Using Discrete Choice Models to Estimate Non-market Values: Effects of Choice Set Formation and Social Networks

by

Minjie Chen

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Agricultural and Resource Economics

Department of Resource Economics and Environmental Sociology
University of Alberta

Abstract

Discrete choice models are often used to estimate non-market values. In standard models, individuals make choices considering all possible alternatives. However, in reality, the set of alternatives individuals consider may differ. Moreover, these choice sets may be influenced by the individual's social networks. For instance, Romeo was going to go to the beach; however after talking to Juliet he is also considering the mountains. Recent research has demonstrated that ignoring the choice set formation (CSF) process leads to biased estimates of non-market values. This paper develops a discrete choice model in which the choice set faced by a decision-maker is influenced by her social network. In the model, a network parameter denominated by social propensity determines the weight a decision-maker places on her network when determining what alternatives to consider. We use Monte-Carlo experiments to investigate the effects of ignoring social networks when modeling CSF. We find that when social propensity is relatively low, CSF models that ignore social networks do not lead to significant bias in welfare estimates. However, as social propensity increases and reaches a certain threshold, welfare estimates that ignore networks are significantly biased and estimates of welfare change are significantly higher than the true welfare change.

Dedication

to Shihui Li

Acknowledgments

First and foremost, I would like to thank my family. My Mom and Dad, thank you so much for the endless love and constant support. My lovely sister Minrui, I am really so happy to know that you are going to have a baby, and can't wait to see my niece. Thanks also to all of my extended family for your support and encouragement.

All my friends from the Department of Economics, Zhaowei, Zihan, Shihui, Tianqi, Yue, Xinglin. Thank you all for the happy time we spent together here in Canada. Thanks to Xuecheng and Ding, who gave me the shelter when I came here for the first time and could not find a place to stay. My office mates, Ryan, Yichuan, Angela, Darskatar and Haoluan. It is my great honor to know all of you, and will remember all those beautiful days we stay in that no-window office. Many thanks to all my classmates at the Department of Resource Economics and Environmental Sociology. I really enjoyed the entire study here with you, and have started to miss those days. My teammates in the soccer team, thank you all for every moment we ran and laughed on the pitch. You helped me survive from several suffering days.

My supervisor Dr. Bruno Wichmann. Without your constant help, supports and encouragement, it is just impossible for me to finish this thesis. No single word can clearly express my appreciation for all the work you have done. Special thanks to Dr. Vic Adamowicz. Thank you for your talented guidance and constant encouragement. I really enjoyed every moment talking with you and it always gets my ideas refreshed. Thanks to Dr. James Rude for all the comments on my thesis and they are so helpful for me to write a better thesis. Also thanks to Feng, Sandeep, Henry, Brent, Scott and Sven, you are always reachable and helpful on any questions I encountered either in the class or in my thesis writing.

Finally, thanks to all the administrative staff, especially Robin, Brett, Maria and Gwen. My thanks also go to the others. Your great work and dedicated supports are highly and sincerely appreciated.

Contents

List of Tables

List of Figures

Terms and Abbreviations

τ_0 perceived average quality of alternatives. It is considered as a threshold of alternative selection.

τ_1 average social propensity describing to what extent an individual interacts with others in the social network.

ε random component in random utility function, and follows a Gumbel distribution.

ξ random component in availability function, and follows a logistic distribution.

W sociomatrix, an $n \times n$ row stochastic matrix describing links within a social network.

CSF choice set formation.

CV compensating variation.

DGP data generation process.

IAL independent availability logit.

MNL multinomial logit.

RUM random utility maximization.

SCSF choice set formation process with social interactions.

WTP/WTA willingness to pay/accept.

Chapter 1

Introduction

Non-market valuation of public goods has been an integral component of environmental and resource economics for decades. Research methods applied to this area are dominated by discrete choice models, especially with foundations in random utility maximization (RUM) theories. RUM-consistent discrete choice models capture individual taste heterogeneities that are usually ignored in traditional consumer theories, but also provide a way to identify the marginal value of properties change in public goods and services, such as an increase of travel cost or improvement in environmental quality. Although these properties often are not explicitly priced in markets, they may significantly effect individuals' willingness to pay (WTP) or willingness to accept (WTA). Therefore, revealing non-market values of goods and services properties becomes an important task of public agencies making policies, especially those concerning the public provision or regulation of environment goods and natural resources. There are several empirical challenges in estimating non-market values. This chapter outlines the broad context that our current research lies in. We start by introducing the problem. After that, we show the novelty and importance of our contribution to the literature, which is followed with a brief discussion about how we approach the problem. A summary of the major findings and how we organize the study are put at the end of this chapter.

1.1 Unordered multiple choice problems

The discrete multiple choice problem primarily concerns a decision maker[1] choosing one alternative from a choice set containing multiple alternatives. A broad class of problems, which is more frequently encountered by individuals in daily lives, involves choosing from a set of *unordered* alternatives[2]. For example, an urban commuter decides a travel mode to work; an angler chooses a lake or river site for fishing; a job applicant selects an offer from multiple companies; and a family discusses a summer vacation destination. These unordered alternatives are assumed well-defined with mutual exclusivity, exhaustiveness and finite in numbers (See discussions in Train, 2003, p.15). Environmental and resource economic analysis often focuses on non-market choices like recreational travel, angling, etc., while other literature examines choice in market environments including product choice. In either case the behavior of individuals in the multi-attribute choice environment is being examined, where attributes are properties of goods and services that differ from one alternative to another. For instance, a set of lakes may be characterized by multi-attribute vector including lake size, travel distance, biomass, etc., and attributes of urban travel modes may include fares, travel time and comfort.

1.2 Choice set formation (CSF) process

Early studies of modeling unordered multiple choice problems often start with a multinomial logit model (MNL), which is formalized by McFadden (1974). As the archetype of a series of RUM-consistent models, the MNL model assumes decision makers face a feasible choice set[3] containing the *same* group of alternatives from which they make

[1]It can be an individual or a group of members, such as a family or a firm. These terms are interchangeable throughout this thesis. See discussions in Ben-Akiva and Lerman (1991, p.33).

[2]Refer to Greene (2008, p.842) for a discussion. Readers interested in *ordered* choice problems and their modeling methods are directed to p.831 of the same book.

[3]McFadden (1974) assumes this feasible choice set is a nonempty subset of a universal choice set.

choice decisions by considering all alternatives through a RUM process. The MNL model provides an appealing way to model unordered discrete choices due to its consistency with RUM, tractability in econometric estimation, and simplicity in computation. However, describing individual choice behaviors within a context of predetermined feasible choice sets is questionable. It is often hard for researchers to exogenously predetermine a choice set containing alternatives that are feasible to every individual. Attempting to do so in non-market valuation often results in a misspecified choice set. That is, this predetermined choice set will probably contain alternatives infeasible to some individuals while miss alternatives feasible to others.

In reality, the set of alternatives a specific individual truly considers is often different from that of others. This is a challenge because choice sets misspecification can lead to significant bias in model estimation and welfare measures (see discussions in Chapter 2). In fact, feasible choice sets are usually endogenously formed by individuals themselves rather than predetermined by researchers. These endogenous behaviors are called CSF processes (Manski, 1977), and are observable to individuals but not to researchers. This unobservability to the researcher indicates that choice sets derived from such processes should be considered probabilistic. Assuming choice set certainty may misdirect further analyses. Therefore, in solving non-market valuation problems of unordered choices, a reasonable task for researchers is not to try to predetermine a feasible choice set universally applicable to all individuals, but is to understand the mechanism of CSF process and appropriately infer the probability of a detected choice set being true.

1.3 Social interactions

Statistical inference about choice sets probabilities is a complex task. Part of the challenge comes from the fact that individuals' CSF processes may be connected with each other through a social network. Previous research (Manski, 1977; Swait and Ben-Akiva, 1987a,b; Li, Adamowicz and Swait, 2012, etc.) however consider decision makers as *socially isolated*, ignoring the important fact that human beings are often *socially connected*. Social traits of individuals have long been demonstrated important to shape choice behaviors in studies of psychology, sociology, behavioral and experimental economics, and many other areas. Specific to the CSF discrete choice models where the choice set is inherently unobservable to researchers, assuming that choice sets are formed in social isolation probably does not reflect actual social behavior. It is therefore reasonable for us to think that the CSF process should incorporate social motivations such that choice set probabilities reflect social effects.

In this thesis, we explore the role of social influences in the CSF processes. This social influence may generally be grouped into two broad types. One is objective while the other is subjective. Objective reasons often include individual income and information asymmetry, etc. Personal income reflecting market price acceptability obviously constrains the availability of alternatives to be considered by individuals (Li, Adamowicz and Swait, 2012). Information asymmetry causes limitations on individual judgments under imperfect information and uncertainties. People identifying these limitations tend to seek for help from others each of whom possesses a piece of fragmented information. For example, an angler may not know fish species and policy regulations on a specific lake. Communication with others, however, improves information quality and thus improves choice quality. The existence of large numbers of substitutes also leads them to intentionally compress the feasible choice set based on their own characteristics in market and non-market behaviors. These objective facts have been frequently tabulated within a social isolation setup. For example, the socioeconomic status in Swait and Ben-Akiva (1987b); price constraint in Li, Adamowicz and Swait (2012); awareness, information and/or familiarity in Parsons and Hauber (1998), Peters, Adamowicz and Boxall (1995) and Haab and Hicks (1997), etc.

However, "economists tend to be deeply skeptical of subjective statements" (Manski, 2000, p.131). These subjective statements usually include peer pressure (passive), altruism, egoism and envy (active), all of which have been extensively studied in the literature of psychology and behavioral economics, and are demonstrated having significant effect on shaping individual choice behaviors (see Merton, 1957; Becker, 1974; Akerlof, 1980; Case, 1991; Manski, 1993a,b and 2000; Brock and Durlauf, 2001; Neilson and Wichmann, 2014). For example, peer pressure from neighbors may result in an individual considering a car that originally was expensive to him. Altruism may cause an individual to vote for a public policy improving recreational amenities irrelevant to him, and the only reason he may vote "yes" is because of a friend that likes the amenities. This study considers the role of social networks in shaping social interactions among individuals, which we believe are potentially able to aggregate the above-mentioned objective and subjective reasons, and may implicitly capture the efforts individuals make to reduce transaction costs. Specifically, we propose a CSF process incorporating social interactions (SCSF) to reveal how individuals deal with different objective and subjective reasons simultaneously in forming their individual-specific choice sets.

1.4 Purpose and contribution

Recent research has demonstrated that ignoring individual's CSF process leads to biased estimates of non-market values (see Swait, 1984; Swait and Ben-Akiva, 1987a,b; Peters, Adamowicz and Boxall, 1995; Parsons and Hauber, 1998; Haab and Hicks, 1997; Li, Adamowicz and Swait, 2012; Truong, 2013). Each approach applied previously has its own limitations. More flexibility can be added to standard CSF processes since different specific constraints imposed on those models result in very different estimates and welfare measures. The central purpose of this study is to examine social constraints[4] on individual's CSF process through an explicit modeling approach.

The contribution to the literature is twofold. First, to the best of our knowledge,

[4]Note the difference from socio-economic constraints discussed in Swait and Ben-Akiva (1987a).

all previous studies on unordered choice problems with explicit two-stage models (see Manski, 1977) typically considered individuals as socially isolated in the formation of their individual-specific choice sets. Some studies, either applying one-stage or two-stage models (e.g. Brock and Durlauf, 2001; Soetevent and Kooreman, 2007; Barucci and Tolotti, 2012; Neilson and Wichmann, 2014; Wichmann, 2014) investigate the influence of social effects on discrete choice analysis, but they all focused on the study of models' RUM-bases (utility/preference). Modeling unordered choice behavior without addressing the CSF process *a priori* seems to be inappropriate as the literature has suggested that choice set formation is indeed important. This study adds to the literature by considering the effects of social networks into individual's CSF process. Second, the literature tends to separate the studies built upon traditional one-stage models (such as MNL model in McFadden (1974)) from those based on explicit two-stage models. Our study, a two-stage model with SCSF process, finds a connection between the two and provides a possibility to integrate them. Specifically, we show that for the specification we employ, if social interactions among individuals are strong enough, traditional explicit two-stage models with CSF processes that do not capture social interactions in forming choice sets actually collapse into one-stage models without considering CSF processes.

1.5 Summary of the approach and results

This study develops a two-stage unordered discrete choice model in which the CSF process is assumed to be influenced by social interactions. Based on that model, we investigate the bias of standard CSF models due to the misspecification of network effects. Specifically, we use Monte-Carlo simulations to study the bias of parameters estimation, distribution of probabilistic choice sets and welfare measures in standard CSF models, when the data is actually generated by a SCSF process. We extend the traditional two-stage independent availability logit (IAL) model[5] to incorporate social interactions in the CSF process. These social interactions are described by a social network matrix. Individual's CSF process depends on own perceptions of goods attributes and on perceptions of one's social network. The strength of the network effect is captured by the level of social propensity (see subsection 3.1.2). Given parametric assumptions imposed on random components of the model, a series of Monte-Carlo experiments are conducted. The experiments are as follows. First, we use the SCSF model to generate "true" data. Second, we test and compare parameter estimates and probabilistic choice set distributions under two situations, one in which the SCSF model is estimated and another in which a standard CSF process is estimated. In order to test the bias of welfare measures, we hypothetically assume a policy change concerning an exogenous quality improvement on one of the alternatives. Monte-Carlo experiments enable us to observe the draw of random components, therefore, we are able to compute and compare welfare measures (WTP or compensating variations) from models under the above two situations to the true welfare measure.

The main results are as follows. First, parameter estimates from the correctly specified model (with SCSF process) in general outperform those of the incorrectly specified model (with standard CSF process), but not always. Specifically, when social propensities are weak, both models estimates can well approximate the true parameters with proportional root mean square errors under 5%. In these cases, although the correctly specified model has better approximation in general, the incorrectly specified model does not lead to very significant bias. However, if social propensities are strong, the incorrectly specified model delivers very biased estimates while the correctly specified model continues showing consistency with the true parameters. The bias of the incorrectly specified model reaches 1,952% under certain conditions.

Second, estimated distributions of probabilistic choice sets are quite different be-

[5]This model was first developed by Manski (1977), Swait (1984), and Swait and Ben-Akiva (1987a,b) and further explored by Li, Adamowicz and Swait (2012)

tween the two models. The correctly specified models can well approximate the distributions of probabilistic choice sets from the data generation process. This is not true, however, for the incorrectly specified model. This phenomenon is directly connected with our model estimation since computation of these distributions is largely based upon estimated parameters. Interestingly, we find that if social interactions are strong enough, the traditional two-stage model where social interactions are ignored from individual's CSF process ultimately collapse into the standard MNL model.

Third, given the hypothetical quality improvement we applied to one of the alternatives, we find that the welfare measures estimated from the correctly specified model provide good approximations to the true welfare change. However, welfare estimates from the incorrectly specified model have bias ranging from 17% to 30% when social interactions are strong. We also find that in general all the above-mentioned findings do not vary significantly across social networks with different densities (i.e. number of links). Hence, we argue that the sizes of social networks are not a major component of the standard CSF model bias. Rather, what really matters is to what extent individuals care about these networks.

1.6 Research organization

Following this introduction, the thesis is organized as follows. Chapter 2 provides a literature review on how previous studies address the problem of modeling unordered choices, the CSF process and what they have attempted to consider social interactions in discrete choice analysis. Chapter 3 sets up the empirical model for our study. We explain in detail how we model social interactions, and how we incorporate them into our study. Chapter 4 performs a series of Monte-Carlo experiments to test the empirical models. Chapter 5 presents the experimental results. In Chapter 6 we conclude our study and briefly lay out potential problems for future research.

Chapter 2

Literature Review

This chapter provides the context and theoretical background of applying choice theories to model the unordered individual choice behaviors for non-market valuation purposes. The literature addressing this type of choice problem discusses both the mechanism of making unordered choices and the relevant estimation techniques. We organize this chapter as follows. In section 2.1, we make a brief introduction to the framework of unordered discrete choice theories. Section 2.2 and 2.3 review the dominant theoretical and empirical models, their evolution path, and present some highlighted evidence from previous empirical studies. In section 2.4 we address the social aspects of discrete choice analysis. This structure of organization intends to make it clear the fundamentals we need to postulate the central concerns of this study, which will be theoretically expanded in Chapter 3 and empirically tested in Chapter 4.

2.1 Introduction

As we previously noted in Chapter 1, an unordered discrete choice problem primarily concerns the decision maker choosing a single utility-maximizing alternative from a choice set consisting of multiple unordered alternatives. The alternatives comprising these choice sets are assumed well-defined, and decision makers are always able to compare them and choose the one that maximizes their random utilities. To the central concerns of this research, we are especially interested in how decision makers generate their feasible choice sets in such a choice problem. Two types of choice sets will be encountered in the following discussions. They are either global if they contain all possible alternatives (McFadden, 1974, 1981; Train, 2003), or local (individual specific) if they are some generated nonempty subsets of the global choice set through a CSF process. We call the former type as the global choice set, and alternatives in it are global alternatives[6]. The latter type are the feasible choice sets to individuals. The structure, such as the number of alternatives contained or their categorizability[7], is demonstrated able to influence the probability of an alternative being selected. Under certain circumstances, this shift in choice probability is large enough to change individual welfare measures. Since the non-market valuation of policy changes ultimately considers an aggregation of individual welfare measures, exploration into individual choice behaviors is essential to the successful modeling of social welfare.

Models inferring discrete choice behaviors are RUM-consistent if their intuitions are exclusively established upon random utility maximization theories. Literature applying these models dates back to Thurstone (1927) and Luce (1959), with primary focus on psychological choice problems. Random utilities are slightly different from the traditional consumer utilities, where the latter usually assumes representative utilities and the former assumes that utilities are decomposable with a representative utility V plus an additive random component ε capturing some unobservable idiosyncratic utilities. On one hand this decomposition results in randomness in utilities and the choice of alternatives being probabilistic, on the other hand it provides a powerful way to mimic individual choice behaviors if their choice sets are known with certainty.

[6]Global alternatives are those that exist for all individuals in a community regardless of whether they are available for a particular individual. For example, in a local community the residents only have three models of cars to choose from. These cars may not be affordable for every individual, or individuals not necessarily like all types of cars.

[7]This means that the alternatives included in a choice set can or cannot be grouped into subgroups. An usual treatment of this may use a nested logit model. See Greene (2008, p.848).

Mathematically, the random utility for an individual n to choose alternative j from a known feasible choice set B_n containing J alternatives can be represented as $U_{nj} = V_{nj} + \varepsilon_{nj}$, where $j \in B_n$ and $j = 1, 2, ..., J$. V_{nj} is the representative or systematic term observable to researchers, and ε_{nj} is a stochastic component reflecting unobservable individual and alternative specific heterogeneities. V_{nj} is usually assumed to be a linear function of measured attributes of alternative j, and $V_{nj} = V_n(x_j) = x'_{nj}\beta$. x_{nj} is a vector of measured attributes of alternative j for individual n. β is defined as the corresponding vector of unknown parameters. Random utility theory tells us that alternative j will be chosen if and only if $U_{nj} \geq max(U_{nk}) \, \forall k \neq j$ and $k, j \in B_n$. The unconditional probability of choosing j is then $P_n(j) = Prob[U_{nj} \geq max(U_{nk})]$. Insert the random utility equation, and an expansion then leads to

$$\begin{aligned} P_n(j) &= Prob(V_{nj} + \varepsilon_{nj} \geq V_{nk} + \varepsilon_{nk}) \\ &= Prob(\varepsilon_{nk} - \varepsilon_{nj} \leq V_{nj} - V_{nk}), \forall k \neq j \text{ and } k, j \in B_n \end{aligned} \tag{2.1}$$

Different unordered discrete choice models are able to be derived from equation (2.1) given a specific parametric assumption on the joint distribution function $f(\varepsilon_{n1}, \varepsilon_{n2}, , \varepsilon_{nJ})$ of the random components ε_{nj}.

2.2 Multinomial logit model and its extension

2.2.1 Multinomial logit model

The MNL model is perhaps the most prevalent RUM-consistent model studied in the literature of unordered choice analysis. It is formalized in McFadden (1974), where the author assumes an additive disturbance ε_{nj} that is independent and identically distributed (*i.i.d.*) with a Gumbel (Type I Extreme Value) distribution[8]. Specifically, if $\varepsilon \sim \mathbf{G}(0, 1)$, where 0 is the location parameter and 1 is the positive scale parameter,

[8]See Ben-Akiva and Lerman (1991, p.104) for basic properties of the Gumbel Distribution.

the unconditional choice probability $P_n(j)$ in equation (2.1) can be further derived for the MNL model as

$$(2.2)\quad P_n(j) = \frac{exp(V_{nj})}{\sum\limits_{k \in B_n} exp(V_{nk})}$$

This simple MNL model provides an appealing way to approximate empirical models of individual's discrete choice behavior and non-market valuation problems. But we also need to notice that, the success of this RUM-consistent MNL model inevitably depends on the correct specification of the choice set B_n since its dimension affects choice probabilities. In the MNL model, each individual is assumed to face the same set of alternatives, and this choice set is known with certainty to researchers. Empirically, "the sample is obtained by a sequence of independent trials, with or without *replications* in which a sequence of choices are observed for individuals with the same measured attributes and alternatives sets" (McFadden, 1974, p.107). In this process of random trials, the MNL model ignores the fact that in reality the choice set B might be available to a specific individual but not necessarily to another. In other words, when individuals are presented with an arbitrary choice set B by the researcher, some alternatives may never even be considered by individuals (B would have too many alternatives in this case) or some relevant alternatives may not be included in B (too few alternatives in this case). This choice set formulation may significantly affect utility maximizing choices.

From the above discussions we can see that, the fundamental assumption on feasible choice sets in the MNL model is quite implausible. In fact, there always exists some endogenous CSF process that is observable only to the individual but not to the researcher. This hidden process has at least two direct effects on the empirical study of unordered discrete choice problems. First, if a strictly preferred alternative is forced to be excluded in the CSF process due to certain reasons for a specific individual, then the observed sequence of choices by researchers tends to violate the basic assumption of rationality in random utility theory. Secondly, from equation (2.2) we can see that, extra inclusion of alternatives leads to underestimated unconditional choice probabilities for the researcher since the denominator $\sum_{k \in B_n} exp(V_{nk})$ is a summation of exponential systematic utilities on all alternatives where $k \in B_n$. Incorrectly computed choice probabilities $P_n(j)$ may lead to very biased inferences about the marginal utility and welfare measures valued upon external policy changes.

2.2.2 Two-stage model

Though including a CSF process in the modeling of unordered discrete choices looks so important, extensive studies concentrating on it start only very recently. A great number of early work (Debreu, 1960; Luce and Suppes, 1965; Tversky, 1972; McFadden, 1974, 1981 and 2001; and recently Hensher and Greene, 2002, 2003, etc.) tends to put more weight on relaxing the assumption of independence from irrelevant alternatives (IIA) introduced by Luce (1959). The major breakthrough focusing on the CSF process is attributed to Manski's (1977) seminal study, where the author alternatively extended the MNL model into a separable two-stage consider-then-choose model, and attempted to "reformulate the random utility model at a position logically prior to that hitherto adopted" (Manski, 1977, p.230).

Let us assume there is a finite population N faces a *primitive finite set*[9] B containing J alternatives, and each of the population chooses one alternative to maximize random utility. The CSF problem of individual n at the first stage in Manski's two-stage model concerns drawing a nonempty subset $C \subseteq B$ according to a probability measure $Q_n(C)$ that describes the CSF process. In the second stage, a random utility function defined on choice set C rather than on the *primitive finite set* B itself is postulated to derive a RUM-consistent model. The outcome probability of this two-stage choice process is

[9] Here the *primitive finite set* shares similarity with the *universal choice set* defined in McFadden (1974) or the global choice set we defined in this thesis.

mathematically described as

$$(2.3)\ \ P_n(i) = \sum_{i \in C, C \subseteq B} Q_n(C) P_n(i \mid C)$$

where i is an alternative belonging to C; $P_n(i)$ is unconditional probability of observing i being chosen from the perspective of a researcher unknowing of the specific alternatives forming individual n's choice set; and $P_n(i \mid C)$ is the conditional probability of choosing alternative i given C is individual n's feasible choice set. This two-stage structure has extensive implications on modeling unordered discrete choice problems and estimating welfare measures. Treatments of $P_n(i \mid C)$ are unanimously coherent throughout the literature studying CSF process since it is consistent with McFadden's (1974) MNL model. Differences (progresses) mainly lie in using some different econometric approaches. In general, very similar to equation (2.2), $P_n(i \mid C)$ can be defined as

$$(2.4)\ \ P_n(i \mid C) = \begin{cases} \frac{exp(V_{ni})}{\sum_{h \in C} exp(V_{nh})}, & \text{if } i \in C \\ 0, & \text{if } i \notin C \end{cases}$$

where V_{ni} is the systematic indirectly utility of U_{ni}. Specifically, as defined in section 2.1, U_{ni} takes the additive random utility form with

$$(2.5)\ \ U_{ni} = V_{ni} + \varepsilon_{ni} = x_{ni}^{'}\beta + \varepsilon_{ni}$$

where ε_{ni} is *i.i.d.* Gumbel distributed with the scale parameter fixed at unity. Equation (2.4) differs from equation (2.2) by including a zero probability if alternative i is eventually not included in n's consideration set. This is possible since the two-stage model incorporates a CSF process where a situation exists that alternatives do not belong to a specific true choice set.

2.3 Empirical approaches of two-stage model

How to specify of $P_n(i \mid C)$ in equation (2.3) and (2.4) has few disputes throughout the literature, while discussions about $Q_n(C)$ are still arguable due to endogeneity and probabilistic properties. The empirical approaches applied to generate individual specific choice sets are often very different in the literature of valuing non-market goods, especially in studies of recreational site choice and transportation mode choice. But we must note that, no matter how broad or narrow these arguments and approaches are, they all aim to find a way to account for the true choice sets for each decision maker. Among of them two approaches are frequently discussed: survey questions and explicit CSF models.

2.3.1 Survey approach

In a few empirical studies of recreation site choice, researchers apply survey questionnaires to collect information that helps narrow down the number of alternatives potentially faced by individuals. Peters, Adamowicz and Boxall (1995) use direct elicitation of choice set by means of a survey question asking recreational anglers to identify the set of sites they consider or are aware of when making fishing trip decisions. This set of sites identified by the respondent himself is considered as their actual choice set. As an analogy to the explicit modeling of CSF process in equation (2.3), we may think that, in this survey approach $Q_n(C)$ is actually implicitly embedded in the survey question. With the survey responses at hand, researchers would know C with certainty, and $Q_n(C) = 1$ in the second-stage computation of unconditional probability $P_n(i)$. Using data from the Southern Alberta Sportfishing survey collected in 1991, the authors compare per-trip welfare measures arising from water quality changes (such as to close a popular site[10], plant trees and increase fish stock) in three different CSF models. They

[10] Results from using site closures to estimate choice probabilities and ultimately the welfare changes indicate that the IIA assumption in RUM-consistent models can be frequently violated in empirical

find that welfare measures with individual specific choice sets are quite different in scale from those two models with research-predefined choice sets, respectively a choice set containing all fishing sites and a choice set consisting of four randomly drawn sites and one that is actually visited by the angler. Their arguments indicate that the former provides a more consistent behavioral choice model and therefore may infer a more reliable welfare measure.

Parsons and Hauber (1998) use travel time data to set up spatial boundaries for individual fresh-water fishing trips to the lakes and rivers located within Maine, USA. The boundary constraint in their study serves as the first step to rule out fishing sites outside individuals' travel time boundaries. After that, 11 randomly drawn sites[11] within the spatial boundary plus one site that is actually visited are selected to construct a individual specific choice set. The authors eventually find that, using spatial boundaries to define feasible choice sets is less important than one would expect. Paramter estimates based on this CSF process and welfare measures with three scenarios of water quality change are not very significantly different from that without spatial boundaries, especially when the boundaries are expanded to more than 4 hours, but inflated welfare estimates are found when the the spatial boundary is confined within 1.6 hours. Another survey approach applied by Hicks and Strand (2000) alternatively[12] ask respondents about their familiarities to 12 public Chesapeake Bay beaches in Maryland. Unfamiliar alternatives are excluded from respondents' individual specific choice sets. They compare this familiarity-based CSF model to a full choice set model (similar to the MNL model) and a distant-based CSF model (similar to the spatial boundary model we just reviewed), and find that parameter estimates and welfare measures are indeed sensitive to choice set definition.

These representative survey approaches we discussed above imply the importance of

studies

[11]Because a very large number of fishing sites still exists even with the application of preliminary spatial constraints

[12]See dicussions in this paper about the difference from that of Peters, Adamowicz and Boxall (1995).

an appropriate choice set definition in non-market valuation of public goods. However, they all share a common problem of simultaneously modeling the choice sets and choices, which may lead to further problems especially in welfare measures of quality change. As Peters, Adamowicz and Boxall (1995) point out later in their study, their approach, by determining an individual specific choice set first and then conduct welfare analysis, does not consider the dynamics of choice set formation. That is, the individual specific choice set elicited by surveys is assumed constant with exogenous policy changes. But the policy change itself in fact may change both the structure of an individual specific choice set and the site selected. Therefore, a more complete approach may be one that incorporates both changes simultaneously. This problem usually can be addressed with explicit models of the CSF process.

2.3.2 Explicit model approach

A few empirical studies that explicitly model the CSF process have been conducted on transportation choice problems. They[13] provide a possibility to deal with the simultaneity problem encountered in many survey approaches. Among of those studies, the independent availability logit (IAL) model — derived from Manski's (1977) two-stage choice paradigm by Swait (1984) and Swait and Ben-Akiva (1987a) — is of our special interest. It defines $Q_n(C)$, again as in equation (2.3), as the probability of the true feasible choice set being C, and is explicitly modeled as

$$(2.6)\quad Q_n(C) = \frac{\prod_{h \in C} A_n(h) \prod_{k \in B-C} [1 - A_n(k)]}{1 - \prod_{j \in B} [1 - A_n(j)]}$$

where h, k and j are alternatives belonging to different choice sets as denoted in equation (2.6). This IAL model in general assumes IIA and therefore the choice probability ratio between two alternatives in C is independent of the inclusion/exclusion of an

[13]Also include the theoretical model suggested by Manski (1977). See subsection 2.2.1.

irrelevant alternative h' to/from C. For a specific individual n, $A_n(j)$ captures the availability of alternative j, or the probability of alternative j being included in his own feasible choice set. As suggested by Swait and Ben-Akiva (1987a), $A_n(j)$ describes the aggregate impact of constraints on alternative attributes (for example, an external price constraint or a quality constraint imposed on public goods and services). These constraints are also affected by individual specific tastes, which are often assumed *i.i.d.* and follow a certain type of distribution, such as a family of normal distributions or exponential distributions. The simultaneity problem arising in Peters, Adamowicz and Boxall (1995) and many other survey approaches is addressed in equation (2.6) by assuming a dynamic $Q_n(C)$ that varies with $A_n(j)$, which is certainly further connected to the varying alternative attributes and individual tastes under exogenous policy changes (see equation(3.7) in Chapter 3 for discussion).

Equation (2.3) - (2.6) fully define the IAL model that gains popularity in several empirical studies in transportation and marketing research. In a consecutive study, Swait and Ben-Akiva (1987b) assumes $A_n(j)$ (they call it the *probability of captivity*) as an exponential function of individual characteristics and alternative attributes. Based on that, they derive a parameterized logit captivity (PLC) model to analyze the work mode choice data collected from São Paulo, Brazil. They find that the PLC model is statistically superior to the MNL model in parameter estimation. With simulated policy changes in travel time, travel cost and income, the PLC model also demonstrates a great potential that parameterization of the choice set probabilities holds for better explaining observed choice behaviors, which, however, are unstable such that, under certain situations it "predicts less sensitivity to changes than does the MNL model, but in other situations the opposite is seen to occur" (Swait and Ben-Akiva, 1987b, p.115). Another test of Swait and Ben-Akiva (1987a)'s IAL model is conducted by Li, Adamowicz and Swait (2012) with a series of Monte-Carlo experiments. They study the choice set misspecification effect on welfare measures through postulating an exogenous price constraint on environmental public goods. By alternatively parameterizing the availability function with a logistic distribution, $A_n(j)$ is defined as

$$(2.7)\quad A_n(j) = Prob(p_{nj} < \tau_0 + \xi_{nj}) = 1 - Prob(\xi_{nj} < p_{nj} - \tau_0) = \frac{1}{1 + exp[\mu(p_{nj} - \tau_0)]}$$

where p_{nj} is the price attribute of alternative j for individual n; τ_0 is a threshold cutoff condition such as a representative price level; random component ξ_{nj} is assumed to be *i.i.d.* and follows a logistic distribution with location of zero and scale of μ. It is independent of ε_{nj} in equation (2.5). With equation (2.3) - (2.7) and a suggested exogenous price increase, the authors find that the IAL model outperforms other models thought to be misspecified in the CSF process. Specifically, the welfare measures from the traditional MNL model, random parameter logit (RPL) model and choice set generation logit model (see Swait, 2001) are underestimated by about 50%, while the IAL model does not bias welfare measures.

Haab and Hicks (1997) develop an endogenous choice set model to incorporate CSF process. Their approach applies a variation of Manski's (1977) two-stage model and shares similarities to the IAL model. It is also capable of addressing the simultaneity issue of dynamic choice set formation. Specifically, the endogenously generated choice set has a probability (slightly change their notations to be consistent with our thesis)

$$(2.8)\quad Q_n(C) = \prod_{i \in C} P_n(i \in C) \prod_{i \notin C} [1 - P_n(i \in C)]$$

where $P_n(i \in C)$ is the probability of alternative i is included in subset $C \subseteq B$. This $P_n(i \in C)$ is analogous to the availability function $A_n(j)$ and is defined on a vector of individual and/or alternative specific information. Empirical applications of this model are first implemented by the authors later in that paper with two simple examples of recreational beach visits, and they find that parameter estimates and welfare measures differ greatly between the endogenous choice set model and the traditional MNL model.

A recent application of this model is found in Hicks and Schnier (2010) where the authors study the spatial choice problems encountered by commercial fisheries around Alaska, Bering Sea and Aleutian Islands. Spatial choice problem in commercial fisheries is a quite special example and different from many previous problems encountered in the modeling of recreational site and transportation mode choices. Specifically, the unconditional probability of choosing an alternative (commercial fishing location) indicated by Manski (1977) is now defined as

$$(2.9)\quad P_n(i) = \sum_{i \in m, m \subseteq M} P_n(m \in M) P_n(i \mid i \in m)$$

where m is specifically defined as some macro-region, and s_i is some micro-region included in m. M is a set space comprising m macro-regions. For example, a large geographic fishing area containing I micro-regions may be divided into M macro-regions based on certain spatial information such as fish stock, marine reserve and other biophysic information. One can imagine that M is substantially smaller than J. Each of the M macro-regions may contain different number of micro-regions. Since an area concerning spatial choices is preliminary regulated to M macro-regions, formation of individual specific choice sets is no longer arbitrary to include all nonempty subsets of the full choice sets consisting of all s_i. In our example, the number of possible true choice sets to be analyzed is therefore only $2^M - 1$, rather than $2^I - 1$. This specification substantially reduces the large number computation problem that is frequently met in other researches. However, application of this nesting structure to other topics may be limited since spatial choice itself is an special topic, and moreover, as Adamowicz, Glenk and Meyerhoff (2013) state, it usually requires very large data sets that are also longitudinal or at least include some temporal dimensions.

2.4 Social interactions

As we have noticed, none of the CSF models we discussed so far has explicitly addressed the problem of social effect, especially those explicit models discussed in section 2.3.2. They certainly take the individual CSF process as *socially isolated*. For us, social isolation means that individuals generate their choice set through several self-motivated constraints like their own prices or quality perceptions. They are taken as *selfish* without considering perspectives from their social networks. Swait and Ben-Akiva (1987a) point out that the consideration of availability constraints should not only include income and infrastructure in transportation mode research, but must also account for informational, psychological, cultural and social restrictions. However, their social consideration stands more on individuals' own socio-economic characteristics such as gender and education rather than social interactions among individuals (Swait and Ben-Akiva, 1987b).

Social interactions have long been shown important to shape individual behaviors (see Merton, 1957; Becker, 1974; Akerlof, 1980; Case, 1991; Manski, 1993a,b, etc. an extensive literature review can be found in Manski, 2000). But its application to RUM-based discrete choice analysis and non-market valuation problems is scarce. One group of papers, including Brock and Durlauf (2001), Soetevent and Kooreman (2007), Barucci and Tolotti (2012) and others, uses socially interacted RUM models to address the problem of multiple equilibria in discrete choices. Another group, including Neilson and Wichmann (2014) and Wichmann (2014), alternatively focus on the welfare measure problem in non-market valuation. The difference is that, random utility in the former group is individual-based without explicitly considering alternative attributes, while the latter is more intuitively alternative attributes-based. Obviously, the purpose of our research indicates more interest will be put on the second group, where the author(s) explicitly applied a social network approach to model social interactions. This approach can be traced back to DeGroot (1974) and DeMarzo, Vayanos and Zwiebel (2003), and shares similarities to the spatial econometric models indicated by Anselin (1988). Specifically, Neilson and Wichmann (2014) suggest a social utility model as follows

$$(2.10)\quad v_i(g) = (1 - \lambda_i)u_i(g) + \lambda_i \sum_j a_{ij}v_j(g)$$

where i and j are individuals; g is a public good; λ_i is the social propensity of individual i, indicating the weight this individual puts on his social network. $v_i(g)$ is the social utility of individual i over g; $u_i(g)$ is the self utility of i. a_{ij} is an element of A, a row stochastic matrix representing the social network. a_{ij} is positive if j is a friend of i, and zero otherwise[14].

In general, equation (2.10) indicates that an individual's social utility can be described with a weighted utility model capturing both self utility and social utility of other individuals. A follow-up Monte-Carlo simulation study of welfare measures with this specification by Wichmann (2014) finds that, when utilities are influenced by social networks, traditional RUM-consistent WTP models are misspecified. This misspecification results in a bias of 123% in WTP measures at high social network effect (denoted by λ). However, we should note that a common ground all these papers shares is that both groups disentangle their modeling problems by introducing additive social utilities into the random utility specified in section 2.1. None has addressed the problem of how to explicitly model a CSF process with social interactions. This thinking leads us to the modeling and estimation of such a question in the following chapters.

[14]The authors eliminate the possibility of existing enemies in the social network. See more discussions about this and other properties in that paper.

Chapter 3

Empirical Model

In this chapter, we attempt to address the modeling problem given the considerations raised in Chapter 2. An explicit empirical model is developed to capture individual social interactions in the CSF process. The procedure is realized by two steps: First, we define the sociomatrix and social propensity to describe the social context of a community in section 3.1. Second, the standard IAL model presented in sections 2.2 to subsection 2.3.2, is extended within the social context. We call this extended IAL model as the social IAL model. Model specification with these two procedures enables us to conduct a welfare analysis involving non-market valuation of public goods, and procedures are presented in the last section of this chapter.

3.1 Measuring a social context

The social context is measured with two components. One is the sociomatrix, and another is the social propensity. The social IAL model we are going to derive in next section is always build in a social context, rather than solely with a sociomatrix or a social propensity.

3.1.1 Sociomatrix

As early as 1946, Forsyth and Katz measured social interactions using matrices to describe social networks. We call these matrices the sociomatrices (Beum and Brundage, 1950). The social network within a group of decision makers, as defined by Wasserman and Faust (1994, p.20), "consists of a finite set or sets of actors and the relation or relations defined on them". In this thesis, the actors are decision makers in the group that make unordered discrete choices, and relations are social links among them. Sociomatrices provide a clear way to visualize these internal social links and at the same time enbale researchers to derive tractable results. Following techniques applied by Neilson and Wichmann (2014), the sociomatrix for a community with N decision makers is formally defined as

Definition 1. *(Sociomatrix) The sociomatrix W of a community with population size N is represented as a $N \times N$ row stochastic matrix. Matrix entry w_{ij} describes the social relation between decision maker i and j, which has the following properties:*

1. *(Nonnegativity) $w_{ij} \geq 0, \forall w_{ij} \in \mathbb{R}$ and $i, j \in 1, 2, ..., N$, is the weight that decision maker i places on the social connection with decision maker j,*
2. $\sum_{j=1}^{N} w_{ij} = 1, \forall i \in 1, 2, ..., N$, *and*
3. $w_{ii} = 0, \forall i \in 1, 2, ..., N$.

Properties (1) and (2) are standard definitions of a row stochastic matrix. Specifically in a sociomatrix, nonnegativity (Property (1)) indicates that any two decision makers in the community either have no relation if $w_{ij} = 0 \, \forall i \neq j$, or have a positive relation (friendship) if $w_{ij} > 0 \, \forall i \neq j$. The value of w_{ij} is larger if this positive relation is stronger. Property (1) rules out the possibility that two decision makers being enemies (envious relation), under which situation one would expect $w_{ij} < 0$. We suppress this for analytic simplicity (refer to Becker (1974) and Bramoullé (2001) for discussion).

Property (2) indicates that the row entries of a sociomatrix are weights assigned to other network members by individual i. Since these weights sum to 1, a newcomer entering the network tends to trigger a weights reassignment among the existing individuals. Property (3) means that a decision maker is assumed to have no social relation with himself. That is, an individual i is not considered to be a friend of himself in a social context.

3.1.2 Social propensity

The sociomatrix defined above objectively describes the social network status existing in the community. But it indicates nothing about how important the social network is to the individual. For now, let us denote a *social propensity* as the weight a decision maker places on the social network. We use a real number $\tau_1 \in [0, 1]$ to represent it. The spread $[0, 1]$ in fact implies that we allow the social network to be as important as one's own weight (pure altruism) or not important at all (pure selfishness). In other words, the two extremes are: *i*) $\tau_1 = 0$ indicating a decision maker that does not consider his social network in making choices, or he is socially isolated, and; *ii*) $\tau_1 = 1$ meaning that the network influence is as important as own influence. The parameter τ_1 we defined here is analogous to the parameter λ defined in Neilson and Wichmann (2014). This construction is in line with the idea that a non-market good may have smaller or larger effect on individual's personal taste depending on to what extent the individual values himself relative to others in his social network.

We must note that in this thesis the social propensity is actually defined for the entire community. That is, we assume each individual has the same level of social propensity to a specific social network in which they are connected. However, in reality, the social propensity should vary both with individuals and the choice set of alternatives. Different individual personalities may lead to very different social propensities placed upon each element of the social network. We could imagine that an altruistic individual usually has a stronger social propensity (greater τ_1) than those selfish individuals. Moreover, the set of alternatives individuals face may also shift the social propensity. For example, if the individual thinks that the good is very important to his own welfare, then he tends to value less his social network. But this is not necessarily to say that individuals have a stronger social propensity on pure public goods (such as air) than pure private goods. It also depends on the dimension of the social network, such as the size, boundary and links. One would expect a stronger social propensity if the network is used to measure interactions within a household or a rural village in developing countries, and weaker if the network is constructed within an urban community.

3.2 A social IAL model

In this section, we set forward to build an explicit choice model within the social context to capture the influence of social interactions on individual CSF process. Throughout the discussions of section 2.2 in Chapter 2, we have demonstrated step by step how an explicit CSF model is established (see equation (2.3)-(2.5), a work of Manski (1977)). The major argument left in previous studies is how one explicitly constructs $Q_n(C)$ in equation (2.3) such that it can incorporate as much information as possible to appropriately reflect choice behaviors. Explicit CSF models developed by Swait and Ben-Akiva (1987a, equation (2.6)) and Haab and Hicks (1997, equation (2.8)) have no significant differences. Similar modeling procedures are conducted. In this thesis, we will extend the one, equation (2.6), that is extensively studied by Swait (1984), Swait and Ben-Akiva (1987a,b) and Li, Adamowicz and Swait (2012). We call the model based upon equation system (2.3)-(2.6) plus a socially isolated choice constraint (such as equation (2.7) or those originally applied in Swait and Ben-Akiva (1987b) the standard IAL model. Therefore, to socially solve the standard IAL model only requires us to pinpoint the availability function $A_n(j)$ for individual n on alternative j within a social context.

Let us consider an unordered multiple choice problem where a group of individuals face a global choice set B consisting of some global alternatives $j = 1, 2, ..., J$. Each of these global alternatives are defined on two attributes, price p and quality q. The CSF process, in particular the availability function $A_n(j)$ is now assumed to be constrained by a social context characterized with a sociomatrix $\boldsymbol{W}$ and a social propensity τ_1. Specifically, this indicates that the setup of a price or quality cutoff threshold is no longer solely determined by the individual's own perceptions on the price or quality as is shown in equation (2.7). It is also influenced by the price or quality perceptions given by his friends in the social network. By using the goods price as an alternative selection criterion in a social context, equation (2.7) demonstrated by Li, Adamowicz and Swait (2012) can be extended as

$$
\begin{aligned}
A_n(j) &= Prob(p_{nj} < \tau_0 + \tau_1 \sum_{nm} w_{nm} p_{mj} + \xi_{nj}) \\
&= 1 - Prob(\xi_{nj} < p_{nj} - \tau_0 - \tau_1 \sum_{nm} w_{nm} p_{mj}) \qquad (3.1) \\
&= \frac{1}{1 + exp[\mu(p_{nj} - \tau_0 - \tau_1 \sum_{nm} w_{nm} p_{mj})]}
\end{aligned}
$$

where $\tau_1 \sum_{nm} w_{nm} p_{mj}$ is the socially perceived price levels from individual n's neighbors. Other terms are as denoted in equation (2.7). One needs to note that, nonnegativity of the additive social price $\tau_1 \sum_{nm} w_{nm} p_{mj}$ indicates that even if every other individual has a higher price perception than individual n, the price cutoff condition still gets more restrained. This seems odd since, for example, if all others perceive that the good deserves a high price, the individual n should also value that good at a higher price rather than a lower price. Simply changing the addition sign to a minus sign before that term certainly would again encounter other problems to explain rational choice behaviors. Therefore, in this study, we are particularly interested in solving the model by using quality as the selection criterion given its popularity and importance in non-market valuation of environmental and resource public goods. Also, it seems more intuitive that the quality perceptions of friends may influence the availability of alternatives while it is not so clear how prices faced by friends would influence own availability. In a social context, quality of non-market goods is usually not the real quality perceived by individuals solely based on their own experiences, knowledge and cognitive abilities. But rather, it is a perceived quality that also summarizes information obtained from others, including considerations of peer pressures, altruistic propensities, etc. Specifically, the social constraint imposed upon quality selection criterion leads to an availability function $A_n(j)$ where

$$
\begin{aligned}
A_n(j) &= Prob(q_{nj} > \tau_0 - \tau_1 \sum_{nm} w_{nm} q_{mj} + \xi_{nj}) \\
&= Prob(\xi_{nj} < q_{nj} + \tau_1 \tilde{q_{nj}} - \tau_0) \\
&= \frac{1}{1 + exp[-\mu(q_{nj} + \tau_1 \tilde{q_{nj}} - \tau_0)]}
\end{aligned}
\tag{3.2}
$$

where ξ_{nj} is parametrized by a logistic distribution with location 0 and scale μ. τ_0 now is the perceived average quality of alternative j. It is considered as a threshold of alternatives selection. $\tau_1 \in [0, 1]$ represents the level of social propensity. The greater value τ_1 takes, the individual is considered more socially interacted with the rest of the community. $\tilde{q_{nj}} = \sum_{nm} w_{nm} q_{mj}$ is the adjusted or perceived quality for individual n from all of his neighbors $m = 1, 2, ..., N$. Given the property (2) in Definition 1, for individual n, $\tilde{q_{nj}}$ is therefore a weighted average of quality attributes from all n's neighbors, with weights determined by the n^{th} row of the sociomatrix $\boldsymbol{W}$. One can easily see that the term $\tau_1 \tilde{q_{nj}}$ is nonnegative given definitions concerning τ_1, quality attribute q_{nj} and the sociomatrix $\boldsymbol{W}$. Notice that, in this formulation network quality makes it easier for a certain alternative to cross the minimum quality threshold for availability. One also can imagine that, if the network effect is large enough, then the availability of each alternative or probability of each alternative to be included in the individual specific choice set may be collapsed into 1. Under this condition, the entire two-stage model would eventually fall back to a standard MNL model (see more discussions in Chapter 5).

Inserting equation (3.2) back into equation (2.6) enables us to derive the probability of a choice set being true ($Q_n(C)$), and the unconditional probability of a global alternative being choosing ($P_n(j)$) can be subsequently computed with equation (2.4). Therefore, equation (2.3), (2.4), (2.6) and (3.2) describe the calibration procedures we need to derive an IAL model incorporating social interactions in its CSF process. Combining these with the RUM theory indicated in equation (2.5), we are able to establish an explicit social IAL model and conduct welfare analysis via some exogenous quality change policies.

Although the inclusion of social interactions in the CSF process shifts availabilities of alternatives $A_n(j)$, it does not change the model structure of equation (2.3) - (2.6). The intuition here is that, individuals form their feasible choice set by considering their social constraints at the first stage. But after that, making the final choice from the feasible set becomes a personal or private problem and the social network effect does not pass forward to the second stage. The social interaction structure added to the availability function (see equation (3.1) and (3.2)) adds flexibility to the IAL model by allowing for the possibility of social network effects. Notice that in equation (3.1) when τ_1 is equal to zero, the social IAL model collapses to the standard IAL model applied by Li, Adamowicz and Swait (2012).

3.3 Welfare

We assume an exogenous quality change in order to test the bias of the standard IAL welfare measures. Following the traditional welfare measure procedures, we implicitly define the compensating variation (CV) for changes that happen only in goods quality as,

$$(3.3)\ \ V(p^0, q^1, y^0 - CV) = V(p^0, q^0, y^0)$$

where V is the indirect utility function; p is the price; q is the quality level; and y is individuals income level. Superscripts 0 and 1 denote the status-quo and subsequent level of attributes after policy change, respectively. Since the price and income levels are assumed fixed before and after policy changes, equation (3.3) implies that $CV > 0$ if $q^1 > q^0$ and vice versa. We further assume an explicit functional form for V, which consists of a linear indirect utility and an additive error, then equation (3.3) can be explicitly represented by,

$$(3.4)\ \ \beta_y(y^0 - p^0 - CV) + \beta_q q^1 + \varepsilon = \beta_y(y^0 - p^0) + \beta_q q^0 + \varepsilon$$

where β_y and β_q are parameters of the utility function. There is no income effect here since it will be canceled out. Note that β_y is the marginal utility of income, or the marginal disutility of price change. Therefore, we have $\beta_y = -\beta_{price}$. Continue to assume ε_n is a stochastic error term following a Gumbel distribution $\mathbf{G}(0,1)$. Then given the properties of Gumbel distribution (see Ben-Akiva and Lerman, 1991, p.104-106), $V_n + \varepsilon_n \sim \mathbf{G}(V_n, 1)$, $max(V_1 + \varepsilon_1, V_2 + \varepsilon_2, ..., V_n + \varepsilon_n) \sim \mathbf{G}(\ln \sum_{j \in B} exp(V_j), 1)$, and $V^* = max(V) = \ln \sum_{j \in B} exp(V_j)$. Manipulation to equation (3.4) leads to,

$$\begin{aligned} CV &= \frac{1}{\beta_y}[V^*(q^1) - V^*(q^0)] \\ &= \frac{1}{\beta_y}[\ln \sum_{j \in B} exp(V_j^1) - \ln \sum_{j \in B} exp(V_j^0)] \qquad (3.5) \\ &= \frac{1}{\beta_y}[\ln \sum_{j \in B} exp(x_j^1 \hat{\beta}) - \ln \sum_{j \in B} exp(x_j^0 \hat{\beta})] \end{aligned}$$

where x_j^1 and x_j^0 are attribute levels after and before the policy change; $\hat{\beta}$ is the corresponding estimated attribute parameters. However, we should note that equation (3.5) is just the CV function for a standard MNL model, where choice set B is fixed with certainty and contains the same bundle of available alternatives across individuals. The IAL model (socially involved or not) is a two-stage model where the choice sets for individuals are formed probabilistically. Since the variation is choice-set-specific, the CV should be weighted by corresponding choice set probabilities. Specifically,

$$(3.6)\ \ CV = \frac{1}{\beta_y}\{\sum_{C\subseteq B}[Q(C)^1 \ln \sum_{i\in C} exp(x_i^1\hat{\beta})] - \sum_{C\subseteq B}[Q(C)^0 \ln \sum_{i\in C} exp(x_i^0\hat{\beta})]\}$$

where $Q(C)^1$ and $Q(C)^0$ are respectively the probability of choice set C being true after and before the policy change (see equation (2.6)). The CV based on the IAL model is a function not only of the alternatives' utilities but also of the choice set probabilities.

Chapter 4

Monte Carlo Experiments

In this chapter, we test the bias of the standard IAL model when social interactions are ignored in individual's CSF process. A series of Monte Carlo experiments are carried out under different social contexts, each of which is uniquely defined by two parameters: the density of the social network and its corresponding level of social propensity[15]. In general, these experiments are conducted by assuming a true model (the social IAL model) as defined in Chapter 3. The sociomatrix is measured with *Erdos-Renyi* networks (Erdos and Renyi, 1959) in which links are *i.i.d.* and each pair of individuals is connected with a fixed probability (Wichmann, 2014). Using data generated by the true model, a correctly specified model that incorporates social interactions in the CSF process is estimated, and then compared to an estimated model that ignores these interactions. The purpose of doing this is threefold. First, we test the bias of parameter estimates when social interactions are or are not ignored in the CSF process of the IAL model. Second, distributional variations of probabilistic choice sets are calculated. We will discuss intuitions explaining these variations. Lastly, we impose an exogenous quality improvement on one of the alternatives. This enables us to compute and compare welfare measures from these three sets of data, *i.e.* the true data and the data predicted by the correctly and incorrectly specified models. Details

[15]See section 4.1 for a detailed discussion

are unfolded in the following sections.

4.1 Real data generation process

Assume a hypothetical community of size N=2000 where the community members are connected with each other through an *Erdos-Renyi* social network. These networks assume that each link between any two nodes (individuals) is independently formed with given probability. In practice, each network is a $N \times N$ square matrix with rows *i.i.d.* drawn from a binomial distribution $\mathbf{B}(n, p)$, where $n = 1$ is the number of trials, and $p \in [0, 1]$ is the probability of success on each trial and equals the expected social network density[16]. In our Monte Carlo experiments, the network density is assumed to be a value in the set {0.05, 0.10, 0.25, 0.50}. The level of social propensity τ_1 is represented by a number belonging to the set {0.05, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80}. To simplify the discussion of our results, we divide τ_1 into two groups. We say the community has a low level of social propensity if $0.05 \leq \tau_1 \leq 0.40$ and a high level if $0.50 \leq \tau_1 \leq 0.80$. Since a social context is uniquely defined by a sociomatrix with certain network density and its corresponding level of social propensity, we therefore have a total number of 4×9 social contexts, which means we will conduct 36 Monte Carlo experiments.

For each Monte Carlo experiment, 400 replications are performed with draws of random components ξ (equation (3.2)) and ε (equation (2.5)), respectively from a logistic distribution and a Gumbel distribution. In all experiments, each individual is assumed to face a global choice set consisting of three global alternatives {1, 2, 3}. Each alternative for each individual is uniquely defined by two exogenous attributes, price p and quality q. Both p and q are independently drawn from a standard uniform distribution $\mathbf{U}(0, 1)$, and are assumed to be constant across different social contexts

[16]To see why, note that the density of a network is defined as the number of existing links divided by the number of possible links.

and replications. The basic setup for the real data generation process (DGP) of our Monte Carlo experiments is summarized in the following table.

Table 4.1 Basic setup of the real data generation process

Social Structure:	
Number of Individuals (N)	2000
Social Network Density (d)	0.05, 0.10, 0.25, 0.50
Level of Social Propensity (τ_1)	0.05, 0.10, 0.20, 0.30, 0.40 (low level)
	0.50, 0.60, 0.70, 0.80 (high level)
Discrete Choice Structure:	
Number of global alternatives in each set	3
Attributes for each alternative	price, quality
Monte Carlo Experiments:	
Number of replications for each experiment	400
True parameters (ASC1, $ASC2^*$, p, q)	1.5, 1, -3, 5

Note *: ASC1 and ASC2 are alternative specific constants for Alternative 1 and Alternative 2, by setting Alternative 3 as the base choice. They reflect the normalized utilities not revealed by price and quality attributes. Their parameters are interpreted as the magnitudes of average utility differences to Alternative 3. In this setting, these magnitudes are 1.5 and 1.

Following the model defined in section 3.2, we generate in each replication the observed individual choices through two steps. In the first step, each individual forms their own choice set by excluding from their global choice set the alternatives whose qualities are lower than a random threshold. In the correctly specified model, we assume this threshold is determined by three components (see equation (3.2)). The first component is the average perceived quality $\tau_0 = 0.5$, a constant across individuals and the real DGP. The second component is a socially interacted term $\tau_1 \sum_{nm} w_{nm} q_{mj}$, indicating how and to what extent other individuals' perceptions of the quality would influence a specific individual's quality perception of the same alternative. The third is a random component ξ_{nj} drawn from a logistic distribution with the scale of $\mu = 10$ and location of zero[17].

[17]Changing scales of the logistic distribution obviously affects the availability of alternatives, but this does not have a direct impact on our analysis of incorporating social interactions in the CSF process. See further discussions in Li, Adamowicz and Swait (2012).

After the alternative(s) rule-out or the CSF process, in the second step the individual compares the utilities of the alternatives remaining in his choice set through random utility maximization rules. Recall that utilities being compared are random utilities (see equation (2.5)), which consist of a systematic part V plus a random error ε that follows a Gumbel distribution with the scale of unity[18]. For simplicity, we assume that V takes a linear functional form $V = x'\beta$, where x is a vector of independent variables, which includes the two alternative-specific constants ASC1, ASC2, price and quality; β is a vector of the corresponding true parameters, which is set to equal {1.5, 1, -3, 5} for each independent variable respectively (see Table 4.1). Variations in random error draws lead to variations in observable choices in each replication, allowing us to perform maximum likelihood estimations given the unconditional choice probabilities calculated from equation (2.3), (2.4), (2.6) and (3.2).

4.2 Distribution of probabilistic choice sets

Distributional sensitivities of choice sets to the network parameters (*i.e.* network density and τ_1) are of much interest in this research. As we stated above, each individual initially faces a global choice set containing three global alternatives. In a two-stage IAL model that incorporates social interactions, the individual chooses first to rule out some alternatives that have quality attribute levels lower than his random threshold cutoff criterion. This alternative(s) rule-out process determines the distribution of probabilistic choice sets. To not lose data, we control this cutoff criterion through redrawing the error ξ_n for individual n in the real DGP if all alternatives are ruled out[19]. Therefore, each individual is guaranteed to have at least one alternative left after the CSF process. That is, subsets expected for each individual are strictly nonempty. Following the

[18]Note that in the real DGP the error term is generated using full sample size not the reduced size after the first step.

[19]This will not significantly change the overall distribution of ε since they are *i.i.d.* drawn from the same distribution. It is especially true because we redraw only a small percentage of errors, which mostly is less than 5%.

quality cutoff (CSF) process, we count the number of subsets in each category to form their distributions. For example, assume each person initially faces a global choice set containing three global alternatives {1, 2, 3}. The CSF process may lead to: *i*) the first individual forms a subset {2, 3}; *ii*) the second individual forms a subset {1, 3}; *iii*) the third individual has a subset {1, 2, 3}, and so on. Whatever these subsets are, they are inevitably nonempty subsets of {1, 2, 3}, and may vary as a function of our social context settings. These variations lead to further variations in observable choices made by individuals in each replication and in each social context.

The CSF process with social interactions indicates that individuals can compress the initial global choice set into a simpler choice set to, for instance, reduce transaction costs. Comparison and selection of alternative for utility maximization will be much easier. One also can easily imagine that the CSF process will be much more necessary if the number of global alternatives in the initial global choice set is large, say 7 fishing lakes, 10 possible vacation destinations, or 20 types of cars, etc. Under these circumstances, CSF process obviously helps reduce the informational complexity and transaction costs. It can also result in more variation of the probabilistic choice sets distribution and its corresponding observable choices.

4.3 Estimation, welfare measures and bias

4.3.1 Parameter estimates and its measurement of bias

Based on the true model and real DGP described respectively in section 3.2, 3.3 and 4.1, we estimate two models for each social context. The first is an IAL model that incorporates social interactions in its CSF process. We call this the correctly specified model. Another is an incorrectly specified model that ignores social interactions, or a pure IAL model as indicated in section 2.2 and subsection 2.3.2. Two pairs of parameter estimates are obtained, one for each model. To evaluate performances of the correctly and incorrectly specified models, we compute the proportional Root Mean Square Errors (RMSE), which is defined as

$$(4.1)\quad Proportional\, RMSE = \frac{1}{R}\sum_{r=1}^{R}\sqrt{(\frac{\hat{\beta}_{ri}-\beta_i}{\beta_i})^2}$$

where $r = 1, 2, ..., R$ and $R = 400$ is the total number of replications for each social context setting; $\hat{\beta_{ri}}$ is the i^{th} parameter estimates in replication r; β_i represents the i^{th} true parameter. In the correctly specified model, we have a total number of seven parameters to be estimated. Hence $i = 1, 2, 3, ..., 7$ respectively indexes parameters for ASC1, ASC2, price, quality, τ_0, τ_1 and μ. In the incorrectly specified model we do not estimate τ_1 since there is no social effect. We therefore only have six parameters to be estimated and use $i = 1, 2, 3, ..., 6$ instead to index the corresponding parameters. The proportional RMSE directly measures the percentage deviation of the values estimated by our models from the true parameter values.

4.3.2 True welfare and bias

For Monte Carlo simulations with a real DGP, we are able to calculate the true welfare measure (CV) since we observe the draws of error terms. Therefore, the true CV is just the difference of random utilities of the alternatives that have been chosen after and before the policy change, a 0.4 quality improvement on Alternative 2 in our paper. The true CV for individual n in replication r is calculated as

$$(4.2)\quad \begin{aligned} CV_{true,n,r} &= \frac{U^1_{chosen,n,r} - U^0_{chosen,n,r}}{-\beta_{price}} \\ &= \frac{(\mathbf{x}^1_{chosen,n,r}\beta + \varepsilon^1_{chosen,n,r}) - (\mathbf{x}^0_{chosen,n,r}\beta + \varepsilon^0_{chosen,n,r})}{-\beta_{price}} \end{aligned}$$

where 0 and 1 respectively represent the states of the world before and after the quality improvement. $CV_{true,n,r}$ is the true CV for individual n in replication r. $U^1_{chosen,n,r}$ is the random utility of the chosen alternative after the quality improvement, while $U^0_{chosen,n,r}$ is that of before quality improvement[20]. $\mathbf{x}^1_{chosen,n,r}$ is a vector of attribute levels of the chosen alternative after the quality change, while $\mathbf{x}^0_{chosen,n,r}$ is that of before quality change; β are their corresponding true parameters and β_{price} is the true parameter of price.

Estimated CVs are calculated based on equation (3.6) in section 3.3. To evaluate the CV performance of the correctly and incorrectly specified models, we compute the RMSE of mean CVs, which is slightly different from the one defined in equation (4.1). Here the RMSE is

$$(4.3)\quad RMSE = \sqrt{\frac{\sum_{r=1}^{R}[E(\hat{CV_r}) - E(CV_r)]^2}{R}}$$

where $E(\hat{CV_r})$ is the estimated (correctly or incorrectly specified) mean CV in replication r, and $E(CV_r)$ is the corresponding true mean CV. Specifically $E(\hat{CV_r}) = \sum_{n=1}^{N} \hat{CV}_{n,r}/N$, where N is the total population number, and $\hat{CV}_{n,r}$ is the predicted CV for individual n in replication r. RMSE here is measured in dollars, the same as the measurement of price and therefore can be directly compared with the true CV.

[20]Note that the alternatives chosen before and after the policy change could be different or be the same.

Chapter 5

Results and Discussion

In this chapter we present the results and discuss their implications based on our Monte-Carlo experiments in Chapter 4. In summary, these results demonstrate that we can successfully estimate an IAL model that incorporates social interactions in its CSF process. We compare estimates from the correctly and incorrectly specified models with their true DGP values. We find that failure to account for network effects may lead to biased estimates of our models' parameters, particularly when the strength of social interactions is high.

5.1 Correctly specified model: the IAL model that incorporates social interactions

Since the setup in the true model incorporates social interactions, we therefore refer to the estimated model that also incorporates social interactions as the correctly specified model. The estimated results of the IAL model incorporating social interactions are shown in various tables below with clear comparisons to the incorrectly specified model, which we will discuss in the next section. Each upper part of Table A.1.1-A.1.4 (see Appendix A) presents the parameter estimates of an IAL model incorporating social interactions in the CSF process under all social contexts, that is, four social network densities with nine different levels of social propensity for each of them.

We find that in each setting of social network densities, the parameter estimates deviate little from the true parameters with the level of social propensity τ_1 varying from low to high. The stable and good performance of the correctly specified IAL model is further confirmed by the proportional RMSEs showed in the upper parts throughout Table A.2.1-A.2.4. Proportional RMSE is a direct measure of percentage differences between values estimated by our model and the true parameter values. The results indicate that on average almost all parameter estimates deviate from their true values less than 5% with only two exceptions, that is, $\hat{\tau_1}$ at density=0.05 and $\tau_1 = 0.10$, and $\hat{\tau_0}$ at density=0.50 and $\tau_1 = 0.80$. In general, the estimate of τ_1 in correctly specified models becomes more precise as the true τ_1 value increases[21].

Table A.3.1-A.3.4 present a comparison of choice set distributions. Notice that in each setting of social network densities, the levels of social propensity are divided into two groups, with the upper columns grouped as low level social propensity ($0.05 \leq \tau_1 \leq 0.40$) and the lower columns grouped as high level social propensity ($0.50 \leq \tau_1 \leq 0.80$). For each τ_1, the distributions of choice sets from the three models are compared. The first column shows the distribution of the true model. The second column shows the distribution of the correctly specified model. The third column shows the distribution of the incorrectly specified model. We can easily find that under all circumstances of social context, the distributions of choice sets from the correctly specified model are very close to that of the true model. In other words, the choice sets estimated by the correctly specified model are likely the individuals true choice sets. This result is robust to different social network densities and variations in their corresponding level of social propensity.

Another interesting finding from Table A.3.1-A.3.4 is that in each social network

[21]Notice that the **NA**s in the table indicate that we do not estimate τ_1 in the incorrectly specified models.

density, when τ_1 changes from a lower level to a higher level, the number of alternatives included in the choice set tend to increase. For example, when social network density equals 0.25 and $\tau_1 = 0.05$, the distribution of choice sets spread quite evenly on each choice set. That is, for all samples, the number of individuals having the true choice set containing only Alternative 1 is very similar to the number of individuals having true choice set containing Alternative 1, 2 or 1, 2, 3. But when τ_1 increases to 0.80 with network density unchanged, the number of individuals having true choice set containing only Alternative 1 has a very sharp decrease while the number of sets having all three alternatives increases significantly. The reason for this pattern can be easily derived from equation (3.2), the availability function. When τ_1 increases, the availability of an alternative ($Prob(\xi_{nj} < q_{nj} + \tau_1 \tilde{q_{nj}} - \tau_0)$) also increases since $\tilde{q_{nj}}$ is a nonnegative term. It leads to an increase in the likelihood of this alternative been included into individuals choice sets, and therefore we see the number of true choice sets including all three available alternatives increase. In other words, when levels of social propensity are very high, the individual tends to put a lot of weight into the quality perception from his network. The network quality is not discounted much. It is added to his personal quality making it easy for the alternative to achieve the minimum quality threshold.

5.2 Incorrectly specified model: the IAL model that ignores social interactions

Lower parts of Table A.1.1-A.1.4 show the results from estimating an incorrectly specified model, an IAL that ignores social interactions in its first-stage CSF process. In general, we find that parameter estimates vary much more than those of the correctly specified model. The most obvious one is τ_0, for which estimates tend to get smaller along with the increase of τ_1. Specifically, a sharp change in $\hat{\tau}_0$ happens when τ_1 increases from 0.40 to 0.50: $\hat{\tau}_0$ decreases more than 6 times and changes sign. The other parameters signs keep consistent with the true values. However, they deviate significantly from the true ones after $\tau_1 \geq 0.50$. In general, under high level social propensity ($\tau_1 \geq 0.50$), parameters for ASC1, ASC2, τ_0 and μ are underestimated, while parameters for price and quality are overestimated. The magnitudes of these underestimations and overestimations are indicated by the proportional RMSEs presented in the lower parts of Table A.2.1-A.2.4. They show that parameter estimates for ASC1, ASC2 and τ_0 deviate from the true ones by 5.26% to 46.54% when the levels of social propensity τ_1 are set at lower levels ($\tau_1 \leq 0.40$). But the deviation changes from 7.16% to a surprisingly 1,952% when τ_1 are set at higher levels ($\tau_1 \geq 0.50$). These patterns are similar across different social network densities. Combining these results with those of Table A.2.1-A.2.4, we find that the incorrectly specified model underestimates the true parameter values. Given a typically allowed 5% level of deviation, the parameter estimates for these three variables are all considered as inaccurate in our Monte-Carlo simulations.

The behavior of the μ estimate is totally different and all estimates are robust even in the incorrectly specified model, with proportional RMSEs ranging from 0.74% to 2.97%. Discussions for proportional RMSEs of the two alternative attributes are divided into two parts: low and high levels of social propensity. At the lower levels of social propensity, proportional RMSEs indicate that the deviations are mostly under 5% with only very few exceptions. However, if $\tau_1 \geq 0.50$, then the parameters on average are overestimated by 7.56% to 17.9%, which again imply significant inaccuracies.

These inaccuracies are visualized throughout Figure B.1.1-B.3.2 (see Appendix B), where we draw cross-model[22] kernel densities of three key variables (price, quality, and τ_0) for all social contexts. Figure B.1.1-B.1.2 and Figure B.2.1-B.2.2 show that, when $\tau_1 \leq 0.40$, the parameter estimates for price and quality have kernel densities evenly distributed around the true values. The distributions of the correctly specified model in

[22]Cross-model means three models, the true, correctly specified and incorrectly specified models.

general outperform, though slightly, those of the incorrectly specified models, indicating more efficient estimation. But when $\tau_1 \geq 0.50$ at the higher level of social propensity, kernel densities indicate that in the incorrectly specified model the distributions are highly skewed to the right-hand side, which again confirm the overestimation previously discussed. Figure B.3.1-B.3.2 indicate that, for each social network density, when τ_1 increases the kernel densities of $\hat{\tau_0}$ from the incorrectly specified model skew more to the left-hand side, indicating a significant downward bias of the parameter estimates.

For most variables discussed above, a clear threshold is found when $0.40 \leq \tau_1 \leq 0.50$[23]. This threshold gets especially apparent when the sign of estimated τ_0 changes from positive to negative in Table A.1.1-A.1.4. We confirm this with the choice set distribution outlined in Table A.3.1-A.3.4. From the lower parts of these tables, we can see that if the model is incorrectly specified, the number of choice sets containing only part of the three alternatives (strict subsets of B) are all zeros. Individuals' choice sets will always contain three alternatives, and no alternatives will be excluded. This is reasonable if we trace back to Table A.1.1-A.1.4 and the *socially isolated* quality-based availability function $A_n(j) = Prob(\xi_{nj} < q_{nj} - \tau_0)$. If we put the negative $\hat{\tau_0}$ back into this equation, right-hand side probability equals almost 1 given drawn distributions of ξ_{nj} and q_{nj}. In general, we can conclude that, under those representative social network densities we tested, if the level of social propensity is high enough ($\tau_1 \geq 0.50$ in our experiments), the incorrectly specified model would just collapse into the simple one-stage MNL model as specified by McFadden (1974).

5.3 Welfare measures and their biasness

An overview of Table 5.1 indicates that, in a correctly specified model, individuals' willingness to pay for a 0.4 quality improvement on Alternative 2 tends to decrease

[23]Notice that the specific location of the threshold may change with the parameterization of the Monte-Carlo experiment. The most important fact is to notice that a threshold behavior indeed exists.

Table 5.1 Welfare measures (in dollars): a comparison of the true, correctly and incorrectly specified models

		Network Density=0.05			**Network Density=0.10**		
		True CV	*CV Net*	*CV NoNet*	*True CV*	*CV Net*	*CV NoNet*
Low level social propensity	τ_1=0.05	0.3860	0.3845(0.0282)	0.3790(0.0347)	0.3856	0.3843(0.0288)	0.3799(0.0373)
	τ_1=0.10	0.3854	0.3824(0.0250)	0.3785(0.0385)	0.3854	0.3819(0.0245)	0.3765(0.0393)
	τ_1=0.20	0.3824	0.3780(0.0215)	0.3728(0.0370)	0.3823	0.3794(0.0238)	0.3760(0.0376)
	τ_1=0.30	0.3750	0.3726(0.0185)	0.3698(0.0291)	0.3756	0.3718(0.0202)	0.3704(0.0298)
	τ_1=0.40	0.3668	0.3667(0.0178)	0.3677(0.0314)	0.3670	0.3628(0.0176)	0.3622(0.0272)
High level social propensity	τ_1=0.50	0.3584	0.3566(0.0156)	0.4308(0.0965)	0.3582	0.3559(0.0155)	0.4372(0.1011)
	τ_1=0.60	0.3496	0.3489(0.0136)	0.4492(0.1028)	0.3493	0.3497(0.0152)	0.4500(0.1045)
	τ_1=0.70	0.3421	0.3408(0.0139)	0.4139(0.0754)	0.3422	0.3405(0.0143)	0.4134(0.0752)
	τ_1=0.80	0.3365	0.3361(0.0142)	0.3898(0.0588)	0.3363	0.3354(0.0160)	0.3902(0.0596)
		Network Density=0.25			**Network Density=0.50**		
		True CV	*CV Net*	*CV NoNet*	*True CV*	*CV Net*	*CV NoNet*
Low level social propensity	τ_1=0.05	0.3857	0.3836(0.0262)	0.3816(0.0366)	0.3852	0.3820(0.0270)	0.3773(0.0344)
	τ_1=0.10	0.3851	0.3833(0.0234)	0.3802(0.0403)	0.3852	0.3827(0.0274)	0.3796(0.0401)
	τ_1=0.20	0.3825	0.3783(0.0189)	0.3744(0.0328)	0.3831	0.3794(0.0236)	0.3719(0.0366)
	τ_1=0.30	0.3755	0.3725(0.0174)	0.3706(0.0316)	0.3754	0.3724(0.0202)	0.3705(0.0306)
	τ_1=0.40	0.3667	0.3660(0.0166)	0.3668(0.0301)	0.3664	0.3647(0.0142)	0.3644(0.0255)
High level social propensity	τ_1=0.50	0.3582	0.3566(0.0156)	0.4367(0.1016)	0.3583	0.3554(0.0163)	0.4343(0.1000)
	τ_1=0.60	0.3501	0.3488(0.0151)	0.4502(0.1037)	0.3495	0.3494(0.0132)	0.4506(0.1045)
	τ_1=0.70	0.3423	0.3417(0.0159)	0.4159(0.0774)	0.3421	0.3415(0.0177)	0.4138(0.0755)
	τ_1=0.80	0.3361	0.3353(0.0129)	0.3868(0.0557)	0.3363	0.3369(0.0142)	0.3896(0.0585)

Note:
1.CVs are measured in mean welfares in each social environment.
2.Correctly and incorrectly specified model respectively indicates an IAL model incorporating and ignoring social interactions.
3.Root Mean Square Errors (RMSEs) are in the parenthesis. Refer to equation (4.3) for specifications.

from around \$0.386 to \$0.336 when the social propensity level increases. This trend has no significant relationship with the density of the social network. In general, network density is not influencing the results. This result suggests that it is not about the number of links in the social network, but how important the links actually are (*i.e.* τ_1) to individuals.

The correctly specified model ($CVNet$) shows an accurate approximation to the true model with precise average welfare measures. RMSEs are all well behaved with average deviation of \$0.03, indicating very little deviation from the true model. Its goodness of fitting to the true model does not vary significantly with variations in social network densities and τ_1. The incorrectly specified model ($CVNoNet$) in general also gives good approximation to the true welfare measures when τ_1 is set at a lower level, however with slightly higher RMSEs. A decreasing trend is found when τ_1 increases within the low level. Both the correctly and incorrectly specified models tend to insignificantly underestimate WTPs while the latter one performs worse.

However, if τ_1 is at a high level where $0.50 \leq \tau_1 \leq 0.80$, the incorrectly specified model tends to significantly overestimate individuals WTPs for the quality improvement. The estimated WTPs vary between \$0.387 and \$0.451 with an average deviation of \$0.0557 to \$0.1045 from the true WTPs. Since the RMSEs for CVs are measured in dollars they can be directly compared with true CVs. A measure called *coefficient of variation* of the RMSE (CoVRMSE) by normalizing the RMSE to the mean of true CVs can give us more intuition on the bias analysis. It is defined as $RMSE/E(CV)$, and calculated results are presented in Table 5.2.

In Table 5.2, The correctly specified model in general exhibits good approximations with CoVRMSEs varying under 8% level and mostly under 5%. The incorrect model has CoVRMSEs slightly higher than that of the correct model at the lower level of social propensity, with a few CoVRMSEs over 10% level. When τ_1 are high, however, the overestimated bias from the incorrect model measured in CoVRMSE are very large, ranging from approximately 17% to 30%. The bias shows no particular pattern across different social network densities.

Table 5.2 Coefficient of variation for correctly and incorrectly specified models

		Network Density=0.05		Network Density=0.10	
		CV Net	*CV NoNet*	*CV Net*	*CV NoNet*
Low level social propensity	τ_1=0.05	0.0731	0.0899	0.0747	0.0967
	τ_1=0.10	0.0649	0.0999	0.0636	0.1020
	τ_1=0.20	0.0562	0.0968	0.0623	0.0984
	τ_1=0.30	0.0493	0.0776	0.0538	0.0793
	τ_1=0.40	0.0485	0.0856	0.0480	0.0741
High level social propensity	τ_1=0.50	0.0435	0.2693	0.0433	0.2822
	τ_1=0.60	0.0389	0.2941	0.0435	0.2992
	τ_1=0.70	0.0406	0.2204	0.0418	0.2198
	τ_1=0.80	0.0422	0.1747	0.0476	0.1772
		Network Density=0.25		**Network Density=0.50**	
		CV Net	*CV NoNet*	*CV Net*	*CV NoNet*
Low level social propensity	τ_1=0.05	0.0679	0.0949	0.0701	0.0893
	τ_1=0.10	0.0608	0.1046	0.0711	0.1041
	τ_1=0.20	0.0494	0.0858	0.0616	0.0955
	τ_1=0.30	0.0463	0.0842	0.0538	0.0815
	τ_1=0.40	0.0453	0.0821	0.0388	0.0696
High level social propensity	τ_1=0.50	0.0436	0.2836	0.0455	0.2791
	τ_1=0.60	0.0431	0.2962	0.0378	0.2990
	τ_1=0.70	0.0465	0.2261	0.0517	0.2207
	τ_1=0.80	0.0384	0.1657	0.0422	0.1740

Note:
1.Coefficient of variation is calculated with $CVRMSE = \frac{RMSE}{E(CV)}$.
2.Correctly and incorrectly specified model respectively indicates an IAL model incorporating and ignoring social interactions.

Combining Table 5.1 and 5.2, the abrupt change of welfare measures in the incorrectly specified models when $0.40 \leq \tau_1 \leq 0.50$ is best explained with the distribution of probabilistic choice sets and the estimated marginal utility of income ($-\hat{\beta_{price}}$). The incorrectly specified IAL model collapses to the standard MNL model when $0.40 \leq \tau_1 \leq 0.50$, as we discussed at the end of section 5.2. Notice in equation (3.6), section 3.3, that the estimated CVs are weighted by a probabilistic CSF process. If $Q(C)$ has a probability 1 $\forall C \subseteq B$ and $C \neq \emptyset$, then equation (3.6) is reduced to equation (3.5), the estimated CV function for standard IAL model. The change in difference between indirect utilities from chosen alternatives after and before quality improvement is unclear when social interactions are completely ignored from the choice set formation process (term in the square brackets of equation (3.5)). However, we do notice that, at the high level social propensity, the underestimated marginal utility of income reaches the lowest when $\tau_1 = 0.60$. This is consistent with our welfare measures where the bias correspondingly reaches the highest value.

Chapter 6

Conclusion and Discussion

6.1 Conclusion

In this study we first discussed potential problems existing in traditional modeling of individuals' unordered discrete choice behaviors. Starting from a standard MNL model, we showed the necessity of including a CSF process such that the MNL model is decomposable with two stages, a CSF stage and a RUM stage. In the first stage, individuals apply certain constraints (rules) to compress the universe of choice sets they face. In the second stage, individuals compare alternatives remaining in the choice sets through RUM theory and select the one that maximizes utility.

Given the benefits of CSF models and their improved welfare estimates, it is important to further explore the choice set formation processes focusing on relaxing standard (but potentially restrictive) assumptions. One of these assumptions is that individuals make decisions in social isolation when forming their choice sets. Although some studies had tried to model social interactions in discrete choice analysis, they all focused on the RUM-base. Incorporating networks to the RUM step is an interesting idea; however, it does not address the social effects in the choice set formation step. This would still lead to bias of standard CSF model estimation and welfare measures when networks are indeed part of the choice set process.

We address this issue by introducing social interactions into the CSF process (we denote this process by SCSF) to set up a social IAL model. This SCSF process is able to capture additive social interaction effects. Our approach assumes that the social network is captured by a row-stochastic sociomatrix that weights the influence of connections in the CSF process. We explore different social context with varying network densities and levels of social propensity. A series of Monte-Carlo experiments were conducted under different social contexts. We find that the correctly specified model with a SCSF process in general outperformed the incorrectly specified model in which social interactions were totally ignored in the CSF process. Estimation bias is particularly strong in scenarios of high social interaction. In these cases, the distributions of probabilistic choice sets and welfare measures are also significantly biased. These biases, however, have little relationship with the densities of social networks, but they vary with the levels of social propensity. Specifically, biases were significantly larger beyond a certain threshold of social propensity. After reaching this threshold, the incorrectly specified IAL model would collapse into a standard MNL model. This indicated that, in a highly interactive society, efforts made to infer individuals' CSF process in social isolation are not necessary. It only adds burdensome computations to the analysis since a standard MNL model delivers the same estimates. Recall, however, that these estimates would be significantly biased and the consistent model is the social IAL model.

In general, this study shows that traditional two-stage choice models without considering social interactions in the CSF process would lead to biased parameter estimates and welfare measures. Public policy makers who frequently deal with non-market valuation problems using discrete choice analysis can have a high risk of misinterpreting the real data collected in field studies, if the implicit social network effects are ignored. Welfare measures derived from such a misinterpretation tend to significantly overestimate individual's WTPs if the social network effects are high (see Table 5.1). This would largely increase the cost of implementing public policies due to failures in a benefit-cost analysis. Our results also indicate that, a two-stage CSF model is not always superior to a standard MNL model. Specifically, if social network effects are ignored, welfare measures derived from a CSF model could be much worse than that from a MNL model.

6.2 Discussion for future research

Both the approach in which we treat social interactions using the CSF process and our subsequent empirical findings add to the choice set formation literature. However, with respect to social interactions, this literature is just taking off and there are still several unanswered questions. First, this thesis shows one possible way of incorporating social interactions in the CSF process. But for a two-stage IAL model, should the social interactions be considered in both stages, *i.e.* the CSF stage and RUM stage? In our model, the probabilistic choice sets derived from a SCSF process incorporate social effects in the first stage. Following that, the second stage RUM considers a pure private behavioral model. That is, after SCSF process, alternatives left in the choice set carry social information, and individuals no longer consider social constraints when maximizing random utilities through comparing these remaining alternatives. Moreover, we also found that a threshold exists at a certain level of social propensity, which finally leads to connections between a standard IAL model and MNL model. But it is still unclear why this happened. Future modeling approaches may benefit from a better understanding of this threshold behavior.

Secondly, an unavoidable challenge for the application of our model to field studies is the data requirement. Specifically, how to appropriately measure the social network and construct the sociomatrix are still questions to be answered in sociology and behavioral economics. Social links within a network are complicated, especially when the size of the network is large. Network boundaries in reality are usually vague. Social propensity may cause another problem since it is very difficult to model. For instance, as we discussed in subsection 3.1.2, they might change with individuals and random coefficient models may be more appropriate. Also, individuals facing different choice contents may exhibit very different interactions within their social network. An index approach that measures social propensity as a function of individual and choice specific characteristics could be explored. While this work takes a first step towards incorporating social structure into unordered discrete choice models with choice set formation, further research is needed to shed light on these additional questions.

Thirdly, like any other parameterized explicit model approaches, the functional form of availability function $A_n(j)$ has effects on parameter estimates and welfare measures. One problem is that, in our social IAL model (equation 3.2) the quality perception from neighbors enters the availability function as a nonnegative term. This indicates that adding network qualities would always make it easier for an alternative to pass the selection threshold, even if every other individual has a lower quality perception. This contradicts the intuition that if all neighbors have relatively lower quality perceptions on a specific alternative, then one would also downgrade his quality perception on that alternative in a social context. A more appropriate approach could be one that includes such a possibility by taking the perception differences rather than the perception itself into consideration. That is, the $\tilde{q_{nj}}$ in equation 3.2 is no longer equals $\sum_{nm} w_{nm} q_{mj}$ but should be $\sum_{nm} w_{nm}(q_{mj} - q_{nj})$. Perception differences are modeled with $(q_{mj} - q_{nj})$. The latter term says that, if the neighbors have higher quality perceptions where $q_{mj} > q_{nj}$, it would be easier for one alternative to pass individual n's threshold; and when $q_{mj} < q_{nj}$, this process becomes harder. Another problem might happen in deciding the distribution of random component ξ_n (equation 3.2), which is assumed to be logistically distributed with location 0 and scale μ. Both parameters have effects on the availability. Higher locations and scales lead to a harder alternative exclusion. This needs to be further explored and decided with the individual specific data characteristics in field experiments.

Fourthly, our social IAL model still encounters the burdensome computation problem as in many other explicit models of CSF process (Manski, 1977; Swait and Ben-Akiva, 1987a,b; Haab and Hicks, 1997, etc.). Since all nonempty subsets of the global choice set will be used to model the probabilistic true choice sets, a slight increase in the number of alternative, e.g. to 10, would substantially increase the number of nonempty subsets to 1023 ($2^{10} - 1$). A possible way to address this problem might be the nesting approach applied to spatial choice problems in Hicks and Schnier (2010). But that approach is still limited since many discrete choice problem does not exhibit the spatial patterns as in commercial fisheries. In the field experiment of common choices among public goods and services, the survey approach suggested by Peters, Adamowicz and Boxall (1995) might be an alternative because the responses to their designed survey questions could have implied individuals' social characteristics. However, if the problems frequently encountered in choice experiment design are not controlled properly, a truthful revealing of individual specific choice sets is not possible, and this could again lead to biased welfare measures.

Bibliography

Adamowicz, W. L., Glenk, K., and Meyerhoff, J. (2013). Choice modelling research in environmental and resource economics: Applications to sub-fields of environmental and resource economics. *Unpublished*.

Akerlof, G. A. (1980). A theory of social custom, of which unemployment may be one consequence. *Quarterly Journal of Economics*, 94(4):749–775.

Anselin, L. (1988). *Spatial Econometrics : Methods and Models*. Dordrecht ; Boston : Kluwer Academic Publishers.

Barucci, E. and Tolotti, M. (2012). Social interaction and conformism in a random utility model. *Journal of Economic Dynamics & Control*, 36(12):1855–1866.

Becker, G. S. (1974). A theory of social interactions. *Journal of Political Economy*, 82(6):1063–1093.

Ben-Akiva, M. and Lerman, S. R. (1991). *Discrete Choice Analysis: Theory and Application to Travel Demand*. The MIT Press, Cambridge, 4th ed. edition.

Beum, C. O. and Brundage, E. G. (1950). A method for analyzing the sociomatrix. *Sociometry*, 13(2):141–145.

Bramoullé, Y. (2001). Interdependent utilities, preference indeterminacy, and social networks. *Documents de travail*.

Brock, W. A. and Durlauf, S. N. (2001). Discrete choice with social interactions. *Review of Economic Studies*, 68(2):235–260.

Case, A. C. (1991). Spatial patterns in household demand. *Econometrica*, 59(4):953–965.

Cherry, T. L., Frykblom, P., Shogren, J. F., List, J. A., and Sullivan, M. B. (2004). Laboratory testbeds and non-market valuation: The case of bidding behavior in a second-price auction with an outside option. *Environmental and Resource Economics*, 29(3):285–294.

Debreu, G. (1960). Individual choice behavior: A theoretical analysis by r. duncan luce. *American Economic Review*, 50(1):186–188.

DeGroot, M. H. (1974). Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121.

DeMarzo, P. M., Vayanos, D., and Zwiebel, J. (2003). Persuasion bias, social influence, and unidimensional opinions. *Quarterly Journal of Economics*, 118(3):909–968.

Erdos, P. and Renyi, A. (1959). On random graphs. i. *Publicationes Mathematicae*, 6:290–297.

Forsyth, E. and Katz, L. (1946). A matrix approach to the analysis of sociometric data: preliminary report. *Sociometry*, 9(4):340–347.

Greene, W. H. (2008). *Econometric Analysis*. Pearson/Prentice Hall, Upper Saddle River, N.J., 6th ed. edition.

Haab, T. C. and Hicks, R. L. (1997). Accounting for choice set endogeneity in random utility models of recreation demand. *Journal of Environmental Economics and Management*, 34(2):127–147.

Hensher, D. A. and Greene, W. H. (2002). Specification and estimation of the nested logit model: Alternative normalisations. *Transportation Research: Part B: Methodological*, 36(1):1–17.

Hensher, D. A. and Greene, W. H. (2003). The mixed logit model: The state of practice. *Transportation*, 30(2):133–176.

Hicks, R. L. and Schnier, K. E. (2010). Spatial regulations and endogenous consideration sets in fisheries. *Resource and Energy Economics*, 32(2):117–134.

Hicks, R. L. and Strand, I. E. (2000). The extent of information: its relevance for random utility models. *Land Economics*, pages 374–385.

Li, L., Adamowicz, W. L., and Swait, J. (2012). The effect of choice set misspecification on welfare measures in random utility models. *Unpublished*.

Luce, R. D. (1959). *Individual Choice Behavior: a Theoretical Analysis*. Wiley, New York.

Luce, R. D. and Suppes, P. (1965). In Luce, R. D., Bush, R. R., and Galanter, E., editors, *Preference, Utility, and Subjective Probability*, Handbook of Mathematical Psychology, Vol. III, pages 252–410. Wiley, New York.

Manski, C. F. (1977). The structure of random utility models. *Theory and Decision*, 8(3):229–254.

Manski, C. F. (1993a). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies*, 60(3):531–542.

Manski, C. F. (1993b). Dynamic choice in social settings: Learnings from the experiences of others. *Journal of Econometrics*, 58(1-2):121–136.

Manski, C. F. (2000). Economic analysis of social interactions. *Journal of Economic Perspectives*, 14(3):115–136.

McFadden, D. (1974). In Zarembka, P., editor, *Conditional Logit Analysis of Qualitative Choice Behavior*, Frontiers in Econometrics, pages 105–142. Academic Press, New York and London.

McFadden, D. (1981). In Manski, C. F. and McFadden, D. L., editors, *Econometric Models of Probablistic Choice*, Structural Analysis of Discrete Data with Econometric Applications, pages 198–272. The MIT Press, Cambridge.

McFadden, D. (2001). Economic choices. *American Economic Review*, 91(3):351–378.

Merton, R. K. (1957). *Social Theory and Social Structure*. Glencoe, Ill. : Free Press, 1957; Rev. and enl. ed.

Neilson, W. and Wichmann, B. (2014). Social networks and non-market valuations. *Journal of Environmental Economics and Management*, 67(2):155–170.

Parsons, G. R. and Hauber, A. B. (1998). Spatial boundaries and choice set definition in a random utility model of recreation demand. *Land Economics*, 74(1):32–48.

Peters, T., Adamowicz, W. L., and Boxall, P. C. (1995). Influence of choice set considerations in modeling the benefits from improved water quality. *Water Resources Research*, 31(7):1781–1787.

Soetevent, A. R. and Kooreman, P. (2007). A discrete-choice model with social interactions: With an application to high school teen behavior. *Journal of Applied Econometrics*, 22(3):599–624.

Swait, J. (1984). Probabilistic choice set formation in transportation demand models. *Unpublished PhD Thesis*.

Swait, J. (2001). Choice set generation within the generalized extreme value family of discrete choice models. *Transportation Research: Part B: Methodological*, 35(7):643–666.

Swait, J. and Ben-Akiva, M. (1987a). Empirical test of a constrained choice discrete model: Mode choice in sao paulo, brazil. *Transportation Research: Part B: Methodological*, 22(2):103–115.

Swait, J. and Ben-Akiva, M. (1987b). Incorporating random constraints in discrete models of choice set generation. *Transportation Research: Part B: Methodological*, 22(2):91–102.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, 34(2):273–286.

Train, K. E. (2003). *Discrete Choice Methods with Simulation*. Cambridge University Press, New York.

Truong, T. D. (2013). Economic analysis of choice behavior: Incorporating choice set formation, non-compensatory preferences and perceptions into the random utility framework. *Unpublished PhD Thesis*.

Tversky, A. (1972). Elimination by aspects: a theory of choice. *Psychological Review*, 79(4):281–299.

Wasserman, S. and Faust, K. (1994). *Social Network Analysis : Methods and Applications*. Cambridge University Press, Cambridge; New York.

Wichmann, B. (2014). Estimation of non-market values in the presence of social network effects: the case of advisory referenda. *Unpublished*.

Appendix A

Tables

1. Each table is measured in a unique density of the social network. We test four different densities {0.05, 0.10, 0.25, 0.50}.

2. Table A.1.1-A.1.4, Table A.2.1-A.2.4 and Table A.3.1-A.3.4 are three groups of tables respectively display the parameter estimates of Monte-Carlo experiments, corresponding proportional RMSEs, and calculated distributions of probabilistic choice sets.

3. In each table of Table A.1.1-A.1.4 and A.2.1-A.2.4, the upper part displays the results estimated from an IAL model incorporating social interactions, while the lower part is from a model that ignores social interactions.

4. Each table in Table A.3.1-A.3.4 categorizes the results into two groups. The upper group is labeled with low level social interactions where $0.05 \leq \tau_1 \leq 0.40$. The lower group is for estimations in high level social interactions where $0.50 \leq \tau_1 \leq 0.80$.

Table A.1.1 Parameter estimates under 0.05 social network density

Social Network Density = 0.05									
	CSF process incorporates social interactions								
Par \ τ_1	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
ASC1 (1.5)	1.5005	1.5047	1.5095	1.4993	1.4959	1.5002	1.5011	1.4998	1.4975
ASC2 (1)	1.0069	1.0057	1.0067	1.0042	0.9993	0.9991	0.9989	0.9945	0.9950
Price (-3)	-3.0042	-3.0075	-3.0055	-3.0031	-2.9849	-3.0000	-3.0064	-3.0028	-3.0042
Quality (5)	5.0115	5.0062	4.9930	5.0062	5.0042	4.9981	5.0049	4.9955	5.0110
τ_0 (0.5)	0.4997	0.5006	0.5024	0.4996	0.4978	0.4995	0.5014	0.4967	0.4961
τ_1 (varied)	0.0497	0.1005	0.2008	0.2996	0.4002	0.5001	0.6000	0.7001	0.8000
μ (10)	10.0095	9.9908	10.0108	10.0176	10.0127	10.0039	10.0019	10.0018	10.0016
	CSF process ignores social interactions								
Par \ τ_1	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
ASC1 (1.5)	1.4995	1.5099	1.5048	1.4916	1.4838	1.3531	1.3200	1.3645	1.3894
ASC2 (1)	1.0060	1.0036	0.9931	0.9854	0.9788	0.8922	0.8823	0.9071	0.9302
Price (-3)	-3.0343	-3.0197	-3.0174	-3.0045	-2.9644	-2.6741	-2.5803	-2.6779	-2.7617
Quality (5)	5.002	5.0088	5.019	5.0582	5.0567	5.6248	5.8950	5.6737	5.5214
τ_0 (0.5)	0.4765	0.4522	0.4022	0.3494	0.2673	-1.5422	-7.1086	-8.5690	-9.2110
τ_1 (varied)	NA	NA	NA	NA	NA	NA	NA	NA	NA
μ (10)	10.0144	9.9956	10.0074	9.9854	9.9499	9.9445	9.8096	9.7457	9.7032

Note: True parameters are in the parenthesis.

Table A.1.2 Parameter estimates under 0.10 social network density

Social Network Density = 0.10									
	CSF process incorporates social interactions								
Par \ τ_1	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
ASC1 (1.5)	1.5036	1.5090	1.5023	1.5046	1.5068	1.4989	1.4981	1.5077	1.5033
ASC2 (1)	1.0068	1.0027	1.0012	1.0026	1.0041	1.0017	0.9963	1.0009	1.0020
Price (-3)	-3.0000	-3.0125	-2.9979	-3.0044	-3.0168	-3.0087	-2.9986	-3.0103	-3.0097
Quality (5)	5.0046	5.0135	5.0052	4.9978	4.9987	4.9985	5.0098	4.9994	5.0001
τ_0 (0.5)	0.5003	0.5012	0.5008	0.5011	0.5010	0.4997	0.4986	0.4985	0.4954
τ_1 (varied)	0.0496	0.1004	0.1991	0.2993	0.4000	0.5002	0.6001	0.7000	0.8000
μ (10)	9.9798	10.0165	10.0171	10.0306	9.9953	9.9964	10.0091	10.0013	10.0022
	CSF process ignores social interactions								
Par \ τ_1	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
ASC1 (1.5)	1.5012	1.5069	1.5151	1.4984	1.4915	1.3415	1.3122	1.3727	1.3981
ASC2 (1)	1.0033	0.9996	0.9988	0.9904	0.9849	0.8902	0.8743	0.9132	0.9324
Price(-3)	-3.0125	-3.0311	-3.0047	-2.9961	-2.9998	-2.6460	-2.5632	-2.6901	-2.7588
Quality(5)	4.9820	5.0061	5.0419	5.0434	5.0331	5.6550	5.8688	5.6883	5.5241
τ_0 (0.5)	0.4761	0.4522	0.4017	0.3507	0.2943	-1.5003	-7.0395	-8.6223	-9.2174
τ_1 (varied)	NA	NA	NA	NA	NA	NA	NA	NA	NA
μ (10)	10.0044	10.0393	10.0041	9.9600	9.9704	9.9285	9.8084	9.7466	9.7051

Note: True parameters are in the parenthesis.

Table A.1.3 Parameter estimates under 0.25 social network density

Social Network Density = 0.25									
	CSF process incorporates social interactions								
Par \ τ_1	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
ASC1 (1.5)	1.5048	1.5078	1.5002	1.5032	1.4993	1.5060	1.5030	1.5076	1.4956
ASC2 (1)	1.0087	1.0042	1.0008	1.0043	1.0061	1.0035	1.0061	1.0072	0.9994
Price (-3)	-3.0137	-3.0043	-2.9970	-3.0013	-3.0013	-2.9986	-3.0058	-3.0055	-3.0060
Quality (5)	5.0181	5.0194	4.9920	5.0028	5.0165	4.9937	4.9956	4.9954	5.0014
τ_0 (0.5)	0.5012	0.5009	0.5003	0.5001	0.4986	0.5005	0.5010	0.5012	0.4944
τ_1 (varied)	0.0498	0.1008	0.1999	0.3001	0.4001	0.5001	0.5999	0.7000	0.8000
μ (10)	9.9857	10.0005	10.0131	10.0042	10.0117	9.9886	9.9959	9.9990	10.0002
	CSF process ignores social interactions								
Par \ τ_1	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
ASC1 (1.5)	1.5090	1.5031	1.5024	1.4923	1.4800	1.3499	1.3153	1.3754	1.4037
ASC2 (1)	1.0099	0.9951	0.9937	0.9900	0.9749	0.8914	0.8848	0.9227	0.9393
Price (-3)	-3.0217	-3.0161	-2.9996	-2.9877	-2.9708	-2.6433	-2.5642	-2.6819	-2.7875
Quality (5)	5.0181	5.0347	5.0141	5.0251	5.0593	5.6553	5.8604	5.6940	5.5319
τ_0 (0.5)	0.4763	0.4515	0.4008	0.3504	0.2814	-1.3914	-7.1379	-8.5985	-9.2605
τ_1 (varied)	NA	NA	NA	NA	NA	NA	NA	NA	NA
μ (10)	9.9971	9.9982	9.9763	9.947	9.942	9.9483	9.8071	9.7488	9.7039

Note: True parameters are in the parenthesis.

Table A.1.4 Parameter estimates under 0.50 social network density

Social Network Density = 0.50									
	CSF process incorporates social interactions								
Par \ τ_1	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
ASC1 (1.5)	1.5082	1.5019	1.5069	1.4962	1.5015	1.5003	1.4985	1.5020	1.4993
ASC2 (1)	1.0052	0.9972	1.0095	1.0068	0.9982	0.9921	0.9964	0.9994	1.0052
Price (-3)	-3.0213	-2.9988	-3.0051	-3.0087	-3.0020	-3.0085	-2.9938	-3.0072	-3.0016
Quality (5)	5.0156	5.0042	5.0110	5.0060	5.0122	5.0074	4.9979	5.0000	5.0110
τ_0 (0.5)	0.5007	0.4997	0.5003	0.4990	0.4989	0.4993	0.4996	0.4995	0.4950
τ_1 (varied)	0.0508	0.1004	0.1995	0.3000	0.3998	0.5000	0.6000	0.7000	0.8001
μ (10)	10.0223	10.0027	10.0193	9.9893	10.0108	10.0105	9.9968	10.0035	9.9950
	CSF process ignores social interactions								
Par \ τ_1	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
ASC1 (1.5)	1.5012	1.4980	1.5186	1.4911	1.4940	1.3487	1.3123	1.3595	1.3973
ASC2 (1)	0.9977	0.9942	1.0098	0.9941	0.9797	0.8871	0.8728	0.9017	0.9415
Price (-3)	-3.0445	-3.0201	-3.0434	-2.9915	-2.9901	-2.6537	-2.5617	-2.6763	-2.7662
Quality (5)	5.0128	5.0315	5.0399	5.0285	5.0664	5.6474	5.8773	5.6712	5.5194
τ_0 (0.5)	0.4758	0.4507	0.4026	0.3499	0.2897	-1.4747	-6.9164	-8.5924	-9.2268
τ_1 (varied)	NA	NA	NA	NA	NA	NA	NA	NA	NA
μ (10)	10.0079	10.0065	9.9880	9.9269	9.9661	9.9431	9.8108	9.7471	9.7031

Note: True parameters are in the parenthesis.

Table A.2.1 Proportional RMSEs of parameter estimates: density=0.05

Social Network Density = 0.05									
	CSF process incorporates social interactions								
Par \ τ_1	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
ASC1	0.0310	0.0298	0.0345	0.0304	0.0295	0.0251	0.0247	0.0270	0.0267
ASC2	0.0446	0.0451	0.0456	0.0443	0.0401	0.0397	0.0355	0.0393	0.0382
Price	0.0211	0.0196	0.0185	0.0134	0.0156	0.0153	0.0138	0.0146	0.0146
Quality	0.0117	0.0136	0.0121	0.0107	0.0107	0.0095	0.0082	0.0084	0.0074
τ_0	0.0329	0.0331	0.0329	0.0324	0.0360	0.0356	0.0379	0.0395	0.0494
τ_1	0.0488	0.0522	0.0129	0.0060	0.0026	0.0016	0.0005	0.0004	0.0003
μ	0.0053	0.0063	0.0057	0.0053	0.0040	0.0031	0.0010	0.0008	0.0006
	CSF process ignores social interactions								
Par \ τ_1	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
ASC1	0.0603	0.0773	0.0718	0.0634	0.0570	0.1176	0.1208	0.0943	0.0814
ASC2	0.0844	0.1152	0.1009	0.0928	0.0851	0.1368	0.1285	0.1083	0.0932
Price	0.0411	0.0516	0.0503	0.0403	0.0420	0.1232	0.1400	0.1077	0.0842
Quality	0.0235	0.0318	0.0359	0.0258	0.0278	0.1311	0.1790	0.1347	0.1050
τ_0	0.0545	0.0970	0.1955	0.3011	0.4654	4.0845	15.2172	18.1380	19.4220
τ_1	NA	NA	NA	NA	NA	NA	NA	NA	NA
μ	0.0117	0.0158	0.0158	0.0136	0.0122	0.0087	0.0190	0.0254	0.0297

Note: Proportional RMSEs are calculated based on equation (4.1) in Chapter 4.

Table A.2.2 Proportional RMSEs of parameter estimates: density=0.10

Social Network Density = 0.10									
	CSF process incorporates social interactions								
Par \ τ_1	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
ASC1	0.0344	0.0334	0.0260	0.0330	0.0245	0.0258	0.0279	0.0243	0.0273
ASC2	0.0478	0.0466	0.0431	0.0488	0.0343	0.0379	0.0378	0.0388	0.0381
Price	0.0218	0.0194	0.0160	0.0174	0.0165	0.0153	0.0159	0.0137	0.0174
Quality	0.0148	0.0115	0.0136	0.0145	0.0093	0.0076	0.0088	0.0076	0.0082
τ_0	0.0324	0.0314	0.0340	0.0354	0.0316	0.0339	0.0377	0.0398	0.0626
τ_1	0.0352	0.0155	0.0102	0.0047	0.0016	0.0009	0.0004	0.0002	0.0003
μ	0.0070	0.0059	0.0049	0.0053	0.0023	0.0022	0.0017	0.0006	0.0006
	CSF process ignores social interactions								
Par \ τ_1	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
ASC1	0.0611	0.0771	0.0678	0.0652	0.0549	0.1194	0.1263	0.0877	0.0767
ASC2	0.0848	0.1095	0.1011	0.0985	0.0790	0.1337	0.1348	0.1031	0.0887
Price	0.0402	0.0476	0.0486	0.0440	0.0394	0.1312	0.1456	0.1044	0.0851
Quality	0.0264	0.0338	0.0343	0.0270	0.0219	0.1375	0.1738	0.1379	0.1050
τ_0	0.0546	0.0966	0.1965	0.2986	0.4115	4.0006	15.0790	18.2446	19.4349
τ_1	NA	NA	NA	NA	NA	NA	NA	NA	NA
μ	0.0124	0.0165	0.0163	0.0123	0.0113	0.0074	0.0192	0.0253	0.0295

Note: Proportional RMSEs are calculated based on equation (4.1) in Chapter 4.

Table A.2.3 Proportional RMSEs of parameter estimates: density=0.25

Social Network Density = 0.25									
	CSF process incorporates social interactions								
Par \ τ_1	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
ASC1	0.0330	0.0314	0.0327	0.0264	0.0293	0.0255	0.0252	0.0279	0.0284
ASC2	0.0454	0.0415	0.0451	0.0388	0.0392	0.0344	0.0380	0.0395	0.0425
Price	0.0199	0.0187	0.0159	0.0147	0.0149	0.0154	0.0149	0.0172	0.0132
Quality	0.0151	0.0098	0.0087	0.0082	0.0082	0.0084	0.0078	0.0091	0.0071
τ_0	0.0301	0.0310	0.0319	0.0328	0.0350	0.0339	0.0347	0.0409	0.0537
τ_1	0.0217	0.0130	0.0048	0.0027	0.0021	0.0009	0.0006	0.0003	0.0002
μ	0.0074	0.0044	0.0062	0.0041	0.0043	0.0026	0.0015	0.0009	0.0005
	CSF process ignores social interactions								
Par \ τ_1	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
ASC1	0.0629	0.0750	0.0683	0.0617	0.0602	0.1199	0.1235	0.0854	0.0716
ASC2	0.0891	0.1022	0.1002	0.0892	0.0865	0.1293	0.1256	0.0969	0.0854
Price	0.0457	0.0496	0.0448	0.0432	0.0397	0.1290	0.1453	0.1068	0.0756
Quality	0.0282	0.0318	0.0283	0.0229	0.0258	0.1371	0.1721	0.1391	0.1072
τ_0	0.0526	0.0980	0.1985	0.2992	0.4371	3.7828	15.2758	18.1969	19.5211
τ_1	NA	NA	NA	NA	NA	NA	NA	NA	NA
μ	0.0130	0.0153	0.0126	0.0104	0.0088	0.0082	0.0193	0.0251	0.0296

Note: Proportional RMSEs are calculated based on equation (4.1) in Chapter 4.

Table A.2.4 Proportional RMSEs of parameter estimates: density=0.50

Social Network Density = 0.50									
	CSF process incorporates social interactions								
Par \ τ_1	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
ASC1	0.0315	0.0284	0.0305	0.0279	0.0267	0.0284	0.0248	0.0259	0.0250
ASC2	0.0467	0.0457	0.0461	0.0428	0.0377	0.0393	0.0383	0.0359	0.0404
Price	0.0218	0.0183	0.0190	0.0148	0.0120	0.0152	0.0133	0.0161	0.0130
Quality	0.0146	0.0118	0.0140	0.0098	0.0082	0.0095	0.0072	0.0084	0.0078
τ_0	0.0308	0.0312	0.0323	0.0334	0.0316	0.0353	0.0349	0.0389	0.0504
τ_1	0.0260	0.0102	0.0051	0.0014	0.0011	0.0007	0.0002	0.0002	0.0003
μ	0.0073	0.0059	0.0048	0.0033	0.0046	0.0034	0.0011	0.0008	0.0010
	CSF process ignores social interactions								
Par \ τ_1	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
ASC1	0.0620	0.0768	0.0668	0.0647	0.0555	0.1156	0.1256	0.0960	0.0764
ASC2	0.0838	0.1130	0.1017	0.0908	0.0809	0.1356	0.1320	0.1117	0.0867
Price	0.0422	0.0512	0.0496	0.0400	0.0352	0.1247	0.1467	0.1082	0.0812
Quality	0.0267	0.0339	0.0362	0.0249	0.0216	0.1363	0.1755	0.1352	0.1044
τ_0	0.0544	0.1002	0.1949	0.3002	0.4206	3.9493	14.8329	18.1849	19.4536
τ_1	NA	NA	NA	NA	NA	NA	NA	NA	NA
μ	0.0121	0.0181	0.0133	0.0137	0.0118	0.0076	0.0189	0.0253	0.0297

Note: Proportional RMSEs are calculated based on equation (4.1) in Chapter 4.

Table A.3.1 Probabilistic choice sets distribution in the True, Network and Non-network models: density=0.05

Social Network Density = 0.05

Low level social propensity $(0.05 \leq \tau_1 \leq 0.40)$

	$\tau_1 = 0.05$			$\tau_1 = 0.10$			$\tau_1 = 0.20$			$\tau_1 = 0.30$			$\tau_1 = 0.40$		
Alts	*True*	*Net*	*NoNet*	*True*	*Net*	*NoNet*	*True*	*Net*	*NoNet*	*True*	*Net*	*NoNet*	*True*	*Net*	*NoNet*
{1}	0.1489	0.1488	0.1495	0.1372	0.1373	0.1381	0.1132	0.1139	0.1143	0.0910	0.0909	0.0908	0.0698	0.0688	0.0571
{2}	0.1363	0.1362	0.1371	0.1251	0.1252	0.1265	0.1029	0.1036	0.1047	0.0823	0.0822	0.0832	0.0634	0.0626	0.0525
{3}	0.1526	0.1525	0.1527	0.1408	0.1409	0.1408	0.1166	0.1174	0.1163	0.0941	0.0940	0.0924	0.0735	0.0725	0.0595
{1,2}	0.1251	0.1252	0.1253	0.1282	0.1282	0.1288	0.1338	0.1337	0.1354	0.1363	0.1364	0.1390	0.1345	0.1344	0.1361
{1,3}	0.1482	0.1483	0.1477	0.1542	0.1542	0.1531	0.1629	0.1627	0.1608	0.1666	0.1666	0.1639	0.1665	0.1665	0.1594
{2,3}	0.1370	0.1370	0.1370	0.1409	0.1408	0.1408	0.1473	0.1472	0.1475	0.1486	0.1486	0.1495	0.1456	0.1455	0.1431
{1,2,3}	0.1518	0.1519	0.1507	0.1737	0.1734	0.1719	0.2233	0.2215	0.2210	0.2810	0.2813	0.2812	0.3467	0.3498	0.3923

High level social propensity $(0.50 \leq \tau_1 \leq 0.80)$

	$\tau_1 = 0.50$			$\tau_1 = 0.60$			$\tau_1 = 0.70$			$\tau_1 = 0.80$		
Alts	*True*	*Net*	*NoNet*	*True*	*Net*	*NoNet*	*True*	*Net*	*NoNet*	*True*	*Net*	*NoNet*
{1}	0.0516	0.0514	0.0000	0.0360	0.0363	0.0000	0.0234	0.0227	0.0000	0.0144	0.0138	0.0000
{2}	0.0466	0.0464	0.0000	0.0325	0.0328	0.0000	0.0211	0.0205	0.0000	0.0129	0.0123	0.0000
{3}	0.0554	0.0552	0.0000	0.0397	0.0400	0.0000	0.0264	0.0256	0.0000	0.0167	0.0160	0.0000
{1,2}	0.1280	0.1280	0.0000	0.1180	0.1182	0.0000	0.1033	0.1023	0.0000	0.0861	0.0848	0.0000
{1,3}	0.1603	0.1603	0.0000	0.1498	0.1502	0.0000	0.1341	0.1329	0.0000	0.1141	0.1123	0.0000
{2,3}	0.1386	0.1385	0.0000	0.1269	0.1273	0.0000	0.1122	0.1111	0.0000	0.0937	0.0922	0.0000
{1,2,3}	0.4194	0.4202	1.0000	0.4972	0.4951	1.0000	0.5795	0.5849	1.0000	0.6621	0.6685	1.0000

Note:

1.Alts represents choice sets containing alternatives labeled as 1, 2 and 3.

2.Numbers displayed are proportions, indicating the ratio of that subset to the summation of all possible non-empty subsets..

Table A.3.2 Probabilistic choice sets distribution in the True, Network and Non-network models: density=0.10

Social Network Density = 0.10

Low level social propensity $(0.05 \leq \tau_1 \leq 0.40)$

	$\tau_1 = 0.05$			$\tau_1 = 0.10$			$\tau_1 = 0.20$			$\tau_1 = 0.30$			$\tau_1 = 0.40$		
Alts	*True*	*Net*	*NoNet*	*True*	*Net*	*NoNet*	*True*	*Net*	*NoNet*	*True*	*Net*	*NoNet*	*True*	*Net*	*NoNet*
{1}	0.1498	0.1500	0.1503	0.1373	0.1377	0.1383	0.1133	0.1137	0.1141	0.0908	0.0914	0.0915	0.0699	0.0703	0.0681
{2}	0.1367	0.1369	0.1375	0.1252	0.1255	0.1266	0.1032	0.1036	0.1047	0.0828	0.0832	0.0842	0.0629	0.0632	0.0622
{3}	0.1524	0.1526	0.1523	0.1401	0.1405	0.1403	0.1166	0.1171	0.1161	0.0937	0.0942	0.0927	0.0733	0.0736	0.0700
{1,2}	0.1245	0.1245	0.1248	0.1287	0.1286	0.1292	0.1341	0.1341	0.1359	0.1359	0.1360	0.1385	0.1346	0.1347	0.1378
{1,3}	0.1478	0.1478	0.1472	0.1541	0.1540	0.1529	0.1632	0.1631	0.1612	0.1675	0.1675	0.1645	0.1669	0.1671	0.1624
{2,3}	0.1371	0.1370	0.1371	0.1410	0.1410	0.1411	0.1470	0.1469	0.1472	0.1487	0.1488	0.1493	0.1462	0.1464	0.1465
{1,2,3}	0.1516	0.1512	0.1507	0.1736	0.1727	0.1718	0.2225	0.2215	0.2208	0.2806	0.2789	0.2793	0.3462	0.3449	0.3531

High level social propensity $(0.50 \leq \tau_1 \leq 0.80)$

	$\tau_1 = 0.50$			$\tau_1 = 0.60$			$\tau_1 = 0.70$			$\tau_1 = 0.80$		
Alts	*True*	*Net*	*NoNet*	*True*	*Net*	*NoNet*	*True*	*Net*	*NoNet*	*True*	*Net*	*NoNet*
{1}	0.0512	0.0511	0.0000	0.0359	0.0355	0.0000	0.0236	0.0232	0.0000	0.0143	0.0135	0.0000
{2}	0.0467	0.0466	0.0000	0.0325	0.0321	0.0000	0.0211	0.0207	0.0000	0.0128	0.0122	0.0000
{3}	0.0551	0.0549	0.0000	0.0394	0.0390	0.0000	0.0268	0.0265	0.0000	0.0167	0.0159	0.0000
{1,2}	0.1286	0.1286	0.0000	0.1184	0.1180	0.0000	0.1039	0.1034	0.0000	0.0866	0.0851	0.0000
{1,3}	0.1610	0.1609	0.0000	0.1500	0.1496	0.0000	0.1341	0.1335	0.0000	0.1150	0.1128	0.0000
{2,3}	0.1385	0.1385	0.0000	0.1270	0.1266	0.0000	0.1109	0.1105	0.0000	0.0939	0.0922	0.0000
{1,2,3}	0.4189	0.4195	1.0000	0.4968	0.4992	1.0000	0.5797	0.5822	1.0000	0.6607	0.6683	1.0000

Note:
1.Alts represents choice sets containing alternatives labeled as 1, 2 and 3.
2.Numbers displayed are proportions, indicating the ratio of that subset to the summation of all possible non-empty subsets..

Table A.3.3 Probabilistic choice sets distribution in the True, Network and Non-network models: density=0.25

Social Network Density = 0.25

Low level social propensity ($0.05 \leq \tau_1 \leq 0.40$)

Alts	$\tau_1 = 0.05$			$\tau_1 = 0.10$			$\tau_1 = 0.20$			$\tau_1 = 0.30$			$\tau_1 = 0.40$		
	True	*Net*	*NoNet*	*True*	*Net*	*NoNet*	*True*	*Net*	*NoNet*	*True*	*Net*	*NoNet*	*True*	*Net*	*NoNet*
{1}	0.1500	0.1504	0.1505	0.1377	0.1379	0.1383	0.1137	0.1138	0.1141	0.0908	0.0908	0.0913	0.0700	0.0693	0.0628
{2}	0.1367	0.1371	0.1376	0.1248	0.1250	0.1261	0.1029	0.1030	0.1042	0.0819	0.0820	0.0834	0.0632	0.0626	0.0577
{3}	0.1525	0.1530	0.1526	0.1402	0.1404	0.1400	0.1168	0.1169	0.1161	0.0941	0.0941	0.0931	0.0734	0.0727	0.0652
{1,2}	0.1246	0.1245	0.1249	0.1281	0.1280	0.1287	0.1335	0.1335	0.1352	0.1363	0.1363	0.1387	0.1345	0.1344	0.1369
{1,3}	0.1480	0.1478	0.1474	0.1542	0.1542	0.1532	0.1632	0.1631	0.1612	0.1672	0.1671	0.1643	0.1662	0.1661	0.1608
{2,3}	0.1367	0.1365	0.1366	0.1413	0.1413	0.1413	0.1470	0.1470	0.1473	0.1488	0.1488	0.1493	0.1457	0.1456	0.1446
{1,2,3}	0.1515	0.1506	0.1505	0.1736	0.1733	0.1724	0.2229	0.2226	0.2218	0.2810	0.2809	0.2798	0.3471	0.3492	0.3720

High level social propensity ($0.50 \leq \tau_1 \leq 0.80$)

Alts	$\tau_1 = 0.50$			$\tau_1 = 0.60$			$\tau_1 = 0.70$			$\tau_1 = 0.80$		
	True	*Net*	*NoNet*	*True*	*Net*	*NoNet*	*True*	*Net*	*NoNet*	*True*	*Net*	*NoNet*
{1}	0.0514	0.0515	0.0000	0.0361	0.0364	0.0000	0.0237	0.0239	0.0000	0.0143	0.0134	0.0000
{2}	0.0461	0.0463	0.0000	0.0322	0.0325	0.0000	0.0209	0.0212	0.0000	0.0131	0.0123	0.0000
{3}	0.0553	0.0554	0.0000	0.0397	0.0400	0.0000	0.0266	0.0269	0.0000	0.0168	0.0158	0.0000
{1,2}	0.1290	0.1290	0.0000	0.1178	0.1181	0.0000	0.1037	0.1040	0.0000	0.0862	0.0844	0.0000
{1,3}	0.1601	0.1602	0.0000	0.1491	0.1494	0.0000	0.1339	0.1344	0.0000	0.1141	0.1117	0.0000
{2,3}	0.1384	0.1385	0.0000	0.1267	0.1270	0.0000	0.1111	0.1115	0.0000	0.0932	0.0911	0.0000
{1,2,3}	0.4197	0.4191	1.0000	0.4983	0.4966	1.0000	0.5801	0.5783	1.0000	0.6623	0.6712	1.0000

Note:
1.Alts represents choice sets containing alternatives labeled as 1, 2 and 3.
2.Numbers displayed are proportions, indicating the ratio of that subset to the summation of all possible non-empty subsets..

Table A.3.4 Probabilistic choice sets distribution in the True, Network and Non-network models: density=0.50

Social Network Density = 0.50

Low level social propensity ($0.05 \leq \tau_1 \leq 0.40$)

	$\tau_1 = 0.05$			$\tau_1 = 0.10$			$\tau_1 = 0.20$			$\tau_1 = 0.30$			$\tau_1 = 0.40$		
Alts	*True*	*Net*	*NoNet*	*True*	*Net*	*NoNet*	*True*	*Net*	*NoNet*	*True*	*Net*	*NoNet*	*True*	*Net*	*NoNet*
{1}	0.1495	0.1496	0.1498	0.1377	0.1374	0.1380	0.1132	0.1134	0.1145	0.0908	0.0903	0.0911	0.0699	0.0695	0.0661
{2}	0.1362	0.1363	0.1369	0.1250	0.1248	0.1259	0.1035	0.1036	0.1057	0.0822	0.0818	0.0835	0.0632	0.0629	0.0608
{3}	0.1526	0.1527	0.1525	0.1405	0.1403	0.1401	0.1169	0.1170	0.1167	0.0943	0.0937	0.0931	0.0734	0.0730	0.0687
{1,2}	0.1251	0.1250	0.1254	0.1281	0.1282	0.1288	0.1343	0.1343	0.1358	0.1355	0.1356	0.1381	0.1349	0.1349	0.1379
{1,3}	0.1481	0.1481	0.1476	0.1539	0.1540	0.1530	0.1628	0.1628	0.1608	0.1672	0.1673	0.1644	0.1666	0.1665	0.1617
{2,3}	0.1369	0.1369	0.1368	0.1409	0.1410	0.1410	0.1465	0.1465	0.1465	0.1489	0.1489	0.1493	0.1461	0.1460	0.1456
{1,2,3}	0.1516	0.1514	0.1510	0.1738	0.1743	0.1732	0.2229	0.2223	0.2200	0.2811	0.2823	0.2805	0.3460	0.3473	0.3592

High level social propensity ($0.50 \leq \tau_1 \leq 0.80$)

	$\tau_1 = 0.50$			$\tau_1 = 0.60$			$\tau_1 = 0.70$			$\tau_1 = 0.80$		
Alts	*True*	*Net*	*NoNet*	*True*	*Net*	*NoNet*	*True*	*Net*	*NoNet*	*True*	*Net*	*NoNet*
{1}	0.0518	0.0515	0.0000	0.0362	0.0361	0.0000	0.0237	0.0236	0.0000	0.0145	0.0137	0.0000
{2}	0.0460	0.0458	0.0000	0.0320	0.0319	0.0000	0.0207	0.0206	0.0000	0.0129	0.0122	0.0000
{3}	0.0552	0.0550	0.0000	0.0396	0.0395	0.0000	0.0265	0.0264	0.0000	0.0169	0.0160	0.0000
{1,2}	0.1285	0.1284	0.0000	0.1184	0.1183	0.0000	0.1031	0.1029	0.0000	0.0858	0.0840	0.0000
{1,3}	0.1605	0.1603	0.0000	0.1491	0.1490	0.0000	0.1335	0.1333	0.0000	0.1135	0.1114	0.0000
{2,3}	0.1384	0.1383	0.0000	0.1269	0.1268	0.0000	0.1117	0.1115	0.0000	0.0936	0.0916	0.0000
{1,2,3}	0.4196	0.4206	1.0000	0.4979	0.4985	1.0000	0.5808	0.5816	1.0000	0.6628	0.6710	1.0000

Note:
1.Alts represents choice sets containing alternatives labeled as 1, 2 and 3.
2.Numbers displayed are proportions, indicating the ratio of that subset to the summation of all possible non-empty subsets..

Appendix B

Figures

1. All groups of figures, Figure B.1.1-B.1.2, Figure B.2.1-B.2.2 and Figure B.3.1-B.3.2, are kernel density plots of estimated parameters, respectively for the price, quality and τ_0.

2. First plot in each group is for that of low level social interactions, while the second is for high level social interactions.

3. The vertical dashed line represents true parameter values.

Figure B.1.1 Price estimator under low level social interactions

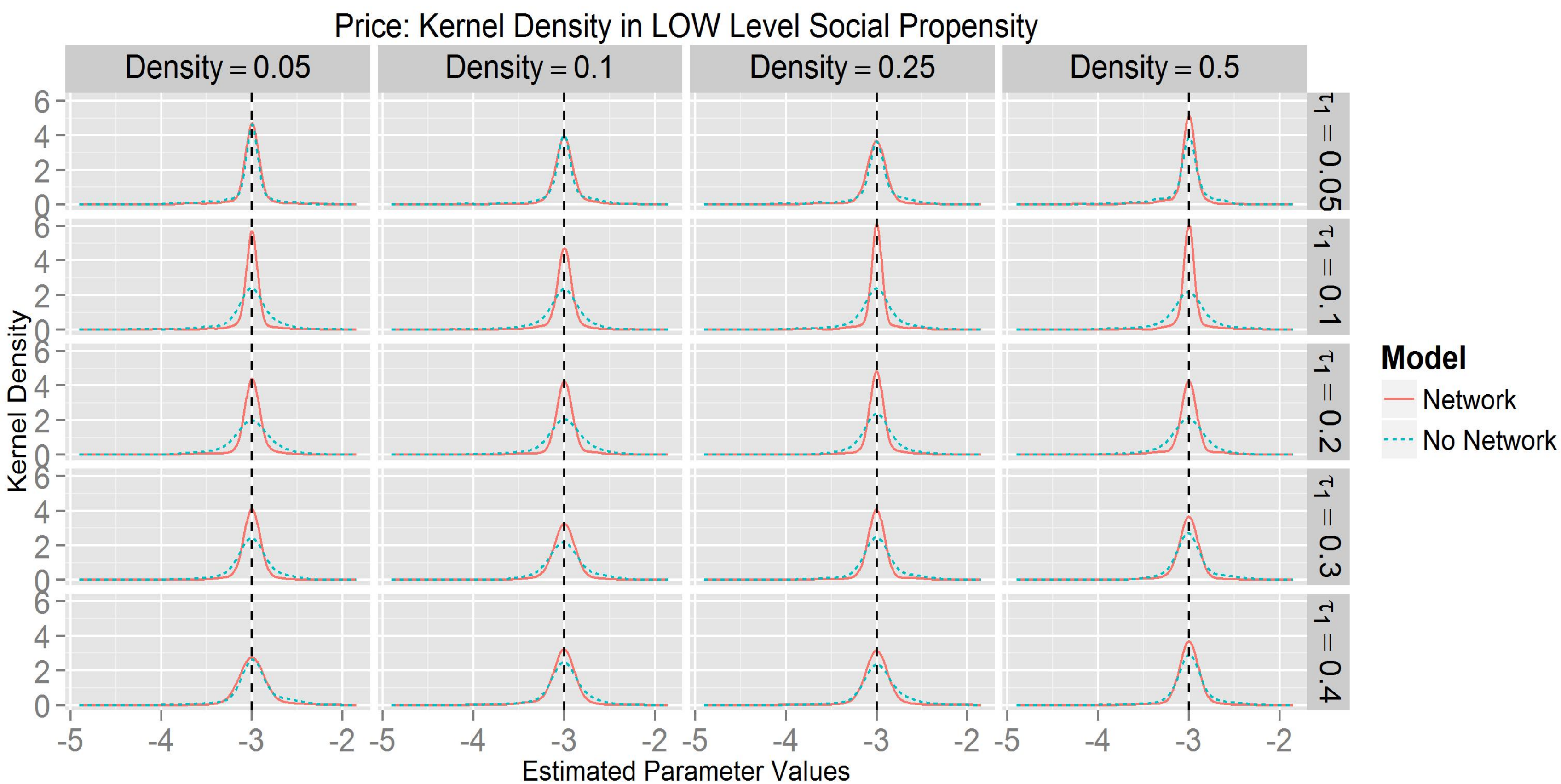

Figure B.1.2 Price estimator under high level social interactions

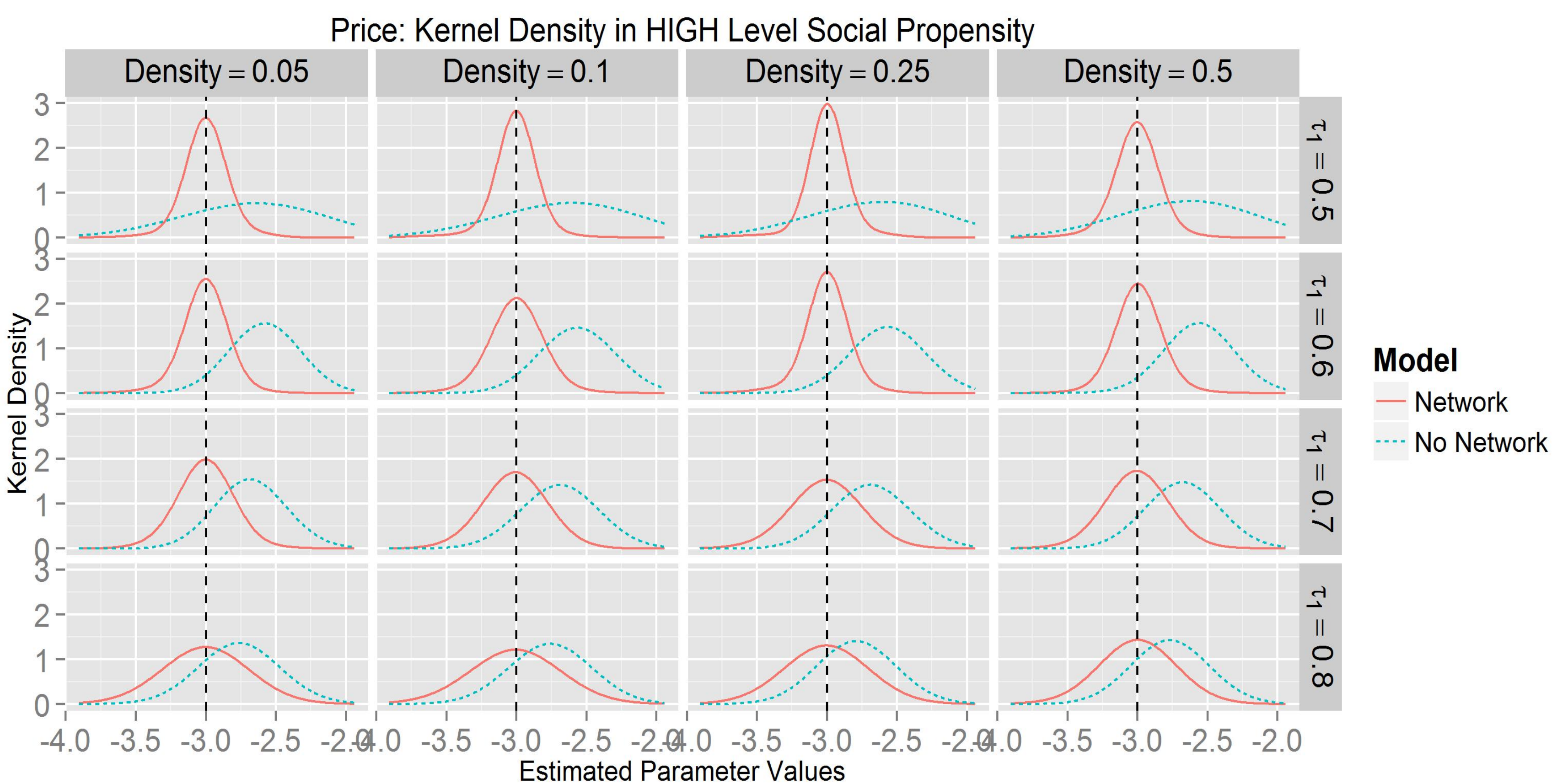

Figure B.2.1 Quality estimator under low level social interactions

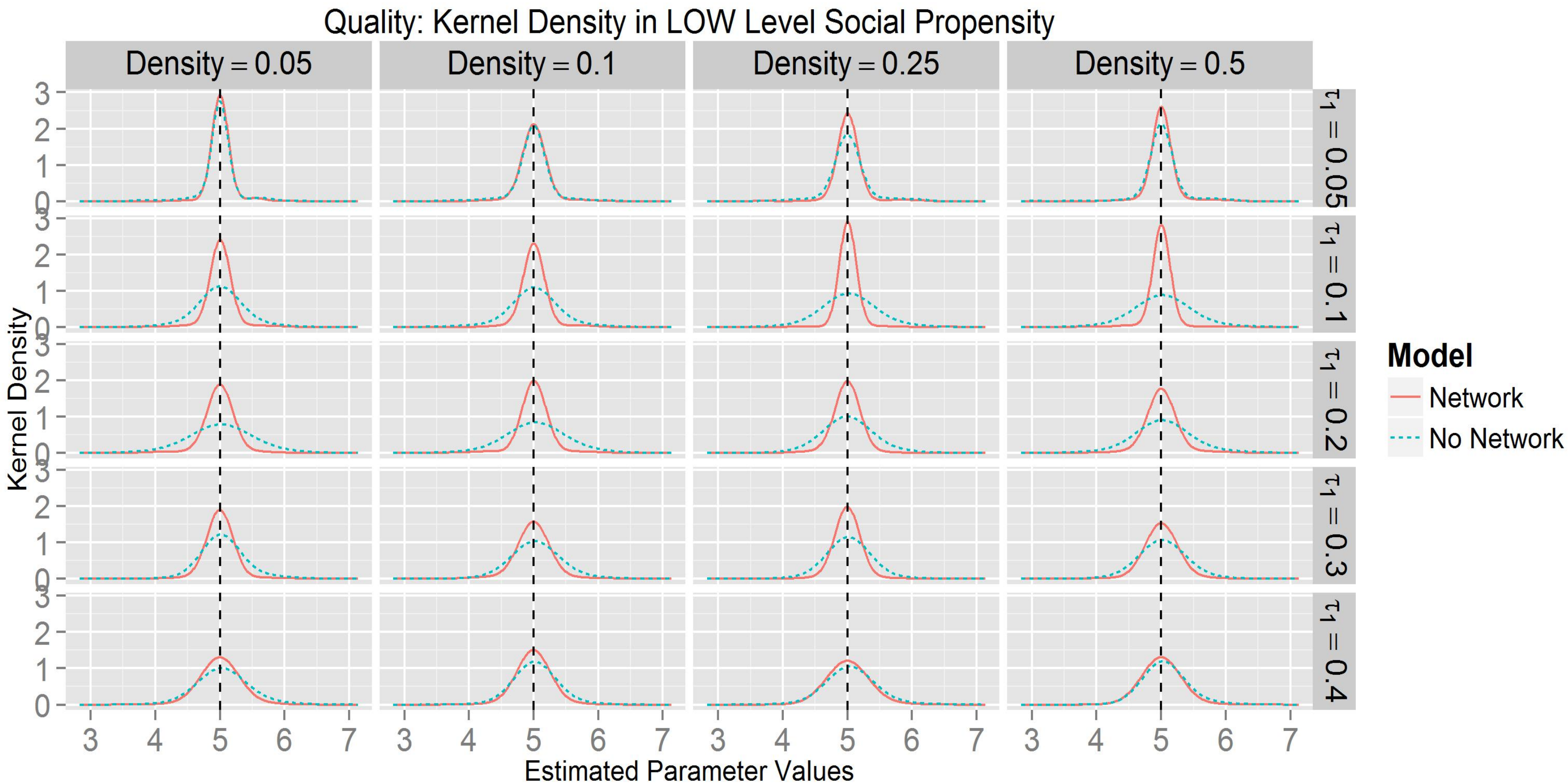

Figure B.2.2 Quality estimator under high level social interactions

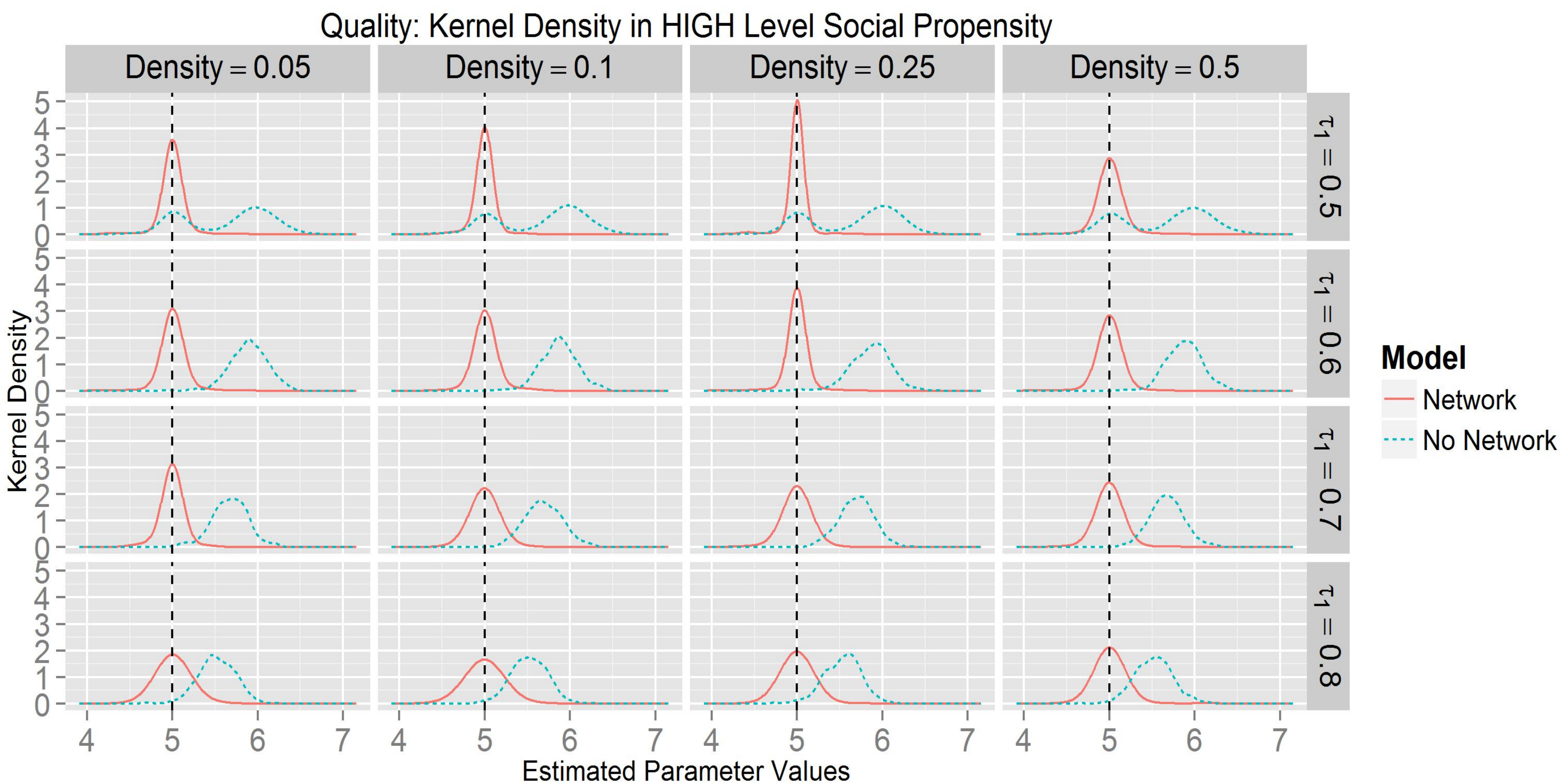

Figure B.3.1 τ_0 estimator under low level social interactions

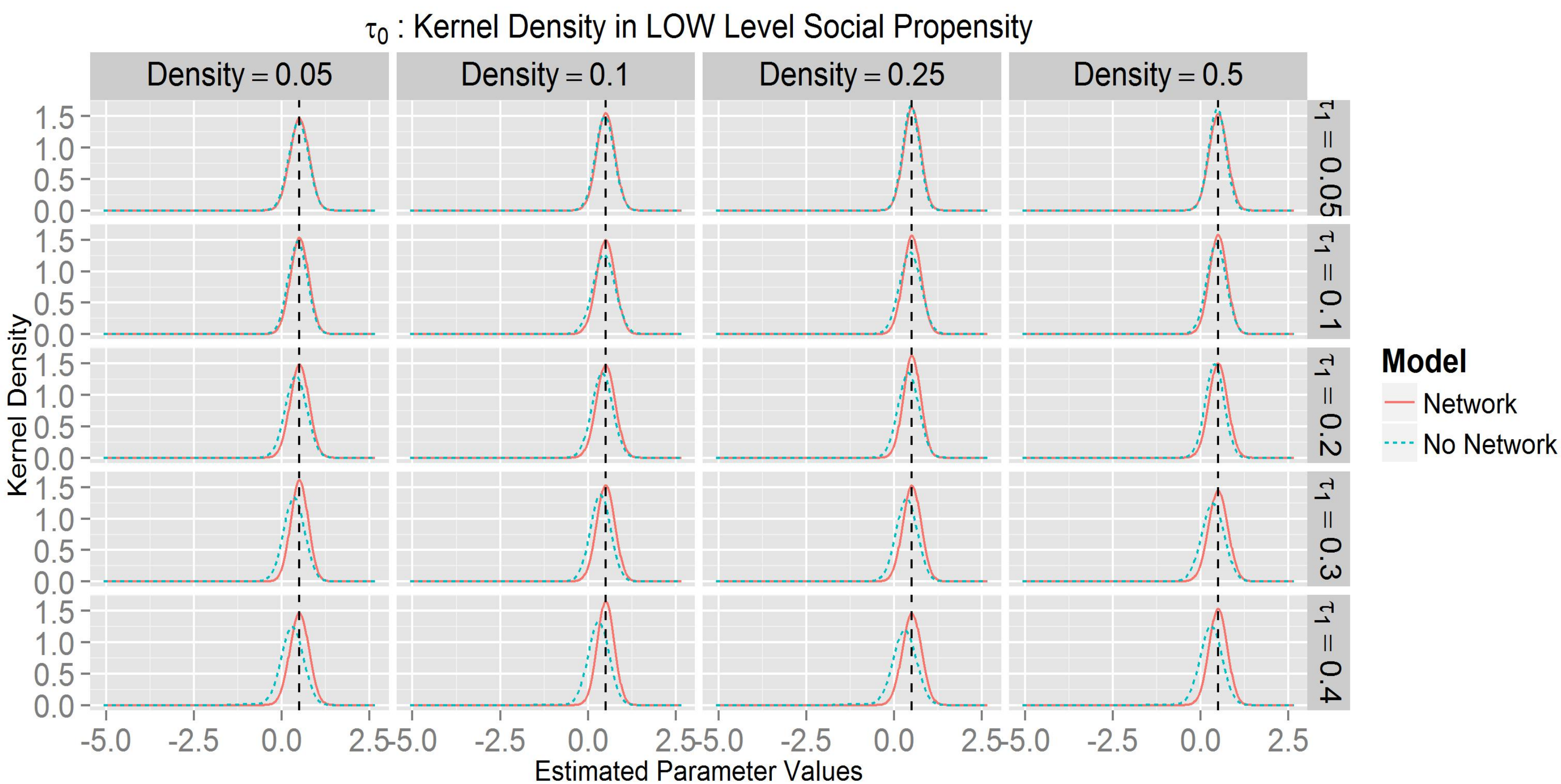

Figure B.3.2 τ_0 estimator under high level social interactions