Incorporating social impact on new product adoption in choice modeling: A case study in green vehicles

Lin He, Mingxian Wang, Wei Chen, Guenter Conzelmann

Environmental Energy Technologies Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road MS 9084000, Berkeley, CA 94720, United States
Mechanical Engineering, Northwestern University, 2145 Sheridan Road, Tech B224, Evanston, IL 60208, United States
Center for Energy, Environmental, and Economic Systems Analysis, Argonne National Laboratory, Argonne, IL 60439, United States

Abstract
While discrete choice analysis is prevalent in capturing consumer preferences and describing their choice behaviors in product design, the traditional choice modeling approach assumes that each individual makes independent decisions, without considering the social impact. However, empirical studies show that choice is social – influenced by many factors beyond engineering performance of a product and consumer attributes. To alleviate this limitation, we propose a new choice modeling framework to capture the dynamic influence from social networks on consumer adoption of new products. By introducing social influence attributes into a choice utility function, social network simulation is integrated with the traditional discrete choice analysis in a three-stage process. Our study shows the need for considering social impact in forecasting new product adoption. Using hybrid electric vehicles as an example, our work illustrates the procedure of social network construction, social influence evaluation, and choice model estimation based on data from the National Household Travel Survey. Our study also demonstrates several interesting findings on the dynamic nature of new technology adoption and how social networks may influence hybrid electric vehicle adoption.

Introduction

While the use of Discrete Choice Analysis (DCA) is prevalent in modeling consumer preferences and describing their choice behaviors in product design (Frischknecht et al., 2010; Li and Azarm, 2000; Michalek et al., 2006; Wassenaar and Chen, 2003; Williams et al., 2008), individuals’ choices are studied without social contexts in most cases. Empirical studies show that social context, such as “neighbor” effects may impact consumer choice behavior (Case, 1992). Often times, social context influences consumer attitudes towards new products, such as those involving green technology. As an example, a consumer’s decision in choosing an eco-friendly alternative fuel vehicle, such as a hybrid electric vehicle (HEV) or plug-in hybrid electric vehicle (PHEV), may be influenced by neighbors and friends or others who share similar social status or profile. In the broad market of consumer products, a large amount of product reviews and recommendations are now made available through rapidly growing online shopping websites and social networking sites which accelerate the social impact on product adoption. Integrating social network simulation into consumer choice modeling and developing methods for...
predicting the social influence on consumer choices and their attitudes towards adopting new green products is the focus of this research.

While the existing work demonstrated the benefits of using DCA in modeling consumer choice behavior (He et al., 2012b; Hoyle et al., 2011; MacDonald et al., 2009; Shiau and Michalek, 2009; Sullivan et al., 2011), the merits of DCA are limited due to its assumption of consumers making individual decisions in isolation of each other. As many behavioral economists and psychologists have noted, choice is social. In other words, an individual’s decisions are not immune to the influence of others. This is especially the case in forecasting the adoption (first-time purchases) of new green products, which is a critical but challenging task. A handful of recent research projects focus on forecasting the HEV/PHEV market potential as the vehicle design evolves and the technology matures. A few pilot projects have been conducted to better understand consumers’ knowledge and awareness of PHEV (Axsen and Kurani, 2008). For choice modeling of alternative fuel vehicles, He et al. (2012a) quantitatively assessed the impact of vehicle usage on HEV choice and demonstrated that consumers driving locally tend to prefer HEV more than consumers with longer commutes. Sullivan et al. (2005) suggested that consumers make purchasing decisions based on their own personal attributes as well as vehicle attributes. They later developed an agent-based simulation approach for modeling market penetration of PHEVs under a variety of consumer, economic, and policy conditions (Sullivan et al., 2009). Struben and Sterman (2008) simulated word-of-mouth effects in alternative fuel vehicle diffusion using a Logit-like choice model. However, existing studies still mostly focus on understanding the impact of marketing attributes and largely ignore the social impacts on consumers’ choices. The effects of peer influence on product attribute preference have been studied in market science by Narayan et al. (2011) who modeled three different mechanisms of social influence. By combining traditional conjoint analysis on product features with peer influence, their work showed that peer influence causes people to change perspective on product importance, and that some product attributes are more sensitive to change than others.

The research objective of this work is to develop an alternative choice modeling framework considering the social impact on new product adoption by integrating methods rooted in social network theories, agent-based modeling, and discrete choice analysis. In contrast to Narayan’s approach that heavily relies on customer survey data to evaluate the attitude change before and after exposure to peer influence, our approach acknowledges the lacking of customer survey data and employs agent-based simulations to simulate social network influence. This framework can be used by product designers to estimate willingness-to-pay for new technology, illustrate consumer preferences and tradeoffs among multiple product design attributes, and forecast product market share for a target market with given social-demographic attributes. By introducing the social influence attributes into the choice utility function, the social network simulation is integrated with the traditional discrete choice analysis by following the procedure of social network construction, social influence evaluation, and choice model estimation. To the authors’ knowledge, this work is the first in literature that integrates network simulation for assessing social influence into choice modeling. In the rest of this paper, we will first provide a review of the literature on social network theories and existing work on integrating social interactions in choice models in section ‘Social network theories and integration with choice modeling’. The proposed choice modeling framework considering social impacts is presented in section ‘Choice modeling framework considering social impact’, followed by a case study of modeling hybrid electric vehicle

<table>
<thead>
<tr>
<th>Nomenclature</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A</strong></td>
</tr>
<tr>
<td><strong>z</strong></td>
</tr>
<tr>
<td><strong>β</strong></td>
</tr>
<tr>
<td><strong>d_{ij}</strong></td>
</tr>
<tr>
<td><strong>E</strong></td>
</tr>
<tr>
<td><strong>ε_{ik}</strong></td>
</tr>
<tr>
<td><strong>HEV</strong></td>
</tr>
<tr>
<td><strong>L_{ij}</strong></td>
</tr>
<tr>
<td><strong>l_{ij}</strong></td>
</tr>
<tr>
<td><strong>MNL</strong></td>
</tr>
<tr>
<td><strong>Mt</strong></td>
</tr>
<tr>
<td><strong>N_{ik,t}</strong></td>
</tr>
<tr>
<td><strong>PHEV</strong></td>
</tr>
<tr>
<td><strong>γ</strong></td>
</tr>
<tr>
<td><strong>S</strong></td>
</tr>
<tr>
<td><strong>W_{ik,t}</strong></td>
</tr>
<tr>
<td><strong>X</strong></td>
</tr>
<tr>
<td><strong>X_{in}</strong></td>
</tr>
</tbody>
</table>
ownership in California in section ‘Case study of green product adoption: HEV ownership in California’. Discussions and conclusions are included in section ‘Conclusions’.

Social network theories and integration with choice modeling

With the growing public awareness of the complex “connectedness” of modern society, the idea of social networks has been gaining more attention (Easley and Kleinberg, 2010; Faust and Wasserman, 1994). A social network is defined as a group of people who are connected to some or all of the others following a random or particular pattern in graph. For example, Facebook, the leading online social network site has so far attracted 1 billion active users (Facebook), demonstrating the power of interpersonal connections in our daily lives. In the following subsections, fundamentals in social network theories, as well as the integration of social network simulations into choice modeling will be discussed.

Social network basics

There are two key elements of a social network which includes nodes, representing members of the network, i.e. consumers in the context of product design, and links, illustrating the connections between members, i.e. social interactions between linked consumers. For example, the graph in Fig. 1 consists of 4 nodes labeled A, B, C, and D, with four links between them. Two nodes are “friends” if they are connected by a link, e.g., A and B, B and D. The term “friends” has a broad meaning in our proposed social network simulation. As to be explained in section ‘Choice modeling framework considering social impact’, the axes $x^{(1)}$ and $x^{(2)}$ are example of consumer attributes that are used to describe a social space. In this research, social distance based on the similarity of consumer attributes is used to determine whether a link between two consumers exists or not.

Homophily refers to the principle that friends tend to be similar. Typically, friends are similar either along their social-economic attributes, such as racial and ethnic background, age, where they live, occupations, income, or their behaviors, interests, beliefs, and opinions. The tendency of people to form friendships with others who are like them is often defined as selection, in that people are selecting friends with similar characteristics. The process of selection is the basis for the social network construction (stage I of the proposed choice modeling framework described in section ‘Choice modeling framework considering social impact’). On the other hand, people may modify their behaviors to bring them more closely into alignment with the behaviors of those who they connect with in the network. This process has been described as social influence (Friedkin, 2006), which can be viewed as the reverse of selection: with selection, the individual characteristics drive the formation of links, while with social influence, the existing links in the network serve to shape people’s behavior, attitudes, etc. In the context of product design, consumers are often influenced by the choices of those who they connect with in the network when they purchase a new product. The process of social influence is the basis for stage II (social influence evaluation) of the proposed choice modeling framework described in section ‘Choice modeling framework considering social impact’.

Depending on the specific network structure, distinctive influences through social networks have been observed, modeled, and researched in numerous domains, including social science and humanities. The meaning of “connectedness” encompasses two related issues in social network modeling and simulation: one is the network structure – the media of social impact; the other is the behavioral interactions – the mechanism of social impact. How to integrate these two key elements of social network into consumer behavior simulation and choice modeling are discussed next in more detail.

Integration with choice modeling: literature

The importance of incorporating social influence in modeling choice behavior has been stated by McFadden (2010) and many other scholars. McFadden decomposed the causes of the sociality of choice by stating that choice is influenced by information from a peer group, heuristics rooted in the behavior of others, analogies or anecdotal information garnered from associates, and constraints imposed by others.

Economists are among the pioneers in quantitatively modeling the interdependence of choices. Social interactions are introduced in binary logit (Brock and Durlauf, 2001), multinomial logit (Brock and Durlauf, 2002), and nested logit models
by allowing a given consumer’s choice for a particular alternative to be dependent on the overall market share, that is, the global (aggregated) social network effects. A more general framework is presented in Dugundji and Gulyas (2003b) for studying local (disaggregated) social network effects in discrete choice models, where network effects are calculated within each market segment. Modeling preference interdependence among consumers was also studied by Yang and Allenby (2003) based on a Bayesian spatial autoregressive multivariate binomial probit modeling approach.

In the transportation domain, Carrasco et al. demonstrated approaches for collecting social network data (Carrasco et al., 2008), and incorporated social network information into the activity–travel behavior modeling framework to better understand social activities and key aspects of the underlying behavioral process (Carrasco and Miller, 2006). Tran (2012) developed an agent-based modeling (ABM) framework to demonstrate the important role that network influence plays in accelerating energy innovation diffusion. Although ABM can provide a distribution of possible outcomes, calibration of model parameters is challenging which makes the prediction even more difficult. Dugundji and Gulyas (Dugundji and Walker, 2005) utilized simulated network data for modeling intercity travel behavior by introducing local social network effects into the choice utility with varying network densities. Páez et al. (2008) presented a multinomial logit model of residential location choice using simulated network data with varying degrees of distributions and clustering parameters. To capture social influences without explicit knowledge of the individual networks, Walker et al. (2011) introduced a local social network effect in their choice model, namely, the percent of population choosing a specific alternative within their peer group defined based on the socio-economic status (income, education, age) and the spatial proximity of residential locations. Although these approaches address the local social network effect, they only capture semi-disaggregate level preferences and are subject to uncertainty depending on the level of segmentation (or definition of peer groups) and the condition of sharp boundary – people at the boundary of peer groups are assumed to be isolated from neighboring peer groups.

While the growing number of publications in the economics and transportation fields provided the theoretical foundation to incorporate social interactions into choice modeling, a major limitation of the aforementioned methods is the assumption that the social impact is the primary critical factor in the choice model. In many fields, including product design, other factors such as product performance, characteristics of consumers, as well as the product usage contexts, are shown to be critical in consumers’ choices. When social network effects are present, taking them into account in the choice model can lead to different product designs decisions. Consider a simple social network with 3 people where all consumers are related to each other. There are two product options available and the designer is to decide which product to offer in the market. Consumer can choose between product 1 and not-to-buy (outside goods), or between product 2 and not-to-buy. A Logit model with utilities and choice probabilities is given in Table 1. A product is considered as a better option for the designer if the aggregate choice probability across all consumers is higher. Without considering social impact, we only focus on time \( t = 1 \), where product 2 is the better option (66.66% aggregate choice probability versus 33.34% of product 1). Note the individual choice probability is assessed based on binomial logit model. Now, suppose person 2 makes a choice at \( t = 1 \) and persons 1 and 3 were to follow at \( t = 2 \), person 2’s purchase at \( t = 1 \) influences persons 1 and 3 by increasing their individual utility of product 1 by 5 respectively (underlined values in the table). As a result, product 1 becomes the better choice, because its aggregate choice probability becomes higher than that of product 1 (58.68% versus 41.32%) due to the increased chance of person 3 choosing product 1.

As demonstrated in the above conceptual example, capturing the social impact and modeling the dynamic and stochastic nature of product adoption in a systematic way does matter in product design. Hence, research is needed to address the unique challenges in incorporating social influences in choice modeling for product design applications.

### Table 1
Illustration of social impact on individual utility and choice probability.

<table>
<thead>
<tr>
<th>Person</th>
<th>Utilities</th>
<th>Choice probability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Product 1</td>
<td>Product 2</td>
</tr>
<tr>
<td>( t = 1 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>20</td>
<td>28</td>
</tr>
<tr>
<td>2</td>
<td>32</td>
<td>28</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>24</td>
</tr>
<tr>
<td>Aggregated</td>
<td>62.69</td>
<td>66.67</td>
</tr>
<tr>
<td>( t = 2 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>25</td>
<td>28</td>
</tr>
<tr>
<td>2</td>
<td>32</td>
<td>28</td>
</tr>
<tr>
<td>3</td>
<td>25</td>
<td>24</td>
</tr>
<tr>
<td>Aggregated</td>
<td>94.42</td>
<td>66.67</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>98.50</td>
<td>100.00</td>
</tr>
<tr>
<td>2</td>
<td>88.08</td>
<td>11.92</td>
</tr>
<tr>
<td>3</td>
<td>85.32</td>
<td>88.08</td>
</tr>
<tr>
<td>Aggregated</td>
<td>90.63</td>
<td>66.67</td>
</tr>
</tbody>
</table>
Choice modeling framework considering social impact

An alternative three-stage choice modeling framework is developed in this research to quantitatively capture the impact of social network on consumer choice behavior for product design. The adoption of alternative fuel vehicles under social influence is used as an example to demonstrate the proposed framework.

The proposed choice modeling framework consists of the following three stages: (I) social network construction, in which a virtual environment is created where consumers interact with their linked network members; (II) social influence evaluation, where the network influence is evaluated in the form of social influence attributes by simulating how consumers communicate with their linked network members and the accumulated influence each receives; and (III) choice model estimation, in which consumers’ rational decisions based on the utility maximization theory are studied to quantify the impact from social influence modeled in stage II together with other product and consumer attributes. The details of each stage are discussed as follows.

Stage I: social network construction

Social network construction is the process of creating network links based on existing data – the predefined network structure and the descriptive information associated with each node (consumer). Due to the complexity in social network data collection, often there is limited data, if any, available on the real social network structure among the sample population. An alternative is to construct the network through simulations based on certain hypotheses of a network structure using collected consumer attributes, such as the socio-demographic and usage context attributes. In our case study of hybrid electric vehicle (HEV), we use the National Household Travel Survey (NHTS) to obtain the demographic and usage context attributes of the population. Our network construction follows commonly used social network structures, such as the small-world network with short average path length (Watts and Strogatz, 1998a) and the scale-free network with power-law degree distribution (Barabási and Albert, 1999). The former is demonstrated in our case study (section ‘Case study of green product adoption: HEV ownership in California’), while the latter is suitable for explaining network mechanisms such as preferential attachment – one classic example being the World Wide Web.

Within the network construct, links are generated to simulate interpersonal interactions. Rogers (1995) suggested that interpersonal influence can occur among individuals who are homophilous (i.e., similar to each other) or who are heterophilous (i.e., dissimilar to each other). Homophilous connections, or close links, represent the connections among neighbors, friends, and other regular contacts (e.g. coworkers), whereas heterophilous connections, or distant links, represent the links created due to acquaintances and other information sources (e.g. online reviewers). Foundational to forming the links is the concept of social distance, which is defined as the distance between locations of two nodes (consumers) in a social space. The term social space is inspired by Krugman’s work on economic geography (Krugman, 1990). A social space can be constructed based on the attributes that are used to describe the consumer social dimensions.

\[ d_{ij} = \left( \sum_{m} |x_i^m - x_j^m|^p \right)^{1/p} \]

(1)

As shown in Eq. (1), \( d_{ij} \) is defined as the \( p \)-norm distance in social space between consumer \( i \) and \( j \), while \( x_i^m \) represents the attributes in the \( m \)-dimensional social space. For product choice modeling, attributes \( x \) of consumer social dimension include, but are not limited to, consumer profile attributes \( S \) and usage context attributes \( E \). For categorical attributes, such as gender and race, data transformations are needed to calculate the distance \( d \). Referring back to Fig. 1 which shows a simple network graph with four nodes A, B, C, and D, in a two dimensional space formed by \( x_1 \) and \( x_2 \), there are four links between the nodes. The length of the link, in this case, represents the Euclidean distance between two nodes, distance \( d_{AB} \) between A and B for example. This basic form of distance function is employed in our case study (section ‘Case study of green product adoption: HEV ownership in California’).

Based on the homophily principle, two nodes with shorter social distance are more likely to be connected. The distance-decay function method reflects the hypothesis that the degree of influence between nodes should decrease as their attributes become more dissimilar (Festinger, 1954). As shown as follows, the strength of a connection in social space is a function of the distance between nodes (Páez et al., 2008):

\[ l_{ij} = \begin{cases} \gamma_1 \exp(-\gamma_2 d_{ij}^p), & \text{for } i \neq j \\ 0, & \text{for } i = j \end{cases} \]

(2)

In Eq. (2), \( \gamma_1 \) is the parameter controlling the magnitude of the effect and \( \gamma_2 \) is controlling the rate of decay. To determine whether a link exists (1) or not (0), the distance decay function in Eq. (2) is replaced with a significance criterion of a given threshold, as shown in Eq. (3).

\[ L_{ij} = \begin{cases} 1, & \text{if } l_{ij} \geq b_i \\ 0, & \text{otherwise} \end{cases} \]

(3)

where \( b_i \) is the threshold value determined by the modeler. Other treatments, such as relative connection strength, can be used to model more complex network structures.
Stage II: social influence evaluation

Once a network is constructed, social influences on consumer preferences are evaluated by simulating the interactions between consumers in the network. As a result, social influence attribute $N_{i,t}$ is defined as the collection of influences from all other consumers linked to a focal consumer $i$ at time $t$. Because the social influence is a function of time $t$, $N$ is evaluated for each of the time periods. This attribute is then added into the consumer choice utility function to capture the social influence on product adoption. Three types of social network influence are studied in Snijders (2001): structural effects on network dynamics, effects on network dynamics associated with covariates, and effects on behavior evolution. In our study, for simplicity, we assume the structure of the social network constructed in Stage I is stable over time, that is, links among consumers do not change over time. Hence, only the third type, effect on behavioral evolution, is relevant in our case. Formulations for evaluating the social influence attributes associated with several popular effects on behavior evolution are presented in Table 2. For example, the average friend effect is defined as the average degree of impact from linked contacts with similar behavior. Because social behavior $y$ (in this case choice behavior) of other consumers in the network changes over time, the average friend effect is updated at each time iteration $t$. This process will be discussed in more details in section ‘Case study of green product adoption: HEV ownership in California’.

Market surveys or empirical studies, such as interviews, are often needed to better understand the importance of social influences and which effect in Table 2 is the most relevant to the problem of interest. For instance, some individuals are highly influenced by their neighbors, coworkers, or other close contacts with whom they communicate daily, while other people are likely to trust suggestions and advice from their remote contacts, such as online product reviews, or blog posts from people with similar lifestyle. These effects are modeled as close links and distant links, respectively, in the small world network tested in our case study presented in section ‘Case study of green product adoption: HEV ownership in California’. In essence, the social influence attributes $N$ are modeled as a function of the binary link variable $L_{ij}$ based on network simulations.

Stage III: choice model estimation

In the traditional choice modeling framework, a predictive model of choice share $Q$ is established using Discrete Choice Analysis, DCA, which is based upon the assumption that individuals seek to maximize their personal consumer choice utility, $u$, when selecting a product from a choice set. The choice utility is derived by assuming that the individual’s (i) true choice utility, $u$, for a design alternative, $k$, consists of an observed part $W$, and an unobserved random disturbance $\epsilon$ (unobserved utility):

$$u_{ik} = W_{ik} + \epsilon_{ik}$$  \hspace{1cm} (4)

As shown in Eq. (5), the observed or deterministic part of the utility $W_{ik}$ is expressed as a function of consumer desired product attributes $A_{ik}$, of respondent $i$, alternative $k$; usage context attributes $E_{i}$; and consumer profile attributes $S_i$ of respondent $i$.

$$W_{ik} = W(\beta : A_{ik}, E_{i}, S_i)$$ \hspace{1cm} (5)

In this work, we introduce a new element, the social influence attributes $N_{ik}$, into the utility function as shown in Eq. (6).

$$W_{ik,t} = W(\beta : A_{ik,t}, E_{i}, S_i, N_{ik,t})$$ \hspace{1cm} (6)

Not only does $N_{ik}$ change for different respondent $i$, it is also time dependent of time variable $t$, which reflects the dynamic nature of social network influence: in each time period, consumers make choice decisions, which may change the social influence attributes of their linked network members. Hence, the coefficients $\beta$ are estimated based on the collected market data over a period of time assuming the growth of market share of a product is a result of social network impact. From the observed part of utility, $W_{ik,t}$, the probability $P_{ik}$ of an individual $i$ choosing a given alternative $k$, and the resulting choice behavior $y_{ik}$ can be estimated.

<table>
<thead>
<tr>
<th>Behavioral effects</th>
<th>Definition</th>
<th>Mathematical formulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tendency effect</td>
<td>Individual constants representing basic tendency</td>
<td>$N_i = C_1$</td>
</tr>
<tr>
<td>Average similarity effect</td>
<td>Average degree of consumer being similar to their friends</td>
<td>$N_i = \sum_{j} L_{ij} / \sum_{j} L_{jj}$</td>
</tr>
<tr>
<td>Total similarity effect</td>
<td>Total degree of consumer being similar to their friends</td>
<td>$N_i = \sum_{j} L_{ij} \cdot \frac{1 - y_{ij} y_{ij-1}} {\sum_{j} L_{jj}}$</td>
</tr>
<tr>
<td>Average friend effect</td>
<td>Average degree of impact from linked contacts with similar behavior $y$</td>
<td>$N_i = \sum_{j} L_{ij} \cdot \frac{1}{\sum_{j} L_{jj}}$</td>
</tr>
<tr>
<td>In-degree effect</td>
<td>Number of friends linked to a consumer</td>
<td>$N_i = \sum_{j} L_{ij}$</td>
</tr>
<tr>
<td>Out-degree effect</td>
<td>Number of friends linked from a consumer</td>
<td>$N_i = \sum_{j} L_{ij}$</td>
</tr>
</tbody>
</table>
The information flow in the three-phase diagram (Fig. 2) shows how the usage context $E$ and consumer profile $S$ are first mapped to the interpersonal links $L$ (Eq. (3)) and then to social influences $N$. In the last stage of choice model estimation, the product attributes $A$, together with consumer profile $S$ and usage context $E$, as well as newly introduced social influence attributes $N$, comprise the explanatory variables of a choice utility function. For each time period, Stages II and III are repeated to capture the changing dynamics of social influence; the loop will stop when $t$ reaches the maximum time period in given data.

The coefficients of the explanatory variables in the discrete choice models are obtained through the maximum-likelihood estimation using market data to capture consumer behavior and model consumer choices for new products. The procedure is similar to that of the conventional choice model estimation except that the coefficients of social influence attribute $N_{i,t}$ need to be calibrated to match with the sales data from multiple time periods. Determining the time dependent social influence attribute $N_{i,t}$ for each individual is challenging as often no empirical data of network links $L_{ij}$ exist at the individual (disaggregate) level to support the model estimation process. Instead, aggregate sales data $M_{r,t}$ throughout multiple time periods can be used to identify aggregated social influence attribute $N_t$ to ensure that the integrated choice model reflects the real market at the aggregate level.

**Case study of green product adoption: HEV ownership in California**

Alternative fuel vehicles have received increased attention in the past few years because of their potential to reduce greenhouse-gas emissions and oil imports and utilize renewable energy resources (Shiau et al., 2009). However, understanding consumer choices of alternative fuel vehicles is challenging because their preference construction process involves many aspects beyond traditional engineering considerations, which calls for a comprehensive modeling framework to incorporate social impact into engineering design. Taking PHEVs as an example, research in Axsen and Kurani (2009) has found that consumers attitudes towards green technology are strongly influenced by the social network. In this section, we present a case study of hybrid electric vehicle (HEV) ownership to illustrate our proposed choice modeling framework. We consider consumers being heterogeneous, differing in their preferences for vehicles based on their socio-demographics (e.g., income, age, region), the influence of social networks (e.g., neighbors, family), and the vehicle usage (e.g., miles-driven, transportation mode).

We use the National Household Travel Survey (NHTS) data as the market data for our model estimation. The data covers the demographic characteristics of households, people, vehicles, and detailed information on daily travel in the US for all modes (FHWA, 2009). The data was collected from a sample of US households and expanded to provide the national estimates of trips and miles by travel mode, trip purpose, and household attributes. Information about vehicles owned by sample households are recorded, including hundreds of car models from multiple years. The sample population of 41,330 respondents from California is used as data source for the case study presented in this section. We decided to focus on the data from California because it has the largest population of HEV owners, which is evident in the NHTS 2009 data.\(^1\) Vehicles purchased in brand-new conditions during 2002–2009 are selected for choice model estimation. It should be noted that in our current study, the impact of alternative fuel policies and other purchase incentives are not considered.

\(^1\) In the NHTS 2009 data, the market share of HEVs (number of hybrids/total number of sampled vehicle fleet) is 2.54%. Among 1300 2008-model-year cars, 15.54% of them are hybrids. With total population of 41,330 respondents from California, there are 4,887 new hybrids per 1000 residents for the model year 2008. As a comparison, in the real market, the total market share of HEV was 2.8 percent in 2009 (source: http://www.hybridcars.com/december-2009-dashboard/). And in California, there are 1.888 new Hybrids per 1000 residents, while the national average is about 0.820 new Hybrids per 1000 residents in 2008 (source: http://www.hybridcars.com/december-2008-dashboard-focus-production-numbers-25416/).
Social network simulation

In the first two stages of the proposed choice modeling framework (Stage I – social network construction and Stage II – social influence evaluation), a small-world network is employed based on the NHTS data from the California sample population.

In Stage I – social network construction, depending on the predefined average number of friends \( n \), each consumer is connected to its \( n \) nearest neighbors. All links are undirected, meaning all connections are mutual, \( L_{ij} = L_{ji} \), due to data limitation (when possible, directed links could be used to model network hierarchy during innovation spread, which could have a large impact on the diffusion process). Every existing link is then rewired to a random consumer with the rewiring probability \( x \). Fig. 3 shows the information flow in the social network simulation. The network parameters, such as the number of friends and the rewiring probability, are inputs to the social network simulation, while social influence attributes and the network properties, such as clustering coefficients and average path length, are outputs of this stage. The clustering coefficient is defined as the probability that two randomly selected friends of a focal consumer are friends with each other. In other words, it is the fraction of pairs of friends that are connected to each other by links. In general, the clustering coefficient of a node ranges from 0 (when none of the consumer’s friends are friends with each other) to 1 (when all of the consumer’s friends are friends with each other). Path in a social network is simply defined as a sequence of nodes with each consecutive pair in the sequence connected by a link. Things often travel along a path – this could be a passenger taking a sequence of airline flights, or a trend of adopting a new technology being passed from one person to another in a social network.

The calculation of the average path length reflects the average distance between two nodes randomly selected in the network. In a social network with short average path length, the social distance between the customers at two nodes is short, meaning the customers are more easily to be influenced by the connected customers. The clustering coefficient increases with the decrease of rewiring probability and the increase of number of friends – the more friends a focal consumer has, the more likely his/her friends are linked to each other. On the other hand, the average path length decreases with the increase of rewiring probability and the number of friends. In the following section, choice modeling results based on \( n = 10 \) and \( x = 0.01 \) (a common choice for social network in the literature Kossinets and Watts (2006), Stonedahl et al. (2010) and Watts and Strogatz (1998b)) are presented, followed by the sensitivity analysis on the impact of varying number of friends and rewiring probabilities.

The Small-World Network (Watts and Strogatz, 1998a) is a type of mathematical graph in which most nodes are not friends of one another, but can be reached from every other node by a small number of steps, as shown in Fig. 4(b). Literature has shown that many empirical networks exhibit the small-world phenomenon. Two other types of networks are provided for comparison: the regular nearest neighbor network in (a) and the random network in (c).

The small-world network offers a mechanism to represent interpersonal influences of both close and distant links within a social network – consumers are connected to their near “neighbors” (in their social space), as well as a small number of consumers far away from them. Such networks have two important characteristics: (1) they have a high clustering coefficient, that is, two randomly chosen consumers in the network who happen to be linked to another consumer have a high probability of also being linked to each other; and (2) they have a small path length, that is, the average distance between any two consumers in the network, measured as the number of links of the shortest path (with the fewest links) connecting them, is small. The popular notion that any two people in the world are connected by short chains of connections (that is, an average of six degrees of separation (Watts, 2004)) is a reflection of the short path lengths in the small-world networks. The interpersonal links \( L_{ij} \) (0 or 1) are determined by following the social distance evaluations and the threshold criteria in Eq. (2) and Eq. (3), respectively.

Following the Watts–Strogatz mechanism (Watts and Strogatz, 1998a) and using a random graph generation model that produces graphs with small-world properties, we create a small world network of California respondents from NHTS using NetLogo (Wilensky, 1999), an agent-based simulation tool. The geographic location, i.e. the latitude and longitude data mapped from the zip code information, is used as the attributes of social dimension for consumer \( \sum_{i} \). In Fig. 5, hybrid electric vehicle owners are shown in black, while the conventional vehicle owners are shown in gray. In the right figure, the dark blue lines represent the original links to nearest neighbors, while the light blue lines represent rewired links to random consumers in the network.

In product design, the small-world phenomenon implies that consumers not only consider the choices of close friends, but are also influenced by remote contacts, such as online reviews from people outside their regular social proximity. In
Stage II – social influence evaluation, we employ a variation of average friend effect described in Table 2, as shown in Eq. (7):  

$$N_{ij} = \frac{\sum_j L_{ij} y_{jt-1}}{\sum_j L_{ij}}$$  

(7)

where $y_{jt-1}$ is a binary variable that represents the choice behavior at previous time period $t-1$: 1 for a hybrid electric vehicle owner, and 0 for a conventional vehicle owner. This social influence attribute can be regarded as the percentage of hybrid electric vehicle owners in the focal consumer’s circle of friends. The simulated value of social influence attribute in Eq. (7) is passed as an input to the choice model. The evaluation process iterates for each of the time periods $t = 1 \ldots T$, to capture the dynamic nature of the social influence attribute.

**Estimation of discrete choice models**

Data from 13,802 respondents in California who owned vehicles of model years 2002–2009 are selected for building the discrete choice model. Note that 2002 is the first year HEVs appeared in the NHTS data. While vehicles from multiple years are considered in our choice modeling, consumer profile and usage context attributes are based on data collected in 2009. Therefore, the choice model estimation results shown below are valid under the assumption that there is no significant change in consumer profile and usage context from 2002 to 2009. Because the NHTS data provides no information about other vehicles considered by the respondents, three vehicles other than the one purchased are randomly selected from a set of 262 car models based on a uniform distribution to compose an individual choice set of four vehicles. McFadden (1978) has shown that a multinomial logit model estimate using choice sets composed of randomly selected members drawn
with a uniform distribution from the set of all choice alternatives will result in consistent estimates of the model parameters. Later in the sensitivity analysis, we study the impact of the number of vehicles in a choice set.

As shown in Fig. 6, seven consumer-desired vehicle attributes $A$ are selected including price, miles per gallon (mpg), vehicle origin, vehicle size, vehicle type, area, acceleration (torque/vehicle weight), and an HEV indicator. The attribute “price” represents the money respondents paid, excluding taxes, registration, trade-in value, etc. The miles per gallon comes from the combined mpg published by the US Environmental Protection Agency (EPA, 2008) and used as $100$/miles per gallon in the model specification to represent number of gallons required to drive 100 miles. Vehicle origins are categorized as domestic, European, Japanese, and Korean. Vehicle sizes are grouped into compact, midsize, large, and premium. Vehicle type includes mini (such as compact vehicles), car (such as sedans), sport utility vehicles (SUV), minivan, van, multi-activity vehicles (MAV), and pickup. Vehicle area is defined as the product of vehicle length and vehicle width, reflecting the general size of the car, while the power, i.e. torque, divided by vehicle weight, is used as an approximate measure of the acceleration feature. The hybrid electric vehicle indicator, is coded as 1 for hybrids, and 0 for conventional vehicles.

Gender, age, household income, number of children under age 18, education level, are included as consumer profile attributes $S$. Numerous combinations of consumer attributes are tested in the choice modeling process. In the final choice model, only three consumer profile attributes, household income, number of children, and education level, are statistically significant. As for the vehicle usage attributes $E$, the most commonly considered vehicle usage attributes miles driven daily is included.

The choice probability function of multinomial logit model is shown in Eq. (8). The log-likelihood function shown in Eq. (9) is maximized in Maximum Likelihood Estimation. The structure of the utility function is shown in Eq. (10), where interactions between $A$, $E$, $S$, and $N$ are explicitly modeled. Alternative specific constants (ASC) for each of the car models are not included in the utility function. While this may decrease the goodness-of-fit of the model, it allows choice prediction of new vehicles in product design.

$$\begin{align*}
P_{ik,t} &= \frac{\exp W_{ik,t}}{\sum_{j} \exp W_{ij,t}} \\
\bar{\beta} &= \arg \max_{\beta} \left( \sum_{t=1}^{T} \sum_{i=1}^{I} \sum_{k=1}^{K} y_{ik,t} \log P_{ik,t} \right) \\
W &= \beta_A \cdot \mathbf{A} + \beta_{AE} \mathbf{A} \cdot \mathbf{E} + \beta_{AS} \mathbf{A} \cdot \mathbf{S} + \beta_{AN} \mathbf{A} \cdot \mathbf{N}
\end{align*}$$

Following the choice modeling procedure in stage III described in section ‘Choice modeling framework considering social impact’, interactions between consumer-desired product attributes $A$, consumer profile $S$, usage context attributes $E$, and the social influence attribute $N$ are explicitly modeled in the utility function. The coefficients for all attributes and their interactions based on a multinomial logit model estimation (MNL with $N$) are compared to the estimation results from a multinomial logit model without “social network influence” (MNL without $N$). In this research, we propose to measure the “HEV preference”, which reflects a consumer’s attitude toward new hybrid technology, as a collective effect of all utility terms involving the hybrid electric vehicle (HEV) indicator, that is, HEV, interaction between HEV and high education level, interaction between HEV and fuel price, and interaction between HEV and social impact, as shown in Table 3.

Goodness-of-fit measures based upon the log-likelihood of the converged model, such as the likelihood ratio index $\rho^2$ (also known as pseudo R-square), reflect how well the estimated model predicts actual individual choices in the data set. Higher values of $\rho^2$ indicate better predictions of the choices. As shown at the top of Table 3, a slightly higher $\rho^2$ value of 0.1459 are achieved using the MNL model with social influence attribute $N$ versus the MNL model without $N$. The $\rho^2$ value of 0.1459 means that the MNL model with $N$ has a 14.59% increase in the log-likelihood function, compared to the initial
model with zero information. Even though the proposed choice modeling framework captures the systematic heterogeneity of consumers by explicitly modeling the consumer attributes including the social influence attribute as model inputs, the differences between the two models are relatively small. A closer look into the data set revealed the underlying reason: HEV owners comprise a small percentage (7.59%) of the whole sample population, resulting into a limited number of observations with non-zero social influence attribute values (0.67%).

From the results of the MNL including N attributes in modeling, we note that all coefficients are statistically significant at the 0.01 level. The coefficients for price/income and 100/miles per gallon are negative as expected. Only three consumer profile attributes, household income, number of children, and education level, are statistically significant. A negative estimator for 100/miles driven daily + miles driven daily indicates that the usage context attribute has a negative impact on consumer preference on the inverse of MPG measure, gasoline needed to travel 100 miles in this case. Similarly, the positive sign of minivan * number of children suggests that households with children prefer minivans, which is consistent with our expectations. However, SUV * number of children has an unexpected sign, which is likely due to its correlation with other variables included in the model. In cases where many attributes are correlated with each other, analyses of substitution effects for new prediction model. In cases where many attributes are correlated with each other, analyses of substitution effects for new prediction.
two models are identical. This result confirms that social influence is a significant determent of hybrid electric vehicle choice and contributes to explaining the choice heterogeneity.

Sensitivity analysis

In the previous section, the social network is constructed based on the assumption that consumers have 10 linked friends on average, a rewiring probability of 1%, and a choice set of 4 alternative vehicles. To better understand the model dependence upon these assumptions, the sensitivity of the proposed choice modeling in response to changes in average number of friends $n$ and rewiring probability $a$ is tested under three choice set scenarios: choice set size $= 4, 20,$ and 100. For each of the choice set scenarios, three discrete values, 5, 10, and 15, are selected for the average number of friends, while four values of the rewiring probability of 0%, 1%, 2%, and 10%, are tested. A few observations can be generalized from the sensitivity results. First, we note that the increasing number of friends resulted into a stronger social influence. This is because the social influence attribute in Eq. (7) follows a percentage representation: the more friends a consumer is linked to, the larger the social impact, with all other things being equal. This difference becomes clearer with the increasing choice set size, because the randomness decreases with the increase of number of alternatives in individual choice sets. Second, no clear distinction in social influence value is seen between different choice set scenarios, suggesting that the choice set size does not have a significant impact upon parameter estimation of social impact in our case study. However, a closer look into the log likelihood at convergence and under scenarios with different choice set size shows that it matters in terms of the model fit, with $\rho^2$ averages to 0.1463, 0.0929, and 0.0687, respectively. Last but not least, the results from our test case showed no conclusive relation between rewiring probability and social impact.

Choice share forecasting

To illustrate the potential in forecasting new product adoption using our proposed choice model, we provide projections of HEV choice share from 2002 to 2042 under different scenarios (each with three simulation runs) in Fig. 7. The HEV choice shares from 2002 to 2008 are assessed based on the proposed choice model built using the NHTS data. The HEV choice shares from 2009 onwards are forecasted and compared between two models of MNL with $N$ and MNL without $N$. As shown in Fig. 7, predicted choice share from model MNL with $N$ and model MNL without $N$ are consistent from year 2002 to 2008, while choice share forecasting diverges significantly starting from 2010 because the MNL model without $N$ does not take into account the social influence. In forecasting choice share post-2009, we use the same social network simulation model.
as described in section 'Social network simulation', assuming that an average of 7% of the consumer population, following a Poisson distribution, enter the market each year to shop for a new car. This assumption is supported by historical data from 2002 to 2008. With the increasing value of social influence attribute, the HEV choice share is forecasted to grow significantly in the next decades with a gradually decreasing growth rate, while the forecast from MNL without \( N \) stays constant due to the exclusion of dynamic social influence attributes in the utility function. Fig. 7 shows an S-shaped curve from MNL with \( N \) for HEV adoption in the future. The fact that the curve follows an S-shape is consistent with the product diffusion literature (Bass, 1969). The zigzagging in the curve reflects the assumption that the number of new consumers entering the market varies each year following a Poisson distribution. It should be noted that the forecast of HEV over time is intended to illustrate the potential of the proposed method in dynamic demand estimation rather than attempting to provide an accurate prediction of HEV adoption in the next 30 years. There are a number of factors which have not been considered in this forecast, such as gas price trend, inflation in vehicle price and consumer income, new HEV models entering the market and technology advancements.

Conclusions

In this work, we propose an integrated choice modeling framework considering social impact for forecasting new product adoption. To the authors’ best knowledge, this is the first attempt in engineering design research to develop analytical techniques that integrate discrete choice models with social network simulations at the individual consumer level to address simultaneously the interactions between product, consumer, and social influence in product design. Modeling social context within a choice modeling framework introduces a new dimension for understanding consumer attitudes towards new products, and at the same time, provides engineering designers the flexibility of using such models to guide engineering design decisions by optimizing the product attributes in the model. Together with the engineering performance and cost models, consumer choices can be linked to engineering design decisions, such as battery size and fuel economy target, as illustrated in He (2012). This integration offers a dynamic view of consumer choice of new products, in which consumer preferences may change over time due to social influence.

A case study of hybrid electric vehicle owners in California illustrates the potential benefits of the proposed methods in supporting the design of green products. Alternative fuel vehicles have received wide attention lately, and HEVs/PHEVs appear to be the most promising technologies at this time given recent technology developments, government incentives, and importantly, social impact on consumer choice of new products. Before these new vehicles enter the US market, forecasting their full market potential is of great interest to the manufacturers, and government agencies. From a broader system point of view, such choice models can be further integrated into a multi-agent energy market simulation framework to study the impact of consumer vehicle choice on the future electric generation, transmission, and distribution system.

It should be noted that our paper is focused on presenting the general framework and procedure of the proposed methodology with an illustrative case study demonstrating the potential usage of the approach. Due to the lack of real social network data, assumptions had to be made for network construction and social influence evaluation. These assumptions should be validated in the future based on empirical studies of social network characteristics to ensure the accuracy of model predictions. In addition, there exist multiple channels and activism mechanisms of social influence, e.g., advertisement, education programs, and online forum that are beyond the social network influence modeled in this research. Future research will be aimed for integrating findings from social behavior studies in understanding social influence processes such as compliance, identification, and internalization into the modeling of social influence on new product adoption. Due to the independence of irrelevant alternatives assumption, the multinomial logit model presented in this work has limitations which warrant future research investigation by implementing other choice modeling techniques such as the Hierarchical Bayes mixed logit model. Moreover, there are potential confounding issues because both the social impact term \( N \) and the choice utility \( W \) is a function of consumer profile attributes \( S \). To extend and refine the current work, the two-stage Berry, Levinsoh, and Pakes method in Walker et al. (2011) could be considered to correct the endogeneity in a choice model with confounding attributes.

Acknowledgements

Grant supports from National Science Foundation (CMMI-0700585 and DUE-0920047) and ISEN (Initiative for Sustainability and Energy at Northwestern) are greatly appreciated.

References


