World Academy of Science, Engineering and Technology
International Journal of Economics and Management Engineering
Vol:15, No:6, 2021

Strategic Investment in Infrastructure Development to Facilitate Economic Growth in the United States

Arkaprabha Bhattacharyya, Makarand Hastak

***Abstract*—**The COVID-19 pandemic is unprecedented in terms of its global reach and economic impacts. Historically, investment in infrastructure development projects has been touted to boost the economic growth of a nation. The State and Local governments responsible for delivering infrastructure assets work under tight budgets. Therefore, it is important to understand which infrastructure projects have the highest potential of boosting economic growth in the post-pandemic era. This paper presents relationships between infrastructure projects and economic growth. Statistical relationships between investment in different types of infrastructure projects (transit, water and wastewater, highways, power, manufacturing etc.) and indicators of economic growth are presented using historic data between 2002 and 2020 from the U.S. Census Bureau and U.S. Bureau of Economic Analysis (BEA). The outcome of the paper is the comparison of statistical correlations between investment in different types of infrastructure projects and indicators of economic growth. The comparison of the statistical correlations is useful in ranking the types of infrastructure projects based on their ability to influence economic prosperity. Therefore, investment in the infrastructures with the higher rank will have a better chance of boosting the economic growth. Once, the ranks are derived, they can be used by the decision-makers in infrastructure investment related decision-making process.

***Keywords*—**Economic growth, infrastructure development, infrastructure projects, strategic investment.

I. Introduction

THE disruption caused by the COVID 19 pandemic has imposed an unprecedented challenge on the economic growth of all nations across the globe and the U.S. is no exception to that. The recent advance estimate revealed by the U.S. BEA shows that the real U.S. Gross Domestic Product (GDP) has fallen at an annual rate of 32.9% in the second quarter of 2020 [1]. Historically, investment in infrastructure development projects has been touted to boost the economic growth of a nation. Infrastructure projects like transit, water and wastewater, airports, roads or highways, electricity, waterways and ports etc., all contribute to a smoothly functioning economy. In the U.S. the chronic under investment in the infrastructure projects has created inefficiencies in various sectors of the economy. A report published by the American Society of Civil Engineers estimated an investment need of $2 trillion for upgrading the infrastructures to meet the future demands [2]. The report also mentioned that under investment in infrastructure can have cascading impact on economy, GDP, employment etc. The State and Local governments responsible for delivering infrastructure assets work under tight budgets. Therefore, it is required to understand which infrastructure projects have the highest potential of boosting the economic growth in the post-pandemic era. This paper presents relationships between different types of infrastructure projects and the economic growth. To achieve that objective, this paper has adopted multiple linear regression model to predict the nominal GDP utilizing historic infrastructure construction spending data between 2002 and 2020 from the U.S. Census Bureau. To further investigate the short-term and long-term impacts of infrastructure construction, this paper has used six different lag periods between zero and five years. The lag period has been defined as the difference between the year of construction spending and the year when its impact is expected. A shorter lag means an immediate impact whereas a longer lag indicates long-term impact. In this paper six different prediction models were developed for six lag period. The outcomes show that the model with a lag of four years can predict the GDP with the highest accuracy. The models were further used to identify the infrastructures which have the highest potential to influence GDP. It has been found that transportation and highway construction have the maximum potential to influence GDP on both short and long term. These outcomes can be used by the decision makers in infrastructure investment prioritization related decision-making process.

Arkaprabha Bhattacharyya is a PhD student at Lyles School of Civil Engineering, Purdue University, West Lafayette, IN-47907, (e-mail: bhatta23@purdue.edu).

Makarand Hastak is a Professor and Head of Division of Construction Engineering and Management, Professor of Lyles School of Civil Engineering, Purdue University, West Lafayette, IN-47907, (e-mail: hastak@purdue.edu)

II. Literature Review

The U.S. Department of Commerce ranks GDP as one of the most influential economic measures that can influence U.S. financial markets [3]. Therefore, this paper has considered GDP as the indicator of growth and analyzed its correlations with infrastructure construction spending. The analysis of the correlation between infrastructure investment and economic growth has been a topic of research for the last few decades. Canning and Fay [4] have used the physical measures of transportation networks like kilometers of paved roads, railways to estimate the social rates of return. They found that in developed countries the rates of return can be between 5% and 25%. Shi et al. [5] have investigated the relationships between infrastructure capital for electricity, roadways, railways, and telecommunication infrastructures and real GDP per worker in China. They found that infrastructure has not always translated into faster growth. Tian and Li [6] have found that infrastructure construction

facilitated the economic growth and per capita output along the "Belt and Road". They also found an inverted U-shaped relationship between infrastructure construction and economic growth. Kumo [7] has adopted pairwise Granger causality test to investigate causality between economic infrastructure investment which consists of both public and private investment and economic growth in South Africa using historic data between 1960 and 2009. The paper found a strong causality and concluded that economic infrastructure investment drives the long-term economic growth in South Africa. Zhang [8] has also used Granger causality test to understand the relationships between transportation infrastructure construction and GDP in China. The research has found that economic development is of reciprocal causation with railways, inland waterways and civil aviation construction. The test results also show that road construction fails to play a role in promoting economic development. Pradhan and Bagchi [9] have also examined the presence of any nexus transportation infrastructure and economic growth in India using data between 1970 and 2010. They found bidirectional causality between road transport infrastructure and economic growth which means road transport facilitates economic growth and vice versa. They have also found unidirectional causality between railway infrastructure and economic growth. Lombard et al. [10] has used cross sectional multiple regression analysis to investigate the relationship between highway and economic development in the state of Indiana in the U.S. They have found that highway mileage has a significant association with the economic growth.

III. Methodology

The methodology has been shown in Fig. 1. This paper aims to analyze the relationships between construction spending in different types of infrastructure and the GDP. For that, a prediction model has been developed to predict the GDP of a quarter based on the cumulative construction spending in different infrastructure sector during that quarter. In this paper, eight types of construction spending have been considered as the predictor variables for the prediction model. They are health care, transportation, communication, power, highway, sewage and waste disposal, water supply, and manufacturing. The construction spending data between 2002 and 2020 have been collected from the U.S. Census Bureau's survey data [11]. The survey covers construction work done each month on new structures and improvements on existing structures by both public and private sectors. Data estimates include the cost of labor and materials, cost of architectural and engineering work, overhead costs, interest and taxes paid during construction, and contractor's profits. The construction spending in a quarter has been derived as the sum of the construction spending constituting the quarter. The GDP data between 2002 and 2020 were collected from U.S. BEA website [12].

For developing the prediction model, six different lag periods between the year of construction spending and the year of GDP have been used. The lags are 0, 1 year, 2 years, 3 years, 4 years, and 5 years. The lag period of "m" year implies that the GDP of quarter "i" of year "j" has been predicted based on the construction spending of quarter "i" of year "j-m" where m ∈[0, 5]. Therefore six different prediction models have been developed and their performances were compared. Finally, the prediction models were used to compare the relative influence of different variables in predicting the GDP to identify the variables, which have the maximum potential to influence GDP.

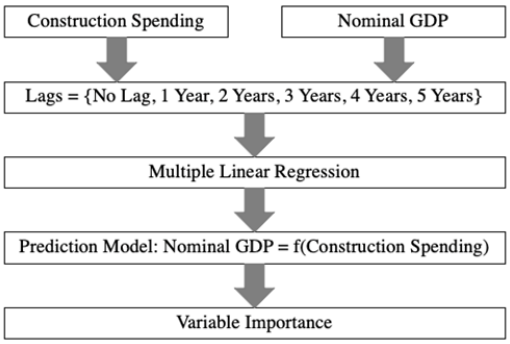

Fig. 1 Methodology of Research

IV. Prediction Model Development

This section discusses about the procedure which has been followed for developing the prediction model to predict the GDP based on the construction spending in different types of infrastructures.

A. Comparison of Correlations

TABLE I
COMPARISON OF CORRELATIONS

Lag	He	Tr	Co	Po	Hi	Se	Wa	Ma
0	0.69	0.96	0.42	0.87	0.92	0.66	0.38	0.91
1	0.61	0.94	0.35	0.84	0.89	0.56	0.18	0.88
2	0.54	0.92	0.23	0.85	0.87	0.52	0.07	0.87
3	0.48	0.92	0.08	0.91	0.87	0.51	-0.05	0.88
4	0.47	0.94	-0.03	0.91	0.86	0.56	-0.04	0.85
5	0.53	0.93	-0.12	0.91	0.83	0.55	0.05	0.79

He = Health Care, Tr = Transportation, Co = Communication, Po = Power, Hi = Highway, Se = Sewage and Waste Disposal, Wa = Water Supply, and Ma = Manufacturing.

Before developing the prediction model, the research analyzed the correlations between the predictor and response variables. For that Pearson's correlation coefficient has been used. Pearson's correlation coefficient is used to measure the strength of a linear association between two variables. The Pearson's correlation coefficient of 1 indicates perfectly positive correlation whereas, -1 indicates a perfectly negative correlation. The correlations between the GDP and eight types of construction spending as discussed in the previous section were computed for all six lags. The correlation coefficient between two variables x (predictor) and y (response) for a particular lag m is computed by tallying y(t) with x(t-m) where t ∈[Quarter 2 of 2020, Quarter 1 of 2002] and m ∈[0, 5]. The results are shown in Table I. The results in Table I show a high positive correlation between GDP and transportation, GDP and power, GDP and highway, GDP and

manufacturing. These four predictors have maintained a very high positive correlation for all lag periods. The correlations between GDP and other four predictors have been found to be moderate when there is no lag, but they have decreased with the increase in lag.

B. Multiple Linear Regression

Multiple linear regression (MLR) model is one of the most popular statistical analysis tools to predict a response variable based on predictor variables. The goal of MLR is to model the linear relationship between the predictor and response variables. In an MLR model, the predictor variables are assumed to be independent of each other. The assumption is applicable to the predictors that have been considered in this research. Equation (1) shows the formulation of the MLR.

$$Y_{GDP} = b_0 + b_{He}x_{He} + b_{Tr}x_{Tr} + b_{Co}x_{Co} + b_{Po}x_{po} + b_{Hi}x_{Hi} + b_{Se}x_{Se} + b_{Wa}x_{Wa} + b_{Ma}x_{Ma} + \epsilon \quad (1)$$

where, Y_{GDP} is the response variable which is GDP. "b_0" is the intercept term for the MLR. "b_i" is the coefficient for a predictor variable x_i and ϵ is the residual error. The same abbreviations as in Table I have been followed in the naming of the predictor variables.

For six different lag, six different MLR models have been developed. A predictor variable has only been used in the constituting an MLR model, if it showed a good correlation (Pearson's correlation coefficient value > 0.3). Therefore, when the MLR model for lag 1 was created, water supply (Wa) was removed from the list of predictors and the prediction model was created using the remaining seven predictor variables as shown in (2):

$$Y_{GDP} = b_0 + b_{He}x_{He} + b_{Tr}x_{Tr} + b_{Co}x_{Co} + b_{Po}x_{po} + b_{Hi}x_{Hi} + b_{Se}x_{Se} + b_{Ma}x_{Ma} + \epsilon \quad (2)$$

From lag 2 and onwards, both communication (Co) and water supply (Wa) started illustrating weak correlations. Hence, for lag 2 and onwards, the remaining six predictors showing strong correlation as shown in (3) were used in the MLR model.

$$Y_{GDP} = b_0 + b_{He}x_{He} + b_{Tr}x_{Tr} + b_{Po}x_{po} + b_{Hi}x_{Hi} + b_{Se}x_{Se} + b_{Ma}x_{Ma} + \epsilon \quad (3)$$

C. Model Performance

The goodness of fit of different MLR models were compared using 3 metrics: Adjusted Coefficient of Determination (R^2), Root Mean Square Error (RMSE), and Mean Average Error (MAE). The six prediction models have different number of predictors. To overcome the influence of the number of predictors, adjusted R^2 has been preferred over R^2. Adjusted R^2 is a modified version of R^2 that has been adjusted for the number of predictors.

$$\text{Adjusted } R^2 = 1 - \frac{(1-R^2)(n-1)}{n-p-1} \quad (4)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (5)$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (6)$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (7)$$

where, "n" is the number of observations or datapoints in the dataset, "p" is the number of predictors in the model, y_i is the original values of the response variable GDP, $\hat{y}_i$ is the fitted values of the response variable based on the MLR model and $\bar{y}$ is the mean of the original values of response variable.

D. Variable Importance

For analyzing the relative influence of different predictors on the accuracy of the MLR models, this paper has adopted the method suggested by [13]. The method computes the change of loss function (loss of sum of squared error, loss of RMSE, loss of accuracy etc.) of the model by randomly permuting the predictor variables. The variable, for which the change of the loss function from the original model is maximum, is considered to have the maximum influence among the predictor variables. This type of approach is model agnostic, and the outcomes are compact and easy to interpret. Let X be a matrix of "n" observations with "p" explanatory variables. For this paper, "n" is number of quarters between January 2002 and June 2020 which is 74 and "p" is the number of predictor variables which is 8 for lag zero. Now, the response variable Y is predicted based on the predictors using a function *f()*. Therefore,

$$\hat{y} = (f(x_1), f(x_2), \ldots\ldots, f(x_n)) \quad (8)$$

Let, L ($\hat{y}$, X, y) be the loss function for the prediction model. The loss function can be value of log likelihood or any other model performance measures. The following algorithm is followed for quantifying relative influence of different variables.

Step1. L^0 ($\hat{y}$, X, y) is calculated for the original model.

Step2. The matrix X^{*j} is created by randomly permuting the j^{th} column of matrix X. By doing this, basically the observations corresponding to the j^{th} predictor variable are being permuted.

Step3. A prediction model is developed using this permuted dataset X^{*j} and prediction of the response variable is computed $\hat{y}^{*j}$.

Step4. The loss function is computed for the new set of predictions.

$$L^{*j} = L(\hat{y}^{*j}, X, y) \quad (9)$$

Step5. The deviation of the loss function (L^{diff}) is calculated using (10).

$$L^{diff} = L^{*j} - L^0 \quad (10)$$

For each variable, step 2 to step 5 are repeated and the change of loss function is computed. Lastly, the change of loss function due to permutation of all predictor variables was compared. A higher change of loss function indicates a relatively higher importance in the prediction model. It should be noted that the use of resampling or permuting the data in step 2 involves randomness. Therefore, to achieve consistent outcomes, multiple permutations need to be performed. In this paper, 50 permutations were performed and the average of 50 L^{diff}s was used for the comparison.

V. Results

The performance of the six prediction models are shown in Table II. The rationale for adopting adjusted R^2 over normal R^2 has been explained in the previous section. It can be seen that the number of predictors has changed with the change of lag period. It can be seen from Table II that in terms of adjusted R^2, the MLR model without any lag has performed the best. But when the model errors were compared, the MLR model with the lag of four years has outperformed the remaining models.

TABLE II
Performance of Prediction Models

Lag Period	No of Predictors	Adjusted R^2 Value	RMSE	MAE
No Lag	8	0.94	664.9	546.7
1 Year	7	0.89	892.7	719.7
2 Years	6	0.90	813.2	656.2
3 Years	6	0.92	706.7	576.8
4 Years	6	0.92	645.7	526.9
5 Years	6	0.91	663.3	527.7

For all six models, the residual errors were tested. No definite pattern was observed in the residual plot, which supports the accuracy of the model.

Next the variable importance plots were created using the method discussed in the previous section. One of the major disadvantages of this approach is its dependence on the random nature of the permutations. Hence, 50 permutations were conducted, and the average loss of function was compared. Figs. 2 (a)-(f) show the outcomes of the variable importance analysis.

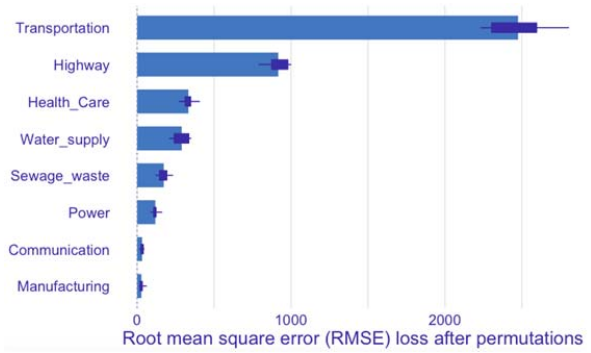

Fig. 2 (a) Variable Importance Plot (No Lag)

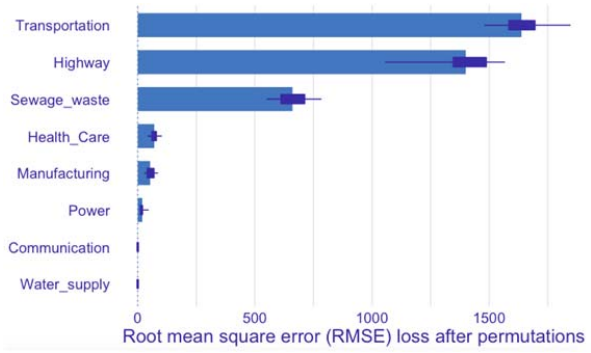

Fig. 2 (b) Variable Importance Plot (Lag = 1 Year)

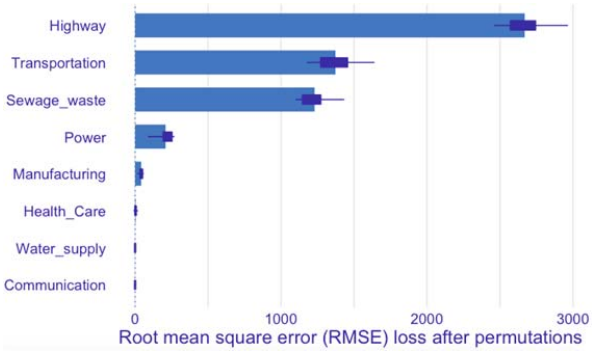

Fig. 2 (c) Variable Importance Plot (Lag = 2 Years)

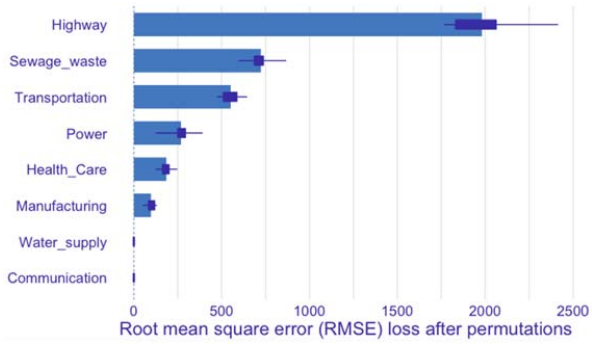

Fig. 2 (d) Variable Importance Plot (Lag = 3 Years)

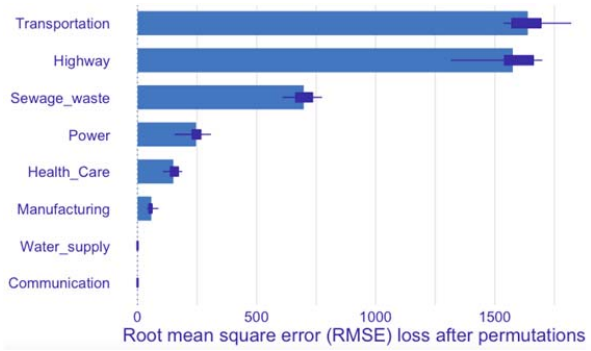

Fig. 2 (e) Variable Importance Plot (Lag = 4 Years)

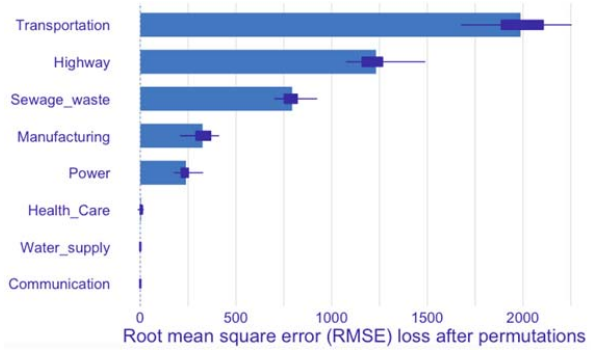

Fig. 2 (f) Variable Importance Plot (Lag = 5 Years)

It can be seen from the figures that the importance of the predictors has changed with the change of lag period. Transportation has been the most influential variable for four out of six models. Highway is another one of the most influential variables. It has been the most influential for two models and 2nd most influential for four models. The importance of sewage and waste disposal has increased with the increase in the lag period. The same has happened for power infrastructure as well. Despite having very strong correlation with the GDP, manufacturing has failed to be at the top half of the list of influential predictors. Another interesting outcome is the relatively lower influence of power infrastructure spending on the GDP. The influence of health care infrastructure construction has been consistently low. The remaining two types of infrastructures: water supply and communication did not show any strong correlation from for lag 2 and onwards. Therefore, they were not used in the last four prediction models.

VI. Discussion

This paper has investigated the relative influence of different types of infrastructures spending on the nominal GDP in the U.S. using historical data between 2002 and 2020. The research has found that out of the eight types of infrastructures which were considered for the research, transportation and highway construction spending have the maximum potential to influence the prediction accuracy of GDP. These two have performed better than the others for all six lags. The consistent higher position indicates both short-term and long-term benefit from this type of construction. Again, the construction of sewage and waste disposal furnishes moderate influence on both short and long term. The construction spending in water supply and communication infrastructures only has short term benefits. They did not furnish sufficiently strong correlation with the GDP when the lag has been more than two years. Mishra et al. [14] have found a positive causality between energy consumption and GDP in the pacific island countries. But this paper shows a relatively weaker influence of power infrastructure construction spending over GDP. Manufacturing and health care infrastructure construction spending has also furnished relatively lower potential to influence the GDP. Interestingly manufacturing has shown very high positive correlation with GDP (Table I). But when multiple variables are considered, its influence has somehow diminished.

Out of the six prediction models developed in this paper, the model with the lag of four years has performed the best. The error for the model with four years of lag is the minimum and the adjusted R^2 is also on the higher side. Therefore, in future if GDP needs to be predicted based on the construction spending on different types of infrastructure a lag of four years might be considered to minimize the error of prediction.

VII. Conclusion

Historically infrastructure spending has been touted to bolster the economic growth. This paper has investigated the relative influence of different types of infrastructure construction on GDP. The paper has adopted MLR for predicting GDP based on the infrastructure construction spending. Six different regression models were developed for six different lag period to investigate both short-term and long-term benefits. The prediction models were then used to estimate the relative influence of different predictors. It has been found transportation and highway construction has the highest potential to influence GDP on both short and long term. Therefore, if the decision makers want to prioritize investment in a particular type of infrastructure to boost the economy, they should prioritize investing in transportation and highway construction. Moreover, if the decision makers want to predict the impact of infrastructure construction spending on GDP, they should consider a lag of four years to minimize the errors of prediction. The outcomes can be used by the decision makers for infrastructure investment related decision-making process.

References

[1] U.S. Bureau of Economic Analysis (https://www.bea.gov/news/2020/gross-domestic-product-2nd-quarter-2020-advance-estimate-and-annual-update) accessed on October 15, 2020

[2] American Society of Civil Engineers (https://www.infrastructurereportcard.org/solutions/investment/) accessed on October 15, 2020

[3] U.S. Department of Commerce https://www.commerce.gov/data-and-reports/economic-indicators Accessed on October 13, 2020

[4] Canning, D., & Fay, M. (1993). The effects of transportation networks on economic growth. doi: https://doi.org/10.7916/D80K2H4N

[5] Shi, Y., Guo, S., & Sun, P. (2017). The role of infrastructure in China's regional economic growth. *Journal of Asian Economics*, *49*, 26-41. doi: https://doi.org/10.1016/j.asieco.2017.02.004

[6] Tian, G., & Li, J. (2019). How Does Infrastructure Construction Affect Economic Development along the "Belt and Road": By Promoting Growth or Improving Distribution?. *Emerging markets finance and trade*, *55*(14), 3332-3348. doi: https://doi.org/10.1080/1540496X.2019.1607725

[7] Kumo, W. L. (2012). Infrastructure investment and economic growth in South Africa: A granger causality analysis. *African development Bank Group Working Paper Series*, *160*.

[8] Zhang, Z. (2014). Granger causality analysis on the economy and transportation infrastructure construction. In *ICLEM 2014: System Planning, Supply Chain Management, and Safety* (pp. 766-772). doi: https://doi.org/10.1061/9780784413753.116

[9] Pradhan, R. P., & Bagchi, T. P. (2013). Effect of transportation infrastructure on economic growth in India: the VECM approach. *Research in Transportation Economics*, *38*(1), 139-148. doi: https://doi.org/10.1016/j.retrec.2012.05.008

[10] Lombard, P. C., Sinha, K. C., & Brown, D. J. (1992). Investigation of

the relationship between highway infrastructure and economic development in Indiana. *Transportation Research Record*, (1359).

[11] U.S. Census Bureau (https://www.census.gov/construction/c30/c30index.html) accessed on September 8, 2020

[12] U.S. Bureau of Economic Analysis (https://apps.bea.gov/iTable/iTable.cfm?reqid=19&step=2#reqid=19&step=2&isuri=1&1921=survey) accessed on September 11, 2020

[13] Biecek, P., & Burzykowski, T. (2020). Explanatory model analysis: Explore, explain and examine predictive models.

[14] Mishra, V., Smyth, R., & Sharma, S. (2009). The energy-GDP nexus: evidence from a panel of Pacific Island countries. *Resource and Energy Economics*, *31*(3), 210-220. doi: https://doi.org/10.1016/j.reseneeco.2009.04.002