

WHAT CAN BE LEARNED FROM
OBSERVED BEHAVIOR? :

ESSAYS ON THE PREDICTIVE ABILITY
OF THE UTILITY MAXIMIZATION MODEL AND
TESTING FOR SATISFICING

BY

MARÍA JOSÉ BOCCARDI CHALELA

B.A., UNIVERSIDAD DE MONTEVIDEO, 2006

M.A., BROWN UNIVERSITY, 2011

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
IN THE DEPARTMENT OF ECONOMICS AT BROWN UNIVERSITY

PROVIDENCE, RHODE ISLAND
MAY 2016

© Copyright 2016 by María José Boccardi Chalela

This dissertation by María José Boccardi Chalela is accepted in its present form
by the Department of Economics as satisfying the
dissertation requirements for the degree of Doctor of Philosophy.

Date _____
Doctor Mark Dean, Advisor

Recommended to the Graduate Council

Date _____
Doctor Roberto Serrano, Reader

Date _____
Doctor Andriy Norets, Reader

Approved by the Graduate Council

Date _____
Peter Weber, Dean of the Graduate School

VITA

The author was born April 25th, 1985 in Montevideo, Uruguay. Entering the Universidad de Montevideo in 2003, she received her B.A. in 2006. In the same institution she started a Masters in Economics and taught several undergraduate classes. Meanwhile, she worked for the Uruguayan Bureau of Statistics 2006-2010 as the project leader for the new CPI index at the national level, base December 2010=100. In August of 2010 she entered Brown University to pursue her degree in Economics. She received her M.A. in 2011 and Ph.D. in 2016.

Acknowledgements

I would like to express my special appreciation and thanks to my advisor Professor Dr. Mark Dean, you have been a tremendous mentor for me. I would like to thank you for encouraging my research and for allowing me to grow as an economist. Your advice on both research as well as on my career have been priceless. I would also like to thank my committee members, Professor Dr. Roberto Serrano and Professor Dr. Andriy Norets for serving as my committee members and special thanks to you for all the support during the job market experience, and for your brilliant comments and suggestions. I would like to thank the economics department office staff and especially Angelica Vargas, her guidance through the entire graduate school experience and immeasurable support made me successfully survive to the program and particularly to the job market experience.

I would also like to show my gratitude to Emily Oster for her guidance on the job search process, Pedro dal Bo, Geoffroy deClippel, Jack Fanning, Frank Kleibergen, Anja Sautmann, Rajiv Vohra and seminar participants at Brown University, Econcon2015, the 3rd Annual Warwick Economics PhD Conference and Washington University 9th Annual Graduate Student Conference for helpful comments and suggestions that contributed to this thesis.

I want to thank all of my teachers at Instituto Pallotti for their guidance. I would also like to thank my Professors at Universidad de Montevideo especially Juan Dubra,

Marcelo Caffera, Lorenzo Caliendo and Fernando Borraz who pushed me to be more than I could ever imagine I was capable of, and, without their support and encouragement, I would have never thought or tried to pursue my PhD.

Graduate school is, of course, a long and arduous process. Without the right friends and family who can both support your work and balance it with play, it would be impossible to make it through. I owe a special thanks to my friends and family. Words cannot express how grateful I am to my father and my youngest brother Martín, for all of the sacrifices that they have made on my behalf, their encouragement in this endeavour and constant support throughout the process, even when the situation back home was far from ideal. A great factor in my success and that made me who I am is my friends, the family of my own choice. To my high school and college friends I owe my sanity, I would have not been able to make it this far without them. For that I have to thank Valentina Canto, Cynthia Broquetas, Adriana Montero, Florencia Mazzoni and the CATUM group. Also thanks to all my colleagues at INE, who helped me grow as a professional and provided me with their support to apply to graduate school. For keeping me sane and showing me that there is a life outside grad school, I would like to thank to everyone I met through the practice of handball and field hockey, as well as the staff at the GCB. Finally, I would also like to thank all of my "Providence friends" who supported me in writing, and incentiviced me to strive towards my goal. For this I have to thank, among others, James Budarz, Nicholas Coleman, Sean Dinces, Morgan Hardy, Angélica Meinhofer, Nickolai Riabov, Daniela Scidá and Tim Squires.

Contents

1	Predictive Ability and the Fit-Power Trade-Off in Theories of Consumer Behavior	7
1.1	Introduction	7
1.2	The Rationality Model	13
1.2.1	Axiomatization	13
1.2.2	Challenges for Testing	14
1.3	Prediction Based on Axiomatic Models	16
1.3.1	Set up	19
1.3.2	Predictions Based on GARP	20
1.3.3	Stochastic Extension and Predictive Distribution	24
1.3.4	On the Estimation of the Error Process	32
1.4	Predictive Performance Analysis	34
1.4.1	Decision Making and the Predictive Distribution	34
1.4.2	Marginal Likelihood and Model Averaging	36
1.4.3	Measures of Predictive Ability	38
1.4.4	Comparison to other measures in the literature	45
1.4.5	Optimal Design of Experiments	50
1.5	Empirical Application	52
1.5.1	Effect of Additional Information	53
1.5.2	Empirical Performance of Predictive Ability Measures and Comparison to Other Measures in the Literature	56
1.5.3	Model Comparison: GARP vs. Parametric Specification	57
1.6	Relation to the Literature	57
1.7	Conclusion	59
1.A	Proofs	62
1.A.1	Proofs of Section 1.3	62
1.A.2	Proof of Section 1.4.3	63
1.A.3	Proofs of Section 1.4.4	64
1.B	Error Process	65
1.B.1	Algorithms - Estimation of the Error Process in Finite Samples	65
1.B.2	Performance of Algorithms 1.26, 1.27 and 1.28 for Simulated Data	67
1.C	Tables	68
1.C.1	Data	68
1.C.2	Comparison Between Subjects 313 and 516	73

2	Satisficing and Stochastic Choice - with Victor Aguiar and Mark Dean	75
2.1	Introduction	75
2.2	Set Up	79
2.2.1	Data	79
2.2.2	The Satisficing Model	79
2.3	Characterizing the Satisficing Model.	83
2.3.1	A Negative Result	83
2.3.2	Full Support Satisficing Models	84
2.3.3	Fixed Distribution Satisficing Models	89
2.3.4	Comparative Statics	92
2.4	Extensions	96
2.4.1	Fixed Distributions Without Full Support	96
2.4.2	Allowing for Indifference	97
2.4.3	Incomplete Datasets	100
2.4.4	Random Utility and Random Threshold	101
2.5	Relation to Existing Literature	108
2.A	Proofs	110
2.A.1	Proof of Proposition 2.4	110
2.A.2	Proof of Lemma 2.7	110
2.A.3	Proof of Theorem 2.11	111
2.A.4	Proof of Theorem 2.14	113
2.A.5	Proof of Theorem 2.17	114
2.A.6	Proof of Theorem 2.18	116
2.A.7	Proof of Claim 2.19	117
2.A.8	Proof of Proposition 2.23	117
2.A.9	Comparative Statics with Respect to Search Orders	119
2.A.10	Proof of Proposition 2.28	120
2.A.11	Proof of Theorem 2.32	121
2.A.12	Proof of Theorem 2.33	124
2.A.13	Proof of Theorem 2.34	124
2.A.14	Proof of Lemma 2.39	125
2.A.15	Proof of Lemma 2.40	125
2.A.16	Proof of Lemma 2.36	130
2.A.17	Proof of Theorem 2.37	131
2.A.18	Proof of Lemma 2.41	132
2.A.19	Proof of Theorem 2.42	133
2.A.20	Proof of Theorem 2.48	134
	Bibliography	141

List of Tables

1.1	Data for Example 2.3	17
1.2	Data for Example 1.6	21
1.3	Data for Example 1.7	22
1.4	Simulations for the Algorithms	34
1.5	Experimental design	52
1.6	Summary statistics for (Choi, Fisman, Gale, and Kariv 2007) - half and full sample.	54
1.7	Statistics comparison.	56
1.8	Model comparison.	59
1.9	Simulations algorithms	67
1.10	Summary statistics.	68
1.11	Summary statistics -half of observations.	69
1.12	Comparison - half of observations	69
1.13	Correlations	72
2.1	An example of the satisficing model	82
2.2	Changes in the probability of search orders and its effect on DM's well-being.	96

List of Figures

1	Relation between empirical consistency, falsifiability and predictive ability.	5
1.1	The goodness of fit problem.	14
1.2	The power of the rationality test	16
1.3	Relation between empirical consistency, falsifiability and predictive ability	18
1.4	Observed Choices and the "supporting Set"	21
1.5	Observed economic environments and the "supporting set".	22
1.6	"Supporting set" and number of observations	23
1.7	Predictive Distribution - Intuition	27
1.8	Relation between empirical consistency, falsifiability and predictive ability	29
1.9	Predictive densities for Example 2.3	30
1.10	Histogram of $\frac{\text{Projected-SSR}}{\text{Generated-SSR}}$	32
1.11	Relation between empirical consistency, falsifiability and predictive ability.	36
1.12	Comparison BC and predictive ability measures	49
1.13	Comparison BC and predictive ability measures	50
1.14	Optimal experimental design for prediction	53
1.15	Comparison between subjects	73
1.16	Predictive densities - comparison.	74

Introduction

An economic model can be understood as a theoretical construct that, by simplifying a complex reality, seeks to aid in the understanding and/or prediction of economic processes. As such, for the most part, these models are not meant to replicate an intricate reality but intend to help in the understanding of the processes that drive observed outcomes and, if desired, to be used in the prediction of responses to unobserved economic environments.¹

Though economic models can be thought as interesting theoretical exercises to better understand specific economic processes², special interest lies in checking whether these models can be supported by empirical evidence and/or be useful to predict behavior in out-of-sample economic environments. For that, it is important to understand the connection between the theory and its empirical counterpart. Axiomatizations provide the description of the implications of the model in terms of observable behavior, delivering necessary and sufficient conditions that are equivalent to the theoretical model.³ Consider, for example, the utility maximization model, that is the focus of Chapter 1. Utility functions are unobservable, but their behavioral manifestation -choices- are not. The Generalized Axiom of Revealed Preference (GARP henceforth) provides the necessary and sufficient conditions on observed behavior for it to be consistent with a decision

¹For more on the role of models in economics refer to (Gibbard and Varian 1978), (Sugden 2000), (Sugden 2009), (Mäki 2005), (Caplin and Schotter 2010), among others.

²Notable examples are the "The Market for Lemons: Quality Uncertainty and the Market Mechanism" by (Akerlof 1970) and "Job Market Signalling", (Spence 1973).

³(Dekel and Lipman 2010) argue for the value provided to axiomatic derivations.

maker that behaves *as if* her choices maximize her utility function among feasible alternatives.

Understanding the behavioral implications of the model is not only relevant for testing its empirical validity; but it is also important to understand them when selecting among theories for modeling behavior. Two seemingly distinct behavioral models cannot be teased apart if they result in the same observed data; that is, if for a sequence of decision problems the behavioral implications of two competing models are not significantly distinct, then the theoretical differences between the models are irrelevant.⁴ The empirical content of a model depends on the behavior that pretends to explain and available data; i.e. comparisons across competing models can only be made conditional on the behavior that is intending to explain for the considered sequences of economic environments. For example, the satisficing model⁵ is a highly influential and intuitive model of bounded rationality, but it cannot be tested using standard choice data; its behavioral implications are indistinguishable from those of a standard utility maximizer. In Chapter 2, my coauthors and I propose an alternative way to test for the satisficing model based on stochastic choice data. Assuming that preferences are fixed, but search order may change randomly, the model predicts that stochastic choice can only occur amongst elements that are always chosen, while all other choices must be consistent with standard utility maximization. Adding the assumption that the probability distribution over search orders is the same for all choice sets makes the satisficing model a subset of the class of random utility models.

The empirical content of a model and the suitability of it to explain a given data set depends on the sequence of decision problems that is being studied. Consider again the utility maximization model. If the decision maker faces menus of alternatives $A = \{a, b\}$

⁴(Gul, Pesendorfer, et al. 2008)

⁵(Simon 1955)

and $B = \{c, d\}$, the utility maximization model has no empirical content, i.e. any profile of choices from those menus can be rationalized. Consider now a decision maker choosing from $A' = \{a, b, c\}$ and $B' = \{b, c, d\}$, for those sets of feasible alternatives the theory has empirical content.

The empirical content of a model also depends on the behavior that seeks to explain/predict. For example, assume that the decision maker chooses $C(A') = \{a\}$; then any choice from B' would be consistent with the model, i.e. choices from A' are not informative about the predictions of the model for B' . This latter example raises a few questions when considering limited data sets; what information can be learned about preferences? what predictions can be made based on observed behavior?. These questions are addressed, for the case of the utility maximization model, in Chapter 1. As discussed in (De Clippel and Rozen 2014) the latter is not a trivial question, for many bounded rationality models. The authors show that for some models, it is the case that a limited data set may be consistent with a subset of possible menus but no extension for a complete data set would be.

(Gabaix and Laibson 2008) describes the seven key properties of useful economic models: *parsimony*, *tractability*, *conceptual insightfulness*, *generalizability*, *falsifiability*, *empirical consistency*, and *predictive precision*. The last three properties pertain to the relation between the theoretical model and its empirical counterpart. A model is *falsifiable* if it makes non trivial predictions on behavior, i.e. if there is some feasible behavior profile that is inconsistent with the model. As a simplification of the reality, a model is not expected to perfectly explain human behavior in all possible circumstances. However, as researchers, we are interested in the *empirical accuracy* of the model, and when observed behavior is not consistent with the model, the degree to which the model fits the data. This is known as the goodness of fit criterion.

The goodness of fit of a model is not the only relevant criteria when judging the suitability of a model. The limitations of this criterion are specially relevant when dealing with axiomatic models with limited data sets, as discussed in Chapter 1. In particular, a model may be empirically accurate because data was indeed generated by a decision maker that behaves as the theory prescribes; but it could be also the case that it is empirically accurate because, for the considered sequence of decision problems, the model prescribes weak (loose) predictions, such that almost any behavior would be consistent with the theory. This caveat relates to *predictive precision*. A model that delivers more precise predictions is desirable because it facilitates its evaluation and testing, furthermore, "*A model with predictive precision may even be useful when it is empirically inaccurate(...) In general, models that make approximately accurate strong predictions are much more useful than models that make exactly accurate weak predictions.*"⁶. Figure 1 illustrates the importance of predictive precision and how empirically inaccurate models can be useful if they deliver strong predictions.

Chapter 1 exploits the connection between the predictive precision and the power to identify inconsistent behavior conditional on observed choices, proposing a predictive ability approach for model assessment. This approach provides a meaningful answer to the tension between the severity of the violations (fit) and the sensitivity of the test to detect them (power), a long standing problem in the literature when dealing with axiomatic representations for behavioral models. The discussion in the chapter is centered around the utility maximization model, though most results extend to a general class of behavioral models. Utility maximization is a core assumption in economics for which the generalized axiom of revealed preference (GARP) provides a nonparametric test. However, two problems arise when testing GARP. First, it provides an extremely sharp

⁶(Gabaix and Laibson 2008)

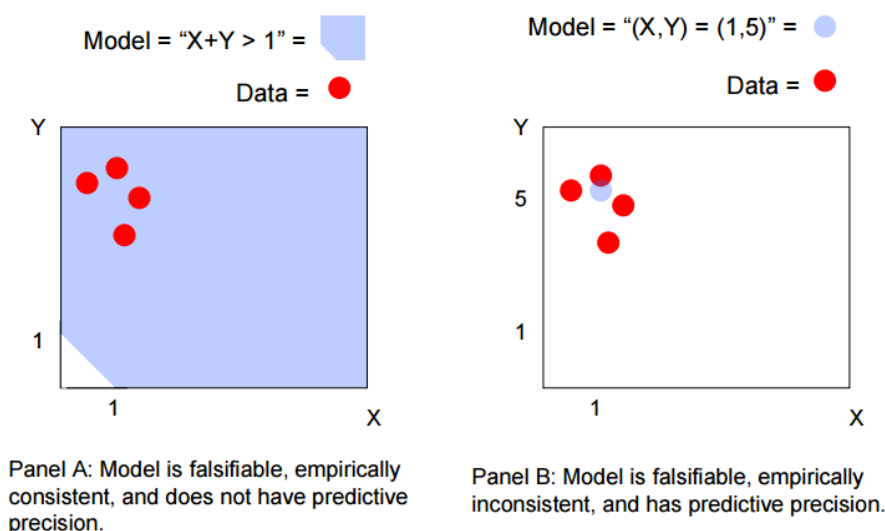


Figure 1: Relation between empirical consistency, falsifiability and predictive ability. Empirical consistency -fit- is important but does not implies usefulness of the model. Even when falsifiable it may not restrict behavior enough to allow for learning about underlying behavior.

test since rejection occurs after only one violation. Second, even when a data set passes the test, it cannot necessarily be understood as a success for the theory since conditions may be so undemanding that any behavior would pass it. Then, the predictive ability approach developed in the chapter naturally establishes a meaningful trade-off between empirical accuracy and falsifiability for models of consumer behavior, a long standing problem in the literature. Intuitively, better predictions are the result of a lot of revealed preference information –better degree of identification of underlying preferences– while requiring small errors for data to be consistent with the model.

This dissertation emphasizes the importance of the empirical implications for models of behavior. Chapter 1 discusses the limitations of goodness of fit approach for axiomatic models when dealing with incomplete data sets. The lack of power to identify inconsistent behavior with limited data sets can be captured by uncertainty to predict behavior based in the model. By accounting for this uncertainty and allowing for empirically inaccurate behavior, the predictive ability approach provides a meaningful criterion to

study the suitability of a theory to model observed behavior. Chapter 2 provides a novel set of conditions to identify whether a decision maker behaves as satisficing, in contrast to standard utility maximizer. This is done by exploiting the structure of stochastic choice data without requiring data on the decision process itself.

Outline The outline of this dissertation is as follows. Chapter 1 presents a predictive approach for the assessment of the suitability of models when these are described by a set of axioms.⁷ Chapter 2 presents necessary and sufficient conditions for stochastic choice data to be consistent with the satisficing model.⁸ Finally Section Conclusion closes presenting a summary of the main results and open lines of future research.

⁷The content of this chapter has been presented before in the paper "Predictive Ability and the Fit-Power Trade-Off in Theories of Consumer Behavior"

⁸The content of this chapter has been circulated under the title "Satisficing and Stochastic Choice" coauthored with Victor H. Aguiar and Mark Dean.

Chapter 1

Predictive Ability and the Fit-Power Trade-Off in Theories of Consumer Behavior

1.1 Introduction

The existence of a utility function that represents preferences is a core assumption in economics. The Generalized Axiom of Revealed Preferences (GARP henceforth) provides an elegant axiomatic test for the validity of this assumption. For GARP, rejection occurs after only one violation, providing a sharp test which most data sets violate. This is a common feature of all models that are described using a set of behavioral axioms. Axioms provide elegant nonparametric tests, but only provide binary information as to whether a particular data set is consistent with the model. In case of inconsistent data, axiomatizations say nothing about the significance of the departures from the model. This is known as the goodness of fit problem in the literature.

A series of goodness of fit measures have been proposed to address the sharpness

of the rationality test. Most of these measures are based on an intuitive moment of the data related to the adjustments to income needed to remove the violations by making them no longer feasible.¹ The significance of these measures is difficult to gauge since the experimental design may be such that no or very loose constraints are imposed on the data. For example, if budget sets are nested, any data set would exhibit perfect consistency with GARP since no constraints are imposed by the model. This is known in the literature as the power problem since it relates to the probability of identifying violations when the data is generated by a non-rational process.²

This chapter assesses the performance of the model by asking whether the model is useful to make precise predictions about behavior. A good model is one which provides precise predictions, reducing the uncertainty of forecasted behavior for unseen economic environments. More precise predictions are the result of a large amount of revealed preference information that can be learned from data, while requiring a small error for observed behavior to be consistent with the model. Given behavior, more stringent environments that impose more demanding constraints on data to satisfy the model, result in more precise revealed preference information learned from data. However, more demanding environments increase the likelihood of detecting violations, leading to overall bigger errors. Hence, the predictive ability approach provides a meaningful way to integrate fit and the degree to which the data constrains the model.

In order to construct the predictive distribution of choices in a new environment, I extend the model to embed it in a statistical framework allowing for an additive error component. Within this framework, I recover the "candidate model" – the most likely sequence of rational choices that generated the data– and an estimate for the error pro-

¹Notable examples are (Afriat 1967), (Varian 1990) and (Echenique, Lee, and Shum 2011)

²See (Andreoni and Harbaugh 2013) for an extensive analysis of the power problem when testing for GARP

cess. Next, I combine these two to compute the predictive distribution generated by the extended model. For an out-of-sample economic environment, I first construct the set of choices that are consistent with the "candidate model". This is the "supporting set" first introduced by (Varian 1982). For an alternative in the "supporting set", the model predicts that choices are given by that alternative plus the error. Then, following the intuition of Bayesian model averaging, I construct the predictive distribution as the distribution of choices on the "supporting set" plus the distribution of the error process. Without prior information, all alternatives in the supporting set are assumed to be equally likely. This distribution can be interpreted as the prediction that can be made by only assuming the nonparametric model, given observed behavior. This construction extends to a general class of behavioral models, provided the availability of an algorithm to recover the "candidate model". For GARP, it is available in the form of a minimum distance estimator provided by (Kocoska 2012).

The predictive distribution reflects two sources of uncertainty when forecasting behavior: "model uncertainty", which derives from the fact that there may be different preference relations that are consistent with observed data³; and "error uncertainty", which derives from the fact that the decision maker may not perfectly maximize her utility. A model is better for prediction if it reduces either or both sources of uncertainty when forecasting behavior. Hence, the predictive distribution implicitly defines a trade-off between the amount of information that can be learned about preferences and the severity of the violations, naturally extending the prediction that would result from the axiomatization to allow for error. As I discuss in Section 1.4, there is a strong link between the amount of recoverable revealed preference information and the power of data set to highlight violations.

³To illustrate this point consider the following simple example. Let the universe of choices be $\{a, b, c\}$, and let observed data be such that a is chosen when $\{a, b\}$ were available, and a chosen from $\{a, c\}$. Then we may have $a \succ b \succ c$ or $a \succ c \succ b$ with equal probability.

For a researcher considering a model for predicting behavior for a particular application, the predictive distribution is a sufficient statistic to compute optimal choices and expected losses due to incorrect predictions. For a set of candidate models, the predictive distribution can be used to calculate an analog of the marginal likelihood of each model, that is, how likely observed data are given the model. Given a prior on the likelihood of each candidate model, marginal likelihoods can be used to construct Bayes-like factors for model selection. Marginal likelihoods can also be used to compute weights for bayesian model averaging (BMA) to account for uncertainty with respect to model selection. The BMA predictive distribution is constructed as the weighted average of the predictive distribution provided for each candidate model, where the weights are given by the likelihood of the model given observed data.

For a researcher that does not have a particular application of interest but seeks to appraise the predictive performance of the model, I provide two intuitive statistics that summarize the information content of the predictive distribution. These statistics define a complete order of data sets and/or models in the domain of predictive distributions. These summary statistics are: (i) the size of the shortest α -level credible set for predicted choices and (ii) the Kullback-Leibler divergence measure between the predictive distribution and an uninformative prior. These measures reflect the extent of both sources of uncertainty, model and error, and can be used for the design of experiments. The selection of the economic environment in which to predict behavior may affect the results, therefore, I propose a leave-one-out construction for these measures to remove any arbitrariness in this regard. The leave-one-out (LOO) predictive precision measures are computed based on the predictive distribution constructed for each observed economic environment considering all remaining observations.

I study the empirical performance of the predictive measures in their application to the experimental data from (Choi, Fisman, Gale, and Kariv 2007). As expected, measures of predictive precision are positively correlated with measures of goodness of fit and with the amount of information that can be extracted from data. I find that more observations lead to more precise predictions since the additional revealed preference information outweighs the potential increase in estimated errors. Finally, I compare the performance of the Cobb-Douglas model, nested into the rationality model, and GARP which does not make any assumption about the shape of the utility function. When comparing the two models, the results show that the nonparametric utility maximization model provides more precise predictions than a model that assumes Cobb-Douglas preferences, since the additional precision of the latter model is outweighed by the misspecification error.

This chapter ties together the goodness of fit and power literatures. The information provided by standard goodness of fit measures is captured by the predictive precision measures through the recovered error process, as shown in Proposition 1.21. Additionally, the LOO predictive precision measures reflect power, through the size of the supporting set. Given the projection using $J - 1$ observations, the supporting set is the subset of feasible choices, in the J th economic environment, that are jointly consistent with the projection. Conditional on the projection, its relative size is the probability of choices being consistent with the model, if they were to be generated uniformly at random from the set of feasible choices. Then, the relative size of the supporting set is the complement of the probability of detecting violations in the J th economic environment under (Becker 1962) definition of irrational behavior; that is, power –as traditionally understood in the literature, (Bronars 1987)– conditional on the projection.⁴ Hence,

⁴To illustrate this point consider a decision maker choosing from $\{a, b, c\}$ and $\{a, c\}$. If a (or c) is chosen from $\{a, b, c\}$ then $\{a, c\}$ imposes demanding constraints since it is possible to violate GARP. On the other hand, if b is chosen from $\{a, b, c\}$ then any choice from $\{a, c\}$ would be consistent with GARP.

this approach provides a meaningful trade-off between fit and power, a long standing problem in the literature.

Current approaches that combine fit and power, as (Beatty and Crawford 2011), measure power consistent with (Bronars 1987) approach. However, this approach fails to account for the actual pattern of choices observed.⁵ Observed choices determine whether the new budget set constrains data or not. This is important since, for the empirical application -50 observations-, standard power is approximately one for all subjects which does not allow to differentiate among the stringency of different designs. I show that two subjects that faced identical economic environments and whose behavior is consistent with GARP, can produce significantly different predictions for the same economic environment due to differences in the supporting set. I show that unconditional power masks significant differences in the amount of preference information that can be inferred for different subjects to predict behavior.

Outline The outline of this chapter is as follows. Section 1.2 presents the main results of revealed preference theory, its testable implications and the challenges that it presents for testing. Section 1.3 introduces the construction of the predictive distribution and discusses its properties. Section 1.4 shows that the predictive distribution is a sufficient statistic for a decision maker that wants to select a model to predict behavior. For a researcher that does not have an application in mind, I show how to use the predictive distribution to construct the marginal likelihood of the model and measures that summarize the predictive performance of the model. Section 1.5 shows the empirical performance of the proposed measures in its application to the experimental data from (Choi, Fisman, Gale, and Kariv 2007). Section 1.6 offers a review of the literature

⁵(Andreoni and Harbaugh 2013) presents a series of power measures and indices whose behavior depend on the characteristics of choices in the population and experimental design.

concerning rationality testing and how the approach followed in this chapter compares to existing approaches. Finally, section 1.7 concludes.

1.2 The Rationality Model

1.2.1 Axiomatization

When facing a set of feasible alternatives, the choices made by a decision maker reveal information about her own preference relation, chosen alternatives are revealed to be better than non-chosen ones. This is the concept of (directly) revealed preference, first introduced by (Samuelson 1938). (Houthakker 1950) extended it by imposing transitivity on the direct revealed preferred relation. Let $\{q^j\}_{j=1}^J$ be the sequence of observed choices, with $q^j \in \mathbb{R}_+^L$. The set of feasible alternatives is determined by the decision maker's budget constraints, that is, for all $j \in \{1, \dots, J\}$, $p^j \cdot q^j \leq x^j$, where $p^j \in \mathbb{R}_{++}^L$ is the price vector faced by the decision maker in environment j , and x^j her income. Then, chosen alternatives q^j are said to be revealed preferred to the other feasible alternatives that were not chosen. Formally,

Definition 1.1 (Directly Revealed Preferred) q^j is directly revealed preferred to \tilde{q} if $p^j \cdot q^j \geq p^j \cdot \tilde{q}$, and it is strictly revealed preferred if $p^j \cdot q^j > p^j \cdot \tilde{q}$

Definition 1.2 (Revealed Preferred) q^j is revealed preferred to \tilde{q} if there is a chain of directly revealed preferred bundles linking q^j to \tilde{q}

(Afriat 1967) theorem shows that revealed preference theory provides a nonparametric condition on consumer's choices that is necessary and sufficient for observed behavior to be consistent with non satiated utility maximization: the generalized axiom of revealed preference (GARP). This axiom imposes internal consistency on the information contained in the revealed preference relation for preferences to be represented by a well behaved utility function. Formally,

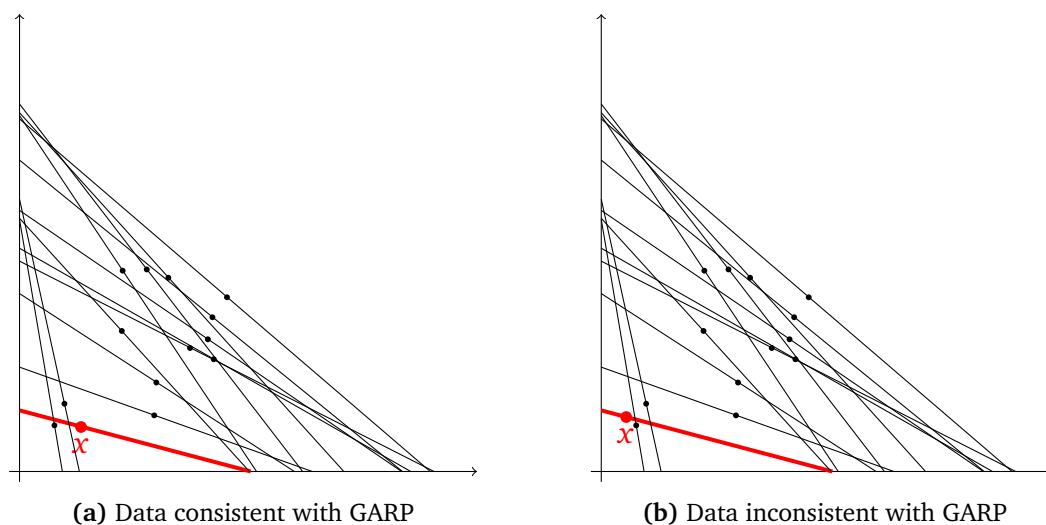


Figure 1.1: The goodness of fit problem. The sharpness of the test is evidenced in the above figure. Panel 1.1a shows data that is perfectly consistent with GARP, while in panel 1.1b one observation x was perturbed. The data in panel 1.1b is not longer consistent with utility maximization.

Definition 1.3 (Generalized Axiom of Revealed Preference (GARP)) *If q^j is revealed preferred to \tilde{q} , then \tilde{q} is not strictly revealed preferred to q^j*

1.2.2 Challenges for Testing

Two main problems arise when testing for the revealed preference conditions: fit and power. First, GARP delivers a sharp test, since rejection occurs after one violation, regardless of the number of consistent observations and/or severity of the violations. Figure 1.1 depicts the case of a decision maker for which we observed 14 choices. Panel 1.1a shows the case of perfectly consistent choices, while panel 1.1b shows the same choices but choice x has been slightly perturbed. The data in panel 1.1b is no longer consistent with GARP. Most empirical studies find violations to the deterministic axiomatization.⁶

⁶From the empirical point of view, several studies have shown the existence of GARP violations in different settings, cross section data ((Blundell, Browning, and Crawford 2003), (Beatty and Crawford 2011) and (Echenique, Lee, and Shum 2011), among others); laboratory experiments ((Andreoni and Miller 2002), (Sippel 1997), (Choi, Fisman, Gale, and Kariv 2007), among others); as well as in field experiments (for example in (Harbaugh, Krause, and Berry 2001)).

Naturally one may ask *how severe these violations are*. This is known in the literature as the *goodness of fit problem*. Measures of fit have been proposed in the literature that are, for the most part, based on the extent of the income adjustment to budget sets required to remove inconsistencies by making them no longer feasible. These provide intuitive measures of the size of the departures, but a high goodness of fit measure can hardly be understood as a success for the model. Good fit could indicate behavior closely consistent with rationality or failure to detect violations.

The stated conditions may deliver very loose or non restrictive constraints on data. For example, if the decision maker faces nested budget sets *any* behavior is consistent with GARP. Then one may ask *how sensitive the test is to detect departures from rationality*. Conditional on observed behavior, the more the data constrains choices, the more precise the identification of the preference relation generating choices. The link between the amount of revealed preference information that can be learned from data and the sensitivity of the test to detect violations of GARP is studied in Section 1.4.4.

The literature has addressed this problem by providing power indices. These indices measure the probability of observing GARP violations if choices were to be generated uniformly at random from the set of feasible alternatives, see (Bronars 1987) and (Beatty and Crawford 2011). However, the sensitivity of the test to detect departures from the model does not only depend on the number and tightness of the constraints imposed by the model, but also on the interaction of these with observed choices. To illustrate this point consider the problem shown in figure 1.2. The budget sets faced by the decision maker can be informative -as in Panels 1.2b and 1.2c- or not informative at all, as in Panel 1.2a. Moreover, given observed budget sets, the probability of detecting violations of GARP depends on observed behavior, and therefore the

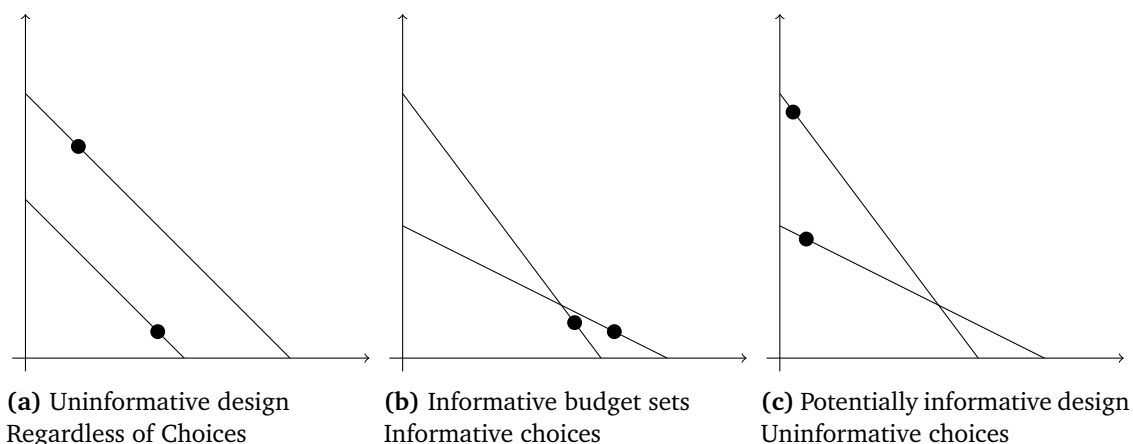


Figure 1.2: The power of the rationality test. This figure presents hypothetical budget sets faced by a decision maker. Panel 1.2a shows two nested budget sets. Independent of observed behavior it is impossible to detect violations of GARP or learn about underlying behavior. Panels 1.2b and 1.2c both depict the same budget sets. This design is potentially informative. The budget sets in panel 1.2b are informative given the distribution of choices, while for panel 1.2c these do not allow to infer significant information about the process generating the choices.

information that can be inferred from choices. This is shown in Panels 1.2b and 1.2c.

1.3 Prediction Based on Axiomatic Models

Economic models are theoretical constructions that simplify and represent economic processes to help in the description, understanding and/or prediction of human behavior. As simplifications models are not meant to replicate a complex reality. Even if the model is not empirically accurate; it may still be useful in predicting human behavior. Consider the following example.

Example 1.4 (GARP) *A researcher wants to predict behavior for budget set, B^0 , defined by prices $p^0 = (1, 1)$ and income $x^0 = 3.5$. She models consumers' behavior by assuming that they are utility maximizers. For that she observes two decision makers A and B, choosing bundles from two budget sets. The information is described in Table 1.1.*

Obs	p_x	p_y	x	q^A	q^B
1	5	1	5	$(\frac{1}{5}, 4)$	$(\frac{5}{8}, \frac{15}{8})$
2	1	2	5	$(4, \frac{1}{2})$	$(\frac{1}{2}, \frac{9}{4})$

Table 1.1: Data for Example 2.3 Data for consumers A and B, choosing from $B(p^1, x^1)$ and $B(p^2, x^2)$

This example is depicted in Figure 1.3. For budget set B^0 , choices q_1^A and q_2^A are not informative even if they are consistent with GARP, since any bundle from B^0 would be consistent with them. On the other hand, even if inconsistent with GARP, choices for decision maker B can potentially provide more information for predicting behavior on B^0 . For either q_1^B and q_2^B there is just a subset of feasible choices on B^0 that would be consistent with them and GARP. Allowing for error would deliver even more precise predictions that result from the intersection of the choices that are jointly consistent with minimally perturbed q_1^B and q_2^B .

This example shows that empirical accuracy may be important but it is not the only relevant attribute of a model. In this regard, (Gabaix and Laibson 2008) recognizes that "*predictive precision is infrequently emphasized in economics research*". They argue that "*strong predictions are desirable because they facilitate model evaluation and model testing (...) Models with predictive precision are useful tools for decision makers who are trying to forecast future events or the consequence of new policies.*".

Ideally, the best model is the one that shows the best fit given observed data and imposes stringent constraints on data, but these two properties tend to be mutually inconsistent. In section 1.4.4, I show that there is a link between the amount of information that can be learned from data about preferences and the extent to which the model imposes demanding constraints in data. Thus, by proposing a predictive approach to assess the suitability of the model, this chapter provides a meaningful compromise be-

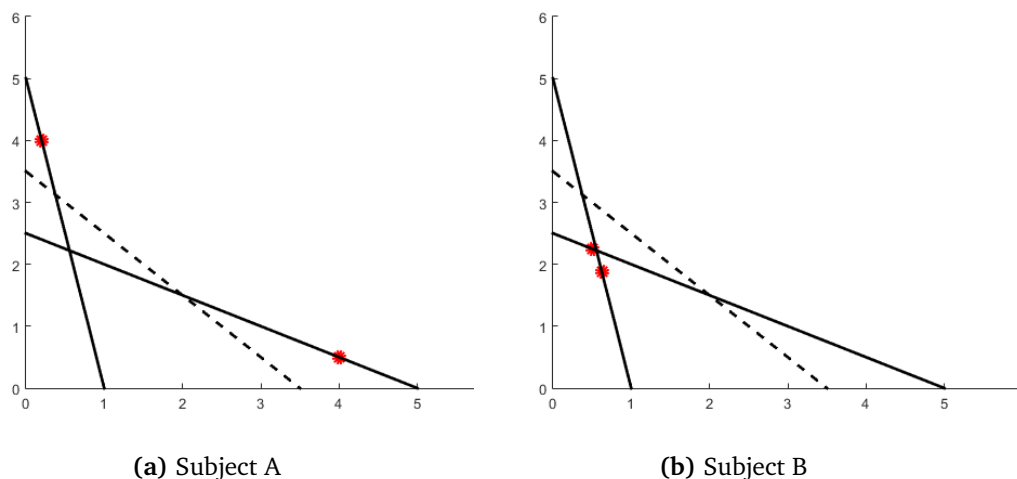


Figure 1.3: Relation between empirical consistency, falsifiability and predictive ability. Empirical consistency -fit- is important but does not implies usefulness of the model. Even when falsifiable it may not restrict behavior enough to allow for learning about underlying behavior.

tween fit and power. First, I extend the model to embed it in a statistical framework, allowing for an additive error to behavior. Within this framework, I recover the revealed preference information and error process from the data. Next, I construct the predictive distribution as the expected distribution of choices for an unobserved economic environment, given the information recovered from data.

The predictive ability of a model is not an inherent property of the model, but it depends on the behavior that seeks to predict. Therefore the analysis is performed for the pair model-data – $\mathcal{M}\mathcal{D}$ -program–, allowing for the comparison across data sets for a given model or across models for a given data set.

The standard practice in economics is to do comparisons based on their accuracy explaining observed behavior (fit). But when dealing with axiomatizations, this criterion favors general models that deliver loose predictions, since models that deliver more precise predictions may fail to explain observed behavior.

Information criteria methods extend goodness of fit measures to account for model complexity⁷, penalizing empirical accuracy to avoid overfitting. These are mostly used in the context of parametric model selection, where increasing the number of regressors almost always improves fit. In the context of axiomatic models, there is a similar trade off between the generality of the model and its empirical accuracy: general models are most likely empirical accurate but provide wide predictions, while more specific models potentially induce bigger errors but providing more precise predictions.

In finite samples, axiomatic models induce an additional source of uncertainty since, conditional on a sequence of rational choices and the error process, the model delivers a set prediction even for perfectly consistent data. In what follows I define this uncertainty as "model uncertainty". In this chapter, I follow the intuition of Bayesian Model Averaging (BMA) to account for "model uncertainty", by averaging across the predictions delivered by the model –that is, all preference relations consistent with the revealed preference information inferred from data–.

1.3.1 Set up

The econometrician observes choices made by an individual that faces J decision problems. Observed data, \mathcal{D} consists of choices $\{q^j\}_{j=1}^J$ and prices and income $\{p^j, x^j\}_{j=1}^J$ with $q^j \in \mathbb{R}_+^L$, $p^j \in \mathbb{R}_{++}^L$ and $x^j \subseteq \mathbb{R}_{++}$ for all $j = 1, \dots, J$, thus $\mathcal{D} \equiv \{q^j, p^j, x^j\}_{j=1}^J$. Assuming local non satiation⁸, $x^j = p^j \cdot q^j$.⁹ Let $B : \mathbb{R}_{++}^{L+1} \rightarrow \mathbb{R}_+^L$ be the feasible consumption correspondence which is defined by imposing the affordability and non-negativity

⁷See (Akaike 1998)

⁸**Definition (Local Non Satiation)** For any $q \in \mathcal{X}$ and every $\varepsilon > 0$ there exists a $y \in \mathcal{X}$ such that $\|q - y\| \leq \varepsilon$ such that y is preferred to q .

⁹Data on income it is not required in the analysis that follows, however, for the sake of uniformity on notation I keep income as a separate variable since new economic environments are defined by prices and income.

constraints, that is

$$B(p^j, x^j) \equiv \{\tilde{q} \in \mathbb{R}_+^L : p^j \cdot \tilde{q} \leq x^j \equiv p^j \cdot q^j\}$$

Let $\mathbf{B}(p, x) = \prod_{j=1}^J B(p^j, x^j)$ be the set of feasible consumption bundles sequences.

The model is a correspondence $\mathbf{M} : \mathbf{B}(p, x) \rightarrow \mathbf{M} \subset \mathbf{B}(p, x)$ that maps from the set of feasible choices to the subset of feasible choices that is consistent with the model. That is

$$\mathbf{M} \equiv \mathbf{M}(\mathbf{B}(p, x)) = \{m \in \mathbf{B}(p, x) : m \text{ satisfies the model}\}$$

Before extending the model to allow for error, I discuss the properties of the prediction provided by the deterministic model. The predictive distribution for the extended model –Section 1.3.3– expands the deterministic prediction to incorporate error in choices, and therefore inherits its properties.

1.3.2 Predictions Based on GARP

Given a vector of choices consistent with the model, $m \in \mathbf{M}(\mathbf{B})$, I define $S \subseteq B^0$ as the subset of B^0 such that $(m, s) \in \mathbf{M}(\mathbf{B}, B^0)$ for all $s \in S$. I call this set the "supporting set". The "supporting set" is the prediction delivered by the axiomatic model, and, coincides precisely with the supporting set defined by (Varian 1982) when the model is GARP. Formally,

Definition 1.5 (Supporting set) *Let $m \equiv (m^1, \dots, m^J)$ be a vector of choices consistent with the model, i.e. $m \in \mathbf{M}(\mathbf{B}(p, x))$. Given a new economic environment B^0 , I define the "supporting set" given m , $S^0(m)$, as*

$$S^0(m) \equiv S(p^0, x^0 | m) \equiv \{s \in B^0 \mid (m, s) \in \mathbf{M}(\mathbf{B}(\{p, p^0\}, \{x, x^0\}))\} \quad (1.1)$$

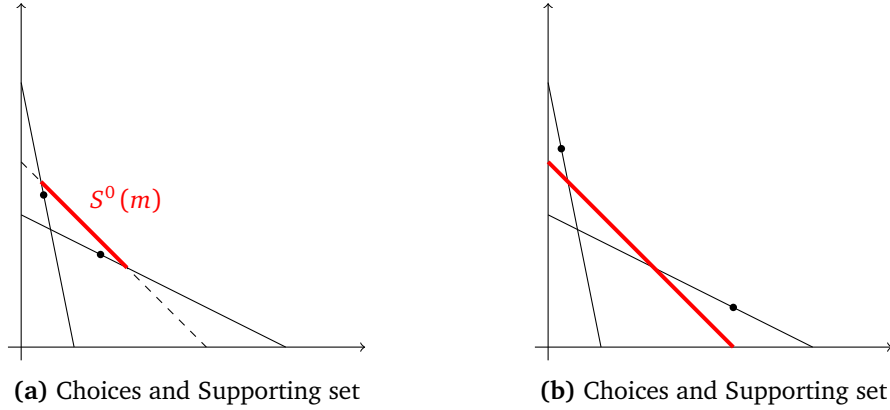


Figure 1.4: Observed choices are a relevant determinant of the "supporting set".

The size of this "supporting set" reflects the tightness of the identification of the underlying behavioral process, since, by definition, each distinct alternative in the "supporting set" represents a behavioral distinguishable and well-behaved preference relation that is consistent with observed data. Therefore, the size of the "supporting set" reflects "model uncertainty", which depends on observed behavior, the distribution of budget sets and the number of observations. To illustrate these dependencies consider the following examples.

Example 1.6 (GARP - "Supporting set" depends on observed choices) *The researcher has information on two decision makers (a) and (b) that have faced the same budget sets. She is considering GARP as a model to predict choices. Their choices are presented in Table 1.2.*

Case	p^A	x^A	q^A	p^B	x^B	q^B	p^0	x^0
(a)	(5, 1)	5	$(\frac{17}{40}, \frac{25}{8})$	(1, 2)	5	(1.5, 1.75)	(1, 1)	3.5
(b)	(5, 1)	5	$(\frac{1}{4}, \frac{15}{4})$	(1, 2)	5	(3.5, 0.75)	(1, 1)	3.5

Table 1.2: Data for Example 1.6

This data is depicted in Figure 1.4. Given observed budget sets, choices for decision maker (a) provide more information for predicting behavior in B^0 than the decision maker's

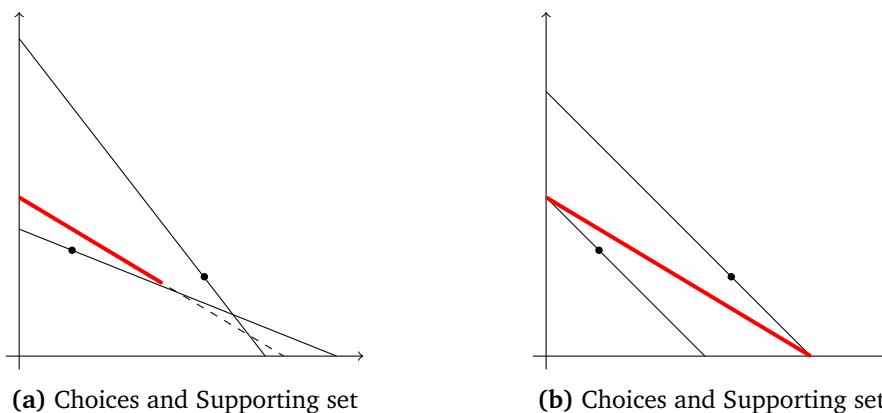


Figure 1.5: Observed economic environments are an important determinant of the "supporting set".

(b) choices, showing that the size of the supporting set depends on observed choices.

Example 1.7 (GARP - "Supporting set" depends on observed budget sets) *The researcher has information on two decision makers, (a) and (b), that have faced different budget sets, but still have chosen the same consumption bundles. She is considering GARP as a model to predict choices. Table 1.3 summarizes the data for these two consumers.*

Case	p^A	x^A	q^A	p^B	x^B	q^B	p^0	x^0
(a)	$(2.15, \frac{5}{3})$	10	(1,2)	(2.5, 1)	6	(3.5, 1.5)	(3,5)	15
(b)	(1, 1)	3	(1,2)	(1, 1)	5	(3.5, 1.5)	(3,5)	15

Table 1.3: Data for Example 1.7

This data is shown in Figure 1.5. For the same observed choices and same objective B^0 , the results are considerably different, showing that the "supporting set" also depends on the environment in where choices were observed.

For any B^0 the "supporting set" is non-empty if and only if it is constructed from a sequence of choices consistent with the model. The size of the "supporting set" (weakly) shrinks with the number of observations¹⁰, since more constraints restrict the set of

¹⁰Figure 1.6 shows simulations results that illustrate the convergence of the "supporting set" for a rational consumer and a randomly selected B^0 when J increases.

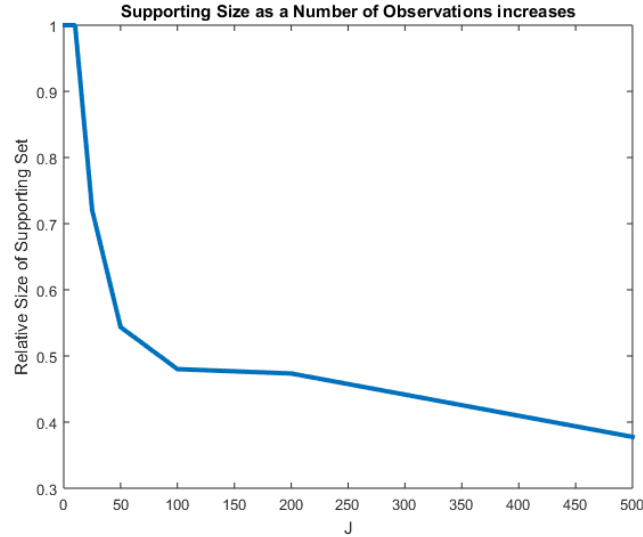


Figure 1.6: "Supporting set" and number of observations Relative size of the supporting set as the number of observations increases. The relative size is defined as $\frac{\|S^0(m)\|}{\|B^0\|}$. Budget sets B were generated at random such that $p_1 \sim U[3, 10]$, $\frac{p_2}{p_1} \sim U[.5, 1.5]$ and $x \sim U[90, 105]$ and choices were generated such that $m^j \equiv \operatorname{argmax}_{q^j \in B^j} \min \{q_1^j, \frac{1}{2}q_2^j\}$. Fixing B^0 , the supporting set is constructed based on $\{m^i, p^i, x^i\}_{i=1}^k$ with $k = 1, 5, 10, 25, 50, 100, 200, 500$.

behavior that is consistent with them. At the limit, when the considered model is GARP, the "supporting set" is a compact and convex set. Finally, for a sequence of budget sets that becomes dense in space of prices and income, the "supporting set" for GARP converges to a point. The latter point implies that, in the limit, if the sequence of budget sets is demanding "enough" –in the sense that they intersect sufficiently enough– the preference relation that generates the data is point identified. These results are formalized in Proposition 1.8.

Proposition 1.8 (Properties of the "Supporting set") *Let B^0 be the economic environment for which the "supporting set" is constructed.*

1. $S^0(m) \neq \emptyset$ if and only if $m \in \mathbf{M}(B)$
2. Let $\mathbf{B}^J, \mathbf{B}^{J+1}$ be a set of J and $J + 1$ budget sets respectively, with $m_j \in \mathbf{M}(B^j)$ and $m_{j+1} = (m_j, m^{j+1}) \in \mathbf{M}(B^{j+1})$ then $S^0(m_{j+1}) \subseteq S^0(m_j)$

Moreover if $\mathbf{M} = \{m \in B \mid m \text{ satisfies GARP}\}$ then

1. $S^0(m)$ is a compact, convex set
2. For an increasing sequence of budget sets C^n such that $C^n \subset C^{n+1} \subset \dots$ and $\bigcup_n \{C^n\}$ is dense in $\mathbb{R}_{++}^{L \times n}$, then $m^n \in \mathbf{B}(C^n)$ such that $m^n \gg 0$ $S^0(m^n) \rightarrow m^0$

Proof. Follows from (Mas-Colell 1978) and (Blundell, Browning, and Crawford 2008). ■

1.3.3 Stochastic Extension and Predictive Distribution

I extend the axiomatic model to allow for error, embedding it in a statistical framework. This stochastic extension is constructed under the assumption that observed choices are the result of decision maker's maximization of her own utility function, $m \in \mathbf{M}$, plus an idiosyncratic error, truncated due to feasibility constraints,

$$q^j = m^j + \tilde{\varepsilon}^j \text{ with } (m^1, \dots, m^J) \in \mathbf{M}$$

This error can be understood as measurement error or error made by consumers when making decisions. The proposed approach does not rely on the nature of the error process; however, the predictive distribution constructed below does so. The reader interested in extending the construction proposed below for alternative error structures—for instance, error with respect to decision maker's perception of her own utility function or with respects to prices—needs to estimate the "candidate model" and the error process consistent with such alternative definition.

By definition, observed and rational choices are feasible, that is, $q \in \mathbf{B}(p, x)$ and $m \in \mathbf{M} \subseteq \mathbf{B}(p, x)$. Therefore the effective error process, $\tilde{\varepsilon}$, is restricted such that perturbed choices are feasible¹¹. Feasibility imposes two sets of constraints, nonnegativity

¹¹Alternatively one can consider the truncation of the distribution of the error process to its domain.

and affordability. First, for each budget set feasibility constraints restrict the distribution of the error process to lie on the budget hyperplane, that is such that $p^j \cdot \varepsilon^j = 0$. Nonnegativity constraints are imposed as follows,

$$q^j = \operatorname{argmin}_{\tilde{q}^j \in B^j} \|m^j + \varepsilon^j - \tilde{q}^j\| \quad (1.2)$$

for all $j = 1, \dots, J$, with $\varepsilon^j \sim F_\varepsilon|_{p^j \cdot \varepsilon^j = 0}$ and $m \in \mathbf{M}(\mathbf{B}(p, x))$ where F_ε is a well behaved distribution. Formally, I assume that, before imposing feasibility, the error process has a symmetric distribution with mean, median and mode 0, such that assigns decreasing probability to extreme values –the probability density function is decreasing in the absolute value of the error–. Formally,

Assumption 1 (Unconstrained Error Process) *The unconstrained process $\varepsilon^j \in \mathbb{R}^L$ is assumed to be i.i.d. with continuous and symmetric p.d.f. with mean, median and mode 0_L and such that $\nabla_\varepsilon f(\varepsilon^j)^T \operatorname{Diag}(\varepsilon^j) < 0$ for all $\varepsilon^j \neq 0$ and $\varepsilon \perp m$, for all $m \in \mathbf{M}(\mathbf{B}(p, x))$ and for all for $j = 1, \dots, J$.*

The stochastic extension is defined for the domain where the empirical implications of the model are defined. In doing so, it allows the recovery of a "candidate model" in the budget sets faced by the decision maker and from it a natural construction of the predictive distribution for choices, extending the "supporting set" to reflect noise. I call this extension \mathbf{M}^ε .

Extending the model to allow for error induces a new source of uncertainty when predicting behavior; I call this uncertainty "error uncertainty".¹² Therefore, I extend

The assumption adopted in this chapter is consistent with a fixed latent error process. The analysis could be carried out with this alternative assumption, but special attention should be paid to the estimation of the error process.

¹²In this chapter I abstract from uncertainty with respect to the systematic component that generated the data. When predicting behavior for an axiomatic model three sources of uncertainty arise: (i) "error uncertainty" as described above; (ii) "model uncertainty" since for a given sequence of rational choices the model yields a multiple prediction; and (iii) "DGP uncertainty" which refers to the uncertainty about the

the "supporting set" -Definition 1.5- to reflect this new source of uncertainty. First, I recover the revealed preference information from data by the way of a minimum distance estimator.

Definition 1.9 (Projection of observed choices onto the model) *The problem is to find the projection $\widehat{m}_{\mathbf{M}\otimes}$ of q onto \mathbf{M} , the minimum distance estimator, as the solution to*

$$\begin{aligned} \widehat{m}_{\mathbf{M}\otimes} &\equiv \operatorname{argmin}_m \|\operatorname{vec}(q) - \operatorname{vec}(m)\| \\ \text{s.t. } & m \in \mathbf{M}(\mathbf{B}) \end{aligned}$$

The estimation defined above cannot be performed following standard mathematical tools due to the topology of the set of sequence of choices that is consistent with the model. For GARP, (Kocoska 2012) proposes a series of derivative-free algorithms to find the "closest" rational sequence of choices to observed behavior, in the economic environments in which choices were observed. This projection requires to minimize the distance between observed behavior and rational choices, for each possible sequence of rational choices, that is, for each possible preference order.¹³

This projection yields an estimate for the "candidate model" and for the error process.¹⁴ Next, I construct the "supporting set" for the estimated "candidate model". As discussed before, its size reflects the "model uncertainty". Finally I compute the predic-

process that generated the data. For the main specification of the predictive distribution in this chapter I fix the data generating process to be the most likely one given data. In section 1.4.2 I proposed a model averaging alternative to account for uncertainty on the DGP

¹³Note that the results of this projection is dependent of measurement units. However, the projection technique is flexible to allow for the inclusion of weights, therefore one could weight by prices, or the ratio of prices to income, to remove the dependency with respect to the units of measurement.

¹⁴In section 1.3.4 I show an algorithm to recover the underlying distribution of the error process even for small samples.

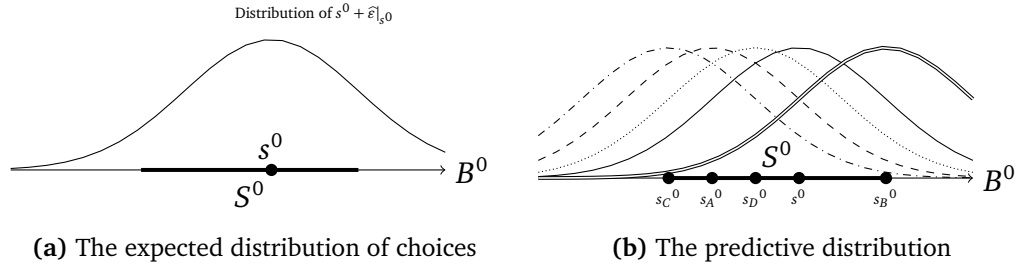


Figure 1.7: Predictive Distribution - Intuition. The predictive distribution in the case of only one dimension.

tive distribution by extending this construction to reflect the estimated error process.

Given assumption 1, the "candidate model" is a vector of choices that most likely generated observed choices, given data. For $\widehat{m}_{\mathbf{M}\mathcal{D}}$, the estimated "candidate model", each distinct sequence of choices $(\widehat{m}_{\mathbf{M}\mathcal{D}}, s^0)$ with $s^0 \in S^0(\widehat{m}_{\mathbf{M}\mathcal{D}})$ represents a distinct behavioral functional among the ones consistent with model \mathbf{M} . Expected choices conditional on $(\widehat{m}_{\mathbf{M}\mathcal{D}}, s^0)$ are given by the distribution of $s^0 + \widehat{\epsilon}|_{s^0}$, where the distribution of $\widehat{\epsilon}|_{s^0}$ is estimated from the residuals of the projection. Figure 1.7 shows the intuition for the predictive distribution conditional on s^0 . Thus, I define the predictive distribution as the distribution of $s + \widehat{\epsilon}|_s$ with $s \sim F_{S^0(\widehat{m}_{\mathbf{M}\mathcal{D}})}$. Without imposing further assumptions¹⁵ all $s^0 \in S^0(\widehat{m}_{\mathcal{D}})$ are equally likely, therefore $F_{S^0(\widehat{m}_{\mathbf{M}\mathcal{D}})}$ is assumed to be uniform on the supporting set. Figure 1.7b shows the expected distribution conditional different choices from the "supporting set". The predictive distribution is the average distribution for all such distributions.

Definition 1.10 (Predictive Distribution for B^0 given $\mathbf{M}\mathcal{D}$) *Let observed choices q be generated by a systematic component $m \in \mathbf{M}$ and an error process ϵ consistent with as-*

¹⁵If the researcher has prior information available, this definition can be extended to reflect this prior information. In particular, I will not longer be the case that all behavioral functionals constructed with different choices from the "supporting set" are equally likely. For a given prior, now the distribution of choices on the supporting set is given by the truncation of the prior to the "supporting set".

umption 1, such that

$$q = m + \tilde{\varepsilon} \text{ such that } m \in \mathbf{M}(\mathbf{B}(p, x)) \text{ and } q \in \mathbf{B}(p, x)$$

where $\tilde{\varepsilon}$ is the truncation of the error process to satisfy feasibility constraints -consistent with equation 1.2-. Let $\widehat{m}_{\mathcal{D}}$ be a consistent estimator of m , and define $\widehat{\varepsilon} = q - \widehat{m}_{\mathcal{D}}$. Then, I define the predictive distribution as the distribution of the random variable

$$S + \widehat{\varepsilon}(S) \text{ with } S \sim U_{S^0(\widehat{m}_{\mathcal{D}})} \text{ and } \widehat{\varepsilon}(S) \sim F_{\widehat{\varepsilon}}^0|S$$

where $F_{\widehat{\varepsilon}}^0|S$ is the truncation of the estimated distribution of the residuals consistent with assumption 1 such that $S + \widehat{\varepsilon}(S) \in B^0$, and $U_{S^0(\widehat{m}_{\mathcal{D}})}$ is an uniform distribution over the supporting set. Let $f_{\mathbf{M}\mathcal{D}}^0$ be the predictive density, then

$$f_{\mathbf{M}\mathcal{D}}^0(y) = \int_{s \in S^0(\widehat{m}_{\mathcal{D}})} \frac{1}{\int_{s \in S^0(\widehat{m}_{\mathcal{D}})} ds} \left[\int_{\eta: p^0 \cdot \eta = p^0 \cdot s} \mathbb{1} \{y = \operatorname{argmin}_{\tilde{y} \in B^0} \|\tilde{y} - \eta\|\} f_{\widehat{\varepsilon}}^0(\eta - s) d\eta \right] ds$$

where $f_{\widehat{\varepsilon}}^0(\xi) \propto \mathbb{1} \{p^0 \cdot \xi = 0\} f_{\varepsilon}(\xi)$

Definition 1.10 combines a (non-standard) frequentist projection tool with a quasi-bayesian¹⁶ model averaging technique to construct the predictive distribution. By definition, each distinct alternative from the "supporting set" represents and observationally distinct behavior among the ones consistent with the model. Then, the predictive distribution is constructed as the average distribution across all these behaviors; where the weights are given by their relative likelihood, following the intuition of BMA. The following example illustrates this construction.

Example (Continued 2.3) *The predictive distributions for the decision makers A and B*

¹⁶A fully Bayesian approach would require the definition of a suitable prior over the set of well-behaved preference orders and a mechanism for updating, this is left for future research.

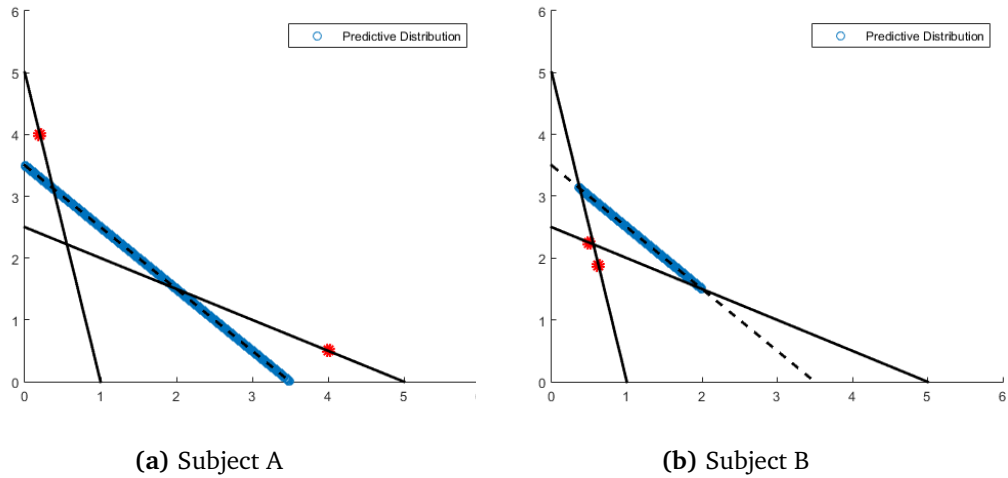


Figure 1.8: Relation between empirical consistency, falsifiability and predictive ability. Empirical consistency -fit- is important but does not imply usefulness of the model. Even when falsifiable it may not restrict behavior enough to allow for learning about underlying behavior.

discussed in Example 2.3 are depicted in Figure 1.8. For consumer A the projection is equal to observed choices, i.e. $\widehat{m}^A = q^A$; then estimated errors are 0. Therefore, the predictive distribution is given by a uniform distribution on the "supporting set" $S^0(\widehat{m}^A) = B^0$. On the other hand, choices observed for consumer B are not consistent with GARP. The closest rational sequence of choices to observed data is $\widehat{m}_1^B = q_1^B$ and $\widehat{m}_2^B = (\frac{5}{9}, \frac{20}{9})$, thus $\widehat{\varepsilon}_1^B = (0, 0)$ and $\widehat{\varepsilon}_2^B = (-\frac{1}{18}, \frac{1}{36})$. The "supporting set" is then defined by $S^0(\widehat{m}^B) = \{q \in \mathbb{R}_{++}^2 : q_1 + q_2 = 3.5 \text{ and } \frac{3}{8} \leq q_1 \leq 2\}$. The resultant predictive distributions confirm the intuition that the data for decision maker A, though consistent with the model, deliver less informative predictions than the data for decision maker B. Figure 1.9 shows the predictive density for q_1^A and q_1^B over B^0 .

The predictive distribution for model \mathbf{M} given observed data \mathcal{D} can be understood as the expected distribution of choices given data \mathcal{D} as prescribed by model \mathbf{M} . Its variance reflects the uncertainty of predicting behavior due to the error process –"error uncertainty"– and the systematic component –"model uncertainty"–. By construction, this distribution extends the "supporting set" to allow for error, and coincides with it if observed behavior is perfectly consistent with the model; case in which the predic-

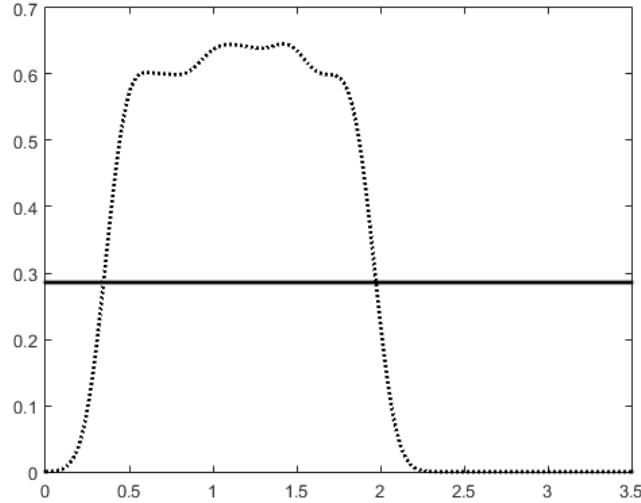


Figure 1.9: Predictive densities for Example 2.3. The predictive densities for good 1 in B^0 constructed from the information in Example 2.3 following definition 1.10.

tive distribution is given by an uniform distribution over the "supporting set". If the "supporting set" converges to a point under a sufficient information condition on the distribution of the budget sets –as it holds for GARP– then the predictive distribution converges to the distribution of the error process centered in the point prediction of the model. These properties are formalized in Theorem 1.11.

Theorem 1.11 (Properties of the Predictive Distribution) Let $f_{\mathbf{M}^\mathcal{D}}^0$ be the predictive density for B^0 based on the extended model \mathbf{M}^ϵ , conditional on data $\mathcal{D} = \{p^j, x^j, q^j\}_{j=1}^J$ constructed as in Definition 1.10. Let $Y_{\mathbf{M}^\mathcal{D}}^0$ be a random variable with probability density $f_{\mathbf{M}^\mathcal{D}}^0$.

1. Reflects recovered error process

If $S^0(\widehat{m}_{\mathbf{M}^\mathcal{D}-A}) = S^0(\widehat{m}_{\mathbf{M}^\mathcal{D}-B})$ and $\|\text{vec}(\widehat{\epsilon}_A)\| > \|\text{vec}(\widehat{\epsilon}_B)\|$ then

$$\text{Diag}(\text{Var}(Y_{\mathbf{M}^\mathcal{D}-A}^0 | Y_{\mathbf{M}^\mathcal{D}-A}^0 \in B(\bar{s}, \delta))) \geq \text{Diag}(\text{Var}(Y_{\mathbf{M}^\mathcal{D}-B}^0 | Y_{\mathbf{M}^\mathcal{D}-B}^0 \in B(\bar{s}, \delta)))$$

for all $\delta > 0$ such that $B(\bar{s}, \delta) \subset B^0$ where $\bar{s} = \int_{S^0(\widehat{m}_{\mathbf{M}^\mathcal{D}-A})} \frac{s}{\int_{S^0(\widehat{m}_{\mathbf{M}^\mathcal{D}-A})} ds}$

2. Reflects the tightness of identification

If $\|\text{vec}(\widehat{\varepsilon}_A)\| = \|\text{vec}(\widehat{\varepsilon}_B)\|$ and $S^0(\widehat{m}_{\mathbf{M}\varnothing-A}) \supset S^0(\widehat{m}_{\mathbf{M}\varnothing-B})$ then

$$\text{Diag}(\text{Var}(Y_{\mathbf{M}\varnothing-A}^0 | Y_{\mathbf{M}\varnothing-A}^0 \gg 0)) > \text{Diag}(\text{Var}(Y_{\mathbf{M}\varnothing-B}^0 | Y_{\mathbf{M}\varnothing-B}^0 \gg 0))$$

3. Limiting Behavior

For an increasing sequence of budget sets C^n such that $C^n \subset C^{n+1} \subset \dots$ and $\bigcup_n \{C^n\}$ is dense in $\mathbb{R}_{++}^{L \times n}$, with $q^n \in \mathbf{B}(C^n)$ such that $q^n \gg 0$, with $S^0(\widehat{m}_{\mathbf{M}\varnothing}^n) \rightarrow m^0$ then

$$F_{\mathbf{M}\varnothing}^0 \rightarrow m^0 + F_e^0 | m^0$$

where F_e^0 is the distribution of the underlying error process centered in m^0 truncated to B^0 .

4. Varian

If $q \in \mathbf{M}$ then $Y_{\mathbf{M}\varnothing}^0 \sim U_{S^0(q)}$ where $U_{S^0(q)}$ is a uniform distribution on $S^0(q)$.

Unless otherwise stated all proofs are relegated to Appendix 1.A.

The effect of the number of observation on the predictive distribution is, in principle, ambiguous. An additional observation potentially enhances predictive precision due to the shrinkage of the supporting set, Proposition 1.8. However, as the identification of the systematic component gets tighter, behavior previously consistent with the model would not longer be consistent, potentially increasing the variance of the error process and therefore increasing the dispersion of the predictive distribution.

This construction provides a predictive distribution that (i) is of interest for a researcher seeking to use the nonparametric model as a way of predicting behavior, this is shown in Section 1.4.1; and, (ii) allows for the assessment of the predictive ability,

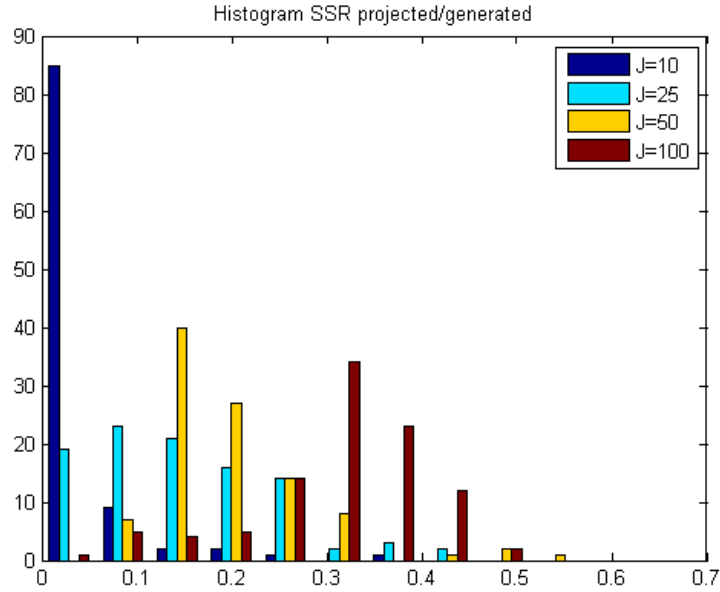


Figure 1.10: Histogram of $\frac{Projected-SSR}{Generated-SSR}$ for $J = 10, 25, 50, 100$, and 100 repetitions. Choices where generated from a Cobb-Douglas utility function $u(x, y) = \sqrt{xy}$ plus ε consistent with the truncation defined in section 1.3.1, where the underlying error process is $\varepsilon \sim N(0, 5)$. $\mathbf{B}(p, x)$ was generated such that $p_1 \sim U[5, 10]$, $\frac{p_2}{p_1} \sim U[.5, 2]$ and $w \sim U[80, 120]$.

which is shown in Sections 1.4.2 and 1.4.3.

1.3.4 On the Estimation of the Error Process

Under some conditions discussed above, GARP is point identified. Therefore, in the limit definition 1.9 provides a consistent estimator of m and consequently for the distribution of ε . Unfortunately, small samples are the norm rather than the exception when testing for axiomatic models of behavior¹⁷.

For any sample size, the residuals from the projection provide the relevant measure of the departures from the model. Nevertheless, the construction of the predictive dis-

¹⁷Without imposing homogeneity assumptions the relevant dimension is the number of observations per decision making unit. Many data sets have information available for a large number of subjects, but cross-sectional studies have been shown to fail when accounting for heterogeneity in behavior, see (Lewbel 2001).

tribution, requires a consistent estimate for the underlying error process.¹⁸ For small samples, the recovered residuals underestimate the underlying error process. Consider figure 1.10, where choices were generated in the context of the rationality model with $\varepsilon \sim N(0, 5)$ and $J = 10, 25, 50, 100$. The underestimation of the residual process is significant for all considered sample sizes. The gap between $\|\varepsilon\|$ and $\|\widehat{\varepsilon}\|$ is jointly defined by the error process and the sample size, given $\mathbf{B}(p, x)$.

In this section I propose a pseudo parametric approach for the estimation of the error process. The approach remains nonparametric on the identification of the systematic component but assumes a parametric assumption on the error process consistent with assumption 1. Let F_ε^θ be the distribution of the error process before the truncation, parametrized by θ . For example, if $F_\varepsilon^\theta = \mathcal{N}(\mu, \sigma^2)$, then $\theta \equiv (\mu, \sigma^2)$. The problem then reduces to consistently estimate θ .

Relying on the structure of GARP, I propose two algorithms to simulate the distribution F_ε^θ without computing the projection in each iteration.¹⁹ The first approach constructs the consistent set \mathbf{M} given \mathcal{D} by intersecting, in a sequential manner, the supporting set for each budget set. For each i chosen at random from the set of observations $\mathcal{J} \equiv \{1, \dots, J\}$, first it constructs $S^i(\hat{m}_{-i})$. Then chooses $k \in \mathcal{J} \setminus \{i\}$ and constructs $S^k(\hat{m}_{-\{i,k\}})$ and do so until \mathcal{J} is exhausted. The results for the estimation of this algorithm in a simulation setting are presented in Table 1.4. The results show that, even for a small sample of $J = 40$, the algorithm significantly improves the estimation

¹⁸Note that the predictive distribution here is constructed by incorporating the error process for each element in the "supporting set". Alternatively one could construct the prediction as the "supporting set" allowing for error in its boundary. If the researcher were to follow this alternative approach, then the relevant error process is the recovered error process.

¹⁹Ideally $\hat{\theta}$ can be estimated from the simulation of the effective (observed) error process for an exhaustive grid on θ , that is $\hat{\theta} = \operatorname{argmin}_{\theta_0} \|F_{\varepsilon}^{\theta_0} F_{\varepsilon}\|$, where F_{ε} is the observed distribution of the error process and F_{ε}^{θ} is the simulated effective distribution of the error process for $\theta = \theta_0$. Estimating $\hat{\theta}$ in this manner implies computing the projection of simulated data to the model for each iteration in the simulation for each possible value of θ which is really expensive computationally.

$u(\cdot)$	σ					
	0.5	1	2	3	5	10
$x_1^{\frac{1}{4}} x_2^{\frac{3}{4}}$	0.9040	0.7618	0.4052	0.6251	0.9218	0.9808
$x_1^{\frac{1}{2}} x_2^{\frac{1}{2}}$	0.9570	0.8816	0.5528	0.5149	0.8918	0.9794

Table 1.4: Simulations for the Algorithms. Kolmogorov-Smirnov two tailed test statistics with respect to measure constructed based on the projection for $\sigma = 3$ for two alternative distributions over a CD utility function.

of the underlying error process with respect a nonparametric estimation of the distribution of the residuals of the projection.

The second approach exploits the symmetry of f_e^θ -Assumption 1- to estimate θ as the parameters that maximize the hits to the set of rational choices from observed choices. These algorithms are formally defined in Appendix 1.B.1.

1.4 Predictive Performance Analysis

1.4.1 Decision Making and the Predictive Distribution

The predictive distribution is a sufficient statistic for a researcher that seeks to predict behavior based on the model. Consider a decision problem where the utility of the decision maker, $v(\cdot)$, depends on her own decision, a , and on the behavior of the consumer (or consumers), q , that could depend on a . The decision maker models consumer's decision by a model \mathbf{M} . Given observed data \mathcal{D} , the expected distribution of choices for a new economic environment, B^0 , conditional on a and \mathbf{M} , is the predictive distribution, $F_{\mathbf{M}\mathcal{D}}^0(q|a)$. The knowledge of the predictive distribution allows the decision maker to compute her optimal choice, $a^*|\mathbf{M} = \operatorname{argmin}_a \mathbb{E}_{F_{\mathbf{M}\mathcal{D}}^0(q|a)}[v(a, q(a))]$, and the expected value of her decision, $v^*|\mathbf{M} = \mathbb{E}_{F_{\mathbf{M}\mathcal{D}}^0(q|a^*)}[v(a^*|M, q(a^*|M))]$. The expected value of the problem can be interpreted as the negative of the loss due to the uncertainty when

predicting behavior by assuming the model \mathbf{M} . Under standard concavity assumptions on the preferences of the decision making, if consumers behavior affects the optimal choice, then higher uncertainty to predict behavior translates in lower expected utility.

To illustrate the relevance of the predictive distribution in decision making, consider the following example.

Example 1.12 (GARP vs Cobb-Douglas) *The producer of good 1 in this economy wants to project her revenue for B^0 . Her utility function is $v : \mathbb{R} \rightarrow \mathbb{R}_+$. For simplicity assume that there is only one consumer, whose choices maximize her utility function $u = \min \{2x, 3y\}$ without error, but this is unknown to the researcher. Choices are observed for the same budget sets as in Example 2.3. The researcher is considering two alternative models of behavior, GARP -model I- and Cobb-Douglas utility function-model II-. Figure 1.11 shows the predictive distribution for both models. Even if empirically inaccurate, the more specific model -model II- provides more information to predict behavior. Specifically, for GARP the Revenue is uniformly distributed in $[0.38, 3.45]$, with expected revenue 1.915 and variance 0.79. On the other hand, the Cobb-Douglas model yield a predictive distribution with two mass points in $\{1.16, 2.52\}$. For model II the expected revenue is 1.84 while the variance is 0.4624. The best model for the decision maker depends on her preferences, that is the best model is the one that maximizes her expected utility, i.e. $\int_{\text{supp}(q_{\mathbf{M}^\circ}^0)} v(p^0 \cdot \tilde{q}) f_{\mathbf{M}^\circ}^0(\tilde{q}) d\tilde{q}$. If $v = \text{Revenue}$ then she prefers GARP, while if the utility is concave in revenue, she may prefer model II.*

Given a set of candidate models, the decision maker can either exploit the information provided by all models to construct a average predictive distribution, where the weights for each candidate model is provided by their marginal likelihood. This construction is discussed in Section 1.4.2. If the decision maker seeks to choose one model from the set of candidates, she chooses the model that maximizes her expected utility

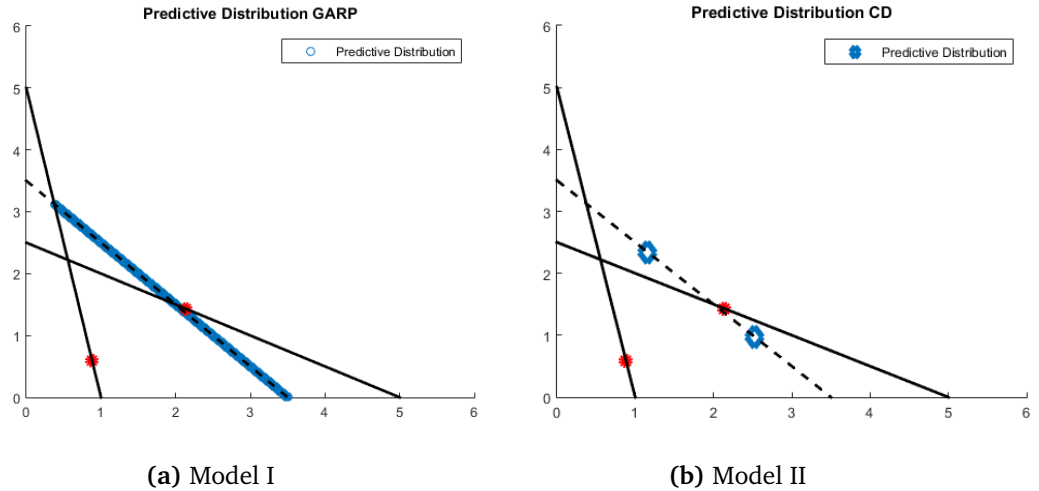


Figure 1.11: Relation between empirical consistency, falsifiability and predictive ability.

–or minimizes the expected loss–. The optimal model depends on the preferences of the decision maker $-v(\cdot)-$, the particular application $-B^0-$ and available data $-\mathcal{D}-$. Conditional on B^0 and \mathcal{D} , sufficient conditions have been proposed in the literature for a model to be better than others in terms of predictive information, see (Blackwell et al. 1951), (Cabrales, Gossner, and Serrano 2010). Such conditions cannot be generalized for any $v(\cdot)$, and do not provide a complete order. To allow for a complete comparison, in Section 1.4.3 I provide a series of statistics that summarize the predictive information extracted from data \mathcal{D} given the model \mathbf{M} .

1.4.2 Marginal Likelihood and Model Averaging

Given a set of k candidate models, $\mathbf{M}_1, \dots, \mathbf{M}_k$, the researcher interested in predicting behavior can use the predictive distribution to compute weights for model averaging. The model averaging predictive distribution (MAPD) is defined as the average of the predictive distribution for each of the candidate models constructed as in Definition 1.10, where the weights are constructed as the relative likelihood of each of these models. The MAPD induces a new source of uncertainty when predicting behavior, "DGP uncertainty", that refers to the fact that there are many plausible models that may

be generating observed behavior.

I follow the intuition of Leave-one-out (LOO) cross validation to compute the LOO marginal likelihood.²⁰ For each observation $j \in \{1, \dots, J\}$ the predictive distribution over B^j is constructed with the observations excluding j , as in definition 1.10. The likelihood to correctly predict behavior can be computed as

$$p(q^j | \mathbf{M}_{\mathcal{D}_{-j}}) = f_{\mathbf{M}_{\mathcal{D}_{-j}}}^j(q^j)$$

where $f_{\mathbf{M}_{\mathcal{D}_{-j}}}^j$ is the probability density function for predictive distribution $F_{\mathbf{M}_{\mathcal{D}_{-j}}}^j$ as defined in 1.10 for data \mathcal{D} for which observation j is not considered. This probability provides a measure of the predictive performance of the model in sample; the better the model the more accurate predictions it delivers and the higher the probability of correctly predicting observed behavior.²¹

Definition 1.13 (Model Evidence ("Marginal Likelihood")) *Let $F_{\mathbf{M}_{\mathcal{D}_{-j}}}^j$ be the predictive distribution as in Definition 1.10 constructed for B^j from data \mathcal{D} for which observation j is not considered and let $f_{\mathbf{M}_{\mathcal{D}_{-j}}}^j$ be its probability density function. Then,*

$$p(q | \mathbf{M}^\varepsilon) = \prod_{j=1}^J f_{\mathbf{M}_{\mathcal{D}_{-j}}}^j(q^j)$$

Given a prior distribution on the set of candidate models $\pi : (\mathbf{M}_1, \dots, \mathbf{M}_k) \rightarrow [0, 1]$,

²⁰By construction, the "posterior marginal likelihood" does not reflect "model uncertainty" since, if I were to include q^j when constructing the predictive distribution for B^j , the supporting set collapses to \widehat{m}^j and the predictive distribution collapses to the distribution of the error process. To reflect the true uncertainty for predicting behavior I proceed with the LOO predictive distribution.

²¹An alternative measure of the predictive performance in sample can be constructed from definition 1.15. For observation j define $C^j(q^j)$ as the biggest HPD set that does not contain q^j , then $1 - \mathbb{P}(C^j(q^j))$ is the confidence at which the researcher can be certain that the interval prediction for the model contains observed data.

the posterior probability of the model can be computed by

$$p(\mathbf{M}_k^\varepsilon | \mathcal{D}) \propto p(q | \mathbf{M}_k^\varepsilon) \times \pi(\mathbf{M}_k)$$

Then, the average predictive distribution, is defined as the weighted average of the predictive distributions for all candidate models. Formally,

$$f_{BMA(\mathbf{M}_1, \dots, \mathbf{M}_k) | \mathcal{D}}^0(y) = \sum_{i=1}^k p(\mathbf{M}_i^\varepsilon | \mathcal{D}) f_{\mathbf{M}_i^\varepsilon | \mathcal{D}}^0(y)$$

1.4.3 Measures of Predictive Ability

Predictive Model Assessment and Model Selection

The marginal likelihood as in Definition 1.13, can be understood as the probability of correctly predicting observed behavior under the extended model, when the prediction is performed considering the other $J - 1$ observations. As constructed, even if slightly different from the standard definition²², the marginal likelihood provides a measure of the evidence for the model. If the decision maker seeks to choose one model from the set of candidate models, this can be done from the marginal likelihood. The Bayes factor provides a summary of the evidence provided by the data in favor of one model

²²Note that if we were to include the observation j when computing $p(q^j | \mathbf{M}_k^\varepsilon)$, this collapses to the distribution of the error process around the "candidate model"; that is,

Definition 1.14 (Model Evidence ("Posterior Marginal Likelihood")) *Let $\widehat{m}_\mathcal{D}$ be a consistent estimator of m as in definition 1.10, and let f_ε be the estimation of the underlying error process from $\widehat{\varepsilon} = q - \widehat{m}_\mathcal{D}$. Then,*

$$\begin{aligned} p(q | \mathbf{M}^\varepsilon; m = \widehat{m}_\mathcal{D}) &= \prod_{j=1}^J \left\{ \mathbb{1}\{q^j \gg 0\} f_\varepsilon(q^j - \widehat{m}_\mathcal{D}^j) \right. \\ &\quad \left. + [1 - \mathbb{1}\{q^j \gg 0\}] \int_{\tilde{q}^j: q^j \equiv \arg \min_{a \in B^j} \|\tilde{q}^j - a\|} f_\varepsilon(\tilde{q}^j - \widehat{m}_\mathcal{D}^j) d\tilde{q}^j \right\} \end{aligned}$$

versus the other.²³

$$\text{Bayes Factor}_{i,k} = \frac{p(q|\mathbf{M}_i^\varepsilon)}{p(q|\mathbf{M}_k^\varepsilon)}$$

The critical value to decide for one model over other depends on the loss function. For interpretation it can be useful to consider twice the logarithm of the Bayes factor since it has the same structure as the familiar deviance and likelihood ratio test statistics.²⁴ Alternatively, the logarithm of the marginal likelihood of the data can be interpreted as a predictive score, then the logarithm of the Bayes factor can be interpreted as the difference on the predictive scores for the considered models.

Measures of Predictive Precision

Given a data set \mathcal{D} , a more informative model delivers tighter predictions. For a given model \mathbf{M} , a data set \mathcal{D}^A provides more information than other \mathcal{D}^B , if it allows for the construction of more informative predictive distributions. Then, a natural way to measure the performance of a $\mathbf{M}\mathcal{D}$ —program is by measuring the dispersion of its predictive distribution. The two summary statistics I propose are: (i) the size of the smallest α -level confidence interval and (ii) the Kullback-Leibler divergence measure with respect to an uninformative prior. The former follows the natural intuition that more useful models deliver "narrower" confidence sets, while the latter collapses on a measure of the entropy of the predictive distribution; better models reduce the uncertainty for predicting behavior.

Size of the Smallest α -level Confidence/Credible Interval The performance of the $\mathbf{M}\mathcal{D}$ program can be gauged by sizing the certainty to predict behavior by imposing the model \mathbf{M} given observed data \mathcal{D} . I measure it by the size of the smallest α -level confidence/credible connected set—that is the highest posterior density set among all

²³For an extensive analysis of Bayes factors and their interpretation refer to (Kass and Raftery 1995)

²⁴Its interpretation as the likelihood ratio test depends on defining one of the models as the null hypothesis and the other as the alternative.

connected subsets of B^0 . The smaller the α -level HPD interval (set) the more precise the predictions delivered by model \mathbf{M} given observed data \mathcal{D} . I define the predictive precision measure as the relative size of the complement of the HPD interval. Formally,

Definition 1.15 (Predictive Precision - Credible Set) *Let C_α^0 be the $100(1-\alpha)\%$ highest posterior density (HPD) connected set for the predictive distribution F^0 constructed given model \mathbf{M} . That is,*

$$C_\alpha^0 \equiv \operatorname{argmin}_{C_\alpha | C_\alpha \text{ is connected}} : \mathbb{P}_{F^0}(q \in C_\alpha) = 1 - \alpha$$

Conditional on data, $\{q^j, p^j\}_{j=1}^J$, the predictive precision α measure for budget set B^0 is defined by

$$PP_\alpha^0 |_{\mathbf{M}\mathcal{D}} \equiv 1 - \mathbb{P}[q^0 \in C_\alpha^0] \quad (1.3)$$

where $\mathbb{P} = U_{B^0}$ is the uniform distribution over the budget set B^0

The smaller C_α^0 , the bigger $[C_\alpha^0]^c$. Under (Becker 1962)'s model of irrational behavior, there is nothing that can be learned from data, therefore $PP_\alpha^0 |_{Becker} = \alpha$.²⁵ Then, $PP_\alpha^0 |_{\mathbf{M}\mathcal{D}} - \alpha$ can be interpreted as the improvement in precision of predicting behavior by using the model \mathbf{M} with respect to not using any model at all.

Information theory based measures An alternative way to measure the usefulness of the $\mathbf{M}\mathcal{D}$ -program is by sizing the gain in information to predict behavior by assuming model \mathbf{M} given data \mathcal{D} with respect to an uninformative prior. Let F^0 be the predictive distribution constructed for B^0 given model \mathbf{M} , and let \mathbb{P} be an uniform distribution over B^0 , $\mathbb{P} = U_{B^0}$.²⁶ I propose to measure the predictive ability of the model by the "distance" between these distributions. The proposed "metric" in this chapter

²⁵This is not the same as data being generated by this alternative data generating process and then imposing GARP as a model. This measure can be constructed to provide further information to the researcher about the power and significance of the measure

²⁶This measure can be constructed for any other probability measure that the researcher consider pertinent, the results and properties are qualitatively the same.

is the Kullback-Leibler divergence measure²⁷, while not a metric in the formal sense, the KL-divergence measure provides an intuitive measure of the gain in information by imposing the model.

Definition 1.16 (Predictive precision index - Information theory) *Let F^0 be the predictive distribution constructed for B^0 given model \mathbf{M} and data \mathcal{D} and let $\mathbb{P} = U_{B^0}$. Then,*

$$PP_{KL}^0 |_{\mathbf{M}\mathcal{D}} \equiv D_{KL}(F^0 || \mathbb{P}) \quad (1.4)$$

The connection between the Kullback-Leibler divergence measure and Shannon entropy,

$$D_{KL}(F^0 || \mathbb{P}) = -H(f^0) - E_{F^0}(\ln \mathbb{P})$$

Since $\mathbb{P} = U_{B^0}$, then $E_{F^0}(\ln \mathbb{P}) = \ln \frac{1}{n} = -\ln n$ where $n = \int_{B^0} dx$, then

$$PP_{KL}^0 |_{\mathbf{M}\mathcal{D}} = \ln n - H(f^0)$$

where $H(\cdot)$ is the Shannon entropy, and f^0 is the density function for F^0 , has been well established in the literature. The entropy of the predictive distribution is the uncertainty of predicting an outcome by using it. Then, maximizing this measure is equivalent to minimizing the entropy of the predictive distribution.

Comparison of $\mathbf{M}\mathcal{D}$ programs

An intuitive way to make the comparison across programs is based on the predictive precision measures. In order to use these measures independently of the selection of B^0 ,

²⁷If the researcher is interested in an alternative measure, e.g. Hellinger or Total Variation, all the results in this chapter follow through.

I propose a "leave-one-out" (LOO) estimation.²⁸ The leave-one-out predictive precision measure is constructed as an average, across observations, of the predictive precision measures constructed for B^j for model \mathbf{M} and data \mathcal{D}_{-j} , i.e. data \mathcal{D} where observation j is excluded. This average can be constructed as a simple arithmetic or geometric mean, weighted mean, median, or any other summary statistic that the researcher considers relevant.

Definition 1.17 (Leave-one-out Predictive Precision (PP^{LOO})) *The leave one out predictive precision measure for program $\mathbf{M}\mathcal{D}$ is defined as that is*

$$PP_{\mathbf{M}\mathcal{D}}^{LOO} = \mathcal{M} \left(\left\{ PP_{-j}^j \right\}_{j=1}^J \right)$$

where \mathcal{M} is an average summary statistic $\mathcal{M}(\cdot) \in \{\text{mean}, \text{median}, \dots\}$ and

$$PP_{-j}^j = PP_{\mathbf{M}\{\mathcal{D}\setminus\mathcal{D}^j\}}^j$$

The leave-one-out predictive precision removes the arbitrariness in the selection of budget sets while allowing to measure the tightness of the identification of underlying behavior for the observed budget sets. Section 1.4.4 shows the relationship between this approach and the power literature.

Let $PP^0|_{\mathbf{M}\mathcal{D}} \in \{PP_\alpha^0|_{\mathbf{M}\mathcal{D}}, PP_{KL}^0|_{\mathbf{M}\mathcal{D}}\}$. The proposed measures inherit the desirable properties from the predictive distribution, favoring more informative predictive distributions. Let \succ_0^P be the linear order induced by $PP^0|_{\mathbf{M}\mathcal{D}}$ on programs, that is $\mathbf{M}^A\mathcal{D}^A$ is said to be more informative than $\mathbf{M}^B\mathcal{D}^B$, i.e. $\mathbf{M}^A\mathcal{D}^A \succ_0^P \mathbf{M}^B\mathcal{D}^B$, if and only if $PP^0|_{\mathbf{M}^A\mathcal{D}^A} >$

²⁸Alternatively a "grid" simulation over a representative sequence of budget sets can be constructed. Computing the predictive precision measures for a sequence of budget sets and summarizing these provide the researcher with an informative measure of the predictive ability of the model given observed behavior. But in order to remove any dependency on the selection of the particular sequence of budget sets, this sequence needs to be exhaustive enough, incurring additional computational costs.

$PP^0|_{\mathbf{M}^A \mathcal{D}^A}$.

Definition 1.18 (Information order) *The information order, $\succeq_0^P P$, is defined by the partial order induced by measure $PP^0|_{\mathbf{M} \mathcal{D}}$ on programs. That is,*

$$\mathbf{M}^A \mathcal{D}^A \succ_0^P \mathbf{M}^B \mathcal{D}^B \Leftrightarrow PP^0|_{\mathbf{M}^A \mathcal{D}^A} > PP^0|_{\mathbf{M}^B \mathcal{D}^B}$$

Conditional on data \mathcal{D} , the information order defines a complete order on the set of models. Then a model \mathbf{M} is said to be predictive-wise more informative than $\tilde{\mathbf{M}}$, for B^0 conditional on \mathcal{D} if

$$\mathbf{M} \succ_{0|\mathcal{D}}^P \tilde{\mathbf{M}} \Leftrightarrow PP^0|_{\mathbf{M} \mathcal{D}} > PP^0|_{\tilde{\mathbf{M}} \mathcal{D}}$$

Analogously, conditional on the model \mathbf{M} , the information order defined a linear order on data sets. \mathcal{D} is said to be predictive-wise more informative than \mathcal{D}' conditional on model \mathbf{M} if

$$\mathcal{D} \succ_{0|\mathbf{M}}^P \mathcal{D}' \Leftrightarrow PP^0|_{\mathbf{M} \mathcal{D}} > PP^0|_{\mathbf{M} \mathcal{D}'}$$

Properties of Measures of Predictive Ability

The leave-one-out (LOO) predictive precision measures allow the researcher to assess the performance of the model-data program in terms of its predictive ability regardless of the particular forecasting application. The LOO predictive precision measures favor $\mathbf{M} \mathcal{D}$ programs that are (i) empirically more accurate and/or (ii) more specific. This is because these measures summarize the fit of the data for the considered model, and the extent of the uncertainty to identify the systematic model component from observed behavior, through the effect of the error process and supporting set respectively.

Theorem 1.19 (Program Selection) *Let $\mathbf{M} \mathcal{D}$ and $\tilde{\mathbf{M}} \mathcal{D}'$ be two alternative model-data programs. Assume that \mathbf{M} and $\tilde{\mathbf{M}}$ are two models with convex supporting sets and assumed*

that the predictive distributions are defined consistent with Definition 1.10. Then,

1. If $S^j(\hat{m}_{-j}|\mathbf{M}\mathcal{D}) = S^j(\hat{m}_{-j}|\tilde{\mathbf{M}}\mathcal{D}')$ for all j and $\|\text{vec}(\hat{\varepsilon})\| < \|\text{vec}(\hat{\tilde{\varepsilon}})\|$. Then there exists $\bar{\alpha} \in (0, 1)$ such that for all $\alpha < \bar{\alpha}$ $\mathbf{M}\mathcal{D} \succ_{\alpha}^P \tilde{\mathbf{M}}\mathcal{D}'$
2. If $\|\text{vec}(\hat{\varepsilon})\| = \|\text{vec}(\hat{\tilde{\varepsilon}})\|$ and $\mathbf{M} \subset \tilde{\mathbf{M}}$. Then there exists $\bar{\alpha} \in (0, 1)$ such that for all $\alpha < \bar{\alpha}$ $\mathbf{M}\mathcal{D} \succ_{\alpha}^P \tilde{\mathbf{M}}\mathcal{D}'$
3. For increasing dense sequences \mathcal{D}^n and \mathcal{D}'^n such that $S^0(\hat{m}^n) \rightarrow \{m^0\}$ $S^0(\hat{\tilde{m}}^n) \rightarrow \{\tilde{m}^0\}$ for any B^0 with $\hat{m}^n = \hat{\tilde{m}}^n$ consistent estimators for the systematic component for sequences of programs $\mathbf{M}\mathcal{D}^n$ and $\tilde{\mathbf{M}}\mathcal{D}'^n$ respectively, if $\|\text{vec}(\varepsilon)\| < \|\text{vec}(\tilde{\varepsilon})\|$ then $\mathbf{M}\mathcal{D}^n \succ^P \tilde{\mathbf{M}}\mathcal{D}'^n$

where $\mathbf{M}^A\mathcal{D}^A \succ^P \mathbf{M}^B\mathcal{D}^B \Leftrightarrow PP_{\mathbf{M}^A\mathcal{D}^A}^{LOO} > PP_{\mathbf{M}^B\mathcal{D}^B}^0$

Understanding the Contribution to Predictive Precision: Fit vs Power

The construction defined in 1.10 allows the disentanglement of the contribution of fit and the certainty of predicting the systematic component, to predictive precision. First, I construct the predictive distribution that would be obtained if observed behavior were equal to the projection. This "zero-error" predictive distribution reflects the uncertainty of predicting behavior only due to "model uncertainty". The predictive precision measures constructed based on this "zero-error" predictive distribution are better than for the overall predictive distribution, due to the effect of the error process. Thus $PP_{\mathbf{M}\mathcal{D}}^0|_{\varepsilon=0} \geq PP_{\mathbf{M}\mathcal{D}}^0|_{\hat{\varepsilon}}$, and

$$\text{Contribution of } \varepsilon \text{ to } PP_{\mathbf{M}\mathcal{D}}^0 = PP_{\mathbf{M}\mathcal{D}}^0|_{\varepsilon=0} - PP_{\mathbf{M}\mathcal{D}}^0|_{\hat{\varepsilon}}$$

1.4.4 Comparison to other measures in the literature

Comparison to goodness of fit measures

To address the sharpness of GARP, several goodness of fit measures have been proposed in the literature. Conditional on the sequence of observed budget sets, the proposed measures reflect other measures in the literature through the effect on the error process. I consider the comparison to three types of measures: (i) Houtman-Maks index (HM), (ii) Afriat and (iii) the Money Pump Index (MPI). While HM is a counting measure that measures the size of the largest subset of choices that is consistent with GARP, (ii) and (iii) are measures of the extent of the violations. Afriat's measure is defined as the minimum perturbation to income required to remove all inconsistencies, therefore reflecting the extent of the worst violation. MPI measures "amount of money that could be extracted from the consumer" that violates GARP, therefore providing a measure of the mean extent of these violations. Formally,

Definition 1.20 (Measures of fit) Set $\mathbf{M} = \text{GARP}$ and data set \mathcal{D} .

The **Houtman-Mask**²⁹ index is defined as

$$HM = \max \frac{|\mathcal{D}^M|}{J} \text{ such that } \mathcal{D}|_{J_M} \in \mathbf{M}|_{J_M}$$

where J_M is a subset of $\{1, \dots, J\}$.

The **Afriat**³⁰ index is defined as

$$e^* = \operatorname{argmax}_e \mathcal{D} \in \mathbf{M}(e)$$

where $\mathbf{M}(e)$ imposes the acyclicity condition on the Revealed Preference- e binary relation

²⁹(Houtman and Maks 1985)

³⁰(Afriat 1967)

where q^j is said to be revealed preferred to q^k at efficiency level e if $ep^j \cdot q^j > p^j \cdot q^k$.

The **Money Pump Index**³¹ for each sequence $(q^{k_1}, \dots, q^{k_n}) \notin \mathbf{M}$ is defined as

$$MPI_{(q^{k_1}, p^{k_1}), \dots, (q^{k_n}, p^{k_n})} = \frac{\sum_{l=1}^n p^{k_l} \cdot (q^{k_l} - q^{k_{l+1}})}{\sum_{l=1}^n p^{k_l} \cdot q^{k_l}}$$

Proposition 1.21 (Connection to Goodness of Fit Measures) Consider $\mathbf{M} = \text{GARP}$ and assume that $\mathcal{D}^j = \mathcal{D}'^j$ for all $j \neq i$, $q, q' \gg 0$, and $\hat{m} = \hat{m}'$. Then, there exists $\bar{\alpha} \in (0, 1)$ such that

1. If $HM(q) > HM(q')$ then $\mathbf{M}\mathcal{D} \succ_{\alpha}^P \mathbf{M}\mathcal{D}'$
2. If $Afriat(q) > Afriat(q')$ then $\mathbf{M}\mathcal{D} \succ_{\alpha}^P \mathbf{M}\mathcal{D}'$
3. If $MPI(q) < MPI(q')$ then $\mathbf{M}\mathcal{D} \succ_{\alpha}^P \mathbf{M}\mathcal{D}'$

for all $\alpha < \bar{\alpha}$

Proposition 1.21 shows that, conditional on the identification of the systematic component, the LOO predictive precision measure reflects the extent of the violations of the model. As discussed in section 1.2.2, high fit can be the result of behavior generated by a rational decision maker or the inability of the program to detect departures from the model. Therefore, the propose information order does not replicate the order defined by measures of fit in the literature but partially reflects the information provided by them.

Comparison to approaches that account for power

For a given model, there are two sources of uncertainty: the uncertainty to predict the systematic component and the error process. Measures of fit only provide information about the first of these sources of uncertainty. Good fit can hardly be interpreted as a

³¹(Echenique, Lee, and Shum 2011)

success for the model since conditions may be so undemanding that "anything goes". The standard approach in the literature is to assess the significance of measures of fit is to compare the results to those that would be obtained if data were to be generated randomly from a uniform distribution over feasible choices—(Bronars 1987)—, consistent with the definition of irrational behavior proposed by (Becker 1962). Given fit, higher power is desirable, and given power higher fit is desirable, but the ranking prescribed by these two properties tends to be negatively correlated and most approaches do not provide a way to combine fit and power.

(Beatty and Crawford 2011) (BC henceforth) incorporates (Bronars 1987) adjusting a measure of fit by the size of the target area, defined as the probability of choices being rational if they were to be generated by a random subject, that is, an ex-ante measure of power. Formally,

Definition 1.22 (Measure of Success ((Beatty and Crawford 2011))) *BC defines the measure of success as $m = r - a$ where $r \in \{0, 1\}$ is a measure of pass/fail with respect, $r = 1 \Leftrightarrow q \in \mathbf{M}(\mathbf{B})$ ³² and $a \in \{0, 1\}$ is a measure of the target area, that is the relative size of the set of consistent choices relative to the whole outcome space.*

BC measure provides a statistic that, by controlling for the stringency of the constraints imposed by the model, allows to compare across behaviors that may not have been observed in the same economic environments. By adopting a measure of power in the spirit of (Bronars 1987) the authors do not account for the actual pattern of choices observed. As it was argued in previous sections, this is a significant determinant when assessing whether the model is imposing demanding constraints on data or not.

³²BC also proposes a continuous version of success that reflects the distance between observed data and the set of sequences of choices that are consistent with the model; that is, $r \in [0, 1]$.

By construction, the predictive ability approach accounts for power conditional on choices, through the effect on the "supporting set". For each observation $j \in \{1, \dots, J\}$, the predictive precision measures constructed using observations $\{1, \dots, J\} \setminus \{j\}$, reflect the size of the "supporting set", $S^j(\hat{m}_{-j})$ –Theorem 1.19–. Then under (Becker 1962) hypothesis of irrational behavior, we have that,

$$\frac{\|S^j(\hat{m}_{-j})\|}{\|B^j\|} = 1 - \mathbb{P}[(\hat{m}_{-j}, q^j) \notin \mathbf{M}]$$

where $\hat{m}_{-j} \in \mathbf{M} | \mathcal{D} \setminus \mathcal{D}^j$ and $\mathbb{P} = U_{B^j}$. That is, the relative size of the "supporting set" is the complement probability of detecting violations to GARP in B^j under the assumption that choices are irrational, conditional on a sequence of choices for budget sets B^i with $i \neq j$ consistent with the model.

Exploiting this relation between the ability to infer information from data to predict behavior and power, the proposed measures tie together the literature on fit and power providing a meaningful trade off in terms of their contribution to predict behavior. By construction, the assessment of power in this chapter is done conditional on observed choices. In doing so, these measures do not replicate the information provided by BC measure. Next example illustrates the relevance of the difference between ex-ante and conditional power.

Example (Continued Example 2.3) Consider again choices presented in Example 2.3. The measures of predictive precision are $PP_\alpha^0 |_{\mathbf{M}^{\mathcal{D}^A}} = 0.05$, $PP_\alpha^0 |_{\mathbf{M}^{\mathcal{D}^B}} = 0.56$ for the measure defined in 1.15 and $PP_{KL}^0 |_{\mathbf{M}^{\mathcal{D}^A}} = 0.01$ $PP_{KL}^0 |_{\mathbf{M}^{\mathcal{D}^B}} = 0.21$ for decision maker A and B respectively. According to these measure GARP provides more precise predictions for \mathcal{D}^B than \mathcal{D}^A . By construction $q^A \in \mathbf{M}$ while $q^B \notin \mathbf{M}$, then success rate for subject A is bigger than for subject B, $r_A > r_B$, while the size of the target area is identical, by construction, for both subjects, $a_A = a_B$. Therefore (Beatty and Crawford 2011) would decide in favor of \mathcal{D}^A .

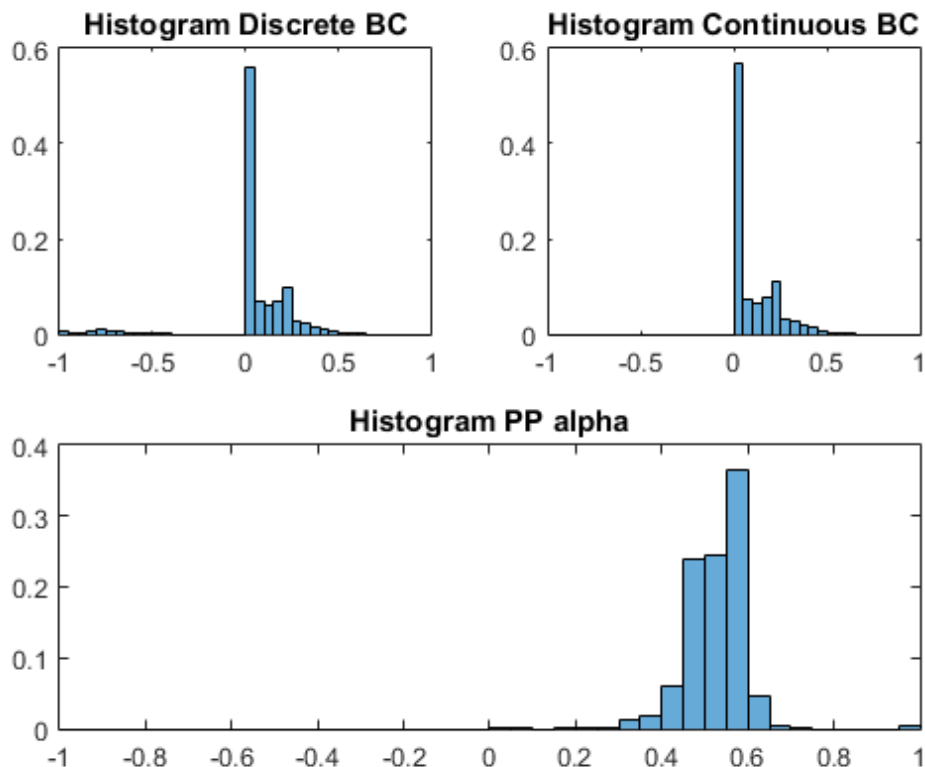


Figure 1.12: Comparison BC and predictive ability measures Data in (Beatty and Crawford 2011).

The difference between ex-ante power and a conditional assessment of power is important when studying the empirical performance of a model. In section 1.5 I show that, for the considered data set, power adjustments in the spirit of (Bronars 1987) become negligible, since ex-ante power rapidly converges to 1. That is, for small samples BC measure allows to account for differences in the design that could lead to differences in the power to identify inconsistencies, as shown in Figure 1.12, but these differences become indistinguishable as the number of observations increases. This seemingly uniform power across different subjects however mask significant differences on the amount of information that can be inferred from data to predict behavior. Figure 1.13 shows that, even for small samples, BC measure masks significant differences on predictive ability that show seemingly identical power.

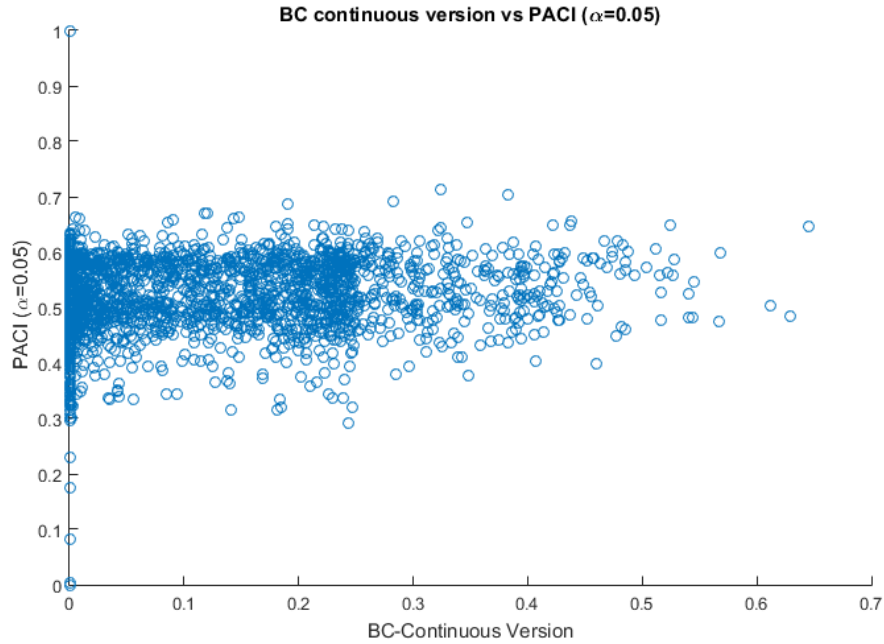


Figure 1.13: Comparison BC and predictive ability measures Even when BC measures allow to differentiate among designs when the sample size is small, this measure still masks significant differences in the predictive ability the model given observed behavior. These computations were realized for the data studied in (Beatty and Crawford 2011).

1.4.5 Optimal Design of Experiments

The proposed measures can be used for optimal experimental design. Consider a researcher that needs to design a sample/experiment to maximize the predictive power of a model \mathbf{M} and has available information about past behavior, $\mathcal{D}^J \equiv \{\mathcal{D}^j\}_{j=1}^J$.

Problem 1.23 Let B^0 be the objective budget set for which the researcher wants to predict behavior³³. The researcher has data on behavior for J economic environments, \mathcal{D}^J . She has been offered a set of experiments to run, and she seeks to optimize precision to predict behavior.

For a set of candidate experimental designs $B_1^{J+1}, \dots, B_d^{J+1}$, assuming the data generating process is stable across all observations, B^{J+1} can be chosen using the proposed

³³If there is not an application in mind the researcher can follow the approach presented in Section 1.4.2 to maximize the amount of information to identify underlying behavior

measures of predictive precision. If the objective is to predict behavior, the optimal selection of B^{J+1} is the one that improves the predictive precision the most. To construct the predictive distribution including B^{J+1} , I first construct the predictive distribution for B^{J+1} conditional on data from observations $\{1, \dots, J\}$. Then, either (i) construct a point prediction from the predictive distribution by computing its mean, median or highest likelihood; or (ii) use the information from the predictive distribution to compute a weighted measure of predictive precision for B^0 where weights are given by their likelihood given $F^{J+1} | \mathbf{M}^{\mathcal{D}^J}$.

Definition 1.24 (Optimal Design - Point Prediction) Let $\hat{q}^{J+1} \equiv \operatorname{argmax}_{\tilde{q}} f^{J+1}(\tilde{q})$, where $f_{\mathbf{M}^{\mathcal{D}^J}}^{J+1}(\cdot)$ is the predictive density for budget set B^{J+1} constructed using \mathcal{D}^J . The optimal selection is defined by

$$B_{\text{optimal}}^{J+1} \equiv \operatorname{argmax}_{\tilde{B}^{J+1}} PP_{\mathbf{M}^{\tilde{\mathcal{D}}^{J+1}}}^0$$

where $\tilde{\mathcal{D}}^{J+1} \equiv \mathcal{D}^J \cup \{\hat{q}^{J+1}, \tilde{p}^{J+1}, \tilde{x}^{J+1}\}$ and let $\tilde{B}^{J+1} \equiv B(\tilde{p}^{J+1}, \tilde{x}^{J+1})$ for some prices \tilde{p}^{J+1} and income \tilde{x}^{J+1} such that $\tilde{B}^{J+1} \neq B^i$ for all $i \in \{1, \dots, J\}$.

Definition 1.25 (Optimal Design - Full Predictive Distribution) Let $f_{\mathbf{M}^{\mathcal{D}^J}}^{J+1}(\cdot)$ be the predictive density for budget set B^{J+1} constructed using \mathcal{D}^J , then the optimal selection is defined by

$$B_{\text{optimal}}^{J+1} \equiv \operatorname{argmax}_{\tilde{B}^{J+1}} \int_{\tilde{B}^{J+1}} f_{\mathbf{M}^{\mathcal{D}^J}}^{J+1}(\tilde{q}^{J+1}) PP_{\mathbf{M}^{\tilde{\mathcal{D}}^{J+1}}}^0 d\tilde{q}^{J+1}$$

where $\tilde{\mathcal{D}}^{J+1} \equiv \mathcal{D}^J \cup \{\tilde{q}^{J+1}, \tilde{p}^{J+1}, \tilde{x}^{J+1}\}$ and let $\tilde{B}^{J+1} \equiv B(\tilde{p}^{J+1}, \tilde{x}^{J+1})$ for some prices \tilde{p}^{J+1} and income \tilde{x}^{J+1} such that $\tilde{B}^{J+1} \neq B^i$ for all $i \in \{1, \dots, J\}$.

The example below illustrates how this criterion works.

Example (Continued 2.3) Consider again the decision maker A from example 2.3. The researcher is considering expanding the sample to consider an additional observation from a third budget set B^3 and wants to select it to improve its forecast on B^0 . She is considering two alternatives B_a^3 and B_b^3 as shown in table 1.5.

Obs	p_x	p_y	x	\hat{q}^3
3_a	1.75	1.25	5	(1.427, 2.002)
3_b	$\frac{3}{4}$	2	5	(1.255, 3.321)

Table 1.5: Experimental design. Data for consumer A, choosing from B^1 and B^2 . The researcher is considering expanding the sample to include a third observation from B^3 to tight the prediction for B^0 . The two alternatives he is considering are B_a^3 and B_b^3 . \hat{q}_a^3 is computed as in definition 1.24

The computation for the expected predictive precision if budget set B^3 is incorporated follows Definition 1.24. The estimated \hat{q}^3 is displayed in the table above. For $\tilde{\mathcal{D}}_a^3$ the predictive precision measures for forecasting on B^0 are given by $PP_\alpha^0 | \mathbf{M}_{\tilde{\mathcal{D}}_a^3} = 0.39$ and $PP_{KL}^0 | \mathbf{M}_{\tilde{\mathcal{D}}_a^3} = 0.12$; while for $\tilde{\mathcal{D}}_b^3$ the predictive precision measures for forecasting on B^0 are given by $PP_\alpha^0 | \mathbf{M}_{\tilde{\mathcal{D}}_b^3} = 0.05$ and $PP_{KL}^0 | \mathbf{M}_{\tilde{\mathcal{D}}_b^3} = 0.01$. For predicting behavior on B^0 , it is expected that B_a^3 aids forecasting to a larger extent than B_b^3 . The respective predictive distributions are depicted in Figure 1.14.

Note that if the researcher seeks to design the sample but does not have any information available, the best it can be done is to maximize the ex-ante probability of detecting violations for B^0 , that is, power as it is traditionally understood in the revealed preference theory.

1.5 Empirical Application

In this section I use data from (Choi, Fisman, Gale, and Kariv 2007) to show that: (i) more observations lead to more precise predictions, (ii) the proposed measures provide additional information to value the usefulness of a model and do not replicate the ranking prescribed by other measures in the literature, and finally (iii) I use the proposed measures to compare the predictive performance of GARP versus the maximization of

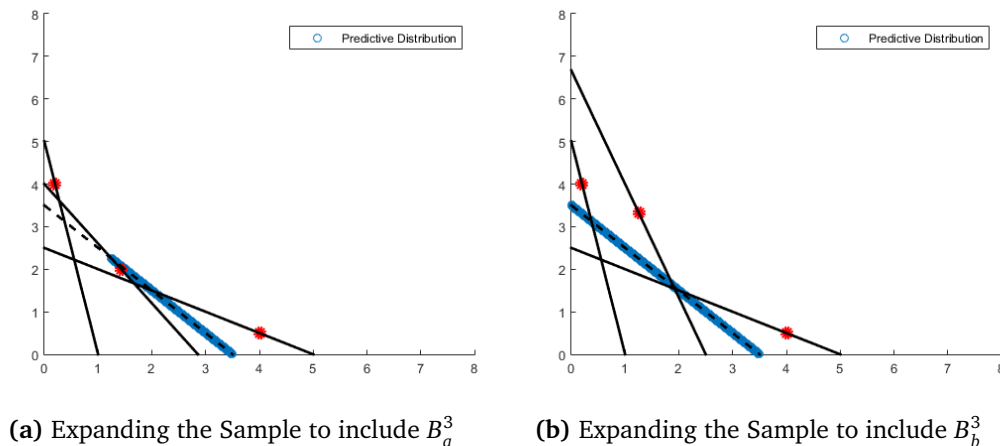


Figure 1.14: Optimal experimental design for prediction

a Cobb-Douglas utility function.

This data set was obtained from a series of experiments designed to study decisions under uncertainty. The 93 subjects in the experiment are presented with a graphical interface displaying standard, two dimensional, budget constraints on the screen. The experiment was conducted at the Experimental Social Science Laboratory at UC Berkeley and each session consisted of 50 independent decision rounds. Summary statistics are shown in Appendix 1.C.

1.5.1 Effect of Additional Information

Additional information enhances the identification of the systematic component which has two effects: (i) improves the precision due to the shrinkage of the supporting set and (ii) improves the identification of the underlying error process which increases uncertainty of predicting behavior. Ex-ante, the net effect is ambiguous. Intuitively, an additional observation from a budget set that intersects at least one of the previously observed ones, may improve the identification of the systematic component by making previously consistent behavior not longer congruent with the model. But, if that is the case, the distance between that observation and the closest consistent one, given by the

	Sample Mean for 25(half) and 50(full) observations per subject				
	Full	Half	Difference	Test-statistic	P-value
R^2	0.9916 (0.0023)	0.9968 (0.0020)	-0.0052 (0.0032)	1.6403	0.1051
Adj R^2	0.1071 (0.0072)	0.1118 (0.0092)	-0.0046 (0.0109)	0.4220	0.6742
Mean ex-ante prob	0.1124 (0.0084)	0.0967 (0.0095)	0.0157 (0.0115)	1.3663	0.1759
BC with r discrete	0.2078 (0.0465)	0.6232 (0.0556)	-0.4154 (0.0725)	-5.7306***	0.0000
BC with r continuous	0.9923 (0.0022)	0.9967 (0.0020)	-0.0044 (0.0030)	-1.4482	0.1492
Size supporting	0.5219 (0.0062)	0.6301 (0.0116)	-0.1081 (0.0108)	-9.9913***	0.0000
PP CI (95%)	0.5186 (0.0056)	0.4246 (0.0072)	0.0941 (0.0063)	14.9986***	0.0000
PP KL	0.2572 (0.0096)	0.1664 (0.0058)	0.0909 (0.0112)	8.1483***	0.0000

Table 1.6: Summary statistics for (Choi, Fisman, Gale, and Kariv 2007) - half and full sample. Data for 77 subjects, (Choi, Fisman, Gale, and Kariv 2007), test for differences in means. Standard error in (). BC with r discrete defines $r = 1$ if and only if observed data is consistent with GARP and 0 otherwise; while BC with r continuous defines $r = R^2$ from the projection.

change in the magnitude of the observed error process.

To study the effect of additional information in the ability to predict behavior, I compare the predictive ability constructed when considering the first 25 observations with the same measures constructed considering the entire sample.³⁴ The results are presented in table 1.6. This table displays the mean across subjects for measures of fit, adjusted fit, BC measures and the predictive precision measures –PP–, for the full sample –first column– and the first half of the sample –second column–. The third column displays the difference between column 1 and column 2; while columns 4 and 5 show the t-statistic and p-value for the hypothesis that these means are not significantly different.

Raw fit and the mean ex-ante probability of detecting violations are not significantly

³⁴For a robustness check, the same exercise is realized by splitting the sample into odds and even observations, using the former to construct the measures and compare to the whole sample. The results are not significantly difference. Table 1.12 in Appendix 1.C

different when comparing 25 versus 50 observations. Therefore there is evidence that preferences are stable and that the process for budget set is stable as well.

The BC measure provides contradictory information when using the discrete or the continuous measure of success. This is explained by the fact that, for the considered data, the relative size of the target set is approximately 0 for all subjects. Thus, the discrete version adjusts a binary measure of whether the data satisfy GARP by a measure that is approximately 0 for all subjects, therefore providing the same information as the deterministic test. Since the number of rational individuals drops from 63% to almost 20% when moving from 25 to 50 observations, the discrete measure favors smaller samples. On the other hand, the continuous version of BC measure collapses to the measure of success considered, therefore it cannot significantly detect differences when considering 25 of 50 observations, since there are not significant differences in fit.

In contrast, the predictive precision measures improve with additional observations. The precision when predicting behavior is significantly improved due to the enhanced identification of the systematic component which is shown in the shrinkage of the "supporting set". Although the difference in the ex ante mean probability of detecting violations is not significant, the overall probability increases with the sample size, which explains the significant difference of the mean size of the supporting set. Nevertheless, the effect of conditional power is not negligible even for the full sample, providing information about the stringency of the constraints across subjects in contrast to BC measure. This is further studied in next section.

	Subject 516	Subject 313
R^2	0.9998	0.9999
Varian	0.9750	0.9700
BC	0.9998	0.9999
$R^2 - (1 - \overline{P(viol)})$	0.2836	0.2931
PP CI (95%)	0.6096	0.4867
PP KL	0.3134	0.2636
Size supporting	0.4335	0.5713

Table 1.7: Statistics comparison. Data for subjects 516 and 313, (Choi, Fisman, Gale, and Kariv 2007)

1.5.2 Empirical Performance of Predictive Ability Measures and Comparison to Other Measures in the Literature

The difference between the predictive measures and BC measure of success can be shown by comparing two subjects. Consider subjects 516 and 313, the choices for these subjects are shown in Figure 1.15 in the Appendix. Figure 1.16 in the Appendix, shows the predictive distribution for good 1, constructed based on the data for these two subjects for four randomly generated environments. For some of these new economic environments, the data indeed provides distinct information to predict behavior –Panels 1.16a and 1.16b– while for others the difference is subtle or almost none –Panels 1.16c and 1.16d–. Table 1.7 shows that, even when standard goodness of fit measures and adjusted ones³⁵ would indicate that the performance of the model is better for subject 313, subject 516 generates a more informative predictive distribution. This is because even if ex ante the design seems to be more demanding for subject 313, a closer look reflects that, given observed choices, more information can be inferred from choices for subject 516, as can be seen from the comparison of the mean size of the supporting set.

³⁵In the spirit of the measure proposed by (Beatty and Crawford 2011), but considering the mean ex-ante probability of detecting probability instead of the overall probability to capturing differences in experimental design across individuals.

1.5.3 Model Comparison: GARP vs. Parametric Specification

I use the predictive approach for the comparison between the nonparametric rationality model and a nested parametric model where choices are generated by a Cobb-Douglas utility function plus error. The parameters that characterize the preferences of each decision maker are allowed to be individual specific.

By assuming a parametric DGP that delivers a point prediction, one may expect the predictive distribution for the Cobb-Douglas model to be more informative than for a more general model (GARP) that comprises it. But in being more restrictive, the parametric model induces misspecification error which increases its entropy. In the considered sample, the latter effect dominates, as can be seen from the results in table 1.8, and the difference is statistically significant. The only case in which the parametric model outperforms GARP is for subject 603, given her irregularities³⁶ throughout all measures of behavior has not been taken into account in the empirical analysis performed in previous sections.

1.6 Relation to the Literature

This chapter relates to two large bodies of literature: (i) revealed preference testing, measures of fit and power adjustments for GARP and (ii) model comparison and model selection. By proposing the use of the predictive distribution to measure the success of a nonparametric model of utility maximization while allowing for error. To my knowledge, the construction of such distribution is novel in the literature. (Adams 2013) allows for error, but identifies only a point forecast estimate by applying the "minimum discrimination information principle". The approach developed in this chapter identifies instead a predictive distribution, which is both more informative and can be used

³⁶Varian Index < 0.3

to measure the quality of the model.

Microeconomic theory has approached the problem of assessing the quality of the model when observed choices are not perfectly rational by proposing goodness of fit measures based on some intuitively appealing moment of the data associated with the monetary cost of the departures from the model. Section 1.4.4 discusses the relation between the predictive performance approach developed in this chapter with the most salient approaches for goodness of fit in the literature.

A common problem that plagues the literature of goodness of fit is the effect of the relative distribution of budget sets on the ability to detect departures from rationality. The literature has approached the problem by studying the "power" of the rationality test; understood as the probability of finding violations given observed economic environments for some alternative behavioral process. (Beatty and Crawford 2011) proposes a measure of predictive success that combines fit and power following the axiomatization provided by (Selten 1991), the relation between this measure and the approach presented in this chapter is shown in Section 1.4.4. In a similar vein, (Hoderlein and Stoye 2009) and (Andreoni and Harbaugh 2013) tackle this problem by proposing alternative measures of power. The predicting precision measures proposed in this chapter combine fit and power by sizing the informational content of the predictive distribution providing a meaningful trade off between the two.

This chapter uses a projection technique to identify the systematic component and the error process from data. Others have proposed the use of similar techniques as (Varian 1985) and (Fleissig and Whitney 2005), but I further impose feasibility constraints. Under the behavioral assumption that choices are rational but observed with error it naturally follows that the rational component must lie in the set of feasible

	GARP	CD	Difference	CI for Difference	
PP CI	0.5140 (0.0059)	0.2436 (0.0157)	0.2702 (0.0149)	0.2406 t-statistic	0.2999 18.14
PP KL	0.3422 (0.0104)	0.2552 (0.0094)	0.0870 (0.0131)	0.0609 t-statistic	0.1131 6.64
$\ vec(\hat{\varepsilon})\ $	558.49 (187.67)	20504.6 (2301.9)	-19946.1 (2271.3)	-24466.2 t-statistic	-15246.1 -8.78

Table 1.8: Model Comparison. Comparison GARP vs a Cobb Douglas utility function - Means choices; and by definition observed choices are feasible as well. (Halevy, Persitz, and Zrill 2014) proposes a projection technique that considers the feasibility constraints but it is based on a parametric assumption over the utility function—that is, the data is projected onto the subset of feasible choices that are consistent with the assumed parametric form, inducing misspecification error. The projection technique employed in this chapter only relies on the nonparametric constraints imposed by GARP.

Finally, this chapter relates to papers in the model selection literature. The information criteria literature deals with trade-off between the goodness of fit of the model and the complexity of the model providing tools for model selection, see (Akaike 1998) and (Geisser and Eddy 1979). (Geisser and Eddy 1979) deals with model comparison in the context of parametric models, Section 1.4.2 shows how the predictive distribution proposed in this chapter can be used for such constructions.

1.7 Conclusion

Rationality is one of the most prevalent assumptions in economics but empirically is almost surely violated. GARP provides an elegant nonparametric test for rationality, but only one violation causes the data to be declared as inconsistent. The aim of this chapter is to assess the quality of these models when the data may not pass the deterministic test relying solely on the necessary and sufficient conditions provided by the axiomatization. I develop a framework that, combining fit and the uncertainty to

predict consistent behavior, gauges the quality of the model by its predictive ability given observed data. By extending the nonparametric prediction for GARP provided by (Varian 1982) to allow for error, I provide a novel construction for the predictive distribution of axiomatic models. This distribution is a sufficient statistic for a decision maker that seeks to select a model to predict behavior. For a researcher that does not have an application in mind the predictive distribution can be used to compute measures of the performance of the model based on marginal likelihood and two intuitive statistics that summarize the predictive performance of the model.

This approach provides a joint measure of the quality of the model-data, therefore allowing for the comparison across data sets and/or behavioral models. The tradeoff between these two features of the model is made conditional on observed data, therefore outperforming other measures of fit prevalent in the literature. Other measures that combine fit and power in the literature do not allow to differentiate among many model-data sets since power adjustments rapidly become insignificant. In contrast, the predictive precision measures allow to differentiate between cases that seemingly have the same power, but deliver significantly different predictive distributions.

The proposed predictive precision measures exhibit, theoretically and empirically, the desired properties. Noisier data performs worse and more demanding environments deliver more informative predictive distributions. Moreover, increasing the number of observations allows the researcher to extract more information from data, which results in more precise predictions.

By combining fit and power, the predictive precision measures are more informative than standard measures in the literature. By relying solely on the axiomatization provided by GARP, without any further assumption on behavior and allowing for individual

specific behavior, I avoid misspecification biases common in the recoverability/demand estimation literature. By imposing the constraints implied by the model when projecting the data, I overcome the concerns raised by (Lewbel 2001) with respect to the common econometric approaches. Finally, the framework proposed in this chapter allows for model comparison featuring a natural trade off between the generality of the model and its empirical accuracy.

1.A Proofs

1.A.1 Proofs of Section 1.3

Proof of Theorem 1.11. (1) From Definition 1.10 the predictive distribution, conditional on data \mathcal{D} is given by

$$f_{\mathbf{M}\mathcal{D}}^0(y) = \int_{s \in S^0(\widehat{m}_{\mathbf{M}\mathcal{D}})} \frac{1}{\int_{s \in S^0(\widehat{m}_{\mathbf{M}\mathcal{D}})} ds} f_\varepsilon(x-s) \mathbb{1}[y = \operatorname{argmin}_{\tilde{y} \gg 0} \|\tilde{y} - x\|] \frac{\mathbb{1}[(x-s) \cdot p^0 = 0]}{\int_{x:(x-s) \cdot p^0 = 0} f_\varepsilon(x-s) dx} ds \quad (1.5)$$

Given assumption 1, the distribution of $Y_{\mathbf{M}\mathcal{D}}^0$ on $B(\bar{s}, \delta) \subset B^0$ with

$$\bar{s} = \int_{S^0(\widehat{m}_{\mathbf{M}\mathcal{D}-A})} \frac{s}{\int_{S^0(\widehat{m}_{\mathbf{M}\mathcal{D}-A})} ds} ds$$

is symmetric around \bar{s} , and its mean is precisely \bar{s} . Then, given that $Y_{\mathbf{M}\mathcal{D}-A}^0$ and $Y_{\mathbf{M}\mathcal{D}-B}^0$, only differ on the variance of the error process, for all $\delta > 0$ such that $B(\bar{s}, \delta) \subset B^0$

$$E[Y_{\mathbf{M}\mathcal{D}-A}^0 | Y_{\mathbf{M}\mathcal{D}-A}^0 \in B(\bar{s}, \delta)] = E[Y_{\mathbf{M}\mathcal{D}-B}^0 | Y_{\mathbf{M}\mathcal{D}-B}^0 \in B(\bar{s}, \delta)] = \bar{s} \quad (1.6)$$

Furthermore, distribution 1.5 truncated to $B(\bar{s}, \delta)$ is given by

$$f_{\mathbf{M}\mathcal{D}}^0(y | y \in B(\bar{s}, \delta)) = \frac{1}{P_{f_{\mathbf{M}\mathcal{D}}^0}(y \in B(\bar{s}, \delta))} \frac{1}{\mathbf{S}} \int_{s \in S^0(\widehat{m}_{\mathbf{M}\mathcal{D}})} f_\varepsilon(y-s) \frac{1}{A(s)} ds \quad (1.7)$$

where $A(s) \equiv \int_{x:(x-s) \cdot p^0 = 0} f_\varepsilon(x-s) dx$ and $\mathbf{S} \equiv \int_{s \in S^0(\widehat{m}_{\mathbf{M}\mathcal{D}})} ds$. Note that, before the truncation the predictive distribution is a mixture of distributions with same variance but different mean, where the weighting scheme is dictated by the uniform distribution on the supporting set. Therefore, the variance of the mixture is given by a weighted average of each of the mixing components, that is the variance of the error process, plus the variance of the distribution for location parameters, which in this case would be the variance of $S \sim U_{S^0}$.

By construction, the density of $Y_{\mathcal{M}^\varnothing}^0$ is symmetric around \bar{s} before the truncation, and it is also symmetric for any $B(\bar{s}, \delta)$, such that $B(\bar{s}, \delta) \subset B^0$. Moreover, by convexity of the supporting set, it follows that the variance of the symmetric truncation is a function of the overall variance and the result follows, that is $\|vec(\hat{\epsilon}_A)\| > \|vec(\hat{\epsilon}_B)\|$, then $Diag(\Sigma_A) \geq Diag(\Sigma_B)$ and the result follows.

(2) From previous point we have that the effect of a bigger supporting set on the untruncated predictive distribution is through the variance the distribution of the location parameter on the supporting set. By the properties of the supporting set it follows that for $S_A \sim U_{S^0(\widehat{m}_{\mathcal{M}^\varnothing-A})}$, and $S_B \sim U_{S^0(\widehat{m}_{\mathcal{M}^\varnothing-B})}$ with $S^0(\widehat{m}_{\mathcal{M}^\varnothing-A}) \supset S^0(\widehat{m}_{\mathcal{M}^\varnothing-B})$ then $Var(S^A) > Var(S^B)$. Therefore the variance of the predictive distribution before the truncation, depends on the variance of the location parameter, therefore on the size of the supporting set.

(3) and (4) follow by Definition 1.10, where the convergence of the distribution to the distribution of the error process follow by standard consistency properties of the least squares estimation when the truncation is not binding ($q \gg 0$). ■

1.A.2 Proof of Section 1.4.3

Proof of Theorem 1.19. From theorem 1.11 we know that the untruncated distribution increases its variance when variance of the error process or the size of the supporting set increases. Then, by definition 1.15, the PP_α is defined as the smallest and connected set C^α such that $P_{F_{\mathcal{M}^\varnothing}^0}(y \in C^\alpha) \geq 1 - \alpha$. Moreover, given convexity of the supporting set, this set would contain $\bar{s} = \frac{\int_{s \in S^0} s ds}{\int_{s \in S^0} ds}$, for all $\alpha < \bar{\alpha}$ with $\bar{\alpha} = 1 - \max_{y \in B^0 \setminus \mathbb{R}_{++}^L} f_{\mathcal{M}^\varnothing}^0(y)$. Then (1) and (2).

Finally, for two point identify models, the predictive distribution converges to the

distribution of the error process, Theorem 1.11, truncated due to feasibility constraints. As in (1), conditionally in the supporting set, in this case given $\widehat{m}^n = \widehat{m}^m$; if $\|\text{vec}(\varepsilon)\| < \|\text{vec}(\tilde{\varepsilon})\|$ then the variance of the untruncated distribution for $\widetilde{\mathbf{M}}^{\mathcal{D}'}$ increases. Defining $\bar{\alpha}$ as above, the result follows. ■

1.A.3 Proofs of Section 1.4.4

Proof of Proposition 1.21.

1. Let $Q^A \subset \{1, \dots, J\}$ be the maximal subset of choices that is consistent with the model, that is $q_{Q^A} \in \mathbf{M}(\mathbf{B}(\{p^i, x^i\}_{i \in Q^A}))$, then $HM(q_A) = \frac{\|Q^A\|}{J}$. $HM(q_A) > HM(q_B) \Rightarrow Q^A > Q^B$, then it must be the case that $\widehat{\varepsilon}_A^i \cdot \widehat{\varepsilon}_A^i < \widehat{\varepsilon}_B^i \cdot \widehat{\varepsilon}_B^i$, and the results follows from Theorem 1.19.
2. $Afriat(q_A) > Afriat(q_B) \Rightarrow \exists j \in \{1, \dots, J\}$. Let $e_A = Afriat(q_A)$ and assume, without loss of generality that the perturbed revealed preference that imposes the cheapest taxation³⁷ on the value of e the efficiency index is such that $e_A = \text{argmax}_e p^i \cdot q_A^i \geq p^i \cdot q_A^j > \text{argmax}_e p^i \cdot q_B^i \geq p^i \cdot q_B^j$, that is if and only if $e_A \equiv \frac{p^i \cdot q_A^j}{p^i \cdot q_A^i} > e_B \equiv \frac{p^i \cdot q_B^j}{p^i \cdot q_B^i}$ which given local non satiation we get $p^i \cdot q_A^j > p^i \cdot q_B^j \rightarrow p^i \cdot (q_A^j - q_B^j) > 0$, which in turn implies that $\widehat{\varepsilon}_A^i \cdot \widehat{\varepsilon}_A^i < \widehat{\varepsilon}_B^i \cdot \widehat{\varepsilon}_B^i$, and the results follows from Theorem 1.19.
3. $-MPI(q_A) > -MPI(q_B)$ if and only if there exists a sequence of choices that involves observation q^i and constitutes a cycle, namely $J_i \subseteq \{1, \dots, J\}$ with $i \in J_i$. That is, $\{q^j\}^{j \in J_i} \notin \mathbf{MB}(\{p^j, x^j\})$. Moreover, from the definition of the MPI implies that for at least one observation $k \neq i$ in J_i if and only if, $p^k \cdot (q_A^k - q_A^i) + p^i \cdot (q_A^i - q_A^k) < p^k \cdot (q_B^k - q_B^i) + p^i \cdot (q_B^i - q_B^k)$, which in turn implies that

$$q_A^i \cdot (p^i - p^k) < q_B^i \cdot (p^i - p^k) \Leftrightarrow (q_B^i - q_A^i) \cdot (p^i - p^k) > 0$$

³⁷The argument follows if $q^j R(e) q^i$

which in turn implies that $\widehat{\varepsilon}_A^i \cdot \widehat{\varepsilon}_A^i < \widehat{\varepsilon}_B^i \cdot \widehat{\varepsilon}_B^i$, and the results follows from from Theorem 1.19.

■

1.B Error Process

1.B.1 Algorithms - Estimation of the Error Process in Finite Samples

Algorithm 1.26 (Simulation of F_ε^θ by simulating M sequentially) *Input: Data \mathcal{D} , projection \widehat{m} and θ*

Output: vector $e^2 \in \mathbb{R}^{M \times L}$ where M is the number of repetitions for the simulation.

1. Choose M and set, $m = 1$
2. Find a randomize order I^m and set $i = 1$
3. Take $I^{m(i)}$ and draw $z_{I^{m(i)}}^m \sim U \left[\text{Supp} \left(p^{I^{m(i)}}, x^{I^{m(i)}} \right) | \widehat{m}_{-I^{m(i)}} \right]$
4. Draw e_j^m from F_ε^θ for $j = 1, \dots, J$, and compute $y^m = z_j^m + e^m$
5. If $y^m \in \mathbf{B}(p, x)$, set $q^m = y^m$ and go to 6. Otherwise compute q^m from y^m consistently with equation 1.2
6. If $q^m \in \mathbf{M}$ then set $e_m^2 = 0$, $m = m + 1$ and $i = i + 1$. Otherwise compute $e_m^2 = \min \{q^m - \widehat{m}, q^m - z^m\}$
7. Repeat steps 3-6 M times

Algorithm 1.27 (Computation of σ_ε^2 assuming $\widehat{q} = m$) *Input: choice data q , observed economic environments $\{p^i, x^i\}_{i=1}^J$, projection \widehat{q} and σ^2*

Output: vector $e^1 \in \mathbb{R}^{M \times L}$ where M is the number of repetitions for the simulation.

1. Choose M and set, $m = 1$
2. Draw e_j^m from $N(0, \sigma^2)$ for $j = 1, \dots, J$, and compute $y^m = \hat{q} + e^m$
3. If $y^m \in \mathbf{B}(p, x)$, set $q^m = y^m$ and go to 4. Otherwise compute q^m from y^m consistently with equation 1.2
4. If $x^m \in \mathbf{M}(\mathbf{B}(p, x))$ then set $e_m^1 = 0$ and $m = m + 1$. Otherwise compute $e_m^1 = q^m - \hat{q}$
5. Repeat steps 2-4 M times

Algorithm 1.28 (Computation of σ_ε^2 by simulating $\mathbf{M}(\mathbf{B}(p, x))$) *Input: data on choices q , observed economic environments $\{p^i, x^i\}_{i=1}^J$, projection \hat{q} and σ^2*

Output: vector $e^3 \in \mathbb{R}^{M \times L}$ where M is the number of repetitions for the simulation.

1. Simulate the rational choice set
 - (a) Choose N , and set $n = 1$ $i = 1$, and set $JN = \emptyset$
 - (b) Choose j at random from the set $JJ = \{1, \dots, J\} \setminus JN$ and set $I^{n(i)} = j$
 - (c) Draw $z_{I^{n(i)}}^n \sim U[\text{Supp}(\mathbf{B}(p, x)_{I^{n(i)}}) | \hat{q}_{JJ}]$
 - (d) Set $JN = JN \cup I^{n(i)}$ and repeat steps 1b-1c until $JJ = \emptyset$
 - (e) Set $M_{sim}^n(\mathbf{B}(p, x)) = z^n$
2. Simulate the distribution of the Q statistic
 - (a) Choose M and set, $m = 1$
 - (b) Choose z^m at random from the set $M_{sim}(\mathbf{B}(p, x))$
 - (c) Draw e_j^m from $N(0, \sigma^2)$ for $j = 1, \dots, J$, and compute $y^m = z_j^m + e^m$

- (d) If $y^m \in \mathbf{B}(p, x)$, set $q^m = y^m$ and go to 2e. Otherwise compute q^m from y^m consistently with equation 1.2
- (e) If $q^m \in \mathbf{M}(\mathbf{B}(p, x))$ then set $e_m^3 = 0$ and $m = m + 1$. Otherwise compute $e = q^m - \hat{q}$
- (f) Repeat steps 2b-2e M times

1.B.2 Performance of Algorithms 1.26, 1.27 and 1.28 for Simulated Data

σ	Algorithms - σ_ϵ^2 - DGP 1			Algorithms - σ_ϵ^2 - DGP 2		
	1	2	3	1	2	3
0.5	0.8850	0.9040	0.5370	0.9470	0.9570	0.5905
1	0.7405	0.7618	0.1794	0.8287	0.8816	0.2901
2	0.3128	0.4052	0.3592	0.4016	0.5528	0.2444
3	0.5876	0.6251	0.6460	0.4786	0.5149	0.5313
5	0.8631	0.9218	0.8087	0.8192	0.8918	0.7196
10	0.9765	0.9808	0.8497	0.9775	0.9794	0.8024

Table 1.9: Simulations algorithms. Kolmogorov-Smirnov two tailed test statistics with respect to measure constructed based on the projection for $\sigma = 3$ for two alternative distributions over a CD utility function.

1.C Tables

1.C.1 Data

Table 1.10: Summary statistics. Summary statistics for the proposed measures and other standard measures from the literature applied to the experimental data from (Choi, Fisman, Gale, and Kariv 2007).

Measure	Mean	Median	Std. dev.	Min	Max
Considering all the observed Budget set $N = 81$					
Rational?	0.1975	0.0000	0.4006	0.0000	1
Afriat	0.9568	0.9810	0.0642	0.6860	1.0000
Varian	0.8898	0.9440	0.1568	0.2290	1.0000
HM index	47.1235	48	3.2418	29	50
# violations WARP	4.8889	4.0000	5.3852	0	27
# violations GARP	43.2099	6.0000	104.2718	0	559
Mean ex-ante prob viol	0.1108	0.0984	0.0731	0.0028	0.3775
Raw R^2	0.9906	0.9994	0.0214	0.8597	1.0000
Adjusted R^2	0.1072	0.1023	0.0640	0.0009	0.2806
Weighted R^2	0.9906	0.9994	0.0207	0.8720	1.0000
Size supporting	0.5231	0.5267	0.0538	0.3693	0.6415
PP CI (95%)	0.5141	0.5106	0.0533	0.3835	0.6358
PP KL	0.2458	0.2236	0.0831	0.1432	0.6064
PP Hellinger	0.3229	0.3179	0.0818	0.1751	0.6006
PP TV	0.4102	0.4001	0.0590	0.2977	0.5574

Raw R^2 , adjusted R^2 and weighted R^2 are defined as in 1.29, 1.30 and 1.31 respectively.

PP Hellinger is defined consistent with 1.16 using the Hellinger distance between the

two distributions, $PP_{info-Hell}^0|_{\mathcal{D}} = \frac{1}{\sqrt{2}} \left\| \sqrt{F^0} - \sqrt{\mathbb{P}} \right\|_2$. PP TV uses the Total Variation

metric, $PP_{info-TV}^0|_{\mathcal{D}} = \frac{1}{2} \|F^0 - \mathbb{P}\|_1$

Table 1.11: Summary statistics - half of the sample. Summary statistics for the proposed measures and other standard measures from the literature applied to the experimental data from (Choi, Fisman, Gale, and Kariv 2007).

Measure	Mean	Median	Std. dev.	Min	Max
Considering half of observations per individual $N = 76$					
Rational?	0.6316	1.0000	0.4856	0.0000	1
Mean ex-ante prob viol	0.0979	0.0791	0.0833	0.0028	0.4704
Raw R^2	0.9968	1.0000	0.0177	0.8516	1.0000
Size supporting	0.6315	0.6317	0.1015	0.0000	0.8091
PP CI (95%)	0.4245	0.4285	0.0639	0.2752	0.6515
PP KL	0.1686	0.1640	0.0478	0.0856	0.4139
PP Hellinger	0.2349	0.2237	0.0930	0.0894	0.6910
PP TV	0.3157	0.3179	0.0729	0.1653	0.6196

Raw R^2 , adjusted R^2 and weighted R^2 are defined as in 1.29, 1.30 and 1.31 respectively. PP Hellinger is defined consistent with 1.16 using the Hellinger distance between the two distributions, $PP_{info-Hell}^0|_{\mathcal{D}} = \frac{1}{\sqrt{2}} \left\| \sqrt{F^0} - \sqrt{\mathbb{P}} \right\|_2$. PP TV uses the Total Variation metric, $PP_{info-TV}^0|_{\mathcal{D}} = \frac{1}{2} \|F^0 - \mathbb{P}\|_1$

	Mean - Odd Half	Mean - First Half	Difference	Test-statistic	P-value
R^2	0.9978 (0.0060)	0.9967 (0.0177)	-0.0011 (0.0022)	0.4927	0.6236
Adj R^2	0.1139 (0.1069)	0.0965 (0.0663)	-0.0174 (0.0140)	1.2401	0.2188
PA CI (95%)	0.4202 (0.0542)	0.4255 (0.0639)	0.0053 (0.0064)	0.8371	0.4052

Table 1.12: Comparison - half of observations. Robustness check for the comparison between full sample and first half of the observations. The table shows that results are not significantly different when considering the first 25 observations or the odd observations

Alternative Goodness of Fit Measures (Blundell, Browning, and Crawford 2008)

proposed that when allowing for a stochastic component when conciliating the data with the revealed preference conditions for utility maximization, the measure of the distance between data and restricted estimators of demand provide a natural formulation for a test statistic for the null of rationality. Alternative, one can construct goodness of fit measures as a measure of the distance of data from the models. I propose to construct: (1) a raw goodness of fit, (2) an adjusted goodness of fit and (3) a weighted goodness of fit. The raw goodness of fit is defined as the coefficient of determination constructed from the residuals of the projection. Formally,

Definition 1.29 (Raw R^2) *Let $\hat{\varepsilon}$ be defined as the residuals from the projection exercise then*

$$R_l^2 = 1 - \frac{\|vec(\hat{\varepsilon}) - vec(\bar{\hat{\varepsilon}})\|}{\|vec(q) - vec(\bar{q})\|} \quad (1.8)$$

As other measures of fit proposed in the literature, the empirical fit of the rationality model is tightly related to the ex-ante probability of detecting violations to the model given the observed economic environments and not only noisy, and the former is particularly significant in small samples. Then I propose to adjust the measure of fit the ex ante probability of detecting violations to the model given observed economic environments. These measure is closely related to (Beatty and Crawford 2011), though the considered measure here is the product of the two factors instead of the difference, and uses the mean ex-ante probability instead the overall probability. If I were to consider the overall probability for the considered sample the adjustment is approx. 1 for all subjects. Formally,

Definition 1.30 (Adjusted measure of goodness of fit) *Define the adjusted measure of goodness of fit as*

$$\bar{R}^2 \equiv Adjusted\ Goodness\ of\ fit \equiv Goodness\ of\ fit \times \overline{Pr}[Violation\ model] \quad (1.9)$$

where Goodness of fit is defined as in 1.30 and $\overline{\text{Pr}}[\text{Violation model}]$ is defined as the mean of the ex-ante probability of detecting a violation to the model given observed economic environments.

Finally, the weighted measure of fit is a measure of fit where each of observation is weighted by how unlikely (ex-ante) is to observed such behavior given observed economic environments; that is,

Definition 1.31 (Adjusted measure of goodness of fit - Weighted R^2) *If there exists a j such that $p_j \neq 0$, then the weighted measure of goodness of fit is given by*

$$R_{l,weighted}^2 = 1 - R_l^2 = 1 - \frac{\|\omega(\text{vec}(\hat{\varepsilon}) - \text{vec}(\bar{\varepsilon}))\|}{\|\omega(\text{vec}(q) - \text{vec}(\bar{q}))\|} \quad (1.10)$$

otherwise it is equal to zero.

The weights are given by

Definition 1.32 (Weights) *Consider the weight function ω such that*

$$\tilde{\omega}_j = \begin{cases} p_j & \text{if } \hat{\varepsilon}_j = 0 \\ 1 - p_j & \text{if } \hat{\varepsilon}_j \neq 0 \end{cases} \quad (1.11)$$

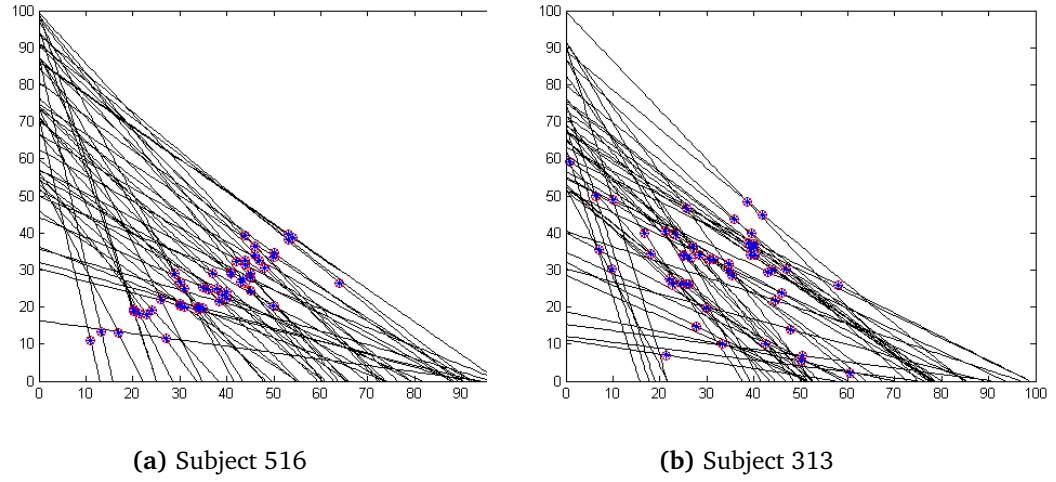
and $\omega_j = \frac{\tilde{\omega}_j}{\sum_j \tilde{\omega}_j}$ with $p_j \equiv \text{Pr}(\text{violation}_{j|-j} | q_1, \dots, q_{j-1}, q_{j+1}, \dots, q_J)$

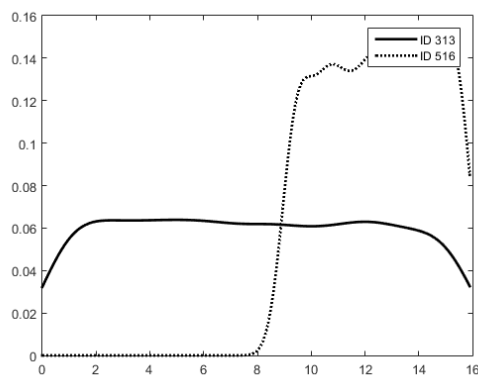
Table 1.13: Correlations. Data for 81 subjects, (Choi, Fisman, Gale, and Kariv 2007).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	
(1) PP CI (95%)	1											
(2) PP Hellinger	0.8789*	1										
(3) PP TV	0.9469*	0.9694*	1									
(4) Raw R^2	0.2227*	0.1212	0.0806	1								
(5) Adjusted R^2	0.0868	0.1235	0.0831	0.1777	1							
(6) Ex ante prob viol	0.1263	0.1928	0.1536	0.0438	0.1674	1						
(7) # viol WARP	-0.3979*	-0.3623*	-0.3021*	-0.7100*	-0.1073	-0.1741	1					
(8) # viol GARP	-0.2469*	-0.1796	0.1236	-0.7321*	-0.0742	0.0038	0.7797*	1				
(9) Afriat	0.2455*	0.1690	0.1125	0.8635*	0.1202	0.0496	-0.8126*	-0.8642*	1			
(10) Varian	0.2440*	0.1696	0.1260	0.8360*	0.1153	0.0428	-0.7984*	-0.8552*	0.9323*	1		
(11) HM	0.3259*	0.2771*	0.2081	0.4642*	0.1576	0.1963*	-0.7825*	-0.7215*	0.6205*	0.5531*	1	
(12) Size supporting	-0.9428*	-0.4293*	-0.8344*	-0.9579*	0.2354*	-0.0148	-0.0254	-0.0538	-0.2092	0.2352*	0.1733	0.1283

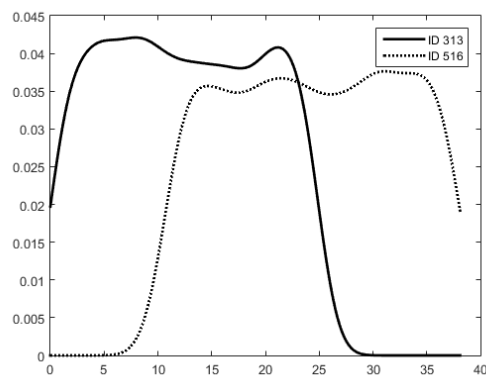
1.C.2 Comparison Between Subjects 313 and 516

Figure 1.15: Comparison between subjects Data from Table 1.7

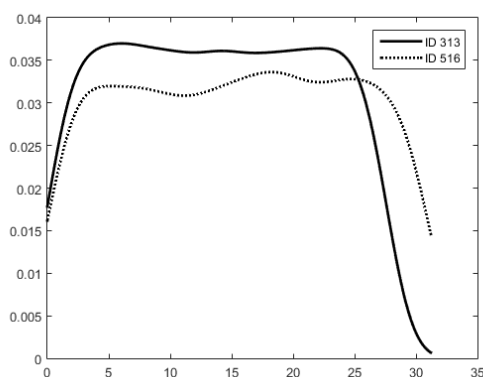




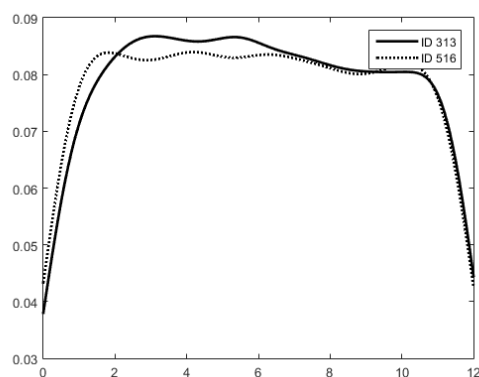
(a) $p^0 = (6.41, 2.05)$ and $x^0 = 102.51$



(b) $p^0 = (2.54, 5.33)$ and $x^0 = 97.13$



(c) $p^0 = (2.65, 0.66)$ and $x^0 = 95.67$



(d) $p^0 = (8.61, 39.38)$ and $x^0 = 103.46$

Figure 1.16: Predictive densities - comparison. Predictive densities for Subjects 313 and 516 where B^0 where randomly generated.

Chapter 2

Satisficing and Stochastic Choice - with Victor Aguiar and Mark Dean

2.1 Introduction

People often do not pay attention to all the available alternatives before making a choice. This fact has led to an extensive recent literature aimed at understanding the observable implications of models in which the decision maker (DM) has limited attention.¹ In an important recent paper, (Manzini and Mariotti 2014) characterize the stochastic choice data generated by a decision maker (DM) who has standard preferences, but only notices each alternative in their choice set with some probability. The chosen item is therefore the best alternative in the ‘consideration set’ of noticed items, which may be a strict subset of the items which are actually available.

The idea that a DM may not search exhaustively through all available alternatives is not new. (Simon 1955) introduced the concept of satisficing: an intuitively plausible choice procedure by which the DM searches through alternatives until they find one that

¹Notable examples include (Masatlioglu, Nakajima, and Ozbay 2012), (Caplin, Dean, and Martin 2011), (Eliasz and Spiegel 2011), and (Salant and Rubinstein 2008).

is ‘good enough’, at which point they stop and choose that alternative.² This model has been hugely influential, both within economics, and in other fields such as psychology ((Schwartz, Ward, Monterosso, Lyubomirsky, White, and Lehman 2002)) and ecology ((Ward 1992)).

Despite the popularity of the satisficing model, testing its predictions can prove challenging. It has long been known that standard choice data, which records only the choices made from different choice sets, cannot be used to disentangle satisficing from utility maximization (see (Caplin, Dean, and Martin 2011) for a discussion). Researchers have therefore typically resorted to richer data sets in order to test the satisficing model. For example, (Caplin, Dean, and Martin 2011) make use of ‘choice process’ data, which records the evolution of choice with decision time, while (Santos, Hortacsu, and Wildenbeest 2012) use the order in which alternatives were searched as recorded from their internet browsing history.

In this chapter, we characterize the observable implications of the satisficing choice procedure for *stochastic* choice data. Such data has been heavily studied in the economics literature.³ We assume that the DM has a fixed utility function and satisficing level. In any given choice set, they search sequentially until they find an alternative which has utility above their satisficing level, at which point they stop and choose that alternative. If they search the entire choice set and do not find a satisficing alternative then they choose the best available option. We assume that search order varies randomly, leading to stochasticity in choice. On the one hand, this chapter is related to the work of (Manzini and Mariotti 2014) (henceforth MM). It specifies a procedure by which attention is allocated, while MM is agnostic in this regard. On the other,

²(Caplin, Dean, and Martin 2011) show that satisficing behavior can be optimal under some circumstances

³See for example (Block and Marschak 1960, Luce and Suppes 1965, Falmagne 1978, Gul and Pesendorfer 2006, Gul, Natenzon, and Pesendorfer 2014, Manzini and Mariotti 2014)

it provides an alternative test of the satisficing model to that of (Caplin, Dean, and Martin 2011) and (Santos, Hortacsu, and Wildenbeest 2012), using a data set which is readily available in many settings.

Our main observation is that the satisficing model implies that choice is stochastic only in choice sets where there are multiple alternatives above the satisficing level. If this is the case, then the order of search will affect the chosen alternative. If not, then either the choice set will be fully searched and the best option deterministically chosen, or the single satisficing alternative will always be chosen. This allows us to behaviorally identify the alternatives that are satisficing for the decision maker.

Without further restriction, any stochastic choice data set can trivially be made commensurate with the satisficing choice procedure by assuming that all alternatives are above the satisficing level, and the resulting distribution of choices reflects the distribution of search orders in that choice set. In order to generate meaningful behavioral implications, we must place further restrictions on the satisficing model. For our main theorem we make the assumption that the distribution of search orders has a full support property (i.e., each item has a positive probability of being searched first), and also rule out the possibility of indifference. This allows us to identify the set of above-reservation alternatives and characterize satisficing with two simple intuitive conditions. The first states that choice can be stochastic only amongst elements that are always chosen (with some probability) when available. The second says that revealed preference, defined via the support of the random choice rule in each set, must satisfy the Strong Axiom of Revealed Preference (SARP). Under these conditions, the data will admit a satisficing representation and the satisficing set, utility function and distribution over search orders can be identified to a high degree of precision.

Our baseline specification puts no restrictions on the relationship between the distribution of search orders across different choice sets. We next consider a refinement of the satisficing model in which the distribution of search order in each choice set is a manifestation of the same underlying search distribution. In order to guarantee such a representation we need a third axiom: the Total Monotonicity condition of (Block and Marschak 1960). This condition on its own is necessary and sufficient for the data to be commensurate with the random utility model (RUM). Thus, the fixed distribution satisficing model is the precise intersection between satisficing and random utility.

We discuss three extensions to our results in which we relax the assumptions of full support, no indifference and the observation of a complete data set. We show that a satisficing model without full support, but with fixed distribution is equivalent to the random utility model. Allowing for indifference (but maintaining the full support assumption) is equivalent to dropping the requirement that stochasticity only take place amongst always chosen alternatives. If data is incomplete, our necessary and sufficient conditions are unchanged, but our ability to identify above satisficing elements is reduced. We finishing studying the extensions to allow for random utility threshold and for random utility. The random utility threshold results in a model that is behaviorally equivalent to the baseline model with full support. We fully characterize the behavioral implications of the extension to allow for random utility and show that it does not provide further information about the utility order among satisficing alternatives.

Section 2.2 describes our set up. Section 2.3 characterizes the satisficing model. Section 2.4 considers the extensions described above, while section 2.5 discusses the related literature.

2.2 Set Up

2.2.1 Data

We consider a finite abstract choice set X , and let $\mathcal{D} \subseteq 2^X \setminus \emptyset$ be the set of menus in which behavior is observed. We assume that data comes in the form of a *random choice rule*, $p : X \times \mathcal{D} \mapsto [0, 1]$, which specifies for each menu $A \in \mathcal{D}$ the probability of choosing each element $a \in A$ (for example, if the DM has a one third probability of choosing x from $\{x, y, z\}$ then $p(x, \{x, y, z\}) = \frac{1}{3}$).

Definition 2.1 (Data set) A data set consists of a set of menus $\mathcal{D} \subseteq 2^X \setminus \emptyset$ and a random choice rule $p : X \times \mathcal{D} \mapsto [0, 1]$ such that $\sum_{a \in A} p(a, A) = 1 \forall A \in \mathcal{D}$. We say a data set is *complete* if $\mathcal{D} = 2^X \setminus \emptyset$.

Random choice rules have been heavily studied in the theoretical, as well as the applied literature.⁴ In practice, while a random choice rule is not directly observable, it can be estimated from observed choice frequencies, pooling either across repeated choices by the same individual, or by aggregating across the choices of different individuals.

2.2.2 The Satisficing Model

The satisficing choice procedure can be described as follows: when faced with a menu of options to choose from, the DM searches through the available alternatives one by one. If, at any point, they come across an alternative which is ‘good enough’, they stop searching and select that alternative from the menu. If they exhaustively search all alternatives without finding an element which satisfies their criteria, then they choose the best available alternative from the set. Note that the standard model of rational choice

⁴Examples of early theoretical work include (Block and Marschak 1960) and (Luce and Suppes 1965). More recent work includes (Gul and Pesendorfer 2006, Manzini and Mariotti 2014, Gul, Natenzon, and Pesendorfer 2014).

is a limiting case of the satisficing model in which no alternative is ‘good enough’.

As a concrete example, consider a DM searching for a book to buy in a bookshop prior to a flight. They examine the available books one by one, looking for one which satisfies their requirements (humorous, has good reviews, long enough to last the flight, not by Dan Brown). If they find such a book, they immediately go to the checkout and buy it. If they search the entire selection and don’t find a book which matches this criteria then they go back and choose the best of the books that they did see.

The satisficing choice procedure therefore has three building blocks. The first is a fixed utility function $u : X \rightarrow \mathbb{R}$, which describes the preferences of the DM. Following (Manzini and Mariotti 2014), for our main results we rule out indifference, and therefore assume that u is injective. We discuss the implications of allowing indifference in section 2.4.2.

The second model element is a utility threshold u^* , which we will refer to as the *reservation utility*. This defines the concept of ‘good enough’: an alternative $x \in X$ is good enough if $u(x) \geq u^*$. We define $U^* = \{a \in X | u(a) \geq u^*\}$ as the set of satisficing elements according to u and u^* . For convenience, we will assume that there is at least one satisficing element: i.e. $u^* \leq \max_{x \in X} u(x)$. This assumption has no behavioral implication: a model in which only the best available alternative is above the reservation utility is indistinguishable from one in which there is no such alternative. However, it will streamline the statement of identification results in section 2.3.

The third element of the satisficing model is the order in which search occurs. A search order for a choice set A is defined by a linear order on that set.⁵ We use R_A to

⁵i.e. a complete, transitive and antisymmetric binary relation on A with the interpretation ‘searched no later than’.

denote the set of linear orders on A , with r_A a typical element in R_A . Our key assumption is that the order of search is determined stochastically: we use $\gamma_A : R_A \rightarrow [0, 1]$ to denote the probability distribution over the set R_A , which we call a ‘stochastic search order’. Abusing notation slightly, we will use $\gamma_A(x \prec_A y)$ to denote the probability of all search orders in which x appears before y : i.e. $\gamma_A(r_A \in R_A | x \prec_A y)$.

We are agnostic about the source of this stochasticity. It could be that the DM randomly decides the order of search - in our example, sometimes the DM search through the books alphabetically, while sometimes they do so by genre. Alternatively, it could be that the random choice rule is generated by a DM who is faced by choice situations which are framed in different ways,⁶ with the framing unobservable to the researcher. For example, sometimes the bookstore puts the thrillers at the front of the store, while sometimes they put the romantic comedies at the front. These ‘frames’ affect the order in which the DM searches (though not their preferences), but are not known to the researcher.

A data set can be represented by the satisficing model if there exists a utility function, satisficing level and family of stochastic search orders which would generate the observed choice probabilities:

Definition 2.2 (General Satisficing Model (GSM)) *A data set (\mathcal{D}, p) has a Generalized Satisficing Model (GSM) representation if there exists an injective $u : X \rightarrow \mathbb{R}$, $u^* \in \mathbb{R}$ such*

⁶In the sense of (Salant and Rubinstein 2008).

Order	(1)	(2)	(3)	(4)	(5)	(6)
1st	a	a	b	b	c	c
2nd	b	c	a	c	a	b
3rd	c	b	c	a	b	a
γ_A	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{24}$	$\frac{1}{8}$	$\frac{1}{4}$

Table 2.1: An example of the satisficing model

that $u^* \leq \max_{x \in X} u(x)$, and $\{\gamma_A\}_{A \in \mathcal{D}}$ such that, for any $A \in \mathcal{D}$ and $a \in A$

$$p(a, A) = \begin{cases} \gamma_A(r_A | a \ r_A \ b \ \forall b \in A \setminus \{a\} \text{ s.t. } u(b) \geq u^*) & \text{if } u(a) \geq u^* \\ 1 & \text{if } a = \arg \max_{x \in A} u(x) < u^* \\ 0 & \text{Otherwise} \end{cases} \quad (2.1)$$

To illustrate how the model works consider the following example.

Example 2.3 Let $A = \{a, b, c\}$ and γ_A be as displayed in table 2.3. Consider first the case where a, b are satisficing alternatives, while c is not, that is $u(a), u(b) > u^* > u(c)$. Then, no matter the search order, c will never be chosen, and so $p(c, A) = 0$. However, as the DM will chose a if a is seen before b and b otherwise, their frequency depends on γ_A . In particular, $p(a, A) = \frac{3}{8}$ and $p(b, A) = \frac{5}{8}$. If instead all alternatives are below the satisficing level (i.e. $u^* > \max_{x \in A} u(x)$) then choice will be independent of search order: all alternatives will always be searched, and the best subsequently chosen.

2.3 Characterizing the Satisficing Model.

2.3.1 A Negative Result

The aim of this chapter is to describe properties of a stochastic choice data set which are necessary and sufficient to guarantee a satisficing representation. However, our first result is negative: without further refinement, the GSM model provides no restriction on such a data set.

Proposition 2.4 *Any data set (\mathcal{D}, p) has a GSM representation.*

All proofs are relegated to the appendix

The GSM is flexible enough to match any data set because it places no restriction on the distribution over search orders in each decision problem. Thus one can always construct a distribution of search orders that will match the data by assuming that all alternatives are above the satisficing level.

To derive testable restrictions for the satisficing model, we introduce a ‘full support’ condition on the distribution of search orders. This restriction will allow us to identify satisficing alternatives as those which are chosen with positive probability in every choice set in which they appear. We can then utilize the underlying structure of the GSM model to derive behavioral restrictions. Intuitively, the stochastic nature of search generates stochastic behavior among satisficing alternatives. In contrast, we expect to observe deterministic utility maximizing behavior among choice sets which consist only of non-satisficing alternatives.

For the remainder of this section we concentrate on the simple case of full support, no indifference and complete data. We identify the behavioral conditions which characterize the resulting model. We also consider a special case of the model in which there

is consistency in the distribution of search orders between choice sets. This additional restriction ensures that our model is behaviorally equivalent to a subset of the class of random utility models (RUMs). In section 2.4 we discuss extensions in which we drop the full support, no indifference, and complete data conditions.

2.3.2 Full Support Satisficing Models

Our model adds to the GSM the assumption that, in each choice set, any item will be searched first with some positive probability.

Assumption 2 (Full Support) For any $a \in A$ and all $A \in \mathcal{D}$: $\gamma_A(r_A \in R_A : a r_A b \quad \forall b \in A \setminus \{a\}) > 0$.

We describe a GSM which additionally satisfies this assumption as a *Full Support Satisficing Model (FSSM)*

Definition 2.5 (Full Support Satisficing Model (FSSM)) A data set (\mathcal{D}, p) has a Full Support Satisficing Model (FSSM) representation if it has a GSM representation in which the stochastic search order satisfies Full Support.

The assumption of Full Support has an important implication: we can identify above-reservation alternatives as those which are always chosen with positive probability in any choice set in which they appear. This is because Full Support implies that, for each such alternative, a search order in which it is searched first occurs with positive probability in each choice set, ensuring that it will be chosen. Furthermore, any alternatives that are *not* above reservation utility will be chosen with zero probability in any choice set which contains an above reservation utility alternative.

We define the set of alternatives which are always chosen:

Definition 2.6 (Always Chosen Set) For any data set (\mathcal{D}, p) , we define the **always chosen** set as $W^* = \{a \in X \mid p(a, A) > 0 \text{ for all } A \in \mathcal{D} \text{ such that } a \in A\}$.

For any complete data set generated by a FSSM, W^* must be equivalent to the set of above-reservation alternatives.

Lemma 2.7 *Assume a complete data set (\mathcal{D}, p) admits an FSSM representation. Then, for any such representation $W^* = U^*$.*

As we discuss in section 2.4.3, if the data set is not complete then W^* may be a strict superset of U^* : a below satisficing alternative may always be chosen because it is only observed in choice sets containing other below satisficing alternatives.⁷

Using the (observable) set W^* , we can define the first of two behavioral conditions which characterize the FSSM. It states that stochastic choice must only occur amongst elements of W^* . This follows from the fact that stochasticity in the satisficing model occurs only from stochasticity in search order.

Axiom 2.8 (Deterministic no satisficing choice) *If $a \in X \setminus W^*$ then either $p(a, A) = 0$ or $p(a, A) = 1$ for all $A \in \mathcal{D}$.*

The second condition ensures that the preference information revealed by a data set is well behaved. In order to state the condition, we introduce the following definitions.

Definition 2.9 (Stochastic Revealed Preference) *Define $C(A) = \{a \in A \mid p(a, A) > 0\}$. a is **stochastically revealed directly preferred** to b if, for some $A \in \mathcal{D}$ $a, b \in A$ and $a \in C(A)$. a is **stochastically revealed preferred** to b if $\{a, b\}$ is in the transitive closure of the stochastically revealed directly preferred relation. a is **stochastically strictly revealed preferred** to b if, for some $A \in \mathcal{D}$, $a \in C(A)$ and $b \notin C(A)$.*

Notice that, for data generated by a FSSM, these revealed preference concepts align with the underlying utility function except in the case of two alternatives above that satisficing level. Such objects will be revealed indifferent to each other, yet may in fact be

⁷Completeness can be replaced for a weaker condition on the richness of the data set which requires observing choices from all two and three element sets.

ranked according to the utility function. It is a defining feature of the satisficing model that utility differences above the threshold u^* are unimportant for behavior. Nevertheless, the FSSM implies that the stochastic revealed preference information must obey the Strong Axiom of Revealed Preference.

Axiom 2.10 (SARP) $C(A)$ must obey SARP: if a is stochastically revealed preferred to b then b must not be stochastically strictly revealed preferred to a .

Our first result is that axioms Axiom 2.8 and Axiom 2.10 are necessary and sufficient for a data set to have a FSSM representation

Theorem 2.11 *The following are equivalent:*

1. A stochastic choice dataset (\mathcal{D}, p) is generated by a FSSM.
2. A stochastic choice dataset (\mathcal{D}, p) satisfies Axiom 2.8 and Axiom 2.10.

To understand the sufficiency of the two axioms -Axiom 2.8 and Axiom 2.10-, note first that SARP allows us, through Afriat/Richter's theorem (Richter 1996), to construct a utility function which represents the stochastic revealed preference relation. Moreover, the elements of W^* will be maximal according to that utility representation, allowing for a u^* such that all elements of the always chosen set can be assigned a utility greater or equal than u^* ; while all the elements that are not always chosen are assigned an utility level below u^* . Axiom 2.8 guarantees deterministic choice in sets which contain at most one above-reservation alternatives, and SARP again ensures that such choices are utility maximizing. For all other choice sets, Axiom 2.8 ensures that alternatives with utility below u^* (and so outside W^*) are not chosen, and a suitable stochastic search order can be constructed from the random choice rules to explain the pattern of choice amongst above satisficing alternatives.

Notice that, while Lemma 2.7 relies on the completeness of the data set, Theorem 2.11 does not. The behavioral content of the FSSM model is the same regardless of whether the data set is complete. However, the degree to which elements of the representation can be identified will be reduced in incomplete data sets, as we discuss in section 2.4.3.

The following examples illustrate the empirical content of the FSSM by presenting data sets which violate each of our axioms.

Example 2.12 (Violation of Axiom 2.8) *Let $X = \{a, b, c\}$, and let $p(a, \{a, b\}) = 1$, $p(b, \{b, c\}) = \frac{1}{2}$, $p(a, \{a, c\}) = 1$ and $p(a, \{a, b, c\}) = 1$. This does not satisfy Axiom 2.8 since $W^* = \{a\}$, but $p(b, \{b, c\}) = \frac{1}{2} \notin \{0, 1\}$. This behavior is incommensurate with the FSSM because the fact that b is chosen probabilistically from $\{b, c\}$ indicates that it must be above the satisficing level, yet this means that it should be chosen some of the time from $\{a, b, c\}$ due to the full support assumption.*

Example 2.13 (Violation of Axiom 2.10) *Let $X = \{a, b, c\}$, and let $p(a, \{a, b\}) = 1$, $p(b, \{b, c\}) = 1$, $p(a, \{a, c\}) = 0$ and $p(a, \{a, b, c\}) = 1$. This does not satisfy Axiom 2.10 since $p(a, \{a, b\}) = 1$ means that a is stochastically strictly revealed preferred to b , $p(b, \{b, c\}) = 1$ means that b is stochastically revealed preferred to c , while $p(c, \{a, c\}) = 1$ means that c is stochastically strictly revealed preferred to a . Such behavior is also incommensurate with the FSSM as, in each case, the uniquely chosen object must have a utility strictly higher than those which are not chosen, either because all are below the satisficing level, in which case the best option is chosen, or because only the chosen object is above the satisficing level.*

Theorem 2.11 shows the extent to which the FSSM can be tested and differentiated from other models. First, note that any data set in which $p(a, A) > 0$ for all $A \in \mathcal{D}$ will trivially satisfy both Axiom 2.8 and Axiom 2.10, and so admit an FSSM representation.

This is because any data set in which the random choice rule has full support in every choice set can be rationalized by an FSSM in which every alternative is above the satisficing level, and the resulting pattern of choice is driven by the choice-set specific distribution over search orders. Second, note that the standard model of utility maximization (without indifference) is a limit case of the FSSM in which $|W^*| = 1$. Third, notice that an alternative interpretation of the FSSM is a model in which attention is complete and choices are governed by a preference relation which has indifference only amongst maximal elements. A model in which such indifference is resolved using a random tie breaking rule with full support amongst maximal elements is equivalent to the FSSM.

Recoverability in the FSSM

In the case of a complete data set which satisfies Axiom 2.8 and Axiom 2.10, many of the elements of the FSSM can be uniquely identified

Theorem 2.14 *Let (\mathcal{D}, p) be a complete data generated by an FSSM $(u, u^*, \{\gamma_A\}_{A \in \mathcal{D}})$. For any FSSM representation of the data $(\bar{u}, \bar{u}^*, \{\bar{\gamma}_A\}_{A \in \mathcal{D}})$*

1. $U^* = \bar{U}^*$
2. For all $a, b \notin U^*$, $u(a) > u(b) \Rightarrow \bar{u}(a) > \bar{u}(b)$
3. $\gamma_A(ar_A b \ \forall b \in \{A \cap U^*\} \setminus \{a\}) = \bar{\gamma}_A(ar_A b \ \forall b \in \{A \cap U^*\} \setminus \{a\})$ for all $A \in \mathcal{D}$,
 $a \in A \cap U^*$

Theorem 2.14 tells us that, in a complete data set we can uniquely identify the above-satisficing elements, the preference ordering over non-satisficing elements, and the probability that one satisficing element will be seen before another in any choice set.

2.3.3 Fixed Distribution Satisficing Models

So far, we have allowed stochastic search order to vary arbitrarily between choice sets: an alternative that is likely to be searched first in choice set A may be very unlikely to be searched first in choice set B . However, in some cases such an assumption may be inappropriate. For example, consider the case in which the probability of search is governed by the ‘saliency’ of different alternatives: a book with a bright pink cover may be more likely to be looked at before one with a dark brown cover regardless of the set of available alternatives.

We now consider the implications of a satisficing model with full support in which the probability distribution over search orders is invariant to the set of available alternatives. We call this the ‘fixed distribution’ property.

Assumption 3 (Fixed Distribution) *There exists a $\Gamma_X : R_X \rightarrow [0, 1]$ such that, for every $A \in \mathcal{D}$ and $r_A \in R_A$*

$$\gamma_A(r_A) = \Gamma_X(r_X | r_A \subset r_X)$$

For every choice set A , it is as if the DM draws a search order from a distribution Γ_X over linear orders on the grand set of alternatives X . They then follow that search order, ignoring any alternatives that are not in fact available in A .

Definition 2.15 (Fixed Distribution Satisficing Model (FDSM)) *A data set (\mathcal{D}, p) has a Fixed Distribution Satisficing Model (FDSM) representation if it has a FSSM representation in which the family of stochastic search orders $\{\gamma_A\}_{A \in \mathcal{D}}$ satisfy Fixed Distribution.*

The conditions Axiom 2.8 and Axiom 2.10 are necessary for FDSM but not sufficient, implying that the FDSM is a strict subcase of FSSM. In order to obtain sufficiency,

we make use of the Total Monotonicity condition of (Block and Marschak 1960). Total Monotonicity by itself is a sufficient and necessary condition for Random Utility Maximization in our environment. This implies that the FDSM is the exact intersection of the FDSM model with the Random Utility model of Block-Marschak and (Falmagne 1978).

In order to define the total monotonicity condition, we first need to define the following function for each $A \in \mathcal{D}$ and $a \in A$:

$$f(a, A) = \sum_{D \in \mathbf{B}(A)} (-1)^{|D \setminus A|} p(a, D).$$

where $\mathbf{B}(A)$ is the class of supersets of A (i.e., $\mathbf{B}(A) \equiv \{D \in \mathcal{D} | A \subseteq D\}$).

(Block and Marschak 1960) and (Falmagne 1978) proved that the following behavioral axiom called Total Monotonicity (or Block-Marschak Monotonicity) is necessary and sufficient for a RUM representation:

Axiom 2.16 (Total Monotonicity) $f(a, B) \geq 0$ for all $a \in X$, for all $B \in \mathcal{D}$.

Observe that f depends only on the data set so this axiom is testable.

Note that Total Monotonicity implies ‘standard’ monotonicity: the probability of choosing any given alternative falls as more alternatives are added to the choice set - that is $p(a, A) \geq p(a, B)$ when $A \subseteq B$.⁸ However, it is also stronger than this condition as we can see in the following example: Set $X = \{a, b, c, d\}$, let $p(a, \{a, b\}) = 0.2$, and $p(a, \{a, b, c\}) = p(a, \{a, b, d\}) = 0.19$ and $p(a, \{a, b, c, d\}) = 0.17$. We check $f(a, \{a, b\}) = p(a, \{a, b\}) + p(a, \{a, b, c, d\}) - [p(a, \{a, b, c\}) + p(a, \{a, b, d\})]$ and observe that $f(a, \{a, b\}) = -0.01$ negative and violating total monotonicity. However, we

⁸To see this take the Mobius inverse representation of $p(a, A) = \sum_{D \in \mathbf{B}(A)} f(a, D)$ and $p(a, B) = \sum_{D' \in \mathbf{B}(B)} f(a, D')$ and note that if $D \in \mathbf{B}(B)$ then $D \in \mathbf{B}(A)$ and $f(a, D) \geq 0$ by total monotonicity, then we have that $p(a, A) \geq p(a, B)$.

can see that standard monotonicity holds in this example.

Clearly, a FDSM cannot lead to a failure of regularity. If it did that would mean that a given satisficing item is more likely to be found first in a bigger menu than in a smaller one, which is not consistent with the idea that the probability of any search order is fixed across menus. The higher order monotonicity conditions implied by total monotonicity can be interpreted as saying that the likelihood of a satisficing item being found first decreases with the size of the menu, but the marginal effect of adding a new option to the menu decreases with its size. In the example above $f(a, \{a, b\}) \geq 0$ means that the impact of adding one additional item c in the menu $\{a, b\}$ on the probability of choosing a – i.e. $p(a, \{a, b\}) - p(a, \{a, b, c\})$ – is bigger than the impact of adding the same item c in the bigger menu $\{a, b, d\}$ on the probability of choosing a – i.e. $p(a, \{a, b, d\}) - p(a, \{a, b, c, d\})$.

We are ready to state the main result of this section.

Theorem 2.17 *The following are equivalent:*

1. *A complete stochastic choice dataset (\mathcal{D}, p) is generated by a FDSM.*
2. *A complete stochastic choice dataset (\mathcal{D}, p) satisfies Axiom 2.8, Axiom 2.10 and Total Monotonicity (Axiom 2.16).*

Note that, unlike Theorem 2.11, Theorem 2.17 requires a complete data set.

Recoverability in the FDSM

In the case of a complete data set which satisfies Axiom 2.8, Axiom 2.10 and Total Monotonicity (Axiom 2.16) several of the elements of the FDSM can be identified. In particular, the identification of the search orderings is improved upon the FSSM recoverability.

Theorem 2.18 *Let (\mathcal{D}, p) be a complete data generated by an FDSM (u, u^*, Γ_X) such that $X \setminus U^* \neq \emptyset$. For any FDSM representation of the data $(\bar{u}, \bar{u}^*, \bar{\Gamma}_X)$*

1. $U^* = \bar{U}^*$
2. For all $a, b \notin U^*$, $u(a) > u(b) \Rightarrow \bar{u}(a) > \bar{u}(b)$
3. $\Gamma_X(ar_X b) = \bar{\Gamma}_X(ar_X b)$ all $A \in D$ and $a, b \in U^*$

Theorem 2.18 tells us that, in a complete data set, we can identify the above-satisficing elements, granted that there is a positive probability of being seen before some above-satisficing element. Moreover, we can identify the preference ordering over the alternatives that are surely non-satisficing elements, and the probability that one revealed satisficing element will be seen before another in any choice set, for those elements that are surely satisficing.

2.3.4 Comparative Statics

In this section we study the comparative statics of the model with respect to the primitives of the model. We first study the behavioral implications of changes on the utility threshold. If we observe two decision makers, I and \tilde{I} , that face the same menus and are such that $u_I^* > u_{\tilde{I}}^*$ and identical otherwise, then we expect to observe that set of always satisficing alternatives for decision maker I is weakly smaller -and a subset- of the set of always satisficing alternatives for decision maker \tilde{I} . Intuitively, lower utility thresholds are associated to a bigger set of satisficing alternatives and, therefore, a bigger set of alternatives that are always chosen when present in a menu. The following claim formalizes this intuition.

Claim 2.19 *Let $I = (u, u^*, \{\gamma_A\}_{A \in \mathcal{D}})$ and $\tilde{I} = (u, \tilde{u}^*, \{\gamma_A\}_{A \in \mathcal{D}})$ be two FSSM (FDSM) decision makers then:*

1. If $u^* > \tilde{u}^*$ then $|W^*| \leq |\tilde{W}^*|$

2. If $|W^*| < |\tilde{W}^*|$ then $u^* > \tilde{u}^*$

To study the effect of the utility threshold on the well being of DMs, we first introduce some definitions.

Definition 2.20 (Utility Distribution) For a FSSM (FDSM) decision maker $I = (u, u^*, \{\gamma_A\}_{A \in \mathcal{D}})$ we define the utility distribution for menu A as the distribution over utility levels implied by the model

$$p_I(\tilde{u}, A) = \sum_{a \in A} \mathbf{1}\{u(a) = \tilde{u}\} p(a, A)$$

and the cumulative distribution function

$$F_I(\bar{u}, A) = \sum_{u \leq \bar{u}} p_I(u, A)$$

Definition 2.21 (Expected Utility) The expected utility for a FSSM (FDSM) decision maker $I = (u, u^*, \{\gamma_A\}_{A \in \mathcal{D}})$ when choosing from menu A is given by

$$E_u^I(A) = \sum_{a \in A} p_I(u(a), A) u(a)$$

We show that a FSSM (FDSM) that has a lower utility threshold would make worse decisions. To understand this, consider two FSSM decision makers $I = \{u, u^*, \{\gamma_A\}_{A \in \mathcal{D}}\}$ and $\tilde{I} = \{u, \tilde{u}^*, \{\gamma_A\}_{A \in \mathcal{D}}\}$ with $u^* > \tilde{u}^*$. Then, for any $A \in \mathcal{D}$, $|A \cap U^*| \leq |A \cap \tilde{U}^*|$. Therefore a DM with a lower threshold is now randomizing including alternatives with lower ex-ante utility level when satisficing alternatives are available menu. The following example illustrates how this mechanism works.

Example 2.22 Let $A = \{a, b, c\}$ and for simplicity assume that $u(a) = 1$, $u(b) = 2$ and $u(c) = 3$. Let $u^* = 2.5$ and $\tilde{u}^* = 1.5$, and consider the search orders from Example 2.3.

For decision maker I , $p(c,A) = 1$ and $p(a,A) = p(b,A) = 0$,

$$F_I(u,A) = \begin{cases} 0 & \text{if } u < 3 \\ 1 & \text{if } u \geq 3 \end{cases}$$

and $E_u^I(A) = 3$. For decision maker \tilde{I} , $p(a,A) = 0$, $p(b,A) = \frac{11}{24}$ and $p(c,A) = \frac{13}{24}$, therefore

$$F_{\tilde{I}}(u,A) = \begin{cases} 0 & \text{if } u < 2 \\ \frac{11}{24} & \text{if } 2 \leq u < 3 \\ 1 & \text{if } u \geq 3 \end{cases}$$

and $E_u^{\tilde{I}}(A) = \frac{61}{24} < 3$.

This result is formalized in Proposition 2.23.

Proposition 2.23 (Utility Threshold) *Let $I = (u, u^*, \{\gamma_A\}_{A \in \mathcal{D}})$ and $\tilde{I} = (u, \tilde{u}^*, \{\gamma_A\}_{A \in \mathcal{D}})$ be two FSSM, with cumulative probability distribution over utility levels $F_I(\bar{u}, A)$ and $F_{\tilde{I}}(\bar{u}, A)$. If $u^* > \tilde{u}^*$ then $F_I(\bar{u}, A) \leq F_{\tilde{I}}(\bar{u}, A)$ for all \bar{u} . Moreover, if $F_I(\bar{u}, A) \leq F_{\tilde{I}}(\bar{u}, A)$ with $F_I(\bar{u}, A) \neq F_{\tilde{I}}(\bar{u}, A)$ then $u^* > \tilde{u}^*$.*

Corollary 2.24 *Under the conditions of Proposition 2.23, $u^* \geq \tilde{u}^*$ then $E_u^I(A) \geq E_u^{\tilde{I}}(A)$ for all $a \in \mathcal{D}$. Moreover, if $E_u^I(A) > E_u^{\tilde{I}}(A)$ then $u^* > \tilde{u}^*$.*

To analyze the overall cost in terms of utility for a decision maker that utilizes a lower utility threshold we note that, from Corollary 2.24, if $u^* \geq \tilde{u}^*$ then

$$\sum_{A \in \mathcal{D}} E_u^I(A) \geq \sum_{A \in \mathcal{D}} E_u^{\tilde{I}}(A)$$

and the last inequality is strict whenever $\{a \in A : \tilde{u}^* \leq u(a) < u^* \text{ for some } A \in \mathcal{D}\} \neq \emptyset$.

Corollary 2.25 *Under the conditions of Proposition 2.23 and assuming complete data sets, if $u^* \geq \tilde{u}^*$ then $|U^*| \leq |\tilde{U}^*|$. Moreover, if $\{a \in A : \tilde{u}^* \leq u(a) < u^* \text{ for some } A \in \mathcal{D}\} \neq \emptyset$ then $|U^*| < |\tilde{U}^*|$*

A direct implication of Corollary 2.25 is that, whenever a satisficing decision maker lowers her utility threshold, the relative size of the set of irrational choices, $\frac{|U^*|}{|X|}$, increases.

Alternatively, we can study the effect of changes in the search orders. Intuitively, if the relative probability of search orders changes in such a way that search orders congruent with preference orders become more likely then the expected utility of their decisions improve. The following example illustrates how this mechanism works.

Example 2.26 *Consider again example 1 with $u(a), u(b) > u^* > u(c)$ and search orders γ_A and $\tilde{\gamma}_A$ as shown in Table 2.2; where $\tilde{\gamma}_A(\{a, b, c\}) = \gamma_A(\{a, b, c\}) + \frac{1}{12}$ and $\tilde{\gamma}_A(\{b, a, c\}) = \gamma_A(\{b, a, c\}) - \frac{1}{12}$. Note that for DM I we have that $p^I(a, A) = \frac{3}{8}$, $p^I(b, A) = \frac{5}{8}$ and $p^I(c, A) = 0$; while for DM \tilde{I} we have that $p^{\tilde{I}}(a, A) = \frac{11}{24}$, $p^{\tilde{I}}(b, A) = \frac{13}{24}$ and $p^{\tilde{I}}(c, A) = 0$. For simplicity assume that $u(a) = 3$ and $u(b) = 2$. Correspondingly, the cumulative distribution functions are given by*

$$F_I(u, A) = \begin{cases} 0 & \text{if } u < 2 \\ \frac{5}{8} & \text{if } 2 \leq u < 3 \\ 1 & \text{if } u \geq 3 \end{cases}$$

with $E_u^I(A) = \frac{19}{8}$.

$$F_{\tilde{I}}(u, A) = \begin{cases} 0 & \text{if } u < 2 \\ \frac{13}{24} & \text{if } 2 \leq u < 3 \\ 1 & \text{if } u \geq 3 \end{cases}$$

Order	(1)	(2)	(3)	(4)	(5)	(6)
1st	a	a	b	b	c	c
2nd	b	c	a	c	a	b
3rd	c	b	c	a	b	a
γ_A	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{24}$	$\frac{1}{8}$	$\frac{1}{4}$
$\tilde{\gamma}_A$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{24}$	$\frac{1}{8}$	$\frac{1}{4}$

Table 2.2: Changes in the probability of search orders and its effect on DM's well-being.

with $E_u^I(A) = \frac{59}{24} > \frac{19}{8}$.

This result is formalized in Appendix 2.A.9.

2.4 Extensions

In this section we extend our model by relaxing in turn the assumptions of full support, no indifference, and complete data

2.4.1 Fixed Distributions Without Full Support

As discussed in section 2.3.1, the GSM model is vacuous without the full support assumption. Here we consider the empirical implication of dropping full support but maintaining the fixed distribution assumption. In such a case the identification of satisficing elements as those that are always chosen breaks down. A satisficing element a may not be chosen in some sets if it is always searched after another satisficing element b . To illustrate this point consider the following example.

Example 2.27 Let $X = \{a, b, c, d\}$, $U^* = \{a, b\}$ and Γ a distribution over search orders on X with full support and where each possible search order is equally likely; moreover assume that $u(c) > u(d)$. Then, for any menu such that $U^* \subseteq A$, $p(a, A) = p(b, A) = \frac{1}{2}$, and if

$A \cap U^* \neq \emptyset$, then $p(U^*, A) = 1$ and $p(c, A) = p(d, A) = 0$. Finally, $p(c, \{c, d\}) = 1$ and $p(d, \{c, d\}) = 0$. Note that this is the standard case describe in section 2.3.3. Now, notice that since we do not assume that Γ needs to have full support on the set of search orders on X , the same data set can be generated by the following fixed distribution satisficing model without full support: $\bar{U}^* = X$, $\Gamma((a, b, c, d)) = \Gamma((b, a, c, d)) = \frac{1}{2}$ and $\Gamma(r_X) = 0$ for all r_X linear order on X , such that $r_X \notin \{(a, b, c, d), (b, a, c, d)\}$. Furthermore, this alternative representation is not unique.

Because it is not possible to identify the satisficing alternatives the only implication of the satisficing model without full support, but with fixed distribution is Total Monotonicity - in other words it is behaviorally indistinguishable from the random utility model. This can be seen by noting that a RUM can be reinterpreted as a satisficing model with fixed distribution by assuming that all alternatives are above the reservation level, and treating the preference orderings from the random utility model as search orders in the satisficing model.

Proposition 2.28 *The following are equivalent:*

1. A complete stochastic choice dataset (\mathcal{D}, p) is generated by a FDSM without Full Support.
2. A complete stochastic choice dataset (\mathcal{D}, p) satisfies Axiom 2.16.

2.4.2 Allowing for Indifference

Here we relax the no indifference assumption while keeping the Full Support conditions. Allowing for indifference potentially introduces stochasticity among non-satisficing alternatives due to the DM's rule to break ties. We assume that tie breaking works as follows, if the DM is indifferent between two or more alternatives, and needs to choose one of them, she chooses at random from the set of indifferent alternatives with prob-

abilities induced by the tie-breaking rule T .

Definition 2.29 (Tie-breaking rule) Let $T : X \times \mathcal{D} \rightarrow \mathbb{R}_{++}$ ⁹ be a function that assigns tie breaking weights to alternatives. In case of indifference between two or more alternatives in menu A , the DM applies the induced tie breaking rule as follows:

$$T(a|A^{\sim a}) = \frac{T(a,A)}{\sum_{b \in A^{\sim a}} T(b,A)} \quad (2.2)$$

where $T(a|A^{\sim a}) > 0$ is the always positive probability that a is chosen when a is indifferent to all the elements in the set $A^{\sim a} \equiv \{b \in A : u(b) = u(a)\}$, and superior to all other elements in A (i.e. $u(a) \geq u(b)$ for all $b \in A$).

Note that if $|A^{\sim a}| = 1$ then $T(a|A^{\sim a}) = 1$, and that $\sum_{b \in A^{\sim a}} T(a|A^{\sim a}) = 1$ in general.

We now extend the Full Support Search Model to allow for indifferences.

Definition 2.30 (Full Support Search Model with Indifferences (FSSMI)) A data set (\mathcal{D}, p) has a Full Support Search Model with Indifferences (FSSMI) representation if there exists $u : X \rightarrow \mathbb{R}$, $u^* \in \mathbb{R}$ such that $u^* \leq \max_{a \in X} u(a)$, stochastic search orders $\{\gamma_A\}_{A \in \mathcal{D}}$ that satisfies Fixed Distribution and tie breaking rule $T : X \times \mathcal{D} \rightarrow \mathbb{R}_{++}$, such that, for any $a \in A$

$$p(a,A) = \begin{cases} \gamma_A(ar_A b \forall b \text{ s.t. } u(b) > u^*) & \text{if } u(a) \geq u^* \\ T(a|A^{\sim a}) & \text{if } a \in \arg \max_{x \in A} u(x) < u^* \\ 0 & \text{Otherwise} \end{cases} \quad (2.3)$$

The following example illustrates the FSSMI.

⁹Note that we rule out deterministic tie breaking since, the behavior of a DM that is indifferent between two alternatives a and b and always chooses a over b is behaviorally indistinguishable from a DM that prefers a over b .

Example 2.31 Consider again example 2.3, and assume that DM's choices can be represented by a Full Support Seach Model with Indifferences. Let $U^* = \{a\}$, and let $u(b) = u(c)$. Then, $p(a, A) = 1$, $p(b, A) = p(c, A) = 0$ for all A such that $a \in A$. Let the tie breaking rule be generated by $T(b, \{b, c\}) = 1$ and $T(c, \{b, c\}) = 2$, then $p(b, \{b, c\}) = \frac{1}{3}$, and $p(c, \{b, c\}) = \frac{2}{3}$

Axiom 2.8 is no longer necessary for the FSSMI model: stochasticity can occur amongst alternatives that are not always chosen due to indifference. In fact, it turns out that the behavioral implication of allowing for indifference is precisely the removal of this axiom from our set of necessary and sufficient conditions.

Theorem 2.32 *The following are equivalent:*

1. A complete stochastic choice dataset (\mathcal{D}, p) is generated by a FSSMI.
2. A complete stochastic choice dataset (\mathcal{D}, p) satisfies Axiom 2.10.

Theorem 2.32 highlights that the satisficing model *without* indifference can be reinterpreted as a standard optimizing model with random tie breaking, but allowing for indifference only amongst the best alternatives.

Recoverability in the FSSMI

Given $u^* \leq \max_{a \in X} u(a)$, the extension of the model to allow for ties, does not obscure the identification result for the satisficing set as in 2.6, where the always chosen set coincides with the satisficing set, i.e. $W^* = U^*$. The following theorem describes the degree to which the other elements of the model cannot be identified.

Theorem 2.33 *Let (\mathcal{D}, p) be a complete data generated by an FSSMI $(u, u^*, \{\gamma_A\}_{A \in \mathcal{D}}, T)$. For any FDSMI representation of the data $(\bar{u}, \bar{u}^*, \{\bar{\gamma}_A\}_{A \in \mathcal{D}}, \bar{T})$*

1. $U^* = \bar{U}^*$

2. For all $a, b \notin U^*$, $u(a) \geq u(b) \Rightarrow \bar{u}(a) \geq \bar{u}(b)$
3. $\gamma_A(ar_A b \ \forall b \in \{A \cap U^*\} \setminus \{a\}) = \bar{\gamma}_A(ar_A b \ \forall b \in \{A \cap U^*\} \setminus \{a\})$ for all $A \in D$,
 $a \in A \cap U^*$
4. $T(a|A^a) = \bar{T}(a|A^a)$ for all $a \in A$, $A \in \mathcal{D}$.

Theorem 2.33 tells us that in a complete data set we can uniquely identify the above-satisficing elements, the preference ordering among non-satisficing elements, the tie breaking rule when used and the probability that one satisficing element is seen before another in any choice set. Its proof follows from Theorem 2.14 when the identification of the tie breaking rule is established. The identification of the former (tie breaker rule) holds because whenever it is used, it is calibrated from the empirical choice probability among elements that are not revealed to be satisficing.

2.4.3 Incomplete Datasets

Here we relax the complete data set assumption while keeping the full support distribution condition and assuming no indifference. We do not work with the fixed distribution assumption since Total Monotonicity is not well defined for incomplete data sets, and the literature on Random Utility Models has not dealt with this extension.

Notice that complete data is not a necessary assumption for Theorem 2.11. Thus, if we drop completeness, the implications of the model are not affected, but identification becomes weaker. To see this note W^* may be a strict superset of U^* since a below-satisficing alternative may be always chosen because it is only observed in choice sets containing below-satisficing alternatives. The accuracy with which we can identify the primitives of the model given observed data depends on the richness of the data set.

Theorem 2.34 *Let (\mathcal{D}, p) be a data set (that needs not to be complete) that satisfies Axiom 2.8 and Axiom 2.10. Then for a FSSM $(u, u^*, \{\gamma_A\}_{A \in \mathcal{D}})$ represents the data if and only if*

1. $\widetilde{W} \subseteq U^* \subseteq W^*$
2. *u must represent the stochastic revealed preference relation on $X \setminus U^*$: that is if a is stochastically strictly revealed preferred to b then $u(a) > u(b)$, and if a is revealed preferred to b then $u(a) \geq u(b)$ for any $a, b \in X \setminus U^*$*
3. $\gamma_A(a \succ_A b \mid \forall b \in (U^* \cap A) \setminus \{a\}) = p(a, A)$ for all $A \in \mathcal{D}$, $a \in U^* \cap A$.

where $\widetilde{W} \equiv \{a \in X \mid \exists A \in \mathcal{D}, \text{ s.t } p(a, A) \in (0, 1)\}$.

Theorem 2.34 tells us that, when dealing with incomplete data sets one can only identify with certainty satisficing choices, if these have been observed chosen when other satisficing alternatives were available as well; that is, if we see them being chosen stochastically. As we established before, the satisficing set is a subset of the set of always chosen alternatives, but with incomplete data it may be a strict subset. Furthermore, one can only partially recover the preference order for the revealed not satisficing alternatives. Finally, the search orders can only be identify up to the set of those that coincides with the one generated from the relative probabilities of the elements that surely are in U^* .

2.4.4 Random Utility and Random Threshold

Now we turn to the study of two variants of the satisficing model where we allow randomness in tastes and in the threshold rule. The first generalizes the constant threshold assumption and allows it to be random but common across all items and menus. The second variant focuses on letting utility taken as fixed before to be also random (i.e., random utility).

Random Threshold

Here we explore the satisficing model when we allow for a random threshold. We establish the somewhat surprising result that assuming a random threshold that is independent of the search process and common across items and menus, adds no generality to the satisficing model. In fact, these variants are indistinguishable from the FSSM/FDSM with constant threshold.

The probability of any item of being satisficing is given by the probability of the common random threshold (CRT) variable of being below this number. It is common because it remains the same across menus and items. We assume throughout the remaining of the section that the random threshold is independent of the random search.

Definition 2.35 (FSSM with Common Random Threshold (FSSM-CRT)) *A data set (\mathcal{D}, p) has a Full Support Satisficing Model with Common Random Threshold (FSSM-CRT) if there exists an injective $u : X \rightarrow \mathbb{R}$, a continuous random variable $u^* \sim F_{u^*}(\cdot)$ such that $\tau(a) = Pr(u(a) > u^*) = 1 - F_{u^*}(u(a))$ with $support(u^*) \subseteq \mathbb{R}$ such that $|U^{*,CRT}| \geq 2$ with $U^{*,CRT} = \{x \in X : \tau(x) > 0\}$, and $\{\gamma_A\}_{A \in \mathcal{D}}$ with the Full Support property, such that, for any $A \in \mathcal{D}$ and $a \in A$:*

$$p(a, A) = \sum_{r_A \in R_A} \tau(a) \prod_{b \in A: br_A a} (1 - \tau(b)) \gamma_A(r_A) + \prod_{c \in A} (1 - \tau(c)) \mathbf{1}(u(a) > u(b) \forall b \in A \setminus \{a\}). \quad (2.4)$$

Notice that the term $\tau(a) \prod_{b \in A: br_A a} (1 - \tau(b))$ measures the probability of item $a \in A$ being satisficing given that all items searched before it in menu A under fixed search r_A where not satisficing, similarly $\prod_{c \in A} (1 - \tau(c))$ is the probability of not finding anything satisficing in A ,¹⁰ and $\mathbf{1}(u(a) > u(b) \forall b \in A) = 1$ when a is maximal under u

¹⁰Notice that the summation of the first term of the expression over all items $P(A) = \sum_{c \in A} \sum_{r_A \in R_A} \tau(c) \prod_{b \in A: br_A c} (1 - \tau(b)) \gamma_A(r_A) = 1 - \prod_{c \in A} (1 - \tau(c))$, thus the $1 - P(A) = \prod_{c \in A} (1 - \tau(c))$ is the residual probability.

in A and zero otherwise. We assume also that the productory over the empty set is 1, $\prod_{\emptyset}(1 - \tau(b)) = 1$.

We first study the implications of CRT assumption with the Full Support assumption on the search process.

Lemma 2.36 *A complete stochastic choice dataset (\mathcal{D}, p) that can be generated by a FSSM-CRT satisfies Deterministic no satisficing choice (Axiom 2.8) and SARP (Axiom 2.10).*

The previous lemma allows us to establish the following equivalence result.

Theorem 2.37 *The following statements are equivalent:*

1. *A complete stochastic choice dataset (\mathcal{D}, p) can be generated by a FSSM.*
2. *A complete stochastic choice dataset (\mathcal{D}, p) can be generated by a FSSM-CRT.*

Now we explore the CRT assumption combined with the fixed distribution assumption over the random search.

Definition 2.38 (FDSM with Common Random Threshold (FDSM-CRT)) *A data set (\mathcal{D}, p) has a Fixed Distribution Satisficing Model with Common Random Threshold (FDSM-CRT) representation if it has a FSSM-CRT representation in which the family of stochastic search orders $\{\gamma_A\}_{A \in \mathcal{D}}$ satisfy the Fixed Distribution property.*

In what follows, we will prove that the FDSM-CRT is behaviorally indistinguishable from the FDSM. But first we need two technical lemmata.

Lemma 2.39 *A weighed sum of totally monotonic mappings $\hat{p}_i : X \times \mathcal{D} \mapsto [0, 1]$ for $i \in \{1, \dots, I\}$ with $I \geq 1$ an integer, evaluated at some (a, A) such that $a \in A$ and $A \in \mathcal{D}$, $\hat{P}(a, A, \omega) = \sum_{i=1}^I \omega_i \hat{p}_i(a, A)$ with $\omega_i \geq 0$ a non-negative weight also satisfies total monotonicity.*

Lemma 2.40 *The mapping $\hat{P} : X \times \mathcal{D} \mapsto [0, 1]$ defined by $(a, A) \mapsto T(A)\mathbf{1}(u(a) > u(b) \forall b \in A \setminus \{a\})$ for a fixed $a \in A$ is totally monotone, where $T(A) = \prod_{c \in A} (1 - \tau(c))$.*

With this in hand we are ready to state and prove the empirical implications of FDSM-CRT.

Lemma 2.41 *A complete stochastic choice dataset (\mathcal{D}, p) that can be generated by a FDSM-CRT satisfies Deterministic no satisficing choice (Axiom 2.8), SARP (Axiom 2.10) and Total Monotonicity (Axiom 2.16).*

The previous lemma allows us to establish the empirical equivalence between FDSM with and without CRT.

Theorem 2.42 *The following statements are equivalent:*

1. *A complete stochastic choice dataset (\mathcal{D}, p) can be generated by a FDSM.*
2. *A complete stochastic choice dataset (\mathcal{D}, p) can be generated by a FDSM-CRT.*

We conclude that adding the CRT does not add generality to the FSSM and FDSM as it does not change the empirical implications of these satisficing models.

Random Utility FDSM

Here we focus on a variant of the FDSM where we also allow for random utility.¹¹ We consider a model where there is a probability measure $\rho : \mathcal{U} \mapsto [0, 1]$ defined over a domain of injective utilities defined on X denoted by \mathcal{U} . For a given element in the support $u \in \mathcal{U}$, we assume that the DM is consistent with FDSM or p_u^{FDSM} with $\{u, u^*, \{\gamma_A\}_{A \in \mathcal{D}}\}$ (i.e. u^* , and $\{\gamma_A\}_{A \in \mathcal{D}}$ common for each $u \in \mathcal{U}$):

¹¹We omit the variant of FSSM with random utility as it is too non-restrictive, in fact, the reader can easily check using the results in this section that the necessary and sufficient condition that characterizes FSSM with random utility is the Degeneracy condition (A 2.44).

Definition 2.43 (Random Utility FDSM representation (RU-FDSM)) A data set (\mathcal{D}, p) has a Random Utility FDSM representation (RU-FDSM) if it is the average of the FDSM models over all possible utilities in (ρ, \mathcal{U}) :

$$p(a, A) = \sum_{u \in \mathcal{U}} \rho(u) p_u^{FDSM}(a, A)$$

We note that the new model does not satisfy SARP (Axiom 2.10) nor Deterministic non satisficing choice (Axiom 2.8). It satisfies however:

Axiom 2.44 (Degeneracy) For $x \in X \setminus W^*$ and $A \cap W^* \neq \emptyset$ then $p(x, A) = 0$.

Notice also the following direct results for RU-FDSM:

Claim 2.45 If the data set (\mathcal{D}, p) has a RU-FDSM representation, then $W^* = \{a \in X : p(a, A) > 0, \forall A \in \mathcal{D}\} \equiv \{a \in X : \exists u \in \mathcal{U}, \rho(u) > 0, u(a) > u^*\}$ corresponds to the set of “sometimes” satisficing items. Also $X \setminus W^*$ corresponds to the set of never satisficing items.

Now we considering the following property, where any item that is satisficing for some fixed utility is satisficing for all utilities in the domain of the random utility (ρ, \mathcal{U}) . This property is closer to the FDSM as it restricts the taste variation for satisficing items.

Definition 2.46 (Always satisficing random utility distribution) A random utility distribution is always satisficing if for any item $x \in X$ such that for some $u \in \mathcal{U}$ with $\rho(u) > 0$, and $u(x) > u^*$ it follows that for any other $\hat{u} \in \mathcal{U}$ with $\rho(\hat{u}) > 0$ it also holds that $\hat{u}(x) > u^*$.

It is clear that this property allows for a sharper interpretation of W^* . In fact, consider the following claim.

Claim 2.47 If complete stochastic choice dataset (\mathcal{D}, p) can be generated by a RU-FDSM with Always satisficing random utility distribution, then if $a \in W^*$ it follows that $Pr(u : u(a) > u^*) = 1$ and $Pr(u : u(b) > u^*) = 0$ for $b \in X \setminus W^*$. Thus $W^* = \{a \in X : u(a) > u^* \forall u \in \mathcal{U}, \rho(u) > 0\}$ is the set of always satisficing items.

We are ready to characterize the RU-FDSM, and we notice already that it is observationally equivalent to the RU-FDSM with Always satisficing random utility distribution.

Theorem 2.48 *The following are equivalent:*

1. *A complete stochastic choice dataset (\mathcal{D}, p) can be generated by a RU-FDSM.*
2. *A complete stochastic choice dataset (\mathcal{D}, p) satisfies Axiom 2.16 and Degeneracy (Axiom 2.44).*
3. *A complete stochastic choice dataset (\mathcal{D}, p) can be generated by a RU-FDSM with Always satisficing random utility distribution.*

Identification of the Utility of Satisficing Items with Random Utility and Random Threshold.

After studying the FSSM and the FDSM a less satisfactory feature of the models is that we cannot recover information about the utility of satisficing elements. It is an interesting question to explore whether this characteristic of the FSSM/FDSM is inherent to the general satisficing procedures or not. In this section we use the partial exploration of random threshold/random utility variants of the satisficing model to try to contribute to answer this question. In general, we find that inferring information about the intensity of the utility of satisficing items is still problematic.

For the case of a common random threshold Theorem 2.42 says that in a complete standard stochastic choice dataset (\mathcal{D}, p) FDSM and FDSM-CRT are indistinguishable from one another. In fact, this gives as a corollary that the random threshold distribution $u^* \sim F_{u^*}$ is not identified. We can only separate the elements that have a positive probability of being over the threshold from those that have a zero probability of surpassing it (this follows from the fact that $W^* \equiv U^*$ in the case of FDSM and $W^* \equiv \{a \in X : \tau(a) > 0\}$ in the FDSM-CRT). This partially identifies the support of F_{u^*}

(up to monotone transformations of the random variable) but not the actual distribution. The following corollary follows trivially from Theorem 2.42.

Corollary 2.49 *If a complete stochastic choice dataset (\mathcal{D}, p) can be generated by a FDSM-CRT, then the utility of satisficing elements such that $U^{*,CRT} = \{a \in X : \tau(a) > 0\}$ is not identified.*

From Theorem 2.42 it is also clear that only restricting the search procedure captured by $\{\gamma_A\}_{A \in \mathcal{D}}$ beyond the fixed distribution assumption we can hope to obtain more information about the random variable u^* and about the utility level of satisficing items. In fact this seems to be true for a uniform search restriction (and the implications of such model are left as an open question).

The RU-FDSM clearly contains the FDSM-CRT/FDSM and is not contained by the latter by Theorem 2.48. In consequence, we expect no gains in allowing random utility to identify utility of satisficing items. However, RU-FDSM is interesting because one could argue that the fact that some items are falling below the fixed threshold u^* randomly could help to identify some form of average utility like the quantity $Pr(u : u(a) > u^*) = \sum_{u \in \mathcal{U}} \rho(u) \mathbf{1}(u(a) \geq u^*)$. However, we have the following corollary of Theorem 2.48 that establishes that $Pr(u : u(a) > u^*)$ is not identified.

Corollary 2.50 *If a complete stochastic choice dataset (\mathcal{D}, p) can be generated by a RU-FDSM (i.e., a list $(\rho, \mathcal{U}, p_u^{FDSM})$ with $Pr(u \in \mathcal{U} : u(a) > u^*) = \sum_{u \in \mathcal{U}} \rho(u) \mathbf{1}(u(a) \geq u^*) > 0$ for $a \in W^*$ and zero otherwise, there is an alternative representation RU-FDSM with Always satisficing property of the dataset (\mathcal{D}, p) (i.e., a list $(\rho^*, \mathcal{U}^*, p_u^{*,FDSM})$) such that $Pr(u \in \mathcal{U}^* : u(a) > u^*) = 1$ if $a \in W^*$ and zero otherwise.*

Recall, that FDSM allows us to recover $\mathbf{1}(u(a) > u^*)$ only, which implies the lack of identification of utility (or utility intensity for satisficing items). Thus the RU-FDSM fares no better than the FDSM in identifying satisficing items utility levels, because

under the Always satisficing property $Pr(u \in \mathcal{U}^* : u(a) > u^*)$ is numerically equivalent to $\mathbf{1}(u(a) > u^*)$, because if $a \in W^*$ then $\mathbf{1}(u(a) > u^*) = 1$ and zero otherwise.

2.5 Relation to Existing Literature

The paper closest in spirit to ours is (Manzini and Mariotti 2014), which characterizes the random choice generated by a DM who makes choices by optimizing on a stochastically generated consideration set. As in our model, preferences are deterministic, with randomness in choice coming from stochastic changes in attention. However, the behavioral implications of the two models are quite different, with the satisficing model being the more general. In the set up of MM, all alternatives are always chosen with positive probability in each set. In such a data set, axioms Axiom 2.8 and Axiom 2.10 are always satisfied, and so the FSSM is trivially more general than the stochastic consideration set model. Moreover, the FDSM also nests the stochastic consideration set model. This follows from the fact that, when restricted to the class of data in which all alternatives are chosen with positive probability, the FDSM model is equivalent to the class of all RUMs, and the model of MM is a strict subset of this class. Moreover, the FSSM and FDSM can accommodate data sets in which not all alternatives are chosen with positive probability.

Our work also contributes to the literature aimed at testing the satisficing model. It is well known that standard deterministic choice data cannot be used to distinguish rational choice from satisficing behavior, implying that richer data is needed. (Caplin, Dean, and Martin 2011, Caplin and Dean 2011) showed how to test the satisficing model using ‘choice process’ data, which records not just final choice made by a decision maker, but also how choices change with contemplation time. (Santos, Hortacsu, and Wildenbeest 2012) utilize data in which the sequence of search is recorded to test the satisficing model. This chapter describes the implication of the satisficing model for

stochastic choice data, which is arguably easier to collect than either choice process or search data.

Another relevant paper is (Rubinstein and Salant 2006) that studies the implications of choices from lists that include as a special case a variant of the satisficing model where the DM chooses the first element in a list (that could be put one to one with a linear search/ordering) that is above a threshold, if none she chooses the last one. We differ from this effort in that we assume we do not observe the lists or linear orderings, instead we infer the distribution of the linear search from the frequency of choice.

Ours is not the first paper to characterize the behavior of random choice rules. Much of the previous work has focused on random utility models (RUMs), in which the DM chooses in order to maximize a utility function, drawn from some distribution (see for example (Block and Marschak 1960, Falmagne 1978, Gul, Natenzon, and Pesendorfer 2014)). As discussed above, the FSSM is behaviorally distinct from the class of RUMs. It is easy to construct examples of FSSMs which violate regularity, and so cannot be modeled as the resulting from random utility maximization. Moreover, RUMs are not guaranteed to satisfy either axioms Axiom 2.8 or Axiom 2.10. In contrast the FDSM is behaviorally a subset of the class of RUMs. Total monotonicity Axiom 2.16 is necessary and sufficient for a RUM representation (Falmagne 1978).

2.A Proofs

2.A.1 Proof of Proposition 2.4

Proof. For any data set (\mathcal{D}, p) , set $U : X \rightarrow [0, 1]$ to be any arbitrary one to one real valued function and set $u^* = -2$; then $U^* = X$. For any menu $A \in \mathcal{D}$, let r_A^a be the set of linear orders on A such that $a r b$ for all $b \in A$. $\{r_A^a\}_{a \in A}$ therefore defines partition on R_A . Define

$$\gamma_A(r_A) = \frac{p(a, A)}{|r_A^a|} \text{ where } r_A \in r_A^a$$

Such a representation will generate p as, for any a , $u(a) \geq u^*$

$$\gamma_A(r_A | a r_A^a b \forall b \text{ s.t. } u(b) \geq u^*) = \gamma_A(r_A \in r_A^a) = p(a, A)$$

■

2.A.2 Proof of Lemma 2.7

Proof. ($W^* \subseteq U^*$) Let $a \in W^*$ then for any given $A \in \mathcal{D}$ we have $p(a, A) > 0$ then either: (i) $a \in U^*$ or (exclusive) (ii) $u(a) > u(b)$ for all $b \in A$ and $U^* \cap A = \emptyset$. Assume $a \notin U^*$ then (since $U^* \neq \emptyset$) there exists $b \in U^*$ and, by completeness of the data, there is a menu $A' \in \mathcal{D}$ such that $a, b \in A'$, and therefore $p(a, A') = 0$, so we have a contradiction.

($U^* \subseteq W^*$) Let $a \in U^* \Rightarrow u(a) > u^*$, by FSSM

$$p(a, A) = \gamma_A(r_A | a r_A b \forall b \in A \text{ s.t. } u(b) > u^*) > 0$$

where the last inequality follows from Full Support assumption. ■

2.A.3 Proof of Theorem 2.11

Proof. First we prove that (1) implies (2). To prove that a data set (\mathcal{D}, p) that admits a FSSM representation satisfies Axiom 2.8 first notice that $U^* \subseteq W^*$ even for incomplete data sets. To see this note that if $a \in U^*$ then $u(a) \geq u^*$ and since the data has a FSSM representation then $p(a, A) = \gamma_A(r_A | a r_A b \forall b \in A \cap U^* \setminus \{a\})$ for all $A \in \mathcal{D}$. Given the full support assumption, $\gamma_A(r_A | a r_A b \forall b \in A \cap U^* \setminus \{a\}) > 0$ for all $A \in \mathcal{D}$, therefore $p(a, A) > 0$ for all $A \in \mathcal{D}$ which in turn implies that $a \in W^*$.

Then if $a \notin W^*$ we have that $a \notin U^*$, which in turn implies that $u(a) < u^*$. Since u is injective, there exists a $a_A^* = \operatorname{argmax}_{a \in A} u(a)$. Then either (i) $a = \operatorname{argmax}_{b \in A} u(b)$ or (ii) $u(a) < \max_{b \in A} u(b)$. If (i) since the data has a FSSM representation $p(a, A) = 1$; while if (ii) $p(a, A) = 0$. In either case Axiom 2.8 follows.

To show that Axiom 2.10 holds, assume, by the way of contradiction, that (i) a is stochastically revealed preferred to b and (ii) b is stochastically strictly revealed preferred to a . From (ii), given that data admits a FSSM we must have (by the full support assumption) that $u(a) < u^*$ and $u(a) < u(b)$. If a is stochastically revealed preferred to b then there must exist a sequence of alternatives c_1, \dots, c_N and choice sets A_1, \dots, A_{N-1} , such that $c_1 = a$, $c_N = b$, $c_n, c_{n+1} \in A_n$ and $c_n \in C(A_n)$. If $u(b) > u^*$ then it must be the case that $u(c_{N-1}) > u^*$ (otherwise c_{N-1} could not be chosen with positive probability when c_N was available). Iterating on this argument implies that $u(a) > u^*$. If $u(b) < u^*$ then this implies that $u(c_{N-1}) > u(b)$. If $u(c_n) < u^*$ for all n then iterating on this argument implies that $u(a) > u(b)$. Otherwise, the previous argument implies that $u(a) > u^*$. Either provides a contradiction.

Now we prove that (2) implies (1). For $A \subseteq X \setminus W^*$, $C(A) \equiv \{a \in A | p(a, A) = 1\}$ from Axiom 2.8. Given Axiom 2.10, we can generate an injective utility function using

Afriat/Richter's theorem, such that $u : X \setminus W^* \rightarrow [0, \bar{u}]$ and $C(A) = \operatorname{argmax}_{a \in A} u(a)$ for all $A \in \mathcal{D}$ such that $A \cap W^* = \emptyset$. Fix $u^* > \bar{u}$, enumerate the elements in W^* as a_1, \dots, a_N and let $u(a_n) = u^* + n$ for all $a_n \in W^*$. Thus $u : X \rightarrow \mathbb{R}$ is injective and $U^* \equiv W^*$. Furthermore, notice that Axiom 2.10 implies that W^* is non-empty, so that $u^* \leq \max_{x \in X} u(x)$.

For every A such that $A \cap W^*$ is non-empty, let $A^* \equiv A \cap W^*$ and define R_{A^*} the set of all linear orders on A^* , and let $R_A(r_{A^*})$ be the set of all linear orders $r_A \in R_A$ that induce the linear order $r_{A^*} \in R_{A^*}$. Then, set the probability of the set of linear orders that generate each r_{A^*} as

$$\gamma_A(r_A \in R_A(r_{A^*}) \mid a_i r_{A^*} a_j \ \forall a_j \in A^* \setminus \{a_i\}) = \frac{p(a_i, A)}{|\{r_{A^*} \in R_{A^*} : a_i r_{A^*} a_j \ \forall a_j \in A^* \setminus \{a_i\}\}|}$$

for all $a_i \in A^*$. Finally, distribute the probability mass above uniformly across the elements in $R_A(r_{A^*})$. For any $r_A \in R_A(r_{A^*})$, for a given r_{A^*} :

$$\gamma_A(r_A) = \gamma_A(r_A \in R_A \mid a_i r_{A^*} a_j \ \forall a_j \in A^* \setminus \{a_i\}) / |R_A(r_{A^*})|$$

Where $|\cdot|$ stands for the cardinality map. Note that, as $r_A \in R_A(r_{A^*}) \mid a_i r_{A^*} a_j \ \forall a_j \in A^* \setminus \{a_i\}$ for some $a_i \in A^*$, and as $p(a_i, A) > 0$ by construction of W^* this distribution will have full support on R_A .

If $A \cap W^* = \emptyset$ then for all $r_A \in R_A$ define

$$\gamma(r_A) = \frac{1}{|R_A|}$$

Thus, between them u , u^* and $\{\gamma_A\}_{A \in \mathcal{D}}$ satisfy the requirements of an FSSM. To verify that we can generate (\mathcal{D}, p) notice that if we face a menu A we have the following cases:

(i) if $A \cap W^* = A$ then $u(a) > u^*$ for all $a \in A$ and, for each a , $p(a, A) = \gamma_A(r_A \in R_A | a \succ b \quad \forall b \in A \setminus \{a\})$, to see that this is true observe that $\gamma_A(r_A \in R_A | a \succ b \quad \forall b \in A \setminus \{a\}) = \sum_{r_A \in R_A} \mathbf{1}[r_A : r_{A^*} \text{ such that } a_i r_{A^*} a_j \quad \forall a_j \in A^*] \gamma_A(r_A) = p(a, A)$.

(ii) If $A \cap W^* = \emptyset$ then $p(a, A) = 1$ if $u(a) > u(b)$ for all $b \in A \setminus \{a\}$ and zero otherwise. This follows directly from the fact that u was constructed to represent choice on such sets.

(iii) If $A \cap W^* \subset A$ and $A \cap W^* \neq \emptyset$ then we have $p(a, A) = \gamma_A(r_A | a \succ b \quad \forall b \in (A \cap W^*) \setminus \{a\})$ if $u(a) \geq u^*$ and $p(a, A) = 0$ if $u(a) < u^*$. To see that this is true observe that by definition of W^* and Axiom 2.8 $p(A \cap W^*, A) = 1$. To see that this is true, observe that if we assume that $p(A \cap W^*, A) < 1$ we must have that $p(a, A) > 0$ for some $a \notin W^*$ but that means by Axiom 2.8 that $p(a, A) = 1$ which is a contradiction of the fact that $p(A \cap W^*, A) > 0$.

Then $p(a, A) = 0$ if $u(a) < u^*$. For $a, b \in A \cap W^*$, the result follows as in (i). ■

2.A.4 Proof of Theorem 2.14

Proof. To prove (1) assume, by contradiction, that $U^* \neq \bar{U}^*$ and let $v \in \bar{U}^* \setminus U^*$. Then, it must be the case that $u(v) < u^*$ and $\bar{u}(v) \geq \bar{u}^*$. By completeness and the fact that $U^* \neq \emptyset$, $\exists a \in U^*$ and a $A \in \mathcal{D}$ such that $A = \{a, v\}$. Given that the data is represented by $(u, u^*, \{\gamma_A\}_{A \in \mathcal{D}})$, $p(a, A) = 1$. On the other hand, since $(\bar{u}, \bar{u}^*, \{\bar{\gamma}_A\}_{A \in \mathcal{D}})$ also represents the data it is the case that $p(a, A) \in (0, 1)$, which establishes a contradiction.

To prove (2) notice that from (1) $U^* = \bar{U}^*$. Since, $(u, u^*, \{\gamma_A\}_{A \in \mathcal{D}})$ and $(\bar{u}, \bar{u}^*, \{\bar{\gamma}_A\}_{A \in \mathcal{D}})$ are generated by the same FSSM, then u and u^* represent the preferences given by definition 2.9 for all $a \notin U^*$. Therefore, it must be the case that u is a strictly increasing

transformation of \bar{u} . on X/U^*

To prove (3) assume, by contradiction, that for some $A \in D$, $a \in A \cap U^*$

$$\gamma_A(ar_A b \quad \forall b \in \{A \cap U^*\} \setminus \{a\}) \neq \bar{\gamma}_A(ar_A b \quad \forall b \in \{A \cap U^*\} \setminus \{a\})$$

then $p(a, A | \gamma_A) \neq p(a, A | \bar{\gamma}_A)$ which in turn implies both, $(u, u^*, \{\gamma_A\}_{A \in \mathcal{D}})$ and $(\bar{u}, \bar{u}^*, \{\bar{\gamma}_A\}_{A \in \mathcal{D}})$ cannot represent the same data. ■

2.A.5 Proof of Theorem 2.17

Proof. First we prove (1) implies (2). If the complete data is generated by a FDSM then there is a triple (u, u^*, Γ_X) , take a realization of Γ_X with support on R_X and call it r_X , then define the linear ordering on $X \succ_X$: (1) $a \succ_X b$ if $ar_X b$ and $a, b \in U^*$, (2) $a \succ_X b$ if $u(a) > u(b)$ and $a, b \notin U^*$ and (3) $a \succ_X b$ if $a \in U^*$, $b \notin U^*$. Now, assign this linear ordering \succ_X the probability $\Gamma_X(r_X)$. It is direct to see that there is a Random Utility Maximization model without indifference with realizations \succ_X with probability $\Gamma_X(r_X)$. By (Block and Marschak 1960) it follows that the generated data set (\mathcal{D}, p) satisfies Total Monotonicity (Axiom 2.16). The fact that Axiom 2.8 and Axiom 2.10 hold follows from Theorem 2.11.

Second we prove (2) implies (1). Because FDSM is a subcase of FSSM and Axiom 2.8 and Axiom 2.10 hold we can build an utility $u : X \mapsto \mathbb{R}$ and a threshold $u^* \in \mathbb{R}$ such that $u(x) \geq u^*$ for all $x \in W^*$ and $u(x) < u^*$ for all $x \in X \setminus W^*$. Finally, if a complete stochastic choice dataset (\mathcal{D}, p) satisfies Axiom 2.16 then the model is a Random Maximization Utility model so we can recover a distribution over linear orders on X $\Gamma_X^Q : \bar{R}_X \mapsto [0, 1]$ such that $p(a, A) = \Gamma_X^Q(a\bar{r}_X b \quad \forall b \in A)$ thanks to (Falmagne 1978). Call its support \bar{R}_X the set of quasi-search ordering, because they are the results of both

deterministic (degenerate) utility maximization choice and fixed random searching and satisficing behavior. To construct the Γ_X we build each element of its support by taking an element of the support of the Quasi-search ordering $\bar{r}_X \in \bar{R}_X$ and we restrict it to W^* . Here we define an equivalence class on \bar{R}_X if they have the same restriction $\bar{r}_X|W^*$, if we have any two elements $\bar{r}_X, \bar{r}'_X \in \bar{R}_X$ that have the same restriction to W^* (i.e., $x\bar{r}_X y \iff x\bar{r}'_X y$ for $x, y \in W^*$) we say $\bar{r}_X \equiv_{W^*} \bar{r}'_X$, the equivalence class set is denoted as $[\bar{r}_X]_{\equiv_{W^*}} = \{\bar{r}'_X \in \bar{R}_X | \bar{r}_X \equiv_{W^*} \bar{r}'_X\}$ then we assign to the representative of the equivalence class or the restriction $r_X|W^*$ the probability corresponding to the sum $\sum_{\bar{r}'_X \in [\bar{r}_X]_{\equiv_{W^*}}} \Gamma_X^Q(\bar{r}'_X)$. For any given restricted ordering $\bar{r}_X|W^*$ we build its transitive closure or the set of transitive extensions to X and call this set $R_X(\bar{r}_X) \subset X \times X$. We assign each of the elements of this set $\hat{r}_X \in R_X(\bar{r}_X)$ the probability $\sum_{\bar{r}_X \in [\bar{r}_X]_{\equiv_{W^*}}} \Gamma_X^Q(\bar{r}_X) / |R_X(\bar{r}_X)|$ where the numerator is the probability of the restricted to W^* quasi-search ordering $\bar{r}_X|W^*$ and the denominator is the cardinality of the previously defined set. Doing this for all elements of \bar{R}_X we build a new support R_X with probabilities as indicated that provide us with Γ_X .

Note that Γ_X has full support due to how W^* is constructed. Because W^* is the always chosen set, we know that any element of the set of restrictions $\bar{r}_X|W^*$ such that for each $a \in W^*$, $a\bar{r}_X b$ for all $b \in W^* \setminus \{a\}$ has positive probability. It follows that by definition, for any $A \in \mathcal{D}$, and for any $x \in W^*$ we have $p(x, A) > 0$, this means that $\Gamma_X^Q(\bar{r}_X \in \bar{R}_X : x\bar{r}_X y \quad \forall y \in X \setminus \{x\}) > 0$. This implies that all representatives of the equivalence class of restricted orderings $r_X|W^*$ where x is searched first in W^* have positive probability. Then we have extended them to X with the uniform distribution for each set $R_X(\bar{r}_X)$ thus preserving the full support for the whole X . The reason is that the transitive closure to X of any $r_X|W^*$ contains linear search orders that have each $x \in W^*$ as the first searched element, because it contains the ordering that preserves the elements in W^* in the top and the rest at the bottom. But also it contains search

orders with each element of $X \setminus W^*$ at the top for each of such elements and W^* at the bottom all with positive probability.

We have extended each restriction $\bar{r}_X|W^*$ such that we can let Γ_X be the fixed distribution of search orders. We have built a FDSM or a triple (u, u^*, Γ_X) . To verify that this FDSM model generate (\mathcal{D}, p) notice that if we face a menu A we have the following cases:

(i) if $A \cap W^* = A$ then $p(a, A) = \Gamma_X(r_X \in R_X | a r_X b \quad \forall b \in A \setminus \{a\})$, to see that this is true observe that $\Gamma_X(r_X \in R_X | a r_X b \quad \forall b \in A \setminus \{a\}) = \Gamma_X^Q(\bar{r}_X \in \bar{R}_X | a \bar{r}_X b \quad \forall b \in A \setminus \{a\})$. In this case the first equality follows from the equivalence to random utility when restricted to W^* .

(ii) If $A \cap W^* = \emptyset$ then $p(a, A) = 1$ if $u(a) > u(b)$ for all $b \in A \setminus \{a\}$ and zero otherwise. This is direct from the fact that u represents C in such sets.

(iii) If $A \cap W^* \subset A$ then we have $p(a, A) = \Gamma_X(r_X | a r_X b \quad \forall b \in (A \cap U^*) \setminus \{a\})$ if $u(a) \geq u^*$ and $p(a, A) = 0$ if $u(a) < u^*$. To see that this is true observe that by definition of W^* and Axiom 2.8 $p(A \cap W^*, A) = 1$ then $p(a, A) = 0$ if $u(a) < u^*$. For $a, b \in A \cap W^*$ observe that by construction $\Gamma_X(r_X \in R_X | a r_X b \quad \forall b \in (A \cap U^*) \setminus \{a\}) = \Gamma_X^Q(\bar{r}_X \in \bar{R}_X | a \bar{r}_X b \quad \forall b \in (A \cap W^*) \setminus \{a\})$. Finally, observe that $\Gamma_X^Q(\bar{r}_X \in \bar{R}_X | a \bar{r}_X b \quad \forall b \in (A \cap W^*) \setminus \{a\}) = \Gamma_X^Q(\bar{r}_X \in \bar{R}_X | a \bar{r}_X b \quad \forall b \in A \setminus \{a\}) = p(a, A)$ because the facts that $p(c, A) = 0$ for any $c \notin W^*$ and that Γ_X^Q represents choice implies that $\Gamma_X^Q(\bar{r}_X \in \bar{R}_X | c \bar{r}_X a) = 0$. ■

2.A.6 Proof of Theorem 2.18

Proof. For (1) and (2) we use the results of Theorem 2.14.

(3) follows from the fact that $W^* \equiv U^*$ and the FDSM behaves as Random Utility in this for the elements in W^* . Assume by contradiction that $\Gamma_X(xr_X y) \neq \bar{\Gamma}_X(xr_X y)$ for some $x, y \in W^*$ but that means that in the menu $\{x, y\}$, $p(x, \{x, y\} | \Gamma_X) \neq p(x, \{x, y\} | \bar{\Gamma}_X)$ which is a contradiction. ■

2.A.7 Proof of Claim 2.19

Proof. To prove (1) note first that it is trivial that $W^* \subseteq \tilde{W}^*$. Moreover, a difference in the utility threshold level only affects decisions if $\{a \in X : u(a) \in [\tilde{u}^*, u^*]\} \cap A \neq \emptyset$, for some $A \in \mathcal{D}$, otherwise $W^* = \tilde{W}^*$. Assume now that $\tilde{A} \equiv \{a \in X : u(a) \in [\tilde{u}^*, u^*]\} \neq \emptyset$. By the definition of the FSSM (FDSM), it is the case $\tilde{A} \subset \tilde{W}^*$ and we get the desired result.

To prove (2) note that DM are identical up to u^*, \tilde{u}^* then if $W^* \neq \tilde{W}^*$, it must be the case that $u^* \neq \tilde{u}^*$. Assume by the way of contradiction that $u^* \leq \tilde{u}^*$, then $\tilde{u}^* > u^*$. Then, by the same argument as above either $W^* = \tilde{W}^*$ or $\{a \in X : u(a) \in [u^*, \tilde{u}^*]\} \neq \emptyset$ and then $W^* \setminus \tilde{W}^* \neq \emptyset$ which induces a contradiction. ■

2.A.8 Proof of Proposition 2.23

Proof. First we prove that if $u^* \geq \tilde{u}^*$ then $F_I(\bar{u}, A) \leq F_{\tilde{I}}(\bar{u}, A)$ for all \bar{u} .

For any menu $A \in \mathcal{D}$, a change in the utility threshold level only affects decisions if there exists $a \in A$ such that $u(a) \in [\tilde{u}^*, u^*]$, otherwise $F_I(\cdot, A) = F_{\tilde{I}}(\cdot, A)$ and the result is trivially true. Now consider the case where there exists at least one $a \in A$ such that $u(a) \in [\tilde{u}^*, u^*]$. We have two cases, (i) $A \cap U^* = \emptyset$ or (ii) $A \cap U^* \neq \emptyset$.

If (i) and $|\{u : u = u(a) \in [\tilde{u}^*, u^*] \text{ for some } a \in A\}| = 1$ then $F_I(\cdot, A) = F_{\tilde{I}}(\cdot, A)$ and we have the desired result. If $|\{u : u = u(a) \in [\tilde{u}^*, u^*] \text{ for some } a \in A\}| > 1$ then the

result follows by noticing that

$$F_{\tilde{I}}(\bar{u}, A) = \begin{cases} 0 & \text{for } \bar{u} < \tilde{u}^* \\ \sum_{a: \tilde{u}^* \leq u(a) < \max_{a \in A} u(a)} p^{\tilde{I}}(a, A) & \text{for } \tilde{u}^* \leq \bar{u} < \max_{a \in A} u(a) \\ 1 & \text{for } \max_{a \in A} u(a) \leq \bar{u} \end{cases}$$

with $\sum_{a: \tilde{u}^* \leq u(a) < \max_{a \in A} u(a)} p^{\tilde{I}}(a, A) > 0$ given full support assumption. On the other hand,

$$F_I(\bar{u}, A) = \begin{cases} 0 & \text{for } \bar{u} < \tilde{u}^* \\ 0 & \text{for } \tilde{u}^* \leq \bar{u} < \max_{a \in A} u(a) \\ 1 & \text{for } \max_{a \in A} u(a) \leq \bar{u} \end{cases}$$

If (ii) then we have that for any $a \in A \cap [\tilde{u}^*, u^*)$, it is the case that $u(a) < u(b)$ for all $b \in U^* \cap A \neq \emptyset$. Since $A \cap U^* \neq \emptyset$ and $A \cap \tilde{U}^* \neq \emptyset$, then

$$\begin{aligned} \sum_{b \in U^* \cap A} p^I(b, A) &= \sum_{b \in \tilde{U}^* \cap A} p^{\tilde{I}}(b, A) \\ \sum_{b \in \tilde{U}^* \cap A} p^{\tilde{I}}(b, A) &= \sum_{b \in U^* \cap A} p^{\tilde{I}}(b, A) + \sum_{b \in (\tilde{U}^* \setminus U^*) \cap A} p^{\tilde{I}}(b, A) \end{aligned}$$

By the full support assumption $\sum_{b \in (\tilde{U}^* \setminus U^*) \cap A} p^{\tilde{I}}(b, A) > 0$. Then

$$F_{\tilde{I}}(\bar{u}, A) = \begin{cases} 0 & \text{for } \bar{u} < \tilde{u}^* \\ \sum_{a: \tilde{u}^* \leq u(a) \leq \bar{u}} p^{\tilde{I}}(a, A) & \text{for } \tilde{u}^* \leq \bar{u} < u^* \\ (1 - \beta)F_I(\bar{u}, A) & \text{for } \tilde{u}^* \leq \bar{u} < u^* \\ 1 & \text{for } \bar{u} \geq u^* \end{cases}$$

with $\beta = \sum_{a: \tilde{u}^* \leq u(a) < u^*} p^{\tilde{I}}(a, A)$.

Therefore,

$$F_{\tilde{I}}(\bar{u}, A) - F_I(\bar{u}, A) = \begin{cases} 0 & \text{for } \bar{u} < \tilde{u}^* \\ \sum_{a: \tilde{u}^* \leq u(a) \leq \bar{u}} p^{\tilde{I}}(a, A) > 0 & \text{for } \tilde{u}^* \leq \bar{u} < u^* \\ \beta F_I(\bar{u}, A) > 0 & \text{for } \tilde{u}^* \leq \bar{u} < u^* \\ 0 & \text{for } \bar{u} \geq u^* \end{cases}$$

then $F_{\tilde{I}}(\bar{u}, A) \geq F_I(\bar{u}, A)$ for all \bar{u} .

Now we prove that if $F_I(\bar{u}, A) \leq F_{\tilde{I}}(\bar{u}, A)$ for all \bar{u} with $F_I(\bar{u}, A) \neq F_{\tilde{I}}(\bar{u}, A)$ then $u^* \geq \tilde{u}^*$. First notice that if $F_I(\bar{u}, A) \neq F_{\tilde{I}}(\bar{u}, A)$ then it must be the case that $u^* \neq \tilde{u}^*$. Assume by the way of contradiction that $u^* \leq \tilde{u}^*$, then $\tilde{u}^* > u^*$. Following the same argument as in the previous part we have that, if there exists $a \in A$ such that $u(a) \in [u^*, \tilde{u}^*)$ then $F_I(\bar{u}, A) \geq F_{\tilde{I}}(\bar{u}, A)$ for all \bar{u} with $F_I(\bar{u}, A) \neq F_{\tilde{I}}(\bar{u}, A)$ which leads to a contradiction. ■

2.A.9 Comparative Statics with Respect to Search Orders

First, we introduced the following class of equivalent search order for a given $A \in \mathcal{D}$.

Definition 2.51 ($\{a, b\}$ -equivalent search orders) Let $\gamma_A, \tilde{\gamma}_A$ be two probability distributions over search orders R_A for some $A \in \mathcal{D}$. We say that $\gamma_A, \tilde{\gamma}_A$ are $\{a, b\}$ -equivalent search orders for menu A if $\gamma_A(r_A) = \tilde{\gamma}_A(r_A)$ for all r_A such that $r_{A \setminus \{a\}}^* = \tilde{r}_{A \setminus \{a\}}$ and $r_{A \setminus \{b\}}^* = \tilde{r}_{A \setminus \{b\}}$ for some $a, b \in A$.

Proposition 2.52 (Saliency of Satisficing Elements) Let $I = (u, u^*, \{\gamma_A\}_{A \in \mathcal{D}})$ and $\tilde{I} = (u, u^*, \{\tilde{\gamma}_A\}_{A \in \mathcal{D}})$ be two FSSM, with cumulative probability distribution over utility levels $F_I(\bar{u}, A)$ and $F_{\tilde{I}}(\bar{u}, A)$. Let γ_A and $\tilde{\gamma}_A$ be $\{a, b\}$ -equivalent search orders as in Definition 2.51 with $\gamma_A(r_A^*) = \tilde{\gamma}_A(r_A^*) + \varepsilon$ and $\gamma_A(\tilde{r}_A) = \tilde{\gamma}_A(\tilde{r}_A) - \varepsilon$ for some $\varepsilon \in (\tilde{\gamma}(\tilde{r}_A), 1 - \tilde{\gamma}_A(r_A^*))$ and $a r_A^* b$ and $b \tilde{r}_A a$.

If $u(a) \geq u(b)$ then $F_I(\cdot, A)$ FOSD $F_{\tilde{I}}(\cdot, A)$. Moreover, if $F_I(\bar{u}, A) \leq F_{\tilde{I}}(\bar{u}, A)$ with $F_I(\bar{u}, A) \neq F_{\tilde{I}}(\bar{u}, A)$ then $u(a) > u(b) \geq u^*$

Proof. First we prove that if $u(a) > u(b)$ then $F_I(\cdot, A)$ FOSD $F_{\tilde{I}}(\cdot, A)$. There are four possible cases: (i) $a, b \notin U^*$ then $F_I(\cdot, A) = F_{\tilde{I}}(\cdot, A)$. (ii) If $a \in U^*$ and $b \notin U^*$, under the conditions of the proposition then $F_I(\cdot, A) = F_{\tilde{I}}(\cdot, A)$. (iii) If $b \in U^*$ and $a \notin U^*$, under the conditions of the proposition then $F_I(\cdot, A) = F_{\tilde{I}}(\cdot, A)$. (iv) Finally, consider the case where $a, b \in U^*$. Under the conditions of the proposition $p^I(a, A) > p^{\tilde{I}}(a, A)$ and $p^I(b, A) < p^{\tilde{I}}(b, A)$ while $p^I(c, A) = p^{\tilde{I}}(c, A)$ for all $c \in A \setminus \{a, b\}$ and the results follows from $u(a) > u(b)$

Now we prove that if $F_I(\bar{u}, A) \leq F_{\tilde{I}}(\bar{u}, A)$ for all \bar{u} with $F_I(\bar{u}, A) \neq F_{\tilde{I}}(\bar{u}, A)$ then $u(a) > u(b) \geq u^*$. Notice that the change in the probabilities of search orders only affects the relative probability of a, b being seen first, leaving unaltered all other relative probabilities. Therefore, the change from $\{\tilde{\gamma}_A\}_{A \in \mathcal{D}}$ to $\{\gamma_A\}_{A \in \mathcal{D}}$ only has an effect on the distribution over utility levels if $\{a, b\} \subseteq U^*$. Moreover, given the conditions of the proposition, $p^I(a, A) > p^{\tilde{I}}(a, A)$ and $p^I(b, A) < p^{\tilde{I}}(b, A)$. Then it must be that $u(a) > u(b)$. To show the latter, assume by contradiction that $u^* \leq u(a) < u(b)$, then $F_{\tilde{I}}(u(a), A) < F_I(u(a), A)$ which leads to a contradiction. ■

Corollary 2.53 *Under the conditions of Proposition 2.52, $u(a) \geq u(b)$ if and only if $E_u^I(A) \geq E_u^{\tilde{I}}(A)$.*

2.A.10 Proof of Proposition 2.28

Proof. First we prove that (1) implies (2). If a complete data has a FDSM without full support representation then there is a triple (u, u^*, Γ_X) , take a realization of Γ_X with support on some subset or all R_X and call it r_X , then define the linear ordering on X

\succ_X : (1) $a \succ_X b$ if $a r_X b$ and $a, b \in U^*$, (2) $a \succ_X b$ if $u(a) > u(b)$ and $a, b \notin U^*$ and (3) $a \succ_X b$ if $a \in U^*, b \notin U^*$. Now, assign this linear ordering \succ_X the probability $\Gamma_X(r_X)$. It is direct to see that there is a Random Utility Maximization model without indifference with realizations \succ_X with probability $\Gamma_X(r_X)$. By (Block and Marschak 1960) it follows that the generated data set (\mathcal{D}, p) satisfies Total Monotonicity (Axiom 2.16).

Now we prove that (2) implies (1). If a complete stochastic choice dataset (\mathcal{D}, p) satisfies Axiom 2.16 then the model is a Random Maximization Utility model so we can recover a distribution over linear orders on X $\Gamma_X : \bar{R}_X \mapsto [0, 1]$ such that $p(a, A) = \Gamma_X^Q(a \bar{r}_X b \forall b \in A)$ thanks to (Falmagne 1978). Interpret Γ_X as a fixed distribution search order, and select any injective $u : X \rightarrow \mathbb{R}$ and u^* such that $u(x) > u^*$ for all $x \in X$. By construction this triple (u, u^*, Γ_X) generates the observed data. ■

2.A.11 Proof of Theorem 2.32

Proof. First we prove that (1) implies (2). To show that Axiom 2.10 holds, assume, by the way of contradiction, that (i) a is stochastically revealed preferred to b and (ii) b is stochastically strictly revealed preferred to a . From (ii), given that data admits a FSSMI we must have (by the full support assumption) that $u(a) < u^*$ and $u(a) < u(b)$. If a is stochastically revealed preferred to b then there must exist a sequence of alternatives c_1, \dots, c_N and choice sets A_1, \dots, A_{N-1} , such that $c_1 = a$, $c_N = b$, $c_n, c_{n+1} \in A_n$ and $c_n \in C(A_n)$. If $u(b) > u^*$ then it must be the case that $u(c_{N-1}) > u^*$ (otherwise c_{N-1} could not be chosen with positive probability when c_N was available). Iterating on this argument implies that $u(a) > u^*$. If $u(b) < u^*$ then this implies that $u(c_{N-1}) > u(b)$. If $u(c_n) < u^*$ for all n then iterating on this argument implies that $u(a) \geq u(b)$. Otherwise, the previous argument implies that $u(a) > u^*$. Either provides a contradiction.

Now we prove that (2) implies (1). We say two items are revealed stochastically in-

different if a is revealed stochastically preferred to b and b is revealed stochastically preferred to a , in that case we denote aI^*b . We modify this relation $I^* \subseteq X \times X$ by removing its elements that have at least one item from always chosen set W^* that is non-empty by SARP. Formally we define the relation $I = \{(a, b) \in X \times X : (a, b) \in I^* \text{ } a \in X \setminus W^* \text{ or } b \in X \setminus W^*\} \cup D(W^*)$ where $D(W^*)$ is the diagonal ordering in W^* (i.e., it contains only the elements $(a, a) \in W^* \times W^*$). I is an equivalence relation because it is reflexive, symmetric and transitive. Because Axiom 2.10 holds I^* is an equivalence relation, and I is still an equivalence relation because it only eliminates the indifference of the items in W^* except for reflexivity, namely the elements $(a, a) \in I^*$ for $a \in W^*$. By Axiom 2.10 and the definition of W^* no item in $X \setminus W^*$ is revealed indifferent to an item in W^* . The relation I induces an equivalence class that we denote as $[a]$. We concentrate on the quotient set $X^I = X/I$, we define the canonical projection $j : X \mapsto X/I$ and its inverse mapping $j^{-1} : X/I \mapsto X$. We let $\mathcal{D}^I \equiv \{j(A)\}_{A \in \mathcal{D}}$ be the indexed set by \mathcal{D} . In particular define $p_I : X^I \times \mathcal{D}^I \mapsto [0, 1]$ as $p(a^I, A^I) = \sum_{a \in j^{-1}(a^I) \cap A} p(a, A)$ for $A \in \mathcal{D}$ such that $j(A) = A^I$ and $a^I \in A^I$, this mapping is well defined. If Axiom 2.10 holds it follows that the quotient dataset $\{p(a^I, A^I)\}_{a^I \in X^I, A^I \in \mathcal{D}^I}$ also satisfies SARP. Also observe, that in the quotient dataset the always chosen set $W^{I,*} = \{a^I \in X^I : p(a^I, A^I) > 0 \quad \forall A^I \in \mathcal{D}^I\}$ is such that $j^{-1}(a^I) \in W^*$ for all $a^I \in W^{I,*}$, this follows from the construction of I because the equivalence classes in W^* are singletons. Observe also that Axiom 2.8 holds in the quotient dataset $\{p(a^I, A^I)\}_{a^I \in X^I, A^I \in \mathcal{D}^I}$, because if $b^I \in X^I \setminus W^{I,*}$ then either $p(b^I, A^I) = 0$ or (exclusively) $p(b^I, A^I) = 1$. In fact, the set X^I is a finite choice set with elements associated with degenerated probabilities of choice $p([a] \cap A, A) = \sum_{a \in [a] \cap A} p(a, A) \in \{0, 1\}$ if $[a] \subseteq X \setminus W^*$. To see this is true assume that $p([a] \cap A, A) \in (0, 1)$ and $[a] \subseteq X \setminus W^*$, this means that there is a third element $c \in A$ such that $c \in A \setminus [a]$, that is stochastically revealed preferred to all $a \in [a]$ (i.e, $p(c, A) > 0$) and of course all elements $a \in [a]$ are stochastically revealed preferred to c , but that means that aIc for all $a \in [a]$ which means that $c \in [a]$, this is a contradiction. By theorem 2.11 we conclude that the

quotient dataset $\{p(a^I, A^I)\}_{a^I \in X^I, A^I \in \mathcal{D}^I}$ can be generated by a FSSM without indifference, thus we build a triple $(\bar{u}, u^* \{\bar{\gamma}\}_{A^I \in \mathcal{D}^I})$, that generates the quotient dataset.

With this in hand we build the FSSMI in the actual dataset $\{p(a, A)\}_{a \in X, A \in \mathcal{D}}$. (i) We build a utility function $u : X \mapsto \mathbb{R}$, by the composition $u = \bar{u} \circ j$ where j is the canonical projection defined above. By construction $u(a) > u^*$ for all $a \in W^*$ and $u(a) = u(b)$ if $b \in [a]$. Moreover, $u(b) < u^*$ for all $b \in X \setminus W^*$.

(ii) The search probabilities are defined over the quotient set, we build search probabilities for the actual set X . $\bar{\gamma}$ defines a full support search distribution on each $A^I \in \mathcal{D}^I$. Now we define γ by the following algorithm: For each menu $A \in \mathcal{D}$, we obtain the menu $A^I \equiv j(A) \in \mathcal{D}^I$ in the quotient dataset, then take R_{A^I} the support of $\bar{\gamma}_{A^I}$, now for any element $r_{A^I} \in R_{A^I}$ define the restriction $r_{A^I} | W^{I,*}$. Now build the set of linear search orders on A $R_A(r_{A^I} | W^{I,*}) = \{r_A \in R_A : a, b \in W^*, ar_A b \text{ if } j(a)r_{A^I} | W^{I,*} j(b)\}$. We assign to each element $r_A \in R_A(r_{A^I} | W^{I,*})$, the probability $\gamma_A(r_A) = \sum_{r'_{A^I} \in R_{A^I}} \bar{\gamma}_{A^I}(r'_{A^I}) \mathbf{1}(r'_{A^I} | W^{I,*} = r_{A^I} | W^{I,*}) / |R_A(r_{A^I} | W^{I,*})|$. This construction provides as with $\{\gamma_A\}_{A \in \mathcal{D}}$ that defines a FS random linear ordering on each \mathcal{D} .

(iii) To build the menu dependent tie breaking rules we calibrate them as follows $T(a, A) = p(a, A)$ if $a \in X \setminus W^*$ and $a \in [a]$ such that $p([a] \cap A, A) = 1$. If $p([a] \cap A, A) = 0$ then we let $T(a | A^{\sim a}) = 1 / |[a] \cap A|$. This guarantees a tie breaking rule that is always positive and that adds up to 1 as required.

We have generated a tuple $(u, u^*, \{\gamma_A\}_{A \in \mathcal{D}}, T)$ or a FSSMI representation that generates the complete dataset $\{p(a, A)\}_{a \in X, A \in \mathcal{D}}$. To verify this claim, notice that this follows immediately from applying Theorem 2.11 to generate the quotient dataset, and noticing that for non-satisficing elements we can generate the actual dataset using the calibrated

tie breaking rule directly and observing that the elements in $W^{I,*}$ have a one to one correspondence to the elements in W^* . ■

2.A.12 Proof of Theorem 2.33

Proof. (1)-(3) follows from Theorem 2.18. To prove (4) notice that, given $U^* = \bar{U}^*$, for all $a \notin U^*$, if $T(a|A^{\sim a}) \neq \bar{T}(a|A^{\sim a})$ then, from the definition of the model $p(a,A) \neq \bar{p}(a,A)$. ■

2.A.13 Proof of Theorem 2.34

Proof. To prove (1) notice that, since we do not allow for ties, $p(a,A) \in (0,1)$ if $a \in U^*$ then $\widetilde{W} \subseteq U^*$. Moreover if $a \in U^*$ then, given the full support assumption, $p(a,A) > 0$ for all $A \in \mathcal{D}$ then $a \in W^*$ as in Theorem 2.11.

(2) follows from Axiom 2.10. Notice that SARP does not requires complete data sets to guarantee the existence of a utility function that represents the revealed preference relation.

(3) follows from the definition of the model. Note that identification is only possible when surely revealed satiscing elements are available together in a menu. That is, we can identify the probability of a seen first than b in $A \in \mathcal{D}$, i.e. $\gamma_A(ar_A b \quad \forall b \in (A \cap U^*) \setminus \{a\})$ if $a \in A \cap U^*$. Moreover notice that if this is the case, then $a, b \in \widetilde{W}$. ■

2.A.14 Proof of Lemma 2.39

Proof. In order to prove the total monotonicity condition for $\hat{P}(a, A, \omega)$, we first need to define the following function for all $i \in \{1, \dots, I\}$ for each $A \in \mathcal{D}$ and $a \in A$:

$$f_i(a, A) = \sum_{D \in \mathbf{B}(A)} (-1)^{|D \setminus A|} \hat{p}_i(a, D)$$

where $\mathbf{B}(A)$ is the class of supersets of A (i.e., $\mathbf{B}(A) \equiv \{D \in \mathcal{D} | A \subseteq D\}$).

Now define

$$f(a, A) = \sum_{D \in \mathbf{B}(A)} (-1)^{|D \setminus A|} \hat{P}(a, A, \omega)$$

Observe that

$$f(a, A) = \sum_{D \in \mathbf{B}(A)} (-1)^{|D \setminus A|} \sum_{i=1}^I \omega_i \hat{p}_i(a, A)$$

by definition. We can interchange the first summation operator with the the second summation operator because the first does not depend on i , then

$$f(a, A) = \sum_{i=1}^I \omega_i \sum_{D \in \mathbf{B}(A)} (-1)^{|D \setminus A|} \hat{p}_i(a, A)$$

which implies:

$$f(a, A) = \sum_{i=1}^I \omega_i f_i(a, A),$$

by the assumption that each of the mappings \hat{p}_i is totally monotonic we have that $f_i(a, A) \geq 0$ for all $i \in \{1, \dots, I\}$ thus establishing the result. ■

2.A.15 Proof of Lemma 2.40

Before the proof of Lemma 2.40 we need the following preliminaries.

Remark 2.54 *It will be useful to notice that the FDSM-CRT can be written in the following*

form:

$$p(a, A) = \sum_{r_X \in R_X} p_{r_A=r_X|A, \tau}^{MM}(a, A) \gamma(r_X) + p_{\tau}^{MM}(o, A) \mathbf{1}(u(a) > u(b) \forall b \in A \setminus \{a\}),$$

where $p_{r_A=r_X|A, \tau}^{MM}(a, A) = \tau(a) \prod_{b \in A: b r_A a; r_A=r_X|A} (1 - \tau(b))$ is numerically equivalent to the MM probability of $a \in A$ being chosen and $r_X|A$ is the restriction of the ordering r_X to the set A . And $p_{\tau}^{MM}(o, A) = \prod_{c \in A} (1 - \tau(c))$ is numerical equivalent to the MM probability of the default alternative. Here there is no default alternative and we mean by $p^{MM}(o, A)$ the probability of no element in A being satisficing (keeping the MM notation for intuition).

Definition 2.55 (Successive differences for mappings) The successive differences for a mapping $\hat{p} : X \times \mathcal{D} \mapsto [0, 1]$ for a fixed $a \in A$ and $A \in \mathcal{D}$ and the probability $\hat{p}(a, A)$, is defined recursively as:

$$\Delta_{A_1} \hat{p}(a, A) = \hat{p}(a, A) - \hat{p}(a, A \cup A_1) \text{ for } A, A_1 \in \mathcal{D},$$

$\Delta_{A_n} \cdots \Delta_{A_1} \hat{p}(a, A) = \Delta_{A_{n-1}} \cdots \Delta_{A_1} \hat{p}(a, A) - \Delta_{A_{n-1}} \cdots \Delta_{A_1} \hat{p}(a, A \cup A_n)$ for all $n \geq 2$ and for all $A, A_1 \cdots A_n \in \mathcal{D}$.

Definition 2.56 (Weakly increasing successive differences) For a mapping $\hat{p} : X \times \mathcal{D} \mapsto [0, 1]$ and a fixed $a \in X$, and all $A \in \mathcal{D}$ such that $a \in A$ and any $\{A_i\}_{i=1}^n \in \mathcal{D}^n$ the mapping \hat{p} satisfies the weakly increasing successive difference property when the successive differences are non-negative $\Delta_{A_n} \cdots \Delta_{A_1} \hat{p}(a, A) \geq 0$ for all $n \geq 1$.

Proof. Notice that any mapping $\hat{p} : X \times \mathcal{D} \mapsto [0, 1]$ for a fixed (a, A) such that $a \in A$ for all $A \in \mathcal{D}$ is totally monotone if and only if it satisfies the weakly increasing successive differences ((Molchanov 2005)).

Define $\hat{P}(a, A) = T(A) \mathbf{1}(u(a) > u(b) \forall b \in A \setminus \{a\})$ where $T(A) = \prod_{c \in A} (1 - \tau(c))$ is totally monotone by MM (since $T(A) = p_{MM, \tau}^{MM}(o, A)$ is numerically equivalent to

the probability of a default choice in MM as explained in the remark) and thus satisfies for all $A \in \mathcal{D}$ such that $a \in A$ and any $\{A_i\}_{i=1}^n \in \mathcal{D}^n$ the weakly increasing successive differences $\Delta_{A_n} \cdots \Delta_{A_1}(1 - P(A)) \geq 0$ for all $n \geq 1$, the same is true for $C_a(A) = \mathbf{1}(u(a) > u(b) \forall b \in A \setminus \{a\})$ such that $\Delta_{A_n} \cdots \Delta_{A_1} \mathbf{1}(u(a) > u(b) \forall b \in A \setminus \{a\}) \geq 0$.

Notice that

$$\begin{aligned} \Delta_{A_1} \hat{P}(a, A) &= \hat{P}(a, A) - \hat{P}(a, A \cup A_1) \\ &= T(A)C_a(A) - T(A \cup A_1)C_a(A \cup A_1) \\ &= \begin{cases} 0 & C_a(A) = 0 \\ \Delta_{A_1} T(A) & C_a(A \cup A_1) = 1 \\ T(A) & C_a(A) = 1, C_a(A \cup A_1) = 0 \end{cases} \end{aligned}$$

Also for the next difference $\Delta_{A_2} \Delta_{A_1} \hat{P}(a, A) = T(A)C_a(A) - T(A \cup A_1)C_a(A \cup A_1) - [T(A \cup A_2)C_a(A \cup A_2) - T(A \cup A_1 \cup A_2)C_a(A \cup A_1 \cup A_2)]$.

Now

$$\begin{aligned} \Delta_{A_2} \Delta_{A_1} \hat{P}(a, A) &= \Delta_{A_1} \hat{P}(a, A) - \Delta_{A_1} \hat{P}(a, A \cup A_2) \\ &= \begin{cases} 0 & C_a(A) = 0 \\ \Delta_{A_1} T(A) & C_a(A \cup A_1) = 1, C_a(A \cup A_2) = 0 \\ T(A) & C_a(A) = 1, C_a(A \cup A_1) = 0, C_a(A \cup A_2) = 0 \\ \Delta_{A_2} T(A) & C_a(A) = 1, C_a(A \cup A_2) = 1, C_a(A \cup A_1) = 0 \\ \Delta_{A_2} \Delta_{A_1} T(A) & C_a(A \cup A_1 \cup A_2) = 1 \end{cases} \end{aligned}$$

Then we notice that in general the following properties hold:

(i) Null operator: $\Delta_{\emptyset} \Delta_{A_{n-1}} \cdots \Delta_{A_1} T(A) = 0$,

(ii) Absorption: $\Delta_{A_{n-1}} \Delta_{A_{n-1}} \cdots \Delta_{A_1} T(A) = \Delta_{A_{n-1}} \cdots \Delta_{A_1} T(A)$.

$\Delta_{A_1} \hat{P}(a, A) \in \{\Delta_{X_1} T(A)\}$ for $X_1 \in \{\emptyset, A_1\}$ and its value depends on the value of $C_a(A)$ and $C_a(A \cup A_1)$ but in any case we have

$\Delta_{X_1} T(A) \geq 0$ for all $X_1 \in \{\emptyset, A_1\}$. Notice that $\Delta_{\emptyset} T(A) = 0$.

Also, $\Delta_{A_2} \Delta_{A_1} \hat{P}(a, A) \in \{\Delta_{X_{2,2}} \Delta_{X_{2,1}} T(A)\}$ where $X_{2,1} \in \{\emptyset, A_1, A_2\}$ and $X_{2,2} \in \{\emptyset, A_1, A_2\}$ where the actual combination depends on the values of the vector

$$\{C_a(A \cup X_{2,1} \cup X_{2,2})\}_{X_{2,1} \in \{\emptyset, A_1, A_2\}, X_{2,2} \in \{\emptyset, A_1, A_2\}}$$

again in any case:

$\Delta_{X_{2,2}} \Delta_{X_{2,1}} T(A) \geq 0$ where $X_{2,1} \in \{\emptyset, A_1, A_2\}$ and $X_{2,2} \in \{\emptyset, A_1, A_2\}$. Notice that $\Delta_{\emptyset} \Delta_{X_{2,1}} T(A) = \Delta_{X_{2,1}} T(A) - \Delta_{X_{2,1}} T(A) = 0$, $\Delta_{A_2} \Delta_{A_2} T(A) = \Delta_{A_2} T(A) - \Delta_{A_2} T(A \cup A_2) = \Delta_{A_2} T(A)$.

The induction hypothesis is:

$\Delta_{A_{n-1}} \cdots \Delta_{A_1} \hat{P}(a, A) \in \{\Delta_{X_{n-1, n-1}} \cdots \Delta_{X_{n-1, 1}} T(A)\}$ for $X_{n-1, i} \in \{\emptyset, A_1, A_2, \dots, A_{n-1}\}$ for all $i \in \{1, \dots, n-1\}$,

with $\Delta_{X_{n-1, n-1}} \cdots \Delta_{X_{n-1, 1}} T(A) \geq 0$ for any $X_{n-1, i} \in \{\emptyset, A_1, A_2, \dots, A_{n-1}\}$ for all $i \in \{1, \dots, n-1\}$.

We have to prove that

$$\Delta_{A_n} \cdots \Delta_{A_1} \hat{P}(a, A) \geq 0,$$

Notice that $\Delta_{A_n} \cdots \Delta_{A_1} \hat{P}(a, A) = \Delta_{A_{n-1}} \cdots \Delta_{A_1} \hat{P}(a, A) - \Delta_{A_{n-1}} \cdots \Delta_{A_1} \hat{P}(a, A \cup A_n)$ by definition.

Now by the induction step:

$$\Delta_{A_n} \cdots \Delta_{A_1} \hat{P}(a, A) \in \{\Delta_{X_{n-1,n-1}} \cdots \Delta_{X_{n-1,1}} T(A) - \Delta_{X_{n-1,n-1}} \cdots \Delta_{X_{n-1,1}} T(A \cup A_n)\} \text{ for } X_{n-1,i} \in \{\emptyset, A_1, A_2, \dots, A_{n-1}\} \text{ for all } i \in \{1, \dots, n-1\}.$$

Finally by the definition of the difference operator:

$$\Delta_{X_{n-1,n-1}} \cdots \Delta_{X_{n-1,1}} T(A) - \Delta_{X_{n-1,n-1}} \cdots \Delta_{X_{n-1,1}} T(A \cup A_n) = \Delta_{A_n} \Delta_{X_{n-1,n-1}} \cdots \Delta_{X_{n-1,1}} T(A).$$

We notice that $\Delta_{A_n} \Delta_{X_{n-1,n-1}} \cdots \Delta_{X_{n-1,1}} T(A) \geq 0$ is positive by total monotonicity of T , that preserves works for any combination of $X_{n-1,i} \in \{\emptyset, A_1, A_2, \dots, A_{n-1}\}$ for all $i \in \{1, \dots, n-1\}$.

Thus $\Delta_{A_n} \cdots \Delta_{A_1} \hat{P}(a, A) \geq 0$ for any fixed $a \in A$ and all such $A \in \mathcal{D}$ and for all $\{A_i\}_{i=1}^n \in \mathcal{D}^n$ for all $n \geq 1$.

We conclude that \hat{P} is a total monotone mapping. ■

2.A.16 Proof of Lemma 2.36

Proof. (I) If the data is generated by FSSM-CRT then it satisfies SARP

First notice that if $\tau(a) > 0$ then it follows under full support that $p(a,A) > 0$. Furthermore, by full support and common threshold we observe also that $p(a,A) > 0$ when there is a $b \in A \setminus \{a\}$ such that $\tau(b) > 0$ only if $\tau(a) > 0$. Notice that if there is a $b \in A \setminus \{a\}$ with $\tau(b) > 0$ and $\tau(a) = 0$, then by the common threshold and monotonicity of a CDF F_{u^*} we know that it cannot be the case that $u(a) > u(b)$, thus $p(a,A) = 0$ (because $\tau(a) = 0$ implies the term $\mathbf{1}(u(a) > u(b)) \forall b \in A \setminus \{a\} = 0$) under the CRT assumption).

To show that Axiom 2.10 holds, assume, by the way of contradiction, that (i) a is stochastically revealed preferred to b and (ii) b is stochastically strictly revealed preferred to a . From (ii), given that data admits a FSSM-CRT we must have (by the full support assumption) that $\tau(a) = Pr(u(a) > u^*) = 0$ and $u(a) < u(b)$. If a is stochastically revealed preferred to b then there must exist a sequence of alternatives c_1, \dots, c_N and choice sets A_1, \dots, A_{N-1} , such that $c_1 = a$, $c_N = b$, $c_n, c_{n+1} \in A_n$ and $c_n \in C(A_n)$. If $\tau(b) = Pr(u(b) > u^*) > 0$ then it must be the case that $\tau(c_{N-1}) = Pr(u(c_{N-1}) > u^*) > 0$ (otherwise c_{N-1} could not be chosen with positive probability when c_N was available). Iterating on this argument implies that $\tau(a) = Pr(u(a) > u^*) > 0$.

If $\tau(b) = Pr(u(b) > u^*) = 0$ then this implies that $u(c_{N-1}) > u(b)$. If $Pr(u(c_n) > u^*) = 0$ for all n then iterating on this argument implies that $u(a) > u(b)$. Otherwise, the previous argument implies that $\tau(a) = Pr(u(a) > u^*) > 0$. Either provides a contradiction.

(II) A dataset that has a FSSM-CRT representation satisfies the Deterministic no sat-

isficing choice axiom.

Observe that $W^* = \{a \in X | \tau(a) = Pr(u(a) > u^*) > 0\}$ under full support, in that sense when $a \in X \setminus W^*$ and we have a menu A such that $a \in A$: (i) Either there is a $b \in W^* \cap A$ in which case $u(b) > u(a)$ because $\tau(b) > \tau(a) = 0$ only if $u(b) > u(a)$. This is because $Pr(u(b) > u^*) > 0 = Pr(u(a) > u^*)$ only if $u(b) > u(a)$ thus $p(a, A) = 0$ (under the common random threshold). (ii) Or there is no $b \in W^* \cap A$ in which case $1 - P(A) = 1$, thus making the probability $p(a, A) = \mathbf{1}(u(a) > u(c) \forall c \in A)$ which is by definition either $p(a, A) = 0$ or $p(a, A) = 1$. ■

2.A.17 Proof of Theorem 2.37

Proof. First we prove that (1) implies (2).

If a dataset is generated by a FSSM with parameters $\{\{\gamma_A\}_{A \in \mathcal{D}}, u, \bar{u}^*\}$ then it can be generated by a FSSM-CRT $\{\{\gamma_A\}_{A \in \mathcal{D}}, u^{RT}, \tau\}$ with $\tau(a) = Pr(u(a) > \bar{u}^*)$ where \bar{u}^* is a constant random variable, such that $\tau(a) = 1$ for all elements in FSSM such that $u(a) > \bar{u}^*$ and $\tau(b) = 0$ for $u(b) < \bar{u}^*$, with the same utility $u^{RT} \equiv u$ and the same random search function $\gamma_A^{RT} \equiv \gamma_A$. In other words, FSSM is a special case of FSSM-CRT with a constant threshold.

Now we prove that (2) implies (1).

If a dataset is generated by FSSM-CRT by Lemma 2.36 it satisfies satisfies Deterministic no satisficing choice (Axiom 2.8), SARP (Axiom 2.10). Thus by Theorem 2.11 we can build an FSSM that generates the data. Thus (2) implies (1). ■

2.A.18 Proof of Lemma 2.41

Proof. (I) If the data is generated by FDSM-CRT then $p(a, A)$ satisfies total monotonicity for all $a \in A$ and all $A \subseteq X$.

We define the MM model for a fixed linear search r_A as $p_{r_A, \tau}^{MM}(a, A) = \tau(a) \prod_{b \in B_{r_A}(a)} (1 - \tau(b))$ and $p_{\tau}^{MM}(o, A) = \prod_{c \in A} (1 - \tau(c))$. Now notice that the FDSM-CRT can be written as the average of the MM models, plus a correction for non satisficing cases.

$$p(a, A) = \sum_{r_x \in R_X} p_{r_A=r_x|A, \tau}^{MM}(a, A) \gamma(r_x) + p_{\tau}^{MM}(o, A) \mathbf{1}(u(a) > u(b) \forall b \in A \setminus \{a\})$$

It is evident by MM that $p_{r_A, \tau}^{MM}(a, A)$ satisfies total monotonicity for all $r_A \in R_A$ therefore by Lemma 2.39 we know that a summation (here a convex combination) of total monotonic mappings also satisfies total monotonicity, thus the term $\sum_{r_x \in R_X} p_{r_A=r_x|A, \tau}^{MM}(a, A) \gamma(r_x)$ satisfies total monotonicity.¹² The same argument is used for $p_{\tau}^{MM}(o, A)$ that by MM is also totally monotonic.

Also, notice that the indicator function $\mathbf{1}(u(a) > u(b) \forall b \in A \setminus \{a\})$ satisfies total monotonicity because it can be understood a degenerate random utility distribution. We know also that $p_{\tau}^{MM}(o, A) \mathbf{1}(u(a) > u(b) \forall b \in A \setminus \{a\})$ is totally monotonic by Lemma 2.40.¹³

¹²Notice that $\hat{p}(a, A) = \sum_{r_x \in R_X} p_{r_A=r_x|A, \tau}^{MM}(a, A) \gamma(r_x)$ satisfies total monotonicity however this term is a quasi-probability because it does not necessarily adds up to 1, so it is not a RUM in general. RUM is equivalent to total monotonicity, being non negative and adding up to 1.

¹³A non-constructive but equally valid argument is that $1 - p_{\tau}^{MM}(o, A) \mathbf{1}(u(a) > u(b) \forall b \in A \setminus \{a\})$ can be understood as the capacity of a random set that is the union of two other two independent random sets with capacities $1 - p_{\tau}^{MM}(o, A)$ and $(1 - \mathbf{1}(u(a) > u(b) \forall b \in A \setminus \{a\}))$ corresponding to the random sets $\hat{W} = \{x \in X : u(a) > u^*\}$ where u^* is the random threshold and $B(a) = \{b \in X : u(b) > u(a)\}$ that is the deterministic set of better than a items under the utility u . By (Molchanov 2005) $p_{\tau}^{MM}(o, A) \mathbf{1}(u(a) > u(b) \forall b \in A \setminus \{a\})$ is totally monotone, and $p_{\tau}^{MM}(o, A) \mathbf{1}(u(a) > u(b) \forall b \in A \setminus \{a\}) = Pr(\hat{W} \cup B(a) \cap A = \emptyset)$ the probability that A does not have anything satisficing or better than a .

Since $p(a, A)$ is the summation of two totally monotonic mappings again by Lemma 2.39 we conclude that $p(a, A)$ satisfies total monotonicity.

(II) If the data is generated by FDSM-CRT then it satisfies SARP. This follows from the fact that FDSM-CRT is an special case of FSSM-CRT and by Lemma 2.36 we know that it satisfies SARP.

(III) A dataset that has a FDSM-CRT representation satisfies the Deterministic no satisficing choice axiom. This follows from the fact that FDSM-CRT is an special case of FSSM-CRT and by Lemma 2.36 we know that it satisfies the Deterministic no satisficing choice axiom. ■

2.A.19 Proof of Theorem 2.42

Proof. First we prove that (1) implies (2).

If a dataset is generated by a FDSM with parameters $\{\{\gamma_A\}_{A \in \mathcal{D}}, u, \bar{u}^*\}$ then it can be generated by a FDSM-CRT $\{\{\gamma_A\}_{A \in \mathcal{D}}, u^{RT}, \tau\}$ with $\tau(a) = Pr(u(a) > \bar{u}^*)$ where \bar{u}^* is a constant random variable, such that $\tau(a) = 1$ for all elements in FDSM such that $u(a) > \bar{u}^*$ and $\tau(b) = 0$ for $u(b) < \bar{u}^*$, with the same utility $u^{RT} \equiv u$ and the same random search function $\gamma_A^{RT} \equiv \gamma_A$. In other words, FDSM is a special case of FDSM-CRT with a constant threshold.

Now we prove that (2) implies (1).

If a dataset is generated by FDSM-CRT by Lemma 2.41 it satisfies satisfies Deterministic no satisficing choice (Axiom 2.8), SARP (Axiom 2.10) and Total Monotonicity (Axiom 2.16). Thus by Theorem 2.17 we can build an FDSM that generates the data.

Thus (2) implies (1). ■

2.A.20 Proof of Theorem 2.48

Proof. First we prove that (1) implies (2).

First we recall that for a fixed $u \in \mathcal{U}$, $p_u^{FDSM}(a, A)$ satisfies Total Monotonicity (Axiom 2.16). By Lemma 2.39, the weighted average of totally monotonic mappings is also totally monotone $p(a, A) = \sum_{u \in \mathcal{U}} \rho(u) p_u^{FDSM}(a, A)$, thus RU-FDSM is totally monotone.

Second we notice that if $x \in X \setminus W^*$ it is never satisficing that is for all $u \in \mathcal{U}$ with $\rho(u) > 0$, $u(x) < u^*$, this means that for any fixed $u \in \mathcal{U}$, $p_u^{FDSM}(x, A) = 0$ if $A \cap W^* \neq \emptyset$, this in turn implies that $p(x, A) = \sum_{u \in \mathcal{U}} \rho(u) p_u^{FDSM}(x, A) = 0$. Thus degeneracy is established for RU-FDSM.

Now we prove that (2) implies (3).

With Claim 2.45 in hand, we define the following virtual dataset. We define the equivalence class $\sim = \{(a, b) \in X \times X : a \in X \setminus W^*, b \in X \setminus W^*\} \cup \{(c, c) \in X \times X : c \in W^*\}$ (it is symmetric, reflexive and transitive).

We define the set $X^\sim = X / \sim$ as the quotient space with respect to the equivalence relation. In words, we are “shrinking” all never satisficing elements to a singleton. We concentrate on the quotient set $X^\sim = X / \sim$, we define the canonical projection $j : X \mapsto X / \sim$ and its inverse mapping $j^{-1} : X / \sim \mapsto X$. We let $\mathcal{D}^\sim \equiv \{j(A)\}_{A \in \mathcal{D}}$ be the indexed set by \mathcal{D} . In particular define $p_\sim : X^\sim \times \mathcal{D}^\sim \mapsto [0, 1]$ as $p(a^\sim, A^\sim) = \sum_{a \in j^{-1}(a^\sim) \cap A} p(a, A)$ for $A \in \mathcal{D}$ such that $j(A) = A^\sim$ and $a^\sim \in A^\sim$, this mapping is well defined. If Degeneracy (Axiom 2.44) holds it follows that the quotient dataset

$\{p(a^\sim, A^\sim)\}_{a^\sim \in X^\sim, A^\sim \in \mathcal{D}^\sim}$ satisfies Deterministic not satisficing choice (Axiom 2.8) in X^\sim , because defining $W^{*\sim} = \{a^\sim \in X^\sim \mid p_\sim(a^\sim, A^\sim) > 0 \ \forall A^\sim \in \mathcal{D}^\sim\}$, we notice $X^\sim \setminus W^{*\sim}$ corresponds by construction to a singleton $\{z^\sim\} \equiv X^\sim \setminus W^{*\sim}$ when there is non-satisficing element (or empty if not) with probability $p(z^\sim, A^\sim) = \sum_{a \in j^{-1}(z) \cap A} p(a, A) \in \{0, 1\}$ in all $A^\sim \in \mathcal{D}^\sim$. In particular, it is exactly 1 only when $A^\sim \equiv \{z^\sim\}$ and zero otherwise. Also notice that Degeneracy implies SARP in $\{p(a^\sim, A^\sim)\}_{a^\sim \in X^\sim, A^\sim \in \mathcal{D}^\sim}$, because all $x^\sim, y^\sim \in W^{*\sim}$ are declared stochastically revealed preferred to one another (i.e., stochastically revealed indifferent) and $x^\sim \in W^{*\sim}$ is always strictly revealed preferred to z^\sim (or the non-satisficing singleton $\{z^\sim\} \equiv X^\sim \setminus W^{*\sim}$). Thus $C^\sim(A^\sim) = \{a^\sim \in A^\sim : p_\sim(a^\sim, A^\sim) > 0\}$ satisfies SARP. Also it is simple to see that the dataset in the quotient space X^\sim also satisfies total monotonicity because $p(a^\sim, A^\sim) = \sum_{a \in j^{-1}(a^\sim) \cap A} p(a, A)$ for $A \in \mathcal{D}$ such that $j(A) = A^\sim$ and $a^\sim \in A^\sim$ is a sum of total monotonic mappings $p(a, A)$ by (2) Total Monotonicity (Axiom 2.16). By the theorem that characterizes the FDSM (Theorem 2.17) the dataset $\{p(a^\sim, A^\sim)\}_{a^\sim \in X^\sim, A^\sim \in \mathcal{D}^\sim}$ in the quotient space has a FDSM representation, thus we build a triple $(\bar{u}, u^*, \{\bar{\gamma}\}_{A^\sim \in \mathcal{D}^\sim})$, that generates the quotient dataset.

With this in hand we build the RU-FDSM in the actual dataset $\{p(a, A)\}_{a \in X, A \in \mathcal{D}}$. (i) We build a utility function $u : X \mapsto \mathbb{R}$, by the composition $u = \bar{u} \circ j$ where j is the canonical projection defined above. By construction $u(a) > u^*$ for all $a \in W^*$ and $u(b) < u^*$ for all $b \in X \setminus W^*$. We build now (ρ, \mathcal{U}) , by making all elements in it have the following restriction $\hat{u} \in \mathcal{U}$ with $\rho(\hat{u}) > 0$ is such that $\hat{u}(a) = u(a)$ for all $a \in W^*$. For $b \in X \setminus W^*$ we use total monotonicity (Axiom 2.16) that holds for the restricted dataset $\{p(a, A)\}_{a \in X \setminus W^*, A \in \mathcal{D}, A \subseteq X \setminus W^*}$ to obtain a random utility by (Falmagne 1978) defined on $X \setminus W^*$, we notice that under this $p(a, A) = \rho(\hat{u} : \hat{u}(a) > \hat{u}(b) \ \forall b \in A \setminus \{a\})$ for $a \in A \subseteq X \setminus W^*$, we fix an injective utility $\hat{u}_{X \setminus W^*} : X \setminus W^* \mapsto \mathbb{R}$ with $\rho(\hat{u}_{X \setminus W^*})$ compatible with $\rho(\hat{u}_{X \setminus W^*} : \hat{u}_{X \setminus W^*}(a) > \hat{u}_{X \setminus W^*}(b) \ \forall b \in A \setminus \{a\})$ with $A \subseteq X \setminus W^*$ (we omit the trivial construction). We extend $\hat{u}_{X \setminus W^*} : X \setminus W^* \mapsto \mathbb{R}$ to the grand set X , by defining $\hat{u} : X \mapsto \mathbb{R}$

using $\hat{u}(a) = u(a)$ for all $a \in W^*$ and $\hat{u}(b) = \hat{u}_{X \setminus W^*}$ for all $b \in X \setminus W^*$, this is an injective utility in X with mass $\rho(\hat{u}) = \rho(\hat{u}_{X \setminus W^*})$. We have built (ρ, \mathcal{U}) that has the Always satisficing random utility property.

(ii) The search probabilities are defined over the quotient set, we build search probabilities for the actual set X . $\bar{\gamma}$ defines a full support search distribution on each $A^\sim \in \mathcal{D}^\sim$. Now we define γ by the following algorithm: For each menu $A \in \mathcal{D}$, we obtain the menu $A^\sim \equiv j(A) \in \mathcal{D}^\sim$ in the quotient dataset, then take R_{A^\sim} the support of $\bar{\gamma}_{A^\sim}$, now for any element $r_{A^\sim} \in R_{A^\sim}$ define the restriction $r_{A^\sim}|W^*$. Now build the set of linear search orders on A $R_A(r_{A^\sim}|W^*) = \{r_A \in R_A : a, b \in W^*, ar_A b \text{ if } j(a) r_{A^\sim}|W^* j(b)\}$. We assign to each element $r_A \in R_A(r_{A^\sim}|W^*)$, the probability $\gamma_A(r_A) = \sum_{r_{A^\sim} \in R_{A^\sim}} \bar{\gamma}_{A^\sim}(r_{A^\sim}) \mathbf{1}(r_{A^\sim}|W^* = r_A|W^*) / |R_A(r_{A^\sim}|W^*)|$. This construction provides as with $\{\gamma_A\}_{A \in \mathcal{D}}$ that defines a Full Support random linear ordering on each \mathcal{D} and by Total Monotonicity (Axiom 2.16) $\bar{\gamma}$ has the Fixed Distribution property and we build γ on the basis on it, extending it to the original dataset with a uniform rule, it follows that $\{\gamma_A\}_{A \in \mathcal{D}}$ has the Fixed Distribution property.

It is clear that for fixed $u \in \mathcal{U}$ constructed above with $\rho(u) > 0$, we have an FDSM with $(u, u^*, \{\gamma_A\}_{A \in \mathcal{D}})$ where $u, \{\gamma_A\}_{A \in \mathcal{D}}$ are constructed above and u^* is the same as the FDSM in the quotient space. Thus we have built a RU-FDSM

$$p(a, A) = \sum_{u \in \mathcal{U}} \rho(u) p_u^{FDSM}(a, A)$$

with FDSM with $(u, u^*, \{\gamma_A\}_{A \in \mathcal{D}})$ with the always satisficing property for (ρ, \mathcal{U}) . Finally, this generates the dataset (p, \mathcal{D}) by noticing that for $a \in W^*$ $p(a, A) = p_{\bar{u}}^{FDSM}(a, A)$ for any fixed $\bar{u} \in \mathcal{U}$ with $\rho(\bar{u}) > 0$ (thanks to the always satisficing property), so the fact that the constructed RU-FDSM (with always satisficing property) generates this proba-

bilities follows from Theorem 2.17. For $a \in X \setminus W^*$, we have $p(a, A) = \mathbf{1}(\forall c \in A : \bar{u}(c) \leq u^*) \rho(u : u(a) > u(b) \forall b \in A \setminus \{a\})$ for any fixed $\bar{u} \in \mathcal{U}$ with $\rho(\bar{u}) > 0$ (again due to the always satisficing property), we have two cases $\mathbf{1}(\forall c \in A : \bar{u}(c) \leq u^*) = 1$, in which case the fact that the constructed RU-FDSM generates the data follows from (Falmagne 1978) as these cases are equivalent to random utility in the restricted dataset defined in $X \setminus W^*$. The remaining case is $\mathbf{1}(\forall c \in A : \bar{u}(c) \leq u^*) = 0$ in which case we know that $\exists c \in A$ such that $c \in W^*$ and by Degeneracy (Axiom 2.44) we conclude that $p(a, A) = 0$ as it should be. Thus the constructed RU-FDSM generates the data.

Finally we prove (3) implies (1).

We notice that if complete stochastic choice dataset (\mathcal{D}, p) can be generated by a RU-FDSM with Always satisficing random utility distribution, it is by definition a special case of RU-FDSM, with the additional Always satisficing restriction on (ρ, \mathcal{U}) thus the dataset (\mathcal{D}, p) is also generated by a RU-FDSM. ■

Conclusion

Chapter 1 proposes a predictive ability approach for the assessment of the performance of a model when this is described by a set of axioms. Following this approach, I judge the model for how useful it is in terms of providing precise predictions. Intuitively, higher precision comes as a result of lots of information about preferences that can be inferred from data while requiring small errors to satisfy the model. Therefore, the predictive ability approach establishes a natural trade-off between fit and power conditional on observed behavior, providing a meaningful answer to the fit and power problems, a long standing problem in the literature. Moreover, by considering conditional power the proposed approach reflects also the effect of the actual pattern of choices observed, and I show that traditional power masks significant differences on the amount of information that it can be learned from data to predict behavior.

Additionally, the construction provided for predictive distribution can be used to forecast consumer's responses to policy or market interventions. I study the performance of the proposed measures with the experimental data from Choi et al (2007). As expected my measures are positively correlated with the amount of preference information that can be learned from data and with other measures of fit in the literature only through the error process. I also show that more observations enhance predictive precision -the additional revealed information outweighs the potential increase in errors-. Finally, I find that the non-parametric utility maximization model provides more precise

predictions than a model that assumes Cobb-Douglas preferences: the additional precision of the latter model is outweighed by the larger error process needed to fit the data.

The proposed predictive precision measures are shown to be related to the extent of the identification of the systematic component of the process that generated the data. Therefore these measures can be "inverted" to optimally design an experiment to test for a given set of candidate models. An avenue for future research is to produce an efficient way to conduct experimental design based on these measures, which can be of greatly interest for the experimental field. Concretely this project requires: (i) providing a general, computationally feasible, algorithm for the construction of the predictive distribution; (ii) streamlining the connection between power and predictive precision for general context and providing a feasible algorithm for experimental selection and (iii) studying the finite samples and asymptotic properties of these measures for general models of behavior.

The proposed approach to assess the performance of a model is a broad approach that can be applied to a myriad of behavioral models. However, successfully applying this approach to other models, that may deliver empirical implications in more general contexts, depends on the extension of the predictive distribution to these contexts. In general, the predictive distribution will reflect the two types of uncertainty described above. However, the nature of the error process may differ for different economic environments. For example, the error process should exhibit a different structure when dealing with preferences over lotteries, menus, sets of discrete alternatives or preferences over consumption bundles. One could think as defining the error process for discrete choices as the minimal number of swaps required to make consistent the revealed preference information -in the case of utility maximization- inferred from data, as in (Apesteguia and Ballester 2013).

Chapter 2 develops necessary and sufficient conditions for *stochastic* choice data to be consistent with satisficing, assuming that preferences are fixed, but search order may change randomly. The model predicts that stochastic choice can only occur amongst elements that are always chosen, while all other choices must be consistent with standard utility maximization. Adding the assumption that the probability distribution over search orders is the same for all choice sets makes the satisficing model a subset of the class of random utility models.

Bibliography

- ADAMS, A. (2013): “Globally Rational Nonparametric Demand Forecasts,” .
- AFRIAT, S. N. (1967): “The Construction of Utility Functions from Expenditure Data,” International Economic Review, 8(1), pp. 67–77.
- AKAIKE, H. (1998): “Information theory and an extension of the maximum likelihood principle,” in Selected Papers of Hirotugu Akaike, pp. 199–213. Springer.
- AKERLOF, G. A. (1970): “The Market for "Lemons": Quality Uncertainty and the Market Mechanism,” The Quarterly Journal of Economics, 84(3), 488–500.
- ANDREONI, JAMES, G. B., AND W. HARBAUGH (2013): “The power of revealed preference tests: ex-post evaluation of experimental design,” Working paper.
- ANDREONI, J., AND J. MILLER (2002): “Giving according to GARP: An experimental test of the consistency of preferences for altruism,” Econometrica, 70(2), 737–753.
- ANDREWS, D. W., AND P. GUGGENBERGER (2009): “Validity of Subsampling and "Plu-in Asymptotic" inference for parameters defined by moment inequalities,” Econometric Theory, 25, 669–709.
- APESTEGUIA, J., AND M. A. BALLESTER (2013): “A measure of rationality and welfare,” Working Papers (Universitat Pompeu Fabra. Departamento de Economía y Empresa).
- BEATTY, T. K., AND I. A. CRAWFORD (2011): “How demanding is the revealed preference approach to demand?,” The American Economic Review, 101(6), 2782–2795.

- BECKER, G. S. (1962): "Irrational behavior and economic theory," The Journal of Political Economy, 70(1), 1–13.
- BLACKWELL, D., ET AL. (1951): "Comparison of experiments," in Proceedings of the second Berkeley symposium on mathematical statistics and probability, vol. 1, pp. 93–102.
- BLOCK, H. D., AND J. MARSCHAK (1960): Contributions to Probability and Statistics chap. Random Orderings and Stochastic Theories of Response, in (Olkin 1960).
- BLUNDELL, R., M. BROWNING, AND I. CRAWFORD (2008): "Best Nonparametric Bounds on Demand Responses," Econometrica, 76(6), 1227–1262.
- BLUNDELL, R., D. KRISTENSEN, AND R. MATZKIN (2014): "Bounding quantile demand functions using revealed preference inequalities," Journal of Econometrics, 179(2), 112–127.
- BLUNDELL, R. W., M. BROWNING, AND I. A. CRAWFORD (2003): "Nonparametric Engel Curves and Revealed Preference," Econometrica, 71(1), pp. 205–240.
- BRONARS, S. G. (1987): "The power of nonparametric tests of preference maximization," Econometrica: Journal of the Econometric Society, pp. 693–698.
- CABRALES, A., O. GOSSNER, AND R. SERRANO (2010): "Entropy and the value of information for investors," Available at SSRN 1720963.
- CAPLIN, A., AND M. DEAN (2011): "Search, choice, and revealed preference," Theoretical Economics, 6(1), 19–48.
- CAPLIN, A., M. DEAN, AND D. MARTIN (2011): "Search and satisficing," The American Economic Review, pp. 2899–2922.
- CAPLIN, A., AND A. SCHOTTER (2010): The foundations of positive and normative economics: a handbook. Oxford University Press.

- CHERCHYE, L., I. CRAWFORD, B. DE ROCK, AND F. VERMEULEN (2009): “The revealed preference approach to demand,” Contributions to Economic Analysis, 288, 247–279.
- CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2007): “Estimation and confidence regions for parameter sets in econometric models¹,” Econometrica, 75(5), 1243–1284.
- CHOI, S., R. FISMAN, D. GALE, AND S. KARIV (2007): “Consistency and Heterogeneity of Individual Behavior under Uncertainty,” American Economic Review, 97(5), 1921–1938.
- CRAWFORD, I., AND B. DE ROCK (2014): “Empirical Revealed Preference,” Annual Review of Economics, 6(1), 503–524.
- DE CLIPPEL, G., AND K. ROZEN (2014): “Bounded rationality and limited datasets,” .
- DEAN, M., AND D. MARTIN (2012): “A Comment on How Demanding Is the Revealed Preference Approach to Demand?,” Discussion paper, Working Paper.
- (2013): “Measuring rationality with the minimum cost of revealed preference violations,” Working paper.
- DEKEL, E., AND B. L. LIPMAN (2010): “How (Not) to Do Decision Theory,” Annual Review of Economics, 2(1), 257–282.
- DIEWERT, W. E. (2012): “Afriat’s Theorem and some Extensions to Choice under Uncertainty*,” The Economic Journal, 122(560), 305–331.
- ECHENIQUE, F., S. LEE, AND M. SHUM (2011): “The money pump as a measure of revealed preference violations,” Journal of Political Economy, 119(6), 1201–1223.
- ELIAZ, K., AND R. SPIEGLER (2011): “Consideration sets and competitive marketing,” The Review of Economic Studies, 78(1), 235–262.

- FALMAGNE, J. C. (1978): "A representation theorem for finite random scale systems," Journal of Mathematical Psychology, 18(1), 52–72.
- FAMULARI, M. (1995): "A household-based, nonparametric test of demand theory," The Review of Economics and Statistics, pp. 372–382.
- FLEISSIG, A. R., AND G. A. WHITNEY (2005): "Testing for the Significance of Violations of Afriat's Inequalities," Journal of Business & Economic Statistics, 23(3), 355–362.
- GABAIX, X., AND D. LAIBSON (2008): "The seven properties of good models," The foundations of positive and normative economics. Oxford University Press, Oxford, pp. 292–299.
- GEISSER, S., AND W. F. EDDY (1979): "A Predictive Approach to Model Selection," Journal of the American Statistical Association, 74(365), 153–160.
- GIBBARD, A., AND H. R. VARIAN (1978): "Economic models," The Journal of Philosophy, pp. 664–677.
- GUL, F., P. NATENZON, AND W. PESENDORFER (2014): "Random Choice as Behavioral Optimization," Econometrica, 82(5), 1873–1912.
- GUL, F., AND W. PESENDORFER (2006): "Random Expected Utility," Econometrica, 74(1), 121–146.
- GUL, F., W. PESENDORFER, ET AL. (2008): "The case for mindless economics," The foundations of positive and normative economics, 3, 39.
- HALEVY, Y., D. PERSITZ, AND L. ZRILL (2014): "Parametric Recoverability of Preferences," Discussion paper, Microeconomics. ca Website.
- HARBAUGH, W. T., K. KRAUSE, AND T. R. BERRY (2001): "GARP for kids: On the development of rational choice behavior," American Economic Review, pp. 1539–1545.

- HJERTSTRAND, P. (2013): "A Simple Method to Account for Measurement Errors in Revealed Preference Tests," Working Paper Series 990, Research Institute of Industrial Economics.
- HJERTSTRAND, P., AND J. SWOFFORD (2009): "A stochastic test of the number of revealed preference violations," Working paper.
- HJERTSTRAND, P., AND J. SWOFFORD (2014): "Are the choices of people stochastically rational? A stochastic test of the number of revealed preference violations," Empirical Economics, 46(4), 1495–1519.
- HODERLEIN, S., AND J. STOYE (2009): "Revealed Preferences in a Heterogeneous Population," Review of Economics and Statistics.
- HOETING, J. A., D. MADIGAN, A. E. RAFTERY, AND C. T. VOLINSKY (1999): "Bayesian Model Averaging: A Tutorial," Statistical Science, 14(4), pp. 382–401.
- HOUTHAKKER, H. S. (1950): "Revealed Preference and the Utility Function," Economica, 17(66), pp. 159–174.
- HOUTMAN, M., AND J. MAKS (1985): "Determining all maximal data subsets consistent with revealed preference," Kwantitatieve methoden, 19, 89–104.
- KASS, R. E., AND A. E. RAFTERY (1995): "Bayes factors," Journal of the American statistical association, 90(430), 773–795.
- KOCOSKA, L. (2012): "A non-parametric approach to the estimations of critical inputs to economic models based on consumption data," Ph.D. thesis.
- LEWBEL, A. (2001): "Demand Systems with and without Errors," American Economic Review, pp. 611–618.
- LUCE, R., AND P. SUPPES (1965): Handbook of Mathematical Psychology, Vol. IIIchap. Preference, Utility, and Subjective Probability, in (R. D. Luce and Galanter 1965).

- MÄKI, U. (2005): “Models are experiments, experiments are models,” Journal of Economic Methodology, 12(2), 303–315.
- MANZINI, P., AND M. MARIOTTI (2014): “Stochastic Choice and Consideration Sets,” Econometrica, 82(3), 1153–1176.
- MAS-COLELL, A. (1978): “On revealed preference analysis,” The Review of Economic Studies, pp. 121–131.
- MASATLIOGLU, Y., D. NAKAJIMA, AND E. Y. OZBAY (2012): “Revealed attention,” The American Economic Review, 102(5), 2183–2205.
- MCCLELLON, M. (2015): “Unique Random Utility Representations,” .
- MOLCHANOV, I. (2005): Theory of Random Sets. Springer, London.
- OLKIN, I. (ed.) (1960): Contributions to Probability and Statistics. Stanford University Press, Stanford.
- R. D. LUCE, R. R. B., AND E. GALANTER (eds.) (1965): Handbook of Mathematical Psychology, Vol. III. Wiley, New York.
- REUTSKAJA, E., R. NAGEL, C. F. CAMERER, AND A. RANGEL (2011): “Search dynamics in consumer choice under time pressure: An eye-tracking study,” The American Economic Review, pp. 900–926.
- RICHTER, M. K. (1996): “Revealed preference theory,” Econometrica: Journal of the Econometric Society, pp. 635–645.
- RUBINSTEIN, A., AND Y. SALANT (2006): “A model of choice from lists,” Theoretical Economics, 1(1), 3–17.
- SALANT, Y., AND A. RUBINSTEIN (2008): “(A, f): Choice with Frames,” The Review of Economic Studies, 75(4), 1287–1296.

- SAMUELSON, P. A. (1938): "A Note on the Pure Theory of Consumer's Behaviour," Economica, 5(17), pp. 61–71.
- SAMUELSON, P. A. (1948): "Consumption Theory in Terms of Revealed Preference," Economica, 15(60), pp. 243–253.
- SANTOS, B. D. L., A. HORTACSU, AND M. R. WILDENBEEST (2012): "Testing Models of Consumer Search Using Data on Web Browsing and Purchasing Behavior," American Economic Review, 102(6), 2955–80.
- SCHWARTZ, B., A. WARD, J. MONTEROSSO, S. LYUBOMIRSKY, K. WHITE, AND D. R. LEHMAN (2002): "Maximizing versus satisficing: happiness is a matter of choice.," Journal of personality and social psychology, 83(5), 1178.
- SELTEN, R. (1991): "Properties of a measure of predictive success," Mathematical Social Sciences, 21(2), 153–167.
- SIMON, H. A. (1955): "A behavioral model of rational choice," The quarterly journal of economics, pp. 99–118.
- SIPPEL, R. (1997): "An experiment on the pure theory of consumer's behaviour*," The Economic Journal, 107(444), 1431–1444.
- SPENCE, M. (1973): "Job market signaling," The quarterly journal of Economics, pp. 355–374.
- SUGDEN, R. (2000): "Credible worlds: the status of theoretical models in economics," Journal of Economic Methodology, 7(1), 1–31.
- (2009): "Credible worlds, capacities and mechanisms," Erkenntnis, 70(1), 3–27.
- VARIAN, H. R. (1982): "The nonparametric approach to demand analysis," Econometrica: Journal of the Econometric Society, pp. 945–973.

- (1985): “Non-parametric analysis of optimizing behavior with measurement error,” Journal of Econometrics, 30(1), 445–458.
- (1990): “Goodness-of-fit in optimizing models,” Journal of Econometrics, 46(1), 125–140.
- (2006): “Revealed preference,” Samuelsonian economics and the twenty-first century, pp. 99–115.
- WARD, D. (1992): “The role of satisficing in foraging theory,” Oikos, pp. 312–317.