# ESSAYS ON ECONOMETRIC IDENTIFICATION OF NETWORK AND CHOICE MODELS WITH LIMITED CONSIDERATION

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Matthew Kelly Thirkettle

May 2020

ESSAYS ON ECONOMETRIC IDENTIFICATION OF NETWORK AND CHOICE MODELS WITH LIMITED CONSIDERATION

Matthew Kelly Thirkettle, Ph.D.

Cornell University 2020

This dissertation is comprised of two papers. In the first paper (Chapter 2), I obtain informative bounds on network statistics in a partially observed network whose formation I explicitly model. Partially observed networks are commonplace due to, for example, partial sampling or incomplete responses in surveys. Network statistics (e.g., centrality measures) are not point identified when the network is partially observed. Worst-case bounds on network statistics can be obtained by letting all missing links take values zero and one. I dramatically improve on the worst-case bounds by specifying a structural model for network formation. An important feature of the model is that I allow for positive externalities in the network-formation process. The network-formation model and network statistics are set identified due to multiplicity of equilibria. I provide a computationally tractable outer approximation of the joint identified region for preferences determining network-formation processes and network statistics. In a simulation study on Katz-Bonacich centrality, I find that worst-case bounds that do not use the network formation model are 44 times wider than the bounds I obtain from my procedure.

The second paper (Chapter 3) is concerned about learning decision makers' (DMs) preferences using data on observed choices from a finite set of risky alternatives with monetary outcomes. This chapter is coauthored with Levon Barseghyan and Francesca Molinari. We propose a discrete choice model with unobserved

heterogeneity in consideration sets (the collection of alternatives considered by DMs) and unobserved heterogeneity in standard risk aversion. In this framework, stochastic choice is driven both by different rankings of alternatives induced by unobserved heterogeneity in risk preferences and by different sets of alternatives considered. We obtain sufficient conditions for semi-nonparametric point identification of both the distribution of unobserved heterogeneity in preferences and the distribution of consideration sets. Our method yields an estimator that is easy to compute and that can be used in markets with a large number of alternatives. We apply our method to a dataset on property insurance purchases. We find that although households are on average strongly risk averse, they consider lower coverages more frequently than higher coverages. Finally, we estimate the monetary losses associated with limited consideration in our application.

## BIOGRAPHICAL SKETCH

Matthew Thirkettle joined the graduate program in the Department of Economics at Cornell University in 2014. His research interests include the econometrics of networks and the econometrics of choice under limited consideration. He is particularly interested in answering economic questions using models that make realistic assumptions about observed data. He holds a Masters of Art in Economics from Cornell University. He also holds a Bachelor Honours Degree of Science in Economics and Mathematics from the University of Canterbury in New Zealand, where he graduated with First Class Honours, as well as a Bachelor Degree of Science in Economics and Mathematics from the same institute. While at Cornell, Matthew had the opportunity to join the Office of the Chief Economist at Microsoft Research as a Graduate Research Intern where he investigated why software developers make contributions to social coding platforms such as GitHub. Prior to perusing his doctoral degree, Matthew visited the University of California at Berkeley in 2012 as part of his undergraduate degree requirement through the UCEAP Reciprocal Exchange Program. He also conducted economic research on New Zealand agricultural and environmental policy at Motu Economic and Public Policy Research in Wellington, New Zealand.

This dissertation is dedicated to Jackie and John Thirkettle.

## ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to my special committee Francesca Molinari, Levon Barseghyan, David Easley, and Jörg Stoye for their continued support and enthusiasm throughout the dissertation process. I am especially indebted to Francesca Molinari and Levon Barseghyan for their insurmountable encouragement, for their intellectual guidance, and for their friendship.

I thank everyone who provided me with helpful suggestions on this research, including Isha Agarwal, Abhishek Ananth, Panle Jia Barwick, Giulia Brancaccio, Anne Burton, Angela Cools, Maura Coughlin, Tommaso Denti, Thomas Eisenberg, James Elwell, Amanda Eng, Andrew J. Fieldhouse, Jorgen Harris, Katelyn Heath, Christopher Huckfeldt, Jean-Francois Houde, Malin Hu, Eleonora Patacchini, Sida Peng, Penny Sanders, Seth Sanders, Caroline Walker, and seminar participates at Cornell University.

I gratefully acknowledge financial support provided by Sage Fellowship, the Tapan Mitra endowment, the L.R. “Red” Wilson endowment, the Ernest Liu Family endowment, and the Howard and Abby Milstein foundation.

Finally, I thank Katelyn Heath for her immeasurable love, support, and hours spent discussing and reading this dissertation. I also thank both my immediate and extended family, including (but not limited to) Jackie, John, Jamie, and Daniel Thirkettle and the Heath Family for their inspiration, confidence, and encouragement.

# TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

CHAPTER 1

# INTRODUCTION

This dissertation is comprised of two papers on two seemingly different topics, but are linked through the common theme of *Econometric Identification*. Identification addresses the question of: What can we learn about an economic model from the distribution of observables that generated the data (Lewbel, 2019)? Determining what conditions are required for identification is a fundamental step for any economic model and the answer can guide the practitioner on how to estimate the model.

To illustrate this idea, suppose the researcher is only willing to maintain assumptions that yield a *set identified* model (i.e., there is more than one parameterization of the economic model that is congruent with the distribution of observables). Suppose, however, the researcher is unaware that the model is set identified and mistakenly reports an estimator as if the model is *point identified* (i.e., there is exactly one parameterization of the economic model that is congruent with the distribution of observables). Concretely, suppose that the estimator is based on maximizing a likelihood function. Since the model is set identified, the population likelihood function is flat around a solution and, consequently, the reported estimator depends on the starting value for the optimization routine chosen by the researcher. There are two implications of this. First, the reported estimator may have extremely different economic implications (e.g., welfare calculations) compared to other solutions obtained from different starting points. Second, the statistical properties (e.g., confidence intervals) of the estimator are not correct. Therefore, it is imperative in any economic model that both identification and the type of identification (set or point) is established as a first step.

Chapter 2 is concerned about learning certain features of a social network (data describing friendship relationships) that is not completely observed by the researcher. Partially observed networks could arise due to, for example, partial sampling or incomplete responses in surveys. I explicitly model the formation of the network in order to recover the missing part of the network and hence obtain identification for certain network statistics. As a motivating example, suppose that a new vaccine is invented. A policy maker would like to inform households in a remote village about the new vaccine. However, due to financial constraints, the policy maker can only inform a small number of households about the vaccine. Diffusion centrality is a household level network statistic that can help guide the policy maker on which households to inform about the vaccine (Banerjee, Chandrasekhar, Duflo, & Jackson, 2013). Households with larger diffusion centrality are more likely to spread information throughout the social network. Therefore, to maximize the diffusion of information about the vaccine, the policy maker should target the household with the largest diffusion centrality if they are only able to target a single household.

There is, however, a double-identification issue when the network is partially observed. First, after fixing a parameterization of the network-formation model, there is more than one configuration of the network that is implied by the model (i.e., the model admits multiple equilibria). Therefore, the partially observed network cannot be uniquely completed and statistics of the social network are set identified. Second, it is well known that the primitives of games with multiple equilibria are set identified (Tamer, 2003). That is, more than one parameterization of the network-formation model is congruent with the distribution of observables. This exacerbates the set identification of diffusion centrality and other network statistics. I provide a computationally tractable method to approximate the joint identified region for preferences determining network-formation processes and network statistics.

In contrast, Chapter 3 provides sufficient conditions to point identify the primitives of a discrete choice model under risk that allows for limited consideration. Many of the choices that we make are made in a world with uncertainty about the future. However, the existing models of decision making under risk struggle to explain consumers' market behavior in the context of risky products, such as insurance choice and pension plans. This not only applies to expected utility theory (EUT), but also the behavioral models that are designed to fix the shortcomings of EUT. The intuition behind why these models breakdown is the fact that they smoothly spread choices around the first best alternative. Yet, the alternatives that are close to the first best are chosen at a much lower frequency in many empirical applications. At the same time, there is ample evidence to suggest that decision makers are context dependent, e.g., they have limited consideration. Decision makers have limited consideration when they first reduce the universal choice set to a smaller set (the consideration set) before making a choice. Limited consideration could be the result of a variety of mechanisms, including limited cognitive ability, firm manipulation, and objective reasons such as unobserved budget or liquidity constraints. A model of limited consideration can explain the low choice frequencies of alternatives that are close to the first best by imposing lower consideration probabilities on those alternatives.

We propose a discrete choice model with unobserved heterogeneity in consideration sets (the collection of alternatives considered by decision makers) and unobserved heterogeneity in standard risk aversion. Stochastic choice is driven both by different rankings of alternatives induced by unobserved heterogeneity in risk preferences and by different sets of alternatives considered. We obtain sufficient conditions for semi-nonparametric point identification of both the distribution of unobserved heterogeneity in preferences and the distribution of consideration sets. Our identifying assumptions are weaker than what is required in existing limited consideration models. In particular, we allow for the possibility that the relative prices across alternatives do not vary – a common feature in insurance markets. Our method yields an estimator that is easy to compute and can be used in markets with a large number of alternatives. We apply our method to a dataset on property insurance purchases. We find that although households are on average strongly risk averse, they consider lower coverages more frequently than higher coverages. Intuitively, lower coverage insurance contracts are relatively less expensive, and individuals appear to be focusing on the insurance premium.

CHAPTER 2

# IDENTIFICATION AND ESTIMATION OF NETWORK STATISTICS WITH MISSING LINK DATA

Matthew Thirkettle

## 2.1 Introduction

A wide array of economic outcomes of interest are generated through processes that involve the social interaction of individuals. In a classroom setting, for example, students' schooling effort and subsequent test scores are determined in part by their friends' effort provision through the process of knowledge spillovers. As another example, information about vaccines in developing countries and consequently vaccine uptake is spread through word of mouth. The position of an individual determines, at least in part, her economic outcomes as well as her influence on the economic outcomes of others. Centrality measures are network statistics that allow researchers to parsimoniously capture different features of an individual's network position. In practice, however, parts of the network are often unobserved due to subsampling or low response rates to surveys inquiring about social interactions (e.g., Banerjee et al. (2013)). Consequently, centrality measures and other statistics of the network are not point identified.

I obtain informative bounds on network statistics and their impact on economic outcomes of interest in a partially observed network whose formation I explicitly model. My method to recover bounds on a network statistic applies to *social networks*, e.g., friendships between students in a classroom. I propose a model in which the network is endogenous and individuals choose their friends based on their preferences. A distinctive feature of my network-formation model is that preferences depend on the topology of the network, e.g., the popularity of others. As a result, individuals strategically choose their friends and some individuals may only form friendships with popular individuals. I next use this network-formation model to infer the missing portion of the network. The network-formation model admits multiple equilibria, resulting in partial identification of the model. I recover bounds on the missing portion of the network and hence bounds on network statistics of interest. My main theoretical result is in obtaining a joint outer region for both the preferences determining network formation processes and for network statistics. This result applies broadly to a range of network statistics including intercentrality (Ballester, Calvó-Armengol, & Zenou, 2006), diffusion centrality (Banerjee et al., 2013), and Katz-Bonacich centrality (KBC) (Katz, 1953; Bonacich, 1987).[1]

The problem of partially observed network data is common in applied research. Due to resource constraints, researchers may only elicit social interaction information from a subsample of the population of interest – whether households in remote villages (Banerjee et al., 2013) or students in a school (Add Health, Harris (2009)). Low response rates also result in partially observed networks. For example, the response rate in the Add Health dataset ranges from 77.5% to 88.6%. My model allows for this type of partial network information as well as any setting where network links are missing at random, but the identity of the individuals are known. I assume that the researcher partially observes a large number of undirected social networks. Each network is a cross-sectional snapshot and can be thought of as a market with a fixed population of individuals. These partially observed networks may, for example, come from a random survey.[2] I assume the characteristics that

[1]Many more examples are given in Bloch, Jackson, and Tebaldi (2019).

[2]In the survey example, the researcher obtains information on all the social interactions between individuals in the subsample. In addition, the researcher can observe the connections that individuals in the subsample have with individuals not in the subsample. Connections

influence social connections, e.g, gender, race, ect. are observed for all individuals in the network. While I do allow for characteristics that are unobservable to the researcher, these characteristics must satisfy an independence assumption commonly maintained in the literature.

The network is endogenously formed according to a structural model, which governs how people choose their friends. For example, students in a classroom form friendships with other students based on their characteristics and the popularity of other students. This structural framework allows me to both partially reconstruct the unobserved component of the network and execute counterfactual analysis where, for instance, a set of individuals or links are removed from the network. A distinctive feature of my network-formation model relative to previous work on identification with sampled networks (Chandrasekhar & Lewis, 2016) is that it allows for positive externalities. Social actors derive positive utility from connections with popular individuals and from individuals with whom they have many mutual friends. Allowing for positive externalities is important for obtaining good model fit. Standard models without spillovers predict that triads – three people that are connected with one another – form at much lower frequencies than what is found empirically, and allowing for spillovers is critical in correcting this issue (Jackson et al., 2008; Graham, 2016). Positive externalities also imply strategic behavior; there is ample evidence to suggest that individuals act strategically when forming friendships.[3]

---

between individuals who are not in the subsample are not observed. Consequently, the *star network* is obtained for each network (Chandrasekhar & Lewis, 2016), see Figures A.6.1a and A.6.1b.

[3]See, for example, Jackson and Wolinsky (1996); Bala and Goyal (2000); Echenique, Fryer Jr, and Kaufman (2006); Jackson et al. (2008); Currarini, Jackson, and Pin (2009); Leung (2015b); Mele (2017); Badev (2018); Sheng (2018); Gualdani (2019).

Identification and estimation in my framework is challenging. The network-formation model admits multiple equilibria. Consequently, the model is incomplete (Tamer, 2003) without further restrictions on the function that selects which particular equilibrium is realized in cases of multiplicity – generally called the *selection mechanism* in the partial identification litearture (Tamer, 2010; Molinari, 2019).[4] The *sharp identified region* contains all network-formation parameters such that the parameterized model is consistent with the observed data. I theoretically characterize the sharp identified region for the network-formation model under partially observed network data using an approach similar to that described in Gualdani (2019) and Molinari (2019). The sharp region cannot, however, be feasibly computed, because doing so would entail checking that $O(2^{2^{\frac{n(n-1)}{2}}})$ moment inequalities are satisfied, with $n$ denoting the number of individuals in the network (in a small classroom with 5 students, that amounts to $10^{308}$ moment inequalities). In light of this challenge, I make progress by implementing the following steps. I leverage a useful property of the network-formation model called *strategic complementarity* for the purpose of identification.[5] Strategic complementarity implies that the marginal value of a friendship is increasing as more friendships form in the network. As a result, all equilibria belong to an easily characterizable lattice (Miyauchi, 2016), which I call the *admissible lattice*. I prove that the admissible lattice can be computed in no more than $\frac{n(n-1)}{2} - n + 1$ evaluations of a matrix function. This is important for establishing computational feasibility of the model. I also show how to use the admissible lattice for identification of

[4]The resulting identification problem is similar to the well known one affecting inference in entry games (see, for example, Ciliberto and Tamer (2009); Beresteanu, Molchanov, and Molinari (2011)).

[5]Strategic complementarity is also referred to as supermodularity in the mathematical economics literature (Tarski et al., 1955; Topkis, 1978).

the network-formation model and to obtain bounds on KBC and other centrality measures.

My main theorem establishes an outer region (a set containing the sharp identified region) using these bounds coupled with subnetwork identification (Sheng, 2018). Subnetwork identification bounds the joint probability of a subnetwork and the full characteristic vector that determines friendships by: (1) the probability that the subnetwork is the unique equilibrium; and (2) the probability that the subnetwork is in the set of equilibria. I extend this framework in two ways. First, I take expectations with respect to characteristics of individuals not contained in the subnetwork. This reduces the number of moment inequalities resulting in a feasible method and does not suffer from inference issues related to many moments. Second, I explicitly use the lattice structure of the equilibria set to enhance the feasibility of the model.

To show the applicability of my paper to a general economics context, I present results from a simulation study. My method for identifying the network-formation model is 100 to 1000 times faster than existing methods based on a comparison of run-time between my method and the results reported in Sheng (2018). These comparisons are based on subnetworks of size two, and I expect my method to offer an even larger advantage for larger subnetworks. I report results based on subnetworks of up to size five. In the simulation study, I allow for two channels of spillovers in the network-formation process: popularity spillover and mutual friend spillover. With only one spillover channel, I obtain very tight bounds on the network-formation model and on KBC. In particular, the identified region for the popularity spillover is $[0.996, 1.012]$ when the true value is equal to 1. With respect to KBC, worst-case bounds range from 2.850 to 5.532, while the true value is

4.020.[6] In contrast, I obtain bounds on effort equal to $[4.016, 4.239]$ when applying my framework with one channel of spillovers. This is a 12 fold improvement on the worst-case bounds. When I allow for two channels of spillovers, I find fairly wide bounds on the network-formation parameters. However, the bounds on KBC remain informative, ranging from 3.678 to 3.730 when the true value is 3.690.

The rest of this paper is organized as follows. Section 2.2 describes the related literature. Section 2.3 details the data requirements. Section 2.4 provides details on centrality measures and the model. Section 2.5 discusses identification. Section 2.6 presents the application and Section 2.7 concludes. All proofs are relegated to Appendix A.2. Appendix Table A.1 summarizes all relevant notation for this paper.

## 2.2 Related Literature

I make contributions to two distinct literatures: (1) estimating network models and statistics with missing network data; and (2) identification, estimation, and computation of partially identified, structural network formation models. To do so, I leverage useful results from games with strategic complementaries.

There is a growing literature on estimating network models with missing network data, ranging from partial network data (e.g., survey data or misclassified links) to completely unobserved networks. My paper belongs to the literature on

[6] The worst-case bounds result from imposing no assumptions on the unobserved links. These can be thought of as Manski worst-case bounds (Manski, 1989, 2003). In the case when the marginal return to effort is positive for all individuals, these bounds are obtained by evaluating the unobserved entries in the network to 0 and 1. from which no assumptions are imposed.

partial network data. Chandrasekhar and Lewis (2016) is the closest paper to mine. They assume the researcher has network survey data and propose a two-step method for identifying network statistics. The network is generated according to an exogenous process and is applied to a model where economic outcomes are a linear function of a network statistic.[7] I extend their framework by endogenizing the network and allowing for strategic network formation. Liu (2013) provides conditions under which the structural parameters of the linear social interactions model are identified with a sampled network.

The focus of my paper is on identifying network statistics from a partially observed network in a more general context and not on the linear social interactions model *per se*. The equilibrium level of effort in the social interactions model is equal to KBC. Therefore, my framework easily extends to obtaining bounds on peer effort subject to availability of outcome data and maintaining assumptions outlined in Liu (2013). Partial network data may also result from link misclassification where, for example, two individuals are friends and the friendship is incorrectly recorded as not existing in the data. Lewbel, Qu, and Tang (2019) allow for link misclassification and maintain linear restrictions on the structural parameters of the linear social interactions model to obtain point identification. I assume that the links are correctly reported, but I do not impose linear restrictions on the structural parameters.

There is a separate literature where the network is completely unobserved.

[7] There is a large literature using exogenous processes to impute networks with misclassified nodes (i.e., individuals) and links, e.g., see Robins, Pattison, and Woolcock (2004); Smith and Moody (2013); Huisman (2014); Krause, Huisman, Steglich, and Sniiders (2018) and references within. These methods are not suitable for social network formation as they do not allow for strategic network formation.

Various assumptions have been proposed to obtain point identification for the data generating process governing network formation or the underlying structural parameters for a game played on the network. I do not maintain these assumptions. For example, de Paula, Rasul, and Souza (2018), Rose (2015), Gautier and Rose (2016), and Manresa (2016) identify and estimate the structural parameters of the social interactions model and de Paula et al. also estimate the network links when the network is completely unobserved. These papers assume that the network is sparse and require panel data consisting of multiple draws of outcomes on a fixed network. Boucher and Houndetoungan (2019) assume observability of aggregate network statistics and obtain point identification for the structural parameters of the social interactions model. Battaglini, Patacchini, and Rainone (2019) propose a model where one observes only the outcomes of legislators (legislative effectiveness) and put forward a new network competitive equilibrium concept. This can be thought of as a general market equilibrium where effectiveness is analogous to market-clearing prices. They assume that legislators optimally choose friends while taking effectiveness as given (i.e., they are price takers).

The second literature that I contribute to is on identification and estimation of network-formation models. I assume that the network forms according to a model with complete information and strategic complementarity. There are a variety of strategies that have been proposed to identify the structural parameters for this model, see Graham (2014), Chandrasekhar (2016), de Paula (2017), de Paula (2019), or Graham (2019) for an overview. I use subnetwork identification, which was first proposed by Sheng (2018). I extend this procedure, as mentioned in the introduction, by (1) integrating out characteristics of individuals not in the subnetwork, and (2) bounding the distribution of subnetworks using the admissible lattice (a set defined below that contains all network equilibria). Other identification strategies have also been suggested. For example, Miyauchi (2016) proposes the use of monotone network statistics to partially identify the structural parameters of the model. I can augment my identification procedure with monotone network statistics. However, these statistics may require full knowledge of the network, which is not feasible in my data setting. de Paula, Richards-Shubik, and Tamer (2015) consider a single, growing network. Restrictions are imposed on the richness of unobserved heterogeneity and the number of direct friends (i.e., the network is sparse). I do not impose these restrictions and, in particular, I allow each dyadic pair to receive an i.i.d taste shock. Menzel (2015) also considers a growing network and shows that the asymptotic probability of forming a link is summarized by a conditional inclusive value.[8] Menzel proposes a maximum likelihood estimator based on the asymptotic distribution after imposing additional assumptions for point identification. Mele (2017) proposes that individuals meet sequentially at random and obtains point identification for the network-formation parameters. Specifying the meeting process completes the model and results in a unique equilibrium. Christakis, Fowler, Imbens, and Kalyanaraman (2010), Mele and Zhu (2017), Badev (2018), Boucher (2018), and Hsieh, Lee, and Boucher (2019) also assume a similar meeting processes to obtain point identification. In contrast to these papers, I do not impose restrictions on the selection mechanism and, as a result, my method allows for any meeting process.

There are alternate models of network-formation to the one that I propose that can also capture positive externalities from friendship formation. For example, dynamic models can allow friendship formation to depend on past popularity of individuals in the network (Goldsmith-Pinkham & Imbens, 2013; Graham, 2016;

[8]The inclusive value is a sufficient statistic for the choice probabilities of the available alternatives in the standard Logit model for multinomial choice (Train, 2009).

Lee, Fosdick, & McCormick, 2018; Bykhovskaya, 2019). These dynamic models require rich panel network data. They also do not allow for contemporaneous strategic interaction and, as a consequence, admit a unique equilibrium. Incomplete information games with strategic interaction have also been proposed in the literature (Leung, 2015a; Song & van der Schaar, 2015; de Martí & Zenou, 2015; Ridder & Sheng, 2017). Incomplete information assumes that individuals do not observe taste shocks before choosing friends. One criticism of an incomplete information game is that it suffers from *ex-post* regret, where individuals would like to reevaluate the links after the network has formed. I assume that the observed static network is the equilibrium of a long-run game where individuals have all relevant information available to them. Based on their preferences, individuals are unwilling to remove friends or mutually form new friendships. As a result, my model does not suffer from *ex-post* regret.

I show how to theoretically characterize the sharp identified region for the network-formation model with partially observed network data. This theoretical result builds on Molinari (2019) who shows how to obtain the sharp identified region for the network-formation model that I specify in this paper when the network is fully observed. Under a different network-formation model than the one I consider, Gualdani (2019) shows how to theoretically characterize the sharp identified region for her game. While the sharp identified region cannot be feasibly computed in my model, it is an important ingredient for characterizing an outer region that both can be feasibly computed and is informative about the underlying structural parameters and the network statistic of interest.

Using theory from games with strategic complementarities (Topkis, 1978; Tarski et al., 1955; Milgrom & Roberts, 1990), Miyauchi (2016) provides us with a useful characterization of the set of pairwise stable networks. Similar characterizations have been applied in Nash games, finite level rationalizability, and two-sided matching games (Jia, 2008; Molinari & Rosen, 2008; Uetake & Watanabe, 2012; Nishida, 2014). A key insight of my paper is to use the characterization to achieve computationally tractable bounds for the network formation model and obtain bounds on network statistics, such as KBC, intercentrality, and diffusion centrality.

## 2.3 Data

Before diving into the theoretical section of this paper, I discuss data requirements. These are weak in the sense that I only require the observation of a small number of links between a set of individuals to identify the network-formation model (i.e., a subnetwork – see Figure A.6.2a). As a motivating example for how a partially observed network may arise, consider a classroom setting. The researcher collects information about students' social interactions, which are represented by the network adjacency matrix $G$ where $G_{ij} = 1$ if and only if individuals $i$ and $j$ are friends. However, due to financial and time constraints, the researcher subsamples a set of students $\bar{\boldsymbol{n}} \subset \boldsymbol{n}$ and asks who they are friends with. The interviewed students reveal social ties that they have with everyone in the school population. As a result, the connections between individuals in the subsample $\bar{\boldsymbol{n}}$ to individuals in $\boldsymbol{n}$ are revealed (shaded area of Figure A.6.1a). I assume that the network is undirected, so that $i$ is friends with $j$ if and only if $j$ is friends with $i$.[9] As a

[9] In certain contexts, we may believe that the network is directed. For example in a classroom, students may form friendships with the athletic star, but the star does not reciprocate. My framework can be easily extended to the case of directed networks by using directed network-

result, links from individuals in $\boldsymbol{n}$ to $\bar{\boldsymbol{n}}$ are also revealed and hence the shaded area in Figure A.6.1b is observed by the researcher. I partition the network into two components $G = (G^{\bar{\boldsymbol{n}}}, G^{-\bar{\boldsymbol{n}}})$, where $G^{\bar{\boldsymbol{n}}}$ represents information known to the researcher.

More generally, $G^{\bar{\boldsymbol{n}}}$ represents any network with missing links. The minimal requirement is that all links are observed between a small set of individuals in $\bar{\boldsymbol{n}} \subset \boldsymbol{n}$. This is a very weak requirement and holds, for instance, when the researcher asks one student in the classroom to reveal a randomly chosen friend. While links to individuals outside of $\bar{\boldsymbol{n}}$ are not required and the size of $\bar{\boldsymbol{n}}$ can trivially equal two, the bounds on the network statistic of interest are tighter when more information is available. I assume that the researcher observes $T$ partially observed networks $\{G_t^{\bar{\boldsymbol{n}}_t}\}_{t=1}^T$, where $G_t^{\bar{\boldsymbol{n}}_t}$ is a random set of links from the network $G_t$. These can be thought of as a cross-section from many markets. For instance, $T$ may count the number of schools surveyed (Harris, 2009) or the number of remote villages in India (Banerjee et al., 2013). I also assume that the researcher has access to covariates $\boldsymbol{x}_{it}$ for all individuals in $\boldsymbol{n}_t$ (e.g., from a census or school enrollment data).[10]

**Assumption 2.1** (Observational Assumption)**.** *The researcher observes an i.i.d. sequence* $\{G_t^{\bar{\boldsymbol{n}}_t}, \boldsymbol{x}_t\}_{i \in \boldsymbol{n}_t, t=1,\ldots,T}$ *with* $T \to \infty$*. The partially observed network* $G_t^{\bar{\boldsymbol{n}}_t}$ *contains a random sample of links from* $G_t$ *(i.e., links are missing at random).*

**Example 2.1** (Complete Survey Data)**.** *Consider the case where the partially observed network* $G_t^{\bar{\boldsymbol{n}}_t}$ *is constructed from a random survey of individuals in a popu-*

formation models, e.g., Gualdani (2019).

[10]My framework can be extended to the case where we only have survey data on $\boldsymbol{x}_t$. The network is reconstructed using the estimated distribution of $\boldsymbol{x}_t$. This, however, will result in wider bounds as I lose information.

*lation, $\bar{\boldsymbol{n}}_t \subset \boldsymbol{n}_t$. Suppose that all surveyed individuals reveal all of their friendships to the researcher. Assumption 2.1 holds this type of survey data.*

## 2.4 Model

My framework can be thought of as a two-stage model. In the first stage the network is formed according to a social network-formation model parameterized by $\theta \in \Theta$. In the second stage, a network statistic $d(G)$ results, where $d : \mathcal{G} \to B$ is a known function and $B \subseteq \mathbb{R}^p$. The statistic can be a simple function of the network $G$, e.g., the number of links, but also an endogenous outcome from a game played on the network. In the linear social interactions game, for example, individuals simultaneously choose effort. The equilibrium level of effort is equal to a statistic that has received much attention in the literature, the Katz-Bonacich Centrality. Other second-stage applications are also considered in the appendix, such as the game described in Battaglini and Patacchini (2018). The goal is to learn about $(\theta, d(G))$ given a distribution of observables (partially observed networks, characteristics, and outcomes) $P$, under the conditions imposed by the model. I first discuss important centrality measures and I illustrate worst-case bounds for these statistics when no assumptions are imposed on the missing links.

### 2.4.1 Network Statistics and Centrality Measures

Centrality measures are network statistics that allow researchers to parsimoniously capture different features of an individual's network position. For example, KBC has been used to characterize the way peer effects impact student outcomes (Calvó-

Armengol, Patacchini, & Zenou, 2009); intercentrality can identify the key player in a criminal network (Ballester et al., 2006); and diffusion centrality is a strong predictor of how information will spread across the network (Banerjee et al., 2013). For exposition, I first focus on KBC and then generalize to other centrality measures in Section 2.4.1.

### Katz-Bonacich Centrality

KBC was first proposed in Katz (1953) and further developed by Bonacich (1987). It is a weighted measure of all paths leading to a particular individual, with shorter paths receiving larger value. Individuals that have a lot of friends have a higher KBC than those with few or no friends, *ceteris paribus.* Popular individuals that are connected to popular individuals also have a higher KBC than popular individuals connected to isolated individuals.

**Definition 2.1** (Weighted Katz-Bonacich Centrality)**.** *Consider a network $G$ and fix a weighting vector $\boldsymbol{w} \in \mathbb{R}^n$ and a decay parameter $\lambda \in \mathbb{R}$. The* Weighted Katz-Bonacich Centrality *is*

$$d^{kbc}(G; \boldsymbol{w}, \lambda) \equiv \sum_{k=0}^{\infty} \lambda^k G^k \boldsymbol{w}.$$

*The* unweighted Katz-Bonacich Centrality *corresponds to $\boldsymbol{w} = (1, \ldots, 1)'$, so that the expression simplifies to $d^{kbc}(G; \lambda) = \sum_{k=0}^{\infty} \lambda^k G^k$.*

The weighting vector $\boldsymbol{w}$ allows for the possibility that certain individuals have larger influence over their friends' behavior. The weight $w_i$ can be negative in which case the individual is a *negative influencer.* A path is a sequence of friendships between two individuals. The matrix $G^k$ counts all (not necessarily unique) paths of length $k$. The decay parameter $\lambda$ controls the importance of longer paths:

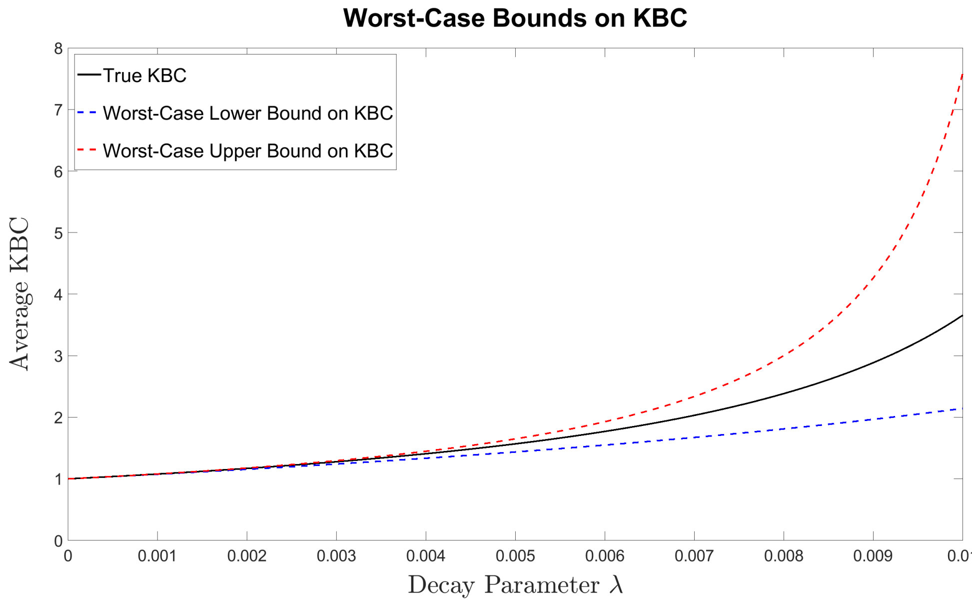



**Figure 2.1:** Worst-case bounds on peer effort. The solid line plots the true level of effort as a function of the social multiplier $\lambda$, and the dotted lines plot worst-case bounds. Results are reported for a network of size $n = 100$ with $\bar{n} = 25$ observed individuals in a Complete Network Survey. The network is generated according to the model described in Section 2.6 with $\gamma_1 = \gamma_2 = 0.5$ and $\theta_0 = 0$.

longer paths contribute less to the value of KBC when $\lambda$ is small. As $\lambda \to 0$, $d^{\text{kbc}}(G; \boldsymbol{w}, \lambda) \to \boldsymbol{w}$ in which case all paths are irrelevant. Denote the largest eigenvalue of $G$ by $\mu(G)$. If $\lambda\mu(G) < 1$, then KBC is finite and given by

$$d^{\text{kbc}}(G; \boldsymbol{w}, \lambda) = (I - \lambda G)^{-1}\boldsymbol{w}$$

Four economic applications relating to KBC are provided in Examples 2.8, 2.9, A.1.

KBC requires full knowledge of the network and cannot be computed from a partially observed network $G^{\bar{\boldsymbol{n}}}$. Intuitively, paths of length two – i.e., friends-of-friends – impact the value of KBC. Even in the case that we observe complete survey data (Example 2.1), not all friends-of-friends are not observed. Worst-case bounds (Manski, 1989, 2003) on KBC are obtained by imposing no restrictions on the unobserved links. Suppose that the weighting vector is positive, $\boldsymbol{w} \in \mathbb{R}^n_+$. Then

KBC is monotone in the network and worst-case bounds are obtained by setting the unobserved links to zero or one. Figure 2.1 illustrates worst-case bounds as a function of the decay parameter $\lambda$. When $\lambda$ is small, long paths have little influence and KBC is driven by the first order term $\lambda^0 G^0 \boldsymbol{w} = \boldsymbol{w}$ (normalized to one in this case). As a result, the discrepancy between $G$ and the worst-case assignments for its unobserved portion is irrelevant and the worst-case bounds' interval is thin. However, when the decay parameter is larger, the discrepancy becomes a serious issue and the worst-case bounds are very wide. This example illustrates the importance of accounting for unobserved links when estimating network statistics, centrality measures, or equilibrium behavior.

While worst-case bounds can be uninformative, they do not rely on a model. However, if we have a small amount of data on friendships, I can dramatically shrink the bounds. To accomplish this, I use a network-formation model and a clever computational approach. The network-formation model is discussed in Section 2.4.2 and the formal identification approach is detailed in Section 2.5. First, I provide generic bounds on network statistics (e.g., KBC) subject to $G$ belonging to a lattice – a set that has useful properties.

#### Bounding Centrality Measures

I later show that, under the assumptions imposed by my network formation model, all possible equilibria networks belong to a lattice. A lattice is a set of networks that are binary and componentwise bounded between two networks $\underline{G}$ and $\overline{G}$.

Formally,

$$\mathcal{L}(\underline{G}, \overline{G}) = \{G \in \mathcal{G} : \underline{G} \leq G \leq \overline{G}\}, \quad \text{where} \quad \underline{G} \leq \overline{G}.$$ [11]

The goal is to solve the following problem

$$\min_{G \in \mathcal{L}(\underline{G}, \overline{G})}/\max d(G), \tag{2.1}$$

where $d(G)$ is a network statistic of interest. The solution to Problem (2.1) improves on the worst-case bounds by using information about the model to restrict the set of possible networks to $\mathcal{L}(\underline{G}, \overline{G})$. Since $\mathcal{L}(\underline{G}, \overline{G})$ belongs to a discrete space, the naïve solution to Problem (2.1) is to check all networks in $\mathcal{L}(\underline{G}, \overline{G})$. If, however, $\underline{G}$ and $\overline{G}$ differ by $k$ elements, then this requires $2^k$ evaluations of $d(G)$ – even for relatively small values of $k$ this is not feasible. Monotonicity is a very useful property of a network statistic that delivers an analytical solution to Problem (2.1).

**Definition 2.2** (Monotonic Network Statistic)**.** *A network statistic $d(G)$ is* monotonically increasing *if and only if $d(G) \leq d(G')$ for all networks $G, G' \in \mathcal{G}$ such that $G \leq G'$. A network statistic $d(G)$ is* monotonically decreasing *if and only if $d(G) \geq d(G')$ for all networks $G, G' \in \mathcal{G}$ such that $G \leq G'$.*

Under monotonicity, the solutions to Problem (2.1) are the extreme points of the lattice.

**Lemma 2.1.** *Fix $\underline{G}$ and $\overline{G}$ and consider a network statistic $d(G)$.*

1. *If $d(G)$ is monotonically increasing, then the solutions to Problem* (2.1) *are $\underline{G}$ and $\overline{G}$, respectively.*

[11]The partial order $\leq$ refers to element-wise dominance. That is, $G \leq G'$ if and only if $\forall i, j$, $G_{ij} \leq G'_{ij}$.

2. *If $d(G)$ is monotonically decreasing, then the solutions to Problem* (2.1) *are* $\overline{G}$ *and* $\underline{G}$*, respectively.*

**Remark 2.1.** Lemma 2.1 is sharp in the sense that it delivers the exact solution to Problem (2.1). That is, if we know from a model and data that the true network $G$ belongs to $\mathcal{L}(\underline{G}, \overline{G})$ and we are unable to further refine this set, then Lemma 2.1 delivers the tightest possible bounds on the network statistic.

Lemma 2.1 establishes bounds on many network statistics, including KBC, provided that the sign of $w_i$ is the same for all individuals. I now provide a list of centrality measures that Lemma 2.1 applies to. Each of these centrality measures are useful in parsimoniously describing various forms of economic activity resulting from social interaction, and are commonly used in many practical applications.[12]

**Example 2.2** (Monotonic Katz-Bonacich Centrality). *Let* $d^{kbc}(G; \boldsymbol{w}, \lambda) \equiv \sum_{k=0}^{\infty} \lambda^k G^k \boldsymbol{w}$*, where* $\lambda \geq 0$*. Suppose that the largest eigenvalue* $\mu(\overline{G})$ *of* $\overline{G}$ *satisfies* $\lambda\mu(\overline{G}) < 1$*. In addition, suppose that the weights have the same sign:* $\boldsymbol{w} \in \mathbb{R}^n_+ \cup \mathbb{R}^n_-$*. Then KBC is well defined and is monotonic in the network. An economic application for the monotonic KBC is given in Example A.2.*

**Example 2.3** (Diffusion Centrality). $d_i^{df}(G; \lambda, K) = \sum_{k=1}^{K} \sum_{j=1}^{n} \lambda^k G_{ij}^k$*, where* $\lambda \geq 0$*. Diffusion centrality is equal to a truncation of the unweighted KBC. In particular, all non-unique paths up to length $K$ enter this measure; KBC is the limiting case with $K \to \infty$. Diffusion centrality was proposed by Banerjee et al.*

[12]These centralities are summarized in Bloch et al. (2019). Bloch et al. show that all centrality measures belong to a particular family that is characterized by set of axiom. That is, all centrality measures satisfies a particular set of axioms and, hence, all are related to one another. These axioms, however, do not imply that the statistic is monotonically increasing in the network. Indeed, eigenvector centrality is a counterexample.

*(2013) and Banerjee, Chandrasekhar, Duflo, and Jackson (2014) to capture the information flow of microfinance uptake in remote rural villages in India.*

**Example 2.4** (Degree Centrality). $d_i^{deg}(G) = \sum_{j=1}^n G_{ij}$. *Degree centrality is a measure of popularity. It is useful for understanding the friendship paradox and its implication on the statistical properties of peer effort (Jackson, 2019). Degree centrality has also been used to understand power laws and disease epidemics (Easley & Kleinberg, 2010).*

**Example 2.5** (Closeness Centrality). $d_i^{cc}(G) = \frac{n-1}{\sum_{j=1}^n \rho_{ij}(G)}$, *where* $\rho_{ij}(G)$ *is the shortest distance between* $i$ *and* $j$ *in network* $G$. *If* $i$ *is isolated, then* $d_i(G) = 0$.

A closely related centrality measure is harmonic centrality.

**Example 2.6** (Harmonic Centrality). $d_i^{hc}(G) = \sum_{j=1}^n \frac{n-1}{\rho_{ij}(G)}$. *Harmonic and closeness centrality have been proposed to capture the speed at which a message is transmitted through a network (Bavelas, 1950; Sabidussi, 1966).*

**Example 2.7** (Decay Centrality). $d_i^{dc}(G;\lambda) = \sum_{k=1}^{n-1}\sum_{j=1}^n \lambda^k \mathbb{1}(\rho_{ij}(G) = k)$. *While diffusion centrality counts all paths up to length* $K$ *and KBC counts all paths, decay centrality only counts shortest paths. The maximum length of the shortest is equal to* $n-1$. *Decay centrality has been proposed for optimal targeting/treatment in a social network (Banerjee et al., 2013; Chatterjee & Dutta, 2016; Tsakas, 2016a, 2016b) and can lead to the greatest amount of information diffusion about, e.g., a new vaccine.*

**Proposition 2.1.** *The centrality measures in Examples 2.2–2.7 are either monotonically increasing or decreasing in the network. Hence, solutions to Problem* (2.1) *for these centrality measures are given by Lemma 2.1.*

There exist network statistics that are not monotonic in the network. Examples include eigenvalue centrality, targeting centrality (Bramoullé & Genicot, 2018), and the nodal neighborhood statistic in Bloch et al. (2019). Despite this, there still exist computationally feasible bounds to Problem (2.1) for certain statistics. That is, I find $\underline{d}(\underline{G}, \overline{G})$ and $\overline{d}(\underline{G}, \overline{G})$ that are informative and satisfy

$$\underline{d}(\underline{G}, \overline{G}) \leq \min_{G \in \mathcal{L}(\underline{G}, \overline{G})} d(G) \quad \text{and} \quad \max_{G \in \mathcal{L}(\underline{G}, \overline{G})} d(G) \leq \overline{d}(\underline{G}, \overline{G}).$$

These bounds are solved on a case-by-case basis. I now present two important examples, for which I derive informative bounds in Propositions 2.2 and 2.3 below.

**Example 2.8** (Non-monotonic Katz-Bonacich Centrality)**.** *Let $d^{kbc}(G; \boldsymbol{w}, \lambda) \equiv \sum_{k=0}^{\infty} \lambda^k G^k \boldsymbol{w}$. and suppose that the largest eigenvalue of $\overline{G}$ denoted $\mu(\overline{G})$ satisfies $\lambda \mu(\overline{G}) < 1$. However, suppose that the weights do not have the same sign, so that $w_i > 0$ and $w_j < 0$ for some $i$ and $j$. KBC is not monotonic in $G$ in this case. As a concrete example for this statistic, consider the linear social interactions model. Taking the network $G$ as given, suppose that the students simultaneously choose schooling effort $y_i \in \mathbb{R}$ to maximize the following utility function:*

$$v_i(\boldsymbol{y}, G; \alpha_i, \phi) = \alpha_i y_i - \frac{1}{2} y_i^2 + \phi \sum_{j=1}^{n} G_{ij} y_i y_j. \tag{2.2}$$

*The first term $\alpha_i y_i - \frac{1}{2} y_i^2$ captures the direct benefit of effort. The term $\alpha_i$ allows for heterogenity in the marginal returns to effort and is typically modeled as a linear function of observable characteristics $\boldsymbol{z}$ and an unobservable idiosyncratic shock. The second term $\phi \sum_{j=1}^{n} G_{ij} y_i y_j$ captures local spillovers from direct friends exerting effort. The unique Nash equilibrium is $\boldsymbol{y}^*(G; \boldsymbol{\alpha}, \phi) = (I - \phi G)^{-1} \boldsymbol{\alpha} = c(G; \boldsymbol{\alpha}, \phi)$. That is, equilibrium effort is equal to KBC with weights $\boldsymbol{\alpha}$ and decay parameter $\phi$. In this model, the marginal returns to effort $\alpha_i$ can be negative or positive. Consequently, KBC is generally not monotonic in the network.*

**Example 2.9** (Intercentrality). *Denote intercentrality by* $d^{int}(G;\phi) \equiv \frac{d_i^{kbc}(\underline{G};\phi)}{(I-\phi G)^{-1}_{ii}}$. *Intercentrality is a non-linear transformation of KBC and is a network statistic that can provide a guide for determining the Key Player (Ballester et al., 2006; Zenou, 2016) – the individual who if removed would result in the largest decrease in total activity. In a criminal network, the Key Player is the person who law enforcement should target if the goal is to reduce criminal activity. The Key Player is the solution to:*

$$\min_{i=1,\dots,n} \sum_{j=1}^{n} y_i^*(G^{[-i]}),$$

*where* $y_i^*(G^{[-i]})$ *is the equilibrium level of effort after removing individual* $i$ *from the network. Under the assumptions on preferences detailed in their paper, Ballester et al. (2006) show that the Key Player is the individual with the largest intercentrality.*

**Proposition 2.2.** *Fix* $\boldsymbol{w}, \phi, \underline{G}$, *and* $\overline{G}$ *and order* $\boldsymbol{w}$

$$\boldsymbol{w} = (w_1, \dots, w_{n_+}, w_{n_++1}, \dots, w_n)$$

*such that* $\forall i \leq n_+ : w_i \geq 0$ *and* $\forall i > n_+ : w_i < 0$. *Suppose that the largest eigenvalue of* $\overline{G}$ *denoted* $\mu(\overline{G})$ *satisfies* $\phi\mu(\overline{G}) < 1$. *Define mappings* $\underline{d} : \mathcal{G} \times \mathcal{G} \to \mathbb{R}^n$ *and* $\overline{d} : \mathcal{G} \times \mathcal{G} \to \mathbb{R}^n$ *by*

$$\underline{d}(\underline{G}, \overline{G}; \phi) = [(I - \phi\underline{G})^{-1}_1, \dots, (I - \phi\underline{G})^{-1}_{n_+}, (I - \phi\overline{G})^{-1}_{n_++1}, \dots (I - \phi\overline{G})^{-1}_n]$$
$$\overline{d}(\underline{G}, \overline{G}; \phi) = [(I - \phi\overline{G})^{-1}_1, \dots, (I - \phi\overline{G})^{-1}_{n_+}, (I - \phi\underline{G})^{-1}_{n_++1}, \dots, (I - \phi\underline{G})^{-1}_n]$$

*where* $(I - \phi G)^{-1}_j$ *is the* $j^{th}$ *column of* $(I - \phi G)^{-1}_1$. *Then* $\forall i = 1, \dots, n$:

$$[\underline{d}(\underline{G}, \overline{G}; \phi)\boldsymbol{w}]_i \leq \min_{G \in \mathcal{L}(\underline{G}, \overline{G})} d_i^{kbc}(G; \boldsymbol{w}, \phi)$$

$$[\overline{d}(\underline{G}, \overline{G}; \phi)\boldsymbol{w}]_i \geq \max_{G \in \mathcal{L}(\underline{G}, \overline{G})} d_i^{kbc}(G; \boldsymbol{w}, \phi)$$

**Proposition 2.3.** *Fix* $\phi, \underline{G}$, *and* $\overline{G}$. *Suppose that the largest eigenvalue of* $\overline{G}$ *de-*

*noted $\mu(\overline{G})$ satisfies $\phi\mu(\overline{G}) < 1$. Then*

$$\frac{d_i^{kbc}(\underline{G};\phi)}{(I-\phi\overline{G})_{ii}^{-1}} \leq \min_{G\in\mathcal{L}(\underline{G},\overline{G})} d^{int}(G;\phi) \quad \textit{and} \quad \max_{G\in\mathcal{L}(\underline{G},\overline{G})} d^{int}(G;\phi) \leq \frac{d_i^{kbc}(\overline{G};\phi)}{(I-\phi\underline{G})_{ii}^{-1}}$$

**Remark 2.2.** There does not always exist a $G \in \mathcal{L}(\underline{G}, \overline{G})$ such that $\overline{d}(\underline{G}, \overline{G}; \phi) = (I - \phi G)^{-1}$, and hence the bounds given in Propositions 2.2 and 2.3 are generally not sharp. See Appendix A.2 for details.

I present one final example for non-monotonic statistics in the appendix, see Example A.1 and Proposition A.1.

When the researcher has access to partial network data, I have shown that worst-case bounds – bounds obtained by imposing no restrictions on the missing links – can be uninformative. I have obtained bounds on a wide range of centrality measures that are commonly used in the applied networks literature subject to the network belonging to a generic lattice. I will now specify a network-formation model that restricts the set of possible networks to a particular lattice. The model can be estimated using the partially observed data to provide informative bounds on the network statistic.

### 2.4.2 Network Formation Model

The network-formation model is off-the-shelf and is featured in many theoretical and structural network-related papers (Jackson & Wolinsky, 1996; de Paula et al., 2015; Currarini et al., 2009; Miyauchi, 2016; Sheng, 2018), none of which focus on partially observed networks. Specifically, I assume that individuals play a complete information, pairwise stable game and that links are undirected. One important feature of the model is that it allows for positive network externalities, resulting in individuals strategically choosing their friends.

Let $\boldsymbol{n} = \{1, \dots, n\}$ be the set of individuals in a population and let $i, j, k, l \in \boldsymbol{n}$ be arbitrary individuals. The network is encoded by the adjacency matrix $G$ where $G_{ij} = 1$ if and only if $i$ and $j$ are linked. The network is undirected, $G_{ij} = G_{ji}$, and there are no self loops, $G_{ii} = 0$. Denote the space of such matrices by

$$\mathcal{G} = \{G \in \mathbb{Z}_2^{n \times n} : G_{ii} = 0,\ G_{ij} = G_{ji} \quad \forall i, j \in \boldsymbol{n}\}.$$

The notation $G + \{ij\}$ denotes the network with the link $ij$ added (i.e., $G + \{ij\}$ has $k, l$ entry equal to $G_{kl}$ for all $kl \neq ij$ and $G_{ij} = 1$). Similarly, $G - \{ij\}$ denotes the network with the link $ij$ deleted.

Each individual $i \in \boldsymbol{n}$ is characterized by a vector of observable characteristics $\boldsymbol{x}_i \in \mathcal{X}$ (collect these in the matrix $\boldsymbol{x} = (\boldsymbol{x}_i)_{i \in \boldsymbol{n}}$), a matrix of preference shocks $\varepsilon \equiv (\varepsilon_{ij})_{ij \in \boldsymbol{n}} \in \mathcal{E}$, and a network utility function $\pi_i : \mathcal{G} \times \mathcal{X} \times \mathcal{E} \times \Theta \to \mathbb{R}$. The network utility function $\pi_i(\cdot, \cdot, \cdot; \theta)$ is parameterized by the same value $\theta_0$ for all individuals and it represents the value that individual $i$ places on network $G \in \mathcal{G}$. Marginal utility is a key ingredient in defining an equilibrium. In contrast to standard economic models, utility is a function of the binary network $G$. Hence, the marginal utility of a link is defined as the difference in utilities when the link is present and when it is not present.

**Definition 2.3** (Marginal Utility)**.** *The marginal utility of individual $i$ over link $ij$ is the mapping $\Pi_{ij} : \mathcal{G} \times \mathcal{X} \times \mathcal{E} \times \Theta \to \mathbb{R}$ defined as:*

$$\Pi_{ij}(G, \boldsymbol{x}, \varepsilon; \theta) \equiv \pi_i(G + \{ij\}, \boldsymbol{x}, \varepsilon; \theta) - \pi_i(G - \{ij\}, \boldsymbol{x}, \varepsilon; \theta).$$

I maintain the assumption of pairwise stability with non-transferable utility as an equilibrium condition.[13] Pairwise stability (Jackson & Wolinsky, 1996) ensures

[13] The game with transferable utility is discussed in Sheng (2018), and the game with directed

that any two individuals are unwilling to mutually create a link and no individual is willing to sever a link. Other equilibrium concepts have been proposed such as Nash equilibrium and coalition equilibrium. The Nash equilibrium concept is not attractive in network games due to coordination failure and, in particular, because the network with no links is a Nash equilibrium (Myerson, 1991; Calvó-Armengol & İlkılıç, 2009). Coalition equilibrium is a refinement of pairwise stability and ensures that no group of individuals are willing to renegotiate the set of links between them (Myerson, 1977; Jackson & Wolinsky, 1996). I do not impose coalition equilibrium as it is stronger than pairwise stability and I aim to impose minimal assumptions on equilibrium behavior.

**Definition 2.4** (Pairwise Stable Network)**.** *Given $\boldsymbol{x}$, $\varepsilon$, and a payoff function $\pi_i$, the network $G$ is said to be* pairwise stable *if the following conditions hold:*

1. *For all $i, j \in \boldsymbol{n}$ such that $G_{ij} = 1$,*

$$\Pi_{ij}(G, \boldsymbol{x}, \varepsilon; \theta) \geq 0 \quad \textit{and} \quad \Pi_{ji}(G, \boldsymbol{x}, \varepsilon; \theta) \geq 0,$$

2. *For all $i, j \in \boldsymbol{n}$ such that $G_{ij} = 0$,*

$$\textit{if} \quad \Pi_{ij}(G, \boldsymbol{x}, \varepsilon; \theta) > 0, \quad \textit{then} \quad \Pi_{ji}(G, \boldsymbol{x}, \varepsilon; \theta) < 0.$$

A key assumption I maintain is that marginal utility is monotonically increasing in the network. Intuitively, this assumes that there are positive externalities when individuals form friends and, in particular, allows students in the classroom to receive larger value from connecting to the popular student. This type of assumption is also referred to as strategic complementarity or supermodularity in

networks is discussed in Gualdani (2019).

the microeconomic theory literature (Tarski et al., 1955; Topkis, 1978). The assumption allows me to refine the set of networks to a lattice and guarantees the existence of an equilibrium.

**Assumption 2.2** (Monotonically Increasing). *Marginal utility is monotonically increasing in $G$: if $G \leq G'$, then $\Pi_{ij}(G, \boldsymbol{x}, \varepsilon; \theta) \leq \Pi_{ij}(G', \boldsymbol{x}, \varepsilon; \theta) \quad \forall i, j \in \boldsymbol{n}$.*

Monotonically increasing marginal utility is a strong assumption; it rules out competition and cannibalization effects present in many Industrial Organization applications (Berry & Jia, 2008; Jia, 2008; Nishida, 2014).[14] Nonetheless it holds in many applications of interest, such as models of social interaction where individuals tend to form connections with popular individuals.

In the application I assume the following functional form for the utility function.

**Assumption 2.3** (Linear Utility). *Utility is given by*

$$\begin{aligned}\pi_i(G, \boldsymbol{x}, \varepsilon; \theta) \equiv & \sum_{j \in \boldsymbol{n}} G_{ij}(u(\boldsymbol{x}_i, \boldsymbol{x}_j; \theta) + \varepsilon_{ij}) \\ & + \gamma_1 \frac{1}{n-1} \sum_{j,k \in \boldsymbol{n}: k \neq i} G_{ij} G_{jk} + \gamma_2 \frac{1}{n-2} \sum_{j,k \in \boldsymbol{n}: k \neq i} G_{ij} G_{ik} G_{jk},\end{aligned}$$

*so that marginal utility is linear in parameters:*

$$\Pi_{ij}(G, \boldsymbol{x}, \varepsilon; \theta) = u(\boldsymbol{x}_i, \boldsymbol{x}_j; \theta) + \varepsilon_{ij} + \gamma_1 \frac{1}{n-1} \sum_{k \in \boldsymbol{n}: k \neq i} G_{jk} + \gamma_2 \frac{1}{n-2} \sum_{k \in \boldsymbol{n}: k \neq i} G_{ik} G_{jk}.$$

This particular specification for the marginal utility function has three components. The first term, $u(\boldsymbol{x}_i, \boldsymbol{x}_j, \theta) + \varepsilon_{ij}$, is the direct benefit that $i$ gets by

[14] Jia (2008) shows that in a two-player network game with competition effects, a simple transformation of the model can be performed to yield entry decisions that satisfy strategic complementarity. This, however, cannot be generalized to markets with more than two players. The pairwise stable network game consists of $n > 2$ players and, therefore, monotonically increasing marginal utility is required.

connecting to individual $j$. This term allows for homophily – that is, individuals who have similar characteristics are more likely to become friends. Second, the popularity spillover is given by $\gamma_1 \frac{1}{n-1} \sum_{j,k \in \boldsymbol{n}: k \neq i} G_{ij} G_{jk}$. The term $\frac{\gamma_1}{n-1}$ is the marginal value that individual $i$ gets from having one more indirect link. Finally, the mutual friend spillover is given by $\gamma_2 \frac{1}{n-2} \sum_{j,k \in \boldsymbol{n}: k \neq i} G_{ij} G_{ik} G_{jk}$. The term $\frac{\gamma_1}{n-2}$ is the marginal value that individual $i$ gets from having one more mutual link. The two spillover terms are normalized by their maximum values $(n-1)$ and $(n-2)$, so that the spillovers takes values in $[0, \gamma_1]$ and $[0, \gamma_2]$, respectively. Under this utility specification marginal utility is monotonically increasing provided that the popularity and mutual friend spillovers are non-negative.

**Equilibrium Results**

There are three useful results related to the characterization and existence of the equilibrium that are based on games with strategic complementarities (Tarski et al., 1955; Topkis, 1978; Milgrom & Roberts, 1990; Jia, 2008), see Miyauchi (2016) for the network case. The first result characterizes pairwise stability in terms of a fixed-point mapping.

**Fixed-point Characterization.** *Fix $\boldsymbol{x}$, $\varepsilon$, and a payoff functions $\pi_i$. Define the mapping $V : \mathcal{G} \to \mathcal{G}$ by $V_{ij}(G) \equiv \mathbb{1}[\Pi_{ij}(G, \boldsymbol{x}, \varepsilon; \theta) \geq 0] \mathbb{1}[\Pi_{ji}(G, \boldsymbol{x}, \varepsilon; \theta) \geq 0]$, where $\mathbb{1}(\cdot)$ is the indicator function. The network $G$ is pairwise stable if and only if $G = V(G)$.*

The fixed-point characterization provides a useful algorithm for checking whether $G$ is pairwise stable. General conditions for the existence of a pairwise stable network are given in Jackson and Watts (2001) and Hellmann (2013). Sheng

(2018) shows by example that if Assumption 2.2 does not hold, then there are cases where no pairwise stable network exists. A sufficient condition for the existence of a pairwise stable network with non-transferable utility is that Assumption 2.2 holds.

**Existence of an Equilibrium.** *Let Assumption 2.2 hold. For all $\boldsymbol{x} \in \mathcal{X}$, $\varepsilon \in \mathcal{E}$, there exists at least one pairwise stable network.*

Under monotonicity, the set of pairwise stable equilibria belong to a lattice.

**Set of Equilibria.** *Let Assumption 2.2 hold. Given $\boldsymbol{x}, \varepsilon$, there exists networks $\underline{G}(\boldsymbol{x}, \varepsilon, \theta)$ and $\overline{G}(\boldsymbol{x}, \varepsilon, \theta)$ such that:*

1. *$\underline{G}(\boldsymbol{x}, \varepsilon, \theta)$ and $\overline{G}(\boldsymbol{x}, \varepsilon, \theta)$ are pairwise stable; and*
2. *If $G$ is pairwise stable, then $\underline{G}(\boldsymbol{x}, \varepsilon, \theta) \leq G \leq \overline{G}(\boldsymbol{x}, \varepsilon, \theta)$.*

I define any network that belongs to the equilibria lattice to be *admissible.*

**Definition 2.5** (Admissible Set of Networks, Pairwise Stable Set of Network)**.** *Let $\mathcal{G}_\theta(\boldsymbol{x}, \varepsilon)$ denote the* Admissible Set of Networks*:*

$$\mathcal{G}_\theta(\boldsymbol{x}, \varepsilon) \equiv \{G \in \mathcal{G} : \underline{G}(\boldsymbol{x}, \varepsilon, \theta) \leq G \leq \overline{G}(\boldsymbol{x}, \varepsilon, \theta)\},$$

*Similarly, let $\mathcal{G}_\theta^{ps}(\boldsymbol{x}, \varepsilon)$ denote the* Pairwise Set of Stable Networks*:*

$$\mathcal{G}_\theta^{ps}(\boldsymbol{x}, \varepsilon) \equiv \{G \in \mathcal{G} : G \textit{ is pairwise stable}\}.$$

As the following proposition shows, the lattice $\mathcal{G}_\theta(\boldsymbol{x}, \varepsilon)$ is very quick to compute while computation of $\mathcal{G}_\theta^{\text{ps}}(\boldsymbol{x}, \varepsilon)$ is infeasible. Typically not all networks in $\mathcal{G}_\theta(\boldsymbol{x}, \varepsilon)$

are pairwise stable. Hence, to compute $\mathcal{G}_\theta^{\text{ps}}(\boldsymbol{x},\varepsilon)$ I need to evaluate $V(G)$ for all networks $G$ in the admissible lattice. If $\underline{G}(\boldsymbol{x},\varepsilon,\theta)$ and $\overline{G}(\boldsymbol{x},\varepsilon,\theta)$ differ by $k$ elements, then this requires $2^k$ evaluations of $V(G)$. This is infeasible even for relatively small values of $k$.

**Proposition 2.4.** *The admissible set of networks $\mathcal{G}_\theta(\boldsymbol{x},\varepsilon)$ can be computed in no more than $\frac{n(n-1)}{2} - n + 1$ evaluations of $V(\cdot)$.*

The proof of Proposition 2.4 also provides a constructive way of computing the lattice. In particular, the lattice can be computed by iteratively applying $V(\cdot)$ to the network of zeros and ones.

Propositions 2.1 and 2.4 as well as the results in this section play an important role in identification of the bounds on the statistics, discussed further in Section 2.5.2. To understand how, consider the case where we have a candidate value for $\theta$ and suppose that all relevant characteristics of network-formation are observed by the researcher. I have shown that all networks belong to an admissible lattice $\mathcal{G}_\theta(\boldsymbol{x})$, which, in this case, depend only on observed characteristics $\boldsymbol{x}$. The lattice $\mathcal{G}_\theta(\boldsymbol{x})$ can be feasibly computed and, provided the model is correctly specified, must contain the observed network $G^{\bar{\boldsymbol{n}}}$. The aforementioned results can then be applied to obtain bounds on a network statistic such as KBC, which is what I set out to achieve. There are two issues with this thought experiment. First, there are typically unobserved characteristics $\varepsilon$ that affect the admissible lattice $\mathcal{G}_\theta(\boldsymbol{x},\varepsilon)$. This is resolved by taking expectations and integrating out the unobserved taste shocks. Second, I do not have a candidate value for $\theta$. I use the observed network data and an identification strategy discussed in the next section to provide a set of candidate values for the network-formation parameter, which can then be used to bound the network statistic of interest.

## 2.5 Identification and Estimation

I first discuss identification of the network formation model. Identification is challenging due to multiplicity of pairwise stable equilibria – typically the admissible lattice contains more than one network $|\mathcal{G}_\theta(\boldsymbol{x}, \varepsilon)| > 1$. If there were a single pairwise stable network, then the network formation model would yield a single model implied distribution for the network $\boldsymbol{G}$, and one would be able to learn $\theta$ by matching the observed distribution $\boldsymbol{P}(\boldsymbol{G} = G_0|\boldsymbol{x})$ with the one implied by the model for all $G_0 \in \mathcal{G}$.[15] Due to multiplicity of equilibria, however, the model implies multiple distributions for the network. This identification problem is further aggravated by the assumption that the networks are only partially observed. That is, the data only reveals information about the distribution over partially observed networks $\boldsymbol{P}(\boldsymbol{G}^{\bar{\boldsymbol{n}}} = G_0^{\bar{\boldsymbol{n}}}|\boldsymbol{x})$. These data limitations and multiplicity of equilibria result in a set identification.[16]

I show two theoretical results in Section 2.5.1. First, I show how to obtain the sharp identified region $\mathcal{H}_P[\theta]$. This set includes all network-formation parameters $\theta$ such that the data-implied distribution $\boldsymbol{P}(\boldsymbol{G}^{\bar{\boldsymbol{n}}} = G_0^{\bar{\boldsymbol{n}}}|\boldsymbol{x})$ is consistent with one of the multiple distributions implied by the model. The sharp identified region, however, cannot be feasibly computed. I propose an outer region $\mathcal{O}_P[\theta]$ that contains $\mathcal{H}_P[\theta]$,

[15] In terms of notation, $\boldsymbol{P}(\cdot)$ is the probability measure and capital bold face letters indicate random variables, such as $\boldsymbol{G}$ and $\boldsymbol{X}$. Realized values for the network are capital non-boldface letters such as $G$. Realized values for covariates are given by lower-case bold face $\boldsymbol{x}$.

[16] There is an additional concern that the frequency estimator for $\boldsymbol{P}(G^{\bar{\boldsymbol{n}}} = G_0^{\bar{\boldsymbol{n}}}|\boldsymbol{x})$ is not precise unless $T$, the number of observed networks, is very large. The reason for this is that the space of networks $\mathcal{G}$ is very large, and so the likelihood of observing the exact same network-configuration twice is very small. My identification strategy relies on subcomponents of the network, which are precisely estimated even for relatively small values of $T$.

which can be computed by only considering subnetworks of $G_0^{\bar{\boldsymbol{n}}}$. By restricting the size of the subnetwork, the outer region can be feasibly computed and remains informative about the underlying parameters of the network-formation model. In Section 2.5.2, I show how to augment $\mathcal{O}_P[\theta]$ with bounds on the network statistic implied by Lemma 2.1 to obtain a joint outer region $\mathcal{O}_P[\theta, \beta_\theta]$ for both the network-formation parameter and the network statistic of interest, where $\beta_\theta = E_\theta(d(G))$ (the expectation is taken with respect to one of the multiple distributions for $G$).

### 2.5.1 Identification: The Network-Formation Model

Identification in networks with complete information and externalities is both data intensive and computationally difficult. The sharp identified region, $\mathcal{H}_P[\theta]$, is defined by moment inequalities – functions that are less than or equal to zero for $\theta \in \mathcal{H}_P[\theta]$. The sharp identified region requires evaluating $O(2^{2^{\frac{n(n-1)}{2}}})$ moment inequalities, which is not computationally feasible. In light of this challenge, I use a subnetwork identification strategy similar to Sheng (2018). The idea is to derive moment inequalities defined over small subnetworks of the observed network. These moment inequalities define an outer region for the network-formation parameters $\mathcal{O}_P[\theta]$. This strategy resolves the computational challenge, since the number of moment inequalities implied by subnetworks is equal to $2\sum_{m=1}^{q} 2^{\frac{m(m-1)}{2}}$. The integer $q$ is chosen by the researcher to control the computational burden of the problem and is typically set to a number less than six. Formally, a subnetwork is defined as follows.

**Definition 2.6** (Subnetwork $A^{\boldsymbol{s}}$ and Complement Subnetwork $A^{-\boldsymbol{s}}$)**.** *Let $\boldsymbol{s} \subset \boldsymbol{n}$ be a subset of the population set of individuals. The* Subnetwork *denoted $A^{\boldsymbol{s}}$ represents all links between individuals in $\boldsymbol{s}$ with other individuals in $\boldsymbol{s}$. The* Complement

Subnetwork *denoted* $A^{-\boldsymbol{s}}$ *represents all other links in the network. That is,* $A^{-\boldsymbol{s}}$ *contains all undirected links between individuals in* $\boldsymbol{n} - \boldsymbol{s}$ *with individuals in* $\boldsymbol{n}$*. The Subnetwork and Complement Subnetwork fully characterize* $G = (A^{\boldsymbol{s}}, A^{-\boldsymbol{s}})$*.*

The dark shaded area in Figure A.6.2a displays a subnetwork $A^{\boldsymbol{s}}$. I choose $A^{\boldsymbol{s}}$ so that its links are contained in the observed part of the network $G^{\bar{\boldsymbol{n}}}$. I select any set $\boldsymbol{s} \subset \boldsymbol{n}$ in the case of complete survey data. The striped area in Figure A.6.2b is the complement subnetwork and this includes all links between individuals not in $\boldsymbol{s}$. See Example A.3 for a matrix representation of a subnetwork. I derive bounds on the joint distribution of a subnetwork $A^{\boldsymbol{s}}$ and the characteristics of the individuals within the subnetwork, $\boldsymbol{x}^{\boldsymbol{s}}$. For that, I require the theoretical distribution of $(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}})$.

**Proposition 2.5.** *The distribution of* $(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}})$ *is given by*

$$\begin{aligned} &\boldsymbol{P}(\boldsymbol{A}^{\boldsymbol{s}} = A^{\boldsymbol{s}}, \boldsymbol{X} = \boldsymbol{x}^{\boldsymbol{s}}; \theta) \\ &\qquad = \sum_{\boldsymbol{x}^{-\boldsymbol{s}}} \int \left[ \sum_{A^{-\boldsymbol{s}} \in \mathcal{G}^{-\boldsymbol{s}}} \psi(\boldsymbol{G} = (A^{\boldsymbol{s}}, A^{-\boldsymbol{s}}) | \boldsymbol{x}, \varepsilon; \theta) \right] dF_{\varepsilon}(\varepsilon) \boldsymbol{P}(\boldsymbol{X}^{-\boldsymbol{s}} = \boldsymbol{x}^{-\boldsymbol{s}}). \end{aligned}$$

In Proposition 2.5, I integrate out characteristics of individuals not in the subnetwork $\boldsymbol{x}^{-\boldsymbol{s}}$. Moving in, I take expectations with respect to unobservable preference shocks $\varepsilon$. Finally, inside the square brackets is the probability that $(A^{\boldsymbol{s}}, A^{-\boldsymbol{s}})$ forms conditional on $\boldsymbol{x}$ and $\varepsilon$. Because the model admits multiple equilibria, I require a rule that selects which network forms in equilibrium. The term $\psi(\boldsymbol{G} = (A^{\boldsymbol{s}}, A^{-\boldsymbol{s}}) | \boldsymbol{x}, \varepsilon; \theta)$ is the network selection mechanism that determines with what probability $(A^{\boldsymbol{s}}, A^{-\boldsymbol{s}})$ occurs in equilibrium. The selection mechanism is formally defined below. It is equal to one if $G$ is the unique equilibrium at $(\boldsymbol{x}, \varepsilon; \theta)$ and zero if $G$ is not in the equilibria set. Otherwise, the selection mechanism is any valid probability measure (Tamer, 2003).

**Definition 2.7** (Network Selection Mechanism)**.** *The* Network Selection Mechanism *is a measurable function* $\psi(\cdot|\boldsymbol{x}, \varepsilon; \theta)$ *satisfying the following conditions*

1. $\forall G \notin \mathcal{G}_\theta^{ps}(\boldsymbol{x}, \varepsilon),\ \ \psi(G|\boldsymbol{x}, \varepsilon; \theta) = 0$*;*

2. $\forall G \in \mathcal{G}_\theta^{ps}(\boldsymbol{x}, \varepsilon),\ \ \psi(G|\boldsymbol{x}, \varepsilon; \theta) \in [0, 1]$*; and*

3. $\sum_{G \in \mathcal{G}_\theta^{ps}(\boldsymbol{x}, \varepsilon)} \psi(G|\boldsymbol{x}, \varepsilon; \theta) = 1$*.*

**Remark 2.3.** The first key distinction between my approach and Sheng (2018) is that I derive bounds on the joint distribution of $(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}})$ and Sheng (2018) derives bounds for $(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}}, F_{\boldsymbol{x}^{-\boldsymbol{s}}}(\boldsymbol{x}^{-\boldsymbol{s}}))$. My method requires a moment inequality for each possible realization of $(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}})$. By limiting the size of the subnetwork I can keep the number of moment inequalities in check. In contrast, Sheng (2018) requires a moment inequality for all combinations of $(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}}, F_{\boldsymbol{x}^{-\boldsymbol{s}}}(\boldsymbol{x}^{-\boldsymbol{s}}))$, where $F_{\boldsymbol{x}^{-\boldsymbol{s}}}(\boldsymbol{x}^{-\boldsymbol{s}})$ is a value for the distribution at a support point $\boldsymbol{x}^{-\boldsymbol{s}}$. If the support is sufficiently rich (e.g., discrete but not binary), the number of moment inequalities implied by all combinations of $(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}}, F_{\boldsymbol{x}^{-\boldsymbol{s}}}(\boldsymbol{x}^{-\boldsymbol{s}}))$ is very large even when $|\boldsymbol{s}|$ is small. The problem persists even after using exchangeability and equivalence classes to limit the number of moment conditions, which is discussed below. There is, however, a cost from averaging over $\boldsymbol{x}^{-\boldsymbol{s}}$ in Proposition 2.5 in that we may lose information. In particular, the set of moment inequalities for $(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}})$ may be satisfied at a particular value of $\theta$, but may fail for those defined by $(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}}, F_{\boldsymbol{x}^{-\boldsymbol{s}}}(\boldsymbol{x}^{-\boldsymbol{s}}))$.

I partition the support for $\varepsilon$ into regions that characterize the equilibrium status of a particular subnetwork $A^{\boldsymbol{s}} \in \mathcal{G}^{\boldsymbol{s}}$. I will use this partition to decompose the probability mass function for each subnetwork.

**Definition 2.8** (Region of Uniqueness, Region of Multiplicity, Region of Admissibility)**.** *Denote* $\mathcal{E}_u(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta}) \subset \mathcal{E}$ *to be the* Region of Uniqueness*. If* $\varepsilon \in$

$\mathcal{E}_u(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})$*, then there exists a complement subnetwork* $A^{-\boldsymbol{s}}$ *with the property that* $(A^{\boldsymbol{s}}, A^{-\boldsymbol{s}})$ *is a pairwise stable equilibrium. Moreover, any pairwise-stable equilibrium* $G$ *has* $A^{\boldsymbol{s}}$ *as its subnetwork:* $G = (A^{\boldsymbol{s}}, A^{-\boldsymbol{s}})$ *for some* $A^{-\boldsymbol{s}}$*. Denote* $\mathcal{E}_m(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta}) \subset \mathcal{E}$ *to be the* Region of Multiplicity*. If* $\varepsilon \in \mathcal{E}_m(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})$*, then there exists a complement subnetwork* $A^{-\boldsymbol{s}}$ *with the property that* $(A^{\boldsymbol{s}}, A^{-\boldsymbol{s}})$ *is a pairwise stable equilibrium. Moreover, there exists a pairwise-stable equilibrium* $G$ *that does not have* $A^{\boldsymbol{s}}$ *as its subnetwork:* $G = (\tilde{A}^{\boldsymbol{s}}, A^{-\boldsymbol{s}})$ *for some* $\tilde{A}^{\boldsymbol{s}} \neq A^{\boldsymbol{s}}$ *and* $A^{-\boldsymbol{s}}$*. Denote* $\mathcal{E}_a(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta}) \subset \mathcal{E}$ *to be the* Region of Admissibility*. If* $\varepsilon \in \mathcal{E}_a(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})$*, then there exists* $A^{-\boldsymbol{s}}$ *such that* $(A^{\boldsymbol{s}}, A^{-\boldsymbol{s}}) \in \mathcal{G}_\theta(\boldsymbol{x}, \varepsilon)$*, i.e.,* $(A^{\boldsymbol{s}}, A^{-\boldsymbol{s}})$ *belongs to the admissible lattice.*

Using these definitions, I decompose the probability mass function for the subnetwork into two terms.

$$
\begin{aligned}
&\boldsymbol{P}(\boldsymbol{A}^{\boldsymbol{s}} = A^{\boldsymbol{s}}, \boldsymbol{X} = \boldsymbol{x}^{\boldsymbol{s}}; \theta) \qquad (2.3)\\
&= \sum_{\boldsymbol{x}^{-\boldsymbol{s}}} \Bigg[ \int \mathbb{1}[\varepsilon \in \mathcal{E}_u(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})] dF_\varepsilon(\varepsilon) \\
&\quad + \int \mathbb{1}[\varepsilon \in \mathcal{E}_m(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})] \sum_{A^{-\boldsymbol{s}} \in \mathcal{G}^{-\boldsymbol{s}}} \psi(\boldsymbol{G} = (A^{\boldsymbol{s}}, A^{-\boldsymbol{s}}) | \boldsymbol{x}, \varepsilon; \theta) dF_\varepsilon(\varepsilon) \Bigg] \boldsymbol{P}(\boldsymbol{X} = \boldsymbol{x}),
\end{aligned}
$$

where $\mathbb{1}[\cdot]$ is the indicator function. The first term involving $\mathbb{1}[\varepsilon \in \mathcal{E}_u(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})]$ is the integral over the region where the subnetwork is unique. The summation $\sum_{A^{-\boldsymbol{s}} \in \mathcal{G}^{-\boldsymbol{s}}} \psi(\boldsymbol{G} = (A^{\boldsymbol{s}}, A^{-\boldsymbol{s}}) | \boldsymbol{x}, \varepsilon; \theta)$ drops out of the first term as it is equal to one – the selection mechanism picks a network $(A^{\boldsymbol{s}}, A^{-\boldsymbol{s}})$ with $A^{\boldsymbol{s}}$ as its subnetwork almost surely. The second term involving $\mathbb{1}[\varepsilon \in \mathcal{E}_m(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})]$ is the integral over the region where the subnetwork is not unique.[17] The selection mechanism is a

[17] This is a similar decomposition to the one applied to the Nash entry game (Ciliberto & Tamer, 2009)

measurable function, thus the following bounds are satisfied

$$0 \leq \sum_{A^{-\boldsymbol{s}} \in \mathcal{G}^{-\boldsymbol{s}}} \psi(\boldsymbol{G} = (A^{\boldsymbol{s}}, A^{-\boldsymbol{s}}) | \boldsymbol{x}, \varepsilon; \theta) \leq \sum_{G \in \mathcal{G}} \psi(\boldsymbol{G} = G | \boldsymbol{x}, \varepsilon; \theta) = 1.$$

Therefore, I can set $\sum_{A^{-\boldsymbol{s}} \in \mathcal{G}^{-\boldsymbol{s}}} \psi(\boldsymbol{G} = (A^{\boldsymbol{s}}, A^{-\boldsymbol{s}}) | \boldsymbol{x}, \varepsilon; \theta) \in \{0, 1\}$ in Equation (2.3) to obtain bounds on the probability mass function:

$$\begin{aligned}
&\sum_{\boldsymbol{x}^{-\boldsymbol{s}}} \int \mathbb{1}[\varepsilon \in \mathcal{E}_u(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})] dF_\varepsilon(\varepsilon) \boldsymbol{P}(\boldsymbol{X}^{-\boldsymbol{s}} = \boldsymbol{x}^{-\boldsymbol{s}}) \\
&\qquad \leq \boldsymbol{P}(\boldsymbol{A}^{\boldsymbol{s}} = A^{\boldsymbol{s}}, \boldsymbol{X} = \boldsymbol{x}^{\boldsymbol{s}}; \theta) \\
&\qquad \leq \sum_{\boldsymbol{x}^{-\boldsymbol{s}}} \int \mathbb{1}[\varepsilon \in \mathcal{E}_u(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta}) \cup \mathcal{E}_m(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})] dF_\varepsilon(\varepsilon) \boldsymbol{P}(\boldsymbol{X}^{-\boldsymbol{s}} = \boldsymbol{x}^{-\boldsymbol{s}}).
\end{aligned}$$

Feasibly computing these bounds is an important consideration as I need to augment this with moment inequalities implied by the network statistic and construct projected confidence intervals. Checking whether $\varepsilon \in \mathcal{E}_u(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta}) \cup \mathcal{E}_m(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})$ is computationally costly and must be repeated for each simulated draw of $\varepsilon$. Sheng (2018) shows that checking $\varepsilon \in \mathcal{E}_u(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta}) \cup \mathcal{E}_m(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})$ for the Transferable Utility case requires solving an optimization routine. Based on times reported in Sheng (2018), it is approximately 100 to 1000 times faster to compute the admissible lattice and work with the following bounds:

$$\begin{aligned}
\boldsymbol{P}(\boldsymbol{A}^{\boldsymbol{s}} = A^{\boldsymbol{s}}, \boldsymbol{X} = \boldsymbol{x}^{\boldsymbol{s}}; \theta) &\geq \sum_{\boldsymbol{x}^{-\boldsymbol{s}}} \int \mathbb{1}[\varepsilon \in \mathcal{E}_u(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})] dF_\varepsilon(\varepsilon) \boldsymbol{P}(\boldsymbol{X}^{-\boldsymbol{s}} = \boldsymbol{x}^{-\boldsymbol{s}}) \\
\boldsymbol{P}(\boldsymbol{A}^{\boldsymbol{s}} = A^{\boldsymbol{s}}, \boldsymbol{X} = \boldsymbol{x}^{\boldsymbol{s}}; \theta) &\leq \sum_{\boldsymbol{x}^{-\boldsymbol{s}}} \int \mathbb{1}[\varepsilon \in \mathcal{E}_a(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})] dF_\varepsilon(\varepsilon) \boldsymbol{P}(\boldsymbol{X}^{-\boldsymbol{s}} = \boldsymbol{x}^{-\boldsymbol{s}}),
\end{aligned}$$

which are valid because $\mathcal{E}_u(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta}) \cup \mathcal{E}_m(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta}) \subset \mathcal{E}_a(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})$. To that end, I define the following moment inequality functions:

$$\tilde{m}_1(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}}; \theta)$$
$$\equiv -\boldsymbol{P}(\boldsymbol{A}^{\boldsymbol{s}} = A^{\boldsymbol{s}}, \boldsymbol{X} = \boldsymbol{x}^{\boldsymbol{s}}) + \sum_{\boldsymbol{x}^{-\boldsymbol{s}}} \int \mathbb{1}[\varepsilon \in \mathcal{E}_u(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})] dF_{\varepsilon}(\varepsilon) \boldsymbol{P}(\boldsymbol{X}^{-\boldsymbol{s}} = \boldsymbol{x}^{-\boldsymbol{s}})$$
$$\tilde{m}_2(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}}; \theta)$$
$$\equiv \boldsymbol{P}(\boldsymbol{A}^{\boldsymbol{s}} = A^{\boldsymbol{s}}, \boldsymbol{X} = \boldsymbol{x}^{\boldsymbol{s}}) - \sum_{\boldsymbol{x}^{-\boldsymbol{s}}} \int \mathbb{1}[\varepsilon \in \mathcal{E}_a(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})] dF_{\varepsilon}(\varepsilon) \boldsymbol{P}(\boldsymbol{X}^{-\boldsymbol{s}} = \boldsymbol{x}^{-\boldsymbol{s}}).$$

These moment functions satisfy $\tilde{m}_j(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}}; \theta) \leq 0$ for all $\theta \in \mathcal{H}_P[\theta]$, where $\mathcal{H}_P[\theta]$ is the sharp identified region for $\theta$. Hence, these moment inequality functions define a valid outer region. However, even for small subnetworks the number of moment inequalities can be quite large. One way to reduce the number of moment inequalities is to aggregate subnetworks into isomorphic equivalence classes. Two subnetworks $(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}})$ and $(\tilde{A}^{\boldsymbol{s}}, \tilde{\boldsymbol{x}}^{\boldsymbol{s}})$ are in the same equivalence class if there is a permutation $\tau(\cdot)$ such that $(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}}) = (\tilde{A}^{\tau(\boldsymbol{s})}, \tilde{\boldsymbol{x}}^{\tau(\boldsymbol{s})})$. Two moment inequalities are required for each equivalence class, which dramatically reduces the number of moment inequalities defining the outer region. I further reduce the computational complexity by summing over equivalence classes of $\boldsymbol{x}^{-\boldsymbol{s}}$. If the covariate is binary and univariate, it is sufficient to count the number of cases such that $x_i = 1$ for $i \in \boldsymbol{n} - \boldsymbol{s}$. Denote the equivalence class of $(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}})$ to be $\mathcal{C}(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}})$. The moment inequalities that I work with are the following:

$$\begin{aligned} m_1(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}}; \theta) \equiv & -\boldsymbol{P}((\boldsymbol{A}^{\boldsymbol{s}}, \boldsymbol{X}^{\boldsymbol{s}}) \in \mathcal{C}(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}})) \\ & + \sum_{\mathcal{C}(\boldsymbol{x}^{-\boldsymbol{s}})} \int \mathbb{1}[\varepsilon \in \mathcal{E}_u(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})] dF_{\varepsilon}(\varepsilon) \boldsymbol{P}(\boldsymbol{X}^{-\boldsymbol{s}} \in \mathcal{C}(\boldsymbol{x}^{-\boldsymbol{s}})) \end{aligned} \tag{2.4}$$

and

$$\begin{aligned} m_2(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}}; \theta) \equiv & \boldsymbol{P}((\boldsymbol{A}^{\boldsymbol{s}}, \boldsymbol{X}^{\boldsymbol{s}}) \in \mathcal{C}(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}})) \\ & - \sum_{\mathcal{C}(\boldsymbol{x}^{-\boldsymbol{s}})} \int \mathbb{1}[\varepsilon \in \mathcal{E}_a(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})] dF_{\varepsilon}(\varepsilon) \boldsymbol{P}(\boldsymbol{X}^{-\boldsymbol{s}} \in \mathcal{C}(\boldsymbol{x}^{-\boldsymbol{s}})). \end{aligned}$$

This aggregation can, in some cases, result in a loss of information. That is, the moment inequalities $\tilde{m}_j(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}}; \theta) \leq 0$ can be violated for some $\theta$, but the corresponding moment inequalities $m_j(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}}; \theta) \leq 0$ might hold. Provided the primitives of the utility function and the selection mechanism satisfy interchangeability (de Finetti, 1929; Chernoff & Teicher, 1958; Kallenberg, 2006; Austin, 2008), the moment inequalities can be aggregated up to equivalence classes without loss of information (Sheng, 2018). These interchangeability conditions, along with a full support assumption on the preference shocks $\varepsilon$, imply that these subnetwork frequency estimators are informative (i.e., bounded away from 0 and 1) as the network grows in size (Sheng, 2018, Proposition 4.2). However, interchangeability is restrictive and implies that the network is dense (Orbanz & Roy, 2014).

I specify below econometric and structural assumptions that I maintain for identification. Under these assumptions as well as Assumptions 2.1 and 2.2, I display two theorems that describe the sharp identified region and the outer region based on the above moment inequalities over all subnetworks up to a particular size chosen by the researcher. The sharp identified region does not require Assumption 2.2.

**Assumption 2.4** (Econometric Assumptions)**.** *There is a sequence of random elements $(\boldsymbol{x}_t, \varepsilon_t)$ such that: (1) $\forall t$, $\boldsymbol{x}_t$ and $\varepsilon_t$ are independent, and (2) $\forall i \neq j$ and $\forall t$, $\varepsilon_{ijt}$ are i.i.d., supported on $\mathcal{E} \subset \mathbb{R}$, and generated from a continuously differentiable parametric distribution $F(\varepsilon; \theta_\varepsilon)$ that depends on the finite-dimensional parameter $\theta_\varepsilon \in \mathbb{R}^{dim(\theta_\varepsilon)}$; (3) $\boldsymbol{x}$ is discrete.*

**Assumption 2.5** (Structural Assumptions)**.** *Each individual $i \in \boldsymbol{n}_t$ receives utility from a network according to a their payoff function $\pi_i(G, \boldsymbol{x}, \varepsilon; \theta)$. Individuals simultaneously choose friendships with complete information over $(\boldsymbol{x}_t, \varepsilon_t)$ with the* *restriction that the resulting network $G_t$ is pairwise stable.*

**Theorem 2.1** (Sharp Identified Region for Network Formation Parameter)**.** *Let Assumptions 2.1, 2.4, and 2.5 hold. Define*

$$\mathcal{G}_\theta^{ps}(\boldsymbol{x}, \varepsilon, \bar{\boldsymbol{n}}) = \{G^{\bar{\boldsymbol{n}}} \in \mathcal{G} : \exists G^{-\bar{\boldsymbol{n}}} \ s.t. \ (G^{\bar{\boldsymbol{n}}}, G^{-\bar{\boldsymbol{n}}}) \in \mathcal{G}_\theta^{ps}(\boldsymbol{x}, \varepsilon))\}$$

*That is, $\mathcal{G}_\theta^{ps}(\boldsymbol{x}, \varepsilon, \bar{\boldsymbol{n}})$ is a random set that consists of all partially observed networks $G^{\bar{\boldsymbol{n}}}$ that can be completed by a pairwise stable network. The sharp identification region for $\theta$ is given by*

$$\mathcal{H}_P[\theta] = \{\theta \in \Theta : \boldsymbol{P}(\boldsymbol{G}^{\bar{\boldsymbol{n}}} \in \mathcal{K}|\boldsymbol{x}) \leq \boldsymbol{T}_{\mathcal{G}_\theta^{ps}(\boldsymbol{x},\varepsilon,\bar{\boldsymbol{n}})}(\mathcal{K}; F_\varepsilon) \ \forall \mathcal{K} \subset \boldsymbol{\mathcal{G}}^{\bar{\boldsymbol{n}}}, \boldsymbol{x} - a.s\},$$

*where $\boldsymbol{T}_{\mathcal{G}_\theta^{ps}(\boldsymbol{x},\varepsilon,\bar{\boldsymbol{n}})}(\mathcal{K}; F_\varepsilon) \equiv \boldsymbol{P}(\{\mathcal{G}_\theta^{ps}(\boldsymbol{x}, \varepsilon, \bar{\boldsymbol{n}}) \cap \mathcal{K} \neq \emptyset\}, \varepsilon \sim F_\varepsilon)$ is the capacity functional.*

Theorem 2.1 characterizes the sharp identification region in a similar way to Molinari (2019, Theorem SIR-3.8). In contrast to what I do, Molinari assumes the researcher observes the complete network. The key difference is the information revealed to the researcher and the random set governing Artstein's inequality (Artstein, 1983). While I do maintain Assumption 2.2 throughout this paper, Theorem 2.1 does not technically require it. Suppose that Assumption 2.2 did not hold so that marginal utility is not monotonically increasing in the network. It is possible that there does not exist an equilibrium for some value of $\theta$ in which case $\mathcal{G}_\theta^{\mathrm{ps}}(\boldsymbol{x}, \varepsilon) = \emptyset$ and $\boldsymbol{T}_{\mathcal{G}_\theta^{\mathrm{ps}}(\boldsymbol{x},\varepsilon,\bar{\boldsymbol{n}})}(\mathcal{K}; F_\varepsilon) = 0$. The sharp identified region in this case would exclude these values of $\theta$. The sharp identified region cannot be feasibly computed as it requires enumerating a doubly exponential number of compact sets. It is, however, useful for establishing an outer region that is computationally feasible. The next Theorem displays an outer region $\mathcal{O}_P[\theta]$ based on subnetwork moment inequalities that contains the set $\mathcal{H}_P[\theta]$.

**Theorem 2.2** (Outer Region for Network Formation Parameter)**.** *Fix an integer $q \leq |\bar{\boldsymbol{n}}|$ and suppose that Assumptions 2.1, 2.2, 2.4, and 2.5 hold. Define*

$$\mathcal{O}_P[\theta] \equiv \{\theta \in \Theta : m_j(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}}; \theta) \ \leq 0 \quad j = 1, 2 \quad \forall A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}} \ s.t.\ |\boldsymbol{s}| \leq q,\ \boldsymbol{s} \subset \bar{\boldsymbol{n}}\}.$$

*where $m_1(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}}; \theta)$ and $m_2(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}}; \theta)$ are given in Equation* (2.4)*. Then $\mathcal{H}_P[\theta] \subseteq \mathcal{O}_P[\theta]$.*

I compute the outer region $\mathcal{O}_P[\theta]$ using simulated methods, which are discussed in detail in Appendix A.1. Next, I show how to use Propositions 2.1, 2.2, and 2.3 to identify a wide class of centrality measures, including Katz-Bonacich Centrality, diffusion centrality, and decay centrality. These results naturally extend to identifying endogenous outcomes of a game played on the network, such as the peer effects game.

### 2.5.2 Identification: Network Statistics

Consider a network statistic $d(G)$. Ideally, I would like to compute the value of $d(\cdot)$ for one of the observed networks $G^{\bar{\boldsymbol{n}}}$. However, computing $d(G)$ typically requires the full network.[18] Using the model, I can take expectations with respect to $d(G)$ conditional on $G^{\bar{\boldsymbol{n}}}$. Due to multiplicity of equilibria, however, the model implies multiple distributions for the network. Therefore, there is a collection of values of $\beta(G^{\bar{\boldsymbol{n}}}) \equiv \boldsymbol{E}_\theta(d(\boldsymbol{G})|G^{\bar{\boldsymbol{n}}}; \theta)$ that are consistent with the model. Using Propositions 2.1, 2.2, and 2.3, I can bound these values to provide informative bounds on the network statistic of interest.

[18] In the case of degree centrality, only the direct friends of an individual is required. With complete survey data, the degree centrality is computable for surveyed individuals. However, this information is not available with arbitrary random missing link data.

A related and easier problem that I address first is obtaining bounds on the unconditional expectation of $d(G)$. These bounds can be applied to out-of-sample networks – networks that I have no prior knowledge over. An out-of-sample network is, for example, a friendship network among students in a school that I have not sampled. As another example, I can use the out-of-sample network bounds to execute counterfactual analysis where I hypothetically remove students from the classroom and allow the remaining students to re-optimize their friendship network according to the model. For exposition, suppose that $d(G)$ is monotonically increasing in the network. I will relax this assumption in the main theorem. Conditional on $\boldsymbol{x}$, the expected value of $d(G)$ is given by

$$\beta \equiv \boldsymbol{E}_\theta(d(\boldsymbol{G})|\boldsymbol{x};\theta) = \int_\varepsilon \sum_{G \in \mathcal{G}} d(G)\psi(\boldsymbol{G} = G|\boldsymbol{x}, \varepsilon; \theta) dF_\varepsilon(\varepsilon). \tag{2.5}$$

The joint sharp identified region for $(\beta_\theta, \theta)$ includes all values given by Equation (2.5) for some $\theta \in \mathcal{H}_P[\theta]$ and for some valid network selection mechanism $\psi(\cdot|\boldsymbol{x}, \varepsilon; \theta)$.[19] Formally,

$$\mathcal{H}_P[\beta, \theta] \equiv \left\{ \begin{array}{ll} (\beta, \theta) \in B \times \mathcal{H}_P[\theta] : & \exists \text{ network selection mechanism } \psi(\cdot) \\ & \text{s.t. } \beta \text{ satisfies Equation (2.5)} \end{array} \right\}. \tag{2.6}$$

Applying Lemma 2.1, monotonically increasing network statistics are maximized by $\overline{G}(\boldsymbol{x}, \varepsilon, \theta)$ on $\mathcal{G}_\theta(\boldsymbol{x}, \varepsilon)$ from which it follows that

$$\beta \leq \int_\varepsilon d(\overline{G}(\boldsymbol{x}, \varepsilon, \boldsymbol{\theta})) dF_\varepsilon(\varepsilon).$$

[19] This sharp identified region can also be derived from the Aumann Expectation of the random set containing admissible values of $d(G)$. The Aumann Expectation is a correspondence from a random set to a set of values that are consistent with one of the multiple underlying distributions (Aumann, 1965; Molchanov, 2005; Molchanov & Molinari, 2018). In this case, it is easier (and isomorphic) to work with the selection mechanism.

Similarly, the lower bound is achieved by loading all of the selection mechanism's mass on $\underline{G}(\boldsymbol{x}, \varepsilon, \boldsymbol{\theta})$ to obtain

$$\beta \geq \int_{\varepsilon} d(\underline{G}(\boldsymbol{x}, \varepsilon, \boldsymbol{\theta}))dF_{\varepsilon}(\varepsilon).$$

Define the following moment inequalities:

$$\begin{aligned} m_3(\boldsymbol{x}; \beta, \theta) &\equiv -\beta + \int_{\varepsilon} d(\underline{G}(\boldsymbol{x}, \varepsilon, \boldsymbol{\theta}))dF_{\varepsilon}(\varepsilon) \\ m_4(\boldsymbol{x}; \beta, \theta) &\equiv \beta - \int_{\varepsilon} d(\overline{G}(\boldsymbol{x}, \varepsilon, \boldsymbol{\theta}))dF_{\varepsilon}(\varepsilon). \end{aligned} \tag{2.7}$$

For non-monotone statistics, the moment inequalities are replaced by

$$\int_{\varepsilon} \underline{d}(\underline{G}(\boldsymbol{x}, \varepsilon, \boldsymbol{\theta}), \overline{G}(\boldsymbol{x}, \varepsilon, \boldsymbol{\theta}))dF_{\varepsilon}(\varepsilon) \quad \text{and} \quad \int_{\varepsilon} \overline{d}(\underline{G}(\boldsymbol{x}, \varepsilon, \boldsymbol{\theta}), \overline{G}(\boldsymbol{x}, \varepsilon, \boldsymbol{\theta}))dF_{\varepsilon}(\varepsilon),$$

where $\underline{d}(\cdot, \cdot)$ and $\overline{d}(\cdot, \cdot)$ are bounds on Problem (2.1). The next theorem displays the outer region for $(\beta, \theta)$ for an out-of-sample network $G$ that I have no information on.

**Theorem 2.3** (Outer Region for a Network Statistic)**.** *Fix an integer $q \leq |\bar{\boldsymbol{n}}|$ and suppose that Assumptions 2.1, 2.2, 2.4, and 2.5. Let $d(G)$ be a network statistic. Assume that there are mappings $\underline{d} : \mathcal{G} \times \mathcal{G} \to \mathbb{R}^n$ and $\overline{d} : \mathcal{G} \times \mathcal{G} \to \mathbb{R}^n$ such that*

$$\underline{d}(\underline{G}, \overline{G}) \leq \min_{G \in \mathcal{L}(\underline{G}, \overline{G})} d(G) \quad \textit{and} \quad \max_{G \in \mathcal{L}(\underline{G}, \overline{G})} d(G) \leq \overline{d}(\underline{G}, \overline{G})$$

*for all network lattices $\mathcal{L}(\underline{G}, \overline{G}) = \{G : \underline{G} \leq G \leq \overline{G}\}$. Define*

$$\begin{aligned} &\mathcal{O}_P[\beta, \theta] \\ &\equiv \left\{ \begin{array}{ll} \beta \in B, \theta \in \Theta : & m_j(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}}; \theta) \leq 0 \quad j = 1, 2 \quad \forall A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}} \ s.t.\ |\boldsymbol{s}| \leq q,\ \boldsymbol{s} \subset \bar{\boldsymbol{n}} \\ & m_j(\boldsymbol{x}; \beta, \theta) \leq 0, \quad j = 3, 4 \quad \boldsymbol{x} - a.s. \end{array} \right\}, \end{aligned}$$

*where*

$$\begin{aligned} m_3(\boldsymbol{x}; \beta, \theta) &\equiv -\beta + \int_{\varepsilon} \underline{d}(\underline{G}(\boldsymbol{x}, \varepsilon, \boldsymbol{\theta}), \overline{G}(\boldsymbol{x}, \varepsilon, \boldsymbol{\theta}))dF_{\varepsilon}(\varepsilon) \\ m_4(\boldsymbol{x}; \beta, \theta) &\equiv \beta - \int_{\varepsilon} \overline{d}(\underline{G}(\boldsymbol{x}, \varepsilon, \boldsymbol{\theta}), \overline{G}(\boldsymbol{x}, \varepsilon, \boldsymbol{\theta}))dF_{\varepsilon}(\varepsilon). \end{aligned}$$

*Then $\mathcal{H}_P[\beta, \theta] \subseteq \mathcal{O}_P[\beta, \theta]$.*

**Remark 2.4.** The above theorem is useful for analyzing policy interventions where we do not observe the network. Consider for example a policy where resources are injected into all schools in a country, but the researcher lacks network data on all of the schools. The researcher has, however, estimated the network-formation model using network data from a sample of students from a sample of schools. Extrapolating the model to non-sampled schools, I can obtain bounds on a centrality measures based on observed characteristics of the students. The centrality measure can then be used to infer how the policy affects outcomes on the unobserved network.

I now turn to obtaining bounds on a network statistic for an in-sample network that is partially observed. Conditional on the realized network $G^{\bar{\boldsymbol{n}}}$, I require that $(G^{\bar{\boldsymbol{n}}}, G^{-\bar{\boldsymbol{n}}})$ is pairwise stable. This is true if only if the subgame played on links in $G^{-\bar{\boldsymbol{n}}}$ is pairwise stable conditional on $G^{\bar{\boldsymbol{n}}}$. That is, for all links with $G^{-\bar{\boldsymbol{n}}}_{ij} = 1$, it must be the case that the marginal utility over $(G^{\bar{\boldsymbol{n}}}, G^{-\bar{\boldsymbol{n}}})$ is positive for both $i$ and $j$. If $G^{-\bar{\boldsymbol{n}}}_{ij} = 0$, then either $i$ has negative marginal utility over $j$ or $j$ has negative marginal utility over $i$. The sharp identified region $\mathcal{H}_P[\beta(G^{\bar{\boldsymbol{n}}}), \theta]$ is analogous to the one displayed in Equation (2.6) with the restriction that $\beta(G^{\bar{\boldsymbol{n}}})$ satisfies

$$\beta(G^{\bar{\boldsymbol{n}}}) = \int_{\varepsilon} \sum_{G^{-\bar{\boldsymbol{n}}}} d(G)\psi(\boldsymbol{G} = (G^{\bar{\boldsymbol{n}}}, G^{-\bar{\boldsymbol{n}}})|G^{\bar{\boldsymbol{n}}}, \boldsymbol{x}, \varepsilon; \theta) dF_{\varepsilon}(\varepsilon).$$

I apply Topkis (1978) to this subgame to obtain the conditional admissible lattice denoted

$$\mathcal{G}_{\theta}(\boldsymbol{x}, G^{\bar{\boldsymbol{n}}}, \varepsilon) = \{G : \underline{H}(\boldsymbol{x}, G^{\bar{\boldsymbol{n}}}, \varepsilon; \theta) \leq G \leq \overline{H}(\boldsymbol{x}, G^{\bar{\boldsymbol{n}}}, \varepsilon; \theta)\},$$

where $\underline{H}(\boldsymbol{x}, G^{\bar{\boldsymbol{n}}}, \varepsilon; \theta)$ and $\overline{H}(\boldsymbol{x}, G^{\bar{\boldsymbol{n}}}, \varepsilon; \theta)$ are obtained by applying the modified mapping

$$\tilde{V}_{ij}(G^{\bar{\boldsymbol{n}}}, G^{-\bar{\boldsymbol{n}}}) = \begin{cases} G^{\bar{\boldsymbol{n}}}_{ij} & \text{if } ij \in G^{\bar{\boldsymbol{n}}} \\ V_{ij}(G^{\bar{\boldsymbol{n}}}, G^{-\bar{\boldsymbol{n}}}) & \text{if } ij \in G^{-\bar{\boldsymbol{n}}} \end{cases}$$

to $(G^{-\bar{\boldsymbol{n}}}, \mathbf{0}^{-\bar{\boldsymbol{n}}})$ and $(G^{-\bar{\boldsymbol{n}}}, \mathbf{1}^{-\bar{\boldsymbol{n}}})$, where $\mathbf{0}^{-\bar{\boldsymbol{n}}}$ and $\mathbf{1}^{-\bar{\boldsymbol{n}}}$ are the zero and one submatrices with dimension equal to $G^{-\bar{\boldsymbol{n}}}$. From this, I obtain the following bounds on a monotonically increasing network statistic $d(G)$:

$$\begin{aligned} \beta(G^{\bar{\boldsymbol{n}}}) &\leq \int_{\varepsilon} d(\overline{H}(\boldsymbol{x}, G^{\bar{\boldsymbol{n}}}, \varepsilon; \theta)) dF_{\varepsilon}(\varepsilon) \\ \beta(G^{\bar{\boldsymbol{n}}}) &\geq \int_{\varepsilon} d(\underline{H}(\boldsymbol{x}, G^{\bar{\boldsymbol{n}}}, \varepsilon; \theta)) dF_{\varepsilon}(\varepsilon). \end{aligned} \tag{2.8}$$

Consequently, I obtain an outer region similar to the one characterized in Theorem 2.3.

**Theorem 2.4** (Outer Region for a Network Statistic)**.** *Fix an integer* $q \leq |\bar{\boldsymbol{n}}|$ *and suppose that Assumptions 2.1, 2.2, 2.4, 2.5 hold. Let* $d(G)$ *be a network statistic. Suppose that a partially observed network* $G^{\bar{\boldsymbol{n}}}$ *is specified. Assume that there are mappings* $\underline{d} : \mathcal{G} \times \mathcal{G} \to \mathbb{R}^n$ *and* $\overline{d} : \mathcal{G} \times \mathcal{G} \to \mathbb{R}^n$ *such that*

$$\underline{d}(\underline{G}, \overline{G}) \leq \min_{G \in \mathcal{L}(\underline{G}, \overline{G})} d(G) \quad \textit{and} \quad \max_{G \in \mathcal{L}(\underline{G}, \overline{G})} d(G) \leq \overline{d}(\underline{G}, \overline{G})$$

*for all network lattices* $\mathcal{L}(\underline{G}, \overline{G}) = \{G : \underline{G} \leq G \leq \overline{G}\}$. *Define*

$$\mathcal{O}_P[\beta(G^{\bar{\boldsymbol{n}}}), \theta] \equiv \left\{ \begin{array}{ll} (\theta, \beta(G^{\bar{\boldsymbol{n}}})) \in \Theta \times B : & m_j(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}}; \theta) \leq 0 \quad j = 1, 2 \quad \forall A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}} \\ & s.t. \ |\boldsymbol{s}| \leq A, \ \boldsymbol{s} \subset \bar{\boldsymbol{n}} \\ & m_j(\boldsymbol{x}; \beta(G^{\bar{\boldsymbol{n}}}), \theta) \leq 0, \quad j = 3, 4 \quad \boldsymbol{x} - a.s. \end{array} \right\},$$

*where*

$$\begin{aligned} m_3(\boldsymbol{x}; \beta(G^{\bar{\boldsymbol{n}}}), \theta) &\equiv -\beta(G^{\bar{\boldsymbol{n}}}) + \int_{\varepsilon} \underline{d}(\underline{H}(\boldsymbol{x}, G^{\bar{\boldsymbol{n}}}, \varepsilon; \theta), \overline{H}(\boldsymbol{x}, G^{\bar{\boldsymbol{n}}}, \varepsilon; \theta)) dF_{\varepsilon}(\varepsilon) \\ m_4(\boldsymbol{x}; \beta(G^{\bar{\boldsymbol{n}}}), \theta) &\equiv \beta(G^{\bar{\boldsymbol{n}}}) - \int_{\varepsilon} \overline{d}(\underline{H}(\boldsymbol{x}, G^{\bar{\boldsymbol{n}}}, \varepsilon; \theta), \overline{H}(\boldsymbol{x}, G^{\bar{\boldsymbol{n}}}, \varepsilon; \theta)) dF_{\varepsilon}(\varepsilon). \end{aligned}$$

*Then* $\mathcal{H}_P[\beta(G^{\bar{\boldsymbol{n}}}), \theta] \subset \mathcal{O}_P[\beta(G^{\bar{\boldsymbol{n}}}), \theta]$.

Theorem 2.4 is the key result of my paper. It provides us with a joint set of parameters for both the network statistic of interest as well as the network formation parameters that are consistent with the partially observed network $G^{\bar{\boldsymbol{n}}}$. I can leverage this outer region to answer many interesting questions, including: who is the most influential player, who should we optimally target to spread information about a new vaccine, and who is the key player in a criminal network? I use calibrated projection Kaido, Molinari, and Stoye (2019) as implemented by Kaido, Molinari, Stoye, and Thirkettle (2017) to construct confidence intervals. I conclude with a simulation study to show the applicability of my framework.

## 2.6 Application

I apply my framework to the Katz-Bonacich Centrality. I specify the marginal utility of links as a linear function of three terms:

$$\Pi_{ij}(G, \boldsymbol{x}, \varepsilon; \theta) = \theta_0 |x_i - x_j| + \gamma_1 \frac{1}{n-1} \sum_{k \in \boldsymbol{n}: k \neq i} G_{jk} + \gamma_2 \frac{1}{n-2} \sum_{k \in \boldsymbol{n}: k \neq i} G_{ik} G_{jk} + \varepsilon_{ij},$$

The first term $\theta_0|x_i - x_j|$ allows for homophily. This is the concept that individuals are more likely to form friends with individuals who have similar characteristics. I set $\theta_0 < 0$ so that if $x_i \neq x_j$, then the marginal utility over the link is smaller and hence a friendship is less likely to occur. I assume $x_i$ is i.i.d, binary (e.g., the individual's sex), and $\boldsymbol{P}(x_i = 1) = 0.5$. This model allows for two channels of spillovers: popularity spillover $\gamma_1 \frac{1}{n-1} \sum_{k \in \boldsymbol{n}: k \neq i} G_{jk}$ and mutual friend spillover $\gamma_2 \frac{1}{n-2} \sum_{k \in \boldsymbol{n}: k \neq i} G_{ik} G_{jk}$.[20] I also assume that the additive error $\varepsilon_{ij}$ is distributed

[20]The simulations in Sheng (2018) is a variation of the model I present. In her paper, $\gamma_1 = 0$ (i.e., no popularity spillovers) and utility is transferable.

i.i.d $N(0,1)$ with $\varepsilon_{ij} = \varepsilon_{ji}$, and $\varepsilon_{ij} \perp\!\!\!\perp x_k$ for all $i, j, k$. I report bounds on the unweighted Katz-Bonacich centrality, $d(G) = (I - \lambda G)^{-1}\mathbf{1}$. The decay parameter $\lambda$ is in $[0, \bar{\lambda}]$, where $\bar{\lambda}$ is selected to ensure that the network statistic is well defined.

I first restrict the model to one spillover with no homophily in order to understand the performance of the model and the variation required to pin down the set of structural parameters. I show that this parsimonious model is nearly point identified and provides tight bounds on KBC.

### 2.6.1 Popularity Spillover Only

To understand the performance of the model, I first shut down observed characteristics and mutual-friend spillovers by setting $\theta_0 = 0$ and $\gamma_2 = 0$. I assume that this is known to the researcher. In this case, marginal utility is given by

$$\Pi_{ij}(G, \varepsilon; \theta) = \gamma_1 \frac{1}{n-1} \sum_{k \in \boldsymbol{n}: k \neq i} G_{jk} + \varepsilon_{ij}.$$

I first report the size and density of the admissible lattice. I vary the size of the spillover $\gamma_1 \in [0, 3]$ and set the population size equal to $n = 50$. Figure A.1c displays the expected fraction of links in $\underline{G}$ (solid line) and $\overline{G}$ (dotted line). The vertical distance between $\overline{G}$ and $\underline{G}$ is proportional to the size of the admissible lattice and counts the fractional difference in the number of links between $\overline{G}$ and $\underline{G}$. On average, the lattice is narrow: for a fixed value of $\gamma_1$, there is little variation in the overall density of the equilibrium network. Increasing the size of the spillover results in a denser network. As $\gamma_1$ approaches 3, the network becomes a complete network where all individuals are friends. Next, I increase the population size to $n = 100$ in Figure A.1d. For every value of $\gamma_1$, the lattice becomes more dense. The intuition here is that individuals now have a larger set of potential friends and experience larger spillovers. In addition, the difference in the density between $\overline{G}$ and $\underline{G}$ also declines. What could be happening here is that the lattice is converging to the complete network as $n$ increases.

Table A.1 reports the outer region for $\gamma_1$. The first column reports the identified region using subnetworks up to size $q = 2$. The second column sets $q = 3$ and the third sets $q = 4$. The rows report results varying the number of individuals in the network $n \in \{50, 100\}$ and the true value for the popularity spiller $\gamma_1 \in \{0, 0.5, 1\}$. The model is effectively point identified when $\gamma_1 = 0$ and is partially identified with $\gamma_1 > 0$. In all cases the identified set is narrow. For example, the identified set is $[0.985, 1.018]$ when $\gamma_1 = 1$. As expected, the identified set tightens as I increase the maximum size of the subnetwork from $q = 2$ to $q = 4$. The identified set is tighter when the population size increases to $n = 100$ as well. Reflecting on Figure A.1d, this makes sense as the admissible lattice tightens when $n$ increases, implying that bounds on the distribution of subnetworks also tighten.

Table A.3 reports the outer region for the average level of Katz-Bonacich Centrality using Theorem 2.4. I vary both the size of the popularity spillover $\gamma_1$ as well as the decay parameter $\lambda$. I report bounds using maximum subnetwork sizes ranging from $q = 2$ to $q = 4$. I also report the true average level of KBC and the worst-case bounds. When the popularity spillovers are zero (i.e., $\gamma_1 = 0$) the model is point identified and equivalent to an exponential random graph model. In this case, the outer region for average KBC is a single point and is equal to the true value of KBC, while worst-case bounds are fairly wide. Increasing the size of the popularity spillovers result in wider bounds on average KBC, but remain tight. In the least preferable case, the worst-case bounds are $[2.85, 5.53]$, which are 12 times wider than the bounds that my model predicts, $[4.02, 4.24]$. Using my framework, therefore, I obtain informative bounds on the network statistic of interest.

I also report results for the case where only the mutual-friend spillover is present, see Appendix A.5 and A.4. These are largely in line with what I find for the model with only popularity spillovers. I next consider the model with two channels of spillovers.

### 2.6.2 Both Spillovers

Suppose now that marginal utility contains both channels of spillovers, but no homophily:

$$\Pi_{ij}(G, \varepsilon, \theta) = \gamma_1 \frac{1}{n-1} \sum_{k \in \boldsymbol{n}: k \neq i} G_{jk} + \gamma_2 \frac{1}{n-2} \sum_{k \in \boldsymbol{n}: k \neq i} G_{ik} G_{jk} + \varepsilon_{ij}.$$

This is a DGP that has not been explored in previous simulation studies – simulation studies in prior literature restrict attention to models with one spillover. Figure A.2 reports the density and the size of the admissible lattice where $(\gamma_1, \gamma_2) \in [0, 2] \times [0, 2]$. The first two panels display the fraction of links in $\underline{G}$ and $\overline{G}$, respectively. Similar to what I find for the single-spillover model, the density approaches 1 as the size of the spillovers increase. The third panel displays the fractional difference between $\overline{G}$ and $\underline{G}$. The maximal difference is only 1%.

Tables A.4 and A.5 report the projected outer region for $(\gamma_1, \gamma_2)$. In contrast to the previous results, the outer region is wide, especially for outer regions constructed using $q = 2$. For example, when $(\gamma_1, \gamma_2) = (0.5, 0.5)$ and setting $q = 2$, I obtain projected outer regions equal to $[0, 0.9]$ for $\gamma_1$ and $[0, 1.16]$ for $\gamma_2$. The size of the outer region rapidly contracts with the maximum subnetwork size. One particularly remarkable case is $(\gamma_1, \gamma_2) = (0, 0.5)$ with $q = 4$, where the projected outer regions are $[0, 0.03]$ and $[0.44, 0.50]$ for $\gamma_1$ and $\gamma_2$, respectively. In reflection of Figure A.2 this is not surprising. I only find a 1% difference in the density of $\overline{G}$ and $\underline{G}$, indicating that the theoretical sharp identified set based on the full network is narrow. I lose information by only considering subnetworks with restricted sizes. Since my model remains tractable for relativity large subnetwork sizes, I am able to feasibly obtain tighter bounds on this particular model.

While the bounds on the preference parameters are wide, the bound on KBC is narrow. Table A.6 and A.7 reports identified regions for KBC when I allow for two channels of spillovers in the network-formation process. I set $\lambda = 0.01$ in Table A.6 and $\lambda = 0.02$ in Table A.7. I vary $(\gamma_1, \gamma_2) \in \{0, 0.25, 0.5\} \times \{0, 0.25, 0.5\}$. Despite the fact that the identified region for the preference parameters are typically wide and seemingly uninformative, I obtain precise bounds on network statistic of interest. The outer region on average KBC is $[3.725, 3.792]$ when $(\gamma_1, \gamma_2) = (0.5, 0.5)$ and $\lambda = 0.02$. In contrast, the worst-case bounds are $[2.376, 5.328]$. As a result my bounds are 44 times narrower than the worst-case bounds. Overall, the bounds on KBC are informative for various values of $(\lambda, \gamma_1, \gamma_2)$.

## 2.7 Conclusion

Social network data with links missing at random is very common in the social sciences. Applied researchers are often faced with the dilemma of estimating a network statistic when the social network is partially observed and formed endogenously. With cross-sectional network data, the literature either: (1) ignores the fact that the network is partially observed in which case the reported network statistic is incorrect; or (2) the missing links are filled in with an exogenous network-formation process. However, in practice, social networks are formed strategically. For example, students in a classroom choose their friends based on the popularity of others. I provide a framework to obtain informative bounds on network statistics in a partially observed network whose formation I explicitly model. I assume that individuals endogenously form links according to the standard social-network formation model of complete information and pairwise stability. I assume that the researcher has access to cross-sectional data from multiple partially observed networks, where the links are missing at random. I obtain a computationally tractable method to obtain bounds on both the preferences determining network formation processes and network statistics. In a simulation study on the Katz-Bonacich centrality measure, I dramatically reduce the worst-case bounds, which do not use the network formation model. I obtain from my procedure bounds that are 44 times narrower than the worst-case bounds.

CHAPTER 3

# DISCRETE CHOICE UNDER RISK WITH LIMITED CONSIDERATION

Levon Barseghyan, Francesca Molinari, and Matthew Thirkettle

## 3.1 Introduction

This paper is concerned with learning decision makers' (DMs) preferences using data on observed choices from a finite set of risky alternatives with monetary outcomes. The prevailing empirical approach to study this problem merges expected utility theory (EUT) models with econometric methods for discrete choice analysis.[1] Standard EUT assumes that the DM assesses a risky alternative by computing its expected utility; evaluates all available alternatives; and chooses the alternative yielding highest expected utility. The DM's risk aversion is determined by the concavity of her underlying Bernoulli utility function. The set of all alternatives – the *choice set* – is assumed to be observable by the researcher.

We depart from this standard approach by proposing a discrete choice model with unobserved heterogeneity in risk aversion and unobserved heterogeneity in *consideration sets*. Each DM evaluates only the alternatives in her consideration set, which is a subset of the choice set. Hence, stochastic choice is driven both by different rankings of alternatives induced by unobserved heterogeneity in risk preferences and by different sets of alternatives considered. Our first contribution is to establish that the requirements of standard economic theory, coupled with a

[1] For a non-exhaustive list of papers in this literature see (Starmer, 2000) and (Barseghyan, Molinari, O'Donoghue, & Teitelbaum, 2018). We discuss an important class of non-expected utility theory models and how our analysis applies to these models later in the paper.

slight strengthening of the classic conditions for semi-nonparametric identification of discrete choice models with full consideration and identical choice sets (see, e.g., Matzkin (2007)), yield semi-nonparametric identification of both the distribution of unobserved heterogeneity in risk aversion and the distribution of consideration sets.[2,3]

Our second contribution is to provide a simple method to compute our likelihood-based estimator, whose computational complexity grows polynomially in the number of alternatives in the choice set.[4] In particular, our method does not require enumerating all possible subsets of the choice set. If it did, the computational complexity would grow exponentially with the size of the choice set.

Our third contribution is to elucidate the applicability and the advantages of our framework over the standard application of random utility models (RUMs) with additively separable unobserved heterogeneity (e.g., Mixed Logit models) and full consideration. First, our model can generate zero shares for non-dominated alternatives. Second, the model has no difficulty explaining relatively large shares of dominated alternatives. Third, in markets with many insurance domains, our model can match not only the marginal but also the joint distribution of choices across domains. Forth, our framework is immune to an important criticism recently raised by Apesteguia and Ballester (2018) against using standard RUMs to study decision making under risk. As these authors note, combining standard EUT

[2]In fact, with a binary consideration set, the former and the latter coincide.

[3]The identification results are semi-nonparametric because we specify the utility function up to a DM-specific preference parameter. We establish nonparametric identification of the distribution of the latter.

[4]The function evaluation time of the log-likelihood objective function grows linearly with the number of alternatives. Provided that the objective function is locally concave, the local rate of convergence of the standard SQP program is quadratic. See, for example, (Boggs & Tolle, 1995).

with additive noise results in non-monotonicity of choice probabilities in the risk preferences, a clearly undesirable feature.

In general, distinguishing heterogeneous preferences from heterogeneous consideration using discrete choice data is a formidable task. When a DM chooses an alternative, this can be either because that alternative yields the highest expected utility among those in her entire choice set or because the DM does not consider some better available alternatives and the chosen one is the best in her consideration set, implying different distributions of preferences. We show that this seemingly inescapable identification problem can be resolved under certain conditions by leveraging standard requirements of economic theory. Specifically, our random preference models satisfy the classic Single Crossing Property (SCP) of Mirrlees (1971); Spence (1974): the preference order of any two alternatives switches only once on the support of the preference coefficient.[5] The SCP is central to important studies of decision making under risk, as well as those in other fields of Economics.[6] More so, as we make clear, the SCP is necessary for nonparametric identification of the preference parameter distribution in the standard model with full consideration and homogeneous observed choice sets. Coupled with three additional requirements (imposed in the literature on point identification of limited consideration models), we show that the SCP delivers nonparametric identification of the preference-parameter distribution even in the presence of unobserved heterogeneity in consideration sets. The first two requirements are: (1) specification of a

[5]The EUT framework with concave Bernoulli utility satisfies the SCP. The SCP requires that if a DM with a certain degree of risk aversion prefers a safer lottery to a riskier one, then all DMs with higher risk aversion also prefer the safer lottery. As we discuss in Section 3.8, many non-EU models, when they feature unidimensional preference heterogeneity, also satisfy SCP.

[6]E.g., (Athey, 2001; Apesteguia, Ballester, & Lu, 2017; Chiappori, Salanié, Salanié, & Gandhi, 2018).

consideration set formation model and (2) independence between unobserved heterogeneity in consideration and in risk preferences, conditional on observable characteristics. The second requirement is part of the standard framework: when all DMs consider the entire (non-stochastic) choice set, consideration is independent of underlying preferences by definition. Requirements (1) and (2) are motivated by Barseghyan, Coughlin, Molinari, and Teitelbaum (2019), who establish that in the absence of restrictions on the consideration set formation and its relation with risk preferences, one can partially but not point identify the distribution of risk preferences (even parametrically). The final requirement is that there exists a DM-characteristic with large support that shifts preferences over alternatives, but does not affect consideration. In RUMs, requiring existence of a regressor with large support (or an equivalent assumption) is necessary for nonparametric identification even without probabilistic consideration (Matzkin, 2007). The additional restriction in our framework, implicit in the full-consideration literature, is that the large-support regressor does not affect the probability of considering any of the alternatives in the choice set. We do, however, allow for the case that the large support regressor is not alternative specific, that is, it may only vary across DMs. Moreover, the consideration distribution may depend on the other characteristics of the DMs and of all alternatives.

With this structure in place, our identification result leverages a simple intuition: as the large-support regressor takes values sufficiently large or small, the alternatives in the choice set are unambiguously ranked for all possible realizations of the unobserved risk-preference coefficient. Hence, the choice frequency observed in the data is a function of only the consideration probabilities and, under weak restrictions, this function admits a unique solution for the consideration probabilities. The SCP then allows us to trace out the distribution of preferences given variation in the large-support regressor.

We describe our identification approach in detail for two probabilistic consideration models, each having up to as many parameters as the size of the choice set. These two models are different in nature and can be used as a blueprint to study the empirical content of many others, as we explain in the paper. The first model, termed the Alternative Specific Random Consideration (ARC) model, is inspired by Manski (1977) and Manzini and Mariotti (2014). In this model, alternative $j$ appears in the DM's consideration set with an alternative-specific probability $\varphi_j$ and each alternative enters the consideration set independent of all other alternatives. The second model, termed the Random Consideration Level (RCL) model, posits that the DM first draws the size of her consideration set, $l$ (her consideration level, possibly determined by her cognitive ability), and then randomly picks $l$ alternatives to consider, with each alternative having the same probability of being picked.

Of course, random preference models like the ones we consider are random utility models as originally envisioned by McFadden (1974) (for a textbook treatment see Manski (2007)). We show that our random preference models with probabilistic consideration can be written as RUMs with unobserved heterogeneity in risk aversion and with an additive error that has a discrete distribution with support $\{-\infty, 0\}$. It is then natural to draw parallels with the mixed (random coefficient) logit model (McFadden & Train, 2000). In our setting, the Mixed Logit model boils down to assuming that, given the DM's risk aversion, her evaluation of an alternative equals its expected utility summed with an unobserved heterogeneity term capturing the DM's idiosyncratic taste for unobserved characteristics of that alternative. However, in some markets it is hard to envision such characteristics:

For example, many insurance contracts are identical in *all* aspects *except* for the coverage level and price.[7] In other contexts, unobservable characteristics may affect choice mostly via consideration – as we model – rather than via "additive noise".[8]

As in the Mixed Logit model, our models assume independence of the additive error with the observable payoff-relevant characteristics and the unobservable heterogeneity in preferences. However, in the ARC model, the additive error is independent across alternatives but is not identically distributed; in the RCL model, the additive error is identically distributed but is not independent across alternatives; and in the Mixed Logit, the additive error is i.i.d. across alternatives. These differences generate contrasting implications in several respects.

First, the RCL model and the Mixed Logit model generally imply that each alternative has a positive probability of being chosen, while the ARC model can generate zero shares by setting the consideration probability of a given alternative to zero. Second, the RCL model and the Mixed Logit model satisfy a *Generalized Dominance Property* that we derive: if for any degree of risk aversion alternative $j$ is dominated by either alternative $k$ or $m$, then the probability of choosing $j$ must be no larger than the probability of choosing $k$ or $m$. The ARC model does not abide Generalized Dominance. Third, in the ARC and the RCL models, choice probabilities depend on the ordinal expected utility rankings of the alternatives, while, in the Mixed Logit, choice probabilities depend on the cardinal expected utility rankings. As we explain in Section 3.5, this difference implies that choice

[7]E.g., employer provided health insurance, auto, or home insurance offered by a single company.

[8]E.g., a DM may only consider those supplemental prescription drug plans that cover specific medications.

probabilities are monotone in risk preferences in our models, while in the Mixed Logit model they are not (Apesteguia & Ballester, 2018). Armed with the identification results obtained for the ARC and RCL models, we show in Section 3.6 that our approach easily extends to the case where consideration sets form based on liquidity constraints or behavioral phenomena such as extremeness aversion.

Our empirical application is a study of households' deductible choices across three lines of insurance coverage: auto collision, auto comprehensive, and home (all perils). The aim of our exercise is to estimate the underlying distribution of risk preferences and the consideration parameters; to assess the resulting fit of the models; and to evaluate the monetary cost of limited consideration. We find that the ARC model does a remarkable job at matching the distribution of observed choices, and because of its aforementioned properties, outperforms both the RCL model and the EUT model with additive extreme value type I (Gumbel) error. Under the ARC model, we find that although households are on average strongly risk averse, they consider lower coverages more often than higher coverages. Finally, the average monetary losses resulting from limited consideration are \$49.

## 3.2 Related Literature

The literature concerned with the formulation, identification, and estimation of discrete choice models with limited consideration is vast. However, to our knowledge, there is no previous work applying such models to the study of decision making under risk, except for the contemporaneous work of Barseghyan et al. (2019). They study models of decision making under risk, where unobserved heterogeneity in preferences as well as in choice and/or consideration sets is allowed for. They additionally allow for arbitrary dependence between consideration sets and preferences, and impose no restrictions on the consideration set formation process. They show that such unrestricted forms of heterogeneity yield, in general, partial but not point identification of the model, even when a parametric distribution for preference heterogeneity is specified. They obtain bounds on the distribution of consideration sets' size, but no other features of the distribution of consideration sets can be learned.

In this paper we take a conceptually different approach. As in the entire related literature on point identification of limited consideration models, we maintain independence of consideration sets and preferences and we focus on specific consideration sets' formation processes. The latter are grounded in a sizable literature spanning experimental economics, microeconomics, behavioral economics, psychology, and marketing which aims to formalize the cognitive process underlying the formation of consideration sets.[9] For the remainder of this section, we discuss the literature on identification of limited consideration models that are closely related to the ARC and RCL models.

To identify parametric models of demand, previous contributions in this area have typically relied on auxiliary data revealing the consideration set composition (Draganska & Klapper, 2011; Conlon & Mortimer, 2013; Honka & Chintagunta,

[9]See, e.g., (Simon, 1959; Tversky, 1972; Howard, 1977; Manski, 1977; Treisman & Gelade, 1980; Hauser & Wernerfelt, 1990; Shocker, Ben-Akiva, Boccara, & Nedungadi, 1991; Roberts & Lattin, 1991; Ben-Akiva & Boccara, 1995; Eliaz & Spiegler, 2011; Masatlioglu, Nakajima, & Ozbay, 2012; Manzini & Mariotti, 2014; Caplin, Dean, & Leahy, 2018). Even when DMs pay full attention, they may face unobserved constraints on what alternatives they can choose e.g., Gaynor, Propper, and Seiler (2016).

2017), or on the existence of regressor(s) that impact utility but not consideration (or *vice versa*) (Goeree, 2008; Gaynor et al., 2016; Heiss, McFadden, Winter, Wuppermann, & Zhou, 2016; Hortaçsu, Madanizadeh, & Puller, 2017).[10] In contrast, we establish semi-nonparametric identification of the distributions of unobserved heterogeneity in preferences and in consideration sets through a combination of the SCP with an exclusion restriction and a large support assumption.

A recent related literature, closest to our own work, studies various departures from the tight parametric structure of the earlier analysis of limited consideration models. Dardanoni, Manzini, Mariotti, and Tyson (2017) consider a stochastic choice model with homogeneous preferences and heterogeneous cognitive types.[11] The cognitive types are implemented through the RCL model and a variant of the ARC model. In the RCL model, the cognitive type is the number of alternatives the DM is able to consider. In the ARC model, the type is the probability with which the DM considers an alternative (which is assumed constant across alternatives). The authors show how one can learn the moments of the distribution of types from a single cross section of aggregate choice shares. A key assumption for identification is that there exists a default alternative and the researcher observes the frequency with which the default alternative is chosen. In our paper, we do not require that the default option is observed and we are flexible on its existence.

[10](Crawford, Grithz, & Iariax, 2017) estimate a Fixed-Effect Logit type model that (1) utilizes observed purchase decisions (in a panel or a group-homogeneous cross-section) to construct "sufficient sets" of alternatives that lie within DMs' feasible sets; (2) given the "sufficient sets", uses classic techniques on estimating logit models on subsets of feasible sets.

[11]Heterogeneous tastes are also explored. To obtain identification, however, one of two strong assumptions are imposed. Either the taste distribution is known or preferences are linear in observable alternative characteristics and there is an additive error term with extreme value type 1 distribution.

Cattaneo, Ma, Masatlioglu, and Suleymanov (2017) propose a general random attention model where the probability of drawing a consideration set decreases as the choice set enlarges. Their model, however, does not allow for unobserved heterogeneity in preferences, and yields partial identification results while requiring rich observable variation in the choice set.

Abaluck and Adams (2017) study identification of an additive error random utility model with consideration formation similar to that in our ARC model, a variant of the RCL model (also considered in Ho, Hogan, and Scott Morton (2017); Hortaçsu et al. (2017); Heiss et al. (2016)), and a mix of the two. Abaluck and Adams (2017) method and ours are distinct and complementary. They require a default option, and the existence of a regressor (e.g., price) that is alternative specific and enters the indirect utility function linearly (or additively separably with shape restrictions). The price of each alternative is required to have large support, to exhibit cross-alternative variation (i.e., independent variation for each alternative), and to be excluded from the consideration probability of all other alternatives.

When modeling choice under risk, concave utility yields that price enters neither linearly nor additively separably. More importantly, our work aims at providing a method to learn DMs' risk preferences and consideration probabilities from their choices of insurance products. Many important recent empirical contributions in this area use data from a single company – either a firm selling insurance or a firm offering health insurance to its employees.[12] In their data, observable characteristics with large support are typically DM specific and not alternative specific, so

[12]See, e.g.,(Cohen & Einav, 2007; Einav, Finkelstein, Pascu, & Cullen, 2012; Barseghyan, Molinari, O'Donoghue, & Teitelbaum, 2013; Handel, 2013; Bhargava, Loewenstein, & Sydnor, 2017).

that the Abaluck and Adams (2017) method does not apply.[13] In contrast, our framework does not require a default option and we only assume that the large support regressor is independent of consideration set formation. This regressor may or may not be alternative specific.

## 3.3 Models

### 3.3.1 Decision Making under Risk in a Market Setting: An Example

To set the stage we consider the following insurance market. There is an underlying risk of a loss with probability equal to $\mu$ that varies across DMs. A finite number of insurance alternatives are available against this loss. Each alternative is a pair $(d_j, p_j), j \in \{1, .., D\}$. The first element is a deductible, which is the DM's out of pocket expense in case a loss occurs. Deductibles are decreasing with index $j$. All deductibles are less than the lowest realization of the loss. The second element is a price, which also varies across DMs. For each DM there is a baseline price $\bar{p}$ that determines prices for all alternatives faced by the DM according to a multiplication rule, $p_j = g_j \cdot \bar{p} + \delta$, where $\delta$ is a small positive amount and $g_j$ increases with $j$: lower deductibles provide more coverage, and hence cost more. Both $g_j$ and $\delta$ are invariant across DMs. The lotteries that the DM faces can be written as follows:

$$L_j(x) \equiv \left(-p_j, 1 - \mu; -p_j - d_j, \mu\right),$$

[13] For example, in the context of health insurance there is a fixed price for each insurance plan offered to all employees and there is large variation in risk across employees.

where $x \equiv (\bar{p}, \mu)$. DMs are expected utility maximizers. Given initial wealth $w$, the expected utility of deductible lottery $L_j(x)$ is given by

$$EU(L_j(x)) = (1-\mu)\,u\,(w-p_j) + \mu u\,(w-p_j-d_j)\,,$$

where $u(\cdot)$ is a Bernoulli utility function defined over final wealth states. We assume that $u(\cdot)$ belongs to a certain family of utility functions that are fully characterized by a single coefficient $\nu$ (e.g. Constant Absolute Risk Aversion (CARA), Constant Relative Risk Aversion (CRRA) or Negligible Third Derivative (NTD)).[14] This coefficient of risk aversion is randomly distributed across DMs and has bounded support.

The relationship between risk aversion, underlying prices, and loss probabilities is standard. At sufficiently high $\bar{p}$ or low $\mu$, less coverage is always preferred to more coverage: $L_1 \succ L_2 \succ L_3 \succ ... \succ L_D$ for all $\nu$ on the support. At sufficiently low $\bar{p}$ or high $\mu$, we have the opposite ordering: $L_D \succ L_{D-1} \succ L_{D-2} \succ ... \succ L_1$. For moderate levels of prices and loss probabilities things are more interesting: for each pair of deductible lotteries $j < k$ there is a cutoff value $c_{j,k}(x)$ in the interior of the risk-preference coefficient support. On the left of this cutoff the higher deductible is preferred and on the right the lower deductible is preferred. In other words, $c_{j,k}(x)$ is the unique coefficient of risk aversion that makes the DM indifferent between $L_j$ and $L_k$. Those with lower $\nu$ choose the riskier alternative $L_j$, while those with higher $\nu$ choose the safer alternative $L_k$. Note that $c_{j,k}(x)$ is a continuous function, since the expected utility from each deductible lottery is continuous in $x$ as well as in $\nu$.

[14]Under CRRA, it is implied that DMs' initial wealth is known to the researcher. NTD utility is defined in (Cohen & Einav, 2007) and in (Barseghyan et al., 2013).

### 3.3.2 The Model under Full Consideration

There is a continuum of DMs who face a choice among a finite number of alternatives, i.e., the choice set, which is denoted $\mathcal{D} = \{1, \ldots, D\}$. Alternatives vary by their utility-relevant characteristics. One characteristic, $d_j \in \mathbb{R}, j \in \mathcal{D}$, is DM invariant. When it is unambiguous, we may write $d_j$ instead of "alternative $j$". Other characteristics, denoted by $x_j \in \mathcal{X}_j \subset \mathbb{R}^q$, may vary across DMs as well as across alternatives. That is, alternative $j$ is fully characterized by $(d_j, x_j)$. We denote $x = (x_1, ..., x_D)$ and $\mathcal{X} = \mathcal{X}_1 \otimes ... \otimes \mathcal{X}_D$. Given these characteristics, each DM's preferences over the alternatives are defined by a utility function $U_\nu(d, x)$. The latter is fully described by a DM-specific index $\nu$ assumed to be distributed according to $F(\cdot)$ over a bounded support $\Gamma = [0, \bar{\nu}]$.[15] We assume that the random variables $\nu$ and $x$ are independent (and that demographic variables, if available, have been conditioned on). The DM's draw of $\nu$ is not observed by the researcher and $F(\cdot)$ is assumed to be continuous and otherwise left completely unspecified. Going forward, we only consider models satisfying a basic Single Crossing Property, as defined below.

**Definition 3.1** (Single Crossing Property)**.** *The DM's preference relation over alternatives satisfies the Single Crossing Property iff the following condition holds: For all $j, k,\ j \neq k$ there exists a continuous function $c_{jk} : \mathcal{X} \to \mathbb{R}_{[-\infty,\infty]}$ (or $c_{kj} : \mathcal{X} \to \mathbb{R}_{[-\infty,\infty]}$) such that*

$$U_\nu(d_j, x) > U_\nu(d_k, x) \quad \forall \nu \in (-\infty, c_{jk}(x))$$

$$U_\nu(d_j, x) = U_\nu(d_k, x) \quad \nu = c_{jk}(x)$$

$$U_\nu(d_j, x) < U_\nu(d_k, x) \quad \forall \nu \in (c_{jk}(x), \infty).$$

[15] We assume that while $\nu$ has bounded support, the utility function is well defined for any real valued $\nu$.

That is, we require that the DM's ranking of alternatives is monotone in $\nu$: if a DM with a certain degree of risk aversion prefers a safer (riskier) asset to a riskier (safer) one, then all DMs with higher (lower) risk aversion also prefer the safer (riskier) asset.[16]

Full consideration is maintained in this subsection: each DM considers *all* alternatives in the choice set and chooses the one with highest utility (that is, consideration and choice sets coincide). Assumption 3.1 is a data requirement which guarantees this model's identification: There must be sufficient variation in a utility-relevant characteristic(s) to move the cutoffs (the single crossing points in Definition 3.1) through the support for the preference coefficient.

**Assumption 3.1** (Large Support). *For all $\nu \in \Gamma$ there exists $x \in \mathcal{X}$ and alternative $j$ such that either: (1) $c_{j,k}(x)$ exists for all $k \neq j$ and $\nu = \min_{k \neq j} c_{j,k}(x)$; or (2) $c_{k,j}(x)$ exists for all $k \neq j$ and $\nu = \max_{k \neq j} c_{k,j}(x)$.*

**Theorem 3.1.** *Suppose Assumption 3.1 holds. Then $F(\cdot)$ is identified.*

*Proof.* Fix any $\nu \in \Gamma$. Find $x$ and alternative $j$ such that $\nu = \min_{k \neq j} c_{j,k}(x)$ or $\nu = \max_{k \neq j} c_{k,j}(x)$. Then one of the following is true:

[16] Since we allow the cutoffs to be infinite, our regularity condition does not exclude strongly dominated choices, i.e. situations in which $d_k$ is preferred to $d_j$ for all values of $\nu$. In the context of risk preferences this definition of strong dominance is equivalent to first order stochastic dominance. When $u(\cdot)$ is restricted to the class of concave utility functions strong dominance is equivalent to second order stochastic dominance.

$$
\begin{aligned}
F(\nu) =& F\left(\min_{k\neq j} c_{j,k}(x)\right) = Pr(d = d_j|x) \\
F(\nu) =& F\left(\max_{k\neq j} c_{k,j}(x)\right) = 1 - Pr(d = d_j|x).
\end{aligned}
$$

Since $Pr(d = d_j|x)$ is identified by the data, $F(\nu)$ is identified. □

Theorem 3.1 is akin to the "full-support" identification result (Chamberlain, 1986; Heckman, 1990; Lewbel, 2019) and the intuition is straightforward: there must be sufficient variation in the underlying exogenous characteristics to trace out the distribution of $\nu$ over its entire support. Some variant of Assumption 3.1 is also necessary for identification. If, for example, there exists an interval $[\nu_*, \nu^*] \subset \Gamma$ such that for all $k, j$ there is no $x$ with $c_{k,j}(x) \in [\nu_*, \nu^*]$, then $F(\cdot)$ will not be identified in this interval. Simply put, the data does not provide any information about the distribution of the preference coefficient in this region. Next, we present two models of limited consideration. Each of them has the same underlying primitives as the benchmark model, except the consideration set formation is stochastic.

### 3.3.3 Alternative Specific Random Consideration Model

In the Alternative Specific Random Consideration (ARC) Model Manski (1977); Manzini and Mariotti (2014), each alternative $d_j$ appears in the consideration set with probability $\varphi_j$ independently of other alternatives. These probabilities are assumed to be the same across DMs. We note that without loss of generality $\varphi_j$ can be interpreted as a function of exogenous characteristics (such as advertisement) that are not utility relevant. In such a case, all of the results below should be interpreted as conditional on a given value of these characteristics.[17] Once the consideration set is drawn, the DM chooses the best alternative according to her preferences. Given that each alternative is considered probabilistically, it is possible that none of the alternatives enter the consideration set. In particular, with probability $\prod_{k=1}^{D}(1-\varphi_k)$ the consideration set is empty. Hence, to close the model, we require a completion rule specifying the behavior of the DM in the case of non-consideration. We offer four possible completion rules, each suited for application in different market settings.

**Coin Toss**: Assume $\varphi_j < 1$ for all $j$. Then there is positive probability that the DM does not consider any alternative. In such a case, the DM randomly uniformly picks one alternative from the choice set, i.e. each alternative has probability $\frac{1}{D}$ of being chosen. Coin Toss is consistent with scenarios in which DMs must choose an alternative (e.g. a deductible when buying home insurance), but lack the desire or ability to meaningfully evaluate them.[18]

**Default Option**: Assume $\varphi_j < 1$ for all $j$. If no alternative is considered, the DM chooses a preset alternative. This completion rule is applicable to scenarios where, without the DM's active choice, she is assigned a pre-specified alternative from the choice set (e.g. employer provided benefits such as 401k allocations and medical insurance).

**Preferred Options** Manski (1977): Some alternative(s) is (are) always considered, i.e. $\varphi_j = 1$ for some $j$. The identity of these alternatives does not have to

[17]More so, these characteristics may include a strict subset of $x$. As explained later, we only need one element of $x$ to have certain properties and be consideration irrelevant.

[18]In a classical IO setting, this type of completion rule is consistent with, for example, a shopper randomly choosing a chip packet from the shelf without carefully evaluating the utility derived from consuming various flavors of chips.

be known to the researcher. However, if there exist multiple $j$'s such that $\varphi_j = 1$, then these alternatives should be adjacent to each other in the following sense. If there exists an $x$ such that for all $\nu \in \Gamma$ some non-preferred alternative dominates a preferred alternative, then it also dominates all other preferred alternatives. This completion rule captures market scenarios in which some alternatives are always discussed or emphasized by the sellers.

**Outside Option** Manzini and Mariotti (2014): Assume $\varphi_j < 1$ for all $j$. The first interpretation of this rule is that all DMs who draw the empty set exit the market and are not part of the data. A second interpretation of this rule is as follows. If the empty consideration set is drawn, then the DM redraws a consideration set according to Equation (3.1) below. The DM continues to draw consideration sets until a non-empty set is obtained.

For all completion rules, the probability that the consideration set takes realization $\mathcal{K}$ is

$$p(\mathcal{K}) \equiv \prod_{d_k \in \mathcal{K}} \varphi_k \prod_{d_k \in \mathcal{D} - \mathcal{K}} (1 - \varphi_k), \quad \forall \mathcal{K} \subset \mathcal{D}. \tag{3.1}$$

The differences in completion rules appear in the formulation of the likelihood function. A computationally appealing way to write the likelihood function is to determine the probability that a DM with preference coefficient $\nu$ chooses alternative $d_j$ conditional on characteristics $x$. Suppose the consideration set is not empty. Then, if $d_j$ is chosen, it is in the consideration set and every alternative that dominates it is not. Denote $\mathcal{B}_\nu(d_j, x)$ the set of alternatives that dominate $d_j$ for a DM with preference coefficient $\nu$ and characteristics $x$:

$$\mathcal{B}_\nu(d_j, x) \equiv \{d_k : U_\nu(d_k, x) > U_\nu(d_j, x)\}.$$

It follows that for the first three completion rules

$$Pr(d_j|x,\nu) = \varphi_j \prod_{d_k \in \mathcal{B}_\nu(d_j,x)} (1-\varphi_k) + r_j,$$

where $r_j$ is the term that accounts for the possibility of an empty consideration set. Under Coin Toss, $r_j = \frac{1}{D}\prod_{d_k \in \mathcal{D}}(1-\varphi_k)$. Under Default Option, $r_j = \prod_{d_k \in \mathcal{D}}(1-\varphi_k)$ if $j$ is the default alternative and is zero otherwise. Finally, under Preferred Options, $r_j = 0$. Integrating over $\nu$ we have that

$$Pr(d_j|x) = \int Pr(d_j|x,\nu)dF = \varphi_j \int \prod_{d_k \in \mathcal{B}_\nu(d_j,x)} (1-\varphi_k)dF + r_j.$$

Similarly, under Outside Option, we have that

$$Pr(d_j|x) = \int Pr(d_j|x,\nu)dF = \frac{1}{1-r}\varphi_j \int \prod_{d_k \in \mathcal{B}_\nu(d_j,x)} (1-\varphi_k)dF,$$

where $r \equiv \prod_{d_k \in \mathcal{D}}(1-\varphi_k)$.

We emphasize that these expressions for $Pr(d_j|x)$ do not require enumerating all possible consideration sets, which for large choice sets can be hard if not infeasible. Computation of $Pr(d_j|x)$ comes down to evaluating

$$\mathcal{I}(d_j|x) \equiv \int \prod_{d_k \in \mathcal{B}_\nu(d_j,x)} (1-\varphi_k)dF.$$

Given $\varphi$, the integrand $\prod_{d_k \in \mathcal{B}_\nu(d_j,x)}(1-\varphi_k)$ is piecewise constant in $\nu$ with at most $D-1$ breakpoints, corresponding to indifference points between alternatives $j$ and $k$ (i.e. $c_{j,k}(x)$ or $c_{k,j}(x)$). There are at least two methods to compute this integral. First, for every $d_j$ and $x$, we can directly compute the breakpoints and hence write $\mathcal{I}(d_j|x)$ as a weighted sum:

$$\mathcal{I}(d_j|x) = \sum_{h=0}^{D-1} \left( (F(\nu_{h+1}) - F(\nu_h)) \prod_{d_k \in \mathcal{B}_{\nu_h}(d_j,x)} (1-\varphi_k) \right),$$

where $\nu_h$'s are the sequentially ordered breakpoints augmented by the integration endpoints: $\nu_0 = 0$ and $\nu_D = \bar{\nu}$. This expression is trivial to evaluate given $F(\cdot)$ and breakpoints $\{\nu_h\}_{h=0}^{D}$. More importantly, since the breakpoints are invariant with respect to consideration probabilities, they are computed only once. This simplifies the likelihood maximization routine by orders of magnitude, as each evaluation of the objective function involves a summation over products with at most $D$ terms.

A second approach is to compute $\mathcal{I}(d_j|x)$ using Riemann approximation:

$$\mathcal{I}(d_j|x) \approx \frac{\bar{\nu}}{M} \sum_{m=1}^{M} \left( f(\nu_m) \prod_{d_k \in \mathcal{B}_{\nu_m}(d_j,x)} (1 - \varphi_k) \right),$$

where $M$ is the number of intervals in the approximating sum, $\frac{\bar{\nu}}{M}$ is the intervals' length, $\nu_m$'s are the intervals' midpoints, and $f(\cdot)$ is the density of $F(\cdot)$. Again, one does not need to evaluate the utility from different alternatives in the likelihood maximization. Instead, one *a priori* computes the utility rankings for each $\nu_m$, $m = 1, \ldots, M$.[19] These rankings determine $\mathcal{B}_{\nu_m}(d_j, x)$. The likelihood maximization is now a standard search routine over $\{\varphi_k\}_{k=1}^{D}$ and density $f(\cdot)$. Our theory restricts $f(\cdot)$ to the class of continuous functions. In practice, the search is over a class of non-parametric estimators for $f(\cdot)$.[20] If the density is parameterized, i.e. $f(\nu_m) \equiv f(\nu_m, \theta)$, then the maximization is over $\{\varphi_k\}_{k=1}^{D}$ and $\theta \in \Theta$. Note also that $\nu_m$'s are the same across all DMs, further reducing computational burden.[21]

[19] The resulting computational gains are exploited in importance sampling (Ackerberg, 2009).

[20] One could, for example, use normalized B-splines or a mixture of flexible distributions.

[21] Depending on the class of $f(\cdot)$, it may be more accurate to compute $\mathcal{I}(d_j|x)$ by substituting $\frac{\bar{\nu}}{M} f(\nu_m)$ with $F(\overline{\nu}_m) - F(\underline{\nu}_m)$, where $\overline{\nu}_m$ and $\underline{\nu}_m$ are the endpoints of the corresponding interval.

### 3.3.4 Random Consideration Level Model

In the Random Consideration Level (RCL) Model, the consideration set forms in two steps. In the first step, each DM draws a consideration level, $l$, that is independent of the preference coefficient $\nu$. The consideration level determines the size of the consideration set and it takes discrete values in $\{1, ..., D\}$ with probability $\phi_l$ such that $\sum_{l=1}^{D} \phi_l = 1$. In the second step, the consideration set is formed by drawing alternatives uniformly without replacement from the choice set until the DM obtains a set with cardinality equal to the consideration level $l$. The probability that the consideration set takes realization $\mathcal{K}$ of size $l$ is $p(\mathcal{K}|l) = \binom{D}{l}^{-1}$.

A computationally appealing way to write the likelihood function is as follows. First, consider a DM with a preference coefficient $\nu$ and characteristics $x$ and suppose that $d_j$ is her $m^{th}$-best alternative. The probability that $d_j$ is chosen is given by

$$Q_{l,m} = \begin{cases} \frac{\binom{D-m}{l-1}}{\binom{D}{l}} & \text{if } 1 \leq m \leq D - l + 1 \\ 0 & \text{otherwise.} \end{cases}$$

Denote $\mathbf{b}_\nu(d_j, x)$ the number of alternatives that dominate $d_j$ for an individual with preference coefficient $\nu$ and characteristics $x$: $\mathbf{b}_\nu(d_j, x) \equiv \text{card}(\mathcal{B}_\nu(d_j, x))$.[22]

[22]In the ARC model the identity of the alternatives dominating $j$ matters, while in this model only the number of dominating alternatives matter. The reason is that here all alternatives have equal probability of being considered. E.g., suppose that $D$ is 5 and $j$ is the second best alternative, so that $\mathbf{b}(j|x, \nu) = 1$. The second-best alternative is never chosen under full consideration, so that $Q_{5,1} = 0$. Under consideration level 4, the second best is chosen when the first best is not considered, which happens with probability $\frac{1}{5}$, that is $Q_{4,1} = \frac{1}{5}$. Under consideration level 1 an alternative is chosen iff it is in the consideration set, i.e. $Q_{1,1} = \frac{1}{5}$.

We can write

$$Pr(d_j|x) = \int Pr(d_j|x,\nu)dF = \int \sum_{l=1}^{D} \phi_l Q_{l,1+\mathbf{b}_\nu(d_j,x)} dF. \tag{3.2}$$

We employ similar techniques as those in Section 3.3.3 to compute the integral in Equation (3.2).

## 3.4 Identification

Identification in both the ARC and RCL models requires only minimal strengthening of assumptions relative to Assumption 3.1 for the full consideration case. We start with an example to illustrate this point and to highlight the intuition behind our results that follow.

### 3.4.1 Identification: An Example

Recall our example in Section 3.3.1. Suppose there are only two alternatives: $d_1$ is the high deductible and $d_2$ is the low deductible. From Theorem 3.1 it is clear that to identify the model under full consideration we need enough variation in $\bar{p}$ (and/or $\mu$) such that the cutoff $c_{1,2}(x)$ covers the entire support of the preference coefficient. Here, this variation is sufficient to identify *both* the consideration parameters and the distribution of the risk preferences.

We start with the ARC Model under one of the first three completion rules.[23] For each value of $x$ we have a single moment identified by the data:

[23]Identification under the Outside Default follows similar reasoning.

$$Pr(d = d_1|x) = \varphi_1\varphi_2 F(c_{1,2}(x)) + \varphi_1(1 - \varphi_2) + r_1. \tag{3.3}$$

The first term on the RHS of the Equation (3.3) captures the case where both alternatives are considered, and hence $d_1$ is chosen only if the preference coefficient is below the cutoff. The second term captures the case where only $d_1$ is considered, and thus it is chosen for all values of $\nu$. The third term is zero if at least one option is always considered, and otherwise captures the possibility that the consideration set is empty, and, depending on the completion rule, $r_1$ is equal to either $\frac{1}{2}(1 - \varphi_1)(1 - \varphi_2)$ or $(1 - \varphi_1)(1 - \varphi_2)$. Hence, rather than having a one-to-one mapping between $Pr(d = d_1|x)$ and $F(c_{1,2}(x))$ that would identify the latter as in Theorem 3.1, we have two additional unknown parameters. However, when Assumption 3.1 holds, we can find $x^0$ and $x^1$ such that $c_{1,2}(x^0) = 0$ and $c_{1,2}(x^1) = \bar{\nu}$. This implies that the consideration parameters are the solution to the following system of equations:

$$\begin{aligned}
Pr(d = d_1|x^0) &= \varphi_1(1 - \varphi_2) + r_1 \\
Pr(d = d_1|x^1) &= \varphi_1 + r_1.
\end{aligned}$$

It is straightforward to show that this system has a unique solution. Hence, identification relies on the assumption that variation in $x$ is sufficient to generate values for the cutoff $c_{1,2}(x)$ at the extremes of the support for the preference coefficient, which is also needed for identification in the model with full consideration as discussed in Section 3.3.2.[24] Once the consideration parameters are known, identification of $F(\cdot)$ follows from Equation (3.3), as long as both $\varphi_1$ and $\varphi_2$ are strictly positive.[25]

[24]Since observed variation in characteristics $x$ identifies the distribution of a latent variable, $x$ is referred to as the Lewbel special regressor (Lewbel, 2000, 2014).

[25]If either $\varphi_1 = 0$ or $\varphi_2 = 0$ then choice frequencies do not depend on $x$ and nothing can be learned about the distribution of $\nu$ as $F(c_{1,2}(x))$ drops out of Equation (3.3).

Similarly, under the RCL Model, for each value of $x$ we have a single moment identified by the data:

$$Pr(d = d_1|x) = \phi_2 F(c_{1,2}(x)) + \frac{1}{2}\phi_1 = (1 - \phi_1)F(c_{1,2}(x)) + \frac{1}{2}\phi_1. \tag{3.4}$$

Again, under Assumption 3.1, we can drive $c_{1,2}(x)$ either to zero or to $\bar{\nu}$, which turns the expression above into an equation with one unknown, namely $\phi_1$. Hence the consideration parameters are identified. As long as $\phi_1 < 1$, identification of $F(\cdot)$ follows.

The notable difference between Equations (3.3) and (3.4) is that the latter contains only one consideration parameter, while the former contains two. The reason follows from the restriction that the $\phi$'s sum to one, while no restriction is imposed on the sum of $\varphi$'s. To compensate for this missing moment condition in the ARC Model, the identification argument requires the use of an additional moment. This is the reason why identification of the ARC model will require somewhat stronger conditions.

In sum, using values of $x$ that put the cutoff at the extremes of the preference-coefficient space allows for identification of the consideration parameters. Once the consideration parameters are known, variation in $x$ pins down the preference-coefficient distribution. In the next two sections we proceed with formal arguments for identification of limited consideration parameters in both models. The conditions for identification of the preference-coefficient distribution are described in Section 3.4.3.

### 3.4.2 Identification of Consideration Parameters

We begin with the ARC model. We relegate all proofs to Appendix B.1.

**Theorem 3.2.** *Consider the Coin Toss or Outside Option completion rule. Suppose that there exist $x^0$, $x^1$, and a non-identity permutation $\{o_1, o_2, ..., o_D\}$ of the choice set such that $\forall \nu \in [0, \bar{\nu}]$*

$$L_1(x^0) \succ L_2(x^0) \succ \cdots \succ L_D(x^0),$$

$$L_{o_1}(x^1) \succ L_{o_2}(x^1) \succ \cdots \succ L_{o_D}(x^1).$$

*Then the consideration parameters $\{\varphi_1, \varphi_2, ..., \varphi_D\}$ are identified.*

While it appears that Theorem 3.2 (and the other results in this section) make use of the particular ordering of alternatives at $x^0$, the theorem can be stated with respect to any ordering of the available alternatives. The intuition for Theorem 3.2 is as follows. We need to identify $D$ parameters. Since the preference ordering is deterministic at $x^0$, the observed choice frequencies provide $D-1$ distinct moments. The last *distinct* moment is obtained from the choice frequency evaluated at $x^1$ for an alternative that moved position in the preference order (guaranteed by the permutation). Note that the conditions of the theorem allow for the presence of both dominated and dominating choices (choices that are better or worse than another alternative(s) regardless of the value of $x$). For example, in both preference orderings the best (or the worst) $D-2$ choices may be the same at $x^0$ and $x^1$, but the remaining two alternatives switch places.

**Theorem 3.3.** *Consider the Default Option completion rule. Denote $d_n$ the default option. Suppose that: (1) There exist $x^0$, $x^1$ and a non-identity permutation $\{o_1, o_2, ..., o_D\}$ of the choice set such that $\forall \nu \in [0, \bar{\nu}]$*

$$L_1(x^0) \succ L_2(x^0) \succ \cdots \succ L_D(x^0),$$

$$L_{o_1}(x^1) \succ L_{o_2}(x^1) \succ \cdots \succ L_{o_D}(x^1).$$

*(2) There exists an alternative $d_j \neq d_n$ such that $L_n(x^0) \succ L_j(x^0)$, $L_j(x^1) \succ$*

*$L_n(x^1)$, and $Pr(d_j|x^1) > 0$.*[26] *Then the consideration parameters $\{\varphi_1, \varphi_2, ..., \varphi_D\}$ are identified.*

The only difference between the conditions in Theorems 3.2 and 3.3 is in the latter case we additionally require an alternative, which is considered with positive probability, that switches rankings with the default option between $x^0$ and $x^1$. This is necessary to obtain information about the consideration parameter for the default option. To see why this must be the case, suppose that the default option is $d_D$. Given the ranking of options at $x = x^0$, it is immediate to see how $\varphi_1, \ldots, \varphi_{D-1}$ are identified sequentially since $Pr(d = d_j|x^0) = \varphi_j \prod_{k<j}(1 - \varphi_k)$.[27] However, we cannot learn $\varphi_D$ since the last moment at $x^0$ is redundant and in particular does not reveal any information about $\varphi_D$:

$$\begin{aligned} Pr(d = d_D|x^0) &= \varphi_D \prod_{k<D}(1 - \varphi_k) + r \\ &= \varphi_D \prod_{k<D}(1 - \varphi_k) + \prod_{k\leq D}(1 - \varphi_k) \\ &= \prod_{k<D}(1 - \varphi_k). \end{aligned}$$

Now suppose $d_D$ is dominated by all other alternatives at $x^1$ (so that there does not exist a $d_j \neq d_n$ satisfying the assumption in Theorem 3.3). By the same logic as above, $Pr(d = d_D|x^1)$ does not reveal $\varphi_D$.[28]

**Theorem 3.4.** *Consider the Preferred Options completion rule. Denote $d_{\underline{n}}, d_{\underline{n}+1} \ldots, d_{\overline{n}}$ to be the preferred options for some $\underline{n}$ and $\overline{n}$. Suppose that: (1) There exist $x^0$, $x^1$ and a non-identity permutation $\{o_1, o_2, ..., o_D\}$ of the choice set*

[26] The condition $Pr(d_j|x^1) > 0$ (or $Pr(d_j|x^0) > 0$) is equivalent to assuming $\varphi_j > 0$. A restriction on the data is testable, so the assumption $Pr(d_j|x^1) > 0$ is more appealing.

[27] We have that $\varphi_1 = Pr(d = d_1|x^0)$, $\varphi_2 = \frac{Pr(d=d_2|x^0)}{1-\varphi_1}$, etc.

[28] Of course, in this example, if $d_D$ is always dominated, one may not care about learning $\varphi_D$, since it does not affect the probability of any other alternative being chosen.

*such that* $\forall \nu \in [0, \bar{\nu}]$

$$L_1(x^0) \succ L_2(x^0) \succ \cdots \succ L_D(x^0),$$
$$L_{o_1}(x^1) \succ L_{o_2}(x^1) \succ \cdots \succ L_{o_D}(x^1).$$

*(2) That* $\forall j > \overline{n}$ *we have* $j' < \min_{n' \in \{\underline{n}', \ldots, \overline{n}'\}} n'$ *where* $o_{j'} = j, o_{\underline{n}'} = \underline{n}, \ldots, o_{\overline{n}'} = \overline{n}$*. Then the consideration parameters* $\{\varphi_1, \varphi_2, ..., \varphi_D\}$ *are identified.*

The additional restriction for identification under Preferred Options completion rule vis-à-vis Coin Toss is that all alternatives must dominate the preferred options at either $x^0$ or $x^1$. The second condition in Theorem 3.4 identifies the highest ranked preferred option at $x^0$, namely $d_{\underline{n}}$. Since $\varphi_{\underline{n}}$ is equal to one, for $j > \underline{n}$ the choice frequency is zero at $x^0$.[29] If, contrary to the second condition, $d_j$ is also dominated by a preferred option at $x^1$, then by the same logic its choice frequency is zero and hence the consideration parameter $\varphi_j$ is not identified. We close this section by making two remarks. First, Theorems 3.2–3.4 yield the following condition that guarantees identification under any completion rule:

**Corollary 3.1.** *If there exist* $x^0$ *and* $x^1$ *such that* $\forall \nu \in [0, \bar{\nu}]$

$$L_1(x^0) \succ L_2(x^0) \succ \cdots \succ L_D(x^0),$$
$$L_D(x^1) \succ L_{D-1}(x^1) \succ \cdots \succ L_1(x^1),$$

*then the consideration parameters* $\{\varphi_1, \varphi_2, ..., \varphi_D\}$ *are identified under all completion rules.*

Second, Theorems 3.2–3.4 are indeed only sufficient: depending on the completion rule and $x$'s, there are other conditions that yield identification. For example, the following theorem follows from the proofs of Theorems 3.2 and 3.3:

[29] This follows because $Pr(d = d_j | x^0) = \varphi_j \prod_{k<j} (1 - \varphi_k) = \varphi_j \times 0 = 0$.

**Theorem 3.5.** *Consider the Coin Toss, Default Option or Outside Option completion rule. If there exist $x^0$ and $x^1$ such that $\forall \nu \in [0, \bar{\nu}]$*

$$L_1(x^0) \succ L_2(x^0) \succ \cdots \succ L_D(x^0),$$

$$L_j(x^1) \succ L_1(x^1) \ \ \forall j \neq 1, \ \ (or \ L_D(x^1) \succ L_j(x^1) \ \ \forall j \neq D),$$

*then the consideration parameters $\{\varphi_1, \varphi_2, ..., \varphi_D\}$ are identified.*

More generally, the identification of consideration parameters comes down to the following:

**Theorem 3.6.** *Suppose there exist $\{x^0, x^1, ..., x^M\}$ and $S_j^m \subset \{1, 2, ...D\}$ such that $\forall j \in \{1, ..., D\}$ and $\forall \nu \in [0, \bar{\nu}]$*

$$L_i(x^m) \succ L_j(x^m) \succ L_k(x^m), \ \ \forall i \in S_j^m \ \ \& \ \ \forall k \in D \setminus S_j^m,$$

*then the consideration parameters $\{\varphi_1, \varphi_2, ..., \varphi_D\}$ form a system of equations in $D$ unknowns. If this system admits a unique solution, then consideration parameters are identified.*

The theorem requires that for each alternative $d_j$ there is some $x^m$ that preserves the ranking of $d_j$ relative to all other alternatives regardless of the value of the preference parameter. When this is the case, the observed choice frequency of alternative $d_j$ conditional on $x^m$ is a function of the consideration parameters, but not the preference distribution. Hence, for each $d_j$ we obtain an equation(s) in consideration parameters. If the system of these equations has a unique solution, identification follows. Finally, each set of assumptions in Theorems 3.2, 3.3, and 3.4 guarantee that the aforementioned system of equations exists and that it has a unique solution.

The identifying conditions for the RCL model are similar to those of the ARC model. The conditions are, however, weaker as the RCL model imposes the additional restriction that the consideration parameters must sum to one: $\sum_{j=1}^{D} \phi_j = 1$. For example, the following theorem is the analog of Theorem 3.2:

**Theorem 3.7.** *If there exist $x^0$ such that $\forall \nu \in [0, \bar{\nu}]$*

$$L_1(x^0) \succ L_2(x^0) \succ \cdots \succ L_D(x^0),$$

*then the consideration parameters $\{\phi_1, \phi_2, ..., \phi_D\}$ are identified.*

### 3.4.3 Identification of the preference distribution

To set the stage, it is useful to extend our example in Section 3.4.1 to the case of three alternatives: $d_1$ is the high deductible, $d_2$ is the medium deductible, and $d_3$ is the low deductible. We have that

$$\begin{aligned} Pr(d = d_1|x) =& \varphi_1\varphi_2\varphi_3 F(\min\{c_{1,2}(x), c_{1,3}(x)\}) + \varphi_1\varphi_2(1-\varphi_3)F(c_{1,2}(x)) + \\ & \varphi_1\varphi_3(1-\varphi_2)F(c_{1,3}(x)) + \varphi_1(1-\varphi_2)(1-\varphi_3) + r_1. \end{aligned} \tag{3.5}$$

The first term in the sum above captures the case where all three alternatives are considered, and hence alternative $d_1$ is chosen only if the preference coefficient is below both $c_{1,2}(x)$ and $c_{1,3}(x)$. The second term is the case that alternatives $d_1$ and $d_2$ are considered, but alternative $d_3$ is not considered, so that alternative $d_1$ is chosen only if the preference coefficient is below the cutoff between alternatives $d_1$ and $d_2$. Only alternatives $d_1$ and $d_3$ are considered in the third term, only $d_1$ is considered in the fourth term, and no alternative are considered in the last term. Note that even though the consideration parameters are point identified (and hence $\varphi_j$ can be treated as data), one moment of the data, $Pr(d = d_1|x)$, is associated with $F(\cdot)$ evaluated at two different points, $c_{1,2}(x)$ and $c_{1,3}(x)$.[30] If we had variation in $x$ that effectively shut downs one of the cutoffs (e.g. it drives $c_{1,3}(x)$ to either zero or to $\bar{\nu}$) without affecting the other cutoff, then we would restore a one-to-one mapping between a moment in the data and a value of $F(\cdot)$ at a single cutoff. In certain markets this type of variation is possible: For example, the price of the lowest deductible alternative is sufficiently large so that the alternative is strictly dominated. In the insurance context, however, it is rare to observe this type of variation, as the prices for all alternatives tend to move together. We show in Theorem 3.8 that $F(\cdot)$ is identified under much weaker conditions, that do not rely on independent variation in characteristics of single alternatives. The intuition for our result can be gleaned from Equation (3.5). Suppose we start with a value for the characteristics $\tilde{x}^0$ such that $c_{1,2}(\tilde{x}^0)$ is close to the boundary, with $c_{1,2}(\tilde{x}^0) < \bar{\nu}$ but $c_{1,3}(\tilde{x}^0) > \bar{\nu}$. Then, since $F(c_{1,3}(\tilde{x}^0)) = 1$, $Pr(d = d_1|\tilde{x}^0)$ pins down $F\left(c_{1,2}(\tilde{x}^0)\right)$. Next take $\tilde{x}^1$ such that $c_{1,3}(\tilde{x}^1) = c_{1,2}(\tilde{x}^1)$. Since $F(c_{1,3}(\tilde{x}^1))$ is known, $Pr(d = d_1|\tilde{x}^1)$ identifies $F\left(c_{1,2}(\tilde{x}^1)\right)$. Repeat these steps to construct a sequence $\{\tilde{x}^n\}_{n=1}^N$ such that $c_{1,2}(\tilde{x}^N) \leq 0$. For this approach to work, in addition to having sufficient variation in $x$ to cover the support of $\nu$, we must also require that $c_{1,3}(x)$ does not "catch up" to $c_{1,2}(x)$ (i.e. $c_{1,2}(x) < c_{1,3}(x)$ whenever $c_{1,2}(x) \in \Gamma$),[31] so that our iteration reaches the other extreme of the support.

In sum, our strategy for identifying the preference-coefficient distribution is to (1) identify the distribution of the preference coefficient close to one of the extremes of the support and then (2) move iteratively towards the other extreme. We summarize the variation in $c_{j,k}(\cdot)$ induced by $x$ that we need in the following

[30] Bringing into the analysis another moment, e.g. $Pr(d = d_2|x)$, does not help as that brings with itself evaluation of $F(\cdot)$ at another point, $c_{2,3}(x)$.

[31] That is, a DM with a preference coefficient in the interior of the parameter space cannot be indifferent between more than two alternatives.

assumption:

**Assumption 3.2** (Large Support for the Cutoff)**.** *Let* $\underline{c}_j(x) \equiv \min_{k \neq j} c_{j,k}(x)$ *and* $\bar{c}_j(x) \equiv \max_{k \neq j} c_{k,j}(x)$. *There exist* $x^0$, $x^1$, *and a continuous function* $x(t)$, $x(t) \in \mathcal{X}$, $x(0) = x^0$, $x(1) = x^1$, $t \in [0, 1]$, *and an alternative* $j$ *such that*

*A1.* $c_{j,k}(x(t))$ *exists* $\forall k$;

*A2.* $\underline{c}_j(x^0) = 0$ *and* $\underline{c}_j(x^1) = \bar{\nu}$;

*A3.* $\arg\min_{k \neq j} c_{j,k}(x(t))$ *is unique* $\forall t \in [0, 1]$.

*OR*

*B1.* $c_{k,j}(x(t))$ *exists* $\forall k$;

*B2.* $\bar{c}_j(x^0) = 0$ *and* $\bar{c}_j(x^1) = \bar{\nu}$;

*B3.* $\arg\max_{k \neq j} c_{j,k}(x(t))$ *is unique* $\forall t \in [0, 1]$.

**Theorem 3.8** (Identification under Limited Consideration)**.** *Suppose in the ARC model or in the RCL model the limited consideration parameters are identified and let Assumption 3.2 hold for some* $j$. *Furthermore, for the ARC model suppose for some* $k \neq j$ *that* $\varphi_j > 0$ *and* $\varphi_k > 0$. *For the RCL model suppose* $\phi_1 < 1$. *Then* $F(\cdot)$ *is identified.*

Theorem 3.8 relies on variation in the choice probability of one particular alternative to identify $F(\cdot)$. Hence, for $D > 2$, both ARC and RCL models are over-identified and therefore testable: in either one $Pr(d = d_m | x(t))$ is pinned down by given consideration parameters and $F(\cdot)$. If the model is correctly specified, it must then be that the predicted $Pr(d = d_m | x(t))$ coincides with the data.

## 3.5 Models' Properties

### 3.5.1 Parallels with the RUM

We focus on a *standard* application of the RUM with full consideration in the context of our example in Section 3.3.1. The final evaluation of the utility that the DM derives from alternative $j$ now includes a separately additive error term:

$$V_\nu(L_j(x)) = EU_\nu(L_j(x)) + \varepsilon_j, \tag{3.6}$$

where, as before, $\nu$ captures unobserved heterogeneity in preferences. We emphasize that in the standard RUM $\varepsilon_j$ is assumed independent of the random coefficients (in this application, the DM's risk-preference coefficient $\nu$) as well as of the observable covariates (in this application, $x = (\bar{p}, \mu)$).

Typical implementations of this model further specify that $\varepsilon_j$ is identically and independently distributed across alternatives (and DMs) with a Type 1 Extreme Value distribution, following the seminal work of (McFadden, 1974). This yields a Mixed Logit model that differs from, for example, (McFadden & Train, 2000) because in the latter the random coefficient(s) enter the utility function linearly, while in the context of expected utility models the random preference coefficient(s) enter nonlinearly. We now discuss two properties of Model (3.6) that hinder its applicability to the analysis of random expected utility models, and then illustrate how models ARC and RCL are immune from these problems.

Coupling utility functions in the hyperbolic absolute risk aversion (HARA) family, for example, CARA or CRRA, with a Type 1 Extreme Value distributed additive error, yields:

**Property 3.1** (Non-monotonicity of RUM-predicted choice probabilities in the coefficient of risk aversion)**.** *In Model* (3.6) *with $\varepsilon_j$ i.i.d. Type 1 Extreme Value, as the DM's risk aversion increases, the probability that she chooses a riskier alternative declines at first, but eventually starts to increase (Apesteguia & Ballester, 2018).*[32]

To see why, consider two non-dominated alternatives $d_j$ and $d_k$ such that $d_j$ is riskier than $d_k$. A risk neutral DM prefers $d_j$ to $d_k$ and hence will choose the former with higher probability. As the risk aversion increases, the DM eventually becomes indifferent between $d_j$ and $d_k$ and chooses either of these alternatives with equal probability (with probability equal to 0.5 when there are only two alternatives). As the risk aversion increases further, she prefers $d_k$ to $d_j$ and hence chooses the latter with lower probability. However, as the risk aversion gets even larger, the expected utility of any lottery with finite stakes converges to zero. Consequently, the choice probabilities of all alternatives, regardless of their riskiness, converge to each other, again 0.5 with two alternatives.[33] Hence, to "climb back" to 0.5, at some point the probability of choosing $d_j$ becomes increasing in risk aversion. A careful anatomy of this phenomenon reveals that it originates with the variance of the additive error term $\varepsilon_j$ being independent of $\nu$, a feature that is inescapable in Mixed Logit models.

Next, we establish the relation between utility differences across two alternatives and their respective choice probabilities. Because our random expected utility model features unobserved preference heterogeneity, we work with an analog of the

[32] See also (Wilcox, 2008).

[33] Recall that in the Mixed Logit the magnitude of the utility differences is tied to differences in (log) choice probabilities, $EU_\nu(L_k(x)) - EU_\nu(L_j(x)) = \log(Pr(d = d_k|x,\nu)) - \log(Pr(d = d_j|x,\nu))$, so that as $\nu \to \infty$ the choice probabilities are predicted to be all equal.

rank order property in (Manski, 1975) that is conditional on $\nu$:

**Definition 3.2.** *(Conditional Rank Order of Choice Probabilities) The model yields conditional rank order of the choice probabilities if for given $\nu$ and for any DM and alternatives $j, k \in \mathcal{D}$,*

$$EU_\nu(L_j(x)) > EU_\nu(L_k(x)) \Leftrightarrow Pr(d = d_j|x, \nu) > Pr(d = d_k|x, \nu).$$

The standard Mixed Logit model yields conditional rank ordering of the choice probabilities given $\nu$.[34] In turn, we show that the conditional rank order property implies the following upper bound on the probability that suboptimal alternatives are chosen:

**Property 3.2.** *(Generalized Dominance) Consider any characteristics $x$, alternative $s$, and set $J \subset \mathcal{D}\backslash\{s\}$ satisfying: $\forall \nu$, $\exists j_\nu \in J$ s.t. $EU_\nu(L_s(x)) < EU_\nu(L_{j_\nu}(x))$. Then*

$$Pr(d = d_s|x) < \sum_{k \in J} Pr(d = d_k|x).$$

Hence, in the standard Mixed Logit model, where the conditional rank order property holds, if for all preference coefficients an alternative $s$ is dominated either by alternative $j$ or by alternative $k$, then the probability of observing $s$ is predicted to be less than the sum of the probabilities of observing $j$ or $k$. We remark that neither $j$ nor $k$ is required to be optimal in $\mathcal{D}$, hence the upper bound in Property 3.2 is non-trivial.

[34](Manski, 1975) establishes the rank order property for additive error random utility models (without random coefficients) for a broader class of models that only require very weak restrictions on $\varepsilon_j$. Conditional on $\nu$, his results extend immediately to yield the conditional rank order property.

### 3.5.2 Monotonicity in Models ARC and RCL

We now formally prove that both the ARC model and the RCL model yield predicted choice probabilities that are monotone in the coefficient of risk aversion. We begin by defining monotonicity for situations in which there are more than two alternatives in the choice set.

**Property 3.3.** *(Generalized Monotone Preference Property) Consider any $x$ and suppose that $c_{j,k}(x)$ exists for all $1 \leq j < k \leq D$. Then, for any $\nu_1 < \nu_2$ and $J \in \{1, 2, ..., D\}$:*

$$Pr\left(\bigcup_{j=1}^{J} d_j \middle| x, \nu_1\right) \geq Pr\left(\bigcup_{j=1}^{J} d_j \middle| x, \nu_2\right).$$

The property above states that when alternatives are ordered so that those with lower index are more risky (that is, $c_{j,k}(x)$ exists for all $1 \leq j < k \leq D$), the probability of choosing one of the $J$ riskiest alternatives declines as the preference coefficient increases.

**Fact 3.1.** *The ARC and RCL Models satisfy the Generalized Monotone Preference Property.*

The proof of this fact (and of the ones stated in the next section) is given in Appendix B.2.

### 3.5.3 Ordinal Properties of Models ARC and RCL

In the Mixed Logit model, the cardinality of the differences in the (random) expected utility of alternatives plays a crucial role in the determination of choice probabilities, as it interacts with the realization of the additive error whose variance cannot be a function of $\nu$. We now show that both of our models can be recast as an Ordinal Random Utility Model (ORUM) in which only the ordinal and not the cardinal ranking of alternatives based on their expected utility affects DMs' choices. In contrast to the Mixed Logit, we have:

**Fact 3.2.** *The ARC and RCL Models exhibit the following type of scale invariance: any multiplication of $U_\nu(\cdot)$ by an arbitrary non-negative function of $\nu$ leaves the model's predictions unchanged.*

Hence, to turn these models into models with additive error, the errors must have a very particular structure.

**Fact 3.3.** *(ARC Model as ORUM) The ARC Model is equivalent to an additive error random utility model with unobserved preference heterogeneity where all alternatives are considered, the DM's utility associated with each alternative $j \in \{1, ..., D\}$ is given by*

$$V_\nu(d_j, x) = U_\nu(d_j, x) + \varepsilon_j,$$

*and $\varepsilon_j$ is a random variable such that:*

$$\varepsilon_j = \begin{cases} 0 & \textit{with probability } \varphi_j \\ -\infty & \textit{with probability } (1 - \varphi_j). \end{cases}$$

*The error terms are independent of $(x, \nu)$ and across alternatives. Ties, in case $\varepsilon_j$ takes on $-\infty$ value $\forall j$, are broken according to the completion rule as specified in Section 3.3.3.*

**Fact 3.4.** *(RCL Model as ORUM) The RCL Model is equivalent to an additive error random utility model with unobserved preference heterogeneity where* *all alternatives are considered, the DM's utility associated with each alternative* $j \in \{1, ..., D\}$ *is given by*

$$V_\nu(d_j, x) = U_\nu(d_j, x) + \varepsilon_j,$$

*and* $\varepsilon_j$ *is a random variable that takes two values:* $0$ *and* $-\infty$*. The joint distribution of* $\varepsilon$=$(\varepsilon_1, \varepsilon_2, ..., \varepsilon_D)$ *is as follows. For every realization* $e$ *that has at least one zero element:*

$$p(e) = \frac{\phi_l}{\binom{D}{l}}, \text{ where } l = \sum_k \mathbb{1}(e_k = 0).$$

*and for* $e = \{-\infty, -\infty, \ldots, -\infty\}$*:* $p(e) = 0$*.*

The structure of the additive errors derived in Fact 3.3 and 3.4, respectively, allow us to learn which of these models satisfy the conditional rank order property, and hence the Generalized Dominance Property.

**Fact 3.5.** *The ARC Model does not (always) satisfy the Conditional Rank Order Property and, hence, the Generalized Dominance.*

**Fact 3.6.** *The RCL Model satisfies the Conditional Rank Order Property and, hence, Generalized Dominance.*

We summarize this section with Table 3.1, that lists the differences across the Mixed Logit, ARC, and RCL models. The first panel lists the differences in the assumptions and the second panel lists the differences in implied properties.

## 3.6 Beyond the ARC and RCL Models

As shown above, the ARC and RCL models are complimentary in terms of consideration set formation if viewed through the lens of an error structure corresponding

**Table 3.1** Model Comparisons

| | Mixed Logit | ARC | RCL |
|---|---|---|---|
| **Error Distribution** | | | |
| Support | $\mathbb{R}$ | $\{-\infty, 0\}$ | $\{-\infty, 0\}$ |
| Independent of $x$ | Yes | Yes | Yes |
| Independent of $\nu$ | Yes | Yes | Yes |
| Independent across alternatives | Yes | Yes | No |
| Identical across alternatives | Yes | No | Yes |
| **Properties** | | | |
| Monotonicity | No | Yes | Yes |
| Conditional Rank Order Property | Yes | No | Yes |
| Generalized Dominance | Yes | No | Yes |

to an ORUM. Each model relaxes an assumption imposed on the error structure (either independence or identically distributed). Yet identification of these models rests on a similar logic. We conjecture that other consideration models that relax both the independence and identically distributed assumptions can also be identified. While it is beyond the scope of this paper to argue which consideration set formation is the right one in a given context, we offer two additional examples of consideration set formation based on well established economic/behavioral phenomena. The first example captures economic situations in which a DM will consider alternatives with an attribute that is below a certain DM specific threshold (Kimya, 2018). Within the insurance context, a threshold on the deductible level attribute will arise immediately if there are DM specific (unobserved) liquidity constraints: Anticipating that her liquidity constraint might bind if a loss occurs, a forward looking DM discards high deductible alternatives from the consideration set.[35] The second mechanism builds on the notion of extremeness aversion, one

[35]For a discussion on how liquidity constraints affect households' risk aversion see (Chetty & Szeidl, 2007).

way Behavioral Economics and Marketing literature addresses the context dependency of preferences. (Simonson & Tversky, 1992) define extremeness aversion as the situation when "the attractiveness of an option is enhanced if it is an intermediate option in the choice set and is diminished if it is an extreme option". In our framework, the relative location of the alternative in the choice set will determine its likelihood of being considered.

### 3.6.1 The Threshold Model

We return to our example in Section 3.3.1. Suppose that DMs find it prohibitively costly to have out-of pocket expenses above certain DM-specific limit. Hence there is an upper bound on what deductibles they consider. Formally, we assume that, in addition to the preference parameter $\nu$, each DM has an unobserved threshold parameter $\underline{d}$. The DM's draw of the threshold parameter defines her consideration set. In particular, alternative $d_k$ is considered if and only if $d_k < \underline{d}$.[36] We assume that $\nu$ and $\underline{d}$ are independent conditional on observables (e.g., wealth, credit score, and age), and that the threshold parameter is continuously distributed $G(d)$ with support $(d_D, \infty)$.

Let $\xi_1 = 1 - G(d_1), \xi_2 = G(d_1) - G(d_2), \dots, \xi_{D-2} = G(d_{D-3}) - G(d_{D-2})$, and $\xi_{D-1} = G(d_{D-2})$. Then the fraction of DMs considering all deductibles is $\xi_1$, considering all but the highest deductible is $\xi_2$, and considering only the lowest deductible is $\xi_D$. $F(\nu)$ and $\{\xi_1, \xi_2, ..., \xi_{D-1}\}$ are identified provided sufficient variation in $p$ and/or $\mu$, as discussed above, exists. Indeed, consider DMs with $x^0 = (p^0, \mu^0)$ such that the deductibles are ranked from the highest to the lowest for all $\nu$ on

[36] Recall that deductibles are decreasing in $k$, and hence $d_D$ is the smallest deductible providing maximum coverage.

the support. Then $\xi_k$ is equal to $Pr(d = d_k|x^0)$, which is identified by the data. Identification of $F(\nu)$ follows from similar arguments made in Theorem 3.8.

### 3.6.2 Extremeness Aversion Model

Consider again example in Section 3.3.1. To ease notation we assume that the number of alternatives in the choice set, $D$, is odd and we let $m = \frac{D+1}{2}$. All DMs consider the median alternative $d_m$ with probability equal to one. The remaining alternatives are considered with probability that is decreasing in the distance from the median alternative. In particular, the consideration set is formed according to the following product rule:

$$Pr(d_k \text{ is considered}) = \begin{cases} \prod_{j=m+1}^{k} \xi_j & \text{if } k > m \\ \prod_{j=k}^{m-1} \xi_j & \text{if } k < m \end{cases}.$$

As before, consideration set formation is independent $\nu$ conditional on observables. Suppose that there is sufficient variation in $p$ and/or $\mu$ so that there exists DMs with $x^0 = (p^0, \mu^0)$ such that the deductibles are ranked from the highest to the lowest for all $\nu$ on the support. Then $\xi_k$ is equal to the ratio $\frac{Pr(d_k|x^0)}{Pr(d_{k-1}|x^0)}$, which is identified by the data.

## 3.7 Application

### 3.7.1 Data

We study households' deductible choices across three lines of property insurance: auto collision, auto comprehensive, and home all perils. The data come from a U.S.

**Table 3.2** Premiums Quantiles for the \$500 Deductible

| Quantiles | 0.01 | 0.05 | 0.25 | 0.50 | 0.75 | 0.95 | 0.99 |
|---|---|---|---|---|---|---|---|
| Collision | 53 | 74 | 117 | 162 | 227 | 383 | 565 |
| Comprehensive | 29 | 41 | 69 | 99 | 141 | 242 | 427 |
| Home | 211 | 305 | 420 | 540 | 743 | 1,449 | 2,524 |

insurance company. Our analysis uses a sample of 7,736 households who purchased their auto and home policies for the first time between 2003 and 2007 and within six months of each other. We only consider their *first* purchases.[37] Table B.1 provides descriptive statistics for households' observable characteristics, which we use later to estimate households' preference coefficients.[38] For each household and each coverage we observe the exact menu of alternatives available at the time of the purchase. The deductible alternatives vary across coverages but not across households. Table B.2 presents the frequency of chosen deductibles in our data.

Premiums are determined coverage-by-coverage as in the example from Section 3.3.1. For each household, the company determines a baseline price $\bar{p}$ using a coverage-specific rating function, which takes into account the household's coverage-relevant characteristics and any applicable discounts. Given $\bar{p}$, the premium for alternative $j$ is determined based on a coverage-specific rule, $p_j = g_j \cdot \bar{p} + \delta$. Table B.5 reports the average premium by context and deductible, and Table 3.2 summarizes the premium distributions for the \$500 deductible. As the latter table shows, premiums vary dramatically. In each coverage, the 99th percentile of the

[37] The dataset is an updated version of the one used in (Barseghyan et al., 2013). It contains information for an additional year of data and puts stricter restrictions on the timing of purchases across different lines. These restrictions are meant to minimize potential biases stemming from non-active choices, such as policy renewals, and temporal changes in socioeconomic conditions.

[38] These are the same variables that are used in (Barseghyan et al., 2013) to control for households' characteristics. See discussion there for additional details.

\$500 deductible is more than ten times the corresponding 1st percentile.

The underlying loss probabilities are derived from expected claim rates that are estimated using coverage-by-coverage Poisson-Gamma Bayesian credibility models applied to a large auxiliary panel. The unbalanced panel contains over 400,000 households from 1998 to 2007, yielding more than 1.3 million household-year observations for each coverage. We assume that household $i$'s claims under coverage $j$ in year $t$ follow a Poisson distribution with arrival rate $\lambda_{ijt}$. We treat $\lambda_{ijt}$ as latent random variables and assume that $\ln \lambda_{ijt} = \mathbf{W}'_{ijt}\zeta_j + \epsilon_{ij}$, where $\mathbf{W}_{ijt}$ is a vector of observables, $\epsilon_{ij}$ is an unobserved i.i.d. error term, and $\exp(\epsilon_{ij})$ follows a Gamma distribution with unit mean and variance $\eta_j$.[39] Poisson panel regressions with random effects yield estimates of $\zeta_j$ and $\eta_j$ for each coverage $j$. For each household $i$, we use the regression estimates to generate a predicted claim rate $\widehat{\lambda}_{ij}$ for each coverage $j$, conditional on the household's *ex ante* characteristics $\mathbf{W}_{ij}$ and *ex post* claims experience. In the model, we assume that households expect no more than one claim.[40] Hence, we transform $\widehat{\lambda}_{ij}$ into a predicted claim probability $\widehat{\mu}_{ij} = 1 - \exp(-\widehat{\lambda}_{ij})$. Predicted claim probabilities (summarized in Table 3.3) exhibit extreme variation: The 99th percentile claim probability in collision (comprehensive and home) is 4.3 (12 and 7.6) times higher than the corresponding 1st percentile. Finally, the correlation between claim probabilities and premiums for the \$500 deductible is 0.38 for collision, 0.15 for comprehensive, and 0.11 for

[39]We refer to this model as a Bayesian credibity model because $\hat{\lambda}$ corresponds to the Bayesian credibility premium in the actuarial literature (Denuit, Maréchal, Pitrebois, & Walhin, 2007, Ch 3).

[40]The claim rates are small and, consequently, the likelihood of two or more claims is small. For home insurance 86.2% of predicted claim rates in the core sample are less than 0.1 and 97.4% percent are less than 0.15. For collision the frequencies are 79.8% and 98.6%, respectively. For comprehensive – 99.95% and 100%.

**Table 3.3** Claim Probabilities Across Contexts

| Quantiles | 0.01 | 0.05 | 0.25 | 0.50 | 0.75 | 0.95 | 0.99 |
|---|---|---|---|---|---|---|---|
| Collision | 0.036 | 0.045 | 0.062 | 0.077 | 0.096 | 0.128 | 0.156 |
| Comprehensive | 0.005 | 0.008 | 0.014 | 0.021 | 0.030 | 0.045 | 0.062 |
| Home | 0.024 | 0.032 | 0.048 | 0.064 | 0.084 | 0.130 | 0.183 |

home all perils. Hence there is independent variation in both.[41]

### 3.7.2 The Model

The model is identical to the one in Section 3.3.1, augmented with either the ARC or the RCL Model. As in the example, the DM's problem amounts to choice over deductible lotteries of the form $L_k(x) \equiv (-p_k, 1-\mu; -p_k - d_k, \mu)$, where $x = (\bar{p}, \mu)$. The utility function is assumed to be CARA. For all $\nu > 0$:

$$\begin{aligned} EU_\nu(L_k(x)) &= -\left(1-\mu\right) e^{-\nu(w-p_k)} - \mu e^{-\nu(w-p_k-d_k)} \\ &= -e^{-\nu w}\left[\left(1-\mu\right) e^{\nu p_k} + \mu e^{\nu(p_k+d_k)}\right], \end{aligned}$$

where $w$ denotes the DM's initial wealth.[42] Note that $e^{-\nu w}$ enters multiplicatively in the expression above and hence does not affect the rankings of the alternatives.

We now establish the assumptions required for the identification results in Corollary 3.1 and Theorem 3.8.[43] It is immediate to see that for any $\nu$ on a compact support, when $\bar{p}$ is sufficiently large and/or $\mu$ is close to zero, the preference ordering is sequential: $EU_\nu(L_1) > EU_\nu(L_2) > \cdots > EU_\nu(L_D)$. Alternatively, when

[41]See (Barseghyan et al., 2013) (and (Cohen & Einav, 2007) in the context of Israeli auto insurance) for a detailed discussion of where such independent variation comes from.

[42]When $\nu$ is zero, expected utility is simply $-p_k - \mu d_k$.

[43]In Appendix B.3, we show that these assumptions are satisfied also under CRRA.

$\bar{p}$ is small (and $\mu > 0$) or $\mu$ is close to one (and $\bar{p}$ is not very large) we have that $EU_\nu(L_D) > EU_\nu(L_{D-1}) > \cdots > EU_\nu(L_1)$.

Turning to the identification of the preference-coefficient distribution, note that all cutoffs exist and are continuous functions of $x = (\bar{p}, \mu)$. It remains to show that $\arg\min_{k \neq 1} c_{1,k}(x)$ is unique. To establish this, we show in Appendix B.3 that $c_{1,2}(x) < c_{1,m}(x)$ for any $m > 2$.

### 3.7.3 Estimation Results

#### The ARC Model: Collision

We start by presenting estimation results in a simple setting where the only choice is the collision deductible and observable demographics do not affect preferences. We do so to illustrate the key features of our method and to ascertain that multiple contexts and demographic variables play no particular role in identification.

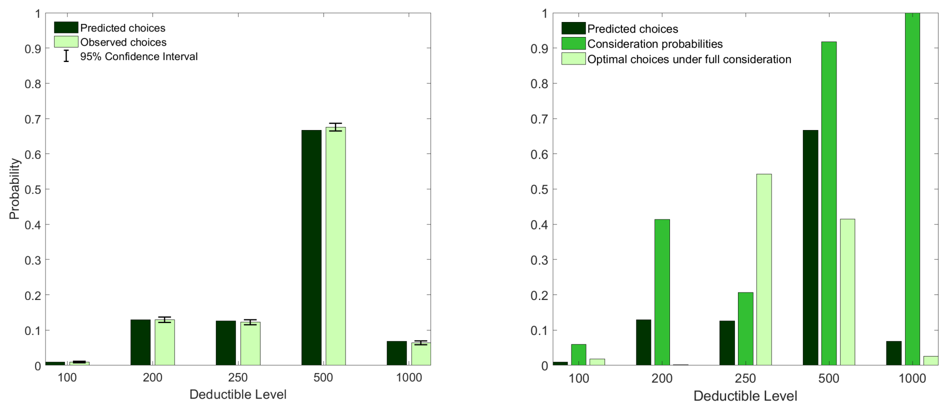
In this market there are no preset defaults for deductibles, which implies that Default Option is not a proper completion rule for these data. We assume the Coin Toss completion rule. The estimation under Coin Toss naturally encompasses Preferred Options – if estimated values of one or more $\varphi_j$'s turn out to be one, then we have Preferred Options.[44] To execute our estimation procedure we need to choose the upper bound of the preference-coefficient support. We set it to $\bar{\nu} = 0.02$, which is conservative see Barseghyan, Molinari, and Teitelbaum (2016). We *ex post* verify that this does not affect our estimation by checking that the density of

[44] We could have also assumed Outside Default (under the second interpretation). The collision only results under this completion rule are nearly identical to those presented in the paper.

the estimated distribution is close to zero at the upper bound. We approximate $F(\cdot)$ non-parametrically through a mixture of Beta distributions. In practice, however, both AIC/BIC criteria indicate that a single component is sufficient for our analysis.

The estimated distribution and consideration parameters are reported in Table B.6. As the first panel in Figure 3.1 shows, the model closely matches the aggregate moments observed in the data. The second panel in Figure 3.1 illustrates side-by-side the frequency of predicted choices, consideration probabilities, and the distribution of households' first-best alternative (i.e., the distribution of optimal choices under full consideration). Predicted choices are determined jointly by the preference induced ranking of deductibles and by the consideration probabilities: Limited consideration forces households' decision towards less desirable outcomes by stochastically eliminating better alternatives. It is noteworthy that the two highest deductibles ($\$1,000$ and $500) are considered at much higher frequency (1.00 and 0.92, respectively) than the other alternatives, suggesting that households have a tendency to regularly pay attention to the cheaper items in the choice set. Yet, the most frequent model-implied optimal choice under full consideration is the $250 deductible, which is considered with relatively low probability. In this application, assuming full consideration leads to a significant downward bias in the estimation of the underlying risk preferences. To see why, consider increasing the consideration probabilities for the lower deductibles to the same levels as the $500 deductible. Holding risk preferences fixed, the likelihood that the lower deductibles are chosen increases and therefore the higher deductibles are chosen with lower probability. Average risk aversion must decline to compensate for this shift and to "push back up" the likelihood function. This is exactly the pattern we find when we estimate a near-full consideration model. In particular, we find

**Figure 3.1:** The ARC Model




The first panel reports the distribution of predicted and observed choices. The second panel displays consideration probabilities and the distribution of optimal choices under full consideration.

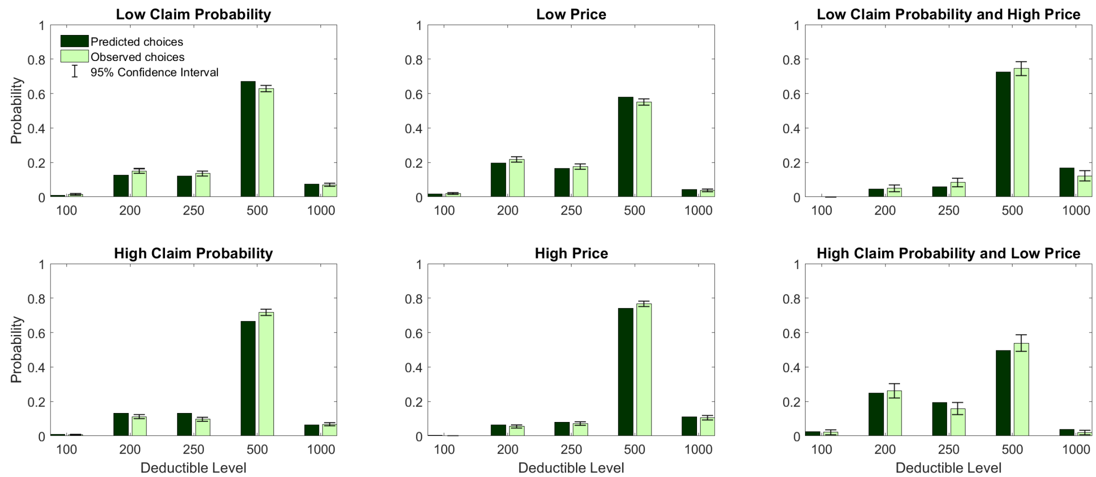
that average risk aversion decreases by about 34% from 0.0036 to 0.0024 when all consideration parameters equal 0.999.[45] To put these numbers into context, a DM with risk aversion equal to 0.0037 is willing to pay \$424 to avoid a $\$1,000$ loss with probability 0.1, while a DM with risk aversion equal to 0.0027 is only willing to pay \$287 to avoid the loss.

The model's ability to match data extends also to conditional moments. The first two panels of Figure 3.2 show observed and predicted choices for the fraction of households facing low and high premiums, respectively, and the next two panels are for households facing low and high claim probabilities.[46] Finally, the last two panels capture households who face both low claim probabilities and high prices and *vice versa*. It is transparent from Figure 3.2 that the model matches closely the observed frequency of choices across different subgroups of DMs facing a variety

[45]We cannot assume that all consideration probabilities are equal to one, since the \$200 deductible is dominated under full consideration and is chosen with positive probability.

[46]Low/high groups here are defined as households whose claim rate (or baseline price) are in the bottom/top third of the distribution.

**Figure 3.2:** The ARC Model: Conditional Distributions




of prices and claim probabilities, even though some of these frequencies are quite different from the aggregate ones.

The ARC model's ability to violate Generalized Dominance is key in matching the data. In our dataset, the \$200 collision deductible is always dominated either by the \$100 deducible or the \$250 deductible. This happens because of the particular pricing schedule in collision. It costs the same to get an additional \$50 of coverage by lowering the deductible from \$250 to \$200 as it does to get an additional \$100 of coverage by lowering the deductible from \$200 to \$100. If a household's risk aversion is sufficiently small, then it prefers the \$250 deducible to the \$200 deducible. If, on the other hand, the household's level of risk aversion is such that it would prefer the \$200 deducible to the \$250 deductible, then it would also prefer getting twice the coverage for the same increase in the premium. That is, for any level of risk aversion, the \$200 deducible is dominated either by the \$100 deducible or by the \$250 deducible.[47] Yet, overall the \$200 deductible is chosen roughly as often as the \$100 and \$250 deductibles combined. More so, for certain

[47]This pattern is at odds not only with EUT but also many non-EU models (Barseghyan et al., 2016).

sub-groups the \$200 deductible is chosen *much more often* than the \$100 and \$250 deducible combined.[48] It follows that a model satisfying Generalized Dominance cannot rationalize these choices.

In the next step of our estimation analysis we relax the assumption that demographic variables do not influence risk preferences. While it is ideal to control for households' observable characteristics non-parametrically, it is data demanding. In practice, it is commonly assumed that household characteristics shift the expected value of the preference-coefficient distribution.[49] We adopt the same strategy here by assuming that for each household $i$, $\log \frac{\beta_{1,i}}{\beta_2} = \mathbf{Z}_i\gamma$, where $\gamma$ is an unknown vector to be estimated. The terms $\beta_{1,i}$ and $\beta_2$ denote the parameters of the Beta distribution, where $\beta_{1,i}$ is household specific and $\beta_2$ is common across households. The preference coefficients are random draws from a distribution with an expected value that is a function of the observable characteristics given by $E(\nu_i) = \frac{\beta_{1,i}}{\beta_{1,i}+\beta_2}\bar{\nu} = \frac{e^{\mathbf{Z}_i\gamma}}{1+e^{\mathbf{Z}_i\gamma}}\bar{\nu}$.[50] The results of this estimation are in line with our first estimation. (See Column 2 in Table B.6, as well as Figures B.1 and B.2 in Appendix B.6.) The new observation here is that the model closely matches the distribution of choices across various sub-populations in the sample including gender, age, credit worthiness, and contracts with multiple drivers. The model's ability to match these conditional distributions can be attributed, in part, to the dependence of risk preferences on household characteristics. The model is, however, fairly par-

[48]For example, the fraction of households facing low $\bar{p}$ and high $\mu$ that choose the \$200 deductible is 0.26, while the fraction that choose the \$100 deductibles or the \$250 deductibles is 0.18.

[49]For exmaple, (Cohen & Einav, 2007) assume that $\log \nu_i = \mathbf{Z}_i\gamma + \varepsilon_i$, where $\mathbf{Z}_i$ are the observables for household $i$ and $\varepsilon_i$ is i.i.d. N(0, $\sigma^2$). Hence, $E(\nu_i) = e^{\mathbf{Z}_i\gamma+\sigma^2/2}$.

[50]If, instead, we assume $\log \frac{\beta_{2,i}}{\beta_1} = \mathbf{Z}_i\tilde{\gamma}$, then we arrive to the same expression for the expected value with the exception that $\tilde{\gamma} = -\gamma$.

simonious as the consideration parameters are restricted to be the same across all households. Finally, estimated consideration probabilities are close in magnitude to those estimated above. In particular, the highest deductibles ($\$1,000$ and $\$500$) are most likely to be considered, with respective frequencies of 0.95 and 0.91. The remaining alternatives are considered at much lower frequencies.

### The RCL Model and the RUM

For completeness, we now discuss estimation results for the RCL Model and the RUM with unobserved heterogeneity. In both cases, we assume that risk-preference coefficient is Beta distributed with support $[0, \bar{\nu}]$, where, as before, $\bar{\nu} = 0.02$. *A priori*, neither of these models should do well in matching the distribution of observed choices. Both of them satisfy the Conditional Rank Order Property and have no ability to direct households' choices in a particular direction. Instead, they smoothly spread households' choices around their respective first bests: the closer the expected utility of a given alternative is to the expected utility of the first best, the higher the frequency at which it will be chosen.

Consequently, these models cannot match the observed distribution and, in particular, are unable to explain the relatively high observed share of the $200 deductible. Table B.7 reports the estimation results for the RCL model and the RUM. Figure B.3 compares the observed distribution of choices and the predicted ones under both models. The predicted distributions are similar to each other, but are a much poorer fit to the data than that of the ARC Model.

To formally assess how well these models fit the data relative to the ARC, we rely on the Vuong test. The latter takes into account both the fact that the models are not nested and that they can have different number of parameters. The test soundly (at 1% level) rejects both the RCL and the RUM in favor of the ARC Model.

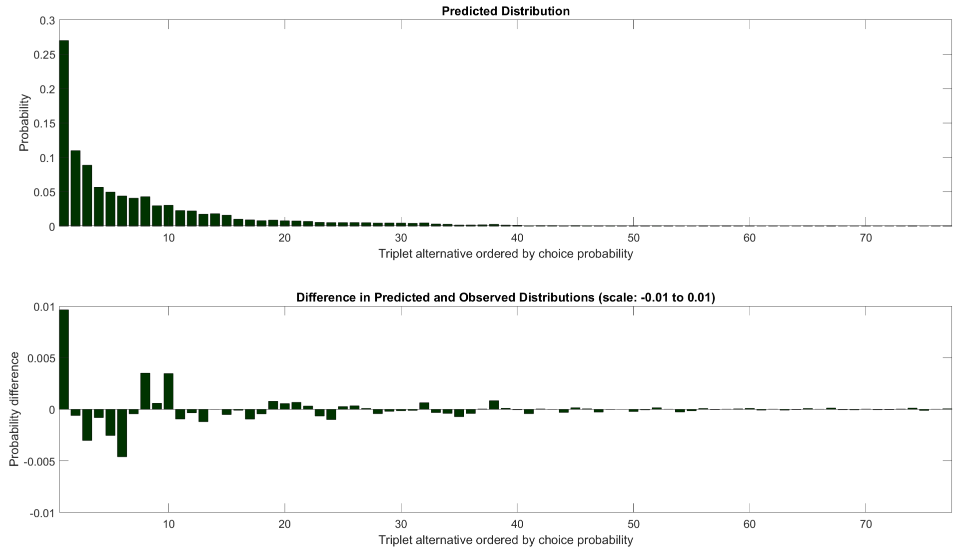
**The ARC Model: All Coverages**

We now proceed with estimation of the full model. We consider two cases. In the first case households' risk preferences are invariant across coverages, but consideration sets form independently within each coverage. There are three sets of consideration parameters $\{\varphi^{coll}, \varphi^{comp}, \varphi^{home}\}$ and the probability that alternative $k$ is considered in one coverage (e.g. collision) is independent of the probability that alternative $j$ is considered in another coverage (comprehensive or home). Hence, within each coverage, the households' problem is identical to that from the previous section.[51] The estimation results are presented in Figure B.4 and Table B.8. Just as in the case of collision coverage only, the model matches well the choice distributions within each coverage. However, the independence of consideration sets across coverages implies that the model does not have the ability to match the joint distribution of choices. For example, the model predicts zero rank correlation across the deductibles and that 12% of households choose an alternative with a larger comprehensive deductible than collision deductible. In the data the rank correlation ranges from 0.35 to 0.61 and only 0.2% of households choose a larger comprehensive deductible.

We next assume that households' consideration sets are formed over the entire

[51] Effectively this scenario amounts to assuming "narrow bracketing" (Read, Loewenstein, & Rabin, 1999), a common approach in the literature. Note that under full consideration there is no loss of generality in assuming narrow bracketing. As it is well known, with CARA preferences the decision in one context is independent of the decisions in other contexts as long as loss events are mutually exclusive.

**Figure 3.3:** The ARC Model, Three Coverages




Triplets are sorted by observed frequency at which they are chosen. The first panel reports the predicted choice frequency and the second panel reports the difference in predicted and observed choice frequencies.

deductible portfolio. There are 120 possible alternative triplets $(d^{coll}, d^{comp}, d^{home})$, each having its own probability of being considered. This model is flexible as it nests many rule of thumb assumptions such as only considering contracts with the same deductible level across the three contexts or only considering contracts with a larger collision deductible than comprehensive deductible. Figure 3.3 and Table B.9 present estimation results. The first panel of the figure shows the predicted distribution of choices across triplets, ranked in descending order by observed frequencies. The second panel plots the differences between predicted and observed choice distributions. Clearly, the predicted distribution is close to the observed distribution.

The largest difference between the predicted and observed distributions is equal to 0.96 percentage points, which occurs at the $(\$500, \$500, \$500)$ triplet that is chosen by 26% of the households. The *integrated* absolute error across all triplets is 4.61%. We note that in our data 43 out of 120 triplets are never chosen (these are omitted from Figure 3.3). It is straightforward to show analytically that likelihood maximization implies that the consideration probabilities for these triplets must be zero, so that their predicted shares are *de facto* zero.[52] Consequently, the likelihood maximization routine is faster and more reliable as we do not need to search for $\varphi_j$ for these alternatives.

Another virtue of the ARC Model is that it effortlessly reconciles two sides of the debate on stability of risk preferences (Barseghyan, Prince, & Teitelbaum, 2011; Einav et al., 2012; Barseghyan et al., 2016). On the one hand, households' risk aversion relative to their peers is correlated across contexts, implying that households preferences have a stable component. On the other hand, analyses based on revealed preference reject the standard models of risk aversion: under full consideration, for the vast majority of households one cannot find a (household-specific) risk aversion parameter that can justify their choices simultaneously across all contexts. Relaxing full consideration allows to match the observed joint distribution of choices, and hence their rank correlations.

Estimated risk preferences are similar to those estimated with collision only data, although the variance is slightly smaller. Turning to consideration, the triplet considered far more frequently than any other alternative is the cheapest one: $(\$1,000, \$1,000, \$1,000)$.[53] Its consideration probability is 0.81, while the next two most considered triplets are ($500, $500, $1,000) and ($500, $500, $500). These are

[52]Since we are estimating the model with the Coin Toss completion rule, these options still can be chosen if the consideration set is empty and $\varphi_j < 1$ for all $j$. In our estimation, the probability that the consideration set is empty is 0.0015, which implies that an alternative with zero consideration probability is chosen with probability 0.0015/120=0.000013.

[53]The first entry is for collision, the second is for comprehensive, and the third is for home.

considered with probability 0.47 and 0.43, respectively. Overall, there is a strong positive correlation (0.54) between the consideration probability and sum of the deductibles in a given alternative. We summarize once more the computational advantages of our procedure. First, estimation of our model remains feasible for a large choice set, since our likelihood calculation does not require summation of probabilities over all possible consideration sets containing each household's choice.[54] Second, the model's parameters grow linearly with the choice set – one parameter per an additional alternative – which keeps the computations in check. Third, enlarging the choice set does not call for new independent sources of data variation. For example, in our model whether there are five deductible alternatives or hundred twenty would not make any difference neither from an identification nor an estimation stand point: with sufficient variation in $\bar{p}$ and/or $\mu$ the model is identified and can be estimated.

As a final remark, once the model is estimated, one can compute the monetary cost of limited consideration. In our data, it is \$49 (see Appendix B.4).

## 3.8 Beyond Expected Utility

Our framework can also be applied to non-EU models as long as risk attitudes are determined by a unidimensional index. Consider, e.g., the probability distortions

[54]This is contrast to Goeree's ((2008)) method, which utilizes the logit structure and hence must keep track of all consideration sets containing the household's choice. In our setting, it is feasible to estimate an additive error RUM assuming the DMs consider each deductible triplet as a separate alternative (Figure B.5 and Table B.10). As the figure shows, the failure to match data is evident. The Vuong test formally rejects it in favor of the ARC model.

model in (Barseghyan et al., 2013).[55] Under this model, the expected utility of a lottery is evaluated using a distorted claim probability $\Omega(\mu)$ instead of $\mu$:

$$EU(L_j(x)) = (1 - \Omega(\mu))\, u\,(w - p_j) + \Omega(\mu) u\,(w - p_j - d_j)\,.$$

Let $u(\cdot)$ be linear across all DMs, and hence $\Omega(\cdot)$ is the only source of risk aversion. Assume that for a given $\mu$, $\Omega(\mu)$ is randomly distributed across DMs with support $[\underline{\Omega}, \overline{\Omega}] \subset [0, 1]$. Since the SCP is trivially satisfied, identification of both the ARC and the RCL models follows under the same conditions as in Section 3.4. If, however, $u(\cdot)$ is concave and varies across DMs, then there are two distinct sources of aversion to risk, $\nu$ and $\Omega(\mu)$. While parametric identification of the joint distribution of $\{\nu, \Omega(\cdot)\}$ under full consideration is straightforward, non-parametric identification is an open question (Barseghyan et al., 2018). If conditions can be derived for non-parametric identification under full consideration, then our identification strategy may be used to obtain identification under limited consideration as well.

## 3.9 Conclusion

In this paper we built a framework where DMs consider only a subset of the available alternatives. We offered two models with different consideration formation mechanisms and established their identification. There are many ways – such as liquidity constraints or extremeness aversion that we discussed – in which limited consideration may arise. While much effort in applied theory has been towards constructing non-EU models that can generate rankings of alternatives that are

[55] In the context of binary lotteries this model incorporates many leading alternatives to EUT. See (Barseghyan et al., 2018) for a discussion.

different from those in EUT, a promising and complementary avenue is to build and test theories that allow for limited consideration in the decision making process. There are many open questions. First, what economic forces determine the formation of consideration sets and how does consideration change with the market setting? Second, once we allow for limited consideration, how do our conclusions about the underlying models of risk change? Would we still need non-EU models to explain DMs' behavior in real market situations, as it is commonly argued in the literature, and if yes which ones?

APPENDIX A

# CHAPTER 1 APPENDIX

## A.1 Outer Region Computation

This section details how to construct the outer identified region for the network-formation parameter, $\mathcal{O}_P[\theta]$, and the joint outer identified region for the target network statistic $\mathcal{O}_P[\beta_\theta, \theta]$. I detail how to construct equivalence classes and also how to simulate the model-implied probability a subnetwork forms.

### A.1.1 Network Formation Moment Inequalities

Recall that the moment inequalities for the network-formation model that I work with are:

$$m_1(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}}; \theta) \tag{A.1.1}$$
$$\equiv -\boldsymbol{P}((\boldsymbol{A}^{\boldsymbol{s}}, \boldsymbol{X}^{\boldsymbol{s}}) \in \mathcal{C}(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}})) + \sum_{\mathcal{C}(\boldsymbol{x}^{-\boldsymbol{s}})} \int_{\varepsilon \in \mathcal{E}_u(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})} dF_\varepsilon(\varepsilon) \boldsymbol{P}(\boldsymbol{X}^{-\boldsymbol{s}} \in \mathcal{C}(\boldsymbol{x}^{-\boldsymbol{s}}))$$
$$m_2(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}}; \theta)$$
$$\equiv \boldsymbol{P}((\boldsymbol{A}^{\boldsymbol{s}}, \boldsymbol{X}) \in \mathcal{C}(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}})) - \sum_{\mathcal{C}(\boldsymbol{x}^{-\boldsymbol{s}})} \int_{\varepsilon \in \mathcal{E}_a(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})} dF_\varepsilon(\varepsilon) \boldsymbol{P}(\boldsymbol{X}^{-\boldsymbol{s}} \in \mathcal{C}(\boldsymbol{x}^{-\boldsymbol{s}})).$$

The first term in Equation (A.1.1) is estimated from data:

$$\hat{\boldsymbol{P}}(\boldsymbol{A}^{\boldsymbol{s}} \in \mathcal{C}(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}})) = \frac{1}{T} \sum_{t=1}^{T} \mathbb{1}(A^{\boldsymbol{s}} \sim A_t^{\boldsymbol{s}}, \boldsymbol{x} \sim \boldsymbol{x}_t^{\boldsymbol{s}}),$$

where $(A^{\boldsymbol{s}} \sim A_t^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}} \sim \boldsymbol{x}_t^{\boldsymbol{s}})$ denotes an isomorphism. That is, there exists a permutation of $\boldsymbol{s}$, say $\tau(\boldsymbol{s})$, such that $A^{\boldsymbol{s}} = G_t^{\tau(\boldsymbol{s})}$ and $\boldsymbol{x}^{\boldsymbol{s}} = \boldsymbol{x}_t^{\tau(\boldsymbol{s})}$. The first computational problem is checking for isomorphisms, which is expensive. I elect to enumerate all possible values of $(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}})$ and construct all equivalence classes outside of the main optimization program. This is feasible provided that I restrict $|\boldsymbol{s}|$ to less than or equal to six. I use the `isisomorphic` function from the Graph and Network Algorithm package in MatLab's base toolbox to check whether two colored networks $(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}})$ and $(\tilde{A}^{\boldsymbol{s}}, \tilde{\boldsymbol{x}^{\boldsymbol{s}}})$ are isomorphic. I compare the realized value of $(A_t^{\boldsymbol{s}}, \boldsymbol{x}_t^{\boldsymbol{s}})$ to the enumerated list of colored networks to obtain its equivalence class number. Last, $\hat{\boldsymbol{P}}(\boldsymbol{A}^{\boldsymbol{s}} \in \mathcal{C}(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}}))$ is computed by summing over all realized networks belonging to class $\mathcal{C}(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}})$.[1]

The second term in Equation (A.1.1) is computed by simulation and from data. The term $\boldsymbol{P}(\boldsymbol{X}^{\boldsymbol{s}} \in \mathcal{C}(\boldsymbol{x}))$ is estimated from the data. I maintain the assumption that $\boldsymbol{x}_i$ is independent of $\boldsymbol{x}_j$, hence $\boldsymbol{P}(\boldsymbol{X}^{-\boldsymbol{s}} = \boldsymbol{x}^{-\boldsymbol{s}}) = \prod_{i \in \boldsymbol{n}-\boldsymbol{s}} \boldsymbol{P}(\boldsymbol{x}_i)$. In the case of binary data $\mathcal{C}(\boldsymbol{x}^{-\boldsymbol{s}})$ is characterized by the number of instances that $\boldsymbol{x}_i = 1$. Let $n_1(\boldsymbol{x}^{-\boldsymbol{s}}) = \sum_{i \in \boldsymbol{n}-\boldsymbol{s}} \mathbb{1}(\boldsymbol{x}_i = 1)$. Then

$$\boldsymbol{P}(\boldsymbol{X}^{-\boldsymbol{s}} \in \mathcal{C}(\boldsymbol{x}^{-\boldsymbol{s}})) = p^{n_1(\boldsymbol{x}^{-\boldsymbol{s}})}(1-p)^{n-|\boldsymbol{s}|-n_1(\boldsymbol{x}^{-\boldsymbol{s}})},$$

where $p$ is the probability that $\boldsymbol{x}_i = 1$. The sum over equivalence classes is then equivalent to summing over all possible values for $n_1(\boldsymbol{x}^{-\boldsymbol{s}})$, i.e., 1 to $n - |\boldsymbol{s}|$. In a richer setting where $\boldsymbol{x}$ is discrete I can simulate values of $\boldsymbol{x}^{-\boldsymbol{s}}$ from its estimated distribution rather than summing up over all possible values.

The heart of the problem is computing $\int \mathbb{1}[\varepsilon \in \mathcal{E}_u(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})] dF_\varepsilon(\varepsilon)$ and $\int \mathbb{1}[\varepsilon \in \mathcal{E}_a(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})] dF_\varepsilon(\varepsilon)$. I simulate these integrals. First, draw $S$ realizations of $\varepsilon_s$ from $F_\varepsilon(\varepsilon; \theta_\varepsilon)$ using an importance sampler or other simulation bias-reducing method. Next, given a pre-defined subnetwork $\boldsymbol{s}$, covariate $\boldsymbol{x}$, and pa-

[1] An alternative method is to check whether $(A_t^{\boldsymbol{s}}, \boldsymbol{x}_t^{\boldsymbol{s}})$ is isomorphic to the representative colored network in each equivalence class. This alternate method does not work for computing isomorphic dominance, which I will discuss shortly.

rameter $\theta$, I compute the lattice of admissible networks using Algorithm A.1 below, which is based on Proposition 2.4 and similar to the algorithm in Jia (2008) and Miyauchi (2016). Define $I_1(A^{\boldsymbol{s}}, \boldsymbol{x}; \theta) \equiv \int \mathbb{1}[\varepsilon \in \mathcal{E}_u(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})]dF_\varepsilon(\varepsilon)$ and $I_2(A^{\boldsymbol{s}}, \boldsymbol{x}; \theta) \equiv \int \mathbb{1}[\varepsilon \in \mathcal{E}_a(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})]dF_\varepsilon(\varepsilon)$. The respective simulated integrals are:

$$\hat{I}_1(A^{\boldsymbol{s}}, \boldsymbol{x}; \theta) = \frac{1}{S} \sum_{s=1}^{S} \mathbb{1}(\underline{G}^{\boldsymbol{s}}(\boldsymbol{x}, \varepsilon_s, \theta) \sim A^{\boldsymbol{s}} \sim \overline{G}^{\boldsymbol{s}}(\boldsymbol{x}, \varepsilon_s, \theta))$$
$$\hat{I}_2(A^{\boldsymbol{s}}, \boldsymbol{x}; \theta) = \frac{1}{S} \sum_{s=1}^{S} \mathbb{1}(\underline{G}^{\boldsymbol{s}}(\boldsymbol{x}, \varepsilon_s, \theta) \preceq A^{\boldsymbol{s}} \preceq \overline{G}^{\boldsymbol{s}}(\boldsymbol{x}, \varepsilon_s, \theta)),$$

where the notation $\underline{G}^{\boldsymbol{s}}(\boldsymbol{x}, \varepsilon_s, \theta) \preceq A^{\boldsymbol{s}}$ means that there exists a permutation of $\boldsymbol{s}$, say $\tau(\boldsymbol{s})$, such that $\underline{G}^{\tau(\boldsymbol{s})}(\boldsymbol{x}, \varepsilon_s, \theta) \leq A^{\boldsymbol{s}}$. I define this condition to be *isomorphic dominance*.

Unfortunately there does not exist a function that checks for isomorphic dominance. I use the isomorphic classes and the enumerated list of networks to check whether two networks are isomorphically dominated. The procedure works as follows. Take two equivalence classes $\mathcal{C}(\tilde{A}^{\boldsymbol{s}})$ and $\mathcal{C}(A^{\boldsymbol{s}})$. For all subnetworks $A^{\boldsymbol{s}} \in \mathcal{C}(A^{\boldsymbol{s}})$, I check whether $\tilde{A}^{\boldsymbol{s}} \leq A^{\boldsymbol{s}}$. If there exists a $A^{\boldsymbol{s}} \in \mathcal{C}(A^{\boldsymbol{s}})$ such that $\tilde{A}^{\boldsymbol{s}} \leq A^{\boldsymbol{s}}$, then all subnetworks in the equivalence class $\mathcal{C}(\tilde{A}^{\boldsymbol{s}})$ are isomorphically dominated by those in $A^{\boldsymbol{s}} \in \mathcal{C}(A^{\boldsymbol{s}})$. I execute this procedure outside of the main algorithm and record in a binary, square matrix with dimension equal to the number of equivalence classes. Using this binary matrix, I can evaluate the term $\mathbb{1}(\underline{G}^{\boldsymbol{s}}(\boldsymbol{x}, \varepsilon_s, \theta) \preceq A^{\boldsymbol{s}} \preceq \overline{G}^{\boldsymbol{s}}(\boldsymbol{x}, \varepsilon_s, \theta))$ inside the main algorithm.

The following algorithm demonstrates how to compute the equilibria lattice.

**Algorithm A.1** (Admissible Lattice)**.** *The input is* $(\boldsymbol{x}, \varepsilon, \theta)$ *and the output is* $\underline{G}(\boldsymbol{x}, \varepsilon; \theta)$ *and* $\overline{G}(\boldsymbol{x}, \varepsilon; \theta)$*. Execute the following steps.*

0. *Initialize: Create the* $n \times n$ *matrices* $G_0$ *and* $G_1$ *as defined in the proof of Proposition 2.4. Define the function*

$$V_{ij}(G) \equiv \mathbb{1}[\Pi_{ij}(G, \boldsymbol{x}, \varepsilon; \theta) \geq 0]\mathbb{1}[\Pi_{ji}(G, \boldsymbol{x}, \varepsilon; \theta) \geq 0].$$

   *Set* $k = 1$. *Set* $V_{ij}^0(G) \equiv G_0$.

1. *Compute* $V_{ij}^k(G_0) \equiv V(V_{ij}^{k-1}(G_0))$.

2. *If* $V_{ij}^k(G_0) = V_{ij}^{k-1}(G_0)$, *set*

$$\underline{G}(\boldsymbol{x}, \varepsilon; \theta) = V_{ij}^k(G_0),$$

   *set* $k = 1$, *and go to step 3. Otherwise, set* $k = k + 1$ *and go to step 1.*

3. *Compute* $V_{ij}^k(G_1) \equiv V(V_{ij}^{k-1}(G_1))$.

4. *If* $V_{ij}^k(G_1) = V_{ij}^{k-1}(G_1)$, *set*

$$\overline{G}(\boldsymbol{x}, \varepsilon; \theta) = V_{ij}^k(G_1)),$$

   *and return* $\underline{G}(\boldsymbol{x}, \varepsilon; \theta)$ *and* $\overline{G}(\boldsymbol{x}, \varepsilon; \theta)$. *Otherwise, set* $k = k + 1$ *and go to step 3.*

### A.1.2 Smoothing Moments

The simulated moments are stepwise and hence the gradient of the moment functions are locally flat. There are many ways of smoothing this problem. I elect to use Kriging. Kriging is an interpolation method that smooths a function $f : \mathbb{R}^Q \to \mathbb{R}^W$ using a Gaussian process governed by prior covariances. I use the DACE MatLab package for this purpose. I first draw $M$ structural network parameters using Latin hypercube sampling. I compute the moment functions using the simulated methods described above for each of these $M$ points. The moments are interpolated using Kriging. Finally, I solve $\min/\max_\theta \theta_j$ subject to the interpolated moment inequalities. The interpolated moment inequalities are smooth and this problem can be solved using any standard gradient based algorithm such as `SQP`. The same Kriging procedure can be applied to get calibrated projected confidence intervals.

## A.2 Theorems and Proofs

*Proof of Lemma 2.1.* I argue that the solution to

$$\min_{G \in \mathcal{L}(\underline{G}, \overline{G})} d(G)$$

is equal to $\underline{G}$ when $d(G)$ is monotonically increasing in $G$. The remaining three cases follow a similar logic. Pick any $G \in \mathcal{L}(\underline{G}, \overline{G})$. By the definition of the lattice, $\underline{G} \leq G$. Moreover, $d(\underline{G}) \leq d(G)$ by the definition of a monotonically increasing statistic. Since $G$ is chosen arbitrarily from $\mathcal{L}(\underline{G}, \overline{G})$ and $\underline{G} \in \mathcal{L}(\underline{G}, \overline{G})$ it follows that $\underline{G} = \arg\min_{G \in \mathcal{L}(\underline{G}, \overline{G})} d(G)$, as required. □

The following Lemma is useful for showing Propositions 2.1 and 2.2.

**Lemma A.1.** *Consider the two binary networks $G$ and $G'$. If $G \leq G'$, then $G^k \leq G'^k$ for all $k$.*

*Proof.* I show this by mathematical induction. For the first step

$$\begin{aligned}
(G^2)_{ij} &= \sum_{k=1}^{n} G_{ik} G_{kj} \\
&\leq \sum_{k=1}^{n} G'_{ik} G'_{kj} \\
&= (G'^2)_{ij}.
\end{aligned}$$

So the base step holds. Now suppose that $G^k \leq G'^k$ holds true and consider the $ij^{th}$ element of $G^{k+1}$

$$\begin{aligned}(G^{k+1})_{ij} &= \sum_{k=1}^{n} G_{ik}(G^k)_{kj} \\ &\leq \sum_{k=1}^{n} G'_{ik}(G'^k)_{kj} \\ &= (G'^{k+1})_{ij}.\end{aligned}$$

Therefore, the claim holds by mathematical induction. $\square$

*Proof of Proposition 2.1.* I show that the centrality measures in Examples 2.2 - 2.7 are monotone in the network.

**Monotonic Katz-Bonacich Centrality**: $d^{\text{kbc}}(G; \boldsymbol{w}, \lambda) \equiv \sum_{k=0}^{\infty} \lambda^k G^k \boldsymbol{w}$. Consider the case when $\boldsymbol{w} \geq 0$. Pick any $G, G'$ such that $G \leq G'$ and consider the $i^{th}$ element of $G^k \boldsymbol{w}$:

$$(G^k \boldsymbol{w})_i = \sum_{j=1}^{n} (G^k)_{ij} w_j \leq \sum_{j=1}^{n} (G'^k)_{ij} w_j = (G'^k \boldsymbol{w})_i.$$

The inequality follows because $w_j \geq 0$ for all $j$, $G \leq G'$, and by the result in Lemma A.1. Thus, for all $k$, $G^k \boldsymbol{w} \leq G'^k \boldsymbol{w}$. It follows that

$$d^{\text{kbc}}(G; \boldsymbol{w}, \lambda) = \sum_{k=0}^{\infty} \lambda^k G^k \boldsymbol{w} \leq \sum_{k=0}^{\infty} \lambda^k G'^k \boldsymbol{w} = d^{\text{kbc}}(G'; \boldsymbol{w}, \lambda),$$

as required.

**Diffusion Centrality**: $d_i(G; \lambda, K) = \sum_{k=1}^{K} \sum_{j=1}^{n} \lambda^k G^k_{ij}$. This follows the same argument for KBC – diffusion centrality is equal to a truncated KBC.

**Degree Centrality**: $d_i(G) = \sum_{j=1}^{n} G_{ij}$. Let $G \leq G'$ and let $\mathcal{N}_i(G) = \{j : G_{ij} = 1\}$. It is clear that $\mathcal{N}_i(G) \subset \mathcal{N}_i(G')$. Hence,

$$d_i(G) = \sum_{j \in \mathcal{N}_i(G)} G_{ij} \leq \sum_{j \in \mathcal{N}_i(G)} G_{ij} + \sum_{j \in \mathcal{N}_i(G') - \mathcal{N}_i(G)} G'_{ij} = d_i(G').$$

**Closeness Centrality**: $d_i(G) = \frac{n-1}{\sum_{j=1}^{n} \rho_{ij}(G)}$. Recall that $\rho_{ij}(G)$ is the length of the shortest path between individuals $i$ and $j$ in network $G$. This term does not have an analytical expression. However, it is clear from the definition that $\rho_{ij}(G) \geq \rho_{ij}(G')$ if and only if $G \leq G'$ – adding links can only shorten the shortest path between individuals. Therefore $d_i(G) \leq d_i(G')$.

**Harmonic Centrality**: $d_i(G) = \sum_{j=1}^{n} \frac{n-1}{\rho_{ij}(G)}$. This follows from the same argument for Closeness Centrality.

**Decay Centrality**: $d_i(G; \lambda) = \sum_{k=1}^{n-1} \sum_{j=1}^{n} \lambda^k \mathbb{1}(\rho_{ij}(G) = k)$. Consider any two networks satisfying $G \leq G'$. Suppose that $\lambda \in [0, 1]$ and that $G$ is fully connected. Then there is a unique number $l_{ij}$ such that $\mathbb{1}(\rho_{ij}(G) = l_{ij}) = 1$. Adding links can only shorten the shortest path, so there exists another unique number $l'_{ij} \leq l_{ij}$ such that $\mathbb{1}(\rho_{ij}(G) = l'_{ij}) = 1$. Using this notation,

$$d_i(G; \lambda) = \sum_{j=1}^{n} \lambda^{l_{ij}} \mathbb{1}(\rho_{ij}(G) = l_{ij}) \leq \sum_{j=1}^{n} \lambda^{l'_{ij}} \mathbb{1}(\rho_{ij}(G) = l'_{ij}) = d_i(G'; \lambda).$$

The inequality follows because $\lambda^{l_{ij}} \leq \lambda^{l'_{ij}}$ when $\lambda \in [0, 1]$. The reverse holds true when $\lambda > 1$. The argument is also true for non-connected networks in which case $\mathbb{1}(\rho_{ij}(G) = k) = 0$ for all $k$ for individuals $i$ and $j$ in different communities. $\square$

*Proof of Proposition 2.2.* I can re-write the expression for $\underline{d}(\underline{G}, \overline{G}; \lambda)$ as follows:

$$\begin{aligned}
&\underline{d}(\underline{G}, \overline{G}; \lambda) \\
&= \left[ \left( \sum_{k=0}^{\infty} \lambda^k \underline{G}^k \right)_1, \ldots, \left( \sum_{k=0}^{\infty} \lambda^k \underline{G}^k \right)_{n_+}, \left( \sum_{k=0}^{\infty} \lambda^k \overline{G}^k \right)_{n_+ + 1}, \ldots, \left( \sum_{k=0}^{\infty} \lambda^k \overline{G}^k \right)_n \right].
\end{aligned}$$

Consider the $i^{th}$ element of $\underline{d}(\underline{G}, \overline{G}; \lambda)\boldsymbol{w}$:

$$\begin{aligned}
&[\underline{d}(\underline{G}, \overline{G}; \lambda)\boldsymbol{w}]_i \\
&= \left\{ \left[ \left( \sum_{k=0}^{\infty} \lambda^k \underline{G}^k \right)_1, \ldots, \left( \sum_{k=0}^{\infty} \lambda^k \underline{G}^k \right)_{n_+}, \left( \sum_{k=0}^{\infty} \lambda^k \overline{G}^k \right)_{n_+ + 1}, \ldots, \left( \sum_{k=0}^{\infty} \lambda^k \overline{G}^k \right)_n \right] \boldsymbol{w} \right\}_i \\
&= \sum_{j=1}^{n_+} \left( \sum_{k=0}^{\infty} \lambda^k \underline{G}^k \right)_{ij} w_j + \sum_{j=n_+ + 1}^{n} \left( \sum_{k=0}^{\infty} \lambda^k \overline{G}^k \right)_{ij} w_j \\
&= \sum_{j=1}^{n_+} \sum_{k=0}^{\infty} \lambda^k (\underline{G}^k)_{ij} w_j + \sum_{j=n_+ + 1}^{n} \sum_{k=0}^{\infty} \lambda^k (\overline{G}^k)_{ij} w_j.
\end{aligned}$$

Now consider any $G \in \mathcal{L}(\underline{G}, \overline{G})$. By Lemma A.1, for all $k$: $\underline{G}^k \leq G^k$ and by assumption $w_j \geq 0$ for $j \leq n_+$, hence:

$$\sum_{j=1}^{n_+} \sum_{k=0}^{\infty} \lambda^k (\underline{G}^k)_{ij} w_j \leq \sum_{j=1}^{n_+} \sum_{k=0}^{\infty} \lambda^k (G^k)_{ij} w_j.$$

Similarly, since $w_j < 0$ for $j > n_+$

$$\sum_{j=n_+ + 1}^{n} \sum_{k=0}^{\infty} \lambda^k (\overline{G}^k)_{ij} w_j \leq \sum_{j=n_+ + 1}^{n} \sum_{k=0}^{\infty} \lambda^k (G^k)_{ij} w_j.$$

It therefore follows that

$$\begin{aligned}
[\underline{d}(\underline{G}, \overline{G}; \lambda)\boldsymbol{w}]_i &= \sum_{j=1}^{n_+} \sum_{k=0}^{\infty} \lambda^k (\underline{G}^k)_{ij} w_j + \sum_{j=n_+ + 1}^{n} \sum_{k=0}^{\infty} \lambda^k (\overline{G}^k)_{ij} w_j \\
&\leq \sum_{j=1}^{n_+} \sum_{k=0}^{\infty} \lambda^k (G^k)_{ij} w_j + \sum_{j=n_+ + 1}^{n} \sum_{k=0}^{\infty} \lambda^k (G^k)_{ij} w_j \\
&= d_i^{\mathrm{kbc}}(G; \boldsymbol{w}, \lambda)
\end{aligned}$$

Since $G$ and $i$ are arbitrary, it follows that

$$[\underline{d}(\underline{G}, \overline{G}; \lambda)\boldsymbol{w}]_i \leq \underset{G \in \mathcal{L}(\underline{G}, \overline{G})}{\arg\min}\, d_i^{\mathrm{kbc}}(G; \boldsymbol{w}, \lambda).$$

A similar argument also shows that $[\overline{d}(\underline{G}, \overline{G}; \lambda)\boldsymbol{w}]_i \geq \arg\max_{G \in \mathcal{L}(\underline{G}, \overline{G})}$. $\square$

*Proof of Proposition 2.3.* This follows from the fact that $d_i^{\text{kbc}}(G; \lambda)$ and $(I - \lambda \underline{G})_{ii}^{-1}$ are monotone increasing in $G$. $\square$

*Proof of Proposition 2.4.* Define $G_0, G_1$ by

$$(G_0)_{ij} \equiv 0 \quad \forall i, j \in \boldsymbol{n}$$

$$(G_1)_{ij} \equiv \begin{cases} 1 & \forall i \neq j \\ \\ 0 & i = j \end{cases}.$$

Apply the mapping $V(\cdot)$ to $G_0$ $k$ times. Under Assumption 2.2, $V^k(G_0) \leq V^{k+1}(G_0)$. Hence after a finite $\underline{k}$ number of iterations we obtain $V^{\underline{k}}(G_0) = V^{\underline{k}+1}(G_0)$. Note that $\underline{k} \leq \frac{n(n-1)}{2} - n$, since at most $\frac{n(n-1)}{2} - n$ elements of $G_0$ can change and at least one element must change each time $V(\cdot)$ is applied before a fixed point is obtained.

It remains to show that $V^{\underline{k}}(G_0) = \underline{G}$. For the purpose of obtaining a contradiction, suppose $V^{\underline{k}}(G_0) \neq \underline{G}$. The network $V^{\underline{k}}(G_0)$ is a fixed point of the mapping $V(\cdot)$, so it is pairwise stable. All pairwise stable networks belong to the admissible lattice, hence $\underline{G} < V^{\underline{k}}(G_0)$. Apply the increasing function $V(\cdot)$ $\underline{k}$ times to both sides of the inequality $G_0 \leq \underline{G}$ to obtain

$$V^{\underline{k}}(G_0) \leq V^{\underline{k}}(\underline{G}) = \underline{G} < V^{\underline{k}}(G_0),$$

a contradiction. Therefore $V^{\underline{k}}(G_0) = \underline{G}$. A symmetric argument can be applied to $G_1$ to show that there exists a $\overline{k}$ such that $V^{\overline{k}}(G_1) = \overline{G}$.

The term $\underline{k} + \overline{k}$ is bounded by $\frac{n(n-1)}{2} - n + 1$. In the worse case scenario $\underline{G} = \overline{G}$, which requires $\underline{k} = \sum_{i \neq j} \mathbb{1}(\underline{G}_{ij} = 1)$ applications of $V(\cdot)$ to $G_0$ plus one additional application to ensure that fixed point is obtained, and $\overline{k} = \sum_{i\neq j} \mathbb{1}(\overline{G}_{ij} = 0)$ applications of $V(\cdot)$ to $G_1$. In this case,

$$\underline{k} + \overline{k} = \sum_{i\neq j} \mathbb{1}(\underline{G}_{ij} = 1) + \sum_{i\neq j} \mathbb{1}(\overline{G}_{ij} = 1) + 1 = \frac{n(n-1)}{2} - n + 1.$$

$\square$

*Formal explanation of Remark 2.2.* Let $\overline{D} \equiv \overline{d}(\underline{G}, \overline{G}; \lambda)$. The claim is that there does not always exist a $G \in \mathcal{L}(\underline{G}, \overline{G})$ such that $\overline{D} = (I - \lambda G)^{-1}$, and hence the bounds given in Theorem 2.2 are not necessarily sharp. For sake of argument, suppose that there did exist a $G$ such that $\overline{D} = (I - \lambda G)^{-1}$. Then

$$\begin{aligned}
\overline{D} &= (I - \lambda G)^{-1} = \sum_{k=0}^{\infty} \lambda^k G^k && \iff \\
I &= (I - \lambda G)\overline{D} && \iff \\
\boldsymbol{e}_j &= \begin{cases} (I - \lambda G)\left(\sum_{k=0}^{\infty} \lambda^k (\overline{G})^k_j\right) & \text{if } j \leq n_+ \\ (I - \lambda G)\left(\sum_{k=0}^{\infty} \lambda^k (\underline{G})^k_j\right) & \text{else} \end{cases},
\end{aligned}$$

where $\boldsymbol{e}_j$ is the $j^{th}$ basis vector. Rewriting this we obtain

$$\begin{aligned}
\sum_{k=0}^{\infty} \lambda^k (\overline{G})^k_j &= \boldsymbol{e}_j + \sum_{k=0}^{\infty} \lambda^{k+1} G (\overline{G})^k_j \quad \forall j \leq n_+ \\
\sum_{k=0}^{\infty} \lambda^k (\underline{G})^k_j &= \boldsymbol{e}_j + \sum_{k=0}^{\infty} \lambda^{k+1} G (\underline{G})^k_j \quad \forall j > n_+
\end{aligned} \tag{A.2.1}$$

Setting $G = \overline{G}$ satisfies the first condition in Equation (A.2.1), since:

$$\begin{aligned}
\boldsymbol{e}_j + \sum_{k=0}^{\infty} \lambda^{k+1} G(\overline{G})_j^k &= \boldsymbol{e}_j + \sum_{k=0}^{\infty} \lambda^{k+1} \overline{G}(\overline{G})_j^k \\
&= \boldsymbol{e}_j + \sum_{k=0}^{\infty} \lambda^{k+1} (\overline{G})_j^{k+1} \\
&= \boldsymbol{e}_j + \sum_{k=1}^{\infty} \lambda^{k} (\overline{G})_j^{k} \\
&= \sum_{k=0}^{\infty} \lambda^{k} (\overline{G})_j^{k}.
\end{aligned}$$

Of course, $G = \overline{G}$ fails the second condition when $j > n_+$ for cases where $\overline{G} \neq \underline{G}$. It follows that $G \neq \overline{G}$ and $G \neq \underline{G}$. I show that no such $G$ exists satisfying Equation (A.2.1) in general. For this purpose, maintain the following assumptions: $\lambda > 0$ and $\overline{G}$ is fully connected so that for all $i, l$ there exists a $k$ such that $(\overline{G}^k)_{il} \neq 0$.

Since $G \neq \overline{G}$, there exists an $i$ and $l$ such that $0 = G_{il} \neq \overline{G}_{il} = 1$. Fix any $j \leq n_+$ and consider $ij$ component for the expression in Equation (A.2.1):

$$\begin{aligned}
\sum_{k=0}^{\infty} \lambda^{k} (\overline{G})_j^{k} &= \left[ \boldsymbol{e}_j + \sum_{k=0}^{\infty} \lambda^{k+1} G(\overline{G})_j^k \right]_i \\
&= \mathbb{1}(i = j) + \sum_{k=0}^{\infty} \lambda^{k+1} \sum_{l=1}^{n} G_{il} (\overline{G})_{lj}^k \\
&< \mathbb{1}(i = j) + \sum_{k=0}^{\infty} \lambda^{k+1} \sum_{l=1} \overline{G}_{il} (\overline{G})_{lj}^k \\
&= \left[ \boldsymbol{e}_j + \sum_{k=0}^{\infty} \lambda^{k+1} (\overline{G})_j^{k+1} \right]_i \\
&= \sum_{k=0}^{\infty} \lambda^{k} (\overline{G})_j^{k}.
\end{aligned}$$

The strict inequality follows since $\lambda > 0$, $G \leq \overline{G}$, $G_{il} < \overline{G}_{il}$, and $(\overline{G}^k)_{lj} \neq 0$ for some $k$. I arrive to a contradiction and hence I conclude that there does not exist a $G$ such that $\overline{D} = (I - \lambda G)^{-1}$ in general $\square$

**Proposition A.1.** *Consider the strategic campaign donation with peer effects model in Example A.1. Suppose that the utility of money is given by $v_i(s) = \kappa \ln(s)$ for all legislators. Define mappings $\underline{d} : \mathcal{G} \times \mathcal{G} \to \mathbb{R}^n$ and $\overline{d} : \mathcal{G} \times \mathcal{G} \to \mathbb{R}^n$ by*

$$\underline{d}(\underline{G}, \overline{G}; \lambda, \phi) = \begin{bmatrix} (I - 2\lambda\phi\overline{G})^{-1}_1 \\ \vdots \\ (I - 2\lambda\phi\overline{G})^{-1}_{i-1} \\ (I - 2\lambda\phi\underline{G})^{-1}_i \\ (I - 2\lambda\phi\overline{G})^{-1}_{i+1} \\ \vdots \\ (I - 2\lambda\phi\overline{G})^{-1}_{i+1} \end{bmatrix} \quad \textit{and} \quad \overline{d}(\underline{G}, \overline{G}; \lambda, \phi) = \begin{bmatrix} (I - 2\lambda\phi\underline{G})^{-1}_1 \\ \vdots \\ (I - 2\lambda\phi\underline{G})^{-1}_{i-1} \\ (I - 2\lambda\phi\overline{G})^{-1}_i \\ (I - 2\lambda\phi\underline{G})^{-1}_{i+1} \\ \vdots \\ (I - 2\lambda\phi\underline{G})^{-1}_{i+1}. \end{bmatrix}$$

*Then*

$$\frac{(\underline{d}(\underline{G}, \overline{G}; \lambda, \phi)\mathbf{1})_i W}{\sum_j (\underline{d}(\underline{G}, \overline{G}; \lambda, \phi)\mathbf{1})_j} \leq s_i^* \leq \frac{(\overline{d}(\underline{G}, \overline{G}; \lambda, \phi)\mathbf{1})_i W}{\sum_j (\overline{d}(\underline{G}, \overline{G}; \lambda, \phi)\mathbf{1})_j}$$

*Proof of Proposition A.1.* Let $\underline{D} \equiv \underline{d}(\underline{G}, \overline{G}; \lambda, \phi)$ and $\overline{D} \equiv \overline{d}(\underline{G}, \overline{G}; \lambda, \phi)$. The following statements are equivalent.

$$\begin{aligned}
&\frac{(\underline{D}\mathbf{1})_i W}{\sum_j (\underline{D}\mathbf{1})_j} \leq s_i^* && \iff \\
&\frac{(\underline{D}\mathbf{1})_i}{\sum_j (\underline{D}\mathbf{1})_j} \leq \frac{d_i^{\mathrm{kbc}}(G'; 2\lambda\phi) W}{\sum_j d_j^{\mathrm{kbc}}(G'; 2\lambda\phi)} && \iff \\
&(\underline{D}\mathbf{1})_i \sum_j d_j^{\mathrm{kbc}}(G'; 2\lambda\phi) \leq d_i^{\mathrm{kbc}}(G'; 2\lambda\phi) \sum_j (\underline{D}\mathbf{1})_j && \iff \\
&(\underline{D}\mathbf{1})_i \sum_{j \neq i} d_j^{\mathrm{kbc}}(G'; 2\lambda\phi) \leq d_i^{\mathrm{kbc}}(G'; 2\lambda\phi) \sum_{j \neq i} (\underline{D}\mathbf{1})_j
\end{aligned}$$

This inequality holds, since

$$(\underline{D}\mathbf{1})_i \leq d_i^{\mathrm{kbc}}(\underline{D}; 2\lambda\phi) \leq (\overline{D}\mathbf{1})_i$$

and

$$\forall j \neq i : \quad (\underline{D}\mathbf{1})_j \geq d_j^{\mathrm{kbc}}(\underline{D}; 2\lambda\phi) \geq (\overline{D}\mathbf{1})_j,$$

by construction of $\underline{D}$ and $\overline{D}$. A similar argument can be applied to the upper bound, as required. $\square$

*Proof of Proposition 2.5.* The distribution the $(\boldsymbol{A^s}, \boldsymbol{X^s})$ is equal to the following marginalized distribution of $(\boldsymbol{G}, \boldsymbol{X})$

$$\begin{aligned}
&\boldsymbol{P}(\boldsymbol{A^s} = A^{\boldsymbol{s}}, \boldsymbol{X} = \boldsymbol{x^s}; \theta) \\
&= \sum_{A^{-\boldsymbol{s}} \in \mathcal{G}^{-\boldsymbol{s}}} \sum_{\boldsymbol{x}^{-\boldsymbol{s}}} \boldsymbol{P}(\boldsymbol{G} = (A^{\boldsymbol{s}}, A^{-\boldsymbol{s}}), \boldsymbol{X} = (\boldsymbol{x^s}, \boldsymbol{x}^{-\boldsymbol{s}}); \theta) \\
&= \sum_{A^{-\boldsymbol{s}} \in \mathcal{G}^{-\boldsymbol{s}}} \sum_{\boldsymbol{x}^{-\boldsymbol{s}}} \boldsymbol{P}(\boldsymbol{G} = (A^{\boldsymbol{s}}, A^{-\boldsymbol{s}}) | \boldsymbol{X} = (\boldsymbol{x^s}, \boldsymbol{x}^{-\boldsymbol{s}}); \theta) \boldsymbol{P}(\boldsymbol{X}^{-\boldsymbol{s}} = \boldsymbol{x}^{-\boldsymbol{s}}) \\
&= \sum_{A^{-\boldsymbol{s}} \in \mathcal{G}^{-\boldsymbol{s}}} \sum_{\boldsymbol{x}^{-\boldsymbol{s}}} \left( \int \psi(\boldsymbol{G} = (A^{\boldsymbol{s}}, A^{-\boldsymbol{s}}) | \boldsymbol{x}, \varepsilon; \theta) dF_{\varepsilon}(\varepsilon) \right) \boldsymbol{P}(\boldsymbol{X}^{-\boldsymbol{s}} = \boldsymbol{x}^{-\boldsymbol{s}}) \\
&= \sum_{\boldsymbol{x}^{-\boldsymbol{s}}} \int \left[ \sum_{A^{-\boldsymbol{s}} \in \mathcal{G}^{-\boldsymbol{s}}} \psi(\boldsymbol{G} = (A^{\boldsymbol{s}}, A^{-\boldsymbol{s}}) | \boldsymbol{x}, \varepsilon; \theta) \right] dF_{\varepsilon}(\varepsilon) \boldsymbol{P}(\boldsymbol{X}^{-\boldsymbol{s}} = \boldsymbol{x}^{-\boldsymbol{s}}),
\end{aligned}$$

where the fourth equality follows from Tonelli's Theorem (Tonelli, 1909). $\square$

Before proving Theorem 2.1, I require the following two tools from random set theory (Molchanov, 2005; Molchanov & Molinari, 2018; Molinari, 2019).

**Definition A.1** (Measurable Selection)**.** Let $\mathcal{Z}$ be a random set. A *measurable selection* of $\mathcal{Z}$ is a random element $\boldsymbol{z} \in \mathbb{R}^d$ such that $\boldsymbol{z}(\omega) \in \mathcal{Z}(\omega)$ almost surely. The set of all selections from $\mathcal{Z}$ is denoted $\mathrm{Sel}(\mathcal{Z})$.

The following Lemma is due to (Artstein, 1983).

**Lemma A.2** (Artstein's Inequality)**.** *A probability distribution $\mu$ on $\mathbb{R}^d$ is the distribution of a selection of a random closed set $\mathcal{Z}$ in $\mathbb{R}^d$ iff*

$$\mu(\mathcal{K}) \leq \boldsymbol{T}_{\mathcal{Z}}(\mathcal{K}) \equiv \boldsymbol{P}\{\mathcal{Z} \cap \mathcal{K} \neq \emptyset\}.$$

*for all compact sets $\mathcal{K} \subset \mathbb{R}^d$.*

*Proof of Theorem 2.1.* Under Assumptions 2.1, the conditional distribution of $G^{\bar{\boldsymbol{n}}}$ for each $G^{\bar{\boldsymbol{n}}} \in \mathcal{G}^{\bar{\boldsymbol{n}}}$ is revealed. The parameter $\theta$ is in $\mathcal{H}_P[\theta]$ if and only if the model-implied distribution coupled with the selection mechanism yields the same conditional distribution of $G^{\bar{\boldsymbol{n}}}$ in the data.

Start by considering any $\theta \in \mathcal{H}_P[\theta]$. By the definition of $\mathcal{H}_P[\theta]$, according to Lemma A.2 the terms $G^{\bar{\boldsymbol{n}}}$ and $\mathcal{G}_\theta^{\mathrm{ps}}(\boldsymbol{x}, \varepsilon, \bar{\boldsymbol{n}})$ can be realized on the same probability space as the random element $G^{\bar{\boldsymbol{n}}'} =^d G^{\bar{\boldsymbol{n}}}$ and random set $\mathcal{G}_\theta^{ps'}(\boldsymbol{x}, \varepsilon, \bar{\boldsymbol{n}}) =^d \mathcal{G}_\theta^{\mathrm{ps}}(\boldsymbol{x}, \varepsilon, \bar{\boldsymbol{n}})$ with the property that $G^{\bar{\boldsymbol{n}}'} \in \mathrm{Sel}(\mathcal{G}_\theta^{ps'}(\boldsymbol{x}, \varepsilon, \bar{\boldsymbol{n}}))$. Since $\mathrm{Sel}(\mathcal{G}_\theta^{ps'}(\boldsymbol{x}, \varepsilon, \bar{\boldsymbol{n}}))$ includes all measurable selections, I can choose the one that assigns probability one to $G^{\bar{\boldsymbol{n}}'}$, which is exactly the selection mechanism needed that yields the same distribution observed in the data.

Now suppose that $\theta \in \Theta$ is such that a valid selection mechanism exists with the property that the model-implied distribution coupled with the selection mechanism yields the same distribution of $G^{\bar{\boldsymbol{n}}}$ observed in the data. It follows that there is a selection of $\mathcal{G}_\theta^{\mathrm{ps}}(\boldsymbol{x}, \varepsilon, \bar{\boldsymbol{n}})$ with the same distribution as the selection mechanism whose conditional distribution is $\boldsymbol{P}(G^{\bar{\boldsymbol{n}}}|\boldsymbol{x}), \boldsymbol{x} - a.s.$; hence $\theta \in \mathcal{H}_P[\theta]$. □

*Proof of Theorem 2.2.* Consider any $\theta \in \mathcal{H}_P[\theta]$, $A^{\boldsymbol{s}}$, and $\boldsymbol{x}^{\boldsymbol{s}}$ such that $|\boldsymbol{s}| \leq q$ and $\boldsymbol{s} \subset \bar{\boldsymbol{n}}$, and fix $\mathcal{K}$ such that $\forall G^{\bar{\boldsymbol{n}}}, \tilde{G}^{\bar{\boldsymbol{n}}} \in \mathcal{K}$ with $A^{\boldsymbol{s}} = \tilde{G}^{\boldsymbol{s}}$, otherwise $G \in \mathcal{K}$ is unrestricted. Let $G^{\bar{\boldsymbol{n}}-\boldsymbol{s}}$ denote the links between individuals in $\boldsymbol{s}$ with those in $\bar{\boldsymbol{n}} - \boldsymbol{s}$. Observe that

$$\begin{aligned}
\boldsymbol{P}(\boldsymbol{A}^{\boldsymbol{s}} = A^{\boldsymbol{s}}, \boldsymbol{X} = \boldsymbol{x}^{\boldsymbol{s}}; \theta) &= \sum_{\boldsymbol{x}^{-\boldsymbol{s}}} \sum_{G^{\bar{\boldsymbol{n}}-\boldsymbol{s}} \in \mathcal{G}^{\bar{\boldsymbol{n}}-\boldsymbol{s}}} \boldsymbol{P}(\boldsymbol{G}^{\bar{\boldsymbol{n}}} = (A^{\boldsymbol{s}}, G^{\bar{\boldsymbol{n}}-\boldsymbol{s}})|\boldsymbol{x}; \theta) \boldsymbol{P}(\boldsymbol{x}^{-\boldsymbol{s}}) \\
&= \sum_{\boldsymbol{x}^{-\boldsymbol{s}}} \boldsymbol{P}(\boldsymbol{G}^{\bar{\boldsymbol{n}}} \in \mathcal{K}|\boldsymbol{x}) \boldsymbol{P}(\boldsymbol{x}^{-\boldsymbol{s}}) \\
&\leq \sum_{\boldsymbol{x}^{-\boldsymbol{s}}} \boldsymbol{T}_{\mathcal{G}_\theta^{\mathrm{ps}}(\boldsymbol{x}, \varepsilon, \bar{\boldsymbol{n}})}(\mathcal{K}; F_\varepsilon) \boldsymbol{P}(\boldsymbol{x}^{-\boldsymbol{s}}),
\end{aligned}$$

since $\theta \in \mathcal{H}_P[\theta]$. Since $\mathcal{K}$ imposes no restrictions on $G_{ij}$ with $i \in \boldsymbol{s}$ and $j \in \bar{\boldsymbol{n}} - \boldsymbol{s}$,

$$\begin{aligned}
\boldsymbol{T}_{\mathcal{G}_\theta^{\text{ps}}(\boldsymbol{x},\varepsilon,\bar{\boldsymbol{n}})}(\mathcal{K}; F_\varepsilon) &= \boldsymbol{P}(\{\mathcal{G}_\theta^{\text{ps}}(\boldsymbol{x}, \varepsilon, \bar{\boldsymbol{n}}) \cap \mathcal{K} \neq \emptyset\}, \varepsilon \sim F_\varepsilon) \\
&= \boldsymbol{P}(\text{there exists a } A^{-\boldsymbol{s}} \text{ such that } (A^{\boldsymbol{s}}, A^{-\boldsymbol{s}}) \text{ is pairwise stable}, \varepsilon \sim F_\varepsilon) \\
&= \int_{\varepsilon \in \mathcal{E}_u(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta}) \cup \mathcal{E}_m(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})} dF_\varepsilon(\varepsilon) \\
&\leq \int_{\varepsilon \in \mathcal{E}_a(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})} dF_\varepsilon(\varepsilon).
\end{aligned}$$

Therefore $m_2(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}}; \theta) \leq 0$ for all $(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}})$. In order to obtain the lower bound, i.e. $m_1(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}}; \theta) \leq 0$, I consider the complement of $\mathcal{K}$ and use the containment functional to relate these inequalities to the capacity functional. Defining $\mathcal{K}$ as above,

$$\begin{aligned}
&1 - \boldsymbol{P}(\boldsymbol{A}^{\boldsymbol{s}} = A^{\boldsymbol{s}}, \boldsymbol{X} = \boldsymbol{x}^{\boldsymbol{s}}; \theta) \\
&= \sum_{\boldsymbol{x}^{-\boldsymbol{s}}} \sum_{G^{\bar{\boldsymbol{n}}-\boldsymbol{s}} \in \mathcal{G}^{\bar{\boldsymbol{n}}-\boldsymbol{s}}} \boldsymbol{P}(\boldsymbol{G}^{\bar{\boldsymbol{n}}} = (A^{\boldsymbol{s}}, G^{\bar{\boldsymbol{n}}-\boldsymbol{s}}) | \boldsymbol{x}; \theta) \boldsymbol{P}(\boldsymbol{x}^{-\boldsymbol{s}}) \\
&= 1 - \sum_{\boldsymbol{x}^{-\boldsymbol{s}}} \boldsymbol{P}(\boldsymbol{G}^{\bar{\boldsymbol{n}}} \in \mathcal{K} | \boldsymbol{x}) \boldsymbol{P}(\boldsymbol{x}^{-\boldsymbol{s}}) \\
&= \sum_{\boldsymbol{x}^{-\boldsymbol{s}}} \left(1 - \boldsymbol{P}(\boldsymbol{G}^{\bar{\boldsymbol{n}}} \in \mathcal{K} | \boldsymbol{x})\right) \boldsymbol{P}(\boldsymbol{x}^{-\boldsymbol{s}}) \\
&= \sum_{\boldsymbol{x}^{-\boldsymbol{s}}} \boldsymbol{P}(\boldsymbol{G}^{\bar{\boldsymbol{n}}} \in \mathcal{K}^c | \boldsymbol{x}) \boldsymbol{P}(\boldsymbol{x}^{-\boldsymbol{s}}) \\
&\leq \sum_{\boldsymbol{x}^{-\boldsymbol{s}}} \boldsymbol{T}_{\mathcal{G}_\theta^{\text{ps}}(\boldsymbol{x},\varepsilon,\bar{\boldsymbol{n}})}(\mathcal{K}^c; F_\varepsilon) \boldsymbol{P}(\boldsymbol{x}^{-\boldsymbol{s}}) \\
&\leq 1 - \sum_{\boldsymbol{x}^{-\boldsymbol{s}}} \boldsymbol{P}(\mathcal{G}_\theta^{\text{ps}}(\boldsymbol{x}, \varepsilon, \bar{\boldsymbol{n}}) \subset \mathcal{K}; F_\varepsilon) \boldsymbol{P}(\boldsymbol{x}^{-\boldsymbol{s}}) \\
&= 1 - \sum_{\boldsymbol{x}^{-\boldsymbol{s}}} \boldsymbol{P} \begin{pmatrix} \text{all pairwise stable networks contain} \\ (A^{\boldsymbol{s}}, \boldsymbol{x}) \text{ as a subnetwork}, \varepsilon \sim F_\varepsilon \end{pmatrix} \boldsymbol{P}(\boldsymbol{x}^{-\boldsymbol{s}}) \\
&= 1 - \sum_{\boldsymbol{x}^{-\boldsymbol{s}}} \int_{\varepsilon \in \mathcal{E}_u(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})} dF_\varepsilon(\varepsilon) \boldsymbol{P}(\boldsymbol{x}^{-\boldsymbol{s}}).
\end{aligned}$$

It follows that $\boldsymbol{P}(\boldsymbol{A}^{\boldsymbol{s}} = A^{\boldsymbol{s}}, \boldsymbol{X}^{\boldsymbol{s}} = \boldsymbol{x}^{\boldsymbol{s}}; \theta) \geq \sum_{\boldsymbol{x}^{-\boldsymbol{s}}} \int_{\varepsilon \in \mathcal{E}_u(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})} dF_\varepsilon(\varepsilon) \boldsymbol{P}(\boldsymbol{x}^{\boldsymbol{s}})$ and hence $m_1(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}}; \theta) \leq 0$. I conclude that $\theta \in \mathcal{O}_P[\theta]$ and so $\mathcal{O}_P[\theta]$ is a valid outer region. $\square$

*Proof of Theorem 2.3.* This follows from a similar argument to Theorem 2.2. The additional step is to show that the network statistic also belongs to the sharp identified set in Equation (2.6). This follows from the fact that moment inequalities for $\boldsymbol{E}_\theta(d(G)|\boldsymbol{x})$ are constructed from the solutions to Problem 2.1.

Formally, consider any $(\beta,\theta) \in \mathcal{H}_P[\beta,\theta]$. Following similar logic to Theorem 2.2, $m_j(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}};\theta) \leq 0$ for $j = 1,2$ for all $(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}})$. It remains to show that $m_j(\boldsymbol{x};\beta,\theta) \leq 0$ for $j = 3,4$. Since $(\beta,\theta) \in \mathcal{H}_P[\beta,\theta]$ it follows that there exists a selection mechanism $\psi(\cdot)$ such that

$$
\begin{aligned}
\beta &= \int_\varepsilon \sum_{G \in \mathcal{G}} d(G)\psi(\boldsymbol{G} = G|\boldsymbol{x},\varepsilon;\theta)dF_\varepsilon(\varepsilon) \\
&\leq \int_\varepsilon \sum_{G \in \mathcal{G}} \overline{d}(\underline{G},\overline{G})\psi(\boldsymbol{G} = G|\boldsymbol{x},\varepsilon;\theta)dF_\varepsilon(\varepsilon) \\
&= \int_\varepsilon \overline{d}(\underline{G},\overline{G}) \sum_{G \in \mathcal{G}} \psi(\boldsymbol{G} = G|\boldsymbol{x},\varepsilon;\theta)dF_\varepsilon(\varepsilon) \\
&= \int_\varepsilon \overline{d}(\underline{G},\overline{G})dF_\varepsilon(\varepsilon),
\end{aligned}
$$

where the last equality follows from Definition 2.7 where the selection mechanism integrates to one. A similar argument shows that

$$
\beta \geq \int_\varepsilon \underline{d}(\underline{G},\overline{G})dF_\varepsilon(\varepsilon).
$$

Hence $m_j(\boldsymbol{x};\beta,\theta) \leq 0$ for $j = 3,4$ and $(\beta,\theta) \in \mathcal{O}_P[\beta,\theta]$, as required. $\square$

*Proof of Theorem 2.4.* The proof parallels Theorem 2.3 with the modification that the network statistics are evaluated at the bounds implied by the subgame admissible lattice. $\square$

## A.3 Examples

**Example A.1** (Strategic Campaign Donations with Peer Effects)**.** *Battaglini and Patacchini (2018) develop a model where legislators vote based on a utility function that is linear in four factors: (1) donations from a special interest group; (2) their friends' voting decision; (3) an idiosyncratic unobserved shock that has bounded support; and (4) her preference over whether the policy is approved or not, which could be negative, positive, or zero. For exposition, I focus on the case where the legislator does not care about whether the policy is approved. There are two special interest groups – one group would like the policy to be approved (group a) and the other would like the policy to not be approved and status quo to hold (group b). The special interest groups each have wealth $W$ to allocate between the $n$ legislators. Group a allocates donations to maximize the probability that the policy is passed conditional on group b's allocation and legislators' voting policy, and group b allocates donations to minimize this probability. The optimal allocation is symmetric and is given by:*

$$\boldsymbol{s}^* = \arg\max_{\boldsymbol{s}\in\mathbb{R}_+} \sum_{j=1}^{n} d_j^{kbc}(G'; 2\lambda\phi) v_j(s_j) \quad s.t. \quad \sum_{j=1}^{n} s_j \leq W,$$

*where $s_j$ is the allocation to legislator $j$, $v_j(s_j)$ is the utility from donations (factor (1) from above), $\lambda$ is the social multiplier (factor (2)), and $\phi$ is the length of the support for the idiosyncratic shock (factor (3)). The function $v_j(\cdot)$ is assumed to satisfy the Inada conditions. For example, if $v_j(s) = \kappa \ln(s)$ for all $j$, then*

$$s_i^* = \frac{d_i^{kbc}(G'; 2\lambda\psi)W}{\sum_j d_j^{kbc}(G'; 2\lambda\phi)},$$

*so that legislators with high KBC (i.e., influential legislators) receive a larger share of the donations relative to legislators with low KBC.*

**Example A.2** (Production Equilibrium with an Intersectoral Network, (Acemoglu, Carvalho, Ozdaglar, & Tahbaz-Salehi, 2012)). *Consider an economy where each firm or sector $i$ produces good $x_i$ using a technology that use labor input $l_i$ and output of sector $j$, $x_{ij}$, as inputs. For example, General Motors uses labor and steel sheets as inputs for the production of automobiles. Let $G$ denote the weighted and directed input-output intersectoral network. Component $G_{ij}$ is positive if and only if sector $j$ is an input supplier to sector $i$. With Cobb-Douglas preferences and a Cobb-Douglas production function, the logarithm of real value added is given by:*

$$\boldsymbol{y} = \boldsymbol{v}'\boldsymbol{e},$$

*where $e_i$ i.i.d log productivity shocks and*

$$\boldsymbol{v} = \frac{\alpha}{n}(I - (1-\alpha)G')^{-1}\mathbf{1}$$

*is the influence vector, which is also the Katz-Bonacich centrality measure with weights $\mathbf{1}$ and social multiplier $(1-\alpha)$, Here $\alpha$ is share of labor hired by each sector. The influence vector plays an important roll for the asymptotic distribution of real value added output.*

**Example A.3.** *As an example, suppose $\boldsymbol{n} = \{1,2,3,4\}$ and $\boldsymbol{s} = \{1,2\}$ and*

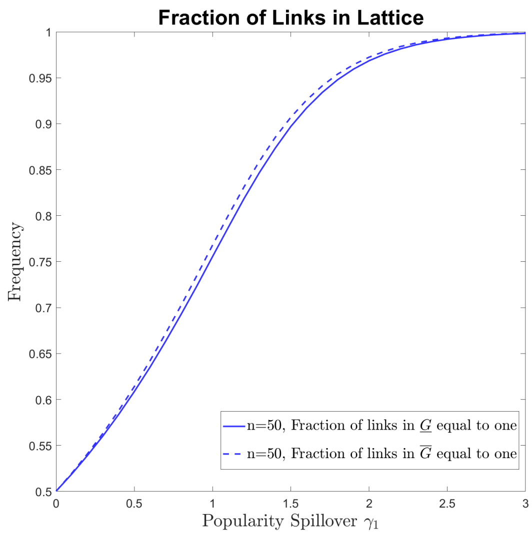
$$G = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix}.$$

*Then*

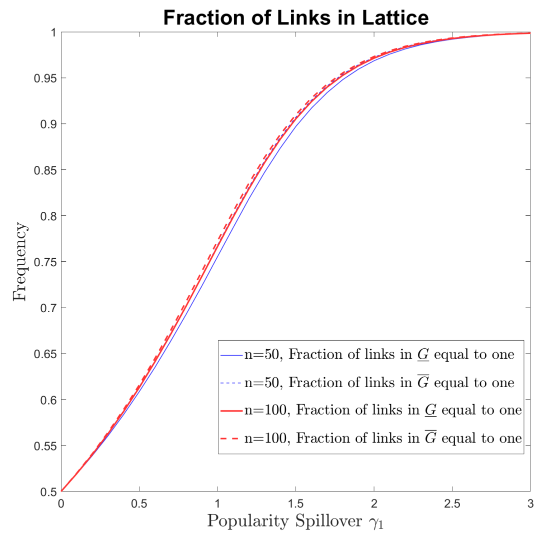
$$A^{\boldsymbol{s}} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \textit{and} \quad A^{-\boldsymbol{s}} = \begin{bmatrix} \cdot & \cdot & 0 & 1 \\ \cdot & \cdot & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix}.$$

## A.4 Simulated Application Results: Figures

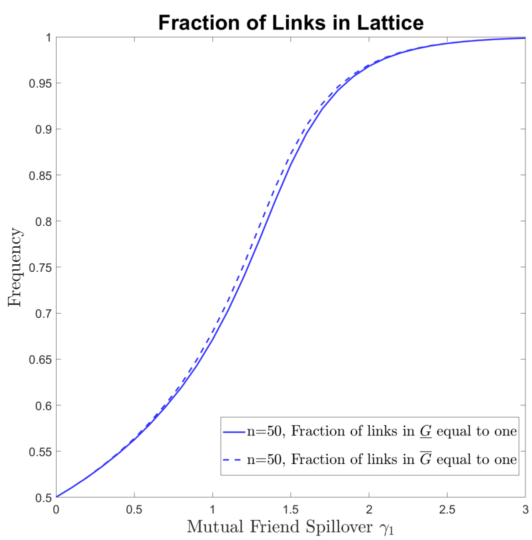
**Figure A.1:** Admissible Lattice Size



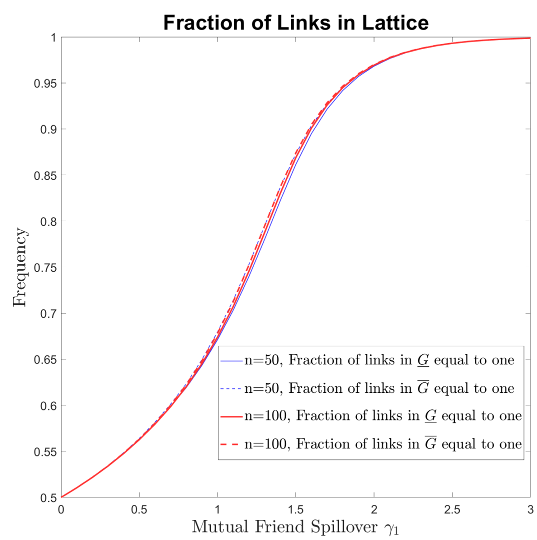

**(a)** Expected fraction of links in the admissible lattice with popularity spillover only. The dotted line is the fraction of links in $\overline{G}$ and the solid line is the fraction of links in $\underline{G}$. These results are reported for $n = 50$.




**(b)** Expected fraction of links in the admissible lattice with popularity spillover only and $n = 100$.

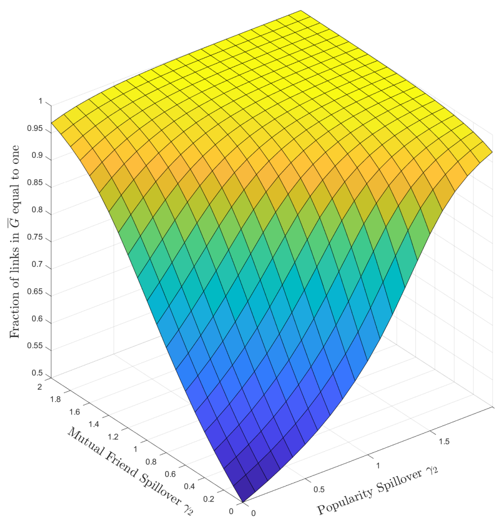



**(c)** Expected fraction of links in the admissible lattice with mutual friend spillover only. The dotted line is the fraction of links in $\overline{G}$ and the solid line is the fraction of links in $\underline{G}$. These results are reported for $n = 50$.

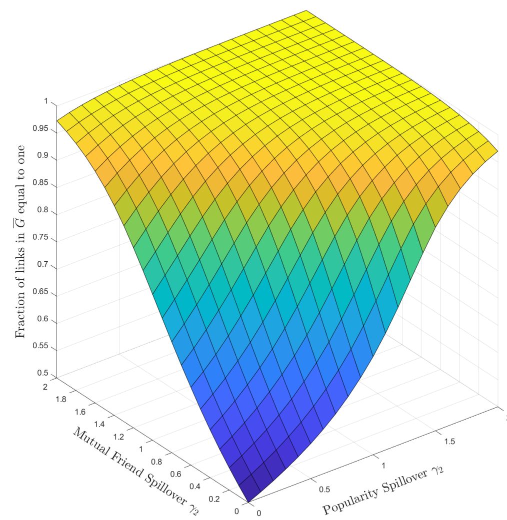



**(d)** Expected fraction of links in the admissible lattice with mutual friend spillover only and $n = 100$.

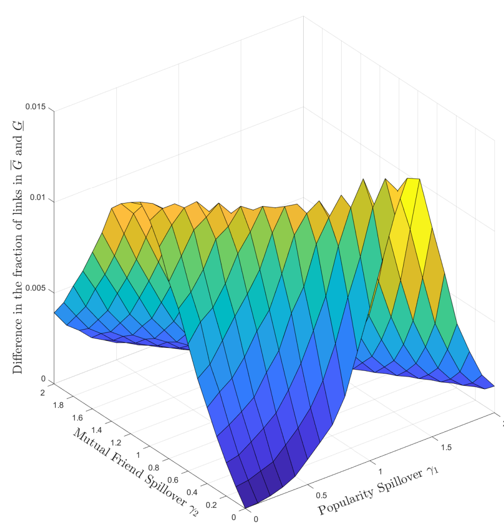
**Figure A.2:** Admissible Lattice Size with Two Channels of Spillovers




**(a)** Expected fraction of links in for the lower bound of the admissible lattice, $\underline{G}$, with both channels of spillovers. These results are reported for $n = 50$.




**(b)** Expected fraction of links in for the upper bound of the admissible lattice, $\overline{G}$, with both channels of spillovers. These results are reported for $n = 50$.




**(c)** Expected fractional difference between the upper and lower bound of the addmissible lattice with both channels of spillovers. These results are reported for $n = 50$.

## A.5 Simulated Application Results: Tables

| Population Size | Popularity Spillover, $\gamma_1$ | Max Subnetwork Size, $s$ $q=2$ | $q=3$ | $q=4$ |
|---|---|---|---|---|
| 50 | 0 | [0.000,0.000] | [0.000,0.000] | [0.000,0.000] |
| | 0.5 | [0.491,0.512] | [0.491,0.512] | [0.491,0.510] |
| | 1 | [0.979,1.018] | [0.984,1.018] | [0.985,1.018] |
| 100 | 0 | [0.000,0.000] | [0.000,0.000] | [0.000,0.000] |
| | 0.5 | [0.494,0.504] | [0.494,0.503] | [0.500,0.503] |
| | 1 | [0.991,1.009] | [0.991,1.006] | [0.992,1.000] |

**Table A.1** Identified set for the popularity spillover with no homophily and one channel of spillovers. The identified set is reported in the square brackets.

| Pop Size | Mutual Friend Spillover, $\gamma_2$ | Max Subnetwork Size $q=2$ | $q=3$ | $q=4$ |
|---|---|---|---|---|
| 50 | 0 | [0.000,0.000] | [0.000,0.000] | [0.000,0.000] |
| | 0.5 | [0.495,0.515] | [0.499,0.501] | [0.499,0.500] |
| | 1 | [0.988,1.014] | [0.988,1.012] | [0.996,1.012] |
| 100 | 0 | [0.000,0.000] | [0.000,0.000] | [0.000,0.000] |
| | 0.5 | [0.498,0.503] | [0.500,0.502] | [0.500,0.501] |
| | 1 | [0.993,1.006] | [0.994,1.002] | [0.995,1.002] |

**Table A.2** Identified set for the mutual-friend friend spillover with no homophily and one channel of spillovers. The identified set is reported in the square brackets.

| $\lambda$ | Pop. Spillover | Max Subnetwork Size $q = 2$ | $q = 3$ | $q = 4$ | True Effort | Worst-Case Bounds |
|---|---|---|---|---|---|---|
| 0.01 | 0 | [1.324,1.324] | [1.324,1.324] | [1.324,1.324] | 1.324 | [1.300,1.356] |
| | 0.5 | [1.421,1.426] | [1.421,1.426] | [1.421,1.426] | 1.422 | [1.383,1.462] |
| | 1 | [1.577,1.592] | [1.579,1.592] | [1.579,1.592] | 1.580 | [1.509,1.626] |
| 0.02 | 0 | [1.972,1.972] | [1.972,1.972] | [1.972,1.972] | 1.972 | [1.803,2.383] |
| | 0.5 | [2.527,2.562] | [2.528,2.560] | [2.528,2.558] | 2.528 | [2.152,3.190] |
| | 1 | [3.991,4.239] | [4.008,4.239] | [4.016,4.239] | 4.020 | [2.850,5.532] |

**Table A.3** Bounds on KBC using my framework. I restrict to one channel of spillovers (popularity) and set $n = 50$ in each network. I assume that 25 individuals are sampled.

| $\gamma_1$ | $\gamma_2$ | Max Subnetwork Size $q = 2$ | $q = 3$ | $q = 4$ |
|---|---|---|---|---|
| 0 | 0 | [0.000,0.000] | [0.000,0.000] | [0.000,0.000] |
| 0.5 | 0 | [0.000,0.510] | [0.054,0.510] | [0.227,0.510] |
| 0 | 0.5 | [0.000,0.312] | [0.000,0.309] | [0.000,0.0308] |
| 0.5 | 0.5 | [0.000,0.904] | [0.000,0.904] | [0.0233,0.883] |

**Table A.4** Identified region for the popularity spillover $\gamma_1$ when both channels of spillovers are present. The popularity spillover is $\gamma_1$ and the mutual friend spillover is $\gamma_2$. Population size $N = 50$.

| $\gamma_1$ | $\gamma_2$ | Max Subnetwork Size $q = 2$ | $q = 3$ | $q = 4$ |
|---|---|---|---|---|
| 0 | 0 | [0.000,0.000] | [0.000,0.000] | [0.000,0.000] |
| 0.5 | 0 | [0.000,0.768] | [0.000,0.673] | [0.000,0.410] |
| 0 | 0.5 | [0.000,0.502] | [0.000,0.502] | [0.437,0.502] |
| 0.5 | 0.5 | [0.000,1.164] | [0.000,1.144] | [0.0233,1.010] |

**Table A.5** Identified region for the mutual-friend spillover $\gamma_2$ when both channels of spillovers are present. The popularity spillover is $\gamma_1$ and the mutual friend spillover is $\gamma_2$. Population size $N = 50$.

| Pop. Spillover | Mutual Spillover | Max Subnetwork Size $q = 2$ | $q = 3$ | $q = 4$ | True KBC | Worst-Case Bounds |
|---|---|---|---|---|---|---|
| 0 | 0 | [1.336, 1.336] | [1.336, 1.336] | [1.336, 1.336] | 1.336 | [1.241, 1.473] |
| 0 | 0.25 | [1.363, 1.364] | [1.363, 1.364] | [1.363, 1.364] | 1.363 | [1.259, 1.497] |
| 0 | 0.5 | [1.396, 1.398] | [1.397, 1.398] | [1.398, 1.398] | 1.398 | [1.282, 1.526] |
| 0.25 | 0 | [1.386, 1.387] | [1.386, 1.387] | [1.386, 1.387] | 1.386 | [1.274, 1.517] |
| 0.25 | 0.5 | [1.428, 1.431] | [1.429, 1.431] | [1.429, 1.430] | 1.430 | [1.304, 1.556] |
| 0.25 | 1 | [1.478, 1.483] | [1.478, 1.482] | [1.479, 1.482] | 1.481 | [1.335, 1.598] |
| 0.5 | 0 | [1.452, 1.455] | [1.452, 1.455] | [1.452, 1.454] | 1.453 | [1.319, 1.576] |
| 0.5 | 0.5 | [1.502, 1.507] | [1.502, 1.507] | [1.502, 1.506] | 1.504 | [1.350, 1.618] |
| 0.5 | 1 | [1.573, 1.580] | [1.573, 1.579] | [1.573, 1.579] | 1.576 | [1.395, 1.676] |

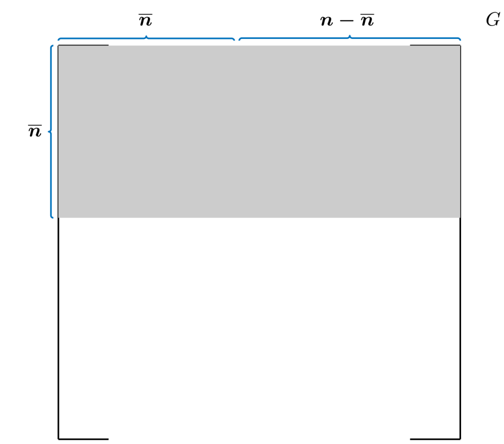
**Table A.6** Bounds on KBC using my framework. I allow for both channels spillovers (popularity) and set $n = 50$ in each network. There are 25 individuals that are sampled in each network. The decay parameter $\lambda = 0.01$.

| Pop. Spillover | Mutual Spillover | Max Subnetwork Size $q = 2$ | $q = 3$ | $q = 4$ | True KBC | Worst-Case Bounds |
|---|---|---|---|---|---|---|
| 0 | 0 | [1.336,1.336] | [1.336,1.336] | [1.336,1.336] | 1.336 | [1.241,1.473] |
| 0 | 0.25 | [1.363,1.364] | [1.363,1.364] | [1.363,1.364] | 1.364 | [1.259 ,1.497] |
| 0 | 0.5 | [1.396,1.398] | [1.397,1.398] | [1.398,1.398] | 1.398 | [1.282, 1.526] |
| 0.25 | 0 | [1.386,1.387] | [1.386,1.387] | [1.386,1.387] | 1.386 | [1.275, 1.517] |
| 0.25 | 0.25 | [2.512,2.535] | [2.519,2.533] | [2.521,2.532] | 2.527 | [1.905 , 3.623] |
| 0.25 | 0.5 | [2.857,2.895] | [2.860,2.889] | [2.863,2.888] | 2.879 | [2.053, 4.108] |
| 0.5 | 0 | [2.670,2.695] | [2.670,2.691] | [2.670,2.690] | 2.679 | [1.975, 3.839] |
| 0.5 | 0.25 | [3.044,3.090] | [3.046,3.087] | [3.048,3.080] | 3.065 | [2.127, 4.368] |
| 0.5 | 0.5 | [3.718,3.800] | [3.721,3.797] | [3.725,3.792] | 3.762 | [2.376, 5.328] |

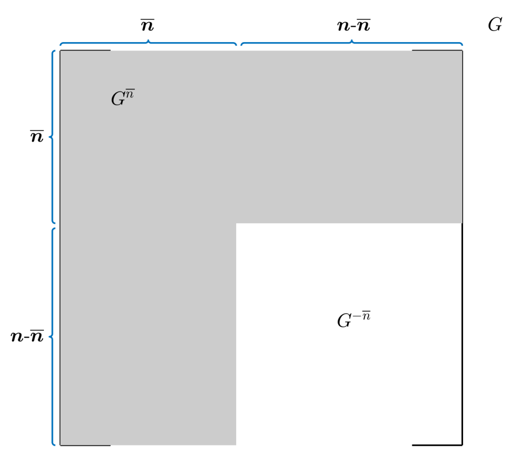
**Table A.7** Bounds on KBC using my framework. I allow for both channels spillovers (popularity) and set $n = 50$ in each network. There are 25 individuals that are sampled in each network. The decay parameter $\lambda = 0.02$.

## A.6 Partially Observed Network and Subnetworks

**Figure A.6.1:** Partially Observed Network

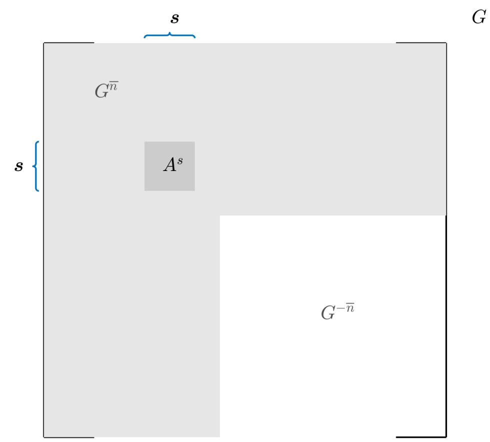



**(a)** The researcher samples $\bar{\boldsymbol{n}}$ individuals and these individuals reveal all of their direct connections. The revealed connections is represented by the shaded area.

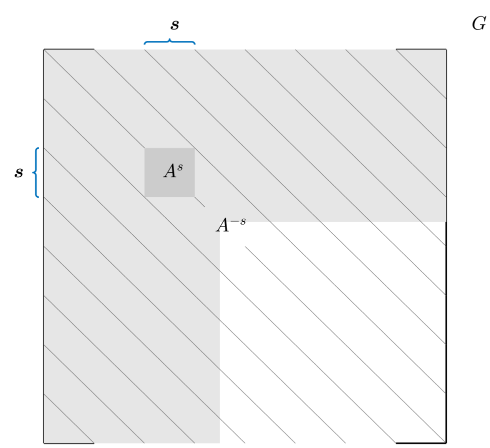



**(b)** The network is assumed to be symmetric. All links between individuals in $\boldsymbol{n} - \bar{\boldsymbol{n}}$ to $\bar{\boldsymbol{n}}$ are also observed. As a result, the observed network is represented by the shaded area. The observed portion of the network is $G^{\bar{\boldsymbol{n}}}$ and the unobserved portion is $G^{-\bar{\boldsymbol{n}}}$.

**Figure A.6.2:** Subnetworks and the Partially Observed Network




**(a)** A subnetwork is constructed from a set of individuals $\boldsymbol{s} \subset \bar{\boldsymbol{n}}$. It contains all links between individuals in $\boldsymbol{s}$. The subnetwork $A^{\boldsymbol{s}}$ is represented by the dark shaded area.




**(b)** The completion to a subnetwork denoted $A^{-\boldsymbol{s}}$ include all links between individuals not in $A^{\boldsymbol{s}}$. The compleition to the subnetwork is represented by the striped area.

## A.7 Table of Notation

**Table A.1** Table of Notation

| | |
|---|---|
| $\boldsymbol{n}_t$ | Set of individuals in environment $t$. |
| $\bar{\boldsymbol{n}}_t$ | Set of individuals interviewed about social connections in environment $t$. |
| $G$, $G_{ij}$ | Adjacency matrix where the $ij$ element is equal to one iff $i$ and $j$ are linked. |
| $A^{\boldsymbol{s}}$, $A^{-\boldsymbol{s}}$ | Adjacency matrix for the subnetwork with individuals in $\boldsymbol{s} \subset \{1, \ldots, n\}$ and its completion. |
| $\boldsymbol{x}$ | Network-formation observable characteristic. |
| $\varepsilon$ | Network-formation unobservable characteristic. |
| $\boldsymbol{G}$, $\boldsymbol{G}^{\bar{\boldsymbol{n}}}$, $\boldsymbol{X}$, $\boldsymbol{A}^{\boldsymbol{s}}$, $\boldsymbol{A}^{-\boldsymbol{s}}$ | Random elements of the above realized values. |
| $G^{\bar{\boldsymbol{n}}}$, $G^{-\bar{\boldsymbol{n}}}$ | Partially observed network and its completion. |
| $G + \{ij\}$, $G - \{ij\}$ | Network $G$ with link $ij$ added and deleted, respectively |
| $\underline{G}(\boldsymbol{x}, \varepsilon, \boldsymbol{\theta}), \overline{G}(\boldsymbol{x}, \varepsilon, \boldsymbol{\theta})$ | Networks that define the admissible lattice. |
| $\theta, \theta_0$ | The network formation parameter and its true value. |
| $\pi_i(\boldsymbol{g}, \boldsymbol{x}, \varepsilon)$ | Agent $i$'s payoff function, where $\boldsymbol{x}$ are observable, $\varepsilon$ unobservable. |
| $\Pi_{ij}(\boldsymbol{g}, \boldsymbol{x}, \varepsilon)$ | Agent $i$'s marginal payoff function over link $ij$. |
| $\mathcal{G}_\theta^{\mathrm{ps}}(\boldsymbol{x}, \varepsilon)$ | Collection of pairwise stable networks. |
| $\mathcal{G}_\theta(\boldsymbol{x}, \varepsilon)$ | Collection of admissible networks. |
| $\mathcal{G}_\theta(\boldsymbol{x}, G^{\bar{\boldsymbol{n}}}, \varepsilon)$ | Collection of conditional admissible networks. |
| $\mathcal{L}(\underline{G}, \overline{G})$ | Generic network lattice containing networks $G$ such that $\underline{G} \leq G \leq \overline{G}$. |
| $\mathcal{G}$ | Collection of undirected networks with no self loops. |
| $\psi(G \vert \boldsymbol{x}, \varepsilon)$ | Selection mechanism for network $G$. |
| | |
| $d(G)$ | Generic network statistic. |
| $d^{\mathrm{kbc}}(G; \boldsymbol{w}, \lambda)$ | Katz-Bonacich Centrality. |
| | |
| $\boldsymbol{P}(\cdot)$ | Probability measure. |
| $m_j(\cdot)$ | Moment inequality |
| $P$ | Joint distribution of observables. |
| $\mathcal{H}_P[.]$ | Sharp identification region for parameter(s) in the square bracket (function of $P$) |
| $\mathcal{O}_P[.]$ | Outer region for parameter(s) in the square bracket (function of $P$) |
| $\mathbb{R}^n_+$, $\mathbb{R}^n_-$ | Positive and negative quadrant of the $n$-dimensional Euclidean space. |
| $\mathbb{Z}$ | Integers. |

An *admissible network* is one that is bounded between the smallest and largest pairwise stable network.

APPENDIX B

# CHAPTER 2 APPENDIX

## B.1 Identification Proofs

**Lemma B.1.** *Consider the ARC Model under the Coin Toss completion rule. If there exists characteristics $x^L$ and permutation $\{o_1, \ldots, o_D\}$ such that $\forall \nu \in [0, \bar{\nu}]$*

$$L_{o_1}(x^L) \succ L_{o_2}(x^L) \succ \cdots \succ L_{o_D}(x^L),$$

*then the following relationship holds for all $j = 1, \ldots, D$:*

$$\varphi_{o_j} = \frac{q_{o_j} - r}{1 + (j-1)r - \sum_{k=1}^{j-1} q_{o_k}},$$

*where $q_{o_k} \equiv Pr(d = d_{o_k} | x^L)$ and $r = \frac{1}{D} \prod_{k=1}^{N} (1 - \varphi_{o_k})$.*

*Proof.* Without loss of generality suppose that $k = o_k$ for all $k$. Fix any $j \in \mathcal{D}$. We first show that

$$\prod_{k=1}^{j-1} (1 - \varphi_k) = 1 + (j-1)r - \sum_{k=1}^{j-1} q_k.$$

On the one hand, by additivity of probability and the definition of $q_k$

$$Pr(d \in \{d_1, d_2, ...d_{j-1}\} | x^L) = \sum_{k=1}^{j-1} q_k.$$

On the other hand, according to the model, it is the probability that at least one of $\{d_1, d_2, ...d_{j-1}\}$ is considered plus the probability one of them is chosen when the consideration set is empty:

$$Pr(d \in \{d_1, d_2, ...d_{j-1}\} | x^L) = 1 - \prod_{k=1}^{j-1} (1 - \varphi_k) + (j-1)r.$$

Finally, due to the assumption of the preference ordering,

$$\begin{aligned} q_j &= \varphi_j \prod_{k=1}^{j-1}(1-\varphi_k) + r \\ &= \varphi_j \left(1 + (j-1)r - \sum_{k=1}^{j-1} q_k\right) + r. \end{aligned}$$

□

**Lemma B.2.** *Consider the ARC Model under the Default Option completion rule. If there exist characteristics $x^L$ and permutation $\{o_1, \ldots, o_D\}$ such that $\forall \nu \in [0, \bar{\nu}]$*

$$L_{o_1}(x^L) \succ L_{o_2}(x^L) \succ \cdots \succ L_{o_D}(x^L),$$

*then the following relationship holds for all $j = 1, \ldots, D$:*

$$\varphi_{o_j} = \begin{cases} \frac{q_{o_j} - r}{1 - \sum_{k=1}^{j-1} q_{o_k}} & \textit{if } d_{o_j} \textit{ is the default option} \\ \frac{q_{o_j}}{1 + r - \sum_{k=1}^{j-1} q_{o_k}} & \textit{if } d_{o_n} \textit{ is the default option for some } n < j \\ \frac{q_{o_j}}{1 - \sum_{k=1}^{j-1} q_{o_k}} & \textit{otherwise} \end{cases},$$

*where $q_{o_k} \equiv Pr(d = d_{o_k} | x^L)$ and $r = \prod_{k=1}^{N}(1 - \varphi_{o_k})$.*

*Proof.* Let $d_n$ denote the default option. The proof follows exactly the same steps as the proof of the previous lemma, except with the following two changes:

$$Pr(d \in \{d_1, d_2, ... d_{j-1}\} | x^L) = \begin{cases} 1 - \prod_{k=1}^{j-1}(1 - \varphi_k) + r & \text{if } n < j \\ 1 - \prod_{k=1}^{j-1}(1 - \varphi_k) & \text{otherwise.} \end{cases}$$

$$q_j = \begin{cases} \varphi_j \prod_{k=1}^{j-1}(1 - \varphi_k) + r & \text{if } n = j \\ \varphi_j \prod_{k=1}^{j-1}(1 - \varphi_k) & \text{otherwise.} \end{cases}$$

The three cases immediately follow depending on whether $n < j$, $n = j$, or $n > j$.

□

**Lemma B.3.** *Consider the ARC Model under the Outside Option completion rule. If there exist characteristics $x^L$ and permutation $\{o_1, \ldots, o_D\}$ such that $\forall \nu \in [0, \bar{\nu}]$*

$$L_{o_1}(x^L) \succ L_{o_2}(x^L) \succ \cdots \succ L_{o_D}(x^L),$$

*then the following relationship holds for all $j = 1, \ldots, D$:*

$$\varphi_{o_j} = \frac{(1-r)q_{o_j}}{1-(1-r)\sum_{k=1}^{j-1} q_{o_k}},$$

*where $q_{o_k} \equiv Pr(d = d_{o_k}|x^L, \mathcal{K} \neq \emptyset)$, $\mathcal{K}$ is the consideration set, and $r = \prod_{k=1}^{N}(1 - \varphi_{o_k})$.*

*Proof.* Without loss of generality suppose that $k = o_k$ for all $k$. Fix any $j \in \mathcal{D}$. We first show that

$$\prod_{k=1}^{j-1}(1-\varphi_k) = 1-(1-r)\sum_{k=1}^{j-1} q_k.$$

On the one hand, by additivity of probability and the definition of $q_k$

$$\begin{aligned} Pr(d \in \{d_1, d_2, ...d_{j-1}\}|x^L) &= Pr(\mathcal{K} \neq \emptyset)Pr(d \in \{d_1, d_2, ...d_{j-1}\}|x^L, \mathcal{K} \neq \emptyset) \\ &= (1-r)\sum_{k=1}^{j-1} q_k. \end{aligned}$$

On the other hand, according to the model, it is the probability that at least one of $\{d_1, d_2, ...d_{j-1}\}$ is considered

$$Pr(d \in \{d_1, d_2, ...d_{j-1}\}|x^L) = 1 - \prod_{k=1}^{j-1}(1-\varphi_k).$$

Finally, due to the assumption of the preference ordering,

$$\begin{aligned} q_j &= \frac{1}{(1-r)}\varphi_j \prod_{k=1}^{j-1}(1-\varphi_k) \\ &= \frac{1}{(1-r)}\varphi_j \left(1-(1-r)\sum_{k=1}^{j-1} q_k\right). \end{aligned}$$

□

## Proof of Theorem 3.2

*Proof.* Start with Coin Toss. Let $d_{o_m}$ be the first alternative in the sequence $L_{o_1}(x^1) \succ L_{o_2}(x^1) \succ \ldots$ such that $o_m \neq m$. Let $j$ be the position of alternative $o_m$ in the sequence $L_1(x^0) \succ L_2(x^0) \succ \ldots$. Note that $m < j$, $d_{o_m} = d_j$, and all lotteries that dominate $d_{o_m}$ at $x^1$ also dominate $d_j$ at $x^0$, since, by construction $o_1 = 1, o_2 = 2, o_3 = 3, \ldots, o_{m-1} = m - 1$.

The assumptions are satisfied in Lemma $B.1$ for $d_j$ at $x^0$ and $d_{o_m}$ at $x^1$. It follows that:

$$\frac{q_j - r}{1 + (j-1)r - \sum_{k=1}^{j-1} q_k} = \varphi_j = \varphi_{o_m} = \frac{s_{o_m} - r}{1 + (m-1)r - \sum_{k=1}^{m-1} s_{o_k}}. \qquad \text{(B.1.1)}$$

where $q_k \equiv Pr(d = d_k | x^0)$ and $s_{o_k} \equiv Pr(d = d_{o_k} | x^1)$. This is a quadratic equation in $r$. Note that

$$s_{o_m} = \varphi_{o_m} \prod_{k=1}^{m-1} (1 - \varphi_{o_k}) + r \geq r.$$

So any admissible solution for $r$ ought to be in the interval $[0, s_{o_m}]$; we show that Equation (B.1.1) has a unique solution in $[0, s_{o_m}]$.

Collecting terms we can write Equation (B.1.1) as follows:

$$\begin{aligned} 0 = g(r) &\equiv ar^2 + br + c \\ &\equiv (m-j)r^2 + \left( s_{o_m}(j-1) - q_j(m-1) + \sum_{k=1}^{j-1} q_k - \sum_{k=1}^{m-1} s_{o_k} \right) r \\ &\quad + s_{o_m}\left(1 - \sum_{k=1}^{j-1} q_k\right) - q_j \left(1 - \sum_{k=1}^{m-1} s_{o_k}\right). \end{aligned}$$

We first show the following

1. $q_j < s_{o_m}$
2. $\sum_{k=1}^{j-1} q_k > \sum_{k=1}^{m-1} s_{o_k}$

3. $s_{o_m}\left(1-\sum_{k=1}^{j-1} q_k\right) < q_j\left(1-\sum_{k=1}^{m-1} s_{o_k}\right)$

from which it follows that the coefficients for the quadratic function satisfy $a < 0$, $b > 0$, and $c < 0$.

Indeed, we have:

1. $q_j < s_{o_m}$:

$$\begin{aligned}
q_j &= \varphi_j \prod_{k=1}^{j-1}(1-\varphi_k) + r \\
&= \varphi_{o_m} \prod_{k=1}^{m-1}(1-\varphi_{o_k}) \prod_{k=m}^{j-1}(1-\varphi_k) + r \quad (\text{since } o_1 = 1, \ldots, o_{m-1} = m-1). \\
&< \varphi_{o_m} \prod_{k=1}^{m-1}(1-\varphi_{o_k}) + r \\
&= s_{o_m}.
\end{aligned}$$

2. $\sum_{k=1}^{j-1} q_k > \sum_{k=1}^{m-1} s_{o_k}$:

$$\sum_{k=1}^{j-1} q_k = \sum_{k=1}^{m-1} s_{o_k} + \sum_{k=m}^{j-1} q_k > \sum_{k=1}^{m-1} s_{o_k}.$$

3. $s_{o_m}\left(1-\sum_{k=1}^{j-1} q_k\right) < q_j\left(1-\sum_{k=1}^{m-1} s_{o_k}\right)$:

$$\begin{aligned}
\left(1-\sum_{k=1}^{j-1} q_k\right) &= \prod_{k=1}^{j-1}(1-\varphi_k)-(j-1)r \\
&= \prod_{k=1}^{m-1}(1-\varphi_{o_k})\prod_{k=m}^{j-1}(1-\varphi_k)-(j-1)r \\
&\equiv uv-(j-1)r, \\
\left(1-\sum_{k=1}^{j-1} s_{o_k}\right) &= \prod_{k=1}^{m-1}(1-\varphi_{o_k})-(m-1)r = u-(m-1)r, \\
s_{o_m} &= \varphi_{o_m}\prod_{k=1}^{m-1}(1-\varphi_{o_k})+r=\varphi_j u+r, \qquad \text{and} \\
q_j &= \varphi_j\prod_{k=1}^{j-1}(1-\varphi_k)+r=\varphi_j uv+r.
\end{aligned}$$

Putting this together:

$$\begin{aligned}
&s_{o_m}\left(1-\sum_{k=1}^{j-1} q_k\right)-q_j\left(1-\sum_{k=1}^{m-1} s_{o_k}\right) \\
&=(\varphi_j u+r)(uv-(j-1)r)-(\varphi_j uv+r)(u-(m-1)r) \\
&=(m-j)r^2+\varphi_j ur(v(m-1)-(j-1))+ur(v-1) \\
&<0
\end{aligned}$$

since $u, v, r, \varphi_j \in [0,1]$ and $m < j$.

We have shown that the coefficients of the quadratic function $g(r)$ have the following signs

$$\begin{aligned}
a =&m-j<0 \\
b =&\left(s_{o_m}(j-1)-q_j(m-1)+\sum_{k=1}^{j-1} q_k-\sum_{k=1}^{m-1} s_{o_k}\right)>0 \\
c =&s_{o_m}\left(1-\sum_{k=1}^{j-1} q_k\right)-q_j\left(1-\sum_{k=1}^{m-1} s_{o_k}\right)<0.
\end{aligned}$$

Thus $g(r)$ is a concave quadratic function. To understand its behavior we show that $g(r)$ evaluated at $r = s_{o_m}$ is positive.

$$\begin{aligned} g(s_{o_m}) &= (m-j)s_{o_m}^2 + \left( s_{o_m}(j-1) - q_j(m-1) + \sum_{k=1}^{j-1} q_k - \sum_{k=1}^{m-1} s_{o_k} \right) s_{o_m} \\ &\quad + s_{o_m}\left(1 - \sum_{k=1}^{j-1} q_k\right) - q_j\left(1 - \sum_{k=1}^{m-1} s_{o_k}\right) \\ &= (m-1)s_{o_m}^2 + \left(1 - q_j(m-1) - \sum_{k=1}^{m-1} s_{o_k}\right) s_{o_m} - q_j\left(1 - \sum_{k=1}^{m-1} s_{o_k}\right) \\ &= (m-1)s_{o_m}(s_{o_m} - q_j) + (s_{o_m} - q_j)\left(1 - \sum_{k=1}^{m-1} s_{o_k}\right) \\ &= (s_{o_m} - q_j)\left(m s_{o_m} + 1 - \sum_{k=1}^{m} s_{o_k}\right) > 0, \end{aligned}$$

since $s_{o_m} > q_j$ and $\sum_{k=1}^{m} s_{o_k} \in [0,1]$. It follows that one root of $g(r)$ is always contained in the interval $(-\infty, s_{o_m})$, say $r^-$, and the other root is always contained in the interval $(s_{o_m}, \infty)$, say $r^+$.

Now since $c$ is negative both roots are positive so that the unique solution to $g(r) = 0$ on $[0, s_{o_m}]$ is $r^-$ and hence $r$ is identified. Once $r$ is known, all $\varphi_j$'s are derived according to the expression in Lemma B.1 applied to $d_1, \cdots, d_D$ at $x^0$.

We now turn to Default Option. Assume W.L.O.G that all $\varphi_j$'s are positive. Indeed, $\varphi_j = 0$ iff alternative $j$ is never chosen for any $x$ and hence is identified. As above, let $d_{o_m}$ be the first alternative in the sequence $L_{o_1}(x^1) \succ L_{o_2}(x^1) \succ \ldots$ such that $o_m \neq m$. Let $j$ be the position of alternative $o_m$ in the sequence $L_1(x^0) \succ L_2(x^0) \succ \ldots$. Note that $m < j$, $d_{o_m} = d_j$, and all lotteries that dominate $d_{o_m}$ at $x^1$ also dominate $d_j$ at $x^0$, since, by construction $o_1 = 1, o_2 = 2, o_3 = 3, \ldots, o_{m-1} = m-1$.

The assumptions are satisfied in Lemma $B.3$ for $d_j$ at $x^0$ and $d_{o_m}$ at $x^1$. It follows that:

$$\frac{(1-r)q_j}{1+(1-r)\sum_{k=1}^{j-1} q_k} = \varphi_j = \varphi_{o_m} = \frac{(1-r)s_{o_m}}{1+(1-r)\sum_{k=1}^{m-1} s_{o_k}}. \tag{B.1.2}$$

where $q_k \equiv Pr(d = d_k|x^0, d \in \{d_1, d_2, ..., d_D\})$ and $s_{o_k} \equiv Pr(d = d_{o_k}|x^1, d \in \{d_1, d_2, ..., d_D\})$.

If $\varphi_j = 0$, it is immediate that $r = 0$. On the other hand, if $\varphi_j > 0$, then Equation B.1.2 implies that

$$r = 1 - \frac{s_{o_m}\sum_{k=1}^{j-1} q_k - q_j \sum_{k=1}^{m-1} s_{o_k}}{s_{o_m} - q_j}.$$

Since $q_j < s_{o_m}$ and $\sum_{k=1}^{j-1} q_k > \sum_{k=1}^{m-1} s_{o_k}$, there is a unique $r \in [0, 1]$ that solves the Equation B.1.2. With known $r$, we can learn $\varphi_j$'s sequentially according to Lemma $B.3$: $\varphi_1 = (1-r)q_1$, $\varphi_2 = \frac{(1-r)q_2}{1-(1-r)q_1}$, and so on.

$\square$

**Proof of Theorem 3.3**

*Proof.* Let $d_n$ denote the default alternative so that it is $n^{th}$ best at $x^0$. Let $r = \prod_{k=1}^{D}(1-\varphi_k)$. We first show that $r$ is identified.

Let $d_j$ and (W.L.O.G) let $j > n$, so that $d_n \succ d_j$ by all DMs with characteristics $x^0$ regardless of their $\nu$. Let $o_{j'}$ and $o_{n'}$ index the position of $d_j$ and $d_n$ at $x^1$. That is, $d_j = d_{o_{j'}}$ and $d_j = d_{o_{n'}}$. By assumption we have $j' < n'$ so that $d_j$ is preferred to $d_n$ by any DM with characteristics $x^1$. The conditions for Lemma B.2 hold, so that

$$\frac{q_j}{r + 1 - \sum_{k=1}^{n-1} q_k} = \varphi_j = \varphi_{o_{j'}} = \frac{s_{o_{j'}}}{1 - \sum_{k=1}^{n'-1} s_{o_k}},$$

Solving for $r$ yields:

$$r = \frac{q_j \left(1 - \sum_{k=1}^{n'-1} s_{o_k}\right) - s_{o_{j'}} \left(1 - \sum_{k=1}^{n-1} q_k\right)}{s_{o_{j'}}}.$$

Note that, by assumption, $q_j, s_{o_{j'}} > 0$ so $r$ is well defined. Once $r$ is known, all $\varphi_j$'s are derived according to the expression in Lemma B.2 applied to $d_1, \ldots, d_D$ at $x^0$. □

**Proof of Theorem 3.4**

*Proof.* Let $q_j = Pr(d_j|x^0)$ and $s_{o_k} = Pr(d_{o_k}|x^1)$. We can learn $\underline{n}$ and $\varphi_j$ for all $j \leq \underline{n}$ as follows. By assumption, there are no preferred options among alternatives $d_1, \ldots, d_{\underline{n}-1}$. Hence,

1. $\varphi_1 = q_1$. If $\varphi_1 = 1$ then $\underline{n} = 1$. Otherwise, set $j = 2$ and proceed to Step 2.
2. $\varphi_j = \frac{q_j}{1-\sum_{k=1}^{j-1} q_k}$. If $\varphi_j = 1$ then $\underline{n} = j$. Otherwise, set $j = j + 1$ and repeat Step 2.

Repeating this argument for the moments evaluated at $x^1$, we find the first $n^*$ such that $\varphi_{o_{n^*}} = 1$ (i.e. $n^* = \min_{n' \in \{\underline{n}', \ldots, \overline{n}'\}} n'$ where $o_{\underline{n}'} = \underline{n}, \ldots, o_{\overline{n}'} = \overline{n}$) and $\varphi_{o_{j'}}$ for all $j' \leq n^*$.

To summarize we have identified $\varphi_j$ for all $j \leq \underline{n}$ and all $j \geq \overline{n}$ (since whenever $j \geq \overline{n}$ it also the case that $j' \leq n^*$ where $o_{j'} = j$). By assumption, Preferred Options are adjacent so that whenever $\underline{n} \leq n \leq \overline{n}$, $d_n$ is also Preferred Options and hence $\varphi_n = 1$. □

**Proof of Theorem 3.7**

*Proof.* We have that

$$Pr(d = d_D|x^1) = \phi_1 Q_{1,D}$$

$$Pr(d = d_{D-1}|x^1) = \phi_1 Q_{1,D-1} + \phi_2 Q_{2,D-1}$$

$$\vdots$$

$$Pr(d = d_1|x^1) = \phi_1 Q_{1,1} + \phi_2 Q_{2,1} + ... + \phi_D Q_{D,1}$$

The $Q$'s in the equations above are known and are strictly positive. It follows that $\phi$'s are identified sequentially. □

**Proof of Theorem 3.8**

*Proof.* We start with Assumption 3.2.A1-A3. Denote $\mathcal{A} = \{k : \varphi_k > 0\}$ under the ARC Model and $\mathcal{A} = \mathcal{D}$ under the RCL Model. Fix $j$ corresponding to Assumption 3.2.A1-A3 and denote $\underline{j}(j,t) \equiv \arg\min_{k\in\mathcal{A}-\{j\}} c_{j,k}(x(t))$. By Assumption 3.2.A3 and the continuity of $c(\cdot)$, the alternative corresponding to $\underline{j}(j,t)$ is the same for all $t \in [0,1]$. Thus, we can write WLOG $\underline{j} = \underline{j}(j,t)$. For any $t \in [0,1]$ we have

$$Pr(d = d_j|x(t)) = w_{j,\underline{j}} F(\underline{c}_j(x(t))) + \sum_{k\in\mathcal{A}-\{j,\underline{j}\}} w_{j,k} F(c_{j,k}(x(t))) + \hat{r_j},$$

where $\hat{r_j} \geq 0$, $w_{j,\underline{j}} > 0$ and all $w_{j,k} > 0$ are known functions of the limited consideration parameters. Since the latter are identified, so are $\hat{r}_j$, $w_{j,\underline{j}}$ and $w_{j,k}$, $k \neq j, \underline{j}$. Next, find the smallest $t_1$ such that $c_{j,k}(x(t_1)) \geq \bar{\nu}$ for all $k \in \mathcal{A} - \{j, \underline{j}\}$. In other words, $t_1$ is the smallest value of $t$ for which only the lowest cutoff is below the upper bound of the support. It follows that for any $t \in [t_1, 1]$,

$$Pr(d = d_j|x(t)) = w_{j,\underline{j}} F(\underline{c}_j(x(t))) + \hat{r}_j,$$

which implies that $F(\cdot)$ is identified for all $\nu \in [\nu_1, \bar{\nu}]$ where $\nu_1 \equiv \underline{c}_j(x(t_1))$. It is clear that if $\mathcal{A} - \{j, \underline{j}\} = \emptyset$ we are done. Otherwise, find the smallest $t_2$ such that $c_{j,k}(x(t_2)) \geq \nu_1$ for all $k \in \mathcal{A} - \{j, \underline{j}\}$. In other words, $t_2$ is the smallest value of $t$ for which only the lowest cutoff is below $\nu_1$. Since all other cutoffs lie in the region where $F(\cdot)$ is known, it follows that $F(\cdot)$ is identified for all $\nu \in [\nu_2, \nu_1]$, and, hence for all $\nu \in [\nu_2, \bar{\nu}]$, where $\nu_2 \equiv \underline{c}_j(x(t_2))$. Proceeding in this way we have that $F(\cdot)$ is identified over $[\nu_n, \bar{\nu}]$. $\{\nu_n\}$ is a strictly monotonically declining sequence defined recursively as

$$\begin{aligned}
\nu_n &\equiv \underline{c}_j(x(t_n)) \\
t_0 &= 1 \\
t_n &= \min_{t \in [0,1]} t \\
&\quad \text{s.t. } c_{j,k}(x(t)) \geq \nu_{n-1} \quad \forall k \in \mathcal{A} - \{j, \underline{j}\}.
\end{aligned}$$

This sequence either eventually converges to 0, crosses to the left of 0, or converges to some accumulation point $\nu^*$ in the interior of $[0, \bar{\nu}]$. In the former cases we have identification. We claim that the latter case cannot arise. For the purpose of obtaining a contradiction, suppose that $\nu^* = \lim_{n\to\infty} \nu_n > 0$. By continuity it follows that

$$\lim_{n\to\infty} \underline{c}_j(x(t_n)) = \underline{c}_j(x(t^*)),$$

where

$$\begin{aligned}
t^* &\equiv \lim_{n\to\infty} t_n \\
&= \lim_{n\to\infty} \left( \min_{t\in[0,1]} t \quad \text{s.t. } c_{j,k}(x(t)) \geq \nu_{n-1} \quad \forall k \in \mathcal{A} - \{j, \underline{j}\} \right) \\
&= \min_{t\in[0,1]} t \quad \text{s.t. } c_{j,k}(x(t)) \geq \lim_{n\to\infty} \nu_{n-1} \quad \forall k \in \mathcal{A} - \{j, \underline{j}\} \qquad \text{(by continuity)} \\
&= \min_{t\in[0,1]} t \quad \text{s.t. } c_{j,k}(x(t)) \geq \nu^* \quad \forall k \in \mathcal{A} - \{j, \underline{j}\}.
\end{aligned}$$

Now since the cutoffs are strictly decreasing in $t$, there is a $k \in \mathcal{A} - \{j, \underline{j}\}$ such that $c_{j,k}(x(t^*)) = \nu^*$. Putting this together we yield

$$c_{j,k}(x(t^*)) = \nu^* = \underline{c}_j(x(t^*)),$$

which contradicts Assumption 3.2.A3. Under Assumption 3.2.B1-3 the proof works in the exactly same way, only we start at the lower end of the preference-coefficient support. □

## B.2 Proofs of Properties

### Proof of Fact 1

*Proof.* Take any non empty consideration set $\mathcal{K}$. For a given preference coefficient $\nu$, let $j_{\mathcal{K}}(\nu)$ denote the identity of the best alternative in this consideration set. Because of the way alternatives are ordered, $j_{\mathcal{K}}(\nu)$ is an increasing step function. Hence, $I(j_{\mathcal{K}}(\nu) \leq J)$ is a decreasing step function, Note, that $Pr\left(\bigcup_{j=1}^{J} d_j \middle| x, \nu\right)$ is the sum of $I(j_{\mathcal{K}}(\nu) \leq J)$ weighted by the probability of $\mathcal{K}$ being drawn. Hence it is decreasing in $\nu$. □

### Proof of Fact 5

*Proof.* Suppose $U_\nu(d_j, x) > U_\nu(d_k, x)$, but $\varphi_k = 1$ and $\varphi_j = 0$. □

### Proof of Fact 6

*Proof.* Suppose $U_\nu(d_j, x) > U_\nu(d_k, x)$. If both $d_j$ and $d_k$ are in the consideration set, then $d_j$ will be chosen. For any consideration set that contains $d_k$ but not

$d_j$, there is an equal size consideration set that contains $d_j$ but not $d_k$, namely $(\mathcal{K} \cup d_j) \setminus d_k$. These two sets have the same probability of being formed. If $d_k$ is preferred to all other alternatives in $\mathcal{K}$, then the same is true for $d_j$ in $\mathcal{K} \cup d_j \setminus d_k$. Summing over all consideration sets delivers the result. □

## B.3 Verifying Identification in Our Application

We start by recalling that CARA and CRRA utility functions satisfy the following basic property (Pratt, 1964; Barseghyan et al., 2018):[1]

**Property B.1.** *For any $y_0 > y_1 > y_2 > 0$, the ratio $R(y_0, y_1, y_2) \equiv \frac{u_\nu(y_1)-u_\nu(y_2)}{u_\nu(y_0)-u_\nu(y_1)}$ is strictly increasing in $\nu$.*

It follows that CARA and CRRA utility functions also satisfy a slightly extended version of the property above:

**Property B.2.** *For any $y_0 > y_1 > y_2 > y_3 > 0$, the ratio $Q_\nu(y_0, y_1, y_2, y_3) \equiv \frac{u_\nu(y_2)-u_\nu(y_3)}{u_\nu(y_0)-u_\nu(y_1)}$ is strictly increasing in $\nu$.*

*Proof.*

$$
\begin{aligned}
Q_\nu(y_0, y_1, y_2, y_3) &= \frac{u_\nu(y_2) - u_\nu(y_3)}{u_\nu(y_0) - u_\nu(y_1)} = \frac{u_\nu(y_2) - u_\nu(y_3)}{u_\nu(y_1) - u_\nu(y_2)} \times \frac{u_\nu(y_1) - u_\nu(y_2)}{u_\nu(y_0) - u_\nu(y_1)} \\
&= R_\nu(y_1, y_2, y_3) R_\nu(y_0, y_1, y_2)
\end{aligned}
$$

□

[1] This property is equivalent to condition (e) in Pratt (1964) Theorem 1. As shown there, it is equivalent to assuming that an increase in $\nu$ corresponds to an increase in the coefficient of absolute risk aversion.

For our application, we show that $c_{1,2}(\bar{p}, \mu) < c_{1,m}(\bar{p}, \mu)$ for any $m > 2$ under both CARA and CRRA preferences.

**Theorem B.1.** *Under either CARA or CRRA expected utility preferences, the cutoff mappings satisfy $c_{1,2}(\bar{p}, \mu) < c_{1,m}(\bar{p}, \mu)$ for any $m > 2$.*

*Proof.* We start with CARA preferences. The existence and the uniqueness of $c_{j,k}(x)$ for all $j < k$ follows directly from the Property B.2. Indeed note that $p_j < p_k < p_k + d_k < p_j + d_j$.[2] At the cutoff the DM is indifferent between lotteries $j$ and $k$. Equating two expected utilities and rearranging we have that

$$\frac{e^{-\nu(w-p_k-d_k)} - e^{-\nu(w-p_j-d_j)}}{e^{-\nu(w-p_j)} - e^{-\nu(w-p_k)}} = \frac{1-\mu}{\mu}, \tag{B.3.1}$$

where $w$ is the DM's initial wealth. By Property B.2, the L.H.S. of Equation B.3.1 is strictly monotone in $\nu$, and it tends to $+\infty$ when $\nu$ goes to $+\infty$ and to zero when $\nu$ goes to $-\infty$. It follows that there exists a unique $\nu$, i.e the cutoff $c_{j,k}(x)$, that solves the Equation B.3.1. Moreover, since the L.H.S. is strictly monotone in $\nu$ it follows from the Implicit Function Theorem that $c_{j,k}(x)$ is continuous in $\mu$ and $\bar{p}$.

The next step is to establish that $c_{1,2}(\bar{p}, \mu) < c_{1,m}(\bar{p}, \mu)$, $m > 2$. First, note that the expected utility of lottery $k$ is proportional to

$$EU_\nu(L_k) \propto -e^{\nu p_k}\left(1 - \mu + \mu e^{\nu d_k}\right)$$

For the purpose of obtaining a contradiction, suppose that there exists $(\bar{p}, \mu)$ and an $m$ such that $c_{1,2}(\bar{p}, \mu) = c_{1,m}(\bar{p}, \mu)$. That is, there exists a $\nu = c_{1,2}(\bar{p}, \mu) = c_{1,m}(\bar{p}, \mu)$ such that

$$\frac{1-\mu+\mu e^{\nu d_1}}{1-\mu+\mu e^{\nu d_2}} e^{\nu(g_1-g_2)\bar{p}} = 1 = \frac{1-\mu+\mu e^{\nu d_1}}{1-\mu+\mu e^{\nu d_m}} e^{\nu(g_1-g_m)\bar{p}}$$

[2]If $p_k + d_k > p_j + d_j$, then alterantive $j$ first order stochastically dominates $k$ and hence the cuttoff is $+\infty$.

Taking logs for each side and rearranging we have that

$$\log\left(\frac{1-\mu+\mu e^{\nu d_1}}{1-\mu+\mu e^{\nu d_2}}\right) = -\nu(g_1-g_2)\bar{p}$$
$$\log\left(\frac{1-\mu+\mu e^{\nu d_1}}{1-\mu+\mu e^{\nu d_m}}\right) = -\nu(g_1-g_m)\bar{p}.$$

Dividing through we have that

$$\frac{\log\left(\frac{1-\mu+\mu e^{\nu d_1}}{1-\mu+\mu e^{\nu d_2}}\right)}{\log\left(\frac{1-\mu+\mu e^{\nu d_1}}{1-\mu+\mu e^{\nu d_m}}\right)} = \frac{g_1-g_2}{g_1-g_m}.$$

The R.H.S. is less than one. The L.H.S. is monotonically decreasing in $\mu < 1$. Indeed, denote $\hat{\mu} = \frac{1-\mu}{\mu}$, $\Delta_1 = e^{\nu d_1}$, $\Delta_2 = e^{\nu d_2}$, and $\Delta_m = e^{\nu d_m}$ to rewrite the L.H.S. as follows

$$f(\hat{\mu}) \equiv \frac{\log(\Delta_1+\hat{\mu}) - \log(\Delta_2+\hat{\mu})}{\log(\Delta_1+\hat{\mu}) - \log(\Delta_m+\hat{\mu})}.$$

We claim that the expression above is monotonically increasing in $\hat{\mu}$. Its derivative is equal to

$$\frac{f'(\hat{\mu})}{f(\hat{\mu})} = \left(\frac{1}{\Delta_1+\hat{\mu}} - \frac{1}{\Delta_2+\hat{\mu}}\right)\frac{1}{\log(\Delta_1+\hat{\mu}) - \log(\Delta_2+\hat{\mu})}$$
$$- \left(\frac{1}{\Delta_1+\hat{\mu}} - \frac{1}{\Delta_m+\hat{\mu}}\right)\frac{1}{\log(\Delta_1+\hat{\mu}) - \log(\Delta_m+\hat{\mu})}$$

Once more relabeling $\Lambda_1 = -\log(\Delta_1+\hat{\mu})$, $\Lambda_2 = -\log(\Delta_2+\hat{\mu})$ and $\Lambda_m = -\log(\Delta_m+\hat{\mu})$ we can write the above as

$$\frac{f'(\hat{\mu})}{f(\hat{\mu})} = \frac{e^{\Lambda_1}-e^{\Lambda_m}}{\Lambda_1-\Lambda_m} - \frac{e^{\Lambda_1}-e^{\Lambda_2}}{\Lambda_1-\Lambda_2}.$$

Since $\Lambda_1 < \Lambda_2 < \Lambda_m$ and exponential function is convex, we have that the expression above is positive. Hence the derivative of $f\left(\frac{1-\mu}{\mu}\right)$ W.R.T. $\mu$ is negative, and hence it achieves its lowest value at $\mu = 1$. When $\mu = 1$, the L.H.S. is equal to $\frac{d_1-d_2}{d_1-d_m}$. Hence, the question is whether the following equality may hold

$$\frac{d_1-d_2}{d_1-d_m} = \frac{g_1-g_2}{g_1-g_m}.$$

It naturally would hold in perfectly competitive markets where additional coverage is simply proportional to its price. In practice, however, one might expect that with some market power the prices increase faster than then coverage, which is exactly what we find in our data (as well as for a larger number of firms appearing in (Barseghyan et al., 2011)). Hence $c_{1,2}(\bar{p},\mu) \neq c_{1,m}(\bar{p},\mu)$, for $m > 2$. Since the cutoffs are continuous, it follows that $c_{1,2}(\bar{p},\mu) < c_{1,m}(\bar{p},\mu)$ for $m > 2$.

Under CRRA, $c_{j,k}(\bar{p},\mu)$ exist and are continuous exactly for the same reasons as under CARA. It remains to establish that $c_{1,2}(\bar{p},\mu) < c_{1,m}(\bar{p},\mu)$ for $m > 2$. Consider the following Taylor expansion for the CRRA Bernoulli utility function $u(w)$ about point $w - p_k$:

$$\begin{aligned} u_\nu(w) &\equiv \frac{w^{1-\nu}}{1-\nu} \\ &= \frac{(w-p_k)^{1-\nu}}{1-\nu} + \frac{w^{-\nu}}{1!}p_k - \nu\frac{w^{-\nu-1}}{2!}p_k^2 \\ &\quad + \nu(\nu+1)\frac{w^{-\nu-2}}{3!}p_k^3 - \nu(\nu+1)(\nu+2)\frac{w^{-\nu-1}}{4!}p_k^4 + \dots \end{aligned}$$

This can be written as follows

$$\frac{(w-p_k)^{1-\nu} - w^{1-\nu}}{w^{1-\nu}} = -\frac{1-\nu}{w}p_k + \frac{(1-\nu)(-\nu)}{2!w^2}p_k^2 - \frac{(1-\nu)(-\nu)(-\nu-1)}{3!w^3}p_k^3 + ...$$

Hence, we can write

$$EU_\nu(L_k) \propto (1-\mu)\sum_{t=1}^{\infty}\omega_t p_k^t + \mu\sum_{t=1}^{\infty}\omega_t\left(p_k + d_k\right)^t.$$

The coefficients $\omega_t \equiv (t!w^t)^{-1}\prod_{t'=0}^{t-1}(1-\nu-t')(-1)^t$ are negative for all $t$, so the two power series above are absolutely convergent. Hence, we take the element-wise difference between $EU_\nu(L_j)$ and $EU_\nu(L_k)$:

$$\begin{aligned} EU_\nu(L_j) - EU_\nu(L_k) \propto (1-\mu)\sum_{t=1}^{\infty}\omega_t\left(p_j^t - p_k^t\right) + \mu\sum_{t=1}^{\infty}\omega_t\left((p_j+d_j)^t - (p_k+d_k)^t\right) \\ = (p_j - p_k)(1-\mu)\sum_{t=1}^{\infty}\omega_t\sum_{h=0}^{t}p_j^h p_k^{t-h} + \\ + ((p_j-p_k)+(d_j-d_k))\,\mu\sum_{t=1}^{\infty}\omega_t\sum_{h=0}^{t}(p_j+d_j)^h(p_k+d_k)^{t-h} \end{aligned}$$

This implies that if $\nu = c_{1,2}(\bar{p},\mu) = c_{1,m}(\bar{p},\mu)$, $m > 2$ we must have that

$$\begin{aligned} \frac{p_1 - p_2}{p_1 - p_m} = & \frac{p_1 - p_2 + d_1 - d_2}{p_1 - p_m + d_1 - d_m}\times \\ & \frac{\sum_{t=1}^{\infty}\omega_t\sum_{h=0}^{t}(p_1+d_1)^h(p_2+d_2)^{t-h}}{\sum_{t=1}^{\infty}\omega_t\sum_{h=0}^{t}(p_1+d_1)^h(p_m+d_m)^{t-h}}\frac{\sum_{t=1}^{\infty}\omega_t\sum_{h=0}^{t}p_1^h p_m^{t-h}}{\sum_{t=1}^{\infty}\omega_t\sum_{h=0}^{t}p_1^h p_2^{t-h}} \end{aligned}$$

Note that $p_m > p_2$. More over, when $\nu = c_{1,2}(\bar{p},\mu) = c_{1,m}(\bar{p},\mu)$ it is also the case that $\nu = c_{1,2}(\bar{p},\mu) = c_{1,m}(\bar{p},\mu) = c_{2,m}(\bar{p},\mu)$.

For the cutoff $c_{2,m}(\bar{p},\mu)$ to be on the support it must be the case that $p_2 + d_2 > p_m + d_m$. Indeed otherwise we have that $p_m - p_2 > d_2 - d_m$, which is a violation of the first order stochastic dominance. Hence if we can show that

$$\frac{p_1 - p_2}{p_1 - p_m} < \frac{p_1 - p_2 + d_1 - d_2}{p_1 - p_m + d_1 - d_m},$$

we would arrive to a contradiction, since it would be mean that the LHS of the equation is smaller than the RHS. Re-arranging we have that

$$\begin{aligned} \frac{p_1 - p_m + d_1 - d_m}{p_1 - p_m} &< \frac{p_1 - p_2 + d_1 - d_2}{p_1 - p_2} \\ \frac{d_1 - d_m}{p_1 - p_m} &< \frac{d_1 - d_2}{p_1 - p_2} \\ \frac{p_1 - p_2}{p_1 - p_m} &< \frac{d_1 - d_2}{d_1 - d_m}. \end{aligned}$$

The latter inequality holds in the data, as discussed in the case of CARA.

□

## B.4 Monetary Cost of Limited Consideration

We view limited consideration as a process that constrains households from achieving their first-best alternative either because the market setting forces some alternatives to become more salient than others (e.g. agent effects) or because of time or psychological costs that prevent the household from evaluating all alternatives in the choice set. Regardless of the underlying mechanism(s) of limited consideration, we can quantify its *monetary* cost within our framework. We ask, *ceteris paribus*, how much money the households "leave on the table" when choosing deductibles in property insurance under limited consideration rather than under full consideration. This is likely to be a lower bound on actual monetary losses arising from limited consideration, because insurance companies might be exploiting sub-optimality of households choices when setting prices or choosing menus.

We measure the monetary costs of limited consideration as follows. For each household we compute (the expected value of) the certainty equivalent of the lottery associated with the households' optimal choice, as well as of the one associated with their choice under limited consideration.[3] We then take the difference between these certainty equivalent values and average them across all households in the sample. On average, we find that households lose \$49 dollars across the three deductibles because of limited consideration. See Table B.7 for variation conditional on demographic characteristics and insurance score. We also find wide dispersion in loss across households (see Figure B.7). In particular, the $10^{th}$ percentile of losses is \$30 and the $90^{th}$ is \$72.

[3]Certainty equivalent of the lottery is defined as the minimum amount they are willing to accept in lieu of the lottery. In our case, for alternative $j$, it is simply $ce_j \equiv \frac{1}{\nu}\ln[(1-\mu)\exp(\nu p_j) + \mu \exp(\nu(p_j + d_j))]$.

## B.5 Data

**Table B.1** Descriptive Statistics

| Variable | Mean | Std. Dev. | 1st % | 99th % |
|---|---|---|---|---|
| Age | 53.3 | 15.7 | 25.4 | 84.3 |
| Female | 0.40 | | | |
| Single | 0.22 | | | |
| Married | 0.55 | | | |
| Second Driver | 0.43 | | | |
| Insurance Score | 767 | 112 | 532 | 985 |

**Table B.2** Frequency of Deductible Choices Across Contexts

| Deductible | 1000 | 500 | 250 | 200 | 100 | 50 |
|---|---|---|---|---|---|---|
| Collision | 0.064 | 0.676 | 0.122 | 0.129 | 0.009 | |
| Comprehensive | 0.037 | 0.430 | 0.121 | 0.329 | 0.039 | 0.044 |
| Home | 0.176 | 0.559 | 0.262 | | 0.002 | |

**Table B.3** Deductible Rank Correlations Across Contexts

| | Collision | Comprehensive | Home |
|---|---|---|---|
| Collision | 1 | | |
| Comprehensive | 0.61 | 1 | |
| Home | 0.37 | 0.35 | 1 |

**Table B.4** Joint Distribution of Auto Deductibles

| | Comprehensive | | | | | |
|---|---|---|---|---|---|---|
| Collision | 1000 | 500 | 250 | 200 | 100 | 50 |
| 1000 | 3.71 | 1.93 | 0.18 | 0.44 | 0.05 | 0.04 |
| 500 | 0 | 40.99 | 6.46 | 17.84 | 1.27 | 1.00 |
| 250 | 0 | 0.04 | 5.42 | 4.55 | 1.28 | 0.94 |
| 200 | 0.01 | 0.05 | 0.03 | 9.99 | 1.07 | 1.78 |
| 100 | 0 | 0 | 0 | 0.04 | 0.23 | 0.66 |

The distribution is reported in percent.

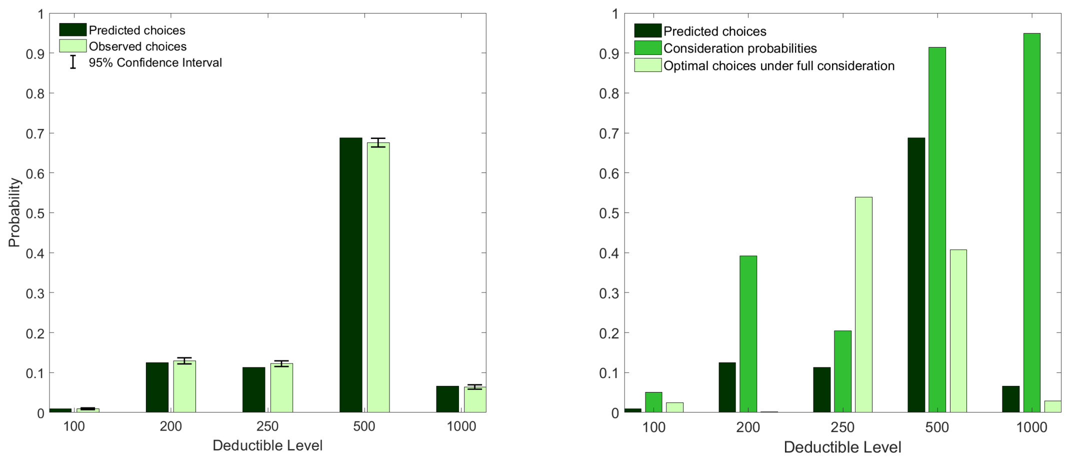
**Table B.5** Average Premiums Across Coverages

| Deductible | 1,000 | 500 | 250 | 200 | 100 | 50 |
|---|---|---|---|---|---|---|
| Collision | 145 | 187 | 243 | 285 | 321 | |
| Comprehensive | 94 | 117 | 147 | 155 | 178 | 224 |
| Home | 594 | 666 | 720 | | 885 | |

## B.6 Empirical Results: Figures and Tables

### B.6.1 Figures

**Figure B.1:** The ARC Model with Observable Demographics




The first panel reports the distribution of predicted and observed choices. The second panel displays consideration probabilities and the distribution of optimal choices under full consideration.

**Figure B.2:** The ARC Model with Observable Demographics: Conditional Distributions

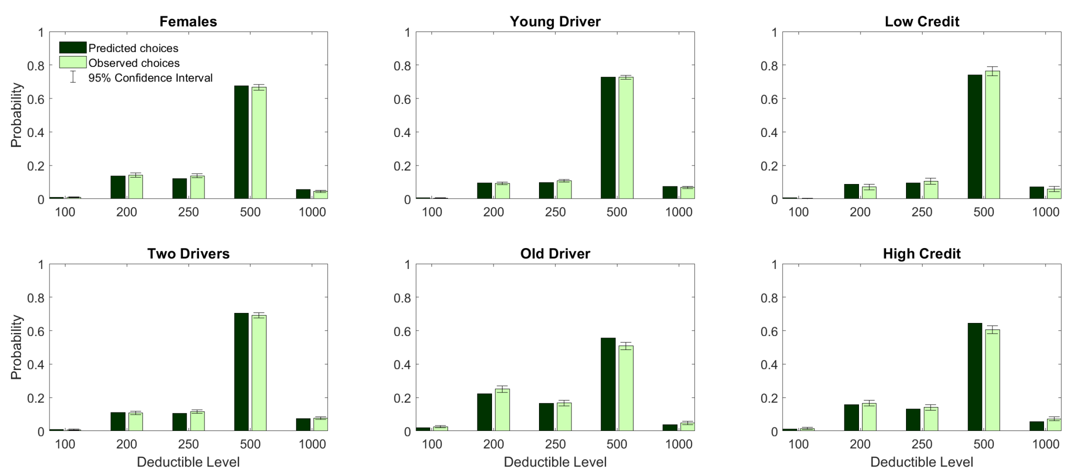


**Figure B.3:** The RCL Model and the RUM

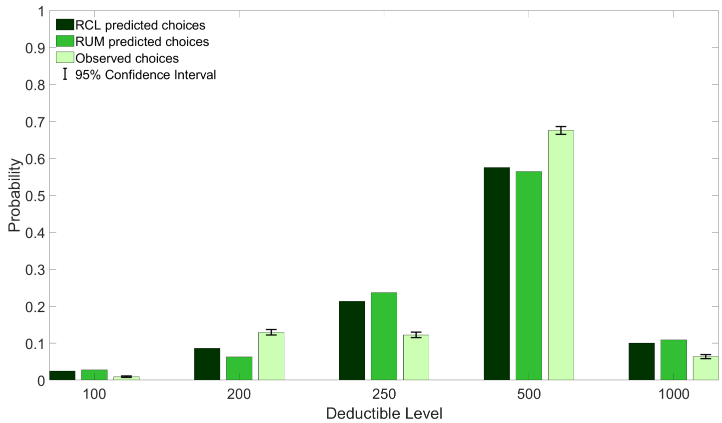



**Figure B.4:** The ARC Model, Three Coverages, "Narrow" Consideration

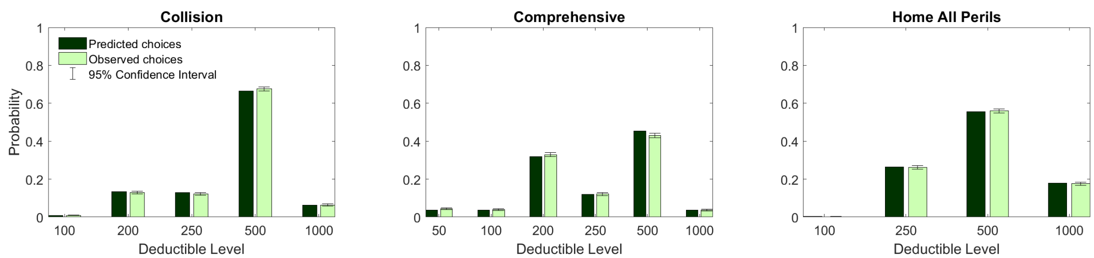
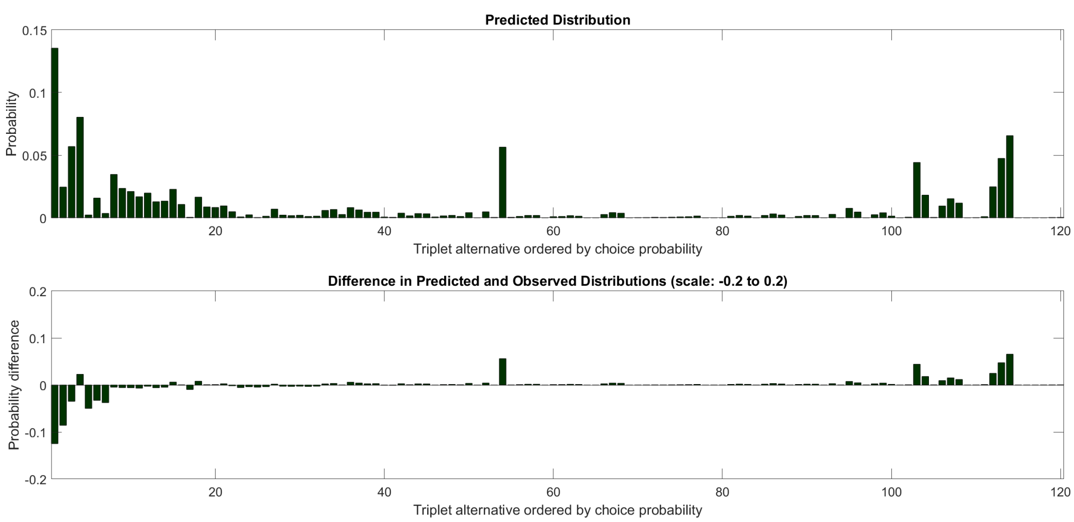
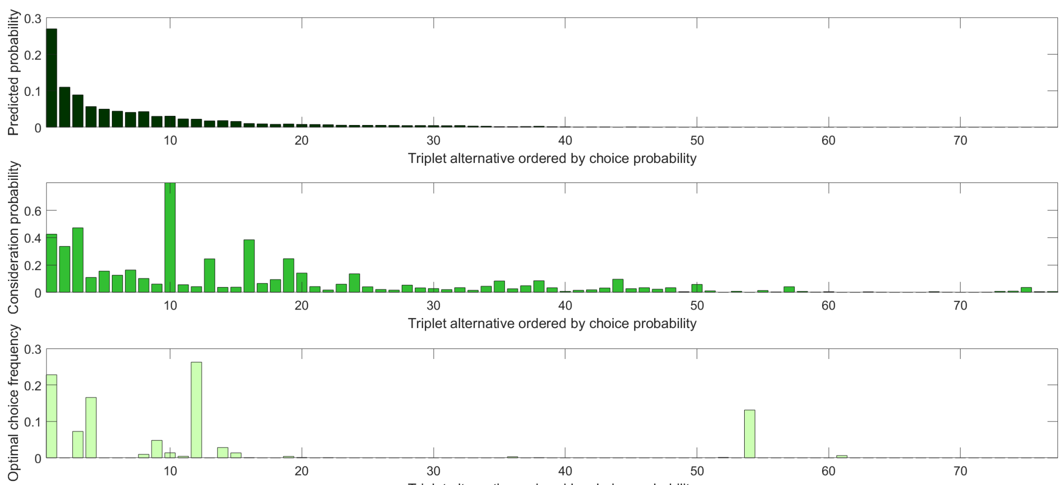



**Figure B.5:** The RUM, Three Coverages




Triplets are sorted by observed frequency at which they are chosen. The first panel reports the predicted choice frequency and the second panel reports the difference in predicted and observed choice frequencies.

**Figure B.6:** The ARC Model, Three Coverages:
Consideration and Optimal Choice Distribution




Triplets are sorted by observed frequency at which they are chosen.

**Figure B.7:** The ARC Model with Three Coverages:
Monetary Loss From Limited Consideration

### B.6.2 Tables

**Table B.6** MLE Estimation Results for the ARC Model

| | ARC Model | | ARC Model with Observables | |
|---|---|---|---|---|
| $\beta_1$ | 1.621 | [1.378, 1.948] | 2.090 | [1.556, 2.816] |
| $\beta_2$ | 7.319 | [5.946, 9.177] | 8.855 | [6.934, 11.758] |
| **Mean of** $\nu$ | 0.004 | [0.003, 0.004] | 0.004 | [0.003, 0.004] |
| **SD of** $\nu$ | 0.002 | [0.002, 0.003] | 0.002 | [0.002, 0.002] |
| **Intercept** | - | - | $-1.432$ | [-1.600, -1.302] |
| **Age** | - | - | 0.211 | [0.149, 0.298] |
| **Age**$^2$ | - | - | 0.047 | [-0.002, 0.106] |
| **Female Driver** | - | - | 0.075 | [0.019, 0.145] |
| **Single Driver** | - | - | 0.050 | [-0.011, 0.114] |
| **Married Driver** | - | - | 0.102 | [0.022, 0.196] |
| **Credit Score** | - | - | 0.137 | [0.078, 0.199] |
| **2+ Drivers** | - | - | $-0.310$ | [-0.479, -0.155] |
| **Collision $100** | 0.059 | [0.041, 0.081] | 0.051 | [0.033, 0.071] |
| **Collision $200** | 0.414 | [0.371, 0.465] | 0.392 | [0.344, 0.453] |
| **Collision $250** | 0.207 | [0.190, 0.224] | 0.205 | [0.188, 0.227] |
| **Collision $500** | 0.918 | [0.904, 0.931] | 0.915 | [0.896, 0.927] |
| **Collision $1000** | 1.000 | [0.972, 1.000] | 0.949 | [0.690, 1.000] |

Note: $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01

**Table B.7** MLE Estimation Results for the RCL and RUM Models

| | **RCL Model** | | **RUM Model** | |
|---|---|---|---|---|
| $\beta_1$ | 6.330 | [5.085, 8.993] | 8.401 | [6.794, 10.650] |
| $\beta_2$ | 100.232 | [80.526, 142.598] | 122.603 | [102.578, 152.511] |
| **Mean of $\nu$** | 0.001 | [0.001, 0.001] | 0.001 | [0.001, 0.001] |
| **SD of $\nu$** | 0.0005 | [0.0004, 0.0005] | 0.0004 | [0.0004, 0.0005] |
| **Intercept** | $-2.569$ | [-2.656, -2.479] | $-2.647$ | [-2.713, -2.586] |
| **Age** | $-0.142$ | [-0.198, -0.090] | $-0.146$ | [-0.178, -0.118] |
| **Age$^2$** | $-0.047$ | [-0.098, 0.003] | $-0.026$ | [-0.051, -0.002] |
| **Female Driver** | 0.011 | [-0.039, 0.064] | $-0.004$ | [-0.032, 0.025] |
| **Single Driver** | 0.038 | [-0.014, 0.092] | $-0.010$ | [-0.039, 0.020] |
| **Married Driver** | 0.027 | [-0.044, 0.101] | $-0.031$ | [-0.069, 0.009] |
| **Credit Score** | 0.232 | [0.180, 0.288] | 0.096 | [0.073, 0.124] |
| **2+ Drivers** | $-0.390$ | [-0.535, -0.233] | $-0.021$ | [-0.101, 0.061] |
| **Attention Level 1** | 0.031 | [0.019, 0.046] | - | - |
| **Attention Level 2** | 0.509 | [0.447, 0.536] | - | - |
| **Attention Level 3** | 0.000 | [0.000, 0.086] | - | - |
| **Attention Level 4** | 0.000 | [0.000, 0.000] | - | - |
| **Full Attention** | 0.460 | [0.423, 0.484] | - | - |
| **Sigma** | - | - | 0.040 | [0.036, 0.043] |

Note: $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01

**Table B.8** MLE Estimation Results for the ARC Model, Three Coverages: "Narrow" Consideration

| | ARC Model | |
|---|---|---|
| $\beta_1$ | 1.152 | [1.010, 1.284] |
| $\beta_2$ | 3.141 | [2.639, 3.694] |
| **Mean of** $\nu$ | 0.005 | [0.005, 0.006] |
| **SD of** $\nu$ | 0.004 | [0.004, 0.004] |
| **Intercept** | $-1.127$ | [-1.225, -1.032] |
| **Age** | 0.198 | [0.164, 0.235] |
| **Age**$^2$ | 0.090 | [0.059, 0.121] |
| **Female Driver** | 0.052 | [0.018, 0.088] |
| **Single Driver** | 0.004 | [-0.037, 0.047] |
| **Married Driver** | 0.008 | [-0.038, 0.062] |
| **Credit Score** | 0.110 | [0.077, 0.145] |
| **2+ Drivers** | $-0.089$ | [-0.186, 0.004] |
| **Collision $100** | 0.033 | [0.023, 0.043] |
| **Collision $200** | 0.324 | [0.299, 0.351] |
| **Collision $250** | 0.199 | [0.185, 0.216] |
| **Collision $500** | 0.953 | [0.945, 0.960] |
| **Collision $1000** | 1.000 | [0.870, 1.000] |
| **Comprehensive $50** | 1.000 | [1.000, 1.000] |
| **Comprehensive $100** | 0.337 | [0.291, 0.384] |
| **Comprehensive $200** | 0.765 | [0.744, 0.790] |
| **Comprehensive $250** | 0.325 | [0.295, 0.357] |
| **Comprehensive $500** | 0.892 | [0.853, 0.928] |
| **Comprehensive $1000** | 0.277 | [0.226, 0.316] |
| **Home $100** | 0.002 | [0.000, 0.010] |
| **Home $250** | 0.387 | [0.368, 0.409] |
| **Home $500** | 0.859 | [0.844, 0.877] |
| **Home $1000** | 0.824 | [0.774, 0.873] |

Note: $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01

**Table B.9** MLE Estimation Results for the ARC Model, Three Coverages

| ARC Model | | | ARC Model (cont.) | | |
|---|---|---|---|---|---|
| $\beta_1$ | 4.515 | [3.432, 6.255] | **(250,200,250)** | 0.037 | [0.029, 0.045] |
| $\beta_2$ | 23.623 | [17.528, 33.251] | **(250,200,500)** | 0.056 | [0.046, 0.067] |
| **Mean of $\nu$** | 0.003 | [0.003, 0.003] | **(250,200,1000)** | 0.045 | [0.025, 0.067] |
| **SD of $\nu$** | 0.001 | [0.001, 0.002] | **(250,250,100)** | 0.001 | [0.000, 0.005] |
| **Intercept** | −1.706 | [-1.792, -1.623] | **(250,250,250)** | 0.042 | [0.035, 0.050] |
| **Age** | 0.166 | [0.130, 0.207] | **(250,250,500)** | 0.061 | [0.051, 0.070] |
| **Age$^2$** | 0.041 | [0.011, 0.073] | **(250,250,1000)** | 0.026 | [0.011, 0.044] |
| **Female Driver** | 0.043 | [0.006, 0.079] | **(250,500,500)** | 0.0007 | [0.000, 0.002] |
| **Single Driver** | 0.011 | [-0.028, 0.052] | **(500,50,250)** | 0.034 | [0.020, 0.049] |
| **Married Driver** | 0.031 | [-0.020, 0.085] | **(500,50,500)** | 0.053 | [0.032, 0.077] |
| **Credit Score** | 0.141 | [0.108, 0.175] | **(500,50,1000)** | 0.034 | [0.007, 0.074] |
| **2+ Drivers** | −0.099 | [-0.196, -0.0004] | **(500,100,250)** | 0.015 | [0.009, 0.022] |
| **(100,50,250)** | 0.041 | [0.026, 0.059] | **(500,100,500)** | 0.042 | [0.029, 0.059] |
| **(100,50,500)** | 0.015 | [0.005, 0.029] | **(500,100,1000)** | 0.049 | [0.022, 0.081] |
| **(100,50,1000)** | 0.013 | [0.000, 0.043] | **(500,200,100)** | 0.008 | [0.000, 0.019] |
| **(100,100,100)** | 0.002 | [0.000, 0.010] | **(500,200,250)** | 0.125 | [0.109, 0.142] |
| **(100,100,250)** | 0.008 | [0.002, 0.014] | **(500,200,500)** | 0.336 | [0.305, 0.370] |
| **(100,100,500)** | 0.005 | [0.000, 0.011] | **(500,200,1000)** | 0.245 | [0.202, 0.296] |
| **(100,100,1000)** | 0.005 | [0.000, 0.019] | **(500,250,100)** | 0.002 | [0.000, 0.008] |
| **(100,200,250)** | 0.0006 | [0.000, 0.003] | **(500,250,250)** | 0.038 | [0.030, 0.046] |
| **(100,200,500)** | 0.0008 | [0.000, 0.003] | **(500,250,500)** | 0.101 | [0.088, 0.118] |
| **(100,200,1000)** | 0.004 | [0.000, 0.016] | **(500,250,1000)** | 0.094 | [0.066, 0.123] |
| **(200,50,100)** | 0.011 | [0.000, 0.025] | **(500,500,100)** | 0.003 | [0.000, 0.011] |
| **(200,50,250)** | 0.065 | [0.047, 0.088] | **(500,500,250)** | 0.109 | [0.097, 0.122] |
| **(200,50,500)** | 0.060 | [0.039, 0.082] | **(500,500,500)** | 0.426 | [0.399, 0.454] |
| **(200,50,1000)** | 0.034 | [0.007, 0.073] | **(500,500,1000)** | 0.472 | [0.435, 0.512] |
| **(200,100,100)** | 0.002 | [0.000, 0.009] | **(1000,50,250)** | 0.008 | [0.000, 0.033] |
| **(200,100,250)** | 0.021 | [0.013, 0.030] | **(1000,50,500)** | 0.009 | [0.000, 0.040] |
| **(200,100,500)** | 0.028 | [0.018, 0.039] | **(1000,50,1000)** | 0.036 | [0.000, 0.150] |
| **(200,100,1000)** | 0.023 | [0.005, 0.048] | **(1000,100,250)** | 0.005 | [0.000, 0.022] |
| **(200,200,100)** | 0.002 | [0.000, 0.007] | **(1000,100,500)** | 0.006 | [0.000, 0.028] |
| **(200,200,250)** | 0.155 | [0.133, 0.178] | **(1000,100,1000)** | 0.041 | [0.000, 0.126] |
| **(200,200,500)** | 0.163 | [0.140, 0.189] | **(1000,200,250)** | 0.032 | [0.007, 0.060] |
| **(200,200,1000)** | 0.135 | [0.090, 0.188] | **(1000,200,500)** | 0.083 | [0.042, 0.135] |
| **(200,250,250)** | 0.0004 | [0.000, 0.001] | **(1000,200,1000)** | 0.096 | [0.021, 0.195] |
| **(200,250,500)** | 0.0005 | [0.000, 0.002] | **(1000,250,250)** | 0.007 | [0.000, 0.022] |
| **(200,500,250)** | 0.002 | [0.000, 0.004] | **(1000,250,500)** | 0.027 | [0.006, 0.057] |
| **(200,1000,1000)** | 0.005 | [0.000, 0.024] | **(1000,250,1000)** | 0.058 | [0.000, 0.134] |
| **(250,50,100)** | 0.002 | [0.000, 0.009] | **(1000,500,250)** | 0.033 | [0.012, 0.060] |
| **(250,50,250)** | 0.020 | [0.013, 0.030] | **(1000,500,500)** | 0.141 | [0.095, 0.188] |
| **(250,50,500)** | 0.033 | [0.021, 0.047] | **(1000,500,1000)** | 0.384 | [0.297, 0.492] |
| **(250,100,250)** | 0.017 | [0.012, 0.023] | **(1000,1000,250)** | 0.085 | [0.037, 0.143] |
| **(250,100,500)** | 0.016 | [0.010, 0.023] | **(1000,1000,500)** | 0.246 | [0.180, 0.324] |
| **(250,100,1000)** | 0.019 | [0.004, 0.037] | **(1000,1000,1000)** | 0.808 | [0.627, 1.000] |
| **(250,200,100)** | 0.001 | [0.000, 0.005] | | | |

Note: $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01

**Table B.10** MLE Estimation Results for RUM, Three Coverages

| | RUM Model | |
|---|---|---|
| $\beta_1$ | 4.363 | [3.953, 4.840] |
| $\beta_2$ | 51.093 | [47.265, 55.484] |
| **Mean of** $\nu$ | 0.002 | [0.002, 0.002] |
| **SD of** $\nu$ | 0.0007 | [0.0007, 0.0007] |
| **Intercept** | $-2.422$ | [-2.469, -2.379] |
| **Age** | $-0.081$ | [-0.103, -0.059] |
| **Age$^2$** | $-0.016$ | [-0.032, 0.002] |
| **Female Driver** | 0.0007 | [-0.018, 0.018] |
| **Single Driver** | $-0.015$ | [-0.034, 0.005] |
| **Married Driver** | $-0.018$ | [-0.047, 0.009] |
| **Credit Score** | 0.037 | [0.020, 0.055] |
| **2+ Drivers** | $-0.049$ | [-0.100, -0.0001] |
| **Sigma** | 0.223 | [0.201, 0.249] |

Note: $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01

**Table B.11** Expected Monetary Loss by Group

| | Expected Monetary Loss | |
|---|---|---|
| **All** | $-49.1$ | [-55.3, -44.7] |
| **Female Driver** | $-53.2$ | [-59.9, -48.0] |
| **Single Driver** | $-44.1$ | [-49.7, -40.2] |
| **Young** | $-44.4$ | [-49.1, -40.9] |
| **Old** | $-64.6$ | [-76.8, -56.1] |
| **Low Credit Driver** | $-46.3$ | [-51.4, -42.5] |
| **High Credit Driver** | $-53.6$ | [-62.0, -47.6] |

Note: $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01

## REFERENCES


Abaluck, J., & Adams, A. (2017). What do consumers consider before they choose? identification from asymmetric demand responses.

Acemoglu, D., Carvalho, V. M., Ozdaglar, A., & Tahbaz-Salehi, A. (2012). The network origins of aggregate fluctuations. *Econometrica*, *80*(5), 1977–2016.

Ackerberg, D. A. (2009). A new use of importance sampling to reduce computational burden in simulation estimation. *Quantitative Marketing and Economics*, *7*(4), 343–376.

Apesteguia, J., & Ballester, M. A. (2018). Monotone stochastic choice models: The case of risk and time preferences. *Journal of Political Economy*, *126*(1), 74–106.

Apesteguia, J., Ballester, M. A., & Lu, J. (2017). Single-crossing random utility models. *Econometrica*, *85*(2), 661–674.

Artstein, Z. (1983). Distributions of random sets and random selections. *Israel Journal of Mathematics*, *46*(4), 313–324.

Athey, S. (2001). Single crossing properties and the existence of pure strategy equilibria in games of incomplete information. *Econometrica*, *69*(4), 861–889.

Aumann, R. J. (1965). Integrals of set-valued functions. *Journal of Mathematical Analysis and Applications*, *12*(1), 1–12.

Austin, T. (2008). On exchangeable random variables and the statistics of large graphs and hypergraphs. *Probability Surveys*, *5*, 80–145.

Badev, A. (2018). Nash equilibria on (un)stable networks. *arXiv preprint arXiv:1901.00373*.

Bala, V., & Goyal, S. (2000). A noncooperative model of network formation. *Econometrica*, *68*(5), 1181–1229.

Ballester, C., Calvó-Armengol, A., & Zenou, Y. (2006). Who's who in networks. wanted: The key player. *Econometrica*, *74*(5), 1403–1417.

Banerjee, A., Chandrasekhar, A. G., Duflo, E., & Jackson, M. O. (2013). The diffusion of microfinance. *Science*, *341*(6144), 1236498.

Banerjee, A., Chandrasekhar, A. G., Duflo, E., & Jackson, M. O. (2014). *Gossip: Identifying central individuals in a social network* (Tech. Rep.). National Bureau of Economic Research.

Barseghyan, L., Coughlin, M., Molinari, F., & Teitelbaum, J. (2019). Heterogeneous choice sets and preferences. *Working paper, Cornell University*.

Barseghyan, L., Molinari, F., O'Donoghue, T., & Teitelbaum, J. C. (2013). The nature of risk preferences: Evidence from insurance choices. *American Economic Review*, *103*(6), 2499-2529.

Barseghyan, L., Molinari, F., O'Donoghue, T., & Teitelbaum, J. C. (2018). Estimating risk preferences in the field. *Journal of Economic Literature*, *56*(2), 501-64.

Barseghyan, L., Molinari, F., & Teitelbaum, J. C. (2016). Inference under stability of risk preferences. *Quantitative Economics*, *7*(2), 367-409.

Barseghyan, L., Prince, J., & Teitelbaum, J. C. (2011). Are risk preferences stable across contexts? Evidence from insurance data. *American Economic Review*, *101*(2), 591-631.

Battaglini, M., & Patacchini, E. (2018). Influencing connected legislators. *Journal of Political Economy*, *126*(6), 2277–2322.

Battaglini, M., Patacchini, E., & Rainone, E. (2019). *Endogenous social connections in legislatures* (Tech. Rep.). National Bureau of Economic Research.

Bavelas, A. (1950). Communication patterns in task-oriented groups. *The Journal of the Acoustical Society of America*, *22*(6), 725–730.

Ben-Akiva, M., & Boccara, B. (1995). Discrete choice models with latent choice sets. *International Journal of Research in Marketing*, *12*(1), 9–24.

Beresteanu, A., Molchanov, I., & Molinari, F. (2011). Sharp identification regions in models with convex moment predictions. *Econometrica*, *79*(6), 1785–1821.

Berry, S., & Jia, P. (2008). Tracing the woes: An empirical analysis of the airline industry.

Bhargava, S., Loewenstein, G., & Sydnor, J. (2017). Choose to lose: Health plan choices from a menu with dominated options. *Quarterly Journal of Economics*, *132*(3), 1319-1372.

Bloch, F., Jackson, M. O., & Tebaldi, P. (2019). Centrality measures in networks. *Available at SSRN 2749124*.

Boggs, P. T., & Tolle, J. W. (1995). Sequential quadratic programming. *Acta numerica*, *4*, 1–51.

Bonacich, P. (1987). Power and centrality: A family of measures. *American journal of sociology*, *92*(5), 1170–1182.

Boucher, V. (2018). Equilibrium homophily in networks. *Available at SSRN 3264071*.

Boucher, V., & Houndetoungan, A. (2019). Estimating peer effects using partial network data. *Draft avaliable at houndetoungan.wixsite.com/aristide/research.*

Bramoullé, Y., & Genicot, G. (2018). Diffusion centrality: Foundations and extensions.

Bykhovskaya, A. (2019). Evolution of networks: Prediction and estimation.

*Available at SSRN 3351454*.

Calvó-Armengol, A., & İlkılıç, R. (2009). Pairwise-stability and nash equilibria in network formation. *International Journal of Game Theory*, *38*(1), 51–79.

Calvó-Armengol, A., Patacchini, E., & Zenou, Y. (2009). Peer effects and social networks in education. *The Review of Economic Studies*, *76*(4), 1239–1267.

Caplin, A., Dean, M., & Leahy, J. (2018). Rational inattention, optimal consideration sets, and stochastic choice. *The Review of Economic Studies*.

Cattaneo, M. D., Ma, X., Masatlioglu, Y., & Suleymanov, E. (2017). A random attention model. *arXiv preprint arXiv:1712.03448*.

Chamberlain, G. (1986). Asymptotic efficiency in semi-parametric models with censoring. *Journal of Econometrics*, *32*(2), 189–218.

Chandrasekhar, A. (2016). Econometrics of network formation. *The Oxford Handbook of the Economics of Networks*, 303–357.

Chandrasekhar, A., & Lewis, R. (2016). Econometrics of sampled networks. *Unpublished manuscript, MIT.[422]*.

Chatterjee, K., & Dutta, B. (2016). Credibility and strategic learning in networks. *International Economic Review*, *57*(3), 759–786.

Chernoff, H., & Teicher, H. (1958). A central limit theorem for sums of interchangeable random variables. *The Annals of Mathematical Statistics*, *29*(1), 118–130.

Chetty, R., & Szeidl, A. (2007, 05). Consumption Commitments and Risk Preferences. *The Quarterly Journal of Economics*, *122*(2), 831-877.

Chiappori, P.-A., Salanié, B., Salanié, F., & Gandhi, A. (2018). From aggregate betting data to individual risk preferences. *Forthcoming in Econometrica*.

Christakis, N. A., Fowler, J. H., Imbens, G. W., & Kalyanaraman, K. (2010). *An empirical model for strategic network formation* (Tech. Rep.). National

Bureau of Economic Research.

Ciliberto, F., & Tamer, E. (2009). Market structure and multiple equilibria in airline markets. *Econometrica*, *77*(6), 1791–1828.

Cohen, A., & Einav, L. (2007). Estimating risk preferences from deductible choice. *American Economic Review*, *97*(3), 745-788.

Conlon, C. T., & Mortimer, J. H. (2013). Demand estimation under incomplete product availability. *American Economic Journal: Microeconomics*, *5*(4).

Crawford, G., Grithz, R., & Iariax, A. (2017). Demand estimation with unobserved choice set heterogeneity. *Work in progress*.

Currarini, S., Jackson, M. O., & Pin, P. (2009). An economic model of friendship: Homophily, minorities, and segregation. *Econometrica*, *77*(4), 1003–1045.

Dardanoni, V., Manzini, P., Mariotti, M., & Tyson, C. J. (2017). Inferring cognitive heterogeneity from aggregate choices. *Work in progress, University of St Andrews*.

de Finetti, B. (1929). Funzione caratteristica di un fenomeno aleatorio. In *Atti del congresso internazionale dei matematici: Bologna del 3 al 10 de settembre di 1928* (pp. 179–190).

de Martí, J., & Zenou, Y. (2015). Network games with incomplete information. *Journal of Mathematical Economics*, *61*, 221–240.

Denuit, M., Maréchal, X., Pitrebois, S., & Walhin, J.-F. (2007). John Wiley & Sons.

de Paula, A. (2017). Econometrics of network models. In *Advances in economics and econometrics: Theory and applications, eleventh world congress* (pp. 268–323).

de Paula, A. (2019). Econometric models of network formation. arXiv:1910.07781.

de Paula, A., Rasul, I., & Souza, P. (2018). Recovering social networks from panel data: identification, simulations and an application.

de Paula, A., Richards-Shubik, S., & Tamer, E. T. (2015). Identification of preferences in network formation games. *Available at SSRN 2577410*.

Draganska, M., & Klapper, D. (2011). Choice set heterogeneity and the role of advertising: An analysis with micro and macro data. *Journal of Marketing Research*, *48*(4), 653-669.

Easley, D., & Kleinberg, J. (2010). *Networks, crowds, and markets* (Vol. 8). Cambridge university press Cambridge.

Echenique, F., Fryer Jr, R. G., & Kaufman, A. (2006). Is school segregation good or bad? *American Economic Review*, *96*(2), 265–269.

Einav, L., Finkelstein, A., Pascu, I., & Cullen, M. R. (2012). How general are risk preferences? Choice under uncertainty in different domains. *American Economic Review*, *102*(6), 2606–2638.

Eliaz, K., & Spiegler, R. (2011). On the strategic use of attention grabbers. *Theoretical Economics*, *6*(1), 127–155.

Gautier, E., & Rose, C. (2016). *Inference in social effects when the network is sparse and unknown.* preparation.

Gaynor, M., Propper, C., & Seiler, S. (2016). Free to choose? Reform, choice, and consideration sets in the english national health service. *American Economic Review*, *106*(11), 3521-57.

Goeree, M. S. (2008). Limited information and advertising in the us personal computer industry. *Econometrica*, *76*(5), 1017–1074.

Goldsmith-Pinkham, P., & Imbens, G. W. (2013). Social networks and the identification of peer effects. *Journal of Business & Economic Statistics*, *31*(3), 253–264.

Graham, B. S. (2014). Methods of identification in social networks.

Graham, B. S. (2016). *Homophily and transitivity in dynamic network formation* (Tech. Rep.). National Bureau of Economic Research.

Graham, B. S. (2019). Network data. , *7A*.

Gualdani, C. (2019). An econometric model of network formation with an application to board interlocks between firms. *Forthcoming at Journal of Econometrics*.

Handel, B. R. (2013). Adverse selection and inertia in health insurance markets: When nudging hurts. *American Economic Review*, *103*(7), 2643-2682.

Harris, K. M. (2009). The national longitudinal study of adolescent to adult health (add health), waves i & ii, 1994–1996; wave iii, 2001–2002; wave iv, 2007-2009. *Chapel Hill, NC: Carolina Population Center, University of North Carolina at Chapel Hill*.

Hauser, J. R., & Wernerfelt, B. (1990). An evaluation cost model of consideration sets. *Journal of Consumer Research*, *16*(4), 393-408.

Heckman, J. (1990). Varieties of selection bias. *The American Economic Review*, *80*(2), 313–318.

Heiss, F., McFadden, D., Winter, J., Wuppermann, A., & Zhou, B. (2016). Inattention and switching costs as sources of inertia in medicare part d. *NBER WP 22765*.

Hellmann, T. (2013). On the existence and uniqueness of pairwise stable networks. *International Journal of Game Theory*, *42*(1), 211–237.

Ho, K., Hogan, J., & Scott Morton, F. (2017). The impact of consumer inattention on insurer pricing in the medicare part d program. *The RAND Journal of Economics*, *48*(4), 877-905.

Honka, E., & Chintagunta, P. (2017). Simultaneous or sequential? search strategies in the u.s. auto insurance industry. *Marketing Science*, *36*(1), 21-42.

Hortaçsu, A., Madanizadeh, S. A., & Puller, S. L. (2017). Power to choose? an analysis of consumer inertia in the residential electricity market. *American Economic Journal: Economic Policy*, *9*(4), 192-226.

Howard, J. A. (1977). *Consumer behavior: Application of theory.* New York: McGraw-Hill.

Hsieh, C.-S., Lee, L.-F., & Boucher, V. (2019). Specification and estimation of network formation and network interaction models with the exponential probability distribution. *Available at SSRN 2967867*.

Huisman, M. (2014). Imputation of missing network data: some simple procedures. *Encyclopedia of Social Network Analysis and Mining*, 707–715.

Jackson, M. O. (2019). The friendship paradox and systematic biases in perceptions and social norms. *Journal of Political Economy*, *127*(2), 777–818.

Jackson, M. O., et al. (2008). *Social and economic networks* (Vol. 3). Princeton university press Princeton.

Jackson, M. O., & Watts, A. (2001). The existence of pairwise stable networks.

Jackson, M. O., & Wolinsky, A. (1996). A strategic model of social and economic networks. *Journal of economic theory*, *71*(1), 44–74.

Jia, P. (2008). What happens when wal-mart comes to town: An empirical analysis of the discount retailing industry. *Econometrica*, *76*(6), 1263–1316.

Kaido, H., Molinari, F., & Stoye, J. (2019). Confidence intervals for projections of partially identified parameters. *Econometrica*, *87*(4), 1397–1432.

Kaido, H., Molinari, F., Stoye, J., & Thirkettle, M. (2017). Calibrated projection in matlab: Users' manual. *arXiv preprint arXiv:1710.09707*.

Kallenberg, O. (2006). *Probabilistic symmetries and invariance principles.* Springer Science & Business Media.

Katz, L. (1953). A new status index derived from sociometric analysis. *Psychometrika*, *18*(1), 39–43.

Kimya, M. (2018). Choice, consideration sets and attribute filters. *American Economic Journal: Microeconomics*.

Krause, R. W., Huisman, M., Steglich, C., & Sniiders, T. A. (2018). Missing network data a comparison of different imputation methods. In *2018 ieee/acm international conference on advances in social networks analysis and mining (asonam)* (pp. 159–163).

Lee, W., Fosdick, B., & McCormick, T. (2018). Inferring social structure from continuous-time interaction data. *Applied stochastic models in business and industry*, *34*(2), 87–104.

Leung, M. (2015a). A random-field approach to inference in large models of network formation. *Available at SSRN 2520272*.

Leung, M. (2015b). Two-step estimation of network-formation models with incomplete information. *Journal of Econometrics*, *188*(1), 182–195.

Lewbel, A. (2000). Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables. *Journal of Econometrics*, *97*(1), 145 – 177.

Lewbel, A. (2014). An overview of the special regressor method. *Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*, 38 – 62.

Lewbel, A. (2019). The identification zoo: Meanings of identification in econometrics. *Journal of Economic Literature*, *57*(4), 835–903.

Lewbel, A., Qu, X., & Tang, X. (2019). Social networks with misclassified or unobserved links.

Liu, X. (2013). Estimation of a local-aggregate network model with sampled networks. *Economics Letters*, *118*(1), 243–246.

Manresa, E. (2016). Estimating the structure of social interactions using panel data. *Unpublished Manuscript. CEMFI, Madrid*.

Manski, C. F. (1975). Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics*, *3*(3), 205 – 228.

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

Manski, C. F. (1989). Anatomy of the selection problem. *Journal of Human resources*, 343–360.

Manski, C. F. (2003). *Partial identification of probability distributions.* Springer Science & Business Media.

Manski, C. F. (2007). *Identification for prediction and decision.* Harvard University Press.

Manzini, P., & Mariotti, M. (2014). Stochastic choice and consideration sets. *Econometrica*, *82*(3), 1153-1176.

Masatlioglu, Y., Nakajima, D., & Ozbay, E. Y. (2012). Revealed attention. *American Economic Review*, *102*(5), 2183-2205.

Matzkin, R. L. (2007). Nonparametric identification. In J. J. Heckman & E. E. Leamer (Eds.), *Handbook of econometrics* (Vol. 6, pp. 5307 – 5368). Elsevier.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*, 105–142.

McFadden, D., & Train, K. (2000). Mixed MNL models for discrete response.

*Journal of Applied Econometrics*, *15*, 447–470.

Mele, A. (2017). A structural model of dense network formation. *Econometrica*, *85*(3), 825–850.

Mele, A., & Zhu, L. (2017). Approximate variational estimation for a model of network formation. *arXiv preprint arXiv:1702.00308*.

Menzel, K. (2015). *Strategic network formation with many agents* (Tech. Rep.). Working papers, NYU.

Milgrom, P., & Roberts, J. (1990). Rationalizability, learning, and equilibrium in games with strategic complementarities. *Econometrica: Journal of the Econometric Society*, 1255–1277.

Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *The review of economic studies*, *38*(2), 175–208.

Miyauchi, Y. (2016). Structural estimation of pairwise stable networks with nonnegative externality. *Journal of econometrics*, *195*(2), 224–235.

Molchanov, I. (2005). *Theory of random sets* (Vol. 19) (No. 2). Springer.

Molchanov, I., & Molinari, F. (2018). *Random sets in econometrics* (Vol. 60). Cambridge University Press.

Molinari, F. (2019). Econometrics with partial identification. *CeMMAP Working Paper CWP25/19*.

Molinari, F., & Rosen, A. (2008). The identification power of equilibrium in games: The supermodular case. *Journal of Business & Economic Statistics*, *26*(3), 297–302.

Myerson, R. B. (1977). Graphs and cooperation in games. *Mathematics of operations research*, *2*(3), 225–229.

Myerson, R. B. (1991). Game theory: analysis of conflict. *The President and Fellows of Harvard College, USA*.

Nishida, M. (2014). Estimating a model of strategic network choice: The convenience-store industry in okinawa. *Marketing Science*, *34*(1), 20–38.

Orbanz, P., & Roy, D. M. (2014). Bayesian models of graphs, arrays and other exchangeable random structures. *IEEE transactions on pattern analysis and machine intelligence*, *37*(2), 437–461.

Pratt, J. W. (1964). Risk aversion in the small and in the large. *Econometrica*, *32*(1/2), 122–136.

Read, D., Loewenstein, G., & Rabin, M. (1999). Choice bracketing. *Journal of Risk and Uncertainty*, *19*(1–3), 171–197.

Ridder, G., & Sheng, S. (2017). *Estimation of large network formation games* (Tech. Rep.). Working papers, UCLA.

Roberts, J. H., & Lattin, J. M. (1991). Development and testing of a model of consideration set composition. *Journal of Marketing Research*, *28*(4), 429–440.

Robins, G., Pattison, P., & Woolcock, J. (2004). Missing data in networks: exponential random graph (p*) models for networks with non-respondents. *Social networks*, *26*(3), 257–283.

Rose, C. (2015). *Essays in applied microeconometrics* (Unpublished doctoral dissertation). University of Bristol.

Sabidussi, G. (1966). The centrality index of a graph. *Psychometrika*, *31*(4), 581–603.

Sheng, S. (2018). *A structural econometric analysis of network formation games* (Tech. Rep.). Conditional Acceptance at Econometrica.

Shocker, A. D., Ben-Akiva, M., Boccara, B., & Nedungadi, P. (1991). Consideration set influences on consumer decision-making and choice: Issues, models, and suggestions. *Marketing Letters*, *2*(3), 181–197.

Simon, H. A. (1959). Theories of decision-making in economics and behavioral science. *The American Economic Review*, *49*(3), 253–283.

Simonson, I., & Tversky, A. (1992). Choice in context: Tradeoff contrast and extremeness aversion. *Journal of Marketing Research*, *XXIX*, 281-95.

Smith, J. A., & Moody, J. (2013). Structural effects of network sampling coverage i: Nodes missing at random. *Social networks*, *35*(4), 652–668.

Song, Y., & van der Schaar, M. (2015). Dynamic network formation with incomplete information. *Economic Theory*, *59*(2), 301–331.

Spence, A. M. (1974). Market signaling: Informational transfer in hiring and related screening processes. *Harvard University Press*, *143*.

Starmer, C. (2000). Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of Economic Literature*, *38*(2), 332-382.

Tamer, E. (2003). Incomplete simultaneous discrete response model with multiple equilibria. *The Review of Economic Studies*, *70*(1), 147–165.

Tamer, E. (2010). Partial identification in econometrics. *Annu. Rev. Econ.*, *2*(1), 167–195.

Tarski, A., et al. (1955). A lattice-theoretical fixpoint theorem and its applications. *Pacific journal of Mathematics*, *5*(2), 285–309.

Tonelli, L. (1909). Sull'integrazione per parti. *Rend. Acc. Naz. Lincei*, *5*(18), 246–253.

Topkis, D. M. (1978). Minimizing a submodular function on a lattice. *Operations research*, *26*(2), 305–321.

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention.

*Cognitive Psychology*, *12*(1), 97 - 136.

Tsakas, N. (2016a). On decay centrality. *The BE Journal of Theoretical Economics*, *19*(1).

Tsakas, N. (2016b). Optimal influence under observational learning. *Available at SSRN 2449420*.

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological review*, *79*(4), 281.

Uetake, K., & Watanabe, Y. (2012). A note on estimation of two-sided matching models. *Economics Letters*, *116*(3), 535–537.

Wilcox, N. T. (2008). Stochastic models for binary discrete choice under risk: A critical primer and econometric comparison. In *Risk aversion in experiments* (pp. 197–292). EGPL.

Zenou, Y. (2016). Key players. *The Oxford Handbook of the Economics of Networks*, *11*.