WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Prediction Model for Gross Domestic Product: A Production Approach

Melanie Aban-Serquina

MST- Mathematics, Saint Mary's University
DOI: https://doi.org/10.55248/gengpi.5.0724.1720

ABSTRACT

This research aims to predict the Gross Domestic Product of Region II, otherwise known as Cagayan Valley Region, in the Philippines, through a production approach. The GDP, being the total value of all goods and services produced, is broken down into the three sectors: Agriculture, Industry, and Services. Several time series models were used for the historical data from 2001 to 2023, gathered from the Philippine Statistics Authority, to determine the best-fit model to predict future values for GDP. These time series models include linear, polynomial degrees 3 and 4, exponential, logarithmic, power, and first autoregression. Multiple regression analysis was used to see the effect of each sector on the GDP; that the Agriculture and Services significantly impact GDP, while the Industry sector has a negligible negative effect. The polynomial degree 4 and first autoregression models gave the highest predictive accuracy with an R-squared of 98.64% and 97.21%, respectively. The conclusion is that these models can be seen as reliable predictors of the region's GDP and a great tool for policymakers, researchers, and other stakeholders in planning and deciding. Future research may consider other independent variables and more complex models to better predict the GDP.

Introduction

Gross Domestic Product (GDP) measures a country's total economic production and performance and Gross Regional Domestic Product (GRDP) is the GDP of a specific region. It reflects the total market value of all the goods and services produced by the economy at a certain period. GDP can be determined based on three different approaches namely the Expenditure, Production and Income approach. This paper is focused on production approach which is categorized into three sectors Agriculture, Industry and Service Sector. Agriculture which is a predictor variable that indicates the primary sector of the economy and in fact is defined as the first transformation of natural resources. Likewise, industry indicates the secondary sector of the economy and is the sectors which process primary goods into more elaborate products for market use. Lastly, service sector indicates the tertiary sector of the economy and can be defined as production of services which can be consumed directly without storing and transforming into other goods (Bajpai, 2023).

One study is "Gross Domestic Product as a Modern-day Economic Indicator" (Tjukanov, 2011), this study seeks to analyze the relevance of Gross Domestic Product (GDP) as an indicator for economic performance and a tool for economic policy-making. His most important source of information was the extensive literature in the field of GDP and primary research which has been done on the topic. An interview with a specialist in his research topic was also conducted to get additional information on the areas which were somehow not addressed by the literature. As the topic of his research is predominantly theoretical, a thorough literature review was a logical way to conduct a majority of the research. To produce a systematic analysis, several different articles have been examined and only the most relevant have been selected. After a deep analysis on his study, he concluded that GDP information clearly has a strong impact on the economy and it is used widely as a tool in economic policy- making. He also noted that GDP is not an adequate measure of standard of living or sustainable economics but it has to be kept in mind that it was never originally developed for this purpose. Instead its use has evolved to a point where its appropriateness can, and even should, be questioned.

In the study entitled "GDP Forecasting: Machine Learning, Linear or Autoregression?" (Maccarrone et.al. 2021), they compare the predictive accuracy of time series models, linear regression models, and machine learning algorithms in forecasting U. S. GDP. The results show that KNN can generate superior forecasts with one step ahead by following repetitive GDP cycles while to the contrary SARIMA cannot generate great forecasts for short and long term forecasting. The SARIMAX also proves to be more accurate with the use of Treasury-bills as covariates and the ARX demonstrates efficient forecasting power for one-step ahead forecasts involving proxies and yield curve data and continued capacity for multi-step forecasts. LR forecast at least 3 steps farthest jumps using proxies for the yield curve and macro variables, and second farthest in 1 step using proxies only. KNN for all three variables for short-term forecasting and LR for financial variables for long-term forecasting. It is also possible to note the time when certain models show better or worse performance for a particular task – time series models work better before the pandemic and LR and KNN models perform better after the pandemic. The findings summarize the importance of continuous forecasting with self-correcting one-step-ahead and multi-step ahead techniques giving short-term replanning as well as strategic planning.

GDP has a strong impact in the economy and it is used widely as a tool in economic policy-making. As a response to this statement, several countries have conducted researches to estimate future GDP for policy making. This could only be possible by using existing or secondary data and applying internationally recognized method, like the ARIMA model, AR model and VAR model. But simple time series models are not enough to offer accurate

forecasting but for a short time forecasting, this models are valid. As a contribution of this research, this paper will serve as stepping stone for future researchers to conduct several studies about GDP for the Region. For this study aim to find the best fit model to estimate the future GDP, this can serve as a basis for them for policy- making.

Statement of the Problem and Objectives

The study is focused on Region II (Cagayan Valley Region's) Gross Domestic Product. The researcher wants to know the behaviors of the graph of the Gross Regional Domestic Product, Agriculture, Industry Sector and the Service sector. The researcher also aims to obtain what mathematical model is fit to obtain specifically what are the estimated values of Gross Regional Domestic Product.

These are the objectives of this paper:

1. Determine the behaviors of the graph of the following:

a. Gross Regional Domestic Product

b. Industry Sector,

c. Agriculture Sector, and

d. Service Sector.

2. Identify the factors that affect the GDP of the Region.

3. Construct a time series model using the following models:

a. Linear

b. Polynomial (degree 3)

c. Polynomial (degree 4)

d. Exponential

e. Logarithmic

f. Power

g. First Autoregression

4. Identify the best fit mathematical model that estimates the Gross Domestic Product of Region II.

Methodology

Research Design

The researcher used an exploratory kind of research in order to come up with an acceptable model for prediction of Region II values through various kinds of regression techniques. This approach entailed conducting research and validations to determine the best-fit regression models by identifying different data patterns and relationships. It was shown that from the comparison between different regression techniques the researchers were able to commence on potential structures and processes in their data and were able to develop preliminary results which can guide further more specific inquiries. Besides the exploratory analysis, the researchers adopted a correlational research design to explain dependency/independence between variables. More specifically they explored or measured the association of change in GDP with the different factors that contribute to the change in GDP. Moreover, by examining the signs and strength of these coefficients, the authors were able to determine how a change in economic, political, social, and other factors affected GDP.

Sample of the Study

Located in the northeastern part of the country, the Cagayan Valley Region, designated as Region II, is the second largest in terms of land area. This region primarily lies in a large valley in northeastern Luzon, between the Cordilleras and Sierra Madre mountain ranges, and includes the Babuyan and Batanes island groups in the Luzon Strait, near the town of Aparri in Cagayan. Region II comprises five provinces, four cities, 89 municipalities, and 2,311 barangays. The researcher chose Region II to base the study on Gross Domestic Product is that their economies underwent a decline in the GDP of the three sectors; most especially agriculture. Other than the availability of many software companies for sampling and the familiarity of the researchers with the area, the region would be an ideal place to conduct data collection.

Source of Data

Annual data from 2001 to 2023 of the GDP of Region II are gathered at the website of Philippines Statistics Authority (PSA) where they release an annual publication of the National Accounts where GDP data were included being reported every year.

Treatment of Data

In applying statistical analysis through Excel in the study of Region II's GDP there were several major stages through which it passed. The first step was to compile data in Excel spreadsheets covering years from 2001 to 2023 taken from the PSA's website and structured in columnar format with Year, GDP, Agriculture, Industry, and Services. Next, the exploratory Data Analysis (EDA) was done using descriptive statistics through the use of mean, median, standard deviation and range using the Excel. Graphics skills involved the preparation of line plots to show countries' total GDP, Agriculture, Industry and Services over the years. On the other hand, regression analysis included a wide range of methods from simple linear regression to 'LINEST' in Excel or 'Regression' in Data Analysis Toolpak. In the present study Agriculture, Industry, and Services appeared as the predictors for GDP which came into the role of the criterion. Time series models tested were linear, exponential, logarithmic, exponential, power, moving average, exponential smoothing and autoregression. A significant part of model evaluation for this study was selecting the best-fit model by comparing the achieved R-squared, adjusted R- squared, and sum of squared error (SSE).

In terms of practices, these analyses in Excel entailed certain steps and actions like identifying data ranges, drawing charts, applying graphical overlays to such charts as two-way data tables or XY scatter plots, applying trendlines, that is, linear or exponential, employing built-in tools like the Data Analysis Toolpak for regression and time series analyses, and applying the F-test to determine the significance between the two-regression lines. Taken together these methods offered a complete statistical analysis needed for the data on GDP for Region II that was critical in economic development and policy decisions.

Results and Discussion

Behaviors of the graph of GDP and its components

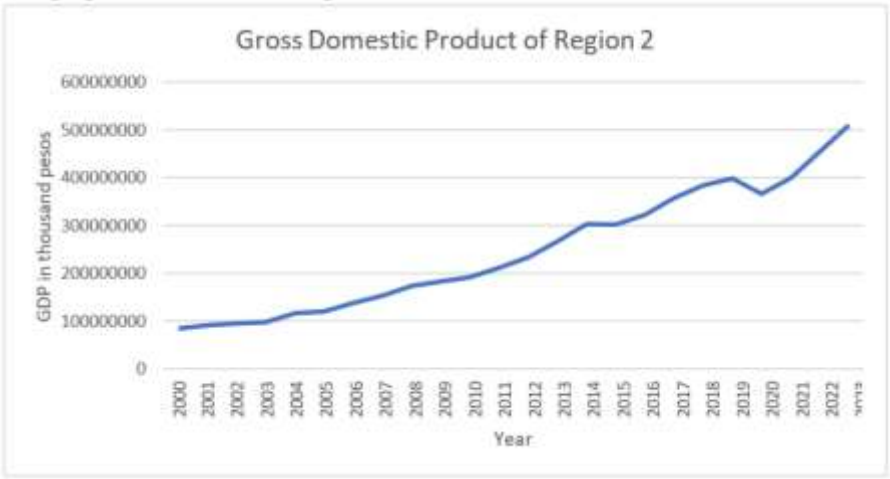

Figure 1. The GDP of Region II

The figure 1 shows the graph of Gross Domestic Product of Region II form 2000 to 2023. This display an upward trend with some fluctuations in the years 2015 and 2020. Its most significant downward trend of 8% was recorded during the year 2020, where Covid-19 virus hit worldwide. While the most high and significant upward trend of 18.7% were recorded in the year 2004.

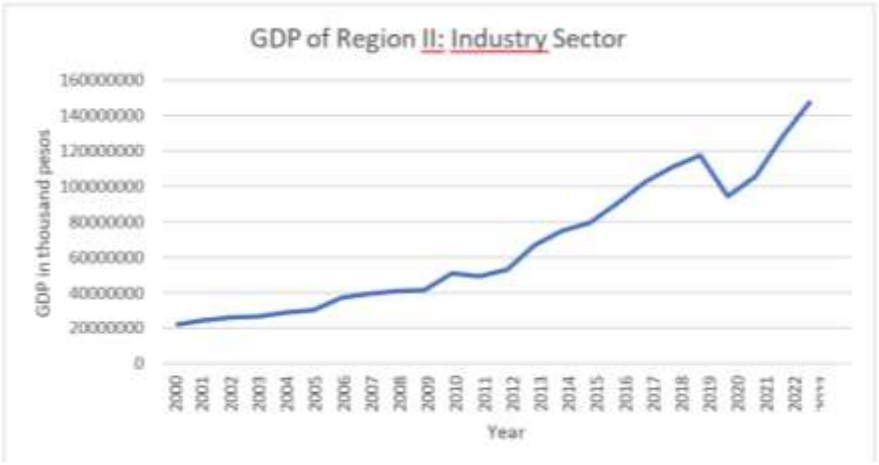

Figure 2. Production in Industry Sector

Figure 2 shows the graph of production in the industry sector of the regions GDP from 2000-2023. The behavior of the graph illustrates an upward trend with noticeable fluctuations in the last 23 years. In the year 2013, it recorded its highest growth rate before declining sharply in the year 2020. The growth of the industry sector during 2013 and before 2020 is due to the intervention of the government resulting to the huge development in the Tilapia industry wherein there are 37.25% since 2003 increase in the production of the species of cichlid fishes, which caused the acknowledgement of Cagayan Valley as the Philippines tilapia capital. (Lagasca, 2008).

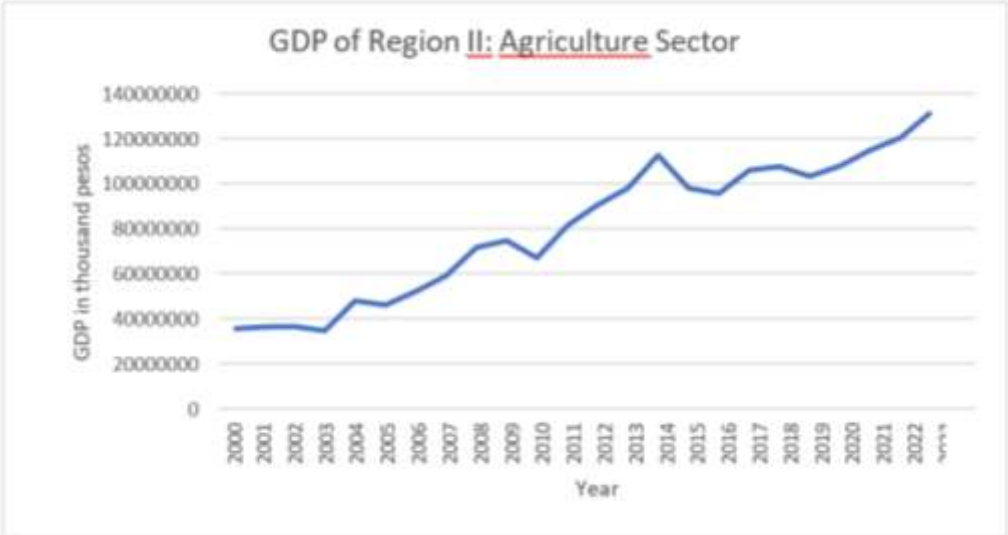

Figure 3. Production in Agriculture Sector

For the past 20 years, some fluctuations revealed because of the recessions that happened in the sector resulting to the instability of the region in its agriculture. Figure 3 illustrates the trend of the sector from 2000- 2023. During the occurrence of typhoon Lando in 2015, it reached its most significant downward trend of 13%.

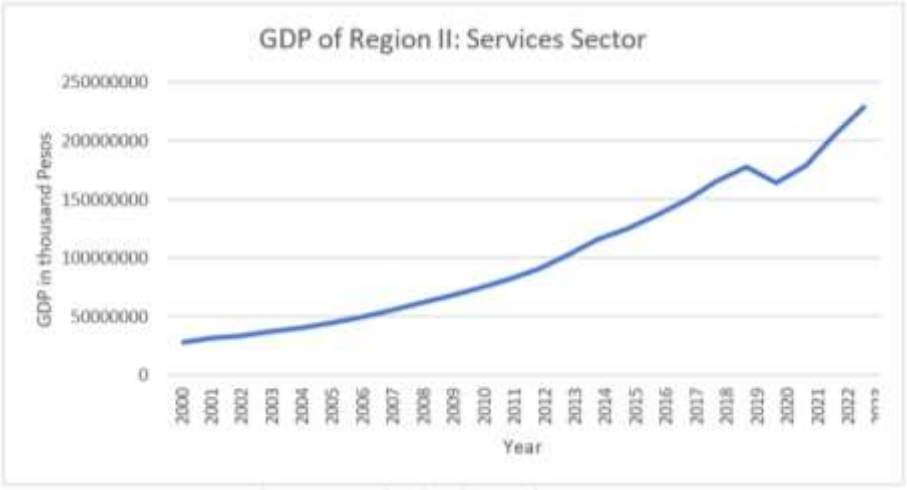

Figure 4. Production in Service Sector

Figure 4 shows the services sector of the GDP of region II for 2000-2023. With the span of 20 years, the graph exhibits an increasing trend with minimal fluctuations that can be noticed in the year 2020 which the COVID-19 pandemic hits. This reflects that before and after the pandemic, improvement and developments of new services in this sector of the region.

Analyzing the Gross Domestic Product (GDP) of the region, one can state that the overall tendency is an upward trend with certain fluctuations seasonally and intermittent blip. The industry sector also rose with a visibly irregular trend, agriculture sector over the last 20 years was characterized with some oscillations, primarily because of the recessions within the sector and services sector maintained singlehandedly a steady growth trajectory without much of a volatility despite a sharp decline in 2020 owing to the pandemic; thus, it presents how there were significant advancements and diversification of services before and after the pandemic hit the world.

The Factors That Affect The GDP Of The Region Table 1. Components of GDP

	Coefficients	*Standard Error*	*t Stat*	*P-value*

Intercept	-1.723327232	0.859127	-2.00591	0.058583
Industry Sector	-9.99888E-08	5.33E-08	-1.87638	0.075273
Agriculture Sector	9.5309E-08	2.07E-08	4.609786	0.000169
Services Sector	1.28934E-07	3.67E-08	3.511098	0.002198

The results of the multiple linear regression analysis highlighted in the above table gives a measure of how the GDP is affected by each sector or component. For the industry sector, the coefficient is -9.99888E-08, with a standard error of 5.33E-08, a t-statistic of -1.87638, and a P-value of 0.075273. This can be interpreted to imply a very low negative association of the industry sector to the GDP. However, it is not significant at .05 level of significance but could be considered to be highly significant at .01 level of significance. The agriculture sector shows a positive relationship with the GDP, with a coefficient of 9.5309E-08, a standard error of 2.07E-08, a t- statistic of 4.609786, and a P-value of 0.000169. This indicates that for each unit increase in the agriculture sector, the GDP increases by 9.5309E-08 units, and this relationship is statistically significant. Similarly, the services sector has a positive coefficient of 1.28934E-07, with a standard error of 3.67E-08, a t-statistic of 3.511098, and a P-value of 0.002198. This means that for each unit increase in the services sector, the GDP increases by 1.28934E-07 units, and this relationship is also statistically significant.

Time Series Models

a. Linear

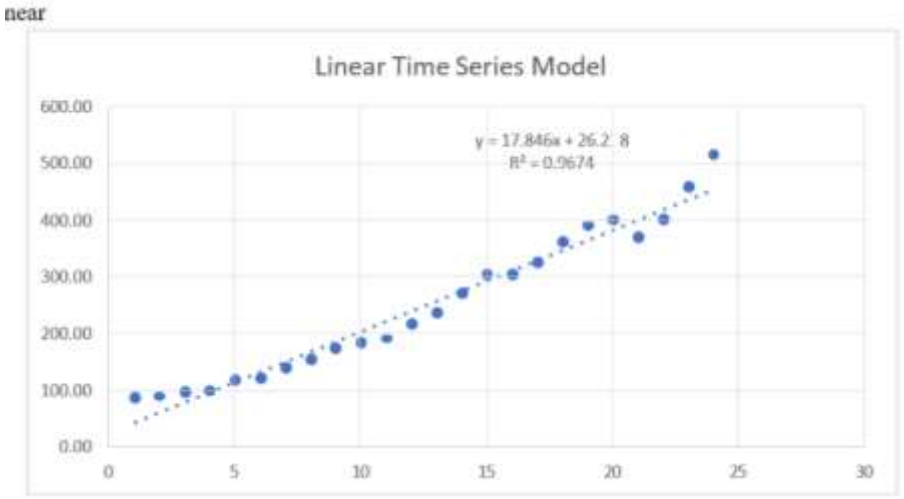

Figure 5. Linear Time Series Model of Region II GDP

Figure 5 illustrates the linear time series model of region II GDP from 2000-2023. The model predicts the GDP of the region with the equation y = 17.846x + 26.218. The coefficient of determination, 0.9674, signifies that 96.74% of the variation in GDP of the region can be explained by a unit change in time.

b. Polynomial (degree 3)

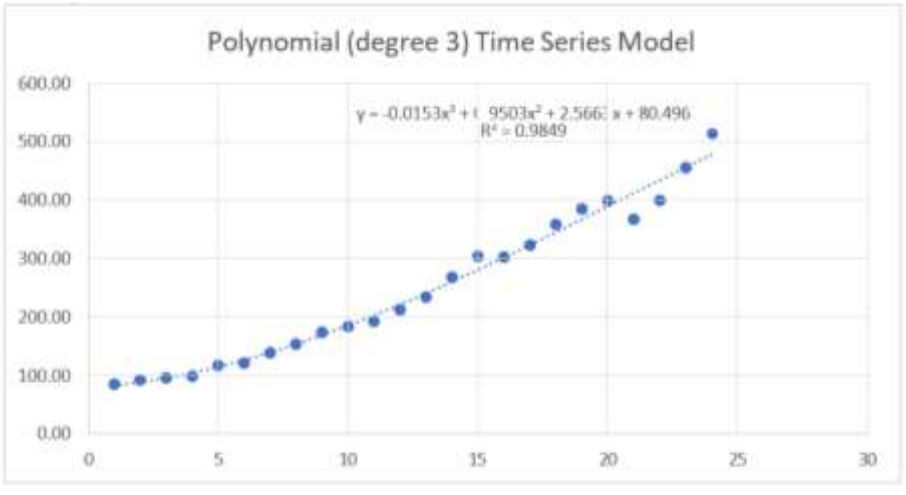

Figure 6. Polynomial (degree 3) Time Series Model of GDP

Figure 6 illustrates the polynomial (degree 3) time series model of region II GDP from 2000-2023. The model predicts the GDP of the region with the equation $y = -0.0153x^3 + 0.9503x^2 + 2.5663x + 80.496$. The

coefficient of determination, 0.9849, signifies that 98.49% of the variation in GDP of the region can be explained by a unit change in time.

c. Polynomial (degree 4)

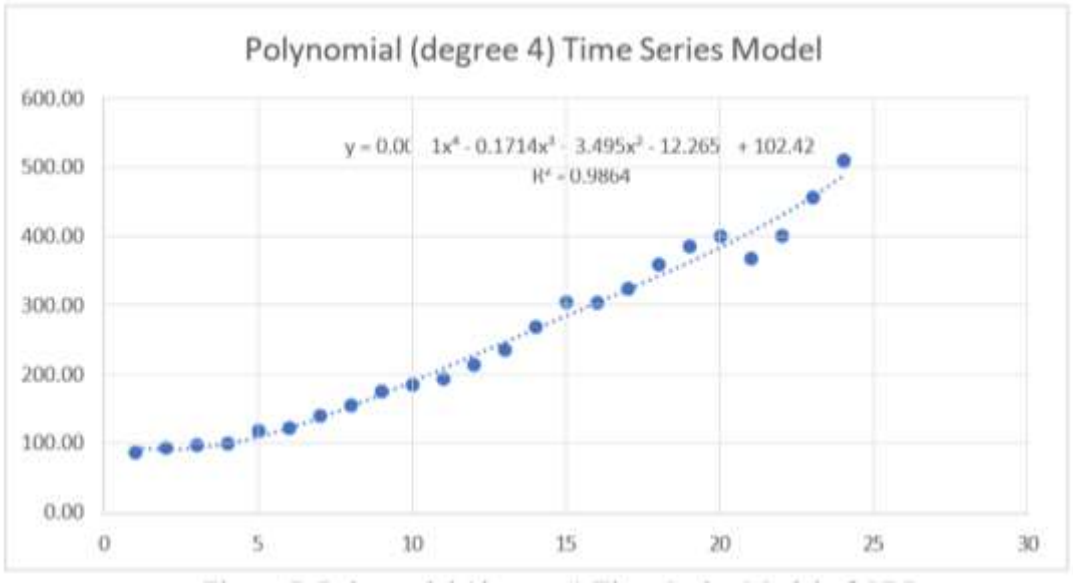

Figure 7. Polynomial (degree 4) Time Series Model of GDP

Figure 7 illustrates the polynomial (degree 4) time series model of region II GDP from 2000-2023. The model predicts the GDP of the region with the equation $y = 0.0031x^4 - 0.1714x^3 + 3.495x^2 - 12.265x + 102.42$. The coefficient of determination, 0.9864, signifies that 98.64% of the variation in GDP of the region can be explained by a unit change in time.

d. Exponential

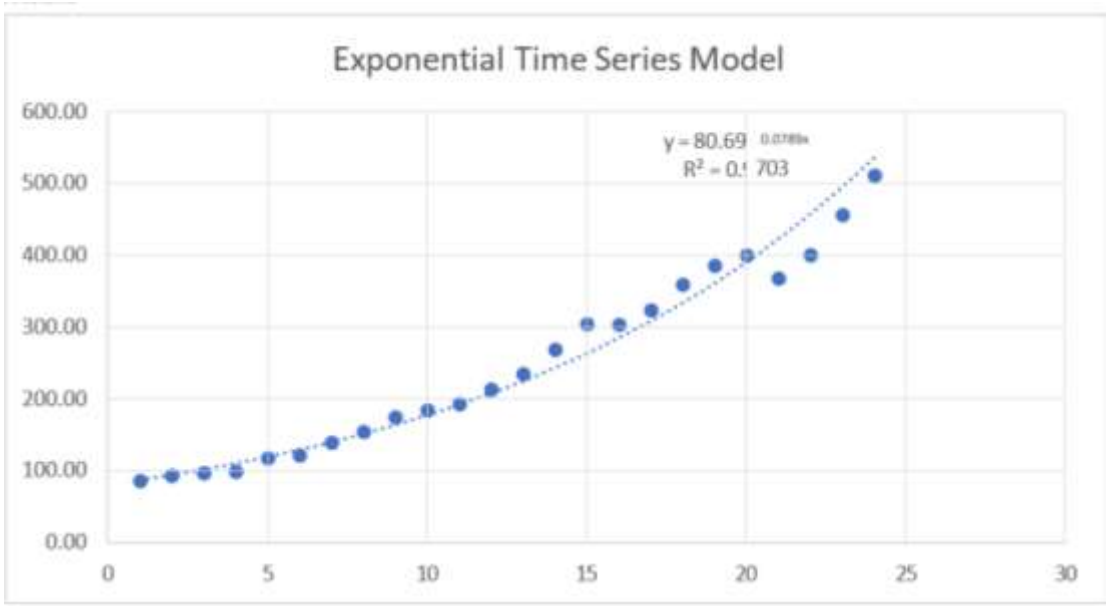

Figure 8. Exponential Time Series Model of GDP

Figure 8 illustrates the exponential time series model of region II GDP from 2000-2023. The model predicts the GDP of the region with the equation $y = 80.69e^{0.0789x}$. The coefficient of determination, 0.9703, signifies that 97.03% of the variation in GDP of the region can be explained by a unit change in time.

e. Logarithmic

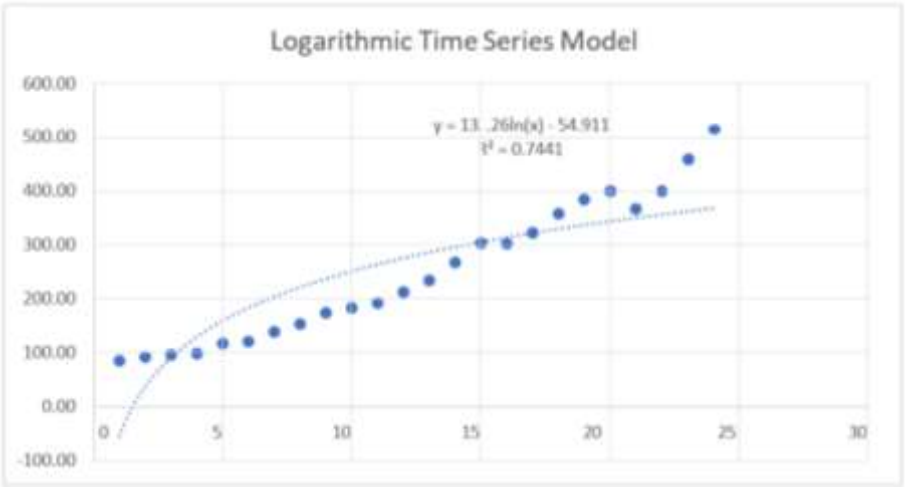

Figure 9. Logarithmic Time Series Model of GDP

Figure 9 illustrates the logarithmic time series model of region II GDP from 2000-2023. The model predicts the GDP of the region with the equation y = 133.26ln(x) - 54.911. The coefficient of determination, 0.7441, signifies that 74.41% of the variation in GDP of the region can be explained by a unit change in time.

f. Power

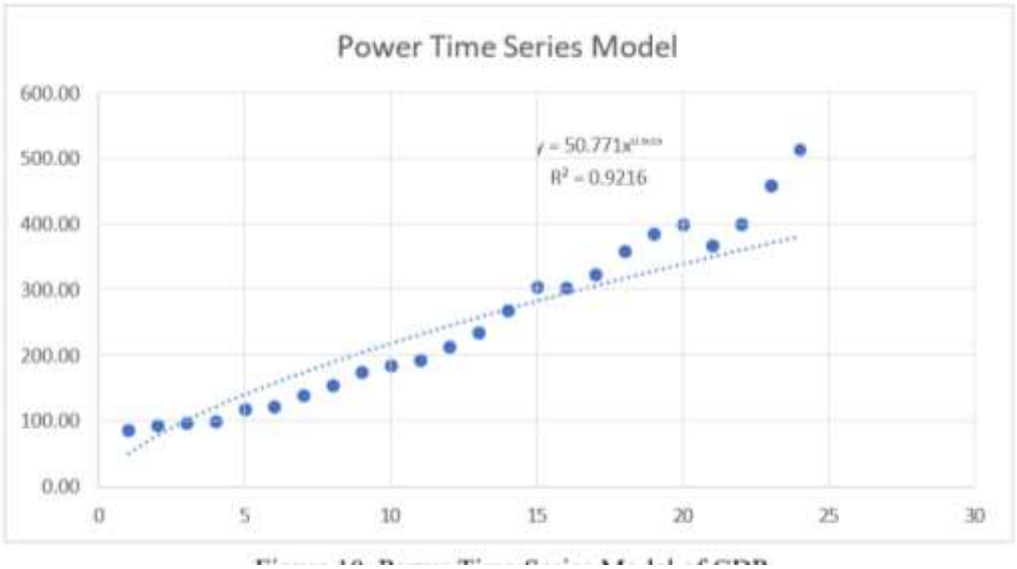

Figure 10. Power Time Series Model of GDP

Figure 10 illustrates the power time series model of region II GDP from 2000-2023. The model predicts the GDP of the region with the equation y = 50.771x0.635. The coefficient of determination, 0.9216, signifies that 92.16% of the variation in GDP of the region can be explained by a unit change in time.

g. First Autoregression

Table 2. First Autoregression of GDP

	Coefficients	Standard Error	T Stat	P-value
Intercept	-0.4637	0.5534	-0.8380	0.4115
X variable	0.05659	0.0021	27.0475	8.33404E-18

r^2 = 0.9721

Table 2 shows the first autoregression result of GDP of the region from the years 2000-2023. The first autoregression model can be used in predicting the GDP of the region. The first autoregression model is $y = 0.05659y_{n-1} - 0.4637$. This can explain 97.21% of the variances.

Summary Table

Table 3. Equation of Time Series Model of GDP and their r^2

	Equation	$r2$
Linear	y = 17.846x + 26.218	0.9674
Polynomial (degree 3)	$y = -0.0153x^3 + 0.9503x^2 + 2.5663x + 80.496$	0.9849
Polynomial (degree 4)	y = 0.0031x4 - 0.1714x3 + 3.495x2 - 12.265x + 102.42	0.9864
Exponential	$y = 80.69e^{0.0789x}$	0.9703
Logarithmic	y = 133.26ln(x) - 54.911.	0.7441
Power	y = 50.771x0.635	0.9216
First Autoregression	$y = 0.05659y_{n-1} - 0.4637$	0.9721

Table 3 illustrates the equation generated by different time series model with their corresponding r- squared of the region GDP from the year 2000-2023.

The Best Fit Mathematical Model That Estimates The Gross Domestic Product Of Region II

Table 4. Best Fit Model for the GDP of the Region

	$r2$	SSE
Linear	96.74%	4935.94
Polynomial (degree 3)	98.49%	599323.15
Polynomial (degree 4)	98.64%	55985392
Exponential	97.03%	286.39
Logarithmic	74.41%	434.65
Power	92.16%	1514.03
First Autoregression	97.21%	275.70

Table 4 presents the table to determine the best fit model for the region II's GDP. The value of each time series r-squared and SSE are shown. Based on the results, the degree 4 polynomial has the highest R-Squared value of 98. 64, pointing to variance explained by this model as the best fit model . However, the first autoregression model also has a high R-squared value of (97. 21%) meaning this model is highly effective in predicting the GDP in the model.

Summary

This study was conducted to forecast the gross domestic product of Region II, otherwise known as the Cagayan Valley Region, in the Philippines through the use of selected regression techniques. It looks at the trend of the GDP and its three major components: Agriculture, Industry, and Services. Different time series models, such as linear, polynomial, exponential, logarithmic, power, and autoregression models, are fitted to determine the best-fit model for prediction purposes of the region's GDP.

The results are that the GDP of Region II generally shows a growing trend with fluctuations, particularly a huge fall during the COVID-19 pandemic in 2020. The Industry sector shows an increasing trend with ups and downs, peaking in 2013 due to the government's intervention in the Tilapia industry. There are fluctuations in the Agriculture sector due to economic recession and natural disasters, like Typhoon Lando in 2015. The Services

sector shows an almost steady growth trajectory with fewer fluctuations but still under the impact of the decline in 2020 due to the pandemic.

Multiple Linear Regression Analysis indicates that the sectors agriculture and services have a significant positive impact on GDP. The industry sector has a very low negative influence on GDP, but this relationship is not statistically significant. A number of time-series models have been tested namely, linear, polynomial of degree 3 and 4, exponential, logarithmic, power, and first autoregression. The polynomial of degree 4 model presents the highest R-squared value of 98.64%, representing the best fit in terms of explained variance. The first autoregressive model also presents a very high R-squared value of 97.21% and a low sum of squared errors (SSE), hence it is the best fit model of prediction for GDP.

Conclusion

The study concluded that the polynomial degree 4 with the equation y = 0.0031x4 - 0.1714x3 + 3.495x2

- 12.265x + 102.42 and first autoregression with the equation y=0.05659y_(n-1)-0.4637 models fit best for the prediction of Region II GDP. These models explain a high portion of the variance in GDP and give reliable predictions based on historical data. The first autoregression model, in particular, balances between high explanatory power and low prediction error.

Recommendations

Based on the findings of the study, these are the recommendation:

1. Policymakers should consider the Agriculture and Services sectors as key drivers of economic growth in Region II and prioritize investments and interventions in these areas.
2. Further studies for future researchers should look into more variables and complex models in order to fine-tune the prediction of GDP.
3. The researcher also propose to seek for data of the subsectors of GDP to predict GDP differently.
4. Informing the NEDA, government and businessmen to use the generated equations to predict GDP to strategize on how to minimize or maximize the growth of the region's GDP.

References

Bajpai, P. (July 12, 2023). Emerging Markets: Analyzing the Philippines's GDP. Investopedia. https://www.investopedia.com/articles/investing/091815/emerging-markets-analyzing-philippines-gdp.asp PSA (n.d.). Gross Regional Domestic Product. GOVPH. https://rssocar.psa.gov.ph/car-grdp

Maccarrone, G., Morelli, G. & Spadaccini S. (October 15, 2021). GDP Forecasting: Machine Learning, Linear or Autoregression? Front Artif. Intell. 4:757864. doi: 10.3389/frai.2021.757864. PMID: 34723174

Tjukanov, T. (October 2011). Gross Domestic Product as a Modern-day Economic Indicator. Helsinki Metropolia University of Applied Sciences, Bachelor of Business Administration, International Business and Logistics, Bachelor's Thesis. https://core.ac.uk/download/pdf/38047645.pdf

Towers, L. (October 27, 2015). Typhoon Lando Causes P5.9 billion Damage to Agriculture. The Pig Site. https://www.thepigsite.com/news/2015/10/typhoon-lando-causes-p59-billion-damage-to-agriculture

OpenStat (n.d.) National Accounts of the Philippines. PSA. https://openstat.psa.gov.ph/Database/Economic-Accounts/National-Accounts-of-the-Philippines

Lagasca, C. (January 13, 2008).Cagayan Valley country's tilapia capital. PhilStar Global. https://www.philstar.com/nation/2008/01/13/38369/cagayan-valley-countryrsquos-tilapia-capital