence against homoscedasticity.

**2. Normality of Residuals (Table 3):**

* All countries show p-values above 0.05, suggesting insufficient evidence to reject the null hypothesis, indicating that residuals likely follow the specified theoretical distribution.

**3. Autocorrelation (Durbin-Watson Statistic):**

* Brazil, India, and Ukraine show no significant autocorrelation.
* Nigeria and Vietnam display positive autocorrelation, especially Vietnam with a significantly low Durbin-Watson statistic, indicating strong positive autocorrelation.

**4. Multicollinearity (Table 6):**

* Brazil, India and Ukraine have low multicollinearity.
* Vietnam exhibit moderate to severe multicollinearity issues, particularly with Education and Income indices.
* Nigeria shows very severe multicollinearity issues, especially with Health and Education indices.

**5. Analysis of Variance (Table 7):**

* Brazil: Health and Education indices are significant predictors, while the Income index is not.
* India: Education and Income indices are significant predictors, while the Health index is not.
* Nigeria: Health index is highly significant, Education index is significant but to a lesser extent, while the Income index is not significant.
* Ukraine: None of the predictors are significant.
* Vietnam: All predictors are highly significant.

The research investigates the impact of household size on economic prosperity across five developing countries: Brazil, India, Nigeria, Ukraine, and Vietnam. Secondary data from 2010 to 2021 were analyzed, focusing on average household sizes and economic indicators like health, education, and income.

Multiple regression analysis revealed varying effects of health, education, and income indices on household size across countries. Tests for homoscedasticity, normality, autocorrelation, and multicollinearity were conducted to assess the validity of regression models. Notably, Nigeria and Vietnam exhibited severe multicollinearity issues, leading to the application of ridge regression for these countries. The results from ridge regression provided insights into the coefficients' effects in the presence of multicollinearity.

The analysis also included interpretation of ANOVA tables, assessing the significance of predictors on average household size. It was observed that different factors significantly influenced household size in each country, emphasizing the importance of considering socio-economic and cultural differences.

**5.2 CONCLUSION**

The findings highlight the complex relationship between household size and economic prosperity in developing countries. While some factors like education may have a strong negative impact on household size in one country, they could have a positive effect in another. These disparities underscore the need for tailored policy interventions addressing unique socio-economic contexts.

Despite methodological challenges such as multicollinearity and heteroscedasticity, the research provides valuable insights into the dynamics of household size and economic indicators. Policymakers can leverage these findings to formulate more effective development strategies aimed at optimizing resource allocation and reducing socio-economic disparities.

**5.3 RECOMMENDATION**

**Building on the analysis, we offer the following recommendations to inform policy and development strategies in developing countries:**

1. **Policy Tailoring**: Develop tailored policies considering the unique socio-economic contexts of each country to address challenges related to household size and economic prosperity.
2. **Data Improvement**: Enhance data collection methods to mitigate issues like multicollinearity and heteroscedasticity, ensuring more robust analyses and reliable results.
3. **Longitudinal Studies**: Conduct longitudinal studies to capture long-term trends and dynamics in household size and economic indicators, providing a comprehensive understanding of the relationship over time.

**REFERENCES**

Adamu, I., Ogbonna, C. J., Ubah, S. O., Irinews, A. A., Muhammad, A., & Ahmed, S. (2023). A comparative study of Bitcoin's price prediction using regression models. *Advances in Research,* 24(6), 259-271. DOI: 10.9734/AIR/2023/v24i61008

Adamu, I., Ogbonna, C. J., Adamu, Y., & Zakari, Y. (2021). Modeling coronavirus pandemic using univariate and multivariate models: The Nigerian perspective. *Asian Journal of Probability and Statistics*, 15(4), 134-143. DOI: 10.9734/AJPAS/2021/v15i430370 ISSN: 2582-0230

Ahmad, M., Atiq, Z., Alam, S., & S Butt, M. (2006). The impact of demography, growth and public policy on household saving: A case study of pakistan. *Asia-Pacific Development Journal, 13(2), 57–71.*

Allen, C.K. (1988). Economic consequences of population change in the third world. Journal of Economic Literature; *Nashville, 26(4), 1685.*

Alvord, C. of A. D. C., LLC. (2023, May 17). Business Corner What is economic prosperity and why should our residents care? Gold Country Media. https://goldcountrymedia.com

Barnes, D.F., Krutilla, K., & Hyde, W.F. (2005). The Urban Household Energy Transition: Social and Environmental Impacts in the Developing World (1st ed.). Routledge.

Bassirou Amadou, D., & Quenton, W. (2007). Demographic transition towards smaller household sizes and basic infrastructure needs in developing countries. *Economics Bulletin, 15(11), 1–11.*

Berry, W.D. & Feldman, S. (1985). Multiple regression in practice. Sage University paper series on quantitative application in the social sciences, series No 07-050 Newbury Park, C.A : Sage publication Inc.

Bhattacharjer, S. (2019). Household Size and Poverty. *Jagannath University Journal of Arts, 9(1), 166–177.*

Bloom, E.D., & G Williamson, J. (1998). Demographic transitions and economic miracles in emerging Asia. *The World Bank Economic Review, 12(3), 419–455.*

Bongaarts, J. (2001). Household size and composition in the developing world in the 1990s.

 *Population Studies*, *55*(3), 263–279.

Bradbury, M., Peterson, M.N. & Liu, J. (2014). Long-term dynamics of household size and their environmental implications. *Popul Environ 36, 73–84.*

Carr, E.R. Development and the Household: Missing the point?. *GeoJournal 62, 71–83 (2005).*

Cohen, J.M. (2021). New conceptions of sufficient home size in High-Income countries: Are we approaching a sustainable consumption transition? *Housing, Theory and Society, 38(2), 173–203.*

Devarajan, S., Swaroop, V., & Zou, H. (1996). The composition of public expenditure and economic growth. *Journal of Monetary Economics, 37(2), 313–344.*

Easterlin, R. A. (1967). Effects of Population Growth on the Economic Development of Developing Countries. *The ANNALS of the American Academy of Political and Social Science, 369(1), 98-108.*

 Ellen, I. G., & O’Flaherty, B. (2007). Social Programs and Household Size: Evidence from York City. *Population Research and Policy Review, 26(4), 387–409.*

Espenshade, T. J., Kamenske, G., & Turchi, B. A. (1983). Family size and economic welfare. *Family planning perspectives, 15(6), 289–294.*

Hafeez-ur-Rehman, E.A R, K., & Muhammad-abrar-ul-haq, M. (2015). Determinants of Rural Household poverty: *The role of household socioeconomic empowerment. American-Eurasian Journal of Agricultural & Environmental Sciences, 15(1), 93–98.*

Hans, H. (2001). Household decisions and equilibrium efficiency. *International Economic Review, 41(4), 835–847.*

Howell, R. T. and Howell, C. (2008). The relation of economic status to subjective well-being in developing countries: a meta-analysis.. *Psychological Bulletin, 134(4), 536-560.*

International Monetary Fund. (2022). World Economic and Financial Surveys: World Economic Outlook Database—WEO Groups and Aggregates Information, April.

Kaelble, H., & Thomas, M. (1991). Income distribution in historical perspective. Cambridge University press.

Krishna, A., Oh, J., Lee, J. K., Lee, H. Y., Perkins, J. M., Heo, J., Ro, Y. S., & Subramanian, S. V. (2015). Short-term and long-term associations between household wealth and physical growth: a cross-comparative analysis of children from four low- and middle-income countries. *Global health action, 8, 26523.*

Kufenko, V., Geloso, V., & Prettner, K. (2018). Does size matter? Implications of household size for economic growth and convergence. Hohenheim Discussion Papers in Business, Economics and Social Sciences.

Lezear, P.E., & T Michael, R. (1980). Family size and the distribution of real per capita income. *The American Economic Review, 70(1), 91–107.*

Maralani V. (2008). The changing relationship between family size and educational attainment over the course of socioeconomic development: *evidence from Indonesia. Demography, 45(3), 693–717.*

Michal, M.J. (1973). Size-Distribution of earnings and household incomes in small socialist countries. *Review of Income and Wealth, 19(4), 407–427.*

Mumford, J.K. (2016). Prosperity, sustainability and the measurement of wealth. *Asia and the Pacific Policy Studies, 3(2), 226–234.*

Peters, H.D., Garg, A., Bloom, G., G Walker, D., R Brieger, W., & M Rahman, H. (2008). Poverty and access to health care in developing countries*. Annals of the New York Academy of Sciences, 1136(1), 161–171.*

Safllios-Rothschild, C. (1980). The role of the family in development: Economic decision making in low-income families and its implications for the design of development programs, Finance & Development, 0017(004), A012.

Tripathi. (2020). Does economic development reduce household size? Evidence From India, *Journal of Policy Modeling, 42(5), 0161–8938.*

United Nations, Department of Economic and Social Affairs, Population Division (2017). Household Size and Composition Around the World 2017.

United Nations, Department of Economic and Social Affairs, Population Division (2019). Patterns and trends in household size and composition: Evidence from a United Nations dataset. (ST/ESA/SER.A/433).

Yue, L., & Martin, R. (2015). Households or locations?: Cities, catchment areas and prosperity in india. *The Policy Working Paper Series, 7473, 39.*

Zhu, Y., Bashir, S. & Marie, M (2022). Assessing the Relationship between Poverty and Economic Growth: Does Sustainable Development Goal Can be Achieved?. *Environ Sci Pollut Res 29, 27613–27623*