Extending the Portfolio and Strategic Planning Horizon by Stochastic Forecasting of Unknown Future Projects

Alireza Shojaei, Ian Flood
M.E. Rinker, Sr. School of Construction Management, University of Florida
Gainesville, Florida, USA.
Email: a.shojaei@ufl.edu, flood@ufl.edu

***Abstract*— Providing a practical and comprehensive methodology to facilitate management and coordination of multiple projects in a company's portfolio is a challenging task. Historically, the focus of research has been limited to the selection and prioritization of the set of known projects, current and near future. It is argued that existing portfolio planning models can be improved by adding a stochastic generator of project streams that extends the portfolio and strategic planning horizon to include future unknown projects. The study both identifies the historical factors in the market that are strong predictors of the profile of future project streams and evaluates alternative modeling approaches to the problem. The outputs from the generator are those parameters most critical to a company, namely the occurrence and letting date of a project, its expected duration, and its expected cost. A preliminary case study is presented developing, validating and testing the project stream generator for design-bid-build highway construction projects let by the Florida Department of Transportation (FDOT).**

Keywords - Project Portfolio Management; Stochastic Forecasting; Time Series Modeling; Strategic Planning; Uncertainty.

I. INTRODUCTION

Construction companies are usually involved in multiple projects at any given time. While different projects progress concurrently, they have different goals and objectives. For instance, some projects may have financial objectives while other projects' may be more focused on marketing or strategic networking. Consequently, a key managerial duty is to allocate resources (such as financial, material, and human resources) between these concurrently ongoing projects and manage their workflow together to maximize the company's performance [1]. The process of coordinating multiple projects as such is a challenging task because each incoming project affects all other ongoing projects in terms of their schedule and progress [2], and without foreseeing these effects, the consequences can be devastating. The goal of this study is to develop a stochastic project stream generator to forecast unknown future projects in order to extend the horizon of strategic planning for construction companies.

The success of a construction company is strongly impacted by its ability to strategically plan for and manage a stream of projects, many of which will overlap in time, and all of which are subject to uncertainty about their occurrence, scope and resource needs. This task can be broadly classified as Project Portfolio Management (PPM). Reference [3] describes PPM 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." Modern portfolio theory was introduced by Markowitz [4] within the finance context. McFarlan [5] introduced the concept of PPM in an information technology project management context. He suggested using projects as the elements of a portfolio (instead of investments) to better achieve an organization's objectives as well as reduce the overall risk that the organization encounters during execution of those projects.

Providing a practical and comprehensive methodology to facilitate management and coordination of multiple projects in a company's portfolio is a challenging task. There are no appropriate analytical solutions available for dynamic scheduling and resource allocation of project portfolios in real-time [2]. Existing proposed mathematical models (such as those of [6]–[9]) cannot handle the complexity of real world challenges due to a limited consideration of significant uncertainties within their models and a lack of provision for dynamic and real-time analysis. The primary focus of PPM research was initially to improve organizational performance by introducing good practices to choose and prioritize projects and ensure that the right mix of projects was adopted. A recurring theme is the alignment of the projects with the organization's strategy. There is also extensive literature on project selection with a mathematical approach. In this research, it is not proposed that developed models are incorrect. Instead, it is argued they can be advanced by adding a stochastic project generator to extend the portfolio and strategic planning horizon by stochastically forecasting the unknown future projects.

The rest of this paper is organized as follows. Section 2 provides a review of the shortcomings of existing PPM models and discusses the impact of uncertainties in PPM. Section 3 describes the project stream generator and the data used for its development. Section 4 discusses the modeling approach and results. Section 5 presents the conclusions and identifies future directions for the research.

II. PROJECT PORTFOLIO MANAGEMENT AND UNCERTAINTY

Selecting projects from available options and planning and scheduling for them have recently received a

 ISBN: 978-1-61208-567-8

considerable amount of attention [10]. For construction related organizations, such as investors, developers, and contractors, it is critical to gather and analyze project information to select the best options according to their strategic goals and schedule them within the required time frame and the financial constraints. This is a complex and multifaceted process, which has many contributing factors, such as the market condition, the organization's structure, resource availability and so on [11]. Research on this topic has come from several different points of view, such as selection model criteria and scheduling mechanisms [12], yet the primary focus has been choosing the most appropriate projects rather than providing a real-time dynamic model to address the project selection and scheduling issues [2]. Another shortcoming has been to disregard the importance of multiple project scheduling and resource allocation under influential factors and uncertainties, such as the economic situation of the construction industry and companies' organizational changes. Despite the available modeling proposals, companies still struggle to optimize and manage changes among their projects [12]. One of the reasons for this is that the proposed mathematical models cannot address the complexity of the real world situation [2]. Excluding uncertainties, such as the impact of possible upcoming projects or changes in the economic and financial situation of the construction industry, are some other noteworthy contributing factors to the poor performance of existing models.

The concept of uncertainty is very significant within the field of project portfolio management. This has led to an extensive literature on uncertainty and the ways to manage it. Duncan [13] and Daft [14] demonstrated that changes in the business environment combined with projects with high complexity always result in an increase in uncertainty in parameters, such as the number of projects, their performance, and their adherence to the project plan.

The impact of uncertainty on organizations is well established across many disciplines from psychology to economics [15]. Environmental uncertainties and their relation to organizations are analogous to the state of a person with a shortage of critical information about the environment. Scott [11] provides an example of the definition of environmental uncertainty as variability or the extent of predictability of the environment where work is executed. They also introduce some measures for uncertainty, such as variability of inputs, the number of deviations in work process, and the number of changes in the main products. In the project management context, uncertainty in a project is defined as the accuracy of predicting the variation of resource consumption, output, and work process. Uncertainty in a project can be seen as a variation from expected performance of the system under investigation.

The Project Management Institute (PMI) standard for portfolio management despite introducing the risk management concept at a portfolio level does not provide much information on how managers should handle uncertainty and risk within their portfolio. They only provide guidelines on categorizing different possible stages and processes plus naming some of the possible techniques available to handle uncertainties. The PMI only suggests monitoring risks and the performance of the project portfolio under the monitoring and control process group. The proposed framework by the PMI also includes monitoring changes in business strategy. This is an important task because when it occurs, it might result in a complete realignment of the portfolio. The mechanisms involved in this realignment are not specified other than restarting the whole PPM process from the beginning. Also, ad-hoc disturbances to the ongoing and approved project portfolios are almost entirely neglected. This oversight is not because the topic lacks interest or that authors assume a stable and predictable environment. Rather, it can probably be explained by the fact that the subject of PPM is relatively young and that the researchers and academics preferred to focus on more pressing issues in this area. For many companies, the environment is unstable, and the high level of uncertainty and unknowns resulting from the dynamic environment lead to some challenges. Upcoming projects significantly affect the performance of a project portfolio [2]. The typical approach when a new project is added to the portfolio is to update the project portfolio's plans and to try to re-optimize everything.

III. PROJECT STREAM GENERATOR

This paper presents an approach to statistically represent unknown future projects to extend the portfolio and strategic planning horizon. Forecasting a company's unknown future projects can be based on the company's past and current portfolio data, or it can use historical data from market to forecast all the upcoming projects as project streams and filter those by bidding success models. In an environment, where the supply of the projects is scarce and very competitive, using just the company's past projects to forecast the future unknown projects is potentially less accurate. Arguably it is more valid to forecast streams of unknown projects (all the available projects in the future) considering the uncertainties in the context and filter those projects by bidding success models to get the final future projects in a company's portfolio. The forecast can statistically generate a single set of outputs or stochastically produce streams of values as output. Considering the uncertainties in the market, the PPM context, and the availability of future projects, stochastic forecasting appears to be the right choice.

A preliminary study is underway developing, validating and testing a project stream generator for design-bid-build highway construction projects let by the Florida Department of Transportation (FDOT). The primary data for this study were obtained from FDOT's historical project lettings database covering 14 years (from 2003 to 2017). The last two years (2015 and 2016) data are withheld to be used as a test set for the final model. The model training and selection are based on the data from 2003 to 2015, which contains 2,816 design-bid-build project-letting reports. The outputs from the generator are those parameters most critical to a company, namely the occurrence and letting date of a project, its expected duration, and its expected cost. Other factors, such as economic condition can have an impact on the project stream. Table 1 shows a pool of candidate variables containing 24 potentially relevant predictors including the macroeconomics metrics and construction indices that were compiled from the related sources and literature [16]. The authors suggest applying a recursive feature elimination with a greedy optimization algorithm to prune down this list. This method iteratively builds models and separates best and worst variables at each iteration. This process continues until all reductions have been made. The result is the ranking of the variables based on their order of elimination.

The data should be split into three sections as a training set, a validating and model selection set, and a testing set for the final model. In the preliminary stage, the test set is the data from 2015 and 2016, and the data from 2003 to 2015 is divided to seventy percent for training and thirty percent for validating the model. However, the aim of the study is to use cross-validation to show the robustness of the model.

TABLE I. POTENTIALLY RELEVANT PREDICTORS.

CANDIDATE VARIABLES	SOURCE
GROSS DOMESTIC PRODUCTS (GDP)	U.S. Bureau of Economic Analysis
GDP IMPLICIT PRICE DEFLATOR	U.S. Bureau of Economic Analysis
INFLATION RATE	World Bank
CONSUMER PRICE INDEX	U.S. Bureau of Labor Statistics
NATIONAL HIGHWAY COST INDEX (NHCCI)	U.S. Department of Transportation
FDOT'S ANNUAL BUDGET	Florida Department of Transportation
FDOT'S PRODUCT BUDGET	Florida Department of Transportation
FEDERAL FUNDS RATE	Federal Reserve Systems
UNEMPLOYMENT RATE	U.S. Bureau of Labor Statistics
FLORIDA UNEMPLOYMENT RATE	U.S. Bureau of Labor Statistics
NUMBER OF EMPLOYEES IN CONSTRUCTION	U.S. Bureau of Labor Statistics
NUMBER OF EMPLOYEES IN CONSTRUCTION IN FL	U.S. Bureau of Labor Statistics
AVERAGE WEEKLY HOURS	U.S. Bureau of Labor Statistics
PRIME LOAN RATE	Federal Reserve System
BUILDING PERMITS	U.S. Bureau of Census
MONEY SUPPLY	Federal Reserve System
AVERAGE HOURLY EARNINGS	U.S. Bureau of Labor Statistics
EMPLOYMENT COST INDEX (ECI) CIVILIAN	U.S. Bureau of Labor Statistics
DOW JONES INDUSTRIAL AVERAGE	Yahoo Finance
CRUDE OIL PRICE	U.S. Energy Information Administration
BRENT OIL PRICE	U.S. Energy Information Administration
PRODUCER PRICE INDEX	U.S. Bureau of Labor Statistics
HOUSINGS STARTS	U.S. Bureau of Census
CONSTRUCTION SPENDING	U.S. Census Bureau

The data under study can be categorized as time series type so the integrity of the data is important and should not be tampered with by randomly dividing into different sections for validation. In this case, as shown in Figure 1, using an evaluation on a rolling forecasting origin method is advisable. This method, in general, has two variations, fixed window (Figure 1-A) where the training (orange bar) and test (blue bar) sets duration is fixed and rolls through time, or the training set (as shown in Figure 1-B) window can be extended in each trial. Using both methods can help better understand the model's performance and give more insights into the characteristics of different time spans of the data.

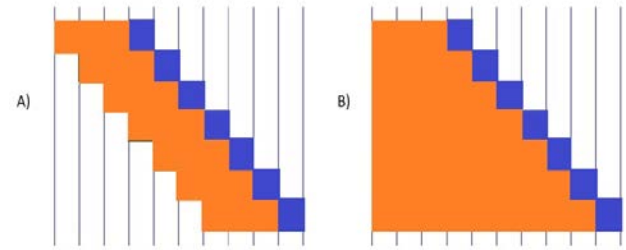

Figure 1. Evaluation of a rolling forecasting.

The sequence of generating information in the proposed model is shown in Figure 2. The first step is to forecast the number of projects (project frequency), for the desired time span, using the optimal model based on the training and validation from historical data. Next, sampling from project cost distribution takes place. At each point in time, the number of samples from the distribution is based on the number of projects forecasted in the previous step. Finally, the same process applies to the duration distribution while the possible correlation between cost and duration should be considered in the sampling process.

Figure 2. The Sequence of Generating Information.

The complete set of results from the proposed framework can be used as an input to any PPM model to consider unknown future projects in strategic planning.

IV. MODELING APPROACH

The scheme used to develop the model is shown in Figure 3. The purpose of this scheme is to look for characteristics of data, to capture them in the model's projections, and then to check to see if the model reproduces them by using the cross-validation models discussed. The univariate model was adopted as a benchmark, which the more complex multivariate models should be compared to it for improvements in forecast accuracy.

The first step is modeling the main variables through univariate modeling methods, such as Autoregressive (AR), Moving Averages (MA), Autoregressive Moving Average (ARMA), and exponential smoothing. More sophisticated approaches such as artificial neural networks can also be implemented considering the availability of the necessary data size. After establishing a benchmark, potentially relevant predictors were identified to populate a pool of candidate independent variables based on a literature review and cognitive theories. This brings in the environmental uncertainties into the forecast with the aim of improving the accuracy of the simulation. These variables are not going to have necessarily a causal relationship with the main variables; the only concern here is to be helpful in forecasting the dependent variable.

 ISBN: 978-1-61208-567-8

The next step is exploratory data analysis. It starts with a graphical comparison of the independent and dependent variables, such as scatterplots of pairs of variables. Pearson correlation, unit root (stationary or non-stationary test), Granger causality (helpful for short term forecasting), and cointegration (helpful for long term forecasting) tests are among diagnosis tests that are relevant.

The last step is to choose a set of multivariate modeling approaches based on the result of the exploratory data analysis and test whether including explanatory variables and models that are more complex can improve the accuracy of the forecast. The range of the models should test for linear and non-linear relationships based on the result of the previous step along with variable selection (pruning), parameter optimization and finding the appropriate lag between variables.

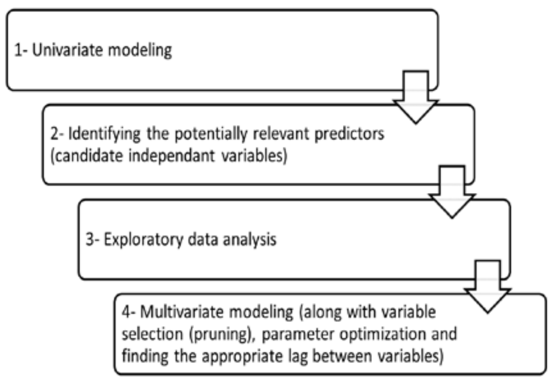

Figure 3. Model Development Scheme.

Models concerning time series data frequently involve using the value from one or more previous time steps to forecast values at the succeeding point in time; in other words, they regress based on past values. In conventional modeling, the assumption is that the independent values are known, and the dependent values are forecast. However, in multivariate time series forecasting, even the independent variables' values in the future are unknown and need to be forecast. As a result, the model contains a system of equations that forecast both independent and dependent variables in the future. This system is recursive when all the causal relationships are unidirectional and non-recursive (simultaneous) when there is reciprocal causation between variables.

Figure 4 shows four of the possible internal structures of the model. Figure 4-A shows the dependencies between the inputs and output in a univariate AR model with a lag of two. In this example, the forecast value at each point in time is based on the two preceding past values. Figure 4-B shows a recursive multivariate model where the dependent variable forecast is based on past values of itself and the independent variables. However, each independent variable is only based on its past values. Figure 4-C shows another recursive model, which differs from model 4-B in that the independent variables also act as input to each other. Figure 4-D shows a sample of a non-recursive (simultaneous) model where all the variables work as inputs for each other. There is no discrimination between dependent and independent variables in this approach.

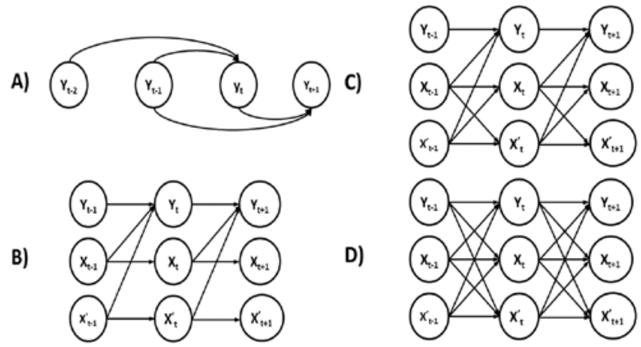

Figure 4. Possible Internal Structures of the Model.

After training and validating the model, some diagnostic tests should be conducted to check the stability of the model. For instance, checking to see if there is an autocorrelation between the residuals of the forecast is an appropriate tool for time series forecasts. Also, checking the way error compounds and undertaking a sensitivity analysis to see how the values of model parameters affect the model's output can give more insight into the performance of the model.

A. Modeling Project Frequency

Before modeling the project frequency, it is necessary to conduct some preliminary data analysis to quantify the data's characteristics. Correlogram of autocorrelation and partial autocorrelation reveals that lag 8 and 12 exceeds the significance bounds.

Testing the stationarity of the project frequency is also important. Figure 5 shows the rolling mean and standard deviation of project frequency plotted along with the actual data. It is visually plausible that the data fluctuate around a fixed mean and variance. It can be further justified by using an Augmented Dickey–Fuller test (ADF) to see if the data is stationary. There are three variations of the ADF test, all with the null hypothesis that a unit root is present in a time series sample (series is not stationary). If under any of the three variations the null hypothesis is rejected it can be inferred that the time series is stationary. The ADF test's result (the appropriate lag is chosen based on the Akaike Information Criterion (AIC)) shows that the null hypothesis can be rejected at 95 percent confidence level. Therefore, the frequency series is stationary.

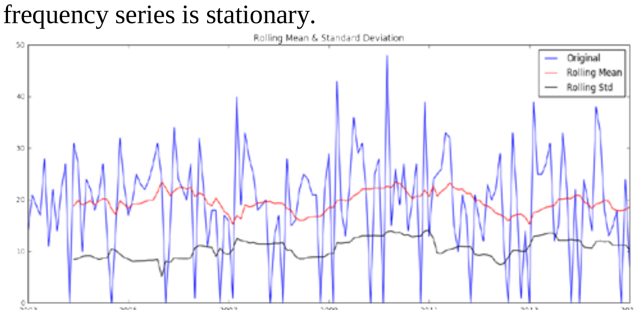

Figure 5. Rolling Mean and Standard Deviation of Project Frequency.

 ISBN: 978-1-61208-567-8

Two approaches can be implemented to forecast project frequency: univariate and multivariate modeling. Table 2 shows the summary of the best univariate models and their performance to forecast the project frequency. ARMA and exponential smoothing are among the most widely used methods to model a univariate time series. ARMA is used to model stationary time series data and is typically represented as ARMA (p,q), where, p is the autoregressive order and q is the moving average order. The order of autoregressive and moving average is selected via autocorrelation and partial autocorrelation correlograms. Based on the preliminary data analysis of project frequency an ARMA (p=8, q=8) is the best choice to model the project frequency series. Also, a set of seasonal ARMA models fitted to the data and the best model is selected via AIC. Moreover, a triple exponential smoothing (Holt-Winters) method is implemented, which takes into account both seasonal changes and trends. This analysis is conducted on two sets, the first containing seventy percent of the data as training data and the second comprising the remaining thirty percent of the data as a test set.

Table 2 presents the performance of the models on training and test sets using Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). The performance on the test set is the critical measure to compare the performance of the models. The results show that the ARMA model outperformed the other models both on the training and, more importantly, test sets.

TABLE II. SUMMARY OF THE UNIVARIATE MODELS' PERFORMANCE.

Model	Set	RMSE	MAE
Holt winter (additive seasonal)	Training	8.10	6.58
	Test	11.73	8.99
ARMA(8,8)	Training	7.93	6.28
	Test	8.82	6.84
ARIMA(0,0,1)(0,0,2)[12]	Training	9.28	7.46
	Test	10.22	8.15

Figure 6 provides a more in-depth understanding of the results by a visual illustration of the performance of the ARMA model, showing the difference between the actual data and the best performing model. The predicted values are shown in blue, and the actual data are plotted in red. Visual inspection of Figure 6 shows that the model performs better forecasting later values (after 2008) and, likewise, better captures the variance of the actual data in these later years. However, it is evident that the model's variance (blue) is less than the actual data (red) through the whole data set. The gray area represents the prediction intervals for the test data set. The dark grey shows the 80% interval and light grey shows 95% interval.

Based on the literature [16]–[18] including explanatory variables and using multivariate models can yield more accurate results. The next step in this research is to continue following the scheme illustrated in Figure 3 using multivariate methods to improve project frequency forecast.

B. Modeling Cost and Duration

Cost and duration are the two variables that are going to be sampled from a fitted distribution from past projects. Checking for the correlation between the two variables is essential. A Pearson correlation test shows 0.662 correlation coefficient with 0.000 significance between the duration and cost at the project level. This shows a moderately linear relationship between the two variables, and it should be incorporated in the model.

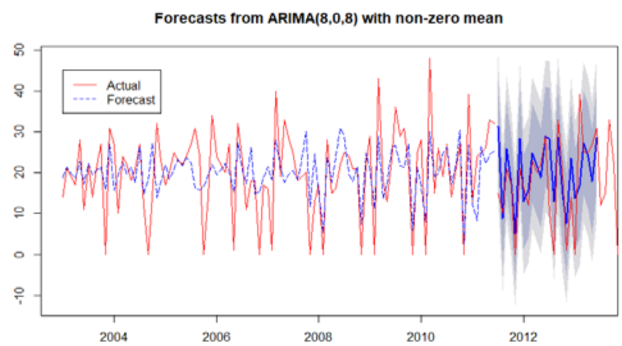

Figure 6. ARIMA (8,0,8) Forecast.

Figure 7 shows the histogram, and the corresponding fitted distribution for the duration and cost of the projects. An Inverse Gaussian distribution with μ= 244.67 and λ= 273.93 was found to provide the best fit using AIC for the duration. A lognormal distribution with (mean log) μ= 14.413319 and (standard deviation log) σ = 1.524961 was found to provide the best fit using AIC for the cost.

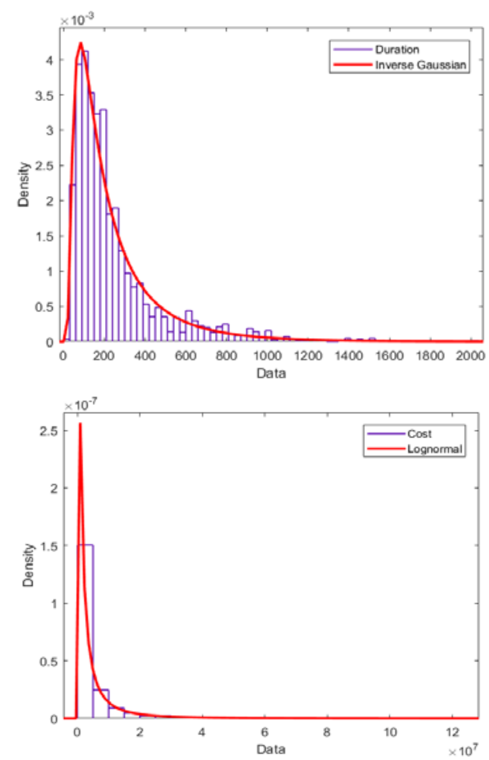

Figure 7. Duration (Up) and Cost (down) Distribution.

The performances of the various model components presented in this section indicate the viability of an integrated project stream forecaster that predicts, within a simulation environment, the frequencies of projects and empirical distributions of project duration and cost. Specifically, the generator will produce stochastic streams of unknown future FDOT projects.

V. CONCLUSION AND FUTURE WORK.

This paper has proposed an extension to the body of existing project portfolio planning models and discussed a methodology for its development. The proposed model will extend the horizon of the portfolio and strategic planning by enabling users to look more into the future and consider unknown (but statistically quantifiable) projects alongside the known and current projects in their planning process.

The proposed model is an additional component to the current portfolio management models. A general modeling approach with different possible training and validating methods is discussed and results of the preliminary research on developing, validating and testing a stream generator to forecast FDOT projects, in terms of time of occurrence, expected duration and expected cost, is presented. It is shown how univariate models can be used to forecast project frequency and the representing distributions for project cost and duration along with the relationship between these two variables is discussed.

A set of potentially relevant predictors including the macroeconomics metrics and construction indices are identified to further improve the model by using multivariate methods in future steps of the research. The next stages of this research include further implementation of the suggested modeling approach and testing using a real case study.

It is also proposed to expand the scope of the research by adding other characteristics to the project stream generator (such as different project types) and implementing it within various environmental contexts.

The complete framework will allow the user to examine different bidding and project selection strategies to see the impact on a company's portfolio and the future resource demands. Furthermore, it will lead to the selection of a closer to optimal strategy and optimal resource distribution for the future. Finally, taking into account uncertainties in future project streams might decrease the required extent of continuous adjustments to a company's portfolio plan resulting from new projects being added to the portfolio.

ACKNOWLEDGMENT

This work was funded by the University of Florida graduate student fellowships program.

REFERENCES

[1] B. S. Blichfeldt and P. Eskerod, "Project portfolio management - There's more to it than what management enacts," Int. J. Proj. Manag., vol. 26, no. 4, pp. 357–365, May 2008.

[2] 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.

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

[4] H. Markowitz, "Portfolio Selection," J. Finance, vol. 7, no. 1, pp. 77–91, Mar. 1952.

[5] W. F. McFarlan, "Portfolio approach to information systems," Harv. Bus. Rev., vol. 59, no. 5, pp. 142–150, 1981.

[6] N. Archer and F. Ghasemzadeh, "An integrated framework for project portfolio selection," Int. J. Proj. Manag., vol. 17, no. 4, pp. 207–216, Aug. 1999.

[7] T. R. Browning and A. A. Yassine, "Resource-constrained multi-project scheduling: Priority rule performance revisited," Int. J. Prod. Econ., vol. 126, no. 2, pp. 212–228, Aug. 2010.

[8] A. F. Carazo, et al., "Solving a comprehensive model for multiobjective project portfolio selection," Comput. Oper. Res., vol. 37, no. 4, pp. 630–639, Apr. 2010.

[9] M. Engwall, "No project is an island: Linking projects to history and context," Res. Policy, vol. 32, no. 5, pp. 789–808, May 2003.

[10] S. S. Liu and C. J. Wang, "Optimizing project selection and scheduling problems with time-dependent resource constraints," Autom. Constr., vol. 20, no. 8, pp. 1110–1119, Dec. 2011.

[11] W. R. Scott, "Organizations: Rational, Natural, and Open Systems," Englewood Cliffs, NJ Pren-tice-Hall.ScottOrganizations Ration. Nat. Open Syst., 2003.

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

[13] R. B. Duncan, "Characteristics of Organizational Environments and Perceived Environmental Uncertainty," Adm. Sci. Q., pp. 313–327, 1972.

[14] R. L. Daft, "Organziation Theory and Design," in South-Western Cengage Learning, 2009, pp. 138–157.

[15] Y. Petit and B. Hobbs, "Project portfolios in dynamic environments: Sources of uncertainty and sensing mechanisms," Proj. Manag. J., vol. 41, no. 4, pp. 46–58, Sep. 2010.

[16] S. M. Shahandashti and B. Ashuri, "Highway Construction Cost Forecasting Using Vector Error Correction Models," J. Manag. Eng., vol. 32, no. 2, p. 4015040, Mar. 2016.

[17] S. Thomas Ng, S. O. Cheung, R. Martin Skitmore, K. C. Lam, and L. Y. Wong, "Prediction of tender price index directional changes," Constr. Manag. Econ., vol. 18, no. 7, pp. 843–852, Oct. 2000.

[18] J. M. W. Wong and S. T. Ng, "Forecasting construction tender price index in Hong Kong using vector error correction model," Constr. Manag. Econ., vol. 28, no. 12, pp. 1255–1268, 2010.