Forecasting Transportation Project Frequency using Multivariate Modeling and Lagged Variables

Alireza Shojaei
Building Construction Science Department
Mississippi State University
Mississippi State, MS, USA
e-mail: Shojaei@caad.msstate.edu

Hashem Izadi Moud
Construction Management Department
Florida Gulf Coast University
Fort Myers, FL, USA
hizadimoud@fgcu.edu

Ian Flood
M. E. Rinker, Sr. School of Construction Management
University of Florida
Gainesville. FL, USA
flood@ufl.edu

***Abstract*—Knowledge of the number of upcoming projects and their impact on the company plays a significant role in strategic planning for project-based companies. The current horizon of planning for companies working on public projects are the latest advertised projects for bidding, which in many cases is reported less than a year in advance. This provides a very short-term horizon for strategic project portfolio planning. In this research, a multivariate regression model with elastic net regularization and a support vector machine are used to forecast the Florida Department of Transportation's (FDOT) number of advertised projects in the future considering economic indices and other environmental factors. Two sets of analyses have been conducted, one with the current values of the independent variables and another one with up to 12 months lag of each independent variable. The results show that, of the predictors considered, the unemployment rate in the construction sector and the Brent oil price are the most significant variables in forecasting FDOT's future project frequency using current values. Also, it is evident that including lagged values of the independent variables increase the model's performance.**

Keywords-Multivariate Regression; Elastic Net Regularization; Strategic Planning; Project Portfolio Management; Forecasting; Support Vector Machine; Time Series.

I. Introduction

Construction companies, as with many other companies working in project-based industries, such as IT, are usually managing multiple projects concurrently while looking for new projects to maintain their business. The task of managing current (ongoing) projects while obtaining projects for continuous business is called Project Portfolio Management (PPM). A crucial part of the management of a portfolio is to make sure that the company resources and on-going projects are optimally balanced to ensure that not only each project meets its objectives but also the whole organization meets its overall goals. Management needs to make sure that they maximize the utilization of their resources by minimizing idle time while not accepting more work than they can complete effectively.

The majority of the literature focuses on internal uncertainties that pertain to PPM. In other words, the most explored aspect of the uncertainties in PPM is the relationships between the projects within the portfolio and the interaction between the current ongoing projects and possible future projects to measure their compatibility in terms of resource demand, and other criteria. However, environmental factors, such as economic conditions and specific industry conditions (for instance, oil price) can have a significant impact on a portfolio and a company's overall performance [1]. This study aims to integrate the environmental uncertainties and uncertainties regarding the unknown future projects, so that companies can apply this approach in their mid-term to long-term strategic planning. Martinsuo's [2] review of PPM frameworks showed that the uncertainty and continual changes in a company's portfolio has a significant negative correlation with its success. As a result, if users can reduce the extent of the uncertainties in their planning and have a more robust portfolio, it could greatly help their success. In summary, this paper proposes a regression model for forecasting the frequency of FDOT's future projects, which helps the user to estimate the number and timing of tendered projects in the future. The novelty of this approach is the consideration of environmental uncertainties in the model and the provision of quantitative insights into the unknown future.

The rest of this paper is organized as follows. Section II describes the impact of uncertainty on PPM and how unknown future projects can impact strategic planning. Section III describes the modeling approach followed in this paper. Section IV addresses the multivariate modeling of FDOT's number of projects in the future. Section V presents the conclusions and identifies future directions for the research.

II. Unknown Future Projects and Portfolio Strategic Planning

Planning is a vital factor in determining the success or failure of construction projects. According to Brown et al. [3] and World Bank [4] most construction projects worldwide do not meet their success targets, in terms of budget, duration or other determining factors, due to poor management practices. While success factors in different sectors of construction, like public, private, commercial, residential, infrastructure, differ, budget and equity remain as one of the main important factors that determine the success of any project. In the public sector, the federal government, as the sole client, forecasts the equity needed for the upcoming fiscal year in advance in order to accurately plan the number of needed projects to meet the society demands. Traditionally, governmental agencies had a short sighted view of the future budget; mainly due to the hardship in accurately estimating the budget needed based on the needs in the future. The process of planning future needs is costly, slow and uncertain. Also, it is usually based on the historical patterns of previously funded projects through earlier years. Using historical data for future prediction is useful, and more accurate when scope, duration, budget and type of future projects are known. Due to the unknown nature of future projects, including a lack of information about future projects' scope, number, and types, using historical data for projection purposes is not always accurate.

In principles of project management, the practice of batching multiple projects under one umbrella and defining target goals for them in a portfolio of a company is usually referred to as PPM. PPM is defined as "dealing with the coordination and control of multiple projects pursuing the same strategic goals and competing for the same resources, whereby managers prioritize among projects to achieve strategic benefit" [5]. Planview a leading Information Technology (IT) firm in project management also defines PPM as "Project portfolio management (PPM) refers to a process used by project managers and project management organizations (PMOs) to analyze the potential return on undertaking a project. By organizing and consolidating every piece of data regarding proposed and current projects, project portfolio managers provide forecasting and business analysis for companies looking to invest in new projects" [6]. PPM handles two important tasks including: (1) ensuring that the investment decisions by managing companies about the projects that participate in the portfolio are based on the single notion of maximizing the return on investment of the portfolio as a whole and minimizing the risks associated with participating projects [7], and (2) assuring that distribution of resources to different projects within the portfolio meets the portfolio goals in maximizing the portfolio and project goals and minimizing the risks [8]. Implementing an effective PPM process is challenging due to various factors involved in PPM. The golden key to a fruitful implementation of PPM in any construction enterprise is information. The future is unknown; thus having the necessary information that can paint a clear picture of the future is crucial in the PPM process. Existence of more accurate knowledge of future enables decision and policy makers to more accurately predict the future events, maximize the goals of the portfolio and projects as a whole and minimize the associated risks. This will result in maximizing the profit of the commercial enterprise [9].

The science and practice of project management is all about managing different kinds of uncertainties. Uncertainty could dramatically harm the success of any construction project regardless of the quality of staff, equipment, plans and drawings, and managers. In project management, uncertainty is defined as the degree of accuracy in determining future work processes, resource variation and work output [6][7][8]. Uncertainty is inherently coupled with risks. In traditional project management, risks and uncertainties have been usually discussed at the project level. However, it is believed that focusing on the totality of risks and uncertainties from a broader perspective might be beneficial to the success of any enterprise. While the Project Management Institute (PMI), one of the leading professional organizations in project management, discusses risks in a more general context of portfolio management, it does not provide any specific details, guidelines, plans, recommendations, directions or procedures on successfully managing future projects and portfolios uncertainties at a portfolio level. In fact, the whole concept of risk management is discussed very briefly by PMI. PMI limits discussions on different risk management to a few risk management techniques and methods and does not go beyond the management of risks and uncertainties at a portfolio level. PMI recommends only a few broad guidelines on detection, monitoring and handling uncertainties [9].

While PMI limits its discussion on risk management and uncertainties, from a scientific perspective, the best method to handle uncertainties and risks in any commercial enterprise is to analyze historical data to predict, model, project and mitigate potential harms of uncertainties and risks. At a scientific level, a variety of methods, techniques, and approaches have been tested to collect historic data, analyze the gathered data and find trends and tendencies in historical data that could help the project and portfolio managers understand the impacts of uncertainties of projects' success, and consequently portfolios. Artificial Intelligence (AI) is found to be a powerful tool in portfolio management [10]. A variety of algorithms have been developed in numerous research that can help to assess and to allocate risks and other types of uncertainties in project selection, execution and portfolio management [11]. Other contemporary analysis and computation techniques, such as multi-agent modeling [12], multi-objective binary programming [13], heuristic methods such as neural networks [14] and use of complicated Bayesian Network models [15] have been proposed and implemented by many scholars to study the nature of uncertainties, risks allocation patterns and process of risk allocation management at the project and/or portfolio levels. It is worth noting that the success rates of the aforementioned methods are not consistent. The success rates of implementing these methods vary based on numerous factors including the type of the project, analysis method, the number of projects in a portfolio and projects and portfolio specifications. Overall, it is still mostly impossible to provide forecast models, to perfectly plan portfolios, while considering unknown projects and environmental uncertainties and risks.

III. Modeling Approach

The literature [5][7][14] has looked at forecasting unknown future projects with a univariate modeling approach where the number of future projects is forecasted solely based on the past values of the number of projects. This study builds upon this work by forecasting unknown future projects using multivariate regression in order to incorporate environmental uncertainties in a forecast. The data used in this case study is obtained by text mining FDOT's historical project letting database. The database covers 12 years (from 2003 to 2015) containing 2816 projects. The features extracted from the database are each project letting date, cost, and duration. Table I provides a pool of candidate independent variables including macroeconomics and construction indices compiled from the literature [5][7][14], which were available at the monthly level and did not have any missing values for the explored time frame. Table I also provides the abbreviation for each variable and the sources from which they have been obtained. These factors are considered in the regression modeling as the dependent (explanatory) variables.

The integrity and continuity of the data are important as it is a time series. As a result, random cross validation was not appropriate, and a rolling forecast origin approach was adopted for cross-validation, as illustrated in Figure 1. The data were divided into two sections, training and testing. The training period starts with three years and increases by one year in each iteration while the testing period remains steady as the three consecutive years after the training set. In other words, seven models are trained, and the average error is considered as the result of cross-validation.

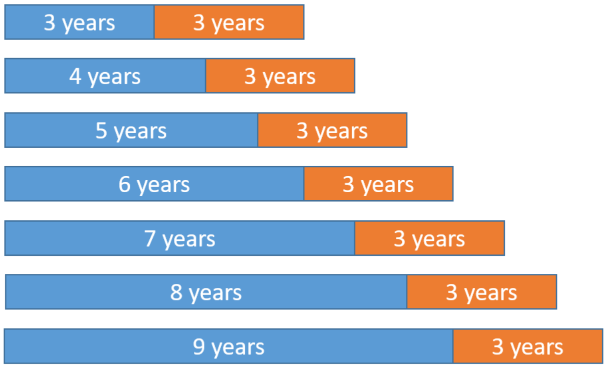

Figure 1. Forecast on a rolling origin cross-validation.

TABLE I. Candidate Independent Variables.

Variable name	Abbreviation of variable	Source
Dow Jones industrial average Vol	DJI	Yahoo Finance
Dow Jones industrial average Closing	DJIC	Yahoo Finance
Money Stock M1	MS1	Federal Reserve System
Money Stock M2	MS2	Federal Reserve System
Federal Fund Rate	FFR	Federal Reserve Systems
Average Prime Rate	APR	Federal Reserve System
Producer Price Index for All Commodities	PPIACO	U.S. Bureau of Labor Statistics
Building Permit	BP	U.S. Bureau of Census
Brent Oil Price	BOP	U.S. Energy Information Administration
Consumer Price Index	CPI	U.S. Bureau of Labor Statistics
Crude Oil Price	COP	U.S. Energy Information Administration
Unemployment Rate	UR	U.S. Bureau of Labor Statistics
Florida Employment	FE	U.S. Bureau of Labor Statistics
Florida Unemployment	FU	U.S. Bureau of Labor Statistics
Florida Unemployment Rate	FUR	U.S. Bureau of Labor Statistics
Florida Number of Employees in Construction	NFEC	U.S. Bureau of Labor Statistics
Number Housing Started	HS	U.S. Bureau of Census
Unemployment Rate Construction	URC	U.S. Bureau of Labor Statistics
Number of Employees in Construction	NEC	U.S. Bureau of Labor Statistics
Number of Job Opening in Construction	JOC	U.S. Bureau of Labor Statistics
Construction Spending	CS	U.S. Census Bureau
Total Highway and Street Spending	THSS	Federal Reserve System

A. *Exploratory data analysis*

To develop the multivariate models, a better understanding of the data characteristics was first necessary, and that information was gained through an exploratory data analysis and the identification of potentially relevant predictors.

The first exploratory analysis consisted of correlation analysis. Figure 2 provides the correlation plot of the variables. The color indicates the magnitude of the correlation, and the direction of the ellipse illustrates the direction of the relationship. Furthermore, the concentration of the ellipse tells us about the degree of the linear relationship between the variables. Project frequency is represented by "freq" in the last row and column. It appears that none of the exploratory variables had a strong linear relationship with the project frequency.

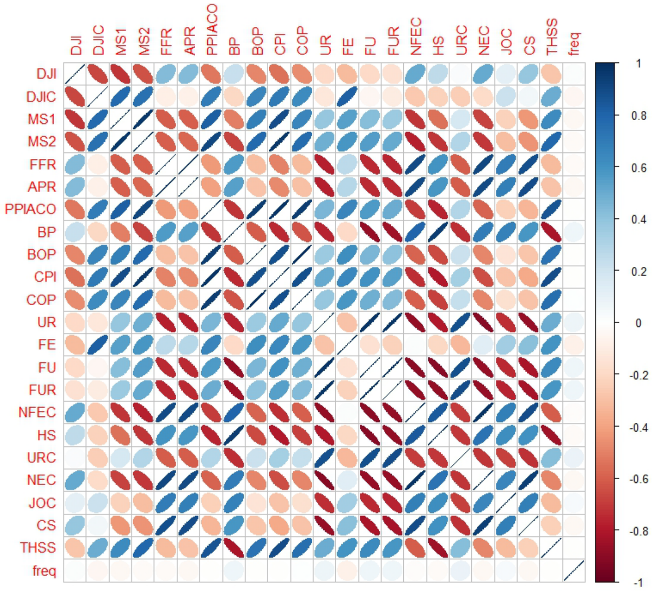

Figure 2. Correlation plot.

B. *Feature selection and feature importance*

Feature selection is the process of selecting the most relevant predictors and removing irrelevant variables from the pool of potentially useful predictors. Depending on the model's structure, feature selection can improve a model's accuracy. This process can be carried out by measuring the contribution of each variable to the model's accuracy, and then removing irrelevant and redundant variables while keeping the most useful ones. In some cases, irrelevant features can even reduce a model's accuracy. In general, there are three approaches to feature selection: the filter method, wrapper method, and embedded method.

Embedded methods implement feature selection and model tuning at the same time. In other words, these machine learning algorithms have built-in feature selection elements. Examples of embedded method implementations include LASSO and elastic net. Regularization is a process in which the user intentionally introduces bias into the training, preventing the coefficients from taking large values. This method is especially useful when the number of variables is high. In such a situation, the linear regression is not stable and in which a small change in a few variables results in a large shift in the coefficients. The LASSO approach uses L1 regularization (adding a penalty equal to the magnitude of the coefficient), while ridge regression uses L2 regularization (adding a penalty equal to the square of the magnitude of the coefficient). The elastic net uses a combination of L1 and L2. Ridge regression is effective in reducing a model's variance by minimizing the summation of the square of the residuals. The LASSO method minimizes the summation of the absolute residuals. The LASSO approach produces a sparse model that minimizes the number of coefficients with non-zero values. As a result, this approach has implicit feature selection. The generalized linear method implemented in the next section uses an elastic net. This approach incorporates both L1 and L2 regularization and thus has implicit feature selection.

Feature reduction methods, such as principal component analysis (PCA), are widely used in studies to reduce the number of independent variables. The output of such methods is a reduced set of new variables extracted from the initial variables while attempting to maintain the same information content. However, using these methods can drastically decrease the ability to interpret the significance of each input, which in itself can be very beneficial. For example, in this study knowing that oil price has a significant impact on the frequency of the projects compared to construction spending can provide valuable insight both for policy makers and contractors. As a result, the authors have chosen not to implement feature reduction methods, such as PCA.

Looking at the correlation between independent variables and the dependent variable, it became evident that a filter method using a correlation analysis was not useful, as all the variables had a nonsignificant relationship with the project frequency. As a result, an elastic net approach is used in the next section.

IV. Multivariate Modeling

The general process of model optimization and feature selection consisted of first defining a set of model parameter values to be evaluated. Then, the data was preprocessed in accordance with a 0-1 scale to make sure the high value in some variables are not skewing the model's coefficient and other variables' importance. For each parameter set, the cross-validation method discussed earlier served to train and test the model. Finally, the average performance was calculated for each parameter set to identify the optimal values for the parameters.

Ordinary linear regression is based on the underlying assumption that the model for the dependent variable has a normal error distribution. Generalized linear models are a flexible generalization of the ordinary linear regression that allows for other error distributions. In general, they can be applied to a wider variety of problems than can the ordinary linear regression approach. Generalized linear models are defined by three components: a random component, a systematic component, and a link function. The random component recognizes the dependent variable and its corresponding probability distribution. The systematic component recognizes the independent variables and their linear combination, which is called the linear predictor. The link function identifies the connection between the random and systematic components. In other words, it pinpoints how the dependent variable is related to the linear predictor of the independent variables.

Ridge regression uses an L2 penalty to limit the size of the coefficient, while LASSO regression uses an L1 penalty to increase the interpretability of the model. The elastic net uses a mix of L1 and L2 regularization, which makes it superior to the other two methods in most cases. Using a combination of L1 and L2, the elastic net can produce a sparse model with few variables selected from the independent variables. This approach is especially useful when multiple features with high correlations with each other exist.

A generalized linear model was fit to the data at the current values using the cross-validation method discussed earlier. Alpha (mixing percentage) and lambda (regularization parameter) were the tuning parameters. Alpha controls the elastic net penalty, where α=1 represents lasso regression, and α=0 represents ridge regression. Lambda controls the power of the penalty. The L2 penalty shrinks the coefficients of correlated variables, whereas the L1 penalty picks one of the correlated variables and removes the rest. Figure 3 illustrates the results of the generalized linear model (for each set of parameters 7 models according to cross-validation method is trained and the average error is assigned to the set of parameters under study), optimized by minimizing the RMSE with controlling alpha and lambda. The optimized parameters were α=1 and λ= 0.56. The authors also tested λ higher than 0.56 up to 1, however, the coefficients were not well-behaved beyond lambda=0.56.

Figure 4 depicts the LASSO coefficient curves. Each curve represents a variable. The path for each variable demonstrates its coefficient in relation to the L1 value. The coefficient paths more effectively highlight why only two variables were significant in the generalized linear model. When two variables were excluded, all other coefficients became zero at the L1 normalization, and this arrangement yielded the best performance. Figure 5 offers the variable importance for the generalized linear model with all the variables. Only the unemployment rate in the construction industry, the Brent oil price, and the unemployment rate (total) had non-zero coefficients. However, the unemployment rate (total) seemed to be relatively insignificant.

To further prune the generalized linear model, another model with only the unemployment rate in the construction sector and the Brent oil price was trained and tested. Table II contains the optimized parameters (coefficients and intercept) for the generalized linear models. The general unemployment rate had a low coefficient and, upon pruning it, the authors saw an improvement in the performance of the model. The most important variable was the unemployment rate in construction having the highest coefficient of 4.03.

Table III illustrates the performance of the optimized general linear model using a different dataset on the cross-validation sections. It was evident that excluding the unemployment rate improved the model's performance over most of the cross-validation data sections. It is notable that the pruned model performed much better in data section 1 which had the highest error and produced a more evenly distributed error among the different data sections tested. The only variables contributing to the final linear model were the unemployment rate in the construction sector and the Brent oil price.

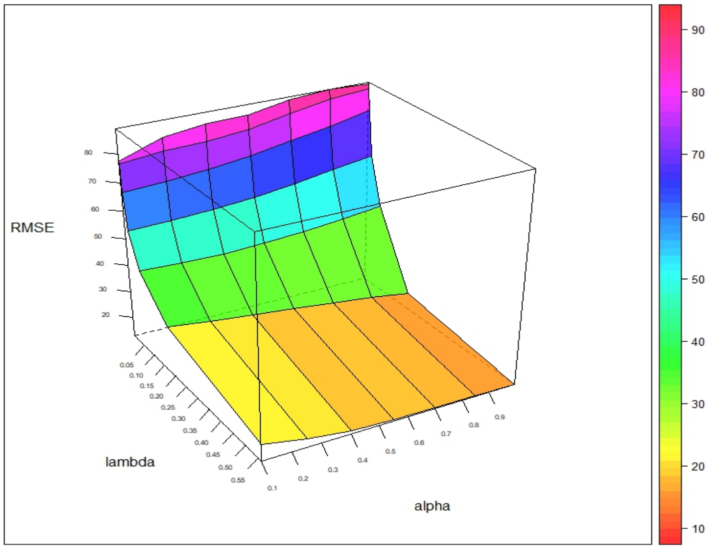

Figure 3. Generalized linear method optimization.

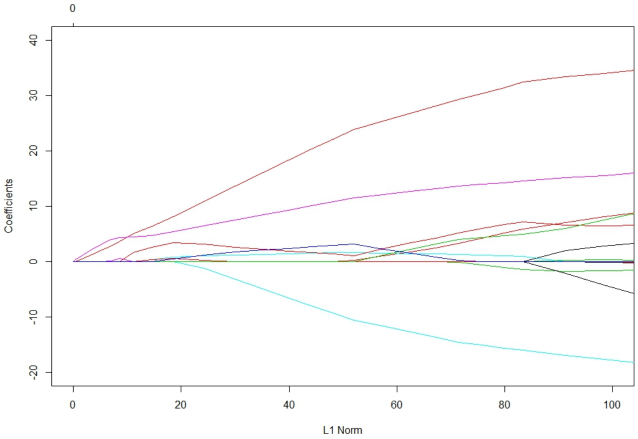

Figure 4. Lasso coefficient curve.

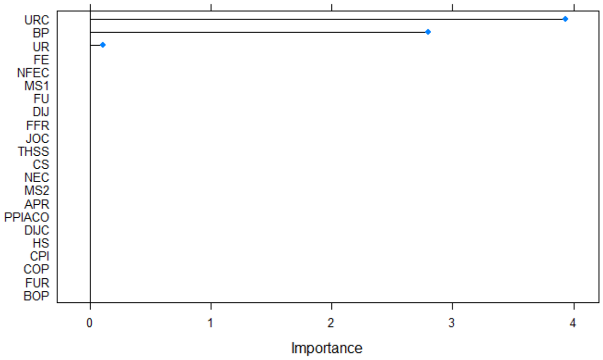

Figure 5. Variable importance of the generalized linear model.

TABLE II. PARAMETERS OF THE GENERALIZED LINEAR MODEL AT CURRENT VALUES.

Variables	Coefficients	Coefficients (Pruned by one variable)
URC	3.94	4.03
BP	2.80	2.77
UR	0.11	-----
Intercept	17.14	17.16

TABLE III. PERFORMANCE OF THE GENERALIZED LINEAR MODEL AT CURRENT VALUES.

Error term	RMSE		MAE	
Cross-validation set	**All**	**Pruned**	**All**	**Pruned**
1	16.13	9.78	13.24	10.8
2	11.58	11.94	9.64	8.56
3	13.86	13.69	11.6	8.01
4	13.16	13.14	10.82	8.25
5	12.07	10.94	9.55	10
6	11.03	10.27	8.53	8.6
7	10.89	10.87	8.6	11.28
Average	12.67	11.52	10.28	9.36

A Support Vector Machine (SVM) with a Radial Kernel were also trained and tested using the cross-validation method adopted in this study to evaluate the possible nonlinear relationship between the variables. Figure 6 depicts the results of the parameter optimization of the SVM model optimized by minimizing the RMSE with controlling sigma and C. The optimal parameters selected were sigma = 0.211 and C= 0.5.

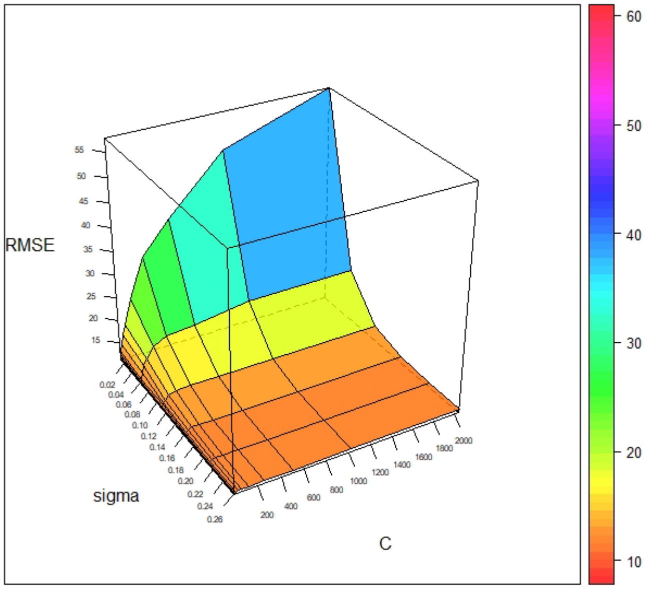

Figure 6. SVM parameter optimization.

Table IV presents the performance of the optimized SVM model on the test sets of the cross validations data sets. The results of the SVM are better than the GLM model considering all the variables at the current values.

The two GLM and SVM models so far were trained and tested on the current values of the independent variables regarding each instance of the project frequency. However, some social and economic indices might impact the dependent variable with some lag, which means that a change in the oil price might take three months to have an impact on the number of projects that FDOT is going to advertise. Figure 7 depicts the possible relationships between the variables. In the next step of this study, GLM and SVM models were trained and tested on each independent variables' current value and past 12 months values to test for both linear and nonlinear relationships between the lags of the independent variables and project frequency.

TABLE IV. PERFORMANCE OF THE SVM MODEL.

Cross-validation set	RMSE	MAE
1	10.93	8.76
2	10.31	8.25
3	9.94	7.19
4	12.06	9.63
5	11.95	9.24
6	11.11	8.38
7	10.98	8.61
Average	11.04	8.58

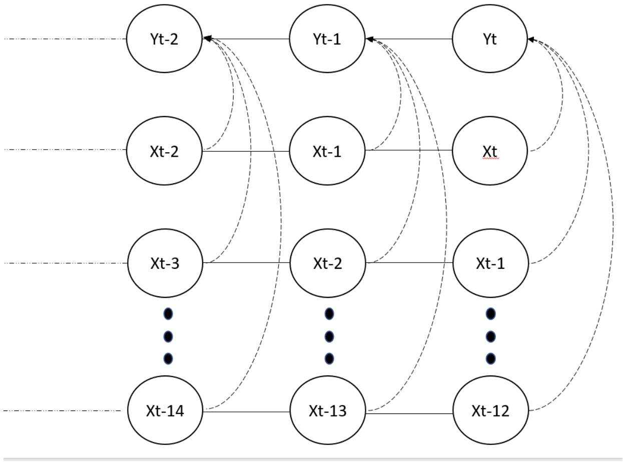

Figure 7. The possible relationship between the variables.

Figure 8 illustrates the results of the generalized linear model, optimized by minimizing the RMSE with controlling alpha and lambda over all the variables with their lagged values. The optimized parameters were α=1 and λ= 3.10. Figure 9 depicts the LASSO coefficient curves of the GLM model. Each curve represents a variable. The path for each variable demonstrates its coefficient in relation to the L1 value. The nature of the lagged value variables make them highly correlated to each other, and as a result, the L1 regularizations removes most of the variables in the process. Table V presents the results of the GLM model with the lagged variables. On the one hand, a comparison of the results with the GLM model's results including only the current values showed that including the lagged variables can increase the performance of the model. On the other hand, GLM is not friendly to variables with high correlation, and other linear models might show higher accuracy for this problem.

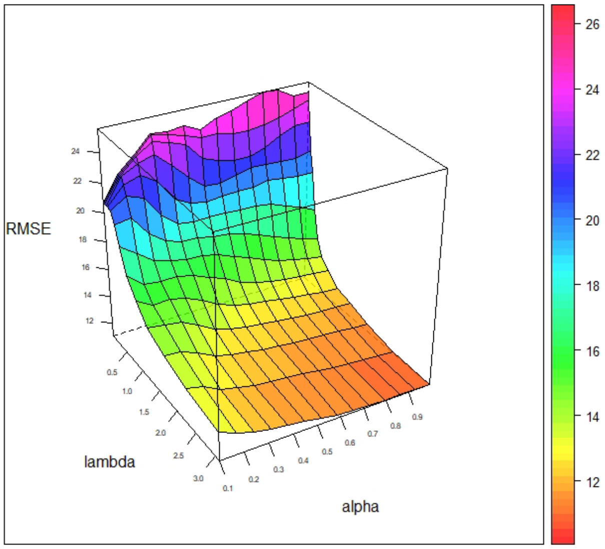

Figure 8. GLM parameter optimization with lagged variables.

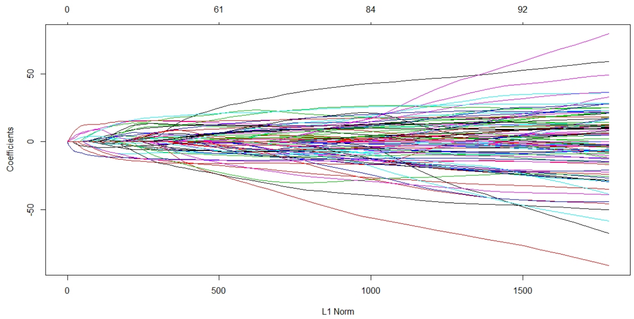

Figure 9. Lasso coefficient curves for GLM with all the variables with lagged values.

TABLE V. PERFORMANCE OF THE GLM MODEL WITH THE LAGGED VARIABLES.

Cross-validation set	RMSE	MAE
1	10.34	8.31
2	12.08	9.65
3	9.92	7.27
4	11.98	9.25
5	11.07	8.43
6	11.07	8.61
7	10.95	8.68
Average	11.06	8.60

To test the nonlinear relationship between the lagged variables and the project frequency an SVM was trained and tested using the same cross validation method. Figure 10 depicts the results of the parameter optimization of the SVM model optimized by minimizing the RMSE with controlling sigma and C. The optimal parameters selected were sigma = 0.004 and C= 0.05. Table VI presents the performance of the optimized SVM model on the test sets of the cross validations data sets. The results of the SVM are very close to the GLM with the lagged variables and SVM with only the current values. As a result, adding the lagged values increased the performance of the GLM close to the SVM model but did not increase the performance of the SVM model.

TABLE VI. PERFORMANCE OF THE SVM MODEL WITH LAGGED VARIABLES.

Cross-validation set	RMSE	MAE
1	11.06	8.22
2	10.22	7.31
3	11.90	9.36
4	11.86	9.03
5	10.99	8.71
6	10.48	8.19
7	11.19	8.61
Average	11.10	8.49

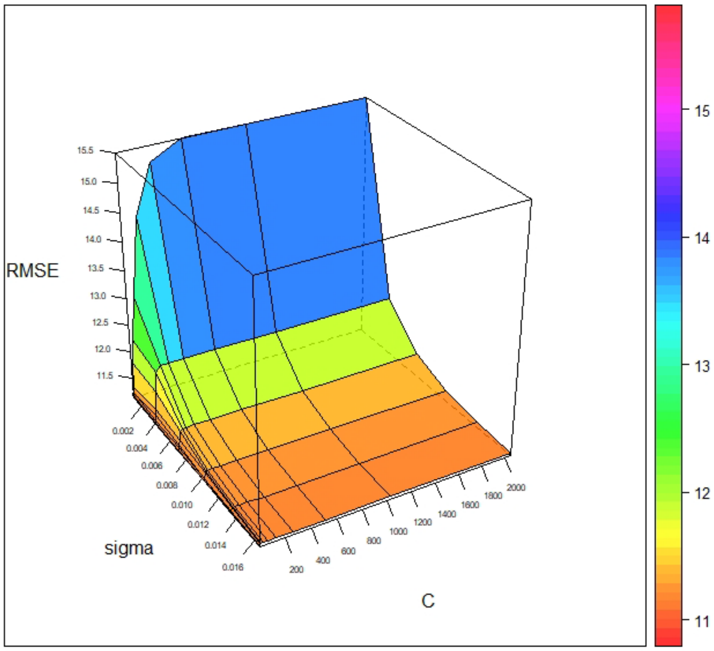

Figure 10. SVM parameter optimization with lagged variables.

Table VII provides a comparison between the multivariate models proposed in this study and some other univariate models studied previously by the authors [9]. Comparing the error terms shows that the multivariate models did not outperform some of the univariate models, such as Autoregressive Moving Average (ARMA). However, it comes close to the best performing example and it provides insight regarding the impact of environmental uncertainties on future project streams and thus could be valuable in long term strategic planning.

TABLE VII. PERFORMANCE COMPARISON OF DIFFERENT MODELS.

Model	RMSE	MAE
GLM Regression with Current variables	11.52	9.36
SVM with Current variables	11.04	8.58
ARMA(8,8)	10.71	8.45
ARMA(12,12)	11.55	9.23
AR(8)	10.92	8.48
Exponential MA (8)	11.4	9.02
GLM regression with lagged Variables	11.05	8.59
SVM with lagged Variables	11.09	8.49

It is important to note that the result of these models is the frequency of FDOT's unknown future projects, about which the user would otherwise have no information. Having reliable estimates with known error margins regarding unknown future projects can arguably provide more insight in strategic planning for a company's future compared to the current conjecture-based decision making. It should be noted that the accuracy of the models as long as the models are stable (the error is not systematic but random) is acceptable. These models are forecasting an unknown-unknown variable in the future for which there is no information available regarding their existence. However, users can use the output of this model including the error margin as inputs to their strategic planning.

The output of this research can provide a quantitative insight as a foundation for future planning. It should be noted that this model is not a standalone portfolio management framework, rather it is a supplement to existing models. For example, knowing that there is likely to be a decrease or increase in the number of projects in the future can help a company prepare in terms of consolidating or expanding its resources and assets. Furthermore, this study is limited to the FDOT's project letting database and applicability of the concept of looking into unknown-unknown projects in the future using historical data should be tested on other datasets in future work.

V. CONCLUSION AND FUTURE WORK

The importance and impact of upcoming projects on a project portfolio have been established in previously published work. However, little work has been done considering the uncertainties regarding incorporating unknown future projects in long term strategic planning. In this paper, an approach for incorporating environmental uncertainties for forecasting the number of unknown future projects is presented. Two multivariate models, generalized linear regression with elastic net regularization and support vector machine were used to forecast FDOT's unknown future projects using economic and construction indices, once with current values and once with both current and lagged values. The results indicate that the approach can reduce the impact of uncertainties on a portfolio and thus enable the development of a more robust plan with a better strategic plan. The generalized linear model with current values indicated that the best explanatory variables were the unemployment rate in the construction sector and the Brent oil price. SVM performed better than the GLM at with the current values variables and thus making a hint at the existence of the nonlinear relationship between the variables. However, adding the lagged values of the variable to the pool of the independent variables resulted in almost the same performance between the SVM and GLM. Meaning that a GLM model with lagged variables performed similarly to the SVM with the current values while adding the lagged values did not increase the performance of the SVM model. The multivariate model's performance is no better than other methods tried earlier by the authors, such as a univariate autoregressive moving average model [9] regressing on project frequency's past value. However, these multivariate models provide insight regarding the impact of environmental uncertainties on future project streams and thus could be valuable in long term strategic planning. Exploring other non-linear modeling techniques, such as neural networks for capturing more complicated relationships between the variables would be the next logical step in this research. The model developed in this study is limited to FDOT projects. However, new forecasting models specific for other databases can be built by following the same steps and adopting appropriate alternative sets of independent variables.

REFERENCES

[1] A. Shojaei, H. I. Moud, and I. Flood, "Forecasting Transportation Project Frequency using Multivariate Regression with Elastic Net Regularization Forecasting Transportation Project Frequency using Multivariate Regression with Elastic Net Regularization," in INFOCOMP 2018, The Eighth International Conference on Advanced Communications and Computation, 2018, no. July, pp. 74–79.

[2] M. Martinsuo, "Project portfolio management in practice and in context," Int. J. Proj. Manag., vol. 31, no. 6, pp. 794–803, Aug. 2013.

[3] A. Brown, J. Hinks, and J. Sneddon, "The facilities management role in new building procurement," Facilities, vol. 19, no. 3/4, pp. 119–130, 2001.

[4] World Bank, "Survey of International Construction Projec," 1996.

[5] R. G. Cooper, S. J. Edgett, and E. J. Kleinschmidt, "Portfolio management in new products: Lessons from the leaders, Part 1," Res. Technol. Manag., vol. 40, no. 5, pp. 16–28, 1997.

[6] "Project Portfolio Management Defined | Planview." [Online]. Available:

https://www.planview.com/resources/articles/project-portfolio-management-defined/. [Accessed: 01-Sep-2019].

[7] F. J. Fabozzi, H. M. Markowitz, P. N. Kolm, and F. Gupta, “Portfolio Selection,” Theory Pract. Invest. Manag. Asset Alloc. Valuation, Portf. Constr. Strateg. Second Ed., vol. 7, no. 1, pp. 45–78, Mar. 2011.

[8] L. D. Dye and J. S. Pennypacker, “Project Portfolio Management and Managing Multiple Projects-Two Sides of the Same Coin?,” in Managing Multiple Projects: Planning, Scheduling and Allocating Resources for Competitive Advantage, CRC Press, 2002, pp. 1–10.

[9] A. Shojaei and I. Flood, “Stochastic forecasting of project streams for construction project portfolio management,” Vis. Eng., vol. 5, no. 1, p. 11, 2017.

[10] R. R. Trippi and J. K. Lee, Artificial intelligence in finance & investing : state-of-the-art technologies for securities selection and portfolio management. McGraw-Hill, Inc., 1996.

[11] A. D. Henriksen and A. J. Traynor, “A practical r&d project-selection scoring tool,” IEEE Trans. Eng. Manag., vol. 46, no. 2, pp. 158–170, May 1999.

[12] J. A. Araúzo, J. Pajares, and A. Lopez-Paredes, “Simulating the dynamic scheduling of project portfolios,” Simul. Model. Pract. Theory, vol. 18, no. 10, pp. 1428–1441, Nov. 2010.

[13] A. F. Carazo, T. Gómez, J. Molina, A. G. Hernández-Díaz, F. M. Guerrero, and R. Caballero, “Solving a comprehensive model for multiobjective project portfolio selection,” Comput. Oper. Res., vol. 37, no. 4, pp. 630–639, Apr. 2010.

[14] F. Costantino, G. Di Gravio, and F. Nonino, “Project selection in project portfolio management: An artificial neural network model based on critical success factors,” Int. J. Proj. Manag., vol. 33, no. 8, pp. 1744–1754, 2015.

[15] R. Demirer, R. R. Mau, and C. Shenoy, “Bayesian Networks: A Decision Tool to Improve Portfolio Risk Analysis.,” J. Appl. Financ., vol. 16, no. 2, pp. 106–119, 2006.

[16] A. Shojaei and I. Flood, “Extending the Portfolio and Strategic Planning Horizon by Stochastic Forecasting of Unknown Future Projects,” in The Seventh International Conference on Advanced Communications and Computation, INFOCOMP 2017, 2017, no. c, pp. 64–69.

[17] A. Shojaei and I. Flood, “Stochastic Forecasting of Unknown Future Project Streams for Strategic Portfolio Planning,” in Computing in Civil Engineering 2017, 2017, pp. 280–288.