ROAD INFRASTRUCTURE AND NATIONAL DEVELOPMENT

A Computer Science Data Analytics Learning Evidence

Presented to the Faculty of College of Information and Computing

University of Southeastern Philippines

Bo. Obrero, Davao City

Learning Evidence Submitted in Partial Fulfillment of the Requirements for
CS226 - DATA ANALYTICS - STAT USING R

De Villena, Francis Nathanael
Bordios, Kent Cyril

Adviser
Remollo, Khalil Stefano

June 2021

APPROVAL SHEET

The Computer Science Data Analytics Project entitled **ROAD INFRASTRUCTURE AND NATIONAL DEVELOPMENT**, prepared and submitted by **Francis Nathanael De Villena** and **Kent Cyril Bordios**, in partial fulfillment of the requirement for the **CS266 – DATA ANALYTICS – STAT USING R** has been examined and is hereby recommended for an oral examination, acceptance and approval.

Kahlil Stefano Remollo

Adviser

ACKNOWLEDGEMENT

First and foremost, we would like to give thanks to our most gracious and Almighty God who guided us and gave us strength in working on this project and in the path of life. Secondly, our parents and family have always loved and supported us throughout our lives. Lastly, we would like to thank our course instructors, Mr. Jamal Kay Rogers and Mr. Kahlil Stefano Remollo, for giving us the knowledge on the fundamental lessons of Data Analysis that is considered one of the most important cornerstones of Computer Science and for helping us finish this project.

ABSTRACT

Road infrastructure is one of the country's underlying issues. This paper provides a study of road infrastructure development in the Philippines. It discusses several factors connected to road infrastructures, such as population and a region's economy. The variables in the study turn out to be directly related to each other. Furthermore, the results suggest that more road projects will lead to a better economy but at the same time road infrastructure needs to be planned thoroughly.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Background of the Study

Road infrastructure and land transport are essential for well-functioning economies, especially in developing countries such as the Philippines, where they play a multifaceted role in the pursuit of development objectives. Moreover, road networks are the driver for a nation's productivity and progress, as they can hasten agricultural and rural growth, stimulate industry and commerce, boost the viability of urban areas, create career opportunities, widen the options of better education, and make social and health services readily reachable. Not only did road networks make land transportation more cost-effective and time-saving but also allow access to various remote rural areas throughout the country, provide mobility of the workforce, encourage the influx of tourism. and promote a wide variety of commercial and social activities [23].

Road infrastructure development positively impacted the relative increase of economic growth and has been proven to provide an impressive return on national investment, however, focusing on road infrastructure alone would be insufficient since the vital factors of economic growth are resources and products available to the people, not to mention the manpower, investments, and other forms of infrastructure. Conversely, extensive road infrastructure is a fundamental precondition for land transport systems, which facilitates quotidian conveyance of the populace and is crucial to the production and distribution of commodities and merchandise [25].

These benefits can only be maximized if the roads are constructed in strategic areas using materials with superior quality, otherwise misplanned and substandard roads would lead to a deficit in the long run and squander the public while generating sparse economic benefits and intense social and political conflict. Inevitably, roads must undergo expansion, upgrade, and proper maintenance to sustain economic growth, or else it will pose a hindrance to the process of urbanization and socio-economic development. [1]

Infrastructure investments can stimulate organizational and management change but it requires long-term financial commitments and well-built roads with impeccable design that adheres to safety standards must be constructed, which in turn, minimize occurrences of accidents as well as fatal casualties and damages of unavoidable mishaps. Public land infrastructure provides a geographic concentration of economic resources and a deeper and wider market for the growth of output and employment. It also reduces costs in most manufacturing industries, boosting resource productivity growth [13].

Insufficient public investment in transport infrastructure causes a significant decrease in productivity as it can affect economic growth by changing aggregate demand; for instance, building transport infrastructure can create and increase demand for intermediate products from other sectors and stimulate multiplier effects in the economy. Conversely, infrastructure should be viewed as the building block of each economy, which provides support to produce goods and services and is not part of the production process. Since macroeconomic growth theories explicitly do not include the concept of infrastructure systems, although road infrastructure plays a very important role in economic development, Nevertheless, road infrastructure paves the way for the betterment of the quality of life and sustains the needs of the people, in turn, the population will flourish and the national economy will prosper [19].

1.2 Statement of the Problem

Road-based transport infrastructure is the lifeblood of the Philippine economy, linking population and economic centers across the archipelago but the quality remained inadequate. Throughout the country, transportation infrastructure is falling into overuse and disrepair. Although roads are used to transport the majority of passengers and freight in the Philippines, much of the network is of poor quality. This results in high transport costs for road users. The extent of the road network in the Philippines is comparable with or better than many neighboring developing member countries [1,4,13, 16].

However, in terms of global performance and ranking of overall infrastructure quality, the Philippines ranks 112th among 138 countries in quality of infrastructure, lagging behind the five pioneer members of the Association of Southeast Asian Nations (ASEAN) [1,4,16]. Despite the improvement and expansion of the transport systems, deteriorating and substandard existent infrastructure will constrict future growth while increasing social and economic burdens. This poses major problems to the country's projected growth trajectory and the vision of increased prosperity for its citizens. Since the significant achievements and critical reforms in infrastructure were deficient to keep up with the rates of population growth and urbanization. For instance, traffic congestion in various metropolitan areas of the country alone hinders economic productivity and degrades ambient air quality [1,4,16,13]. The lack of a reliable and convenient public land transport system, coupled with the rapid growth in the number of registered private-owned vehicles further strained the already limited road space in urban areas. In provinces, primary roads connecting production to commercial centers are typically constrained in terms of safety as they remain vulnerable to the adverse impact of natural disasters such as landslides and earthquakes [10,13,19].

1.3 Significance of the Study

The main goal of this paper is to highlight the effects of road infrastructure development on the economic growth and competitiveness of the Philippine economy. The findings of this study will redound to the benefit of the nation and the populace considering that road infrastructure is imperative for social and economic nationwide development. The study could also add to the wealth of knowledge of the developed methodology in conducting urban-wide change impact analysis and feasibility study of proposed extensive road structural projects by private organizations and government agencies that oversee the planning, funding, design, construction, and maintenance of related infrastructures. This also provides concerned policymakers and economists an insight into the impact of road infrastructure and facilities on the national macroeconomy.

1.4 Scope and Limitations

The study covers the domestic income, census, as well as the extent of road infrastructure of every administrative region of the Philippines, except the Bangsamoro Autonomous Region in Muslim Mindanao (BARMM) due to the lack of valid data gathered from the sources. Mass rapid transit and commuter rail are also ruled out from the context of land transportation since such only operate in Metropolitan Manila and some parts of Luzon whilst railway infrastructure is nonexistent in the rest of the country yet. Even so, the study focus on the density of road networks that facilitate motorized vehicles in every region of the country, encompassing the primary, secondary, and tertiary roads, with the exclusion, however, of local roads, under the jurisdiction of the Department of the Interior and Local Government (DILG).

1.5 Definition of Terms

Causal Research – aims to investigate causal relationships and therefore always involves one or more independent variables and their relationships with one or multiple dependent variables. Causal relationships can be tested using statistical and econometric methods.

Correlation Analysis – is a method of statistical evaluation used to study the strength of a relationship between two, numerically measured, continuous variables. This particular type of analysis is useful to establish if there are possible connections between variables.

Department of Public Works and Highways – abbreviated as DPWH, is the executive department of the Philippine government solely vested with the mandate to be the state’s an engineering and construction arm and responsible for the planning, design, construction, and maintenance of infrastructure facilities and other public works per national development objectives.

Econometrics – uses economic theory, mathematics, and statistical inference to quantify economic phenomena. In other words, it turns theoretical economic models into useful tools for economic policymaking. The objective of econometrics is to convert qualitative statements into quantitative statements.

Generalized Additive Model – abbreviated as GAM, is a generalized linear model in which the linear response variable depends linearly on unknown smooth functions of some predictor variables, and interest focuses on inference about these smooth functions.

Gross Domestic Product – abbreviated as GDP, is the total market value of the goods and services produced by a country's economy during a specified period. It includes all final goods and services—that is, those that are produced by the economic agents located in that country regardless of their ownership and that are not resold in any form. It is used throughout the world as the main measure of output and economic activity.

Gross Domestic Product per capita – is a metric that breaks down a country's economic output per person and is calculated by dividing the GDP of a country by its population. Small, rich countries and more developed industrial countries tend to have the highest per capita GDP.

Keynesian Economics – is the body of ideas set forth by John Maynard Keynes in his General Theory of Employment, Interest and Money and other works, intended to provide a theoretical basis for government full-employment policies. It was the dominant school of macroeconomics and represented the prevailing approach to economic policy among most Western governments until the 1970s.

Philippine Statistics Authority – abbreviated as PSA, is the central statistical authority of the Philippine government that collects, compiles, analyzes, and publishes statistical information on economic, social, demographic, political affairs, and general affairs of the people of the Philippines and enforces the civil registration functions in the country.

Population – is a distinct group of individuals, whether that group comprises a nation or a group of people with a common characteristic.

Regression Analysis – is a set of statistical methods used for the estimation of relationships between a dependent variable and one or more independent variables. It can be utilized to assess the strength of the relationship between variables and for modeling the future relationship between them.

Restricted Maximum Likelihood – abbreviated as REML, is a particular form of maximum likelihood estimation that does not base estimates on a maximum likelihood fit of all the information, but instead uses a likelihood function calculated from a transformed set of data.

Road Density – is the ratio of the length of the country's total road network to the country's land area. The road network includes all roads in the country: motorways, highways, main or national roads, secondary or regional roads, and other urban and rural roads.

Road Infrastructure – is the land used for the transport of goods and passengers via road and may include national, provincial, and municipal roads, facilities to park and stop, road reserves, maintenance, and maneuvering facilities.

Spline Regression – is a non-linear regression that is used to try and overcome the difficulties of linear and polynomial regression algorithms. It is an alternative, and often superior, approach to modeling nonlinear relationships.

.

CHAPTER 2

REVIEW OF RELATED LITERATURE

2.1 Related Literature

The link of land transportation and road infrastructure to economic growth has intrigued market researchers and economists for decades, concerning the theoretical and statistical analysis of the effects in growth theories and their impact on global competitiveness, economic growth, income inequality, industrial productivity, environmental impact, and the overall well-being of the general public. Consequently, their significant effects on economic growth were recognized and proven in many of the studies mentioned thereafter.

According to a study conducted in Slovakia, there was a strong correlation between the expenditures on road infrastructure and GDP. In the case of Croatia, an EIZ study shows that there is a causal link between transport infrastructure, transport services, and the level of international exchange [8]. Ismail and Mahyideen empirically explore the effects of land transport infrastructure on international exchange and economic growth in Asia, and the results have shown positive effects on the rise in international exchange as well as on economic growth [22]. Furthermore, Purwanto et al. analyze the relationship between transport infrastructure investment and its wider economic impacts, namely competitiveness and economic growth, and recommend methodology improvements [12]. Aschauer developed a simple model in which local governments can exert an important influence on the rate of economic growth within their locality. A higher level of highway capacity and quality expands transportation services and, in so doing, raises the marginal product of private capital. The higher marginal product of capital induces higher investment in physical capital and growing per capita incomes and output [6].

Mohmand et al. use the unit root, cointegration, and Granger causality model to estimate the causal linkages between economic growth and transportation infrastructure existing at the national and provincial levels. Their results suggest that there is no causality between the two variables in the short run, at the national level; however, a unidirectional causality from economic development to infrastructure investment exists in the long run [30]. More recently, the relationship and the direction of causality between transport infrastructure, infrastructure investment, and economic growth, using a vector error correction model in the case of Armenia, Turkey, and Georgia. Their results show that gross capital formation and goods transported through roads have a positive and statistically significant impact on economic growth in the short term, and show the existence of bidirectional causality between economic growth and infrastructure investment, and between road passengers carried and infrastructure investment in both the short and longer terms [9]. On the other hand, by using three standard estimators, namely pooled ordinary least squares, fixed effects, and random effects, Lenz et. al. concluded that the results are significant and show that road infrastructure, gross fixed capital formation, population growth, and trade openness have positive effects on economic growth [20].

Carlsson et al. have explored the role of infrastructure in macroeconomic growth theories and confirmed that certain economic functions of infrastructure may be represented in existing macroeconomic models, so new economic geography enables the presentation of transport infrastructure due to a more spatial approach [27]. Ng et.al. analyzed the contribution of road infrastructure development and other socio-economic factors that contribute to economic growth. It was observed that economic growth increases at low urbanization levels but gradually decreases once urbanization exceeds a threshold level. In addition to that, the growth in road length per thousand populations would facilitate export growth [21]. Khanani et. al. investigated how the implementation of road infrastructure projects affects the socio-spatial landscapes and economic development, and social groups within the peri-urban areas of Kisumu, Kenya, and Accra, Ghana. Improved accessibility to basic facilities and services after implementation of the road projects was observed in both communities. Some residents from Kisumu experienced improved access to facilities and services, while in Accra, the biggest factor was an increase in access to potable water [26]. Queiroz and Gautam concluded in their analyses that there is a statistically significant relationship between road infrastructure and economic development on a worldwide basis. Moreover, data between 1950 and 1988 showed significant relationships between per capita gross national product (PGNP) and density (per capita length). [5] Using time-series cross-section data from the manufacturing sector of the German federal states, Bundesländer, from 1970 to 1993, Stephan examined the impact of road infrastructure on private production applying a Cobb-Douglas production function, a translog production function, and a growth accounting approach. Their econometric analysis found that road infrastructure is significant for production in the manufacturing sector [2]. Ke et. al. used a consistent and robust dynamic panel data system generalized method of moments estimation for identification and found strong evidence confirming that transport infrastructure contributes to regional economic growth in China during the period 2007–2015, as the country approaches its upper-middle-income status [29]. Laborda and Sotelsek applied dynamic panel GMM system estimation and found positive effects of road density and road paved on total factor productivity (TFP) in countries with middle-low and low income, using parametric and non-parametric estimations. They also found empirical evidence about a Kuznets curve when they consider the percentage of roads paved and in countries with middle-low and low income between road density and unemployment [15].

However, some studies argue that transport infrastructure alone is not sufficient for reaching higher gross domestic product (GDP) and that infrastructure endowment is a relatively poor predictor of economic growth. It seems that the vast body of evidence is far from being conclusive and that the role of transport infrastructure depends on different circumstances. Therefore, it is important to be aware of other drivers of economic growth because they have important implications on the transport infrastructure's impact on economic growth. [2]

2.2 Research Framework

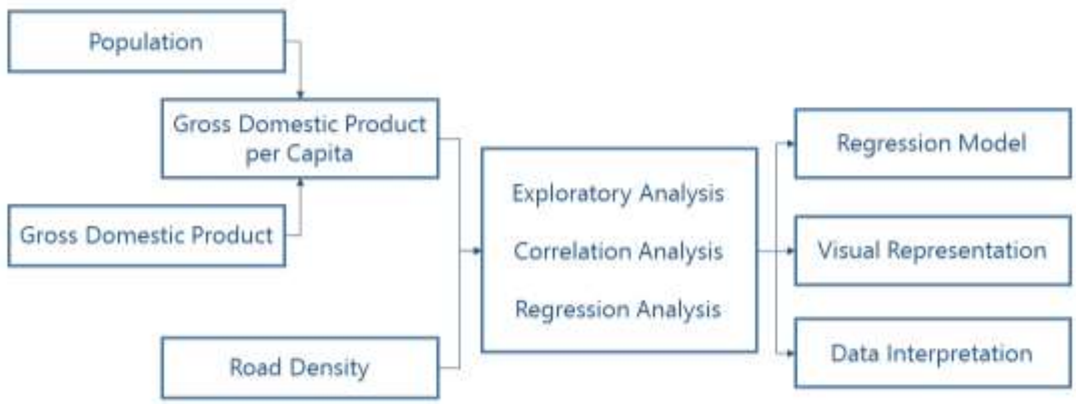

Figure 1 – Research Conceptual Framework

This study recognizes the intellectuals who contributed significantly to the body of knowledge related to road infrastructure and the macroeconomics that lay the foundation for the importance of the problem statement and the research questions. They have deduced the data and the processes involved to achieve the objectives of the study. The variables are gross domestic product, population, and road density. Mainstream theories in economic development include neoclassical growth theory and endogenous growth theory. These theories for economic development imply factors such as the impact of road transport infrastructure on the industrial structure, the connection of transport to the distribution of productive forces, the strength of transport infrastructure on economic ties, and the importance of transportation on the urbanization process.

One theory stands out as providing a framework for the research included in this study. Keynesian economics is a macroeconomic theory of total spending in the economy and its effects on output, employment, and inflation. It was named after John Maynard Keynes, who is regarded as the founder of modern macroeconomics. One of the distinctions of Keynesian economics from other theories is the belief in activist policies to reduce the magnitude of the business cycle, which is ranked among the top priorities in Keynesian economics. Keynes advocated so-called countercyclical fiscal policies that act against the direction of the business cycle, rather than seeing unbalanced government budgets as unjust. One example is that Keynesian economists would advocate deficit-spreading on labor-intensive infrastructure projects to stimulate employment and stabilize wages in case of economic struggles [11].

Keynesian economics suggests that spending boosts aggregate output and generates more income. Keynes advocated for more government spending and fewer taxes to stimulate demand and save the economy from threats of a possible crisis. This theory also implies correcting economic imbalances. Keynes argued that there was no guarantee that the goods produced by individuals would be met with sufficient effective demand, and there is a likelihood of high unemployment periods. He envisioned the economy being unable to maintain itself at full employment automatically and believed that the government needs to step in and put the purchasing power into the hands of the working population through government expenditure. Thus, it is implied that certain individually rational microeconomic-level actions such as not investing savings in the goods and services produced by the economy can likely result in the economy operating below its potential output and growth rate [11].

The idea of infrastructure spending as an economic stimulus is rooted in Keynesian economics. In Keynesian theory, when a recession happens the economy can get stuck with sustained high unemployment and a stagnant GDP for an extended period due to a deficiency of aggregate demand. Because infrastructure spending is usually for a specific budgeted amount to fund particular projects, it tends to meet the criterion of being temporary, though cost overruns and other issues can drag this out. One condition is that infrastructure and regional economic development patterns are strongly correlated. If infrastructure is built solely to provide economic stimulus, not because it provides changes to regional economic development, it could cause significant negative long-term effects. This is doubly important to remember as infrastructure might be rushed for the sake of meeting requirements on time in a way that does not consider longer-term implications. This further limits infrastructure stimulus to projects that are already significantly developed [11].

CHAPTER 3
METHODOLOGY

3.1 Research Design

The proponents delve deeper and applied a non-experimental, causal research design in this quantitative study, primarily aiming to investigate causal relationships—cause and effect—between the extent of road infrastructure and the economic activity in every region of the archipelago with statistical and econometric methods. To this end, the proponents study how commerce and industry, the dependent or response variable, is affected by the expansion of road infrastructure, the predictor or independent variable. Consecutively, correlation and regression among the two variables must be discovered to support the causal inference. Also, temporal sequence, concomitant variation, and nonspurious correlations should occur which are true on road infrastructure and economic activity as supported by the studies mentioned in the previous chapter: Investments on road infrastructure yields economic growth. Since basic econometric and statistical techniques are only utilized in this study, the influence of extraneous variables from the analysis of collected secondary data is nonexistent.

Causal research design is chosen by the proponents as it is appropriate to this study's research objective and it plays an instrumental role in terms of identifying reasons behind a wide range of processes, as well as, assessing the impacts of changes on existing norms, processes, and so on. Moreover, causal studies usually offer the advantages of replication if the necessity arises.

3.2 Data Collection

The data used in this study were gathered directly from the datasets of various government agencies that are deemed overt and readily available to the general public. The data on the regional population and the gross regional domestic product came from the open data platform of PSA, OpenSTAT. In addition, the data on the road density of every region is derived from the Road & Bridge Information Application (RBIA) database of DPWH.

Then, the collected datasets had undergone data cleaning, removing the irrelevant data particularly those outside the scope of this study. The data are classified into three groups based on the years recorded from the quinquennial census of population and housing conducted by PSA, namely the years 2010, 2015, and 2020. The total population of every region is comprised of the population of their respective provinces and chartered cities. Meanwhile, road density pertains to the ratio of the length of the total road network to the land area of every region. The road network includes all roads in the country: motorways, highways, main or national roads, secondary or regional roads, and other urban and rural roads.

In the case of the short-lived Negros Island Region (NIR), the census and national road data from Negros Oriental, Negros Occidental, and Bacolod City in 2015 are reverted to their respective regions ex-ante. As aforementioned, this study excludes the Bangsamoro Autonomous Region in Muslim Mindanao since all regional offices of the executive departments before its inauguration were reconfigured into ministries of Bangsamoro, and unfortunately, the Ministry of Public Works data of Bangsamoro's road infrastructure is not publicly accessible on the web and the infrastructure data of the said region is unavailable in the DPWH annual infrastructure atlas and RBIA database.

3.3 Research Method

Non-experimental causal research involves quantitative methods which are utilized for examination of numerical data and often require the use of statistical tools to analyze collected data. This allows for the measurement of variables and relationships between them can then be established. The overall structure for a quantitative method is based on the scientific method. It uses deductive reasoning, where a hypothesis is formed, collects data in an investigation of the problem, analyzes the data, and then conclusions are presented, to prove the hypotheses are not false or false.

This study leverages the secondary data retrieved from government agencies and omits the first-hand data collection methods. Therefore, we can directly proceed to data cleaning and processing, organizing the secondary data to prepare a dataset suitable for statistical computing. In the dataset, there are 6 columns: region, year, population, GDP, GDP per capita, and road density. The column region contains 16 administrative divisions that primarily serve to coordinate planning and organize national government services across multiple local government units, but excluding the autonomous region, BARMM. The year column contains the period when the data from the said regions were collected. The rest of the columns contain their respective data collected from PSA and DPWH.

After ensuring the dataset was free from erroneous data, the proponents utilized the R programming language in version 4.1.0 and performed exploratory data analysis to summarize its characteristics. Then, regression analysis was executed to ascertain the causal inference of the variables, namely the road density and GDP per capita. An appropriate regression model was formulated based on the characteristics of data to infer causal relationships correctly. However, the coefficients and p-values produced by the regression model were insufficient to conclude whether the model fits well with the data or not. To assess the assumptions of the regression model, the residuals are checked through diagnostic plots. Residuals are leftover of the outcome variable after fitting the model to data and they could reveal underlying patterns in the data by the fitted model.

The standard diagnostic plots validate the multivariate normality, multicollinearity, autocorrelation, and homoscedasticity of the residuals. If these conditions are met, then it is certain that the regression model is reliable and the assumption on causal inference can be accepted.

An R Shiny application was also built to freely interact and view the graphical representation of the dataset as well as the results from a certain fitted regression model.

3.4 Ethical Considerations

This study does not involve any individual and personal information during data collection since the data are not gathered first-hand and only analyze secondary data that are available online for public access from government agencies. Secondary data has been opted as it is easily accessible compared to primary data and requires little to no cost and time to acquire. Besides, data from government agencies guarantee reliability and accuracy while adhering to the policies and laws that protect the rights of data subjects. Hence, it is known for certain that all ethical considerations were being upheld

CHAPTER 4

RESULTS AND DISCUSSIONS

4.1 Exploratory Data Analysis

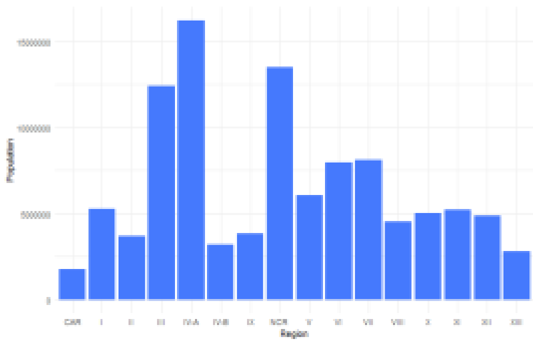

Figure 2 – Population Bar Graph

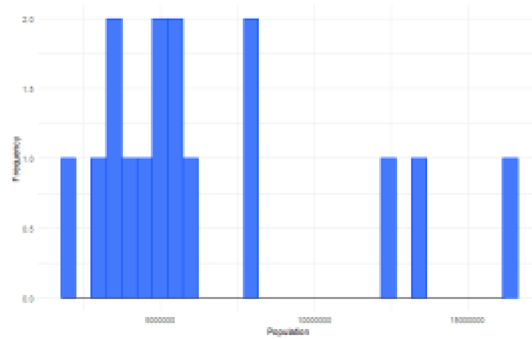

Figure 4 – Population Histogram

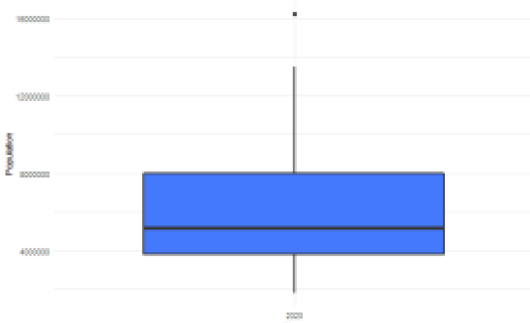

Figure 3 – Population Box Plot

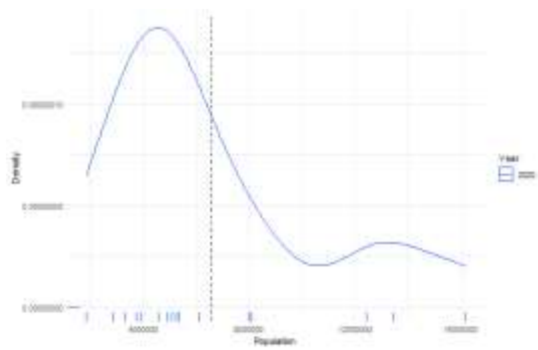

Figure 5 – Population Density Plot

CALABARZON (Region IV-A) is the most populous region in the Philippines having over 16.19 million inhabitants in 2020 and is also the country's second-most densely populated after the National Capital Region (NCR) that has 13.48 million inhabitants, and followed by Central Luzon with 12.42 million inhabitants. The Cordillera Administrative Region (CAR) has the smallest population size of 1.80 million. As seen in Figure 3, the population of CALABARZON is also an outlier amongst all regions. In addition, the distribution of populations in every region appears to be skewed right with the regions, CALABARZON, NCR, and Central Luzon, on its tail, as shown in Figures 4 and 5. Undoubtedly, the said regions comprise most of the highly urbanized cities (HUCS) and densely populated component cities in the country.

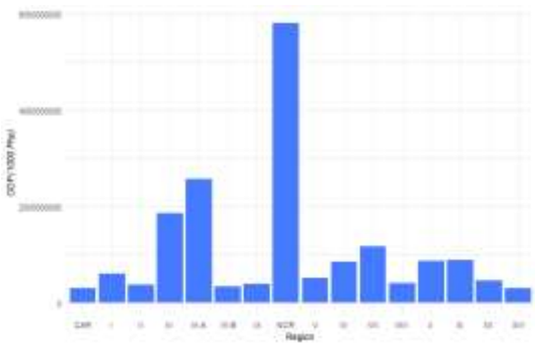

Figure 6 – GDP Bar Graph

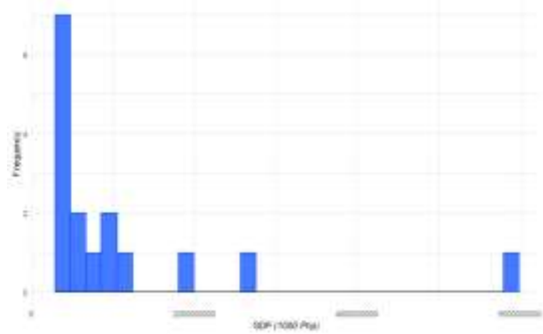

Figure 8 – GDP Histogram

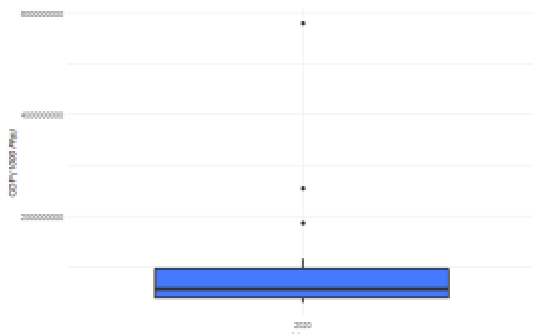

Figure 7 – GDP Box Plot

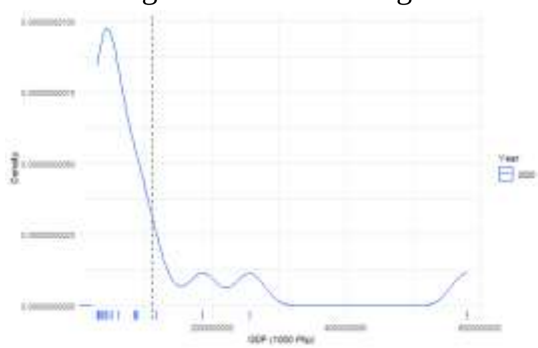

Figure 9 – GDP Density Plot

Figure 6 shows that NCR has the highest nominal gross regional domestic product with approximately ₱5.8 trillion, which accounts for more than 40% of the gross domestic product of the Philippines in 2020. CALABARZON comes far second at ₱2.6 trillion and is followed by Central Luzon Moreover, in Figure 7, the aforementioned regions are the outlier amongst the regions. The distribution of regional GDP is skewed right with NCR, CALABARZON, and Central Luzon on its tail, as shown in Figures 8 and 9. NCR is designated as a global power city, its economic power makes the region the country's premier center for finance and commerce, both locally and internationally. Its economic dynamism is rooted in strong consumer demand supported by a vibrant labor market and robust remittances. Due to the proximity of CALABARZON and Central Luzon to Metro Manila, a large amount of urbanization and industrialization has taken place in the provinces within Greater Manila Area over the years. However, the rest of the region mainly relies on rural incomes, such as agriculture and aquaculture.

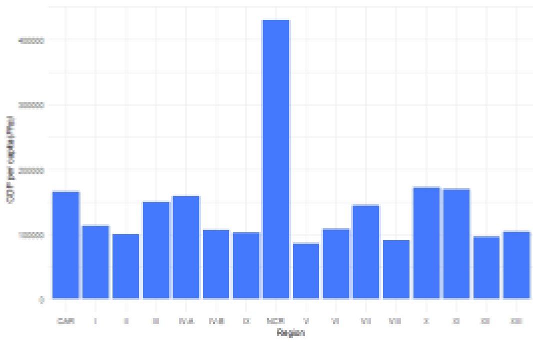

Figure 10 – GDP per capita Bar Graph

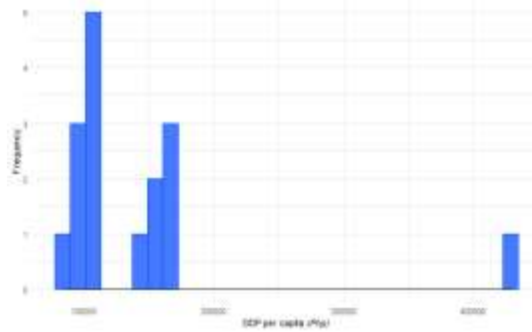

Figure 12 – GDP per capita Histogram

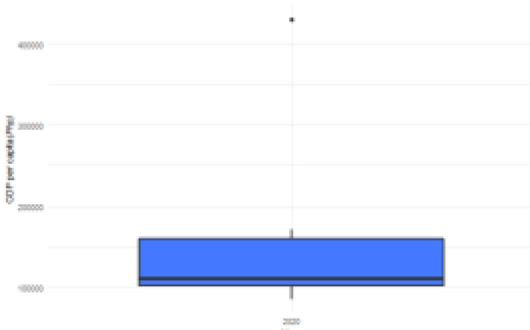

Figure 11 – GDP per capita Box Plot

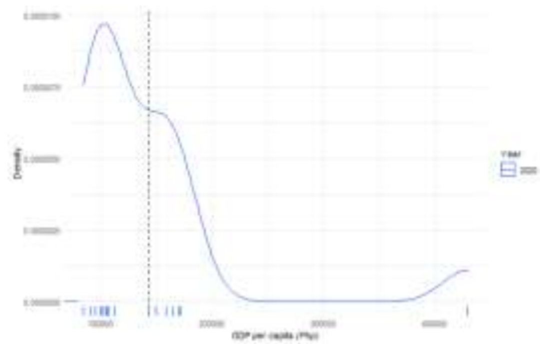

Figure 13 – GDP per capita Density Plot

The nominal GDP per capita of every region was derived from the calculation of the region's marketed goods and services produced within a period when averaged across the total population. In other words, it is the equal apportioning of the gross domestic product for each resident to represent the region's standard of living. Besides, Figures 10 and 11 show that Metropolitan Manila has the highest per capita GDP of the country at ₱429,906.54 in 2020, surpassing other regions by a large margin. In Figures 12 and 13, it is evident that the distribution of GDP per capita in all regions is skewed right, with NCR on the edge. This generally implies that the said region has a predominant upper part middle-income class and expanding upper-income class residents. As it has wide access to commercial and manufacturing establishments, most of the middle class also work in these regions. By contrast, the rest of the country has a larger share of the middle class still belongs to the lower part of the middle-income bracket, between two and four times the poverty line.

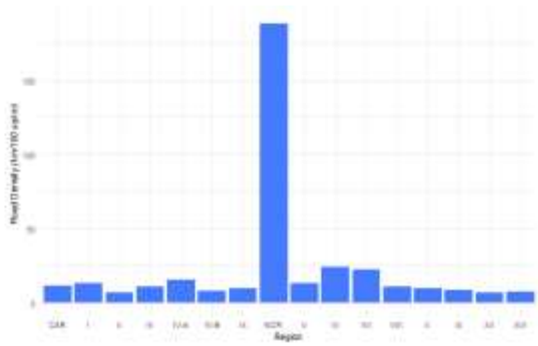

Figure 14 – Road Density Bar Graph

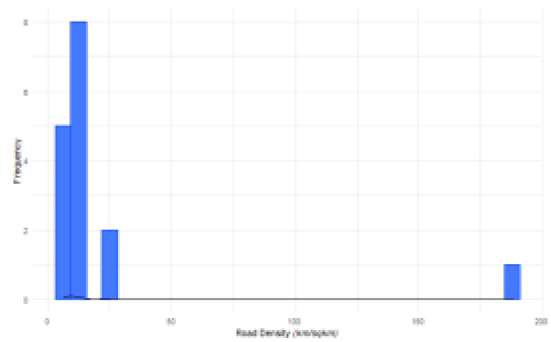

Figure 16 – Road Density Histogram

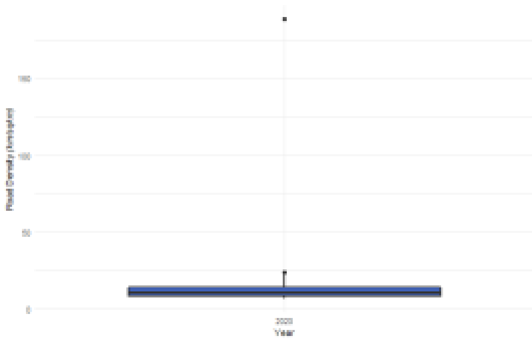

Figure 15 – Road Density Box Plot

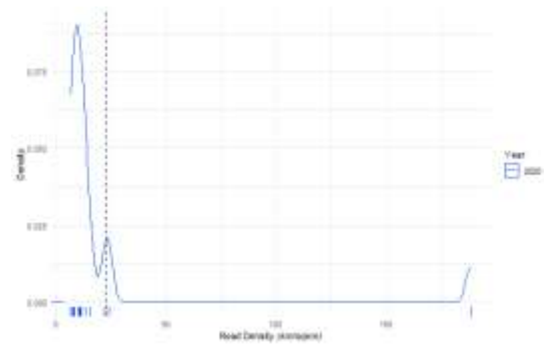

Figure 17 – Road Density Density Plot

In terms of road density, Metropolitan Manila has 188.24 kilometers of roads per square kilometer of land area in 2020. Similar to the case in GDP per capita, the extent of its road infrastructure outclassed other regions of the archipelago, as seen in Figures 14 and 15. Furthermore, in Figures 16 and 17, the distribution of regional road density is skewed right while having only the NCR on the edge. The capital region has a high concentration of road networks, despite it has the shortest road length and the smallest land area amongst the administrative regions, with a total of 1,166.24 km and 619.57 km^2, respectively. On the other hand, the second largest region, Cagayan Valley (Region II), has the lowest road density at 6.73 kilometers per square kilometer. Most of its land area is situated on the valley between the Cordilleras and the Sierra Madre mountain ranges, a stark contrast to the relatively flat alluvial lands of Metropolitan Manila.

4.2 Correlation Analysis

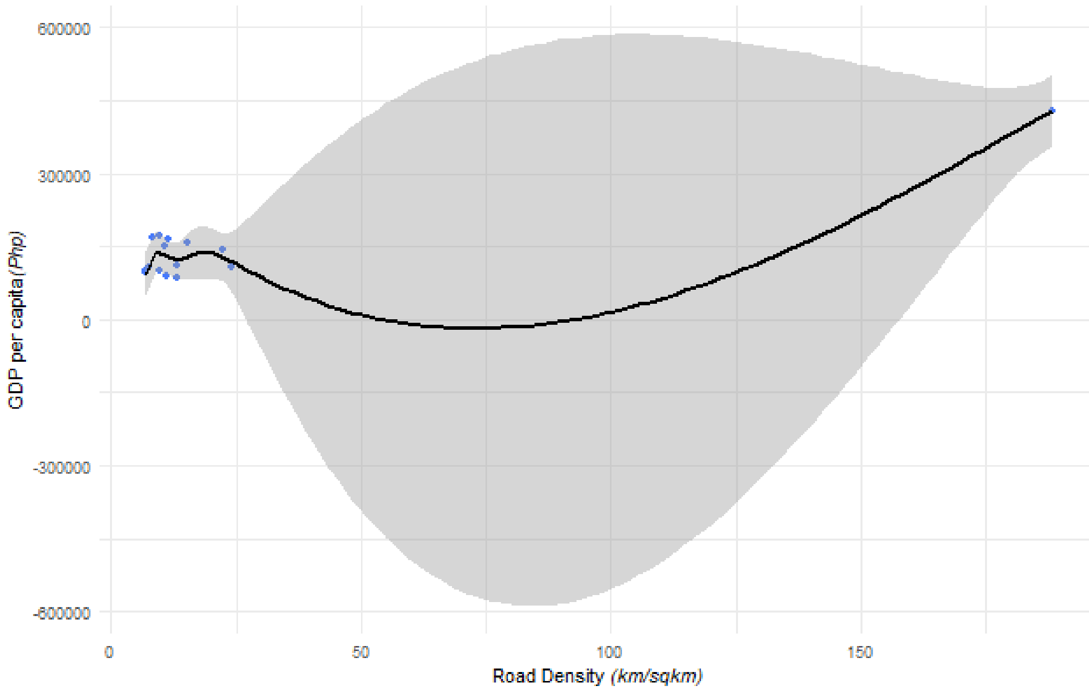

Figure 18 – Road Density and GDP per capita Scatter Plot

Before proceeding to formulate a model to infer causal relationships, it is necessary to check for the directionality and correlation of data. As shown in Figure 18, the dependent variable, GDP per capita, and the independent variable, road density, are plotted in a scatter plot, suggesting a non-linear relationship between the two variables since there is no distinct linear pattern and the distribution of their values is far from normal. Then, the Pearson correlation test is used to evaluate the association between the two variables. As a result, the correlation coefficient for road density and GDP per capita is 0.93, which indicates that the variables have a positive and high degree of correlation. Also, the p-value is 2.375×10^{-7}, which is way below the significant level alpha of 0.05. thus rejecting the null hypothesis and indicates no significant difference. Hence, the data is fit to run a regression analysis.

4.2 Regression Analysis

Regression modeling is simply generating a mathematical model from measured data. This model is said to explain an output value given a new set of input values. In regression modeling when we include a continuous predictor variable in our model, either as the main exposure of interest or as a confounder, we are assuming that the relationship between the predictor variable and the outcome is linear.

Typically, linear regression is used for regression modeling that assumes that the output can be explained using a linear combination of the input values. An increase in the predictor variable is associated with a fixed difference in the outcome. Thus, we make no distinction between an increase in the predictor variable near the minimum value and an increase in the predictor variable near the maximum value. This assumption of linearity may not always be true and may lead to an incorrect conclusion about the relationship between the exposure and outcome, or in the case of a confounder that violates the linearity assumption, may lead to residual confounding.

In this case, the relationship between the outcome and a predictor variable is non-linear but has a significant positive correlation, as explained in the previous section. Therefore, a generalized additive model (GAM) has been chosen. Originally invented by Trevor Hastie and Robert Tibshirani in 1986, it is an extension of the generalized linear model (GLM) in which the linear predictor is given by a user-specified sum of smooth functions of the covariates plus a conventional parametric component of the linear predictor [14].

Generally, GAM has the interpretability advantages of GLMs where the contribution of each independent variable to the prediction is encoded. Additionally, it has substantially more flexibility because the relationships between independent and dependent variables are not assumed to be linear. It discreetly selects what type of predictive functions are needed. From an approximation viewpoint, the use of regularized, nonparametric functions avoids the drawbacks of dealing with higher-order polynomial terms in linear models [14].

In mathematical terms, GAM is an additive modeling technique where the impact of the predictive variables is captured through smooth functions which—depending on the underlying patterns in the data—can be nonlinear [14]:

$$g[E(Y)] = \alpha + s_1(x_1) + \cdots + s_p(x_p)$$

where Y is the dependent variable, E denotes the expected value, and g denotes the link function that links the expected value to the predictor variables $x_1 \ldots x_p$. The terms $s_1(x_1), \ldots, s_p(x_p)$ denotes smooth, nonparametric functions. This can allow for a more flexible estimation of the underlying predictive patterns without knowing upfront what these patterns look like.

The generalized additive model framework is based on the following mental model [14]:

- Relationships between the individual predictors and the dependent variable follow smooth patterns that can be linear or nonlinear.
- It is possible to estimate these smooth relationships simultaneously and then predict linear by simply summing them up.

When a regression model is additive, the interpretation of the marginal impact of the partial derivative does not depend on the values of the other variables in the model. Thus, simple, comprehensible statements about the effects of the predictive variables can be made by simply looking at the output of the model. Furthermore, GAM can capture common nonlinear patterns that a classic linear model would miss, as the predictor functions are automatically derived during model estimation. The GAM framework allows us to control the smoothness of the predictor functions to prevent overfitting. By controlling the wiggliness of the predictor functions, the bias and variance tradeoff can be directly tackled. Also, the type of penalties applied in GAMs have connections to Bayesian regression and l_2 regularization.

To obtain a fitting model that has reliable and stable results, a generalized additive model with univariate penalized cubic regression spline smooths and the restricted maximum likelihood was used to represent the causal relationship of road density and gross domestic product per capita.

A more detailed look at model outputs is necessary to interpret the results of model-fitting. The following figure illustrates the summary of model statistics to better understand the relationships between variables.

```
Family: gaussian
Link function: identity

Formula:
capita ~ s(rdensity, bs = "cr")

Parametric coefficients:
            Estimate Std. Error t value        Pr(>|t|)
(Intercept)   143169       7975   17.95 0.0000000000456 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
               edf Ref.df     F   p-value
s(rdensity) 0.9885      7 12.13 0.0000012 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =   0.85   Deviance explained =   86%
-REML = 188.45  Scale est. = 1.0176e+09  n = 16
```

Figure 19 - Generalized Additive Model Summary

The first part of the summary describes the fitted model. *Family* tells that the model assumes a Gaussian distribution of errors, and the *Link function* indicates that the model doesn't transform the predictions. *Parametric coefficients* describe the parametric terms of the model. In this context, it refers to the linear terms in the model. It shows the coefficients for the linear terms in the model, their values, errors, test statistics, and p-values. The asterisks next to the p-value indicate statistical significance, thus, the model intercept is significant. The *approximate significance of smooth terms* covers smooth terms. For smooths, coefficients are not printed as each smooth has several coefficients—one for each basis function. Instead, *edf*, which stands for effective degrees of freedom, represents the complexity of the smooth. The smooth of road density, with an estimated degree of freedom at 0.9885, denotes an equivalence to a straight line. The terms *Ref.df* and *F* columns are test statistics used in an ANOVA test to check the overall significance of the smooth, where the *p-value* is the result. Since the p-value is 0.0000012, which is less than the alpha level, road density is linear and significant.

Adjusted R^2 denotes the proportion of variance explained, where original variance and residual variance are both estimated using unbiased estimators while *deviance explained* tells the proportion of the null deviance explained by the model. The fitted model has 0.85 Adjusted R^2 and 86% *deviance explained*, indicating that the model fits quite well with the data. *REML* of maximum likelihood (ML) may be used for smoothness selection, by viewing the smooth components as random effects. It is mainly used to compare the smoothness of different models. Lastly, in the case of Gaussian distribution of errors, *Scale estimate* reports the squared value of the residual standard error, which estimates the variance of the errors. Overall, the summary shows that the model fits well with the variables.

One of the most important things to do when interpreting and checking models is a visualization to communicate the results much better. The following figure illustrates the plotted Generalized Additive Model:

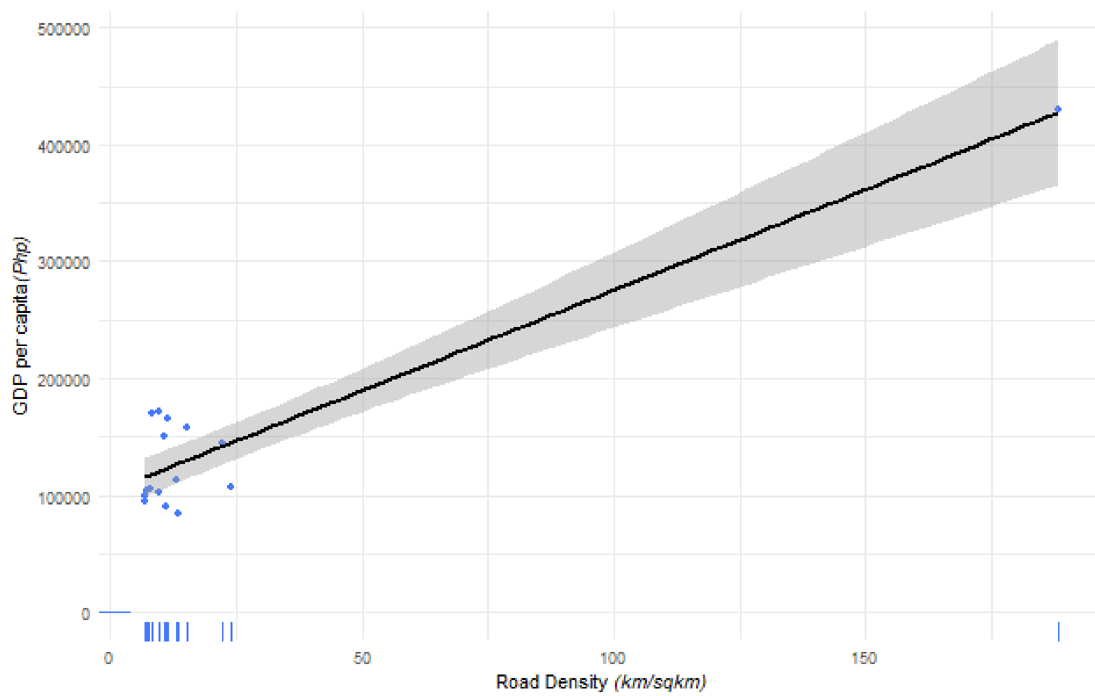

Figure 20 – Generalized Additive Model with Penalized Cubic Regression Splines

```
Method: REML   Optimizer: outer newton
full convergence after 9 iterations.
Gradient range [-0.0001221384,0.0000002090818]
(score 180.4524 & scale 1017610084).
Hessian positive definite, eigenvalue range [0.00002875482,7.53478].
Model rank =  10 / 10

Basis dimension (k) checking results. Low p-value (k-index<1) may
indicate that k is too low, especially if edf is close to k'.

               k'   edf k-index p-value
s(rdensity) 9.000 0.988    1.23    0.76
```

Figure 21 – Generalized Additive Model Check

Figure 20 reports the model convergence and basis checking results. As seen here, it implies full convergence where the best solution was found. If the model has not converged, results are likely not correct. Meanwhile, basis dimension checking results shows a statistical test for patterns in model residuals, which should be random. The test shows that the smooth of road density has 9 basis functions, 0.988 effective degrees of freedom, 1.23 k-index test statistic, and a 0.76 p-value. Therefore, the smooth has no significant patterns in its residuals and has enough basis functions.

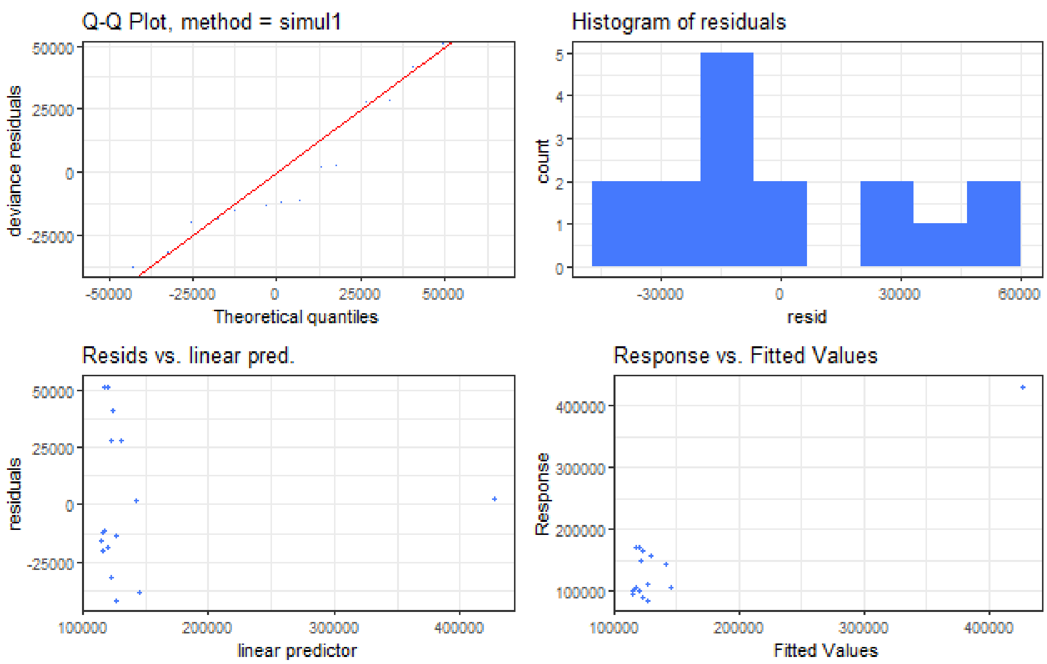

Figure 22 – Generalized Additive Model Diagnostic Plots

Furthermore, the model check also produces four residual diagnostic plots, as seen in Figure 21. These plots reveal how poorly the model fits the variables. On the top-left is a Quantile-Quantile plot, which compares the model residuals to a normal distribution. The model's residuals are close to forming a straight line. On top-right is a histogram of residuals, which shows that the residuals are skewed right. On the bottom left is a plot of residual values, which indicates that they are almost evenly distributed around zero. Finally, on the bottom-right is the plot of response against fitted values, which shows the values cluster around the 1-to-1 line. In Figure 21, the concurvity output metric reports three different scenarios of measuring concurvity: worst, observed, and estimated. Since the values of concurvity are far below 0.5, there is virtually no concurvity and the model has converged.

```
                                             para
worst    0.00000000000000000000000000000004930381
observed 0.00000000000000000000000000000004930381
estimate 0.00000000000000000000000000000004930381
                                       s(rdensity)
worst    0.000000000000000000000000000000041627667
observed 0.000000000000000000000000000000007307095
estimate 0.000000000000000000000000000000004371422
```

Figure 23 – Generalized Additive Model Concurvity

Based on the visual evaluations and prior checking, the generalized additive model is decently fitted to the variables. To infer a causal relationship, the fitted generalized additive model estimates the equation of the line that best describes the association between the independent variable and the dependent variable. The derived equation is formulated as follows:

$$g[E(y_i)] = \beta_0 + f(x_i)$$

where $E(y_i)$ denotes the number of observations of the response variable that represents the per capita gross regional domestic product, g is the Gaussian distributed exponential family with the log link function, β_0 is the mean number of observed prediction value, f denotes the penalized cubic regression spline smooth function, x_i is the predictor variable that represents the extent of road density in a region.

The procedures described above assumes that the association between the independent and dependent variables is linear. The visualized model in Figure 19 suggests that there is a linear relationship, as supported from the results shown from Figure 20 that the road density smooth, with an estimated degree of freedom of 0.998, is linear and significant. Therefore, we reject the null hypothesis and conclude that there is a significant linear relationship between regional economic growth and the magnitude of the locale's public road infrastructure.

CHAPTER 5

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

5.1 Summary

Using a Gaussian generalized additive model approach with penalized cubic regression splines provided a flexible and useful tool for revealing insight into complex relationships and structure in these data. The findings of this study contribute to existing literature and can be used to aid a further understanding of public road infrastructure and its effects on the national macroeconomy. In the previous chapters, different analyses were performed, and it was observed that values in different regions may vary and that the relationship between GDP per capita and road density of every region shows a strong positive correlation and linear relationship. Thus, the expansion of public road infrastructure yields regional economic growth.

5.2 Conclusions

With a growing economy, the Philippines requires more and better-selected infrastructure investments, given its archipelagic landscape, expanding population, and rapid urbanization. To support a higher growth trajectory and improve the quality of life in both urban and rural communities, infrastructure development should remain among the top priorities of the government over the long term. Spending on infrastructure has to be intensified while addressing persistent issues and challenges hampering implementation. Accessibility increases by building new infrastructure and through growth in already connected regions. An important feature of infrastructure is that once a road is in place, continuous economic improvements in one place along its path can create positive spill-overs to other connected places. A well-developed road transport sector, most especially in developing countries, can therefore create lasting stimulus through a variety of activities of the development endeavors of a nation.

Among these, the expansion of road transport facilities plays a significant role in both the production and consumption decisions of every household in their day-to-day activities. Besides, road transport facilities are essential for expanding education, health service provision, trade furtherance – both within the country and the export market and better public as well as private service provisions, including banking and insurance services, to the poor and marginalized rural dweller. Likewise, roads serve as key infrastructural units, which provide linkages to other modes of transportation like railways, shipping, and air network. Hence, we must consider all the factors necessary in coming up with the optimum plan to improve the road infrastructure, one of which is road density.

Economic development and transportation infrastructure are closely related. One reason for this positive relationship is that transportation infrastructure influences regional productivity through the facilitation of efficient movement of goods and labor used in production. Hence, improved transportation infrastructure can enhance the efficiency of goods and labor movement for a production. The reduction in time and effort required to produce goods translates directly into increased regional productivity.

5.3 Recommendations

The increasing complexity in the decision-making process in road infrastructure investments calls for greater dissemination of information and expanded educational efforts. At present, the concepts and the studies concerning the effects of road infrastructure on economic development are not that accessible to this group. Therefore, more effort should be made in subsequent years to make the analysis and the concepts meaningful to practitioners and decision-makers and to educate them about the value of incorporating these concepts and findings into their operations.

Moreover, expenditure on quality local road networks requires not only monies but also importantly strong political will at the national and local levels. Raising an additional public budget to fund local roads is only partly a solution. Much more will depend on the quality and strength of governance and adherence to good planning, budgeting, and procurement practices in the national government and local government units. There are constraints to overcome to have good quality roads but they can be hurdled. In this light, the following recommendations are submitted:

- Raise more local revenues and improve taxation.
- Use local revenues to construct and maintain local public roads that serve the people.
- Improve procurement for local roads.
- Ensure that local roads form part of an efficient road network

REFERENCES

[1] Alexis Fillone. 2015. Evaluating Proposed Transportation Infrastructure Projects in Metro Manila Using the Transport Co-Benefit Analysis.

[2] Andreas Stephan. 1997. The Impact of Road Infrastructure on Productivity and Growth: Some Preliminary Results for the German Manufacturing Sector. *Wissenschaftszentrum Berlin für Sozialforschung Discussion Paper*, 4, (1997) 97-47.

[3] Ben Okwudili Sidiq. 2019. Significance of Road Infrastructure on Economic Sustainability. *American International Journal of Multidisciplinary Scientific Research.* (2019) 5,4;1-9. doi:10.46281/aijmsr.v5i4.405

[4] Caitlin MacLean. 2017. Investing in Transportation Infrastructure in the Philippines. Milken Institute, Santa Monica, California, USA.

[5] Cesar Queiroz and Surhid Gautam. 1992. Road Infrastructure and Economic Development. (1992).

[6] David Aschauer. 1991. Highway Capacity and Economic Growth. *Economic Perspectives.*

[7] Elena Cigu, Daniela Tatiana Agheorghiesei, Anca Florentina Gavrilută Vatamanu, and Elena Toader. 2018. Transport Infrastructure Development, Public Performance and Long-Run Economic Growth: A Case Study for the Eu-28 Countries. *Sustainability* 11, 67 (2019). doi:10.3390/su11010067

[8] Eva Ivanova and Jana Masarova. 2013. Importance of Road Infrastructure in the Economic Development and Competitiveness. *Economics and Management.* 18, 2 (2013). doi:10.5755/j01.em.18.2.4253

[9] Gohar Badalyan, Thomas Herzfeld, and Miroslava Rajcaniova. Transport infrastructure and economic growth: Panel data approach for Armenia, Georgia and Turkey. *Review of Agricultural and Applied Economics* 17, 2 (2014) 22-31. doi:10.15414/raae.2014.17.02.22-31

[10] Gilberto Llanto. 2011. Investing in Local Roads for Economic Growth. *Discussion Paper Series* 2011-38. Philippine Institute for Development Studies.

[11] John Maynard Keynes. 1939. *The General Theory of Employment, Interest, and Money*. Palgrave Macmillan. London, United Kingdom. doi:0.1007/978-3-319-70344-2

[12] Joko Purwanto, Christophe Heyndrickx, Jan Kiel, and Ofelia Betancor. 2017. Impact of Transport Infrastructure on International Competitiveness of Europe. *World Conference on Transportation Research 2017, Transportation Research Procedia* 25C, (2017), 2881-2892. doi:10.1016/j.trpro.2017.05.273

[13] Jose Patrick Rosales. 2020. *Infrastructure Financing Strategies for Sustainable Development*. Economic and Social Commission for Asia and the Pacific.

[14] Kim Larsen. 2015. GAM: The Predictive Modeling Silver Bullet

[15] Leopoldo Laborda and Daniel Sotelsek. 2019. Effects of Road Infrastructure on Employment, Productivity and Growth: An Empirical Analysis at Country Level. Journal of Infrastructure Development 11,1–2, (2019), 81–120. doi: 10.1177/0974930619879573

[16] Luo Xiaofeng, Liu Siwei, Yuan Yue. 2021. Analysis of The Impact of Highway Construction Related to Social and Environment. *E3S Web of Conferences* 233, 01 (2021). doi:10.1051/e3sconf/202123301113

[17] Mohammed Alamgir, Mason Campbell, Sean Sloan, Miriam Goosem, Gopalasamy Reuben Clements, Mahmoud Mahmoud, and William Laurance. Economic, Socio-Political and Environmental Risks of Road Development in the Tropics. *Current Biology* 27, R1130–R1140. doi:10.1016/j.cub.2017.08.067

[18] Mohammed. Aldagheiri. 2009. The Role of the Transport Road Network in the Economic Development of Saudi Arabia. *WIT Transactions on The Built Environment*, Vol 107. doi:10.2495/UT090251

[19] National Economic and Development Authority. Accelerating Infrastructure Development.

[20] Nela Vlahinić Lenz, Helga Pavlić Skender, and Petra Adelaida Mirković. 2018. The macroeconomic effects of transport infrastructure on economic growth: the case of Central and Eastern E.U. member states. *Ekonomska Istraživanja* 31, 1 (2018). doi:10.1080/1331677X.2018.1523740

[21] Ng, Law, Jakarni, and Kulanthayan. 2019. *Materials Science and Engineering* 512, (2019). doi:10.1088/1757-899X/512/1/012045

[22] Normaz Ismail and Jamilah Mohd Mahyideen. 2015. The Impact of Infrastructure on Trade and Economic Growth in Selected Economies in Asia. *SSRN Electronic Journal* (2015). doi:10.2139/ssrn.2709294

[23] OECD. 2020. *Transport Bridging Divides*, OECD Urban Studies, OECD Publishing, Paris, France. doi:10.1787/55ae1fd8-enC

[24] Paul Lombard and Lynette Coetzer. 2007. The Estimation of the Impact of Rural Road Investments on Socio-Economic Development. *International Seminar on Sustainable Road Financing & Investment 2007.*

[25] Randall Eberts. 2000. Understanding the Impact of Transportation on Economic Development. *Transportation in the New Millennium*, 5 (2000).

[26] Risper Sarah Khanani, Emmanuel Junior Adugbila, Javier Martinez, and Karin Pfeffer. 2021. The Impact of Road Infrastructure Development Projects on Local Communities in Peri-Urban Areas: the Case of Kisumu, Kenya and Accra, Ghana. *International Journal of Community Well-Being* 4,1 (2021). doi:10.1007/s42413-020-00077-4

[27] Robert Carlsson, Alexander Otto, and Jim Hall. 2013. The role of infrastructure in macroeconomic growth theories. *Civil Engineering and Environmental Systems* 30,3-4, (2013) 263-273.doi:10.1080/10286608.2013.866107

[28] Sunil Sapra. 2013. Generalized additive models in business and economics. *International Journal of Advanced Statistics and Probability*, 1, 3 (2013) 64-81. doi:10.14419/ijasp.v1i3.1022

[29] Xiao Ke, Justin Yifu Lin, Caihui Fu, and Yong Wang. 2020. Transport Infrastructure Development and Economic Growth in China: Recent Evidence from Dynamic Panel System-GMM Analysis. *Sustainability* 12, 5618 (2020). doi:10.3390/su12145618

[30] Yasir Tariq Mohmand, Aihu Wang, and Abubakr Saeed. The impact of transportation infrastructure on economic growth: empirical evidence from Pakistan. *The International Journal of Transportation Research* 9, 2 (2017), 63-69. doi:10.1080/19427867.2016.1165463