COMPARING UTILITY-BASED AND NETWORK-BASED APPROACHES IN MODELING CUSTOMER PREFERENCES FOR ENGINEERING DESIGN

Sha, Zhenghui (2); Bi, Youyi (1); Wang, Mingxian (3); Stathopoulos, Amanda (1); Contractor, Noshir (1); Fu, Yan (3); Chen, Wei (1)

1: Northwestern University; 2: University of Arkansas; 3: Ford Motor Company

ABSTRACT

Customer preference modeling provides quantitative assessment of the effects of engineering design attributes on customers' choices. Utility-based approaches, such as discrete choice model (DCM), and network analysis approaches, such as exponential random graph model (ERGM), have been developed for customer preference modeling. However, no studies have compared these two approaches. Our objective is to identify the distinctions and connections between these two approaches based on both the theoretical foundation and the empirical evidence. Using the vehicle preference modeling as an example, our study shows that when network structure effects are not considered, results from ERGM are consistent with DCM in most of the test cases. However, in one case where customers have varying choice set with multiple alternatives, inconsistencies are observed, possibly due to the discrepancies of the two models in taking different information when calculating choice probabilities. The insights will lead to valuable guidance for choosing the technique for customer preference modeling and co-developing the two frameworks to support engineering design.

Keywords: Market implications, User centred design, Big data, Network modeling, Customer preference modeling

Contact:
Chen, Wei
Northwestern University
Mechanical Engineering
United States of America
weichen@northwestern.edu

Cite this article: Sha, Z., Bi, Y., Wang, M., Stathopoulos, A., Contractor, N., Fu, Y., Chen, W. (2019) 'Comparing Utility-based and Network-based Approaches in Modeling Customer Preferences for Engineering Design', in *Proceedings of the 22nd International Conference on Engineering Design (ICED19),* Delft, The Netherlands, 5-8 August 2019. DOI:10.1017/dsi.2019.390

1 INTRODUCTION

As a critical link between market research and engineering product development, customer preference modeling quantitatively captures the interrelationship amongst market demand, engineering design attributes, and customer demographics (Chen *et al.*, 2013). Customer preference models support the design of consumer products in many aspects, including conceptual design (Hoyle and Chen, 2009), multidisciplinary design (MacDonald *et al.*, 2009), product configuration (Sha *et al.*, 2017), product innovation (Chang and Chen, 2014; Chen *et al.*, 2015), and design accounting for spatiotemporal heterogeneities (Bi *et al.*, 2018). In the past decade, disaggregate quantitative models, such as utility-based choice models, including the Discrete Choice Model (DCM) (Train, 1986) and conjoint analysis (Tovares *et al.*, 2013), have been widely employed by the engineering design research community (Frischknecht *et al.*, 2010; He *et al.*, 2014; Hoyle *et al.*, 2010) for choice/demand estimation. Disaggregate models provide more accurate representation of preferences and predictions than aggregate models that divide customers into groups sharing similar needs and preferences (Kaul and Rao, 1995), such as Multiple Discriminant Analysis (Johnson, 2011), Factor Analysis (Gorsuch, 1983), and Multi-dimensional Scaling (Green, 1970).

Nevertheless, utility-based preference modeling, such as the DCM, is limited when handling dependency of customer decisions (e.g., their decisions may be influenced by each other because of social relations) and collinearity of design attributes (e.g., vehicles with low prices are more possible to have smaller engine capacity) (Wang *et al.*, 2016). To overcome these limitations, recent studies explored the capability of statistical network models in estimating customer preferences (Fu *et al.*, 2017; Sha *et al.*, 2018). Among existing network-based modeling techniques, the Exponential Random Graph Model (ERGM) is increasingly recognized as one of the most powerful analytic techniques (Snijders *et al.*, 2006). ERGM provides a flexible statistical inference framework that can model the influence of both exogenous effects (e.g., nodal attributes) and endogenous effects (network structures/nodal relations) on the probability of connections among nodes. In our prior work, ERGM has been used to study customers' consideration behaviors (Sha *et al.*, 2017), forecast the impact of technological changes on market competitions (Wang, *et al.*, 2016), model customers' consideration-then-choice behaviors (Fu *et al.*, 2017), and predict products' co-consideration relations (Sha *et al.*, 2018; Wang *et al.*, 2018).

While different techniques have been employed, to the best of our knowledge, no studies have compared the network-based approaches with the existing utility-based approaches in terms of their *consistency*, *interpretability* and *predictive power* in modeling customer choices. There is limited understanding of how well the different methodological approaches perform for a given problem. For example, do these two different families of techniques generate different results when applied to the same problem and dataset; and if the results are different, how can they be interpreted? Do the assumptions underpinning utility-based models still hold true in the network-based models? The families of techniques may have the potential to complement each other by addressing weaknesses or being better suited for specific data challenges. Therefore, **the objective of this study** is to identify the distinctions and connections between network and utility approaches, specifically the ERGM versus the DCM, based on both theoretical foundations and the empirical evidence. This will lead to formulating valuable guidance as to when different modelling techniques should be used, and the potential for integration and co-development of the frameworks. It will contribute to the respective, often separate, literature streams through an improved understanding of utility-based approaches versus network-based approaches in modeling customer preferences for design.

The following Section 2 provides a brief introduction of the two models to be compared and the research approach. In Section 3, we use the Chinese auto market as a case study to illustrate the approach and present the results of the comparison. Finally, Section 4 discusses the implications and concludes the paper with closing thoughts and future research opportunities.

2 THEORETICAL MODELS AND RESEARCH APPRAOCH

2.1 Discrete choice model (DCM)

DCM has been used in choice modeling in many application contexts ranging from understanding commuters' choice of commuting mode of transport (Ben-Akiva and Lerman, 1985), to airlines' behaviors in adding a route or not in air transportation networks (Sha *et al.*, 2016), and to customers' choice of vehicles (He *et al.*, 2012). In discrete choice analysis (DCA), a decision maker obtains utility U_i from choosing an alternative *i*, which consists of two parts, the *observed utility* V_i, which is deterministic in nature from the researcher's point of view, and the *unobserved utility* ε_i representing all possible uncertainties associated with the utility, such as unobserved variations, measurement errors, functional misspecifications. This can be modeled as

$$U_i = V_i + \varepsilon_i \tag{1}$$

In a DCM, V is modeled as a function of explanatory variables, typically represented in a linear additive form (Chen *et al.*, 2013), as shown in Equation (2).

$$V_i = \boldsymbol{x_i}\boldsymbol{\beta_i^T} = \beta_{i1}x_{i1} + \beta_{i2}x_{i2} + \ldots + \beta_{ik}x_{ik} , \tag{2}$$

where $\boldsymbol{x_i} = (x_{i1}, x_{i2}, \ldots, x_{in})$ is a vector that contains *n* variables, and $\boldsymbol{\beta_i} = (\beta_{i1}, \beta_{i2}, \ldots, \beta_{in})$ is the vector of model parameters that quantify preferences in choice making. The DCM is derived based on random-utility maximization, meaning that the alternative *i* is chosen rather than *j* if, and only if, $U_i \geq U_j$, $\forall i \neq j$. So the choice probability of alternative *i* is:

$$P_i = P(U_i \geq U_j) = P(V_i - V_j \geq \varepsilon_j - \varepsilon_i) \quad \forall i \neq j . \tag{3}$$

Equation (3) can be solved as soon as the density function $f(\varepsilon)$ is specified because P_i is the cumulative distribution of $\varepsilon_j - \varepsilon_i$. With different $f(\varepsilon)$, various DCM models can be obtained, such as the probit model if assuming that ε follows the Gaussian distribution or the logit model if ε is identical independent distributed following the Gumbel distribution (Chen *et al.*, 2013). In this study, we adopt the logit model for the comparative analysis for demonstration purpose. The model for the scenario where one alternative is chosen from multiple alternatives is called *multinomial logit model*, as shown in Equation (4)

$$P_i = \frac{e^{x_i \beta_i^T}}{\sum_{j=1}^{J} e^{x_j \beta_j^T}} . \tag{4}$$

If the decision scenario is a binary choice, e.g., 1=Yes or 0=No, the resulting choice probability is defined as a *binary logit model*. As shown in Equation (5), the binary logit model is equivalent to a *logistic regression model*.

$$P_1 = \frac{e^{x\beta_1^T}}{e^{x\beta_1^T} + e^{x\beta_0^T}} = \frac{e^{x(\beta_1 - \beta_0)^T}}{1 + e^{x(\beta_1 - \beta_0)T}} = \frac{e^{x\beta^T}}{1 + e^{x\beta^T}} , \tag{5}$$

where $\boldsymbol{\beta} = \boldsymbol{\beta}_1 - \boldsymbol{\beta}_0$ captures the difference of preferences between choosing yes and no. The vector $\boldsymbol{\beta}$ can be readily estimated by statistical estimation techniques, such as maximum likelihood estimation.

2.2 Exponential random graph model (ERGM)

The exponential family of random graph models (ERGM) interprets the global network structure as a collective result of various local network configurations (Robins *et al.*, 2007). The ERGM defines an observed network, ***y***, as one specific realization from a set of possible random networks, ***Y***, following the distribution in Equation (6).

$$Pr(\boldsymbol{Y} = \boldsymbol{y}) = \frac{\exp(\boldsymbol{\theta}'\boldsymbol{g}(\boldsymbol{y}))}{\kappa(\boldsymbol{\theta})} , \tag{6}$$

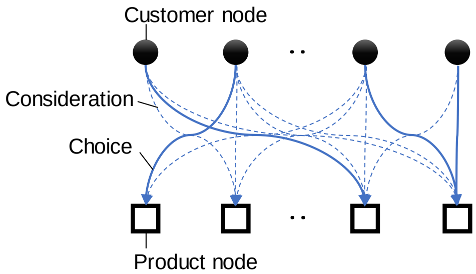

Figure 1. The customer-product bipartite choice network

where $\boldsymbol{\theta}$ is a vector of model parameters, $\boldsymbol{g}(\boldsymbol{y})$ is a vector of the network statistics, and $\kappa(\boldsymbol{\theta})$ is a normalizing quantity to ensure Equation (6) is a proper probability distribution. Equation (6) suggests that the probability of observing any particular network is proportional to the exponent of a weighted combination of network characteristics: one statistic in $\boldsymbol{g}(\boldsymbol{y})$ is more likely to occur if the corresponding $\boldsymbol{\theta}$ is positive. Note that in ERGM, the network itself is a random variable and the probability is

https://doi.org/10.1017/dsi.2019.390 Published online by Cambridge University Press

evaluated on the entire network instead of a link. The major strength of ERGM is its capability of modeling endogenous interdependencies (i.e., relations) among entities with various forms of network structures in addition to exogenous attributes pertaining to nodes and/or edges. Formulating the choice analysis in a network context means that we are interested in understanding what factors (either exogenous or endogenous) drive the formation of a link (choice) between a pair of a customer node and a product node (see Figure 1), and how significant a role each factor plays in that link formation process.

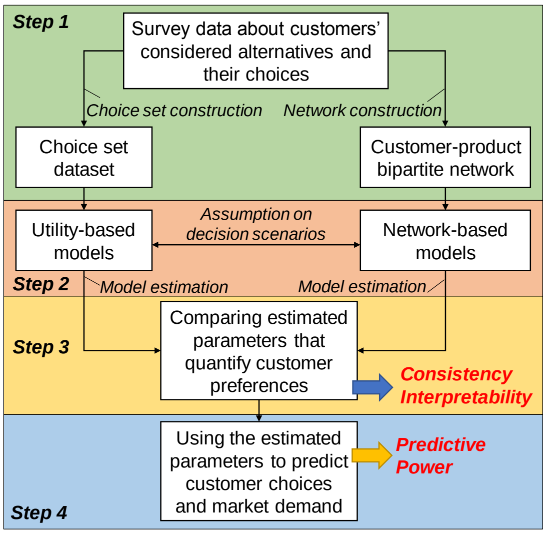

Figure 2. Overview of the research approach

2.3 The research approach

The research approach of this comparative study is shown in Figure 2. A customer survey data (see Section 4.1) containing information about customers' considered products and their final choices, is prepared in Step 1 to provide input for the model comparison. For the utility-based DCM, the input dataset is the choice set along with associated product and customer attributes related to each product alternative and customer. For the network-based ERGM, the input is a customer-product bipartite network, as shown in Figure 1.

In Step 2, the explanatory variables (product and customer attributes) are identified, and the two models are constructed based on the choice scenarios (e.g., binary choices or multinomial choices[1]) and whether interdependencies among choice alternatives will be considered or not (e.g., via different network structures in network-based models). Since the focus of this study is on establishing the equivalency of the two models, no network structure effects are considered. In Step 3, based on the same dataset, the parameters of the two models are estimated and their results are compared to gain knowledge and draw insights into their consistency and interpretability. To establish a common ground for comparison with utility-based approach, special configurations are imposed on the network-based approach in terms of nodal and/or edge constraints, so that the varying consideration set and the "only one final choice" situation can be taken into account in the ERGM (see Section 3.1). Finally, in Step 4, the **predictive power** of these two approaches are compared, e.g., by assessing the predictive accuracy through internal hold-out validation or based on future data. Due to the complexity of comparative predictive modeling for choice behavior, such as missing data of future product attributes and potential biases in declared choice set, our current study is only focused on the first three steps of the proposed approach. When assessing the consistency and interpretability, we tested different settings of choice sets and decision scenarios (see Sections 3.1.2 and 3.2.1 for details).

3 COMPARATIVE STUDY AND ANALYSIS

3.1 Data source and the format of input data

3.1.1 The data source

The dataset used in this study is the 2013 car buyers' survey in the Chinese market. The dataset consists of about 50,000 new car buyers' responses, covering nearly 400 unique car models. The survey has questions covering a variety of topics, including respondent demographics (e.g., age and size of household) and vehicle attributes (e.g., make origin, price and engine size). The respondents were asked to list the car they purchased (i.e., choices), the main alternative car they considered and the second alternative car before making the final choice decision (i.e., choice set).

[1] In this study, binary choice means choosing one from two alternatives, e.g., choosing "Yes" from a choice set containing alternatives of "Yes" and "No". Multinomial choice means choosing one among multiple alternatives.

https://doi.org/10.1017/dsi.2019.390 Published online by Cambridge University Press

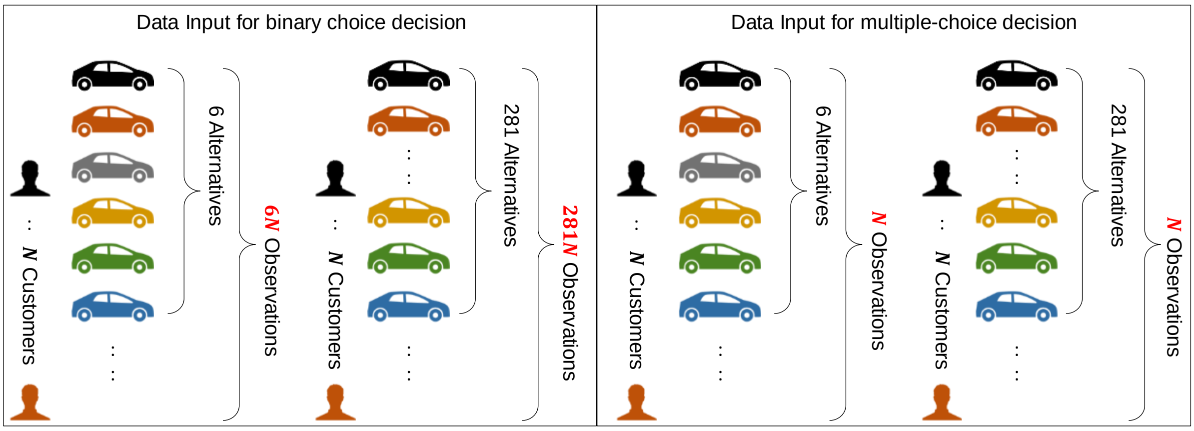

Figure 3. The choice set and the two decision scenarios

As the focus of this paper is a comparative study, to reduce heterogeneity of consumer decision processes, the analysis is limited to the Sedan segment, widespread in the Chinese market. From the original dataset, we create a subset consisting of all the survey respondents who purchased Sedan vehicles. Importantly, even if their final choices are Sedan, the considered vehicles of these respondents may contain car models from different segments. This subset consists of 18,054 respondents and 281 different car models, from which we randomly draw 5000 customers for the sake of computational efficiency.

3.1.2 Two treatments of choice sets

In the original dataset, each customer was restricted to report only one or two considered alternatives. The resulting choice set contains only two or three alternatives, typically with highly correlated vehicle attributes values, which lead to singularity and lack of convergence in the DCM model estimation. To overcome this issue we sample unchosen alternatives from the universal choice set and append them to the original declared choice set for each customer, thereby generating sufficient attribute variation (McFadden, 1978). In this study, a synthetic choice set of six alternatives is created for each customer by sampling and adding three or four car models from the 281 different car models in the original data set. This decision is made based on existing literature (Hauser and Wernerfelt, 1990) reporting that among hundreds of car models in a market, customers typically select three to six models for test driving and consideration. This new dataset is labeled *choiceSet6*, in which each customer's choice set has a fixed number of alternatives (i.e., six) yet varying composition. It should be noted that the use of this artificial choice set does not violate the research objective. For the purpose of comparison, we also adopt another treatment by assuming that each customer considers all possible car models (i.e., 281) before the purchase. This dataset is called *choiceSetAll* where each customer faces the same universal set.

In addition to these two choice set treatments, we also test two different choice scenarios – either a sequence of six binary choices or a single multinomial choices among six alternatives. As shown in Figure 3, for each scenario, two datasets are tested. Given N customers (N=5000 in this study), in the binary choice scenario, we obtain $6N$ observations for the *choiceSet6* dataset and $281N$ observations for the *choiceSetAll* dataset, with each observation of a binary decision. In the multinomial choice, we have N observations of multinomial choice decisions for both datasets.

3.2 Implementation of the models

3.2.1 Scheme of comparison and model settings

Based on whether the decision scenario is modeled as binary or multinomial choices and whether the choice set is sampled or universal, a comparison scheme consisting of four test cases is presented in Table 1.

Table 1. Scheme of comparison

	Binary Choice	**Multinomial Choice**
Varying declared-extracted choice set with 6 alternatives	Test Case 1	Test Case 3
Fixed universal choice set with 281 alternatives	Test Case 2	Test Case 4

https://doi.org/10.1017/dsi.2019.390 Published online by Cambridge University Press

We select four explanatory vehicle attributes based on our prior studies (Fu *et al.*, 2017). These are *Price*, *Fuel Consumption*, *Power*, and *Make Origin country*. For all 281 considered models in the 2013 China market, the average price is 218,163 RMB (~31,617 USD), the average fuel consumption is 9.56 liter per 100 kilometers, and the average power is 147.29 brake horsepower (BHP). The number of car models in each make category is 86 Chinese, 29 American, 88 European, 56 Japanese and 22 Korean, respectively[2]. Due to the large variation in the magnitude of their values and their skewed distributions, log transformation is applied for the three continuous variables. Using log transformation is a common treatment in situations where data exhibits wide and skewed distributions (Galili, 2013). In addition, the z-score normalization is applied to all continuous variables. This enables an across comparison between variables by making their units dimensionless. When using ERGM to model choice probability and estimate customer preferences, the product attributes are added in the network as nodal attributes and the network specification *b2cov* (Morris *et al.*, 2008) that carries nodal attributes are used as input variables $\boldsymbol{g}(\boldsymbol{y})$ for ERGM in Eqn. (6).

While ERGM is a flexible modeling framework, to ensure comparable results from ERGM with those from DCM, two constraint settings are identified and applied for ERGM:

- The *offset* constraint (Hunter *et al.*, 2008) is used to control the nodes to be considered for link formation. Adding this constraint allows us to establish an equivalent network model for the choice set that contains six varying alternatives (*choiceSet6*) in DCM; otherwise, the default setting of ERGM will assume all 281 car models can be linked.
- Degree constraint of a customer node: *constraints=~b1degrees*. This constraint is to ensure the degree of a customer node is equal to one (Handcock and Hunter, 2018), which means each customer chooses one car model as his/her final purchase. DCM models are only normally applicable to such one-final-choice situations.

Corresponding to the test cases shown in Table 1, when applying ERGM, the *offset* needs to be used in Case 1, both *offset* and *constraints=~b1degrees* must be included in Case 3, and *constraints=~b1degrees* needs to be used in Case 4. In Case 2 where all available car models are in consideration, nodal and/or edge constraints are not needed to establish the common ground between ERGM and binary logit model. The estimation of model parameters relies on the algorithm for maximizing the likelihood given the input dataset, and the computations are facilitated by the R packages *mlogit* for DCM and *statnet* for ERGM, respectively.

3.2.2 The results

The estimated parameters of the two models in the four different test cases are shown in Tables 2 and 3, respectively. Results are kept at six-digit decimals for accurate comparison. Table 2 summarizes the results of binary decision representation and provides a comparison between results of two types of models, both ERGM and binary logit. It is observed that under both varying individual choice sets and the fixed universal set, the estimated model parameters of the ERGM are identical to their binary logit counterparts. This indicates that in modeling the binary choice decisions, if only exogenous factors are considered, ERGM provides *consistent* results to those from the binary logit model. By using the *offset* setting to control the alternatives to be considered, the ERGM can easily be applied to the situations where customers have different choice sets.

It is also observed that the models under two different choice set assumptions lead to similar interpretation of attribute significance. For example, among all the continuous attributes investigated, *price* turns out to be the most important one having the largest coefficient magnitude (-3.179 in both cases). All models indicate that higher price makes a car model less desirable to choose. The estimated coefficients of other continuous attributes are consistent with our intuition. In general, a sedan that is less pricey, more powerful and consumes less gas, is more likely to be purchased. For the categorical variable, *Make Origin*, the results show that the cars from Europe are the most attractive to Chinese customers as compared to the ones from other foreign regions.

In the multinomial choice scenario shown in Table 3, we observe different model estimates for the two approaches. Yet, they show similar trends in terms of factor significance and relative importance compared

[2] Value of the vehicle attributes are estimated by averaging the data reported by respondents. We have verified these data against published manufacturing information online. We didn't observe substantial differences.

to their binary choice counterpart. ERGM with the inclusion of *constraints =~b1degrees* in Test Case 4 generates almost the same results as those from the multinomial logit model. In this case, every customer is assumed to have the same choice set including all vehicle models (i.e., 281 car models).

Table 2. Results of test cases 1 and 2 with the comparison between ERGM and binary logit model

	Varying choice set (Case 1)		**Fixed choice set (Case 2)**	
Independent Variables	**ERGM**	**Binary Logit**	**ERGM**	**Binary Logit**
Intercept (or *edge* term in ERGM)	-4.24566***	-4.24566***	-8.42404***	-8.42404***
Price	-3.17890***	-3.17890***	-3.17909***	-3.17909***
Fuel consumption	-0.49813***	-0.49813***	-0.70981***	-0.70981***
Power	2.00439***	2.00439***	2.19884***	2.19884***
Make origin: American	1.82652***	1.82652***	2.41545***	2.41545***
Make origin: European	2.91118***	2.91118***	3.32904***	3.32904***
Make origin: Japanese	1.25843***	1.25843***	1.22876***	1.22876***
Make origin: Korean	1.23018***	1.23018***	1.41615***	1.41615***

*The code *** indicates the 0.001 Level of significance.*

However, in Test Case 3 where each customer choice set contains six varying alternatives, the results from ERGM and the multinomial logit model are different in their magnitudes. Moreover, the estimated coefficients of the variable *fuel consumption* yield contradicting interpretations. In the other three cases, the sign of this coefficient is negative meaning that a car model was more likely to be purchased if it had lower fuel consumption. The positive coefficient (i.e., 0.416578) of the DCM in Test Case 3 seems counterintuitive[3]. This difference may be attributed to the difference in probability calculation governed by Equations (4) and (6), which, in turn, affect the model estimation. In DCM, the choice probability is calculated within the defined choice set. But in ERGM, the probability of a network structure is predicted. This observed difference sheds light on the interpretation of the choice-making process underlying the two models. In DCM, the model indicates that each customer evaluates his/her own choice set and make a final choice by picking the alternative with the maximal utility. The comparison of utilities is done within choice set and customers do not refer to the alternatives outside the choice set. In ERGM, however, there is no concept of utility. The linking probability between two nodes is calculated based on the information from the entire network as well as the specified network structures. The treatment in network models resembles a more realistic decision process where customers choose not only based on product attributes and relations among products within the choice set, but also based on their "awareness" of the aggregate information at market level. Shocker *et al*. (1991) highlighted the role of a more general "awareness set" in choice behaviors in addition to the more restricted concept of "consideration set" and "choice set". The capability of ERGM to model the effect of network structures beyond the choice set of each customer, suggests it can be viewed as a promising tool to model the "awareness set" more explicitly. The above interpretation also explains why in Case 4 the multinomial logit model produces the same results as the ERGM, because every customer considers all car models which is equivalent to knowing the market-level information.

Table 3. Results of test cases 3 and 4 with the comparison between ERGM and multinomial logit model

	Varying choice set (Case 3)		**Fixed choice set (Case 4)**	
Independent Variables	**ERGM**	**Multinomial Logit**	**ERGM**	**Multinomial Logit**
Intercept (or *edge* term in ERGM)	N/A	N/A	N/A	N/A
Price	-3.21187***	-5.113551***	-3.15749***	-3.157432***
Fuel consumption	-0.50015***	0.416578***	-0.70688***	-0.706719***
Power	2.02308***	2.230066***	2.18522***	2.184793***
Make origin: American	1.85362***	3.893254***	2.39982***	2.399382***

[3] We have studied two additional datasets with 5000 samples each and every customer's choice set has six alternatives including both the declared considerations and the randomly generated items. No significant difference is observed, and the same conclusions hold.

https://doi.org/10.1017/dsi.2019.390 Published online by Cambridge University Press

Make origin: European	2.96787***	5.822052***	3.30722***	3.306883***
Make origin: Japanese	1.27344***	4.184239***	1.22007***	1.219463***
Make origin: Korean	1.25281***	3.542386***	1.40699***	1.405878***

*The code *** indicates the 0.001 Level of significance.*

It should be noted that the comparisons thus far are only focused on the investigation of the impact of exogenous effects, e.g., the vehicle and customer attributes, on choice behaviors. However, our prior studies and many exiting works (Fu *et al*., 2017; Rahman *et al*., 2018; Wang *et al*., 2018) have shown that the endogenous effects, such as three-way competition and peer effects, play an important role in customers' decision making. For example, Wang *et al*. (2018) found that peer effects are statistically significant in customers' consideration decisions for luxury vehicle buyers in China. Such interdependencies (i.e., endogenous effects) can be well modelled by ERGM due to its capability of handling correlated nodal attributes and interdependent links by incorporating various effect networks (Figure 4) (Lusher *et al*., 2012). In our future work, we will investigate, if taking exogenous effects into account, how the model results would be different and how well ERGM would perform compared to DCM. In theory, the results from the network-based approach should yield more reliable interpretations of factor effects as it avoids faulty inferences on covariates by considering interdependence (Cranmer and Desmarais, 2011).

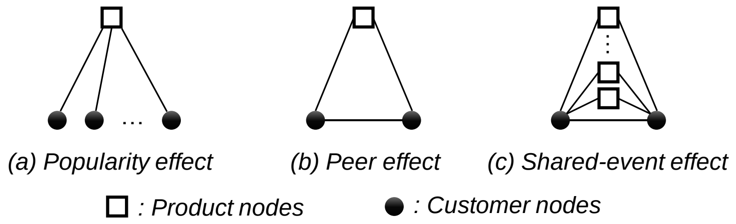

Figure 4. Endogenous network effects that can be added to the model: (a) a customer is more likely to purchase a product favored by majority of customers; (b) customers tend to purchase a product their peers recommended; (c) shared-event effect helps answer the question that if two customers purchased the same product, are they more likely to have a second common product in choice, and a third one and so on?

4 CONCLUSIONS

While both utility-based and network-based approaches are commonly used to model customer choice behaviors and estimate demand in engineering design, the equivalency and consistency of these two modelling approaches have not been reported. Using DCM and ERGM as representative modelling techniques from these two categories of methods, this paper compares the consistency and interpretability of these two modelling techniques. Depending on the choice assumptions studied and whether the choice sets are sampled or universal, four test cases are created and the two models are compared in each case. The results indicate that ERGM provides consistent results and the same factor effects as DCM in most of the test cases when only exogenous variables are considered. As ERGM is not restricted to the number of choices to be estimated and can include both exogenous and endogenous variables as inputs, it provides a more general and flexible modelling framework. As such our results suggest that DCM can be viewed as a special case of ERGM where no non-exogenous (or network structural) effects are specified and only one product node can be chosen from the choice set for link formation. In one case where the models aim to estimate customer preferences in multinomial choice situation with varying choice set, the two models show inconsistencies in the model estimates. The inconsistencies stem from the discrepancies between the two models in the calculation of choice probabilities – while DCM directly predicts the choice probabilities based on the individual customer's decision, the ERGM predicts the probabilities of the structures of the entire network by acknowledging the dependencies in the chose data, and as a result, the linking probability between two nodes is predicted based on the information from the entire network.

The contribution of this study can be summarized in two aspects. First, this study provides an approach to examine the consistency and interpretability of utility-based and network-based choice modelling approaches. Depending on the research hypothesis to be tested or application context, researchers can refer to the more appropriate one for modelling purpose. Second, this study successfully identifies the important constraints that must be applied in order to make ERGM comparable to DCM, for example, the use of *offset* for controlling the varying choice set and the use of *constraints=~b1degrees* for controlling the number of final choices to be one. Such constraints provide clues concerning the distinctions and connections between these two models. Proper use of

https://doi.org/10.1017/dsi.2019.390 Published online by Cambridge University Press

these two constraints provides us with a flexible approach to modelling many decision scenarios that DCM alone has difficulties to represent.

Based on this study, our future work is planned in two directions. First, drawing on the estimated models shown in Tables 2 and 3, prediction analysis will be performed to compare how accurately different approaches can predict customers' choices using test data that contains hold-out information (Step 4 in Figure 2). Second, we plan to add the effect networks, such as those shown in Figure 3, to study the endogenous effects in vehicle market on customer choice behaviors, and investigate whether the inclusion of the effect networks improve the predictive performance of ERGM.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the financial support from NSF-CMMI-1436658 and Ford-Northwestern Alliance Project.

REFERENCES

Ben-Akiva, M.E. and Lerman, S.R. (1985), *Discrete Choice Analysis: Theory and Application to Travel Demand*, Vol. 9, MIT press.

Bi, Y., Xie, J., Sha, Z., Wang, M., Fu, Y. and Chen, W. (2018), "Modeling Spatiotemporal Heterogeneity of Customer Preferences in Engineering Design", *ASME 2018 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference*, Quebec City, Quebec, Canada.

Chang, D. and Chen, C.-H. (2014), "Understanding the influence of customers on product innovation", *International Journal of Agile Systems and Management 20*, Vol. 7 No. 3–4, pp. 348–364, Inderscience Publishers Ltd.

Chen, C.-H., Khoo, L.P. and Chen, N.-F. (2015), "Consumer Goods", *Concurrent Engineering in the 21st Century*, pp. 701–733, Springer.

Chen, W., Hoyle, C. and Wassenaar, H.J. (2013), *Decision-Based Design: Integrating Consumer Preferences into Engineering Design, Decision-Based Design: Integrating Consumer Preferences into Engineering Design*, Springer Science and Business Media.

Cranmer, S.J. and Desmarais, B.A. (2011), "Inferential network analysis with exponential random graph models", *Political Analysis*, Vol. 19 No. 1, pp. 66–86, Cambridge University Press.

Frischknecht, B.D., Whitefoot, K. and Papalambros, P.Y. (2010), "On the Suitability of Econometric Demand Models in Design for Market Systems", *J. Mech. Des*, Vol. 132 No. 12, pp. 121007.

Fu, J.S., Sha, Z., Huang, Y., Wang, M., Fu, Y. and Chen, W. (2017), "Modeling Customer Choice Preferences in Engineering Design using Bipartite Network Analysis", *Proceedings of the ASME 2017 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Cleveland, Ohio.

Galili, T. (2013), "Log Transformations for Skewed and Wide Distributions", *R-Statistics Blog*, available at: https://www.r-statistics.com/2013/05/log-transformations-for-skewed-and-wide-distributions-from-practical-data-science-with-r/.

Gorsuch, R.L. (1983), *Factor Analysis*, Lawrence Erlbaum Associates, Hillsdale.

Green, P.E. (1970), *Multidimensional Scaling and Related Techniques in Marketing Analysis*, Allyn and Bacon.

Handcock, M.S. and Hunter, D.R. (2018), "Reference manual of package ergm", available at: https://cran.r-project.org/web/packages/ergm/ergm.pdf.

Hauser, J.R. and Wernerfelt, B. (1990), "An evaluation cost model of consideration sets", *Journal of Consumer Research*, Vol. 16 No. 4, pp. 393–408, The University of Chicago Press.

He, L., Chen, W. and Conzelmann, G. (2012), "Impact of vehicle usage on consumer choice of hybrid electric vehicles", *Transportation Research Part D: Transport and Environment*, Vol. 17 No. 3, pp. 208–214, Elsevier.

He, L., Wang, M., Chen, W. and Conzelmann, G. (2014), "Incorporating social impact on new product adoption in choice modeling: A case study in green vehicles", *Transportation Research Part D: Transport and Environment*, Vol. 32, pp. 421–434, Elsevier.

Hoyle, C., Chen, W., Wang, N. and Koppelman, F.S. (2010), "Integrated Bayesian hierarchical choice modeling to capture heterogeneous consumer preferences in engineering design", *Journal of Mechanical Design*, Vol. 132 No. 12, pp. 121010, American Society of Mechanical Engineers.

Hoyle, C.J. and Chen, W. (2009), "Product attribute function deployment (PAFD) for decision-based conceptual design", *IEEE Transactions on Engineering Management*, Vol. 56 No. 2, pp. 271–284.

Hunter, D.R., Handcock, M.S., Butts, C.T., Goodreau, S.M. and Morris, M. (2008), "ergm: A package to fit, simulate and diagnose exponential-family models for networks", *Journal of Statistical Software*, Vol. 24 No. 3, pp. nihpa54860, NIH Public Access.

Johnson, R. (2011), *Multiple Discriminant Analysis: Marketing Research Applications*, Marketing Classics Press.

Kaul, A. and Rao, V.R. (1995), “Research for product positioning and design decisions: An integrative review”, *International Journal of Research in Marketing*, Vol. 12 No. 4, pp. 293–320, Elsevier.
Lusher, D., Koskinen, J. and Robins, G. (2012), *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications*, Cambridge University Press.
MacDonald, E.F., Gonzalez, R. and Papalambros, P.Y. (2009), “Preference Inconsistency in Multidisciplinary Design Decision Making”, *J. Mech. Des.*, Vol. 131 No. 3, pp. 031009.
McFadden, D. (1978), “Modeling the choice of residential location”, *Transportation Research Record*, No. 673.
Morris, M., Handcock, M.S. and Hunter, D.R. (2008), “Specification of exponential-family random graph models: terms and computational aspects”, *J. Stat. Softw.*, Vol. 24 No. 4, pp. 1548.
Rahman, M.H., Gashler, M., Xie, C. and Sha, Z. (2018), “Automatic Clustering of Sequential Design Behaviors”, *ASME 2018 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference*, Quebec City, Canada, Paper #: DETC2018-86300.
Sha, Z., Huang, Y., Fu, S., Wang, M., Fu, Y., Contractor, N. and Chen, W. (2018), “A Network-Based Approach to Modeling and Predicting Product Co-Consideration Relations”, *Complexity*, Vol. 2018.
Sha, Z., Moolchandani, K., Panchal, J.H. and DeLaurentis, D.A. (2016), “Modeling Airlines’ Decisions on City-Pair Route Selection Using Discrete Choice Models”, *Journal of Air Transportation*, American Institute of Aeronautics and Astronautics.
Sha, Z., Saeger, V., Wang, M., Fu, Y. and Chen, W. (2017), “Analyzing Customer Preference to Product Optional Features in Supporting Product Configuration”, *SAE International Journal of Materials and Manufacturing*, Vol. 10), No. 2017-1–243.
Shocker, A.D., Ben-Akiva, M., Boccara, B. and Nedungadi, P. (1991), “Consideration set influences on consumer decision-making and choice: Issues, models, and suggestions”, *Marketing Letters*, Vol. 2 No. 3, pp. 181–197, Springer.
Snijders, T.A.B., Pattison, P.E., Robins, G.L. and Handcock, M.S. (2006), “New specifications for exponential random graph models”, *Sociological Methodology*, Vol. 36 No. 1, pp. 99–153, SAGE Publications Sage CA: Los Angeles, CA.
Tovares, N., Cagan, J. and Boatwright, P. (2013), “Capturing Consumer Preference Through Experiential Conjoint Analysis”, *ASME Paper No. DETC2013-12549*.
Train, K. (1986), *Qualitative Choice Analysis: Theory, Econometrics, and an Application to Automobile Demand*, MIT press, Vol. 10.
Wang, M., Chen, W., Huang, Y., Contractor, N.S. and Fu, Y. (2016a), “Modeling customer preferences using multidimensional network analysis in engineering design”, *Design Science*, Cambridge University Press, Vol. 2.
Wang, M., Sha, Z., Huang, Y., Contractor, N., Fu, Y. and Chen, W. (2016b), “Forecasting Technological Impacts on Customers’ Co-Consideration Behaviors: A Data-Driven Network Analysis Approach”, *ASME 2016 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Charlotte, NC, USA, August 21-24, 2016.
Wang, M., Sha, Z., Huang, Y., Contractor, N., Fu, Y. and Chen, W. (2018), “Predicting Products’ Co-Considerations and Market Competitions for Technology-Driven Product Design: A Network-Based Approach”, *Design Science Journal*, Vol. 4.