
SEM for cost estimation in manufacturing: an injection mould case study

Minh-Phuoc Doan^{*1} and Iragaël Joly²

¹University of Bordeaux, CNRS, IMS, UMR 5218, 33405 Talence, France – University of Bordeaux, CNRS, IMS, UMR 5218, 33405 Talence, France – France

²Univ. Grenoble Alpes, CNRS, INRAE, Grenoble INP, GAEL, Grenoble, France – Univ. Grenoble Alpes, CNRS, INRA, Grenoble INP, GAEL, 38000 Grenoble, France – France

Résumé

Estimating the total cost is a challenge for product quotations (1). The final cost is made up of a number of sub-costs that are interdependent. Design and R&D costs vary with the complexity of the product, which in turn affects manufacturing techniques and manufacturing costs. Consequently, the prediction of the total cost should take into account the interdependencies between its components ("subcosts" for example: design and R&D, logistics, and sub-contracting costs). Linear regression is a classic and simple machine learning tool used in data-driven methods of cost estimation. Usual models estimate the total cost assuming a linear equation in which the random terms are assumed to be independent. Nevertheless, such independence assumption can lead to unreliable estimates and predictions (2) when components of the total cost (as the sub-costs) can be interdependent. Multiple regression techniques, like structural equations models (SEM) (3), permit estimating a dedicated equations system including possible covariance between random terms between equations: one equation for the total cost and one equation for each cost component. To the best of our knowledge, there is no application of the SEM method to cost estimation in manufacturing. A second challenge is to identify and select relevant features to be included in the regression model. Expert knowledge gives insight into the variables set to consider. Automatic iterative methods select the features based on their evaluated impact on a performance indicator such as the coefficient of determination (R^2), Comparative Fit Index (CFI), Root Mean Square Error of Approximation (RMSEA), Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) (3), (4).

In this paper, we propose and evaluate several SEM-based machine-learning cost estimation models corresponding to the combinations of the SEM method and several common feature selection methods, including expert selection and three machine-learning selection models (Pearson correlation selection, AIC selection, and BIC selection). Numerical experiments, performed on a case study of an injection mould company, showed first, that the models based on the combination of SEM and AIC feature selection methods give the best solution according to classical performance metrics like RMSEA, CFI, and AIC. Second, the prediction accuracy is different for each of the six considered components of the total cost. Then, some of the features are statistically important for all the cost components while some others are only important for one or several components. After that, the cost drivers for each cost component are often coherent with the quotation experts' vision. However, in some

^{*}Intervenant

cases, it is even strange for the experts to discover these relationships. Finally, the models seem to perform better when the mould cost is lower. This might be due to the lack of observations for high-cost moulds.

References

- (1) A. Niazi, J. S. Dai, S. Balabani, and L. Seneviratne, "Product cost estimation: Technique classification and methodology review," *J. Manuf. Sci. Eng.*, vol. 128, no. 2, pp. 563–575, 2006, doi: 10.1115/1.2137750.
- (2) W. H. Greene, *Econometric Analysis*, vol. 8th edition. Pearson Education, 2018. doi: 10.1007/978-3-030-56239-7_5.
- (3) Y. Fan *et al.*, "Applications of structural equation modeling (SEM) in ecological studies: an updated review," *Ecol. Process.*, 2016, doi: 10.1186/s13717-016-0063-3.
- (4) R. . Hoyle, *Handbook fo Structural Equation Modelling*. The Guilford Press, 2023.