Using Cluster Analysis to Support Commercial Assessment of Equipment Suppliers in the Early Phases of Construction Projects

Marcelo M Azambuja, Ph.D. and Xin Chen, Ph.D.
Southern Illinois University Edwardsville
Edwardsville, Illinois

The main objective of this research is to develop an effective tool to help managers in leading construction firms to select qualified suppliers of engineered equipment. Procurement of equipment is a complex process which requires the evaluation of many suppliers and multiple attributes. Decision support systems that allow managers to select a small group of suppliers from a large pool for bidding and final recommendation are still missing in practice. There is a strong need to develop computer-based tools and techniques to support procurement managers making effective strategic sourcing decisions. This paper presents a methodology to select suppliers through cluster analysis. Two classical cluster analysis techniques were applied to data of 87 suppliers collected from a leading EPC firm and generated three profiles, which are significantly different in all eight attributes. Each profile has a unique signature and managers can rapidly identify a shortlist of suppliers based on their preferences. The long term goal is to develop a real-world tested, easy-to-use system that provides support for the selection of suppliers in the early phases of construction projects.

Key Words: Procurement, Supplier Selection, Decision Support System, Cluster Analysis

Introduction

Major equipment ties up a large proportion of construction costs, has long lead times, and is usually associated with the acquisition of complex or specialized technology (CII 1999). Major equipment is engineered and fabricated specifically for the project (e.g., tanks, heat exchangers, pumps). Engineering, manufacturing and delivery of these items are very uncertain and may disrupt construction schedule. As a consequence, procurement planning for engineered equipment is critical and needs to start during the early stages of capital projects, mainly pre-project planning and conceptual design.

Procurement of long lead engineered equipment is a complex process which requires the evaluation of several suppliers as well as schedule and budget targets. This analysis is usually performed manually and time consuming. As a consequence, certain tradeoffs may be overlooked. Therefore, there is a need to develop or adapt computer-based tools and techniques from other areas (e.g. industrial engineering, computer science) to support and speed up the work of procurement managers in better understanding the tradeoffs involved in the selection of engineered equipment suppliers.

Construction firms – especially the ones building international projects – operate in a very dynamic market. In order to survive in such complex scenario, they need to identify sourcing alternatives, reduce supply risks and improve scheduling and cost performance. Knowledge of market conditions is essential to supporting early planning and supplier selection decisions; however, obtaining, storing, and analyzing volatile market data is challenging. Most firms executing industrial projects have either rich data on market forecasting and/or historical data on supplier performance. However, available data is commonly found in different paper files archived inside procurement managers' drawers or stored in firms' information systems. As a result, data is currently not electronically integrated to support analysis and decision making.

This paper describes the findings of a preliminary study of procurement information processing using cluster analysis. The authors collected actual procurement data from an EPC firm (a company that delivers engineering, procurement, and construction) and built a scenario which was used to illustrate the complexity and the various tradeoffs involved in the selection of fired heaters. The main purpose of this study is to verify whether the chosen cluster analysis technique can provide rapid and insightful information to support commercial assessment of suppliers in the early stages of construction projects. First, we present a brief literature review on decision aid methods for supplier selection in construction. Second, the authors map and describe the current supplier selection process implemented in EPC firms. From this description, the authors identify the opportunity for using cluster analysis. Finally, the findings of the analysis, conclusions, and recommendations for future research are presented.

Decision Aid Methods for Supplier Selection

Decision aid methods are very useful tools used to support managers making complicated selection decisions. There are a number of techniques and software products that have been developed in operations research area that can help solving selection problems in construction. Analytic Hierarchy Process (AHP) and Multi-attribute Utility Theory (MAUT) are examples of common techniques that have been implemented in construction to solve various types of problems (see Elmisalami et al. 2006, Shapira and Goldenberg 2005, Skibniewski and Chao 1992).

However, very little has been published on methods to aid managers making selection decisions on suppliers or supply chain configurations. Jiang et al. (2005) have shown the potential use of AHP to support supply chain selection. These authors proposed a framework based on the AHP for ranking different supply chain configurations. Bernold and Treseler (1991) presented a vendor analysis system that is based on the best-buy concept. A vendor-rating approach to secure the best buy in construction was proposed and analyzed. According to these authors, benefits of the formal vendor-analysis and rating system include the standardization of evaluation criteria, which provided consistency and transparency. The best buy approach allows users to consider and weight their preferences; however, it does not provide the capability to identify the optimal vendor based on project and procurement targets. Azambuja et al. (2009) provided a comprehensive example for the selection of pump suppliers. These authors developed a decision support system that provides an objective and rapid method for tradeoff analysis to compare supplier alternatives against project targets. They used the Aspiration Interactive Method (AIM), a technique used in the operations research area, to conduct their analysis.

The main limitation of these methods is that they are only suited for analysis involving few alternatives. When the decision involves a large set of alternatives, these techniques become too complex for practical use, if not impossible in some cases. Our procurement scenario includes around ninety suppliers and multiple attributes associated to each supplier. Our approach shows a feasible technique that can provide useful information to procurement managers when their decision involves the analysis of several suppliers located in different geographic areas.

Mapping the Selection Process of Engineered Equipment Suppliers

Three EPC firms participated in this study (EPC-1, EPC-2, and EPC-3). Their managers described a very similar – if not the same – process to select engineered equipment suppliers. Figure 1 illustrates the overall process they use to select these suppliers.

During late pre-project planning phase, firms usually prepare an approved suppliers list for the project. This list includes suppliers from both EPC and Owner individual lists. These companies look at their previous purchase orders and the current market for nontraditional suppliers which fulfill their minimum requirements. Managers commented that an initial list with 80 suppliers is not uncommon. Once the list is evaluated, EPCs issue request for inquiries to several suppliers (approximately 20 to 30 suppliers). Suppliers then send information regarding expected lead times, costs, specifications, and interest in participating in the bidding phase. After gathering all this information, EPCs prepare a short list of suppliers which will be requested to send a formal quotation for evaluation (5 to 10 suppliers). According to procurement managers, coming up with the short list of suppliers is a very complex task because of the number of alternatives available at this point of the process.

Finally, after receiving the quotes, EPCs prepare technical and commercial bid tabs for analyzing and recommending the supplier. Technical bid tab is assessed and validated by the engineering department to ensure suppliers are technically capable of producing equipment according to the specifications. The commercial bid tab, which includes data on lead times, costs, and previous performance, is prepared and analyzed together with technical recommendations by the procurement department. Typically, commercial bid tabs are presented in spreadsheet format. A final recommendation is based upon analysis of both tabs.

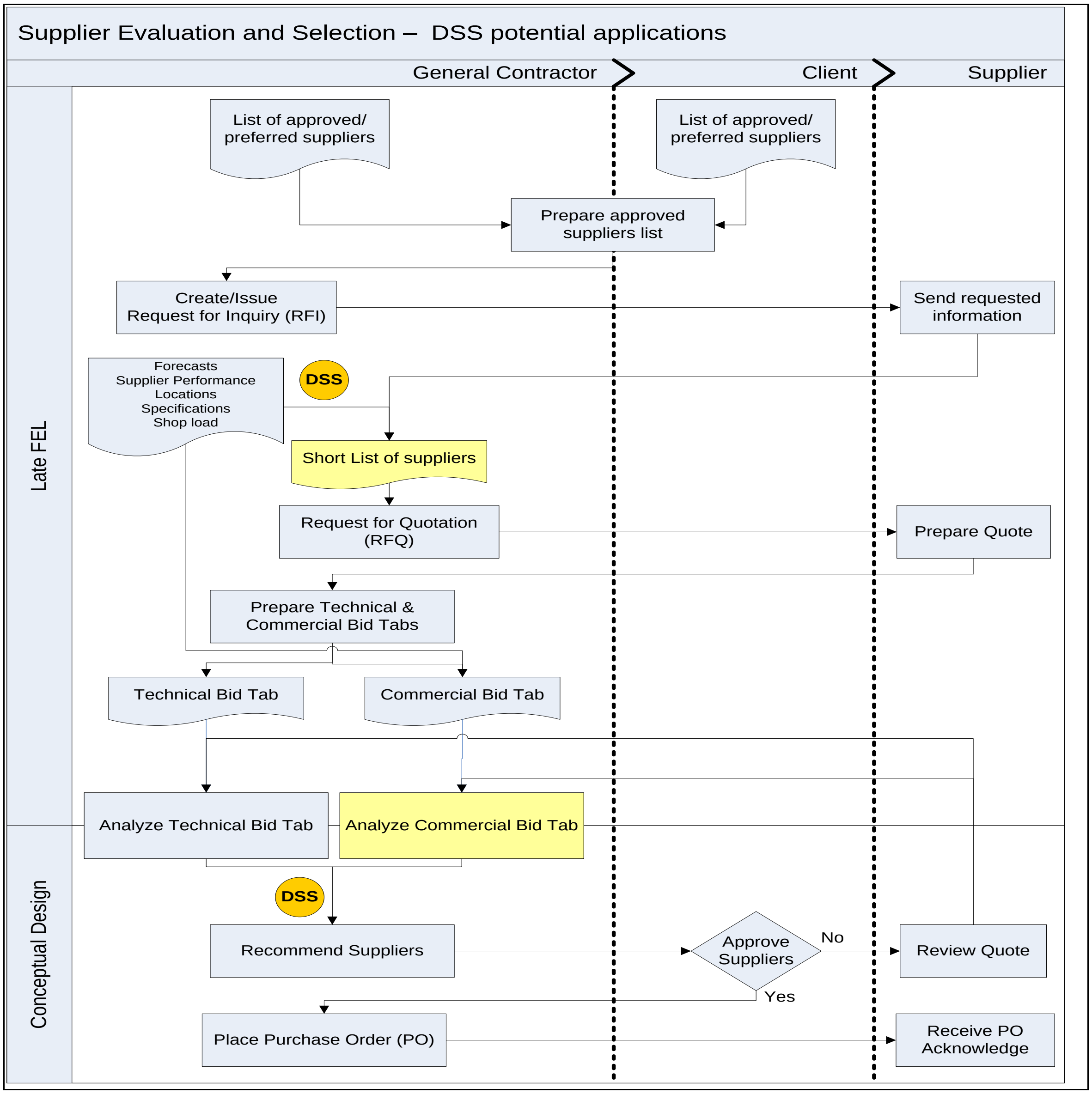

Figure 1: Supplier selection process

According to the studied firms, there are two main potential uses for computer based tools/software to support procurement decisions: 1) Help EPCs narrow down the approved suppliers list in order to come up with an appropriate short list. 2) Support commercial bid tab analysis before final supplier recommendation.
The potential application and implementation for achieving these two purposes depends upon how EPC firms organize their information. EPC-1 for example can use the system for both purposes since information is often available and organized to support analysis during the preparation of the short list. EPC-2 and EPC-3 still need to organize their supplier information – which is available in their computer systems – before considering using the tool to identify which suppliers should be shortlisted.

The authors collected data from EPC-1 which is a leading firm in the industrial construction. Managers at this firm recognized the importance of being able to conduct better analysis of their historical data for short listing suppliers. The other two firms also want to move towards this direction. The authors apply cluster analysis techniques to support the short listing purpose. Commercial data including variables such as material costs, lead times, quality, shop load utilization, and delivery performance was organized and analyzed by the authors. A detailed description of our methodology and findings is presented next.

Methodology

Cluster analysis is a statistical procedure for grouping observations with shared characteristics. The goal of cluster analysis is to derive core profiles that (a) have minimum within-cluster variation, and (b) are highly dissimilar and demonstrate minimal overlap. A combination approach using a hierarchical clustering method followed by a nonhierarchical method is often advisable (McDermott 1998).

As a commonly used hierarchical method, Ward's (Ward 1963) method with the support of simulation studies that have shown its ability to recover known data structures (Kuiper and Fisher 1975, Scheibler and Schneider 1985), was conducted to select the number of clusters and cluster centers that serve as initial cluster seeds in the following nonhierarchical procedure. The nonhierarchical method, k-means method, then clusters all observations using the seed points to provide more accurate cluster memberships.

There is no single objective procedure to determine the "correct" number of clusters. Rather alternative cluster solutions were evaluated on the following considerations to select the ideal solution:

First, statistical fit criteria, based on the rate of change in a total similarity measures as the number of clusters increases or decreases, are used to indicate the number of clusters. One criterion is that there should be an atypical decrease in overall between-cluster variance (R2) with no reverse trend at subsequent steps (Ward 1963), and the other criterion is that there should be a simultaneous elevation of the pseudo-F statistic (Calinski & Harabasz 1974) over pseudo-t2 statistic (Duda & Hart 1973) as demonstrated by Cooper and Milligan (1988).

Second, all clusters should be significantly different across the set of clustering variables. This typically involves the use of the analysis of variance (ANOVA). Cluster solutions failing to show substantial variation indicate other cluster solutions should be examined.

Third, single-member or extremely small clusters are generally not acceptable and should be eliminated. Fourth, the cluster centroid should be assessed for correspondence with the researchers' and/or practitioners' prior expectations based on theory or practical experiences. Ultimately, cluster solutions should have theoretical validity assessed through external validation. This validation can be achieved by examining differences on variables not included in the cluster analysis but for which there is a theoretical and relevant reason to expect variation across the clusters.

Results and Discussion

Table 1 shows the descriptive statistics of the eight variables used in the clustering procedure. The clustering variables were standardized to z scores with mean of 0 and standard deviation of 1 to avoid problems resulting from the use of different scale values.

A three-cluster solution was resolved from the clustering procedure (see Figure 2). The profile 1 represents a group of observations with high values of quality, transportation cost, shop floor utilization, and on-time Shipment, and low values of fabrication lead time and transit time. The Profile 2 has observations with high price, fabrication lead time, transit time, and on-time shipment, and low quality, transportation cost, shop floor utilization, and on-time fabrication. The Profile 3 has everything on average except for a high value of on-time fabrication and a low value of on-time shipment. The ANOVA results indicate that the characteristics of each cluster, i.e., profile, significantly differ on the clustering variables at the nominal level of .05 (see Table 2)

Table 1

Descriptive Statistics of the Clustering Variables (N = 87)

Variable	**Mean**	**Standard Deviation**	**Minimum**	**Maximum**
Price ($)	2,041,940.38	72225.63	1,923,955	2,162,331
Fabrication Lead Time (weeks)	68.99	5.23	61	77
Transit Time (weeks)	2.91	1.45	1	5
Quality (1~5)	3.11	1.50	1	5
Transportation Cost ($)	310202.40	118314.54	84295	511677
Shop Floor Utilization	0.80	0.03	0.75	0.85
On-Time Fabrication	0.90	0.03	0.85	0.96
On-Time Shipment	0.85	0.03	0.80	0.90

Table 2

ANOVA Results across the Eight Clustering Variables (N = 87)

Variable	**F Value (df = 2, 84)**	**p - value**
Price ($)	4.52	0.0136
Fabrication Lead Time (weeks)	7.72	0.0008
Transit Time (weeks)	18.71	<.0001
Quality (1~5)	13.36	<.0001
Transportation Cost ($)	8.65	0.0004
Shop Floor Utilization	3.14	0.0484
On-Time Fabrication	29.54	<.0001
On-Time Shipment	25.09	<.0001

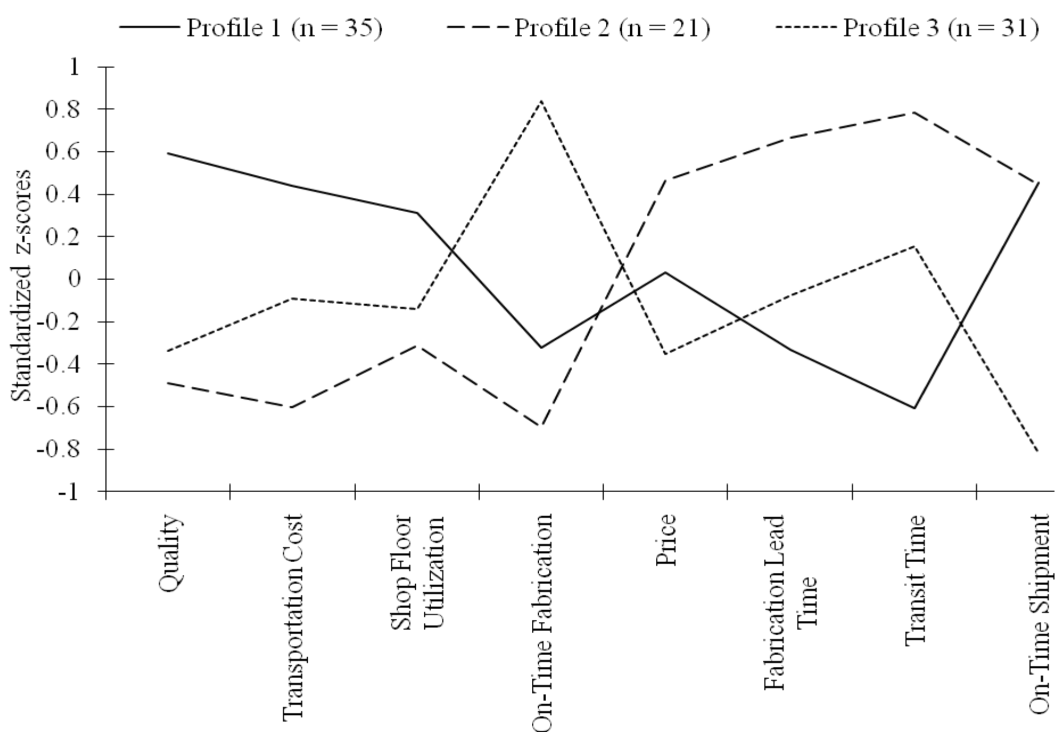

Figure 2: Three profiles of the 87 suppliers based on the eight clustering variables.

Results in Figure 2 show clearly that each of the three profiles has salient signature. For instance, an EPC would select Profile 1 if high quality is required while on-time shipment is of secondary considerations. In this case, the cluster analysis techniques effectively identified 35 potential suppliers from total 87 suppliers; 60% suppliers that do not meet the EPC's requirements were eliminated. Managers could also observe that 21 suppliers in the Profile 2 have relatively high price and low quality, even though these suppliers have high on-time shipment performance. In summary, the output illustrated in Figure 2 can effectively short list the suppliers from the large pool according to managers' preferences.

Conclusions and Future Research

This study presents an approach that can be used to assess equipment suppliers. Computer tools to support procurement decisions have been missing. The results of this research are preliminary steps to fill this gap. Cluster analysis techniques were used to demonstrate their applicability to one selection case. The results indicated that the approach can effectively help EPCs select a small group of suppliers from a large pool, i.e., shortlist qualified suppliers.

The decision support tool through cluster analysis provides a statistically sound and easy to use method to filter suppliers that meet customers' requirements. The tool developed in this research can improve the quality of procurement decisions and reduce the time managers spend carrying out commercial analysis of suppliers in the early phases of capital projects. While the technical assessment is very important and, in some cases, can guide the decision for the selection of best suppliers, the potential users of the cluster analysis outputs are experienced procurement managers who need to carry out commercial assessment of the suppliers to prepare a short list of suppliers which will participate in the bidding phase. The approach used in this study is very useful for such purpose.

Future research will focus on the validation of cluster analysis techniques and results. Additional data will be collected and same cluster analysis techniques are applied to observe if similar profiles can be produced. On the other hand, external variables, for example managers' knowledge and experience selecting suppliers, can be identified to examine if there is theoretical and relevant reasons to validate the profiles obtained from cluster analysis.

References

Azambuja, M.M., O'Brien,W.J., & Rangarajan, S. (2009). A Decision Support Tool for Rapid Assessment and Selection of Engineered Equipment Suppliers. *Proceedings of 2009 ASCE International Workshop on Computing in Civil Engineering*, Austin - TX.

Bernold, L.E., & Treseler, J.F. (1991). Vendor analysis for best buy in construction. *Journal of Construction Engineering and Management*, 117, 645–658.

Calinski, T. & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics*, **3**, pp. 1-27.

Construction Industry Institute (1999). Procurement and materials management: a guide to effective project execution. *Implementation Resource 7-3 Report, CII,* Austin, 566 pp.

Cooper, M. C., & Milligan, G. W.(1988). The effect of error on determining the number of clusters. *Proceedings of the International Workshop on Data Analysis, Decision Support, and Expert Knowledge Representation in Marketing and Related Areas of Research*, Berlin: Springer-Verlag, pp. 319-328.

Duda, R. O., & Hart, P. E. (1973). *Pattern classification and scene analysis*. New York: John Wiley & Sons, Inc.

Elmisalami, T., Walters, R., & Jaselskis, E.J. (2006). Construction IT decision making using multiattribute utility theory for use in a laboratory information management system. *Journal of Construction Engineering and Management*, 132, 1275-1283.

Jiang, A., Issa, R.R., & Cox, R.F. (2005). A Decision Support Model on Cost and Activity-Based Performance Measurement in Steel Construction. *Proceedings of the 2005 ASCE International Conference on Computing in Civil Engineering*, Cancun, Mexico, July 12–15, 12 pp.

Kuiper, R., & Fisher, L. (1975). A Monte Carlo comparison of six clustering procedures. *Biometrics*, **31**, pp. 777-783.

McDermott, P. A. (1998) Megacluster analytic strategy for multistage hierarchical grouping with relocations and replications. *Educational & Psychological Measurement*, **58**, pp. 677-686.

Scheibler, D., & Schneider, W. (1985). Monte Carlo tests of the accuracy of cluster analysis algorithms: A comparison of hierarchical and nonhierarchical methods. *Multivariate Behavioral Research*, **20**, pp. 283-304.

Shapira, A., & Goldenberg, M. (2005). AHP-based equipment selection model for construction projects. *Journal of Construction Engineering and Management*, 131, 1263-1273.

Skibniewski, M., & Chao, L.C. (1992). Evaluation of advanced construction technology with AHP method. *Journal of Construction Engineering and Management*, 118, 577–593.

Ward, J. H., Jr. (1963). Hierarchical grouping to optimize an objective function. *American Statistical Association Journal*, **58**, pp. 236-244.