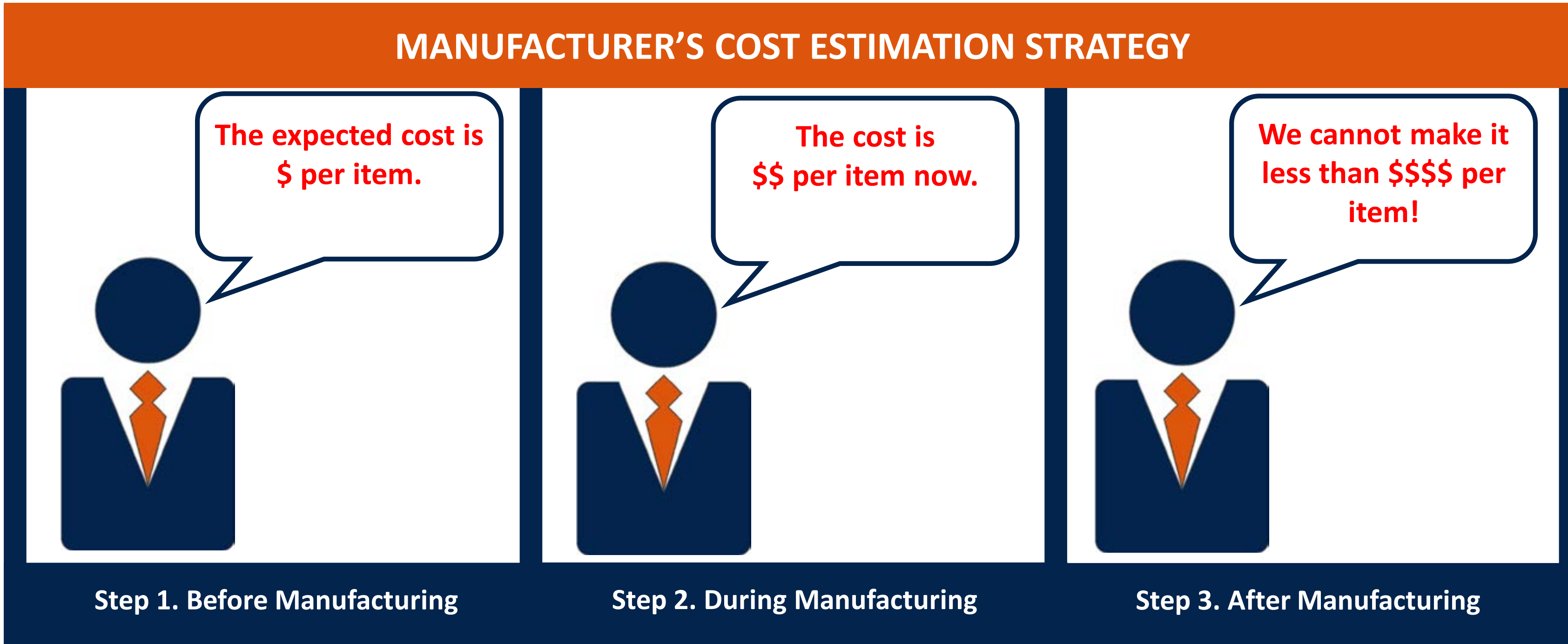
Manufacturing Cost Prediction in the Presence of Categorical and Numeric Design Attributes

By
Dr. Eren Sakinc and Dr. Alice E. Smith
Industrial and Systems Engineering, Auburn University

8th Annual SERC Sponsor Research Review
November 17, 2016
20 F Street NW Conference Center
20 F Street, NW
Washington, DC

www.sercuarc.org

Introduction



Critical Questions:

- What is the negotiation power over the underlying product price? Is the expected cost accurate?
- Is it possible to know the cost of a new and unique design before it is actually manufactured?

Introduction

- When manufacturing a new unique design, the focal point is to establish a price which maximizes customer value while being profitable.
- Since an irreversible and large amount of capital is tied up in production elements, estimating manufacturing costs accurately is critical.
- Final decisions about the product price should be based on analytical approaches, instead of intuitive expectations.

Introduction

- “Cost plus pricing” or “Cost based pricing”

- Poorly established product prices that are a function of product cost may cause two unfavorable consequences:
  - (1) A potential loss of profit due to the gap between the expected cost and the actual cost
  - (2) A loss of customers and goodwill due to higher prices than necessary

Design Attributes (Cost Drivers)

- We need to know the cost structure of a product which consists of a collection of **cost drivers**.
- A **cost driver** is defined as any factor which changes the cost of an activity (according to Chartered Institute of Management Accountants – CIMA).
- From a statistical perspective, **cost drivers** are **explanatory variables** that have a contribution to the manufacturing cost of products.

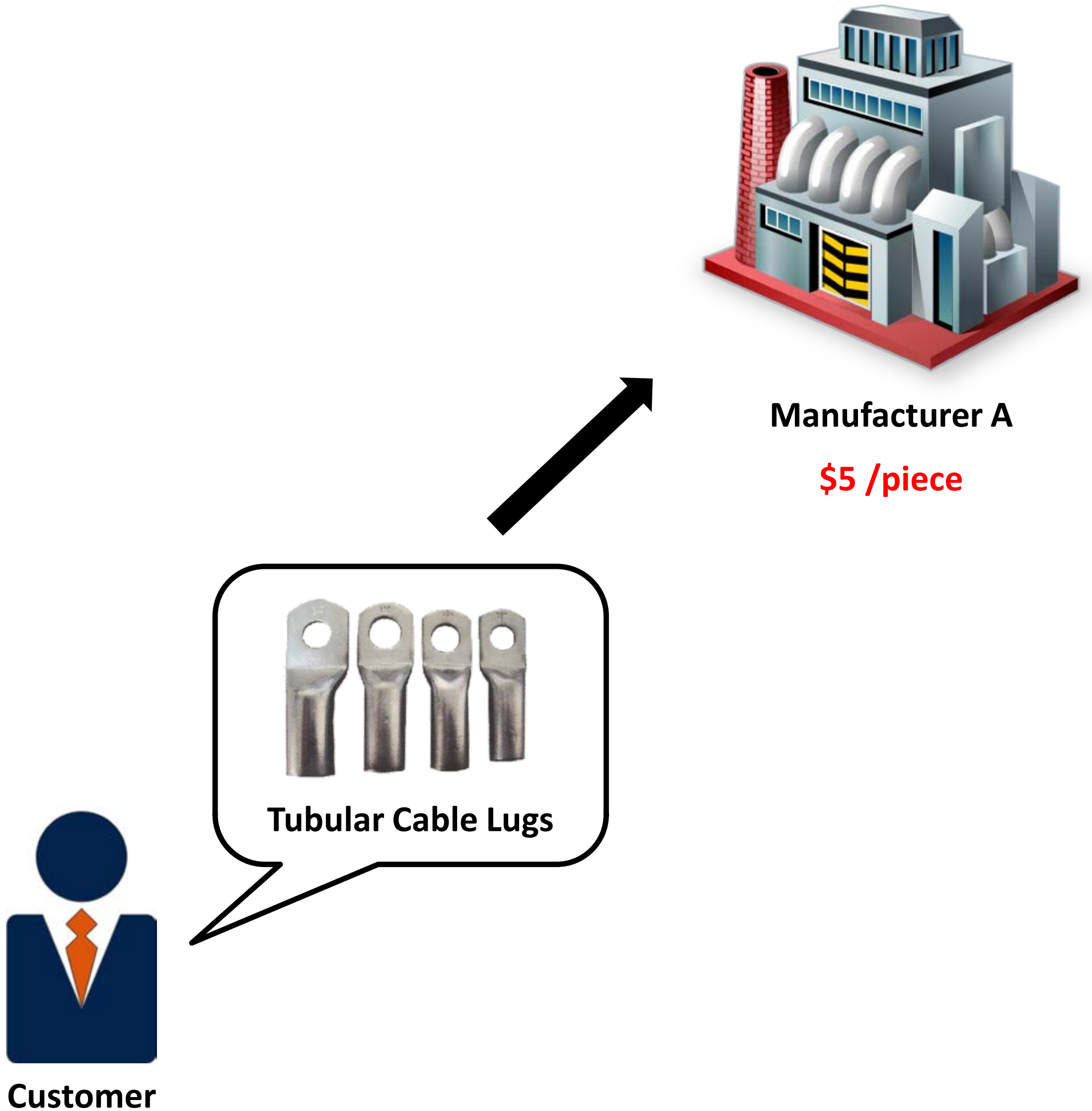
cost drivers = { cost variables, design variables, design attributes, variables, attributes }

Type of Variables

- Categorical (Qualitative / Discrete) Variables
  - Nominal
  - Ordinal
  - Binary – Symmetric and Asymmetric Binary

- Numeric (Quantitative / Continuous) Variables
  - Interval Scaled
  - Ratio Scaled

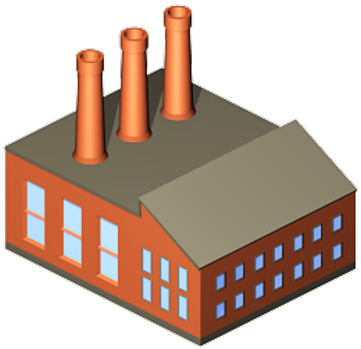
Cost Estimation Challenge vs. Competitive Pricing



Cost Estimation Challenge vs. Competitive Pricing

Manufacturer A

$5 /piece



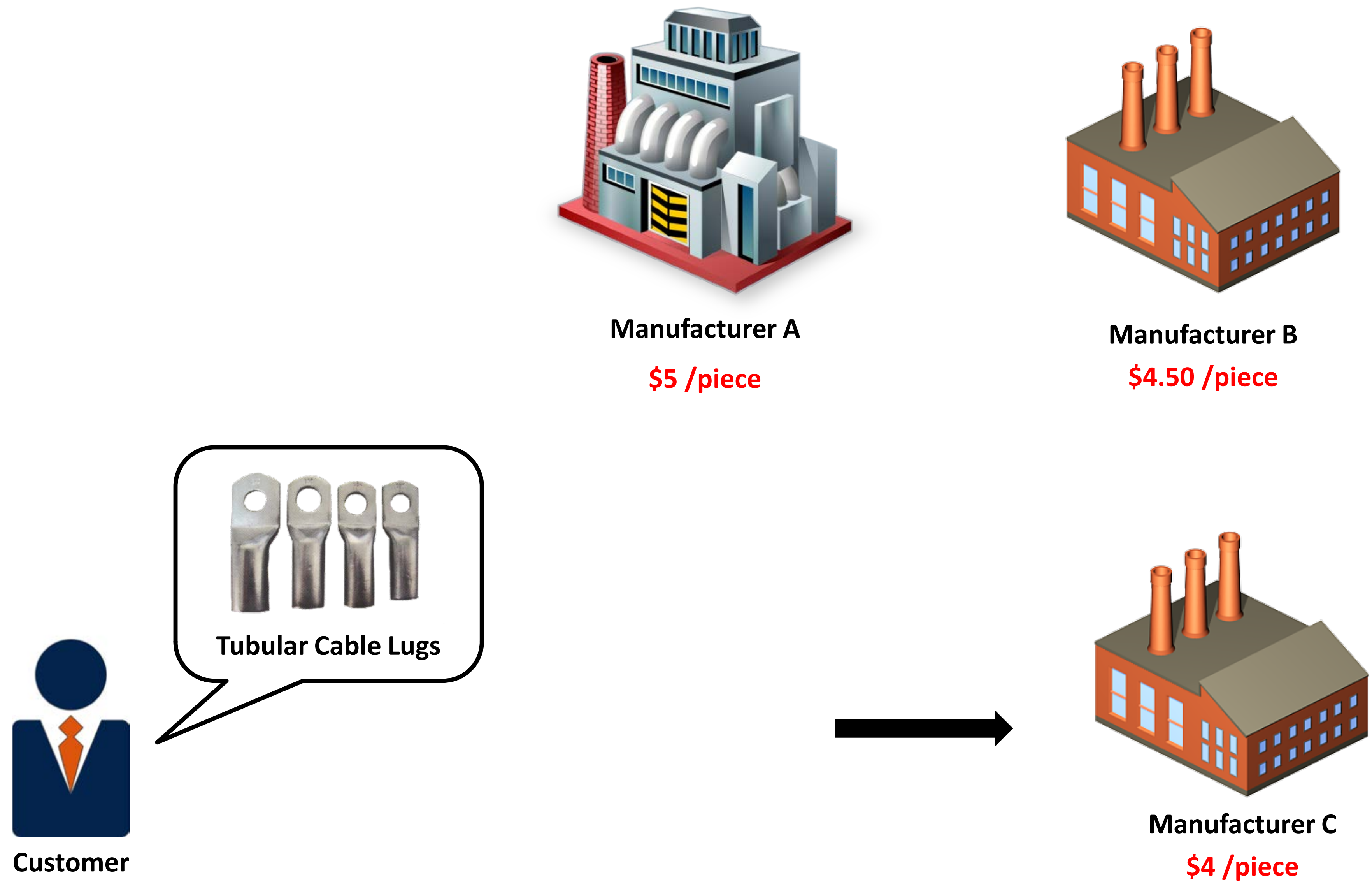
Manufacturer B

$4.50 /piece

Tubular Cable Lugs

Customer

Cost Estimation Challenge vs. Competitive Pricing



Alternative Approaches

Category	Group	Subgroup	Technique	Key Advantages	Limitations
QUALITATIVE TECHNIQUES	Intuitive		Case-Based	Innovative design approach	Dependence on past cases
QUALITATIVE TECHNIQUES	Intuitive	Decision Support Systems	Rule-Based	Can provide optimized results	Time-consuming
QUALITATIVE TECHNIQUES	Intuitive	Decision Support Systems	Fuzzy Logic	Handles uncertainty, reliable estimates	Estimating complex features costs is tedious
QUALITATIVE TECHNIQUES	Intuitive	Decision Support Systems	Expert Systems	Quicker, more consistent and accurate results	Complex programming required
QUALITATIVE TECHNIQUES	Analogical		Regression Analysis Model	Simpler method	Limited to resolve linearity issues
QUALITATIVE TECHNIQUES	Analogical		Back Propagation Neural Networks	Deal with uncertain and non-linear problems	Completely data-dependent, higher establishment cost
QUANTITATIVE TECHNIQUES	Parametric			Utilize cost drivers effectively	Ineffective when cost drivers cannot be identified
QUANTITATIVE TECHNIQUES	Analytical		Operation-Based	For optimized results, alternative process plans can be evaluated	Time-consuming, require detailed design and process planning data
QUANTITATIVE TECHNIQUES	Analytical		Break-Down	Easier method	Detailed cost information required about the resources consumed
QUANTITATIVE TECHNIQUES	Analytical		Cost Tolerance	Cost effective design tolerances can be identified	Require detailed design information
QUANTITATIVE TECHNIQUES	Analytical		Feature-Based	Features with higher costs can be identified	Difficult to identify costs for small and complex features
QUANTITATIVE TECHNIQUES	Analytical		Activity-Based	Easy and effective method using unit activity costs	Require lead-times in the early design stages

J. S. Dai, A. Niazi, S. Balabani and L. Seneviratne, "Product cost estimation: Technique classification and methodology review," *Journal of Manufacturing Science and Engineering,* vol. 128, no. 2, pp. 563-575, 2006.

Alternative Approaches

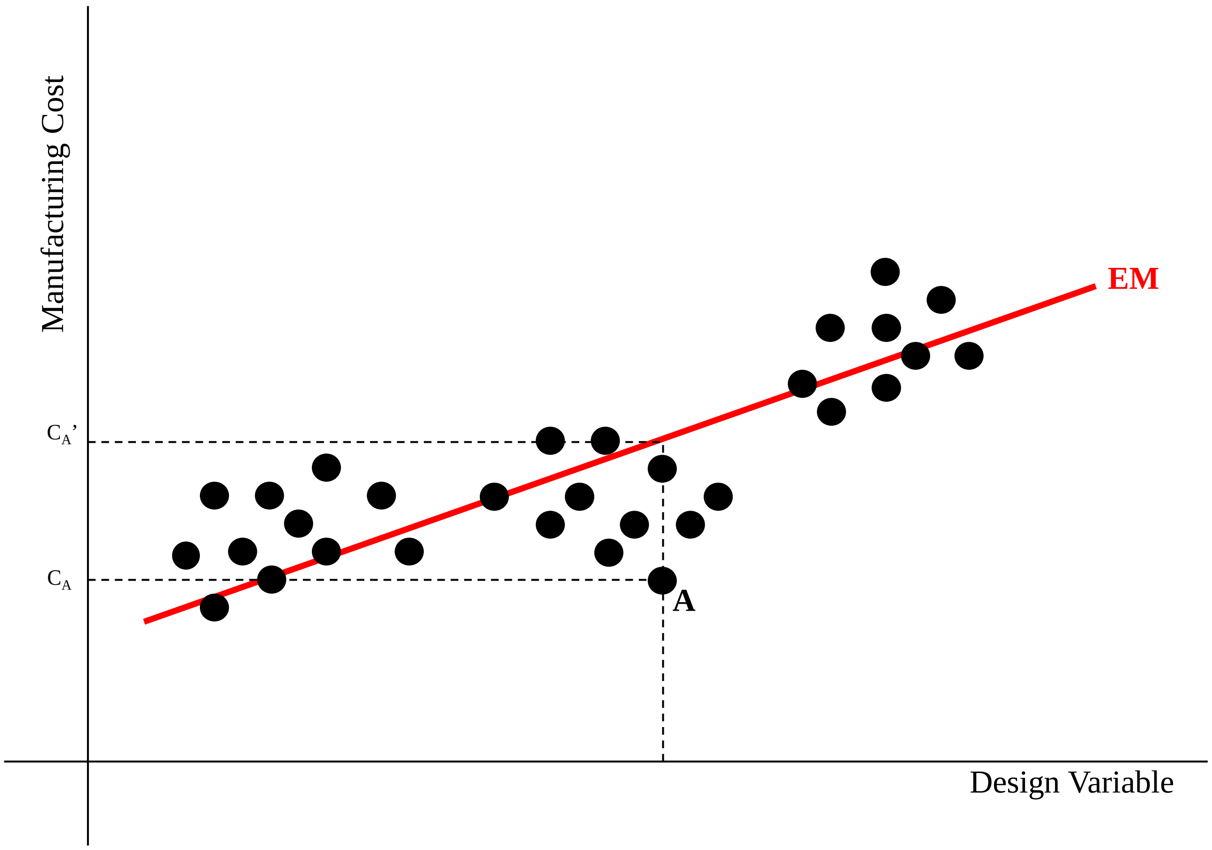
Category	Type	Subtype	Technique	Key Advantages	Limitations
QUALITATITIVE TECHNIQUES	Intuitive		Case-Based	Innovative design approach	Dependence on past cases
QUALITATITIVE TECHNIQUES	Intuitive	Decision Support Systems	Rule-Based	Can provide optimized results	Time-consuming
QUALITATITIVE TECHNIQUES	Intuitive	Decision Support Systems	Fuzzy Logic	Handles uncertainty, reliable estimates	Estimating complex features costs is tedious
QUALITATITIVE TECHNIQUES	Intuitive	Decision Support Systems	Expert Systems	Quicker, more consistent and accurate results	Complex programming required
QUALITATITIVE TECHNIQUES	Analogical		Regression Analysis Model	Simpler method	Limited to resolve linearity issues
QUALITATITIVE TECHNIQUES	Analogical		Back Propagation Neural Networks	Deal with uncertain and non-linear problems	Completely data-dependent, higher establishment cost
QUANTITATIVE TECHNIQUES	Parametric			Utilize cost drivers effectively	Ineffective when cost drivers cannot be identified
QUANTITATIVE TECHNIQUES	Analytical		Operation-Based	For optimized results, alternative process plans can be evaluated	Time-consuming, require detailed design and process planning data
QUANTITATIVE TECHNIQUES	Analytical		Break-Down	Easier method	Detailed cost information required about the resources consumed
QUANTITATIVE TECHNIQUES	Analytical		Cost Tolerance	Cost effective design tolerances can be identified	Require detailed design information
QUANTITATIVE TECHNIQUES	Analytical		Feature-Based	Features with higher costs can be identified	Difficult to identify costs for small and complex features
QUANTITATIVE TECHNIQUES	Analytical		Activity-Based	Easy and effective method using unit activity costs	Require lead-times in the early design stages

J. S. Dai, A. Niazi, S. Balabani and L. Seneviratne, "Product cost estimation: Technique classification and methodology review," *Journal of Manufacturing Science and Engineering,* vol. 128, no. 2, pp. 563-575, 2006.

Parametric / Non-parametric Simulation

- Monte Carlo Simulation
  - —Parametric distribution assignments to cost drivers
  - —Assignments are usually arbitrary
- Efron’s Non-parametric Bootstrapping
  - —Empirical distributions
  - —No benchmark comparison for validity

Industry Practice



At most companies use linear regression models but more often rely on intuition and other ad hoc approaches.

Our Approach

- We would like to predict the manufacturing cost of a product quickly and accurately.
- We investigate ways of using clustering methods to predict the manufacturing cost of products in the presence of complex numeric and categorical design attributes.
- The accuracy of the methodology is assessed in comparison to a traditional approach, a polynomial regression model in absence of a clustering approach.

Motivations

- Many cases, costs are estimated based on primitive intuitive approaches that are far from reality and accuracy.
- Making parametrical distribution assumptions for design attributes can be arbitrary.
- Over a diverse product family, establishing only a single accurate estimation model is challenging and doubtful.

Objectives

- To accurately and quickly estimate the cost of a particular product before it is manufactured
- To deploy clustering techniques to achieve improved accuracy in the prediction
- To find appropriate number of clusters for a given case and series of products

Contributions

- First to introduce a manufacturing costs estimation approach for mixed categorical and numeric variables using clustering methods

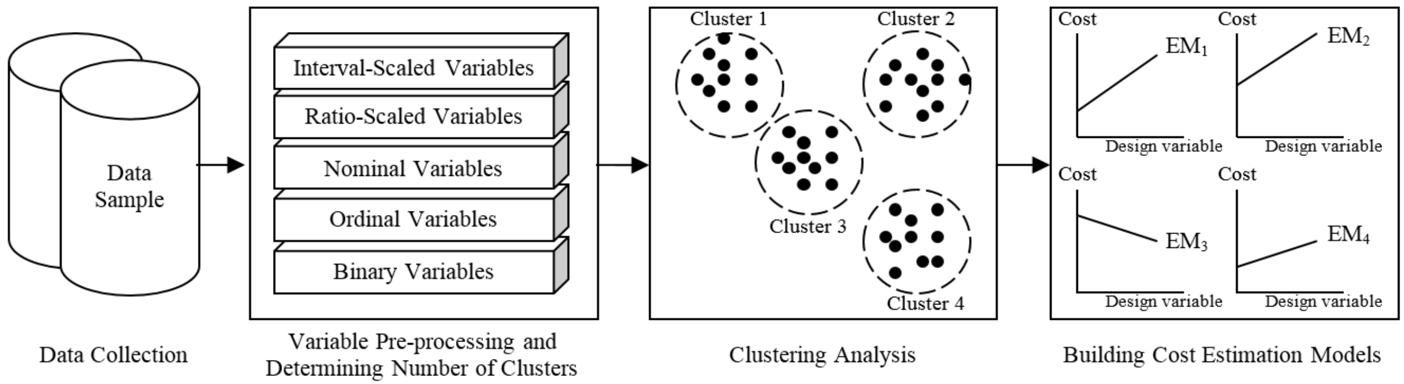
- Implemented a simple heuristic to determine the appropriate number of clusters when there is no prior knowledge about the number of product groups

Assumptions and Limitations

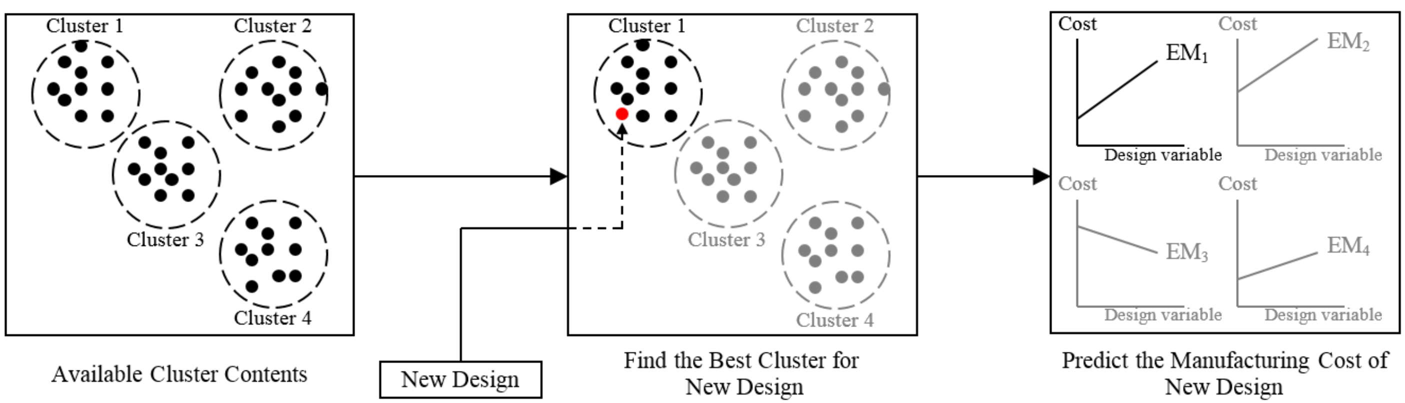
- New products are based on some modifications or variations to existing or historical products
- The clustering contents are not necessarily optimized due to using a clustering heuristic
- Limited to non-parametrical approaches to avoid making assumptions concerning statistical distributions. We assume that all variables come from empirical distributions.
- We assume commodity production where the size of a batch is not important.

Suggested Methodology

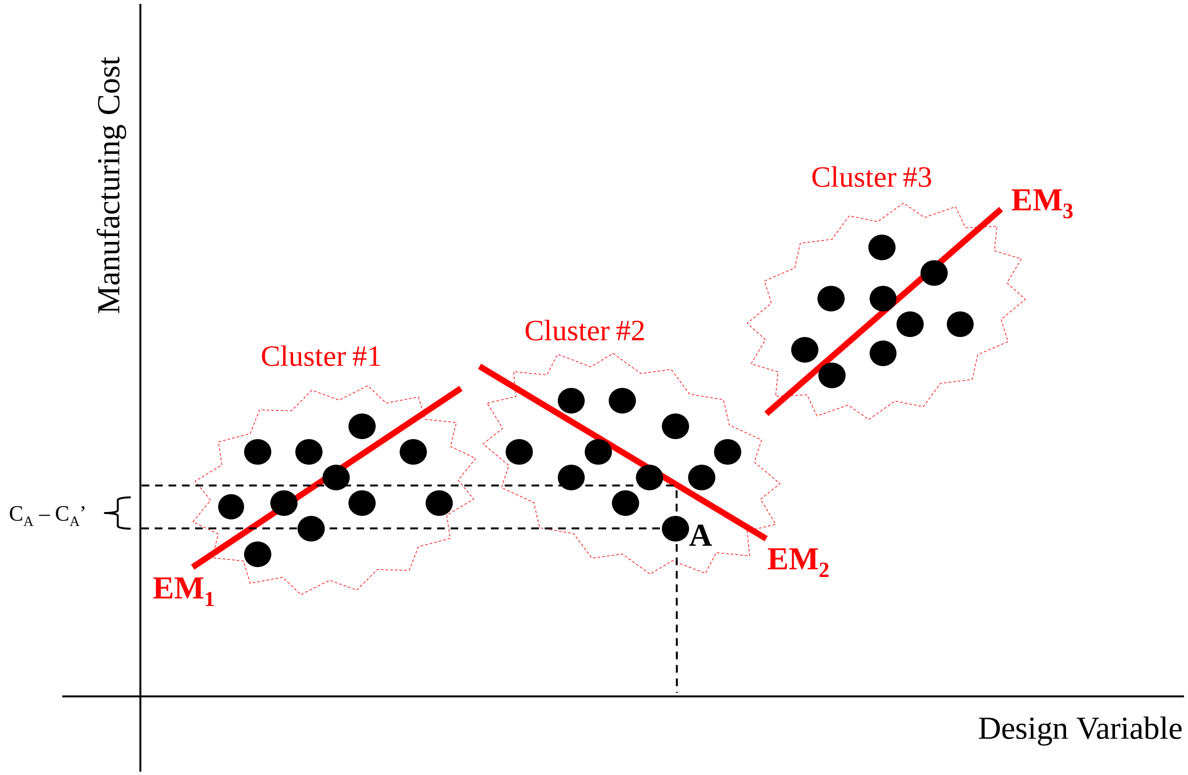
1



2



Suggested Methodology

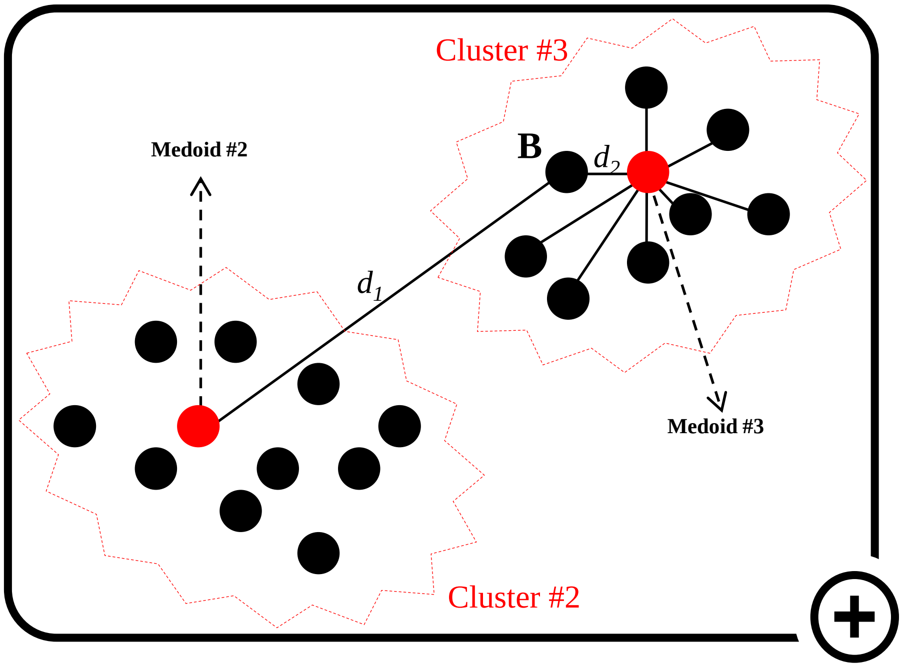


Choice of Clustering Algorithm

- k-means
  - —Squared error based
  - —Limited to continuous variables only
  - —Result is dependable on the initial random solution

- k-prototypes: Modified k-means
  - —Frequency and Squared error based
  - —Euclidean distance and simple matching coefficient
  - —Weighting factor is arbitrary
  - —Combining a quadratic expression with a linear expression

- k-medoids
  - —Operates on a dissimilarity matrix
  - —No randomness: Initial solution (BUILD), Moves (SWAP)
  - —Handles outliers



Choice of Distance Metric

	Consider Correlations	Handle Numeric Data	Handle Categorical Data	Handle Mixed Data	Non-negativity Requirement	Scale for Elliptical Data	Scale for Range	Modifiable Weight	Sensitive to Outliers	Unitless Measure	Compatibility to Our Work
Euclidean Distance		+							+		
Scaled Euclidean Distance		+				+			+	+	
Minkowski Metric		+						+	+		
Mahalanobis Distance	+	+				+				+	
Canberra Metric		+			+		+			+	
Czekanowski Coefficient		+			+		+			+	
Chebychev Distance		+									
Pearson Correlation	+	+				+				+	
Cosine Similarity		+				+	+				
Similarity Coefficients			+				+			+	
Gower’s Dissimilarity Index		+	+	+			+			+	+

Choice of Number of Clusters

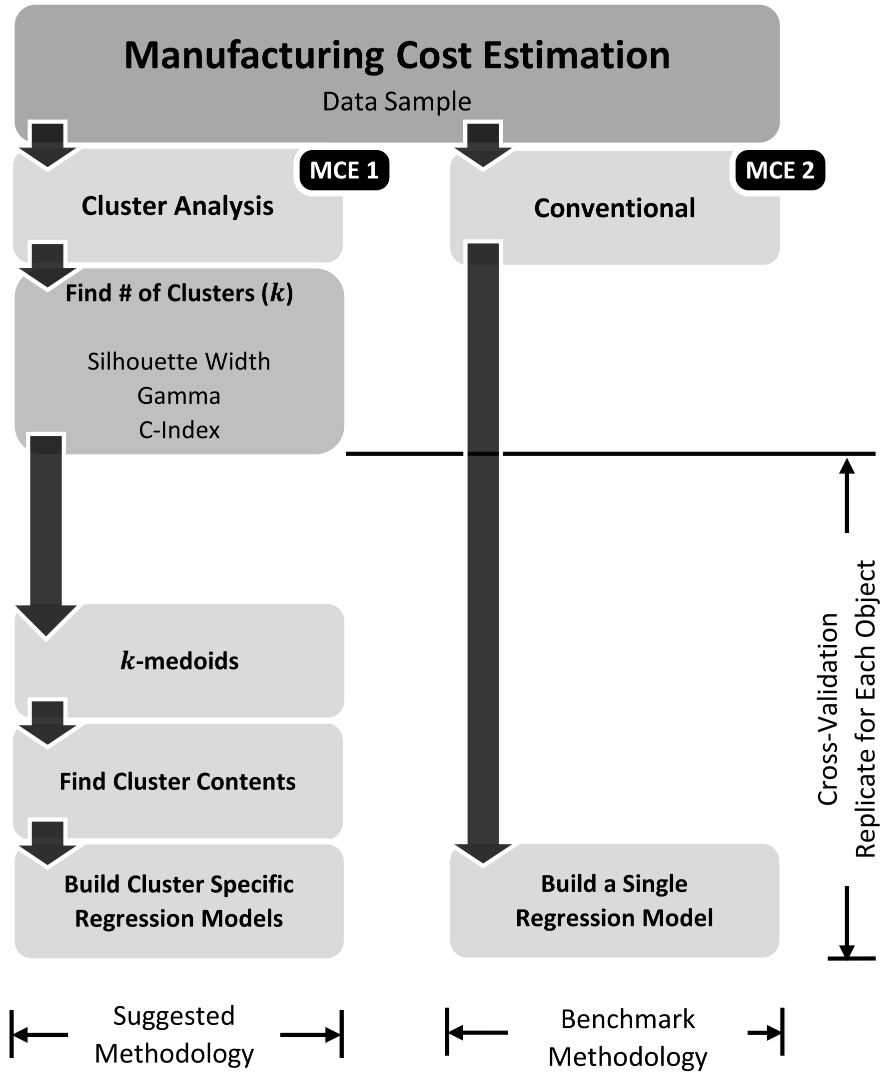
- Top 6 performing indices (Milligan and Cooper):
  - —Calinski and Harabasz's PSF
  - —Duda and Hart's $J_e(2)/J_e(1)$ or PST2
  - —*Dalrymple-Alford's C-index
  - —*Baker and Hubert's Gamma
  - —Beale's F-ratio
  - —Sarle's CCC
- *Rousseeuw's average silhouette width
- Consensus among Gamma (local peaks), silhouette width (local peaks & > 0.5), C-index (local troughs)

Choice of Predictive Model

The most complicated practice in the industry is regression models.

- **Regression Models**
- **Splines**
- Neural Networks
- Kriging

Summary of the Methodology



Performance Metrics

$$ARE_i = \left| \frac{(Actual\ Cost)_i - (Estimated\ Cost)_i}{(Actual\ Cost)_i} \right| \tag{1}$$

$$MARE = \frac{1}{n} \sum_{i=1}^{n} ARE_i \tag{2}$$

$$SE_i = [(Actual\ Cost)_i - (Estimated\ Cost)_i]^2 \tag{3}$$

$$MSE = \frac{1}{n} \sum_{i=1}^{n} SE_i \tag{4}$$

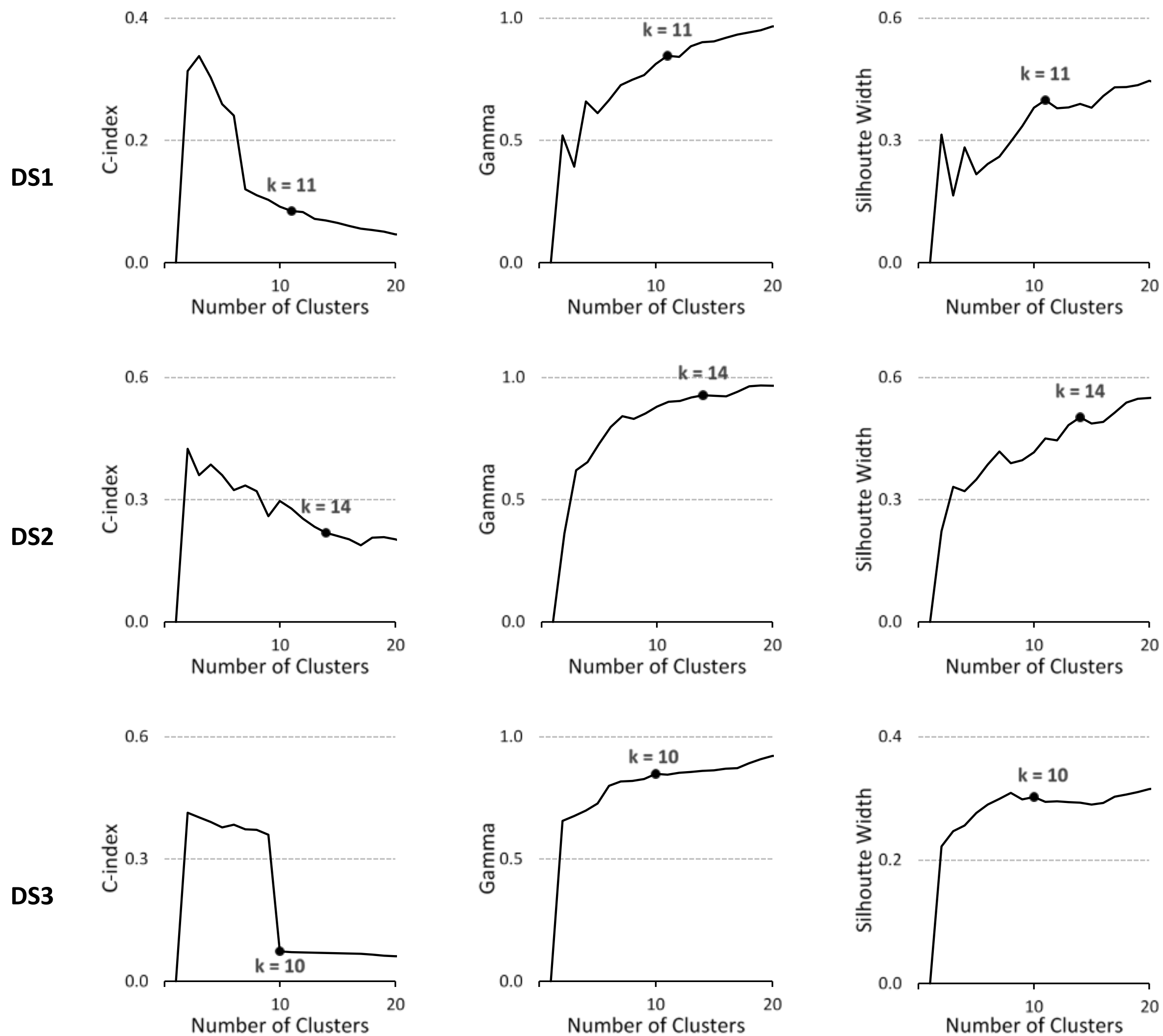
$$RMSE = \sqrt{MSE} \tag{5}$$

Real World Applications

- Electromagnetic and lightening protection parts manufacturer
  - **DS1** Tubular cable lugs: 12 variables
  - **DS2** Air rods: 10 variables

- Plastic kitchen and household products manufacturer
  - **DS3** Plastic parts: 51 variables

Determining the Number of Clusters



Results – Performance Metrics

MARE

	MCE 1	MCE 2
DS 1	**4.98%**	49.82%
DS 2	**5.81%**	15.42%
DS 3	**12.39%**	33.83%

Min ARE

	MCE 1	MCE 2
DS 1	0.00%	0.00%
DS 2	0.00%	0.00%
DS 3	0.00%	0.00%

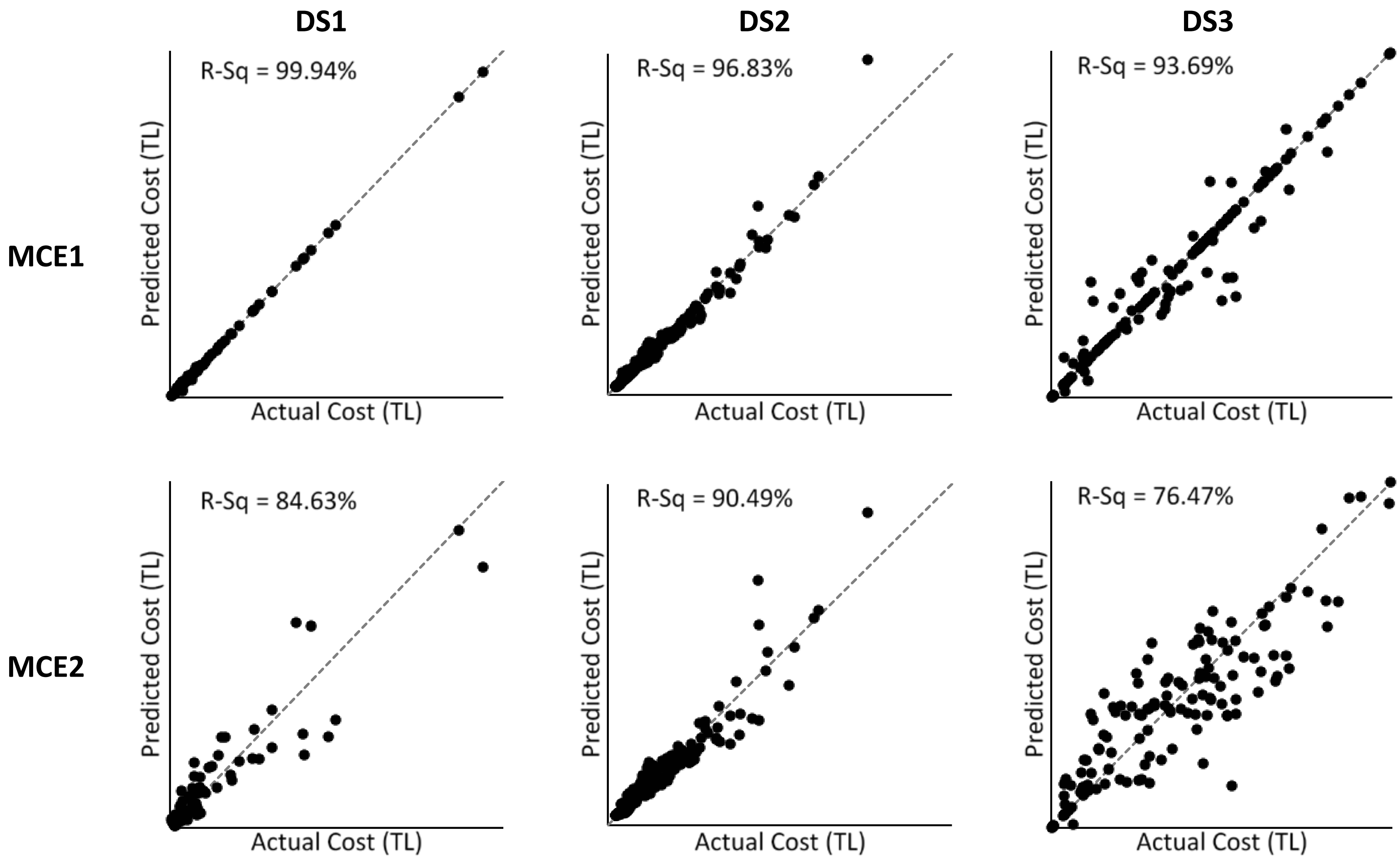
RMSE

	MCE 1	MCE 2
DS 1	**8.86%**	140.26%
DS 2	**355.72%**	615.92%
DS 3	**17.71%**	34.20%

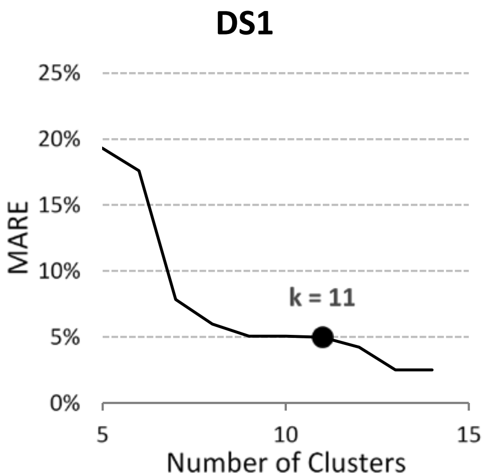
Max ARE

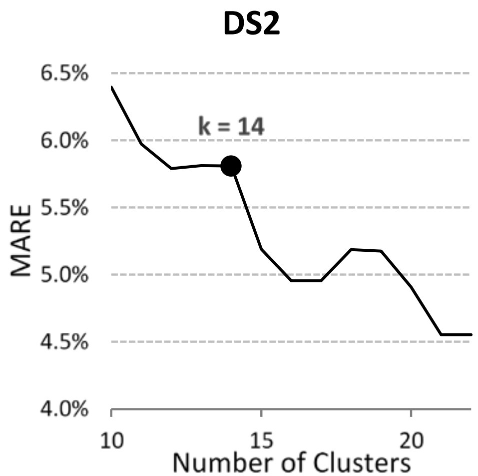
	MCE 1	MCE 2
DS 1	**46.67%**	429.52%
DS 2	**56.04%**	64.36%
DS 3	**203.54%**	233.79%

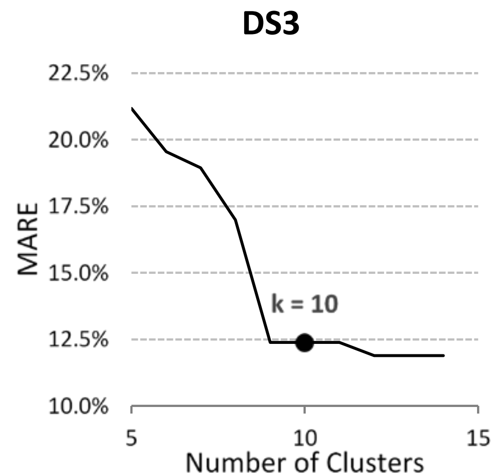
Results – Model Fit



Results – Sensitivity of Number of Clusters







Results – Sensitivity to Polynomial Model

	DS1		DS2		DS3	
	MCE1Q	MCE2Q	MCE1Q	MCE2Q	MCE1Q	MCE2Q
MARE	4.37%	43.22%	2.31%	12.22%	9.69%	36.68%
Min ARE	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Max ARE	81.12%	440.98%	23.26%	44.41%	250.10%	399.26%
MSE	0.42%	93.49%	188.45%	1537.48%	1.34%	6.80%
RMSE	6.50%	96.69%	137.28%	392.11%	11.60%	26.08%

Conclusions I

- We investigated ways of using clustering methods to predict the manufacturing cost of a product without actually manufacturing it.
- The accuracy of the methodology is assessed in comparison to a simple regression model with the absence of clustering approaches.
- The main concern is to predict the manufacturing cost of a product without dealing with arbitrary assignments of statistical distributions to cost related attributes.

Conclusions II

- In real production systems often a variety of products are being manufactured under a single facility roof.
- Over a diverse product family, establishing only a simple accurate estimation model is challenging and even questionable.
- This motivated us grouping products according to their design features, common manufacturing operations or some other factors by dividing the whole database of products into neighborhoods.
- Then for each group of products (clusters), a cost estimation model is developed to predict the manufacturing cost of a new product with using the cluster specific model.

Direction of Future Research

- Developing a comprehensive similarity measure that demonstrates high inter-cluster variability while being able to handle mixed categorical and numeric design attributes.
- A deterministic model such as a mixed integer programming model can be implemented to obtain the optimal cluster results.
- Information gain criterion can be considered when deciding on the inclusion of a candidate predictor (design attribute) in the cost estimation model.

Manufacturing Cost Prediction in the Presence of Categorical and Numeric Design Attributes

By
Dr. Eren Sakinc and Dr. Alice E. Smith
Industrial and Systems Engineering, Auburn University

8th Annual SERC Sponsor Research Review
November 17, 2016
20 F Street NW Conference Center
20 F Street, NW
Washington, DC

www.sercuarc.org