Essays in Behavioural Economics using Revealed Preference Theory and Decision Theory

Gavin Kader

University College London (UCL)

Thesis submitted in partial fulfilment of the degree of

Doctor of Philosophy (PhD) in Economics

London, May 2019

Declaration

I, Gavin Kader, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

London, May 2019

Abstract

This thesis comprises three chapters that provide insights into consumer rationality and consideration sets using revealed preference theory, and a decision theoretic approach to *status quo bias*.

Chapter 1 studies the presence of consideration sets through the lens of economic rationality from the perspective of revealed preference theory. In addition, I propose a new index of rationality (GAV Index) to accompany two commonly used measures (CCEI & MPI), which are applied to a scanner panel dataset and a simulated dataset. Under minimal restrictions, I detect the effects of exogenous consideration set formation on a household's ability to make rational bundle choices. There are also several key demographic factors that correlate well with rationality. This remains true when controlling for the (average) size of the consideration sets households use; these results suggest that a simpler decision-making process with fewer goods can lead to choices that are more rational. Overall, the use of consideration sets as a behavioural heuristic can seemingly benefit consumers by enhancing their decision-making process.

Chapter 2 semi-parametrically estimates costs associated with consideration sets using revealed preference theory. The theorem provided ensures there are testable implications of a parsimonious model of consideration sets. Cost of consideration can be estimated in proportion to expenditure and is heterogeneous across consumers. Using the Stanford Basket Dataset, the model cannot reject the use of consideration sets in the presence of suitable restrictions. On average,

the average consideration set cost is approximately 2% of monthly expenditure. Additionally, there appears to be a strong link between the consumer's cost of consideration and rationality level.

Chapter 3 proposes a choice theory that explains status quo bias (SQB) with the concept of just-noticeable differences (JNDs). SQB comes from an inclination to choose a default option/current choice when decision-making, whereas a JND is the minimal stimulus required to perceive change. JND utility can be considered a general representation of SQB; it is shown that the SQB representation of Masatlioglu and Ok (2005) is a special case. As such, an agent will only move away from a current choice position if there exist other alternatives that are *noticeably* better, otherwise, the agent does not shift away, hence leading to a bias towards the status quo.[1]

[1] Additionally, I show that it is possible to aggregate JND preferences over a finite number of characteristics in ways that are consistent with the final choice of good/s (with or without SQB).

Impact Statement

In terms of the academic impact, all of the essays provide new perspectives and techniques that allow economists to relax many assumptions that are considered standard. Chapters 1 and 2 have a strong focus on consumer rationality and consideration sets. Chapter 3 has direct bearing on the way in which status quo bias is normally thought of in the economics literature.

Chapter 1 provides a new measure of rationality which is easily implementable and interpretable. Chapters 1 and 2 incorporate the notion of consideration sets into the standard revealed preference framework, in which there has been a lack of research. These chapters directly impact the way economists think of rationality because i) size of consideration sets are strongly correlated with rationality ii) a (fixed) cost of consideration can provide a healthy and natural explanation for a lot of what is measured as irrationality.

An exciting sub-field of economics is *behavioural decision theory*. It plays a massive role in the economics profession; to make behavioural economics rigorous in a way that is deep-rooted in economic theory. This is exactly what Chapter 3 does vis-à-vis status quo bias. Chapter 3 provides an alternative and parsimonious way of describing behaviour that is consistent with status quo bias through an axiomatisation.

The cornerstone of my academic career has been to help economists understand consumer behaviour so that we can be better informed and, as such, be more informative. Chapters 1 and 2 highlight an important issue that consumers may not necessarily evaluate every single product they have ever come across. However, it may be welfare improving if i) consumers are more aware of other products, if they are very limited in their choices, or ii) consumers concentrate on a finite set of choices to avoid obfuscation. These chapters show that there is a balance between quality of decision-making and quantity of available choices. This has many beneficial implications for the consumers themselves (commonly referred to as *'nudges'*), marketing departments/firms, and delivery of public services.[2]

Chapter 3 highlights an issue that is far-reaching in both academic and non-academic settings. People are not always able to make perfect comparisons. The decision-making process can be tricky, especially in settings that involve complex processes. Many individuals rely on heuristics or fall back on their status quo. Unless there is a choice that is substantively better than their current decision, people may decide not to switch away. From a welfare stance, it is essential to try to understand why individuals have this switching cost and if there are ways of *'nudging'* people to make objectively better decisions. This chapter analyses one avenue of this in terms of status quo bias.[3]

All of these chapters are self-contained papers and have been presented at various academic conferences and seminars. My ultimate goal for these is to improve and refine them in order to achieve publications in top economics journals.

[2]People can often feel overwhelmed or confused when having to make decisions that are not commonplace, or are difficult e.g. financial decisions, medical health etc...

[3]A common example that often arises is the choice of pensions. Individuals seem to stick with their default pension scheme despite there existing a plethora of information and opportunity to change.

Acknowledgements

Firstly, I would like to extend my gratitude for the funding I have received from numerous sources, in particular, the Economics department at UCL for the WM Gorman Graduate Research Scholarship & numerous tuition fee waivers, and financial support via ECR Grant 509157. The opportunity to work as a PGTA at UCL also served as a reliable financial supplement.

Without Richard Blundell, Syngjoo Choi, and Ran Spiegler, the completion of my PhD would not have been possible. I would like to thank them personally for useful and constructive discussions of ideas, as well as for exemplary supervision and guidance. I have also benefitted hugely from talking with and presenting to Ryan Kendall and Terri Kneeland. All of the aforementioned have helped me navigate through the tumultuousness that is the process of obtaining a PhD, and for this, I cannot extend my gratitude enough. My warmest and deepest thanks to all of you.

Over the years, I have had the privilege of discussing my ideas with various people on the PhD. Although there are many, I would like to make special mention of one particular individual, my friend and colleague, Fernanda Senra de Moura. Without her, our countless lengthy discussions, and her insightful inputs, my job market paper would simply not exist.[4] Thank you, Fernanda.

For a more personal touch, to everyone and anyone who has ever spent any time with me in Drayton G06, all of you have made this arduous and complex process entirely more bearable and manageable.[5] There are far too many of you to mention by name, but I will say: *craziness was kept at bay....with inordinate doses of craziness.* Thanks, everyone!

Last, and most definitely not least, I wish to express an immeasurable indebtedness to my parents to whom I owe lifetimes of gratitude and appreciation. Without them, their love and devotion, none of this would have been possible. My success is your success. Allow me the honour and privilege of dedicating this thesis to you.

[4] Chapter 3 of this thesis

[5] or the kitchen, or the office hour rooms, depending on my mood of the day...

Contents

Preface **2**

1 Consideration Sets and Rationality: Is Revealed Preference Theory Revealing Enough? **5**

1.1 Introduction 5
1.2 Related Literature 7
1.3 Model & Theoretical Foundation 10
1.3.1 Definitions 10
1.3.2 Questions & Hypotheses 14
1.3.3 Rationality Indices 15
1.3.3.1 Critical Cost Efficiency Index and Money Pump Index 15
1.3.3.2 Generalised Axioms Violation Index (GAV Index) . 18
1.3.3.3 Statistical Test for GAV Index 19
1.3.3.4 Similarities and Differences between MPI and GAV Index 21
1.3.4 Theoretics of Using MPI and CCEI 22
1.4 Data Description 24
1.4.1 Stanford Basket Dataset 24
1.5 Results 26
1.5.1 Rationality Indices 26
1.5.2 Demographics 29
1.6 Implications 31

1.6.1 Application & Policy Implications 31
1.7 Extensions & Limitations 32
1.8 Concluding remarks 34

2 Cost of Consideration and Revealed Preference 36
2.1 Introduction 36
2.2 Literature Review 38
2.2.1 Consideration Sets 38
2.2.2 Revealed Preferences 40
2.3 Theoretical Framework 42
2.3.1 Model 42
2.3.2 Convex Optimisation 49
2.3.3 Estimation 52
2.4 Implementation 54
2.4.1 Results 54
2.4.1.1 Cost of Consideration 54
2.4.1.2 Power and Prediction 55
2.4.1.3 Correlation with CCEI 57
2.4.1.4 Empirical Results 58
2.5 Concluding remarks 59

3 Choices with just-noticeable differences or status quo - Is there a (noticeable) difference? 61
3.1 Introduction 61
3.1.1 Example Representations 65
3.1.1.1 Comparisons with representation from Masatlioglu and Ok (2005) 65
3.1.1.2 Example JND representations of SQB 66
3.2 Model 67
3.2.1 Primer 67
3.2.2 Axioms 69

3.2.3 Main Representation . . . 74
3.2.4 Interval Order Representation . . . 79
3.3 Related Literature . . . 82
3.4 Concluding remarks . . . 83

Conclusion **84**

Appendices

Appendix A Appendix for Chapter 1 **95**
A.1 Tables . . . 95
A.2 Regressions . . . 99
A.2.1 Table 7 . . . 99
A.3 Simulations . . . 100
A.3.1 Simulated Dataset . . . 100
A.3.2 Algorithm . . . 101
A.3.3 Results . . . 104
A.4 Statistical Test for GAV Index . . . 107

Appendix B Appendix for Chapter 2 **109**
B.1 Proof of Theorem 1 . . . 109
B.2 Table 1 . . . 111
B.3 Dimensionality Issues . . . 111
B.3.1 Dimensionality Issue . . . 111

Appendix C Appendix for Chapter 3 **113**
C.1 Proofs . . . 113
C.1.1 Theorem 2 . . . 113
C.1.2 Corollary 2.1 . . . 118
C.1.3 Implication 2.1 . . . 118
C.1.4 Lemma 2.1 . . . 118
C.2 Aggregation . . . 121

Preface

In recent years, there has been a surge in the literature on the incorporation of psychology into economics as a way of relaxing the standard assumptions that are normally associated with *homo economicus*, with its commonplace nomenclature being *behavioural economics.*

It is in this respect, that the following thesis chapters contribute directly to the economics literature. Broadly speaking, the essays incorporate a deeper understanding and appreciation of human decision-making processes, in terms of consumer (ir)rationality, and behavioural biases. Methodologically, they contribute to the use of revealed preference theory and decision theory as a means of describing and predicting behaviour. In particular, the use of revealed preference theory allows for rigorous theory to precede naturally to empirical analyses.

Inspired by the parsimony of revealed preference theory, the first chapter of this thesis introduces an alternative measure of the violations of the generalised axiom of revealed preference (GARP), called the GAV Index. It is based on the concept of the exploitation of irrational decisions via a profiteering arbitrager who wishes to extract consumer surplus by essentially buying rational bundles, and selling back irrational bundles. However, one of the main issues of rationality indices is deciding the scale upon which we measure rationality. In essence, *'How irrational does a decision have to be to consider an individual irrational?'*. Exploiting the fact that there is price mis-measurement, the GAV Index can be manipulated and restructured in a way that it can be used to statistically test whether the hypothesis of consumer rationality can be rejected or not.

The main purpose of chapter 1 is to study the presence of consideration sets using revealed preference theory through the lens of economic rationality. Using three measures of rationality, applied to a scanner panel dataset and a simulated dataset, there is evidence to suggest that a household's decision-making process may be substantially improved by breaking up their bundle choices by different consideration sets, as opposed to just one large consideration set. Whether consumers are able to benefit from making decisions, based on consideration sets, can be detected by using these measures of rationality. Using the Stanford Basket Dataset (and simulations), various analyses throughout points to the importance of incorporating the notion consideration sets in revealed preference arguments. There is also evidence to suggest that certain demographic factors can correlate well with rationality levels, as well as the complexity of decision-making via the (average) number of goods per consideration set. Interestingly, this suggests that larger consideration sets do not necessarily yield better results for the household, suggesting some other form of behavioural effects may be at play, e.g. choice overload, or a cost of consideration. This naturally segues into chapter 2.

Using revealed preference theory, this chapter semi-parametrically estimates the costs associated with consideration sets in the decision making process. The theorem provided ensures there are testable implications of a parsimonious model of consideration sets. The model essentially incorporates a fixed cost of consideration for any alternative/good that is purchased. As per the theorem presented, this can be thought of as a price distortion whereby choices can be rationalised by prices that are higher than the observed ones, up to a certain level. The cost of consideration can be heterogeneously estimated in proportion to expenditure. Using the Stanford Basket Dataset, the model cannot reject the use of consideration sets in the presence of suitable and minimal restrictions. On average, the average consideration set cost is approximately 2% of monthly expenditure. Additionally, there appears to be a strong link between the consumer's cost of consideration and their level of rationality.

For the final chapter, I propose a rational choice theory that explains the well-documented phenomenon of status quo bias (SQB) with the concept of just-noticeable differences (JNDs). From behavioural economics, SQB is thought of as an inclination to choose a default option/current choice, for example, individuals tend to stick with their default pensions schemes, or choose the same set of goods in the supermarkets. However, from economic theory (and econophysics), a JND is the minimal level of stimulus required to be able to perceive change, for example, individuals typically had not noticed the reduction in the weights of their commonly purchased foods (e.g. chocolate bars etc...) until there was a significant difference. I show that choice behaviour that is consistent with SQB can also be represented by JND utility. The key notion behind this is that an individual does not move away from their current choice, unless there are other options that are noticeably better. Given that JND utility can yield a potential explanation for SQB.[6]

Following the last essay, there is a remaining section that concludes in terms of summarising the chapters and outlining a route for future research in behavioural economics, both theoretically and empirically. For practicality, the majority of the accompanying tables and figures are resigned to the appendices, alongside the lengthier proofs and derivations.[7]

[6]I also show that it is possible to aggregate JND preferences over characteristics of goods. This is to say that, if goods are thought of as a finite collection of characteristics, this chapter also shows that there is a consistent way to aggregate these preferences over these characteristics to the standard case or the JND utility case.

[7]There may be overlaps in the literature reviews in the respective chapters. They are left as intended as the same papers offer various insights for different chapters. Where there are any other overlaps, this is to ensure that each chapter is also self-contained, allowing the reader to concentrate on specific chapters, if they so wish.

Chapter 1

Consideration Sets and Rationality: Is Revealed Preference Theory Revealing Enough?

1.1 Introduction

[1] Consumer rationality is one of the fundamental assumptions of standard economic theory. The extent to which this can be measured is subjective, in that, there are infinitely many ways to be irrational, but, by definition, only one benchmark for being truly rational. Through revealed preference theory, it is possible to rank bundles through a chain of choices, as described by weak inequality constraints, even if the bundles are not directly compared. If the inequality constraints fail to hold mutually, the data are not consistent with revealed preference theory.

However, there is very little in the current literature that combines the notion of consideration sets with rationality despite their natural links. Commonly seen in the marketing literature, initially proposed by Wright and Barbour (1977), the

[1]For this chapter, I would like to express immense gratitude to Syngjoo Choi for his continuing support and guidance, during and after his tenure at UCL.

consideration set is seen to be a subset of the total number of alternatives with which the consumer makes a choice. As per the description in Horowitz and Louviere (1995), the set of alternatives that the consumer actually uses to make their decision *"need not coincide with the set of all possible alternatives"*. This leads to the natural combination of revealed preference theory and consideration sets, as both concepts deal with consumer choice through the principle of a natural ranking.

Specifically, linking consideration sets and rationality offers a potentially exciting insight into bounded rationality. A reason why economists might observe behaviour consistent with bounded rationality is that there is the notion of a cost that is associated with gaining more information that could aid the consumer in their decision-making process. The role of consideration sets supports this idea as, for whatever reason, the consumer only uses a subset of the total set of alternatives, as the cost of expanding the consideration set is potentially too high. In some sense, the consideration set provides additional restrictions on consumers as it also deals with the subset of alternatives. Revealed preference inequalities would then allow a ranking and testing of rationality on the basis of using only the consideration set. Combining these concepts can provide a more reasonable model of consumer choice and behaviour in general. Chapter 1 will focus on the perspective of rationality and consideration sets, with Chapter 2 concentrating on the cost of consideration.

This chapter attempts to explore the importance of taking into account consideration sets when doing any analysis involving revealed preference theory through commonly used rationality indices. Sections 1.2 and 1.3 explain the relevant topics related to revealed preference theory that will be used throughout and how these tie-in with consideration sets. Section 1.3 goes into further critical analysis and theoretical details of the rationality indices used via consideration sets, as well as providing hypotheses and predictions. Section 1.4 describes the empirical dataset used in this chapter. Sections 1.5 and 1.6 explain the results of the model (including regressions based on demographics), potential applications and implications for policy, and precautions that arise from combining revealed preference and consideration sets. Section 1.7 describes the potential limitations and extensions of the model, with Section 1.8 concluding.[2]

1.2 Related Literature

As mentioned previously, there is an absence of research in the literature linking rationality and consideration sets. The 4 main papers in the literature that go into detail on topics related to consideration sets and their applications to economic theory and revealed preference are Manzini & Mariotti (2014), Masatlioglu, Nakajima & Ozbay (2012), Spiegler & Eliaz (2011), and Demuynck and Seel (2018). Additionally, on the more econometric side of consideration sets, such as Chiang, Chib & Narasimhan (1999), Horowitz & Louviere (1995) etc..., it is widely modelled such that that the consideration set formation in decision-making is part of a dual-staged process, the first stage being the *'consideration stage'* and the second stage being the *'choice stage'*.

Spiegler & Eliaz (2011) apply economic theory to the original application of consideration sets, within a framework of marketing and competition. Their idea behind the construction of the consideration set is via the consumer's unawareness of certain products and that a marketing strategy needs to be introduced to consumers, even after they become aware of new products, in order to expand their consideration set. Marketing is then a tool to overcome this information asymmetry. They devise a model that seeks to encapsulate the consideration set as a means for firms to somehow extract surplus from consumers as a result of asymmetric information. Masatlioglu & Nakajima (2012) design a framework that studies a more general framework of consumer behaviour those using consideration sets in their decision-making process. The consideration set is formed as a function of an exogenous and feasible 'starting point' (F). Let $C_1(F(0), Y) \subseteq Y$ where Y represents budget feasibility. The consumer makes choices based on $C(F, Y)$

[2]For further results and robustness, I refer the reader to Appendix A.3 that replicates the analysis done in Section 1.3 with simulated data.

with a binary relation operator on the complete set of alternatives, A. If the consumer gets (at least) their starting point option, the decision-making process is complete, if not, the consumer constructs a 'smaller' consideration set where the new starting point is the previously rejected alternative i.e. $C_2(F(1), Y)$, where $F(1)$ is the alternative that was rejected in the previous period; the process is then iterated. Masatlioglu & Nakajima (2012) also propose a more realistic version of their iterative model in that the initial starting point is inferred from the choice data. Overall, Masatlioglu and Nakajima (2012) provide one of the more general behavioural models with consideration sets.[3]

Manzini & Mariotti (2014) specifically model a boundedly rational consumer whereby agents have a probability distribution over alternatives. They attempt to infer the preference ordering from consumer choices under the notion that the choices were generated by a consideration set where the consideration set is a function of the larger set of alternatives. This is similar to the idea of Spiegler & Eliaz (2011), however, the constructions of the consideration set are different in both cases. Manzini & Mariotti (2014) put forward a model that suggests the formation of consideration sets is random in the sense that all (feasible) alternatives have a probability of being in the consideration set; they call this the 'attention parameter'.[4] Their decision-making procedure involves a decision-maker with a consideration set with a complete preference relation over the consideration set (not just over the complete set of alternatives).

Masatlioglu, Nakajima & Ozbay (2012) is a highly related paper to Manzini & Mariotti (2014) as both papers devise models that approach choice data and consideration sets from limited attentions. The main idea behind Masatlioglu, Nakajima & Ozbay (2012) is that, as a result of limited attention, there is some form of 'filtering' of alternatives that occurs . A common property of the consideration set is that an alternative that is not within the consideration set cannot affect the consideration set even after it becomes unavailable (or infeasible). Under

[3] Spiegler & Eliaz (2011) can be seen as a special case of Masatlioglu & Nakajima (2012) where the initial marketing device and initial alternative form the basis of the first consideration set. The next stages of consideration set formation are then as a result of marketing.

[4] In the limited attention sense.

this property, they are able to establish how this 'filtering' of alternatives occur through the consumer's choices. Their interpretation is that revealed preference theory may only be a specific example of decision-making under consideration sets. Similar in notion is the paper by Chiang, Chib & Narasimhan (1999) that proposes a basic consideration set choice model that attempts to account for heterogeneity in the construction of consideration sets through random effects. They find that heterogeneity is an important factor (if not accounted for) as it can lead to an over-reliance on pure preference relations and can undervalue the the impact of marketing on consideration set construction. By using scanner data, they estimate and compare results by including/excluding different brands of ketchup to derive heterogeneity factors by iterating over all possible combinations of consideration set based on ketchup brands.

Demuynck and Seel (2018) depart from the economic literature in that the analysis is done from a revealed preference point of view; the most basic difference being that choices á la Afriat are continuous (over a set of discrete goods). They provide a new axiom of revealed preference, namely the Limited Axiom of Revealed Preference (LARP). LARP involves verifying that GARP holds within partitions of a dataset admitted by the same consideration sets. This allows an extension to Afriat's Theorem under limited consideration where goods that are not admitted by a particular consideration set are incorporated via unobserved subjective prices. Consideration set formation is then modelled via the beliefs put on those subjective prices. Applied to a scanner dataset from the Denver area (USA), they largely fail to reject the use of consideration sets in the consumers decision-making process. To the best of my knowledge, this is the only paper that incorporates standard revealed preference analysis with consideration sets.

Horowitz & Louviere (1995) investigate further the notion of a multi-stage process in decision-making. Their particular paper attempts to test the basic hypothesis of whether utility derived from alternatives within the consideration set is higher than utility derived from any alternative not within the consideration set. In this sense, preferences do not depend on the consideration set, rather, the alternatives not in the consideration set yield lower utility. When estimating a utility function, they propose that knowledge of a consumer's consideration set can improve the precision with with a utility function is measured.[5]

As from above, it is clear that there is a rich understanding of consideration sets in the marketing literature and, more recently, a resurgence of research in the fields of econometrics and economic theory. The consideration set, in essence, is an unobservable, which makes it an relevant topic for econometricians to study, whether it be parametric or non-parametric in nature. A more 'micro-founded' and 'psychologically-based' approach can be found in Gabaix (2014). Gabaix introduces an elegant way of modifying basic microeconomic theory into that of a sparsity-based model. The sparsity (based on limited attention) arises from the consumer having to ignore many factors or alternatives in their decision-making process. In effect, the cost of attention is incorporated into the maximisation process, allowing for the manipulation of the standard textbook microeconomic models. This type of sparsity-based model encompasses many classes of behavioural models, particularly related to bounded rationality.

Andreoni et. al (2011) provide much insight and guidance into the use of rationality indices, their corresponding power, and their appropriate interpretation when it comes to using revealed preference theory in empirical works.

1.3 Model & Theoretical Foundation

1.3.1 Definitions

This subsection familiarises the reader with the conventions of revealed preference used throughout the thesis.

Let $\mathbf{T}$ denote time periods such that $\mathbf{T} = \{1, 2, ..., T\}$[6]

Let $\mathbf{p}_t \in \mathbb{R}^n_{++}$ denote prices in period $t \in \mathbf{T}$

Let $\mathbf{q}_t \in \mathbb{R}^n_{+}$ denote quantities in period $t \in \mathbf{T}$

[5] If the utility function is fully observable, information of the consideration set does not aid in modelling choice.

[6] Also commonly referred to as observations.

where n denotes the *total* number of available goods.

Define a finite dataset, $\mathbf{D}$, as a collection of all prices and quantities i.e. $\mathbf{D} = \{\mathbf{p}_t, \mathbf{q}_t\}_{t\in\mathbf{T}}$. This dataset is a collection of observed consumption behaviour, $\mathbf{q}_t$, for a consumer facing prices, $\mathbf{p}_t$, at observation, t.

A dataset $\mathbf{D} = \{\mathbf{p}_t, \mathbf{q}_t\}_{t\in\mathbf{T}}$ is rationalisable if there exists a utility function $u : \mathbb{R}^n \to \mathbb{R}$ and for all observations $t \in \mathbf{T}$, there exists a weakly positive income level y_t such that:

$$\mathbf{q}_t \in \underset{\mathbf{q}}{\operatorname{argmax}}\, u(\mathbf{q}) \text{ subject to } \mathbf{p}_t'\mathbf{q} \leq y_t$$

This says that a dataset is rationalisable if the bundles, $\{\mathbf{q}_t\}_{t\in\mathbf{T}}$, are consistent with utility maximisation with a linear budget constraint.

A bundle $\mathbf{q}_i$ is directly revealed preferred to $\mathbf{q}_j$ if $\mathbf{p}_i'\mathbf{q}_i \geq \mathbf{p}_i'\mathbf{q}_j$. In words, the bundle $\mathbf{q}_i$ was at least as affordable as bundle $\mathbf{q}_j$ at observation i. Let R denote the *directly revealed preferred* binary relation i.e. $\mathbf{q}_i$ is directly revealed preferred to $\mathbf{q}_j$ if $\mathbf{q}_i R \mathbf{q}_j$.

A bundle $\mathbf{q}_i$ is strictly directly revealed preferred to $\mathbf{q}_j$ if $\mathbf{p}_i'\mathbf{q}_i > \mathbf{p}_i'\mathbf{q}_j$. In words, the bundle $\mathbf{q}_i$ was costlier than bundle $\mathbf{q}_j$ at observation i. Let R_S denote the *strictly directly revealed preferred* binary relation i.e. $\mathbf{q}_i$ is strictly directly revealed preferred to $\mathbf{q}_j$ if $\mathbf{q}_i R_S \mathbf{q}_j$.

A bundle $\mathbf{q}_i$ is indirectly revealed preferred to $\mathbf{q}_j$ if there exists a sequence of observations $x, y, ..., z$ in $\mathbf{T}$ such that $\mathbf{q}_i R \mathbf{q}_x, \mathbf{q}_x R \mathbf{q}_y, ..., \mathbf{q}_z R \mathbf{q}_j$. Let P denote the *indirectly revealed preferred* binary relation i.e. $\mathbf{q}_i$ is indirectly revealed preferred to $\mathbf{q}_j$ if $\mathbf{q}_i P \mathbf{q}_j$.[7] The number of bundles in a chain of directly revealed preferred bundles is called the *sequence length*.[8]

The Generalised Axiom of Revealed Preference (GARP) is satisfied by $\mathbf{D} = \{\mathbf{p}_t, \mathbf{q}_t\}_{t\in\mathbf{T}}$, if, for all $\mathbf{q}_i P \mathbf{q}_j$, then it cannot be that $\mathbf{q_j} R_S \mathbf{q_i}$; if $\mathbf{q}_i$ is indirectly preferred to $\mathbf{q}_j$, it cannot be the case that $\mathbf{q}_j$ was purchased even when $\mathbf{q}_i$ is cheaper.

[7] The indirectly revealed preferred binary relation is the transitive closure of R. It is the binary relation that is transitive and minimal with respect to the set it is on.

[8] For example, $\mathbf{q}_a R \mathbf{q}_b, \mathbf{q}_b R \mathbf{q}_c, \mathbf{q}_c R \mathbf{q}_d$ has a sequence length of 4.

The seminal contribution of Afriat (1967) showed that GARP is both necessary and sufficient for a dataset to be rationalisable.[9] Given a dataset $\mathbf{D} = \{\mathbf{p}_t, \mathbf{q}_t\}_{t\in\mathbf{T}}$ the following statements are equivalent:

(i) $\mathbf{D}$ is rationalisable by a locally non-satiated utility function

(ii) $\mathbf{D}$ satisfies GARP

(iii) For all observations $t \in \mathbf{T}$, there exists $u_t, \lambda \in \mathbb{R}$ and $u_t, \lambda_t \in \mathbb{R}_{++}$ such that for all pairs of observations $i, j \in \mathbf{T}$

$$u_i - u_j \leq \lambda_j \mathbf{p_j}'(\mathbf{q_i} - \mathbf{q_j})$$

(iv) $\mathbf{D}$ is rationalisable by a strictly monotone and concave utility function.

where the inequalities of (iii) are called the *Afriat inequalities.* A neat interpretation of these come from the first order conditions of a constrained maximisation problem using KKT conditions, where λ_t is the Lagrange multiplier for the budget constraint at observation t (assuming differentiability and concavity).

Following Demuynck and Seel (2018), suppose a consumer has access to a total of n goods i.e. she chooses a consumption bundle $\mathbf{q}_t$ from a set of goods $G = \{1, ..., n\}$. If the consumer does not necessarily take into account all n goods when purchasing, the consumer is said to have used a consideration set $I_t \subseteq G$. This means that consumption of any goods that lie outside of her consideration set must be equal to zero (without excluding the option that goods within the consideration set are also zero).

This leads to a natural definition of rationalisability with a consideration set. A dataset $\mathbf{D} = \{\mathbf{p}_t, \mathbf{q}_t\}_{t\in\mathbf{T}}$ is rationalisable with consideration set if there exists a utility function $u : \mathbb{R}^n \to \mathbb{R}$ and for all observations $t \in \mathbf{T}$, there exists a weakly

[9] See Afriat (1967), Diewert (1973), Varian (1982b), Fostel et al. (2003) for detailed proofs of Afriat's Theorem.

positive income level y_t and consideration set $I_t \subseteq G$ such that:

$$\mathbf{q}_t \in \underset{\mathbf{q}}{\operatorname{argmax}}\, u(\mathbf{q}) \text{ subject to } \mathbf{p}_t'\mathbf{q} \leq y_t,$$
$$q^i = 0 \text{ for all } i \notin I_t$$

This says that a dataset is rationalisable if the bundles, $\{\mathbf{q}_t\}_{t\in\mathbf{T}}$, are consistent with utility maximisation with a linear budget constraint and additional constraints where the consumption of the i^{th} good in $\mathbf{q}_t$ is zero.[10]

Now consider that within the consideration set, there may be some goods that are considered and yet have zero consumption. Denote this set of goods with positive consumption as J_t. By definition, it must be that $J_t \subseteq I_t$. Lemma 1 of Demuynck and Seel (2018) shows that with rationalisability with limited consideration, it is without loss of generality that $I_t = J_T$ can be assumed. The intuition for the proof of this result is that there is an overlap in the additional negativity constraints. The goods outside the consideration set have zero consumption, and the goods in the consideration set but without positive consumption clearly have zero consumption. With this overlap in constraints, it is without loss of generality that only goods with positive consumption were considered; it is as if any good with zero consumption was not considered.

Further to this, it follows that the defined dataset $\mathbf{D} = \{\mathbf{p}_t, \mathbf{q}_t\}_{t\in\mathbf{T}}$ can be partitioned into observations where the consideration sets are the same as defined by exactly the same goods of positive consumption. Two observations x, y are said to be in the same partition if $I_x = I_y$. Let $E_k \subseteq \mathbf{T}$ denote the k^{th} partition of observations for which the consideration sets are the same such that $\bigcup_{k=1}^{K} E_k = \mathbf{T}$. Hence, a dataset can be defined as $\mathbf{D} = \{\{\mathbf{p}_t, \mathbf{q}_t\}_{t\in\mathbf{E_1}}, ..., \{\mathbf{p}_t, \mathbf{q}_t\}_{t\in\mathbf{E_K}}\}$. For the purposes of this chapter, I will be using this definition of consideration sets for the following analyses.[11][12]

[10]The i^{th} component of $\mathbf{q}_t$ is zero.

[11]See Demuynck and Seel (2018) for further details.

[12]For example, suppose there are 4 time periods and 10 goods. If in the first and second time periods, only the first 5 goods were purchased, and in the third and fourth time periods, only the other 5 goods were purchased, then observations from the first two time periods form a separate partition to the last two time periods.

1.3.2 Questions & Hypotheses

The overarching question of this chapter can be broken down as follows:

Q1 How is rationality affected by incorporating the role of consideration sets?

- Under the definition of consideration sets above, it is clear that there is no role for empty bundles i.e. bundles with zero consumption for every possible good. In essence, it is as if all the zero bundles are partitioned into their own consideration set; any revealed preference test would yield full rationality for this partition. This provides motivation for excluding zero bundles from any revealed preference frameworks as they will only essentially improve observed rationality in a trivial way. Given this, it will be interesting to investigate whether rationality appears to improve/worsen when looking at partitions of the dataset as defined by the consideration set as opposed to over the whole dataset.
- Extending the rationale as above, some analysis can be done by comparing rationality across consideration sets. One might hypothesise that decision making is easier when there are fewer goods in a consideration set.[13] As such, it may be that consumers make bundle choices that are more rational when there are not as many goods to consider.

Q2 Does rationality vary with different sequence lengths?

- Conditional on the same size of consideration set (i.e. same number of goods chosen with positive consumption), it is conceivably possible that a consumer may be more susceptible to an irrational bundle choice given a longer sequence length. For example, if a consumer is using a consideration set with five goods, is a consumer that has another consideration set with five different goods, but over more time periods, more likely to make an irrational choice? It may be that the assumption of acyclicity is more likely to fail over longer time periods. For example,

[13]For example, choice overload, status quo bias, etc...

it may be easier not to form a cycle of preferences if we only look at choices over three time periods, as opposed to, say, twelve.

Q3 Does the explanatory power of demographic variables change when controlling for average consideration set size?

- Rationality can be very much considered an inherent characteristic of an individual/household. On an individual level, the background behind the decision-making process includes many other factors such as education, family size etc... It may be the case that having fewer goods to consider makes it easier for agents to make rational decisions, as such, the average consideration set size may have a positive influence on the measures of rationality. Similarly, given that the dataset is partitioned over observations, it may be possible that the average sequence length per consideration set may also positively effect rationality.

1.3.3 Rationality Indices

1.3.3.1 Critical Cost Efficiency Index and Money Pump Index

Using the definitions of GARP and a sequence length, I can formally define the MPI.

Definition: *Money Pump Index (MPI)*

Given a GARP-satisfying sequence of length T, the associated *Money Pump Cost* (denoted by M_T) is:

$$M_T = \sum_{i=1}^{T} \mathbf{p}_i'(\mathbf{q}_i - \mathbf{q}_{i+1}) \tag{1.1}$$

with $\mathbf{q}_{T+1} = \mathbf{q}_1$ for a full cycle of GARP

Standardising the *Money Pump Cost* yields the MPI:

$$MPI_T = \frac{M_T}{\sum\limits_{i=1}^{T} \mathbf{p}_i'\mathbf{q}_i} \tag{1.2}$$

In words, Echenique et. al (2011) describe a story of a 'fictitious arbitrager' who seeks to buy inefficient and consequently irrationally chosen bundles. This arbitrager then sells the efficient and rational equivalents back to the irrational household. The so-called 'profit' the arbitrager makes is given by the *Money Pump Cost* above. This is, in effect, the extracted surplus from households. The MPI then standardises this cost to make it comparable across all income levels. This indicates that a household that scores a 0 MPI is perfectly rational as there is no exploitable surplus; an MPI of nearly 1 suggests nearly all their income is extracted from the arbitrager.[14]

Using the definitions pertaining to revealed preference theory, I can formally define a relaxed version of GARP[15] devised by Varian (1990, 1991) in order to derive the CCEI.

A bundle $\mathbf{q}_i$ is *directly relaxed revealed preferred* to $\mathbf{q}_j$ if $e\mathbf{p}_i'\mathbf{q}_i \geq \mathbf{p}_i'\mathbf{q}_j$, for $e \in [0,1]$. In words, the bundle $\mathbf{q}_i$ chosen at a weakly lower income was at least as affordable as bundle $\mathbf{q}_j$ at observation i. Let R_e^d denote the *directly relaxed revealed preferred* binary relation i.e. $\mathbf{q}_i$ is directly relaxed revealed preferred to $\mathbf{q}_j$ if $\mathbf{q}_i R_e^d \mathbf{q}_j$. Let R_e^w denote the *indirectly relaxed revealed preferred* binary relation.

With the above definitions, we can now define a relaxed version of GARP, namely, Varian's Generalised Axiom of Revealed Preference (VGARP). If $\mathbf{q}_i R_e^w \mathbf{q}_j$, then $e * \mathbf{p}_j * \mathbf{q}_j \geq \mathbf{p}_j * \mathbf{q}_i$, $e \in [0,1]$. By relaxing the budget at time j, what could have originally been a violation of GARP, may not be considered a violation of VGARP given an appropriate suppression of income, by factor $e > 0$.

Given this relaxed version of GARP, it is now possible to define a metric on the revealed preferences which can measure the extent of (ir)rationality. Specifically, we can define the *Critical Cost Efficiency Index (CCEI)* as the largest value of $e \leq 1$ such that no violations of VGARP exist, denoted e^*. Formally, the CCEI is defined as the supremum over all e such that binary preference relations, R_e^d and R_e^w satisfy VGARP; thus the CCEI is exactly e^*. In words, the CCEI measures how

[14]Further details can be found in Echenique et. al (2011).

[15]For simplicity, I denote this as VGARP.

much income has to be relaxed in order to remove any violations of VGARP. In a sense, it is a measure of how much income is being wasted as a result of choosing GARP violating bundles; the amount of adjustment being equal to $1 - e^*$. Hence, $e^* = 1$ reverts back to no violations of GARP, thus, fully rational as there is no need for any amount of income adjustment.[16]

As with all rationality indices, there is bound to be a certain level of overlap in terms of their underlying idea. However, the main reason for focusing on the CCEI and MPI is that their interpretations are very different. As discussed in the previous section, the CCEI is a measure of 'consumer error' in their consumption choice which gives rise to the interpretation of 'wasted income'. By contrast, the MPI has the interpretation of a monetary value that is extracted from the consumer as a result of irrational behaviour. As a consequence, the reactions of these indices to the above questions can be very different; this allows the following analysis and results to be more robust. In a sense, if the indices had similar interpretations, then their effects through the different analyses could be (approximately) complementary. In essence, if different indices with different interpretations can tell the same story, this leads to more robust and thorough analysis than using indices with similar interpretations. Hence, the use of both the CCEI and MPI is key in trying to answer the above questions.[17] In a similar vein to the MPI, I introduce a third index of rationality that could also be used.

[16]In practice, it is generally accepted to choose a critical value $\overline{e}$ above which a decision is considered rational (Varian (1991) suggests 0.95 but respects its arbitrariness). Echenique et. al (2011) devise a more robust way of deriving a critical value through measurement error in prices (due to store coupons) which allow them to statistically test whether the MPI can reject or not reject rational behaviour.

[17]An important point to note is that, for increasing sequence length, the CCEI and MPI are more likely to differ, as pointed out in Echenique et. al (2011), adding to the robustness of the results.

1.3.3.2 Generalised Axioms Violation Index (GAV Index)

Definition: *Generalised Axioms Violation Index (GAV Index)*

The formula below is the GAV Index for sequence length T.

$$GAV_T = \frac{1}{T}\sum_{i=1}^{T} \frac{\mathbf{p}_i * (\mathbf{q}_i - \mathbf{q}_{i+1})}{\mathbf{p}_i'\mathbf{q}_i} \tag{1.3}$$

with $\mathbf{q}_{T+1} = \mathbf{q}_1$.

Similar to the MPI, an interesting interpretation of exploiting the household irrational bundle choice is to think of the fictitious third party as a business. An efficient "money pump" concept loosely translates to making profit from households in a business-like fashion. However, no real distinction is made between each bundle decision; only the total surplus is considered when attempting to profit from irrationality. Essentially, a business could act more efficiently and achieve higher profits if it established and exploited individual surpluses, but also, vitally, relative to each income. This is the key contribution that the GAV Index highlights. To frame it as a hypothetical story, the fictitious third party now acts as a business with a manager and employees. The manager overlooks the entire process and allocates employees to each household decision. Each employee is then responsible for analysing and extracting relative surplus from their allocated household decision. The sum of those surpluses is then a measure of irrationality. This extraction of these surpluses is done on a bundle-by-bundle basis, as opposed to over the whole cycle. A household that scores a 0 is perfectly rational as there is no hypothetical profit to be made; a GAV Index of nearly 1 suggests almost all their income is extracted from the arbitrager.

To derive the GAV Index, it is sufficient to show the proof for 2 bundles, which extends naturally to the T bundles case. As per the figure combined with the description above, δ_1, δ_2 refer to the relative losses of income that are acceptable to the consumer as a result of their irrational choice. This means that $\delta_2 = \mathbf{p}_1'\mathbf{B}_1 - \mathbf{p}_1'\mathbf{B}_2$ and $\delta_1 = \mathbf{p}_2'\mathbf{B}_2 - \mathbf{p}_2'\mathbf{B}_1$. If these are standardised by expenditure in their respective current periods, and then summed, this yields $GAV_2 = \sum_{i=1}^{2} \frac{\mathbf{p}_i * (\mathbf{q}_i - \mathbf{q}_{i+1})}{\mathbf{p}_i * \mathbf{q}_i}$,

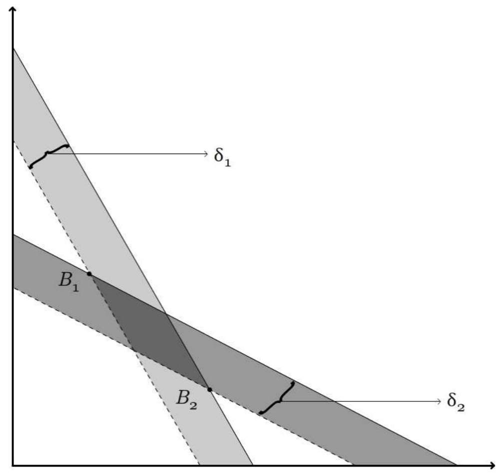

Figure 1.1: Each grey area represents lost income as a result of choosing a GARP violating bundle. The δ's measure the extent to which this income is lost in terms of the area under the budget constraints.

where $\mathbf{q}_3 = \mathbf{q}_1$. The extension to T bundles is then obvious.

1.3.3.3 Statistical Test for GAV Index

As is clear from above, the standard revealed preference analyses are typically binary in their classification for rationality i.e. a consumer is rational if they satisfy GARP, otherwise they are irrational. The literature, including the new measure of rationality in this chapter, addresses this by proposing measures of rationality in order to ascertain how severely a consumer violates GARP. However, there is still difficulty in understanding the magnitudes of any measure of rationality insofar as it is hard to disentangle what is truly the measure of irrationality from statistical noise. I address this by formulating a statistical test for the GAV Index. Following Echenique et al. (2011), it is assumed that there is measurement error (normally distributed) in the observed prices, where the standard deviation of price discounts from coupons is equivalent to the standard error of the mean-zero additive measurement error term, as shown below.[18]

Specifically, the reason for formulating a statistical test for the GAV Index is to establish whether the calculated GAV Index is statistically large enough to

[18]This is approximately 1.11 cents per unit of consumption, see Echenique et al. (2011) for further details.

conclude an irrational decision. The null hypothesis is that of perfect rationality and the following alternative hypothesis being that of irrationality. In order to derive an statistical test for the GAV Index, I use a similar approach as established by Varian (1985) for which there is an assumed measurement error in quantities chosen. Given the data used, measurement error in prices is a much more refined and intuitive way of deriving a statistical test for this context. The statistical test is then applied to each decision per household that incurs a possible GARP violation.

Let $\mathbf{p}_i = \mathbf{K}_i + \mu_i$ where $\mathbf{K}_i$ denotes the true price. Let θ_i denote income shares and assume that the i.i.d $\frac{\mu_i}{\theta_i}$ are normally distributed with mean 0, and variance $\frac{\sigma^2}{\theta_i^2}$. The formula for the variance of the GAV Index with sequence length T is:

$$\sum_{i=1}^{T} \frac{1}{T^2}\sigma^2 ||(\mathbf{q}_i - \mathbf{q}_{i+1})^2|| \frac{1}{\theta_i^2} \tag{1.4}$$

with the derivation resigned to an appendix.[19]

For transparency, the main assumption, beyond the distributional assumptions is that of equal marginal utility of income across observation, for the true prices. In essence, that the marginal change in income results in the same utility change across observation. In the context of grocery shopping, this is far from unreasonable. Recall from above, that a dataset is rationalisable if the bundles chosen are consistent with utility maximisation with a linear budget constraint. This holds true for every time period. Under the null hypothesis of rationality, and Afriat's Theorem, it must be that the Afriat inequalities hold. For simplicity, consider a dataset $\mathbf{D} = \{\mathbf{K}_t, \mathbf{q}_t\}_{t \in \{1,2\}}$. The Afriat inequalities are $u_1 - u_2 \leq \lambda_2 \mathbf{K_2}'(\mathbf{q_1} - \mathbf{q_2})$ and $u_2 - u_1 \leq \lambda_1 \mathbf{K_1}'(\mathbf{q_2} - \mathbf{q_1})$. This implies that $0 \leq \lambda_2 \mathbf{K_2}'(\mathbf{q_1} - \mathbf{q_2}) + \lambda_1 \mathbf{K_1}'(\mathbf{q_2} - \mathbf{q_1})$. By setting $\lambda_1 = \lambda_2$, we get the standard law

[19] As I am mainly concerned with the role of consideration sets and sequence length on rationality, the statistical test is not of huge concern for this particular exercise. This is because the main comparisons will be done across rationality indices, whereas the statistical test is constructed to check whether rationality is violated or not. Hence, as we are concerned with how rationality changes under different decision-making scenarios, the statistical test provides robustness checks for ascertaining rationality. For the CCEI, the standard benchmark of 0.95 is used. The statistical tests are implemented for the different scenarios and do not add (nor take-away) from the ensuing analysis and conclusions.

of demand derived from the Afriat inequalities i.e. $(\mathbf{K_2} - \mathbf{K_1})'(\mathbf{q_1} - \mathbf{q_2}) \leq 0$ with true prices, rather than just the observed prices. Generalised to a larger dataset, these set of equations are crucial in deriving the above formula for the variance of the GAV Index.

1.3.3.4 Similarities and Differences between MPI and GAV Index

As is easily derived, the MPI and GAV Index are only theoretically identical when all the denominators are equal i.e. when $T \cdot \mathbf{p}_1'\mathbf{q}_1 = \sum_{i=1}^{T} \mathbf{p}_i * \mathbf{q}_i = T \cdot \mathbf{p}_2'\mathbf{q}_2 = ... = T \cdot \mathbf{p}_T'\mathbf{q}_T$. This means that, as long as expenditures vary across time periods, then there is no theoretical reason that tie together the MPI and GAV Index. If they are sufficiently close, then these rationality indices will report similar results.

Interestingly, despite the above comparison, the example below shows that the GAV Index and MPI can offer very different conclusions, so, unless expenditure really remains exactly the same across time period, then it is possible for these indices to give opposing results. In a sense, the GAV Index allows for the absolute maximum surplus to be extracted per observation, whereas the MPI seeks to do so over the entire sequence length.

Consider a sequence length of 3 where $\mathbf{p}_1'\mathbf{q}_1 = \alpha$, $\mathbf{p}_2'\mathbf{q}_2 = \alpha$, $\mathbf{p}_3'\mathbf{q}_3 = \alpha + \beta$, $\mathbf{p}_1'\mathbf{q}_2 = \gamma$, $\mathbf{p}_2'\mathbf{q}_3 = \pi$, and $\mathbf{p}_3'\mathbf{q}_1 = \alpha$, where $\alpha \geq \gamma$, $\alpha \geq \pi$.

$$GAV_3 = 1 - \frac{1}{3}\left[\frac{\gamma}{\alpha} + \frac{\pi}{\alpha} + \frac{\alpha}{\alpha + \beta}\right] \tag{1.5}$$

$$MPI_3 = \left[1 - \frac{\gamma + \pi + \alpha}{3\alpha + \beta}\right] \tag{1.6}$$

When β is less than zero (but still sufficiently large), then GARP is not violated as $\mathbf{p}_1'\mathbf{q}_1 \geq \mathbf{p}_1'\mathbf{q}_2$, $\mathbf{p}_2'\mathbf{q}_2 \geq \mathbf{p}_2'\mathbf{q}_3$, and $\mathbf{p}_3'\mathbf{q}_3 \leq \mathbf{p}_3'\mathbf{q}_1$. In this case both MPI and the GAV index are set to 0 indicating no GARP violation. When β is exactly equal to zero (a mild violation of GARP), then the GAV index and MPI coincide, as budgets are equal across all observations. However, when β is greater than zero, β grows larger and larger, the MPI explodes to perfect irrationality whereas the GAV index stays relatively conservative. In this example, the GAV index is not as punishing given that it extracts surplus per observation as opposed to the MPI which accumulates surplus over all observations from the consumer. Even more variability in expenditure over observations can mean the conclusions from the GAV index and MPI can differ further.

Throughout this thesis, due to the data, as observed expenditures are very much constant over time for almost all individuals, I report the results just for the MPI. As shown above, the theoretical results will be similar enough to warrant analysis of either the GAV Index or MPI. Although they are shown to be similar, there are some cases where the GAV Index is better correlated with the CCEI than the MPI. This can be thought of as a consequence of the GAV Index allowing for a more efficient extraction of surplus from the agents. However, overall, the results for the GAV Index can be analysed and evaluated in the same manner as is done for the MPI, and therefore, do not lead to any different conclusions.

1.3.4 Theoretics of Using MPI and CCEI

In order to show robustness of the rationality indices in trying to answer the above questions, it is important to know if there are any potential *theoretical* reasons as to why the MPI or CCEI could react in a predictable way.[20] By examples, it is relatively easy to show that there are no plausible theoretical predictions.

Recall that a dataset, $\mathbf{D} = \{\{\mathbf{p}_t, \mathbf{q}_t\}_{t\in\mathbf{E_1}}, ..., \{\mathbf{p}_t, \mathbf{q}_t\}_{t\in\mathbf{E_K}}\}$, with limited consideration is defined by partitioning a dataset into observations with the same consideration set. Instead of doing the revealed preference analysis over the entire set of observations, rationality indices will be calculated for each partition. For example suppose $\mathbf{E}_1 = \{1, 2, 4, 5\}$, $\mathbf{E}_2 = \{3, 7, 8, 9, 10, 11, 12\}$, and $\mathbf{E}_3 = \{6, 13, 14, 15\}$. So, for $\mathbf{E}_1$, the set of goods with positive consumption were the same in periods 1,2,4, and 5 etc... By the definition of the MPI:

[20] It is important to note that even if some pattern exists, this does not necessarily warrant exclusion of a particular index, rather, it allows the analysis to be completed with even more precision and caution.

$$MPI_{\mathbf{E}_1} = \frac{\sum\limits_{i \in \mathbf{E}_1} \mathbf{p}_i * (\mathbf{q}_i - \mathbf{q}_{i+1})}{\sum\limits_{i \in \mathbf{E}_1} \mathbf{p}_i * \mathbf{q}_i}$$

$$= \frac{\mathbf{p}_1\mathbf{q}_1 + \mathbf{p}_2\mathbf{q}_2 + \mathbf{p}_4\mathbf{q}_4 + \mathbf{p}_5\mathbf{q}_5 - \mathbf{p}_1\mathbf{q}_2 - \mathbf{p}_2\mathbf{q}_4 - \mathbf{p}_4\mathbf{q}_5 - \mathbf{p}_5\mathbf{q}_1}{\mathbf{p}_1\mathbf{q}_1 + \mathbf{p}_2\mathbf{q}_2 + \mathbf{p}_4\mathbf{q}_4 + \mathbf{p}_5\mathbf{q}_5}$$

$$= 1 - \frac{\mathbf{p}_1\mathbf{q}_2 + \mathbf{p}_2\mathbf{q}_4 + \mathbf{p}_4\mathbf{q}_5 + \mathbf{p}_5\mathbf{q}_1}{\mathbf{p}_1\mathbf{q}_1 + \mathbf{p}_2\mathbf{q}_2 + \mathbf{p}_4\mathbf{q}_4 + \mathbf{p}_5\mathbf{q}_5}$$

$$MPI_{\mathbf{E}_2} = 1 - \frac{\mathbf{p}_3\mathbf{q}_7 + \mathbf{p}_7\mathbf{q}_8 + \mathbf{p}_8\mathbf{q}_9 + \mathbf{p}_9\mathbf{q}_{10} + \mathbf{p}_{10}\mathbf{q}_{11} + \mathbf{p}_{11}\mathbf{q}_{12} + \mathbf{p}_{12}\mathbf{q}_3}{\mathbf{p}_3\mathbf{q}_3 + \mathbf{p}_7\mathbf{q}_7 + \mathbf{p}_8\mathbf{q}_8 + \mathbf{p}_9\mathbf{q}_9 + \mathbf{p}_{10}\mathbf{q}_{10} + \mathbf{p}_{11}\mathbf{q}_{11} + \mathbf{p}_{12}\mathbf{q}_{12}}$$

$$MPI_{\mathbf{E}_3} = 1 - \frac{\mathbf{p}_6\mathbf{q}_{13} + \mathbf{p}_{13}\mathbf{q}_{14} + \mathbf{p}_{14}\mathbf{q}_{15} + \mathbf{p}_{15}\mathbf{q}_6}{\mathbf{p}_6\mathbf{q}_6 + \mathbf{p}_{13}\mathbf{q}_{13} + \mathbf{p}_{14}\mathbf{q}_{14} + \mathbf{p}_{15}\mathbf{q}_{15}}$$

Firstly, conditional on all the consideration sets having the same number of goods, there appears to be no natural theoretical pattern that could arise for longer/shorter sequence lengths. In other words, even if the consideration sets used for each partition have exactly the same number of goods, there are no discernible theoretical predictions to be had in terms of sequence length.

Secondly, looking at $MPI_{\mathbf{E}_1}$ and $MPI_{\mathbf{E}_3}$ with the same sequence length, even if they come from consideration sets that have a vastly different number of goods, it is not possible to see any plausible reason for systematic differences in the measures. From a theoretical perspective, even if $\mathbf{E}_1$ came from a consideration set with 5 goods and $\mathbf{E}_1$ with 85 goods, there are no reasons why the calculated MPIs should change systematically.[21]

Recall that the definition of the CCEI is $e^* \in sup(e)$ such that that no violations of VGARP exist (for $e \in [0, 1]$). This implies that comparing any chain of bundles that includes a zero-bundle will result in a CCEI of 1. This is because zero-bundles always trivially satisfy GARP (and thus VGARP). An example would be the following:

[21] A similar analysis can be done with the GAV Index, with the same conclusion.

Going back to the simplest case, suppose there are 2 bundles, the CCEI would be

$$e^* = max\left\{\frac{\mathbf{p}_1'\mathbf{q}_1}{\mathbf{p}_1'\mathbf{q}_2}, \frac{\mathbf{p}_2'\mathbf{q}_2}{\mathbf{p}_2'\mathbf{q}_1}\right\} \tag{1.7}$$

Given that $e \in [0, 1]$, if either $\mathbf{q}_1$ or $\mathbf{q}_2$ were zero implies that $e^* = 1$. The interpretation being that the consumer can always afford the zero bundle, at any prices, so no income adjustment would be required to be considered rational. Although there is a trivial satisfying of GARP, on average, a household may appear more rational if the consideration set has any bundles that are the zero vector. Hence, from the point of view of the CCEI, exclusion of the zero bundles should give a more reasonable picture of rationality, as the chance for trivial satisfying of GARP is removed.[22] Again, by definition of the CCEI, an increasing/decreasing sequence length does not have any predictable effect on the CCEI, nor should the number of goods within consideration sets give rise to any theoretical pattern.

1.4 Data Description

1.4.1 Stanford Basket Dataset

The Stanford Basket Dataset is a a scanner panel dataset based on data from nine major supermarkets located in a large U.S. city. There are 103,345 transactions involving 4,082 (unique) items observed on a weekly basis per household. This dataset[23] comprises expenditure on groceries[24] and demographic data[25] for 494 households covering June 2001 to June 2003 (26 months). It is important to note that analysis on individual households can be performed because of the 'panel' nature of the dataset.

For this particular dataset, I follow Echenique et al. (2011) by focusing on food expenditure.[26] Widely accepted in the literature, food expenditure is not

[22] Obviously, it is still possible to satisfy GARP and score a CCEI of 1 without zero-bundles.
[23] Collected by Information Resources Inc.
[24] Bacon, Barbecue, Butter, Cereal, Coffee, Crackers, Eggs, Ice-Cream, Nuts, Analgesics, Pizza, Snacks, and Sugar.
[25] Age, Income Level, Family Size, and Education Attained
[26] For direct comparisons, I follow their uploaded version of the Stanford Basket Dataset.

expected to fluctuate greatly in response to large changes in income. The main reason for this is that many food types are considered necessities, which means that anticipated changes in income should not reflect in any real changes to food consumption. Excluding luxury goods should escape the issue of potential overestimation of irrationality. The reason for this being that even with minimal changes to income, the consumption response for luxury goods can be volatile.[27] Although this volatility is not necessarily irrational behaviour, when measuring the extent of the irrationality, it can potentially lead to an upward/downward bias of the CCEI, MPI, or GAV Index making consumers seem more irrational/rational than they are. The thirteen products selected, therefore, follow a representative consumer basket as well as avoiding these issues related to luxury goods.[28]

Another issue to be addressed is that of income separability. The assumption held is that, the items chosen form a separable group with respect to household preferences. If this assumption did not hold, the methodology of measuring irrationality using the CCEI and MPI fails as optimal bundle choices would depend on goods external to the ones chosen. This is also true from the framework of consideration sets. Without the income separability of this group, the true consideration set may include goods outside of the representative basket which would not facilitate robust analysis. This assumption is widely accepted in related demand literature e.g. (Blundell (1988) etc...)

Owing to the fact that GARP analysis requires sufficient price observations and variation, any goods that did not have this requirement were dropped from the dataset (approximately 13% of observations dropped). Note also that data is aggregated to the monthly level as in Echenique et al. (2011) and Demuynck and Seel (2018). Given the basket of goods studied, longer time periods would be unnecessary given the nature of the goods and shorter time periods would not take into account issues of storability across short time periods.

[27] Volatile in the sense that any marginal drops in income could result in the consumer substituting away from luxury goods.

[28] Spanning 375 products from which to choose.

1.5 Results

1.5.1 Rationality Indices

Tables 1.1 and 1.2 provide some summary statistics with respect to the partitioning. On average, over the 26 months of data, households bought 56 different goods (from a total of 375 distinct goods) with a fair amount of variation across households. The average number of equivalence classes was around 24.47. For a dataset of 26 observations, this means that, on average, almost every observation is within its own equivalence class. This is exactly true for 217 of the households that have 26 equivalence classes. For the household that has 14 equivalence classes, this means that they had 14 different consideration sets over 26 months, where each consideration set has a different set of goods with positive consumption. As there are several households with many consideration sets, the proceeding results will take into account the trivial satisfying of GARP for those that have 26 different consideration sets over the 26 observations, as these individuals will score perfectly in terms of rationality.

Table 1.3 gives a summary of the rationality indices for the entire dataset. Any average or median calculated was done so for those who violated GARP. This includes all cycle lengths from 2 up to and including 5. As a general comment, although the frequency of GARP violations is high (around 80%), it seems as though the magnitude of these violations is not large.[29]

Table 1.4 provides the same statistics but incorporating consideration sets. The process is exactly the same as in Table 1.3, except the rationality indices are calculated for each equivalence class i.e. for each consideration set. The statistics in this table then show the average over all consideration sets and households. Consider cycle length 2. Suppose a household has 4 consideration sets each with sequence length of 6. I calculate the rationality index for all pairs of bundles *for each of the 4 consideration sets*, with respective cycle length. Grouping each of the

[29] In fact, the MPI and GAV Index are not statistically able to reject GARP at any sensible level of significance. The CCEI is also above 0.95, which is the standard threshold for rejecting GARP using the CCEI.

calculations from each consideration set, I then compute the summary statistics.

As GARP was trivially satisfied for the vast majority of consumers with 25+ equivalence classes, these were omitted for all cycle lengths.[30] Including them would have essentially meant the summary statistics would have exhibited perfect rationality, albeit somewhat superficially. The table shows the statistics for those who had between 14 and 22 consideration sets (depending on cycle length). There are two points of interest showcased by Table 1.4. Firstly, the proportion of households violating GARP has reduced. Secondly, across the board, the consumers appear to be more rational. On the former, as there are essentially less opportunities to fail GARP, given the shorter total sequence lengths, it is not surprising that there is a drop. However, it did drop to a proportion lower than what was expected. On the latter, of those who did violate GARP, there also appears to be a drop in the level of irrationality. This suggests that, in the cases where consumers do violate GARP, they are not doing so as severely. It seems to be the case that when taking into account limited consideration, agents are in fact more rational than originally thought, perhaps as the decision making process is simpler than originally prescribed by the canonical model.

As rationality indices are computed for each consideration set, it is possible to compare consideration sets with different numbers of goods. Consider a household with 4 consideration sets each with sequence length 6. Previously, for each of the consideration sets, I calculated the rationality indices and took summary statistics over the aggregate. However, suppose 2 of the consideration sets had 40 goods used, whereas the other 2 used only 10 goods. It may be more suitable to compare consideration sets of similar sizes. The modal number of goods per consideration set was 3, with the vast majority of consideration sets having between 2 and 18 goods. Note that any household with the specific consideration set size of the column is included. For comparison, Table 1.5 reports summary statistics for cycle length 2, but for consideration sets with 2-7 goods, again excluding those with 25+ consideration sets. Table 1.6 reports the same but for consideration sets

[30] Clearly trivially satisfied for those with 26 equivalence classes.

with 8-13 goods.

Interestingly, for all the rationality indices, there does appear to be an increase in the level of irrationality as the consideration set size increases. This is in keeping with the idea being that it might be costlier to make a rational decision when there are more goods from which to choose. In essence, a household that has larger consideration sets may make somewhat more irrational decisions compared to those with smaller consideration sets. However, it is important to be cautious as the magnitudes of the irrationalities are small, and any increase in these over consideration set size is also fairly minimal; although the pattern clearly exists, with these data, the effect is unsubstantial.[31]

Average sequence length is defined as the average of the average sequence length for a consumer with any consideration set of a specific size. In terms of sequence length, it is not hugely surprising that it is decreasing with consideration set size. As the consideration set size increases, there is an increased chance that these consideration sets belong only to their own equivalence class. As the dataset becomes more partitioned, this means the average sequence length decreases which is what happens as the consideration set size increases. This seems to be offset by the fact that there are many combinations of goods which lead to specific consideration set size. For example, there are many different consideration sets that can have just 2 goods. Overall, this seems to be a modest decrease.[32]

Tables 1.7, 1.8, and 1.9 segment the data both by consideration set size and sequence length.[33] Table 1.7 simply corroborates that there are many consideration sets that are in their own equivalence class i.e. they are unique to their time period. As per the previous tables, Tables 1.8 and 1.9 also show similar patterns across consideration set size do seem to occur, with a slight increase in the level of irrationality. Additionally, there do appear to be some overall decreases in rationality for longer sequence lengths. These suggest that as the decision-making

[31]Based on the statistical tests from the MPI and the GAV index, it is not possible to reject GARP at any sensible significance level.

[32]More generally, it may simply be a consequence of the strict definition of a consideration set.

[33]Any household with that specific sequence length and consideration set size is included.

process becomes more difficult, whether it be from a larger consideration set or a longer sequence length, then there is a greater likelihood of being more irrational. As with with the previous results the magnitudes of these patterns are modest, however, they are definitely present.

1.5.2 Demographics

Tables 7 reports a set of regressions of the CCEI on a selection of demographic variables from the Stanford Basket Dataset.[34] The demographic data collected were dummy variables for age[35], income[36], family size[37], and education[38]. Additionally, Table 7 shows regressions that also control for average consideration set size (ACSS), and average sequence length (ASL).[39]

From Table 7, education is not a statistically significant determinant of rationality. Intuitively, education levels should improve the decision-making process and hence lead to the observation of higher rationality. Although this effect is captured by the negativity of these coefficients, these are not statistically significant at the 10% significance level. In terms of the age dummies, it appears that only MidAge is statistically significant whereas OldAge is not. On one hand, this result could make intuitive sense as it would not be unfair to say that younger individuals are able to access information more efficiently than their older counterparts. As such, being younger could be an important factor in determining rationality. On the other hand, age could signify experience which should lead to more rational decisions. A potential explanation as to why this is not seen in the tables is that rationality is an inherent characteristic; rationality could be seen as a quality that is determined by an inherent ability and availability of resources.

The way in which decisions can also be influenced by their environment. This can be seen though the family size variables. Familial decisions, especially related

[34]Households with missing demographics were omitted from the regressions.

[35]OldAge >65, MidAge $\in$ [30,65]

[36]HighIncome $>$ \$45000, MidIncome $\in$ [\$20000,\$45000]

[37]LargeFamily >4, MidFamily $\in$ [3,4]

[38]High School, College; Average education across partners.

[39]The purpose of the regressions is to easily tabulate conditional correlations with the covariates and rationality. Causal effects would be impossible to argue in this context.

to basics such as food, could be regarded as an extremely important part of the decision-making process. This could be because it forms a large separable group as well as the fact that there may be a very direct and clear desire to avoid wasting income on superfluous food expenditure. In both tables, negative and significant coefficients on the family size variables support this argument. Although, as more factors are involved, the above argument may only be true for the larger families.

In terms of income variables, being a middle-income household is more likely to affect rationality than a high-income household. Relating back to the argument of not wanting to waste income, this makes intuitive sense as high-income households are more likely to make seemingly irrational behaviour when it comes to their necessity goods. Relative to their higher income, there is more scope for wasted income for a basket of necessity goods. As middle-income households may have to be more cautious when it comes to their necessities expenditure, they are less willing to waste any of their income through irrational choices; again, this idea is supported by negative and significant coefficients on MidIncome in both tables.[40]

When controlling for ASL, there do not appear to be any noticeable changes in the other coefficients, and all covariates that were previously significant have stayed so. ASL itself is statistically significant at the 1% level with negative coefficient, suggesting that consumers that have longer (average) sequence lengths tend to make slightly more irrational decisions. However, when also controlling for ACSS, the statistical significance of ASL is lost suggesting that it is ACSS that is a more relevant in explaining (ir)rationality. Again, as the coefficient is negative, this seems to corroborate with the story that agents that have larger consideration sets may be more likely to make mistakes in their bundle choices, as it is more difficult to make decisions from a larger set of goods. In comparison to the demographic variables, it seems as though the coefficient ACSS is in the same order of magnitude as the other main covariates suggesting it can play a large role in the measurement of rationality.

[40] It could be argued that the representative basket for richer households should include luxuries in order to form a proper income separable group.

1.6 Implications

1.6.1 Application & Policy Implications

An important issue that has arisen from all the previous analysis pertains to whether households are consciously aware of the role of consideration sets. This is most likely the case when it comes to the most obvious examples that come from the marketing literature. However, in all likeliness, consumers are using consideration sets in a far more general setting, as suggested by the previous results. Potentially, if more households were aware of this decision-making process, they would be less likely to waste income or be exploited in terms of making rational decisions. This has hugely important implications on research in general where the assumption of completeness is presumed with little additional thought. In fact, from the literature on consideration sets and the results in this chapter, avoiding this issue can lead to misleading and inaccurate conclusions. Although the issue of rationality may seem small in magnitude, this is countered by the fact that even the smallest magnitudes are amplified by population. Overall, there is large potential for more rational decision-making to be made if consumer are more aware of techniques that improve their decision-making process. In this particular scenario of bundle choices, use of consideration sets seems to be one of the most clearcut ways of simplifying decision-making in order to make better choices.

Another related policy issue comes from the simulated dataset analysis. The data imply that the irrational decisions lead to a low scoring rationality (by definition), the natural reverse argument being that those with lower rationality are more susceptible to irrational decisions in other settings. This suggests policy should target those who are regarded as making frequent and/or large irrational decisions. The potential reasons for this are many such as unavailability of resources (both in monetary terms and informational), lack of awareness in relation to other consumers and firms (asymmetric information argument) etc... In essence, policy needs to take into account that there are households that are not as rational as others which leads to irrational decisions, which feeds through adversely through the economy as wasted and exploited income; the positive impact that any policy may have will be dampened if decision-making processes and rationality are not examined properly.

As demonstrated by the regressions, it is not just the intrinsic characteristics of the household that matter but also the more 'human' and 'situational' factors i.e. demographic factors that more accurately describe their 'actual' living situation. It is true to say that variables on income, age, education etc... are all of importance when it comes to explaining rationality. However, 'how' a household lives may also be extremely important when looking at choice data. A stark example could be comparing two modest-income households with several young children, one with family members geographically nearby to help with the daily routine, and a family without. It could be argued that the household with help from additional family members has more time to gain information to improve their decision-making and thus would score higher according to a rationality index. However, in relative terms, this distinction would not have been made by the use of standard demographic variables. The regressions also seem to show that more difficult decisions tend to be more irrational than others, as suggested by the negative relationship between average consideration set size and rationality. The argument being that there is an intrinsic household ability/rationality that is biased upwards or downwards depending on 'living' factors. These kind of factors should be taken into account when it comes to analysis related to choice data, revealed preference or consideration sets.

1.7 Extensions & Limitations

A well known empirical feature of revealed preference analysis is that there tends to be few violations of GARP, this gives rise to the 'low power' of GARP. Andreoni et. al (2011) explain in detail many 'power measures' and 'power indices' that attempt to measure the power of GARP for a given dataset. The concept behind power measures is the measuring of the ex-ante probability of rejecting the null hypothesis of rationality when false. Power indices measure by how much the data generating process, ex-post, needs to be stimulated in order to induce a rationality violation. However, when dealing with power measures, there is flexibility in the definition of the null hypothesis and even more so in how the alternative hypothesis is constructed. This chapter sought to analyse rationality combined with consideration sets and used a simulated dataset as an alternative way of dealing with the potential issue of low power, rather than using power measures for robustness.

Note that choosing the 'most accurate' power measure or power index is somewhat subjective and the choice may change depending on purpose and type of data available. For example, experimental datasets may have given prices and income to subjects and then asked for their choices, therefore, the price and income variability is an issue that has to be taken into consideration when choosing a power measure. Another example would be the use of cross-sectional data which typically exhibit higher income variability which may require a more sensitive power measure (typically, datasets with high income variability tend to not find many GARP violations). However, generally speaking, rationality indices tend to have good power in their ability to reject/not reject rationality. Hence, by simulating an environment similar to the dataset used, this leads to a more robust check of the power of revealed preferences in this particular setting.

Another potential extension would be to modify the typical analysis related to existing rationality indices to directly take into account the use of zero-bundles. The analysis done by Echenique et. al (2011) involved only analysing decisions that were deemed irrational, as such, their calculations on averages did not include any completely rational decisions[41]; this may lead to an interpretation that underestimates consumer rationality. Perhaps an alternative method would have been to take a weighted average over the CCEIs/MPIs instead of a simple arithmetic mean or omitting the completely rational decisions. For example, some form of down-weighting on the purely rational decisions as a function of the number of ra-

[41] Averages were done without those who scored perfectly rational according to the CCEI and MPI.

tional decisions. On average, this should not change the direction of analysis, but may give more meaningful (and perhaps accurate) interpretation to the rationality indices.

Briefly mentioned previously, the issue of how particular goods were aggregated potentially meant that analysis and conclusions drawn were not as specific as they could have been. The publicly available version of this dataset aggregated some of the goods in a precise way. Specifically, the available data aggregated distinct goods of particular brands across different sizes. The aggregation was done across the different sizes of each distinct product for which an average price (according to size) was calculated.[42] Although this type of aggregation is not a significant issue, with access to the more raw data, it would have been interesting to see how rationality changed when including or excluding certain brands of particular products (in a similar vein to Chiang et. al (1999) with different ketchup brands).

1.8 Concluding remarks

The objective of this chapter was to present analysis that combines the role of consideration sets with economic rationality in the decision-making process. Using a scanner panel dataset and a simulated dataset, and applying three separate measures of rationality, there is evidence to suggest a heuristic such as consideration set formation is able to help consumers make more rational decisions. The role of consideration sets is reflected through increasingly irrational consumption choices exacerbated by the 'average size' of the consideration set; in a sense, suggesting that consumers are better at making rational decisions in less complex environments, and thus can be more rational overall. Additionally, there is evidence to suggest that demographic factors play some role in explaining why certain individuals are more likely to make more irrational decisions than others, controlling for variables that are related to consideration set formation. Overall, through the lens of economic rationality, the role of consideration sets can be as a useful heuristic

[42] e.g. Averaging over the different Evian water bottle sizes to get an 'average Evian water product' priced at the weighted average of the prices of the different bottle sizes.

to benefit consumers by enhancing their decision-making process.

In conclusion, there is a clear relationship between the role of consideration sets and revealed preference theory (vis-à-vis rationality) in the decision-making process of the household. In this chapter, I hope to have provided an initial insight into how these concepts can be combined and how that this can improve our ability to study economic behaviour.

Chapter 2

Cost of Consideration and Revealed Preference

2.1 Introduction

[1] Commonly seen in the marketing and management sciences literature, initially proposed by Wright and Barbour (1977) a consideration set is seen to be a subset of the total number of goods with which the consumer makes a choice. Similar to the description given in Horowitz and Louviere (1995), the set of goods that the consumer uses to make their decision "need not coincide with the set of all possible alternatives". In this chapter, I explore a consideration set incorporated revealed preference approach to try determine the cost of consideration.

Since the seminal work of Afriat (1967), the Generalised Axiom of Revealed Preference (GARP) has been used as a way of determining consumer rational behaviour from micro-datasets. Afriat (1967) elegantly proves how GARP can provide us with necessary and sufficient conditions on bundle choices that are consistent with economic rationality. This can boil down to verifying whether the "Afriat Inequalities" hold for a finite dataset. Diewert (1973) provides an easily implementable linear program which yields, as solutions, the Afriat Inequalities.

[1]For this specific chapter, I would like to thank Richard Blundell, Syngjoo Choi, Laurens Cherchye, Bram De Rock, and Frederick Vermeulen for invaluable discussions and comments. I am grateful for financial support from ECR Grant 509157.

If the inequalities are satisfied, the consumer is said to have a utility function that rationalises their behaviour.

What this chapter hopes to highlight is that revealed preference analysis should somehow incorporate the use of consideration sets in the decision-making process. It is well known that consumers must narrow down their set of chosen goods due to some form of cognitive constraint. This is, by no means, a slight on consumers (or even their rationality), rather, it is a way of trying to incorporate more accurate behavioural assumptions into a relatively parsimonious model of consumers. This suggests that observed data may not be consistent with rationality. Given the strict standard economics definition of perfect consumer behaviour, it may be no surprise that consumers do not always appear as standard utility maximisers. I directly attempt to model and estimate this cost associated with consideration in a way that assumes as little as possible whilst still allowing rejections of (bounded) rationality. In principle, I can derive a cost of consideration per good and can be heterogenous across consumers. Mehta et al. (2003) answer a similar question in the sense that they try to establish the cost of consideration using a structural model for laundry detergents through uncertainty in quality of the goods. However, this chapter differs by addressing this issue through revealed preferences without making any distributional assumptions on consideration costs.

The chapter is structured in the following way: Section 2.2 provides a brief outline of the literature as well as a technical overview of revealed preferences. Section 2.3 presents a standard utility maximisation model that incorporates limited consideration, with estimation procedure. Section 2.4 provides the results after applying the model to a scanner dataset. Section 2.5 concludes.[2]

[2]In the appendix, I briefly mention the issue of dimensionality when it comes to analysing consideration sets in a revealed preference approach.

2.2 Literature Review

2.2.1 Consideration Sets

Amongst the economics literature, Spiegler and Eliaz (2011), Manzini and Mariotti (2014), Masatlioglu et al. (2012), Masatlioglu and Nakajima (2013), Demuynck and Seel (2018) are the most prevalent papers.

Spiegler and Eliaz (2011) apply economic theory to the original application of consideration sets, within a framework of marketing and competition. Their idea behind the construction of the consideration set is via the consumer's unawareness of certain products. Some form of marketing strategy is then introduced to consumers in order to alter their consideration set. They devise a model that seeks to encapsulate a means for firms to somehow extract surplus from consumers as a result of asymmetric information from the use of consideration sets. Masatlioglu and Nakajima (2013) design a framework that studies a more general framework of consumers who use consideration sets in their decision-making process. The consideration set is formed as a function of an exogenous and feasible 'starting point' (F). Let $C_1(F(0), Y) \subseteq Y$ where Y represents budget feasibility. The consumer makes choices based on $C(F, Y)$ with a binary relation operator on the complete set of alternatives, A. If the consumer gets (at least) their starting point option, the decision-making process is complete, if not, the consumer constructs a 'smaller' consideration set where the new starting point is the previously rejected alternative i.e. $C_2(F(1), Y)$, where $F(1)$ is the alternative that was rejected in the previous period; the process is then iterated. Masatlioglu and Nakajima (2013) also propose a more realistic version of their iterative model in that the initial starting point is inferred from the choice data.[3]

Manzini and Mariotti (2014) specifically model a boundedly rational consumer whereby agents have a probability distribution over alternatives. They attempt to infer the preference ordering from consumer choices under the notion that the

[3]Spiegler and Eliaz (2011) can be seen as a special case of Masatlioglu & Nakajima (2012) where the initial marketing device and initial alternative form the basis of the first consideration set. The next stages of consideration set formation are then as a result of marketing.

choices were generated by a consideration set where the consideration set is a function of the larger set of alternatives. Manzini and Mariotti (2014) put forward a model that suggests the formation of consideration sets is random in the sense that all (feasible) alternatives have a probability of being in the consideration set; they call this the ‘attention parameter’.[4] Their decision-making procedure involves a decision-maker with a consideration set with a complete preference relation over that consideration set.

Demuynck and Seel (2018) depart from the economic literature in that the analysis is done from a revealed preference point of view; the most basic difference being that choices á la Afriat are continuous (over a set of discrete goods). They provide a new axiom of revealed preference, namely the Limited Axiom of Revealed Preference (LARP). LARP involves verifying that GARP holds within partitions of a dataset admitted by the same consideration sets. This allows an extension to Afriat’s Theorem under limited consideration where goods that are not admitted by a particular consideration set are incorporated via unobserved subjective prices. Consideration set formation is then modelled via the beliefs put on those subjective prices. Applied to a scanner dataset from the Denver area (USA), they largely fail to reject the use of consideration sets in the consumers decision-making process. To the best of my knowledge, this is the only paper that incorporates standard revealed preference analysis with consideration sets.

As Demuynck and Seel (2018) is closest to this chapter, I specifically highlight the differences in concept. In Demuynck and Seel (2018), they initially model a consumer with a fixed consideration set. For the goods within this consideration set, the marginal utilities are as in the standard case. However, goods beyond the consideration set can have marginal utilities that are larger or smaller than those goods within the consideration set. The consideration set formation can be endogenised with different ways of thinking about how those marginal utilities differ from the standard case with the notion of subjective prices e.g. use of average price, previous correct price etc... This is distinctive from the model

[4] In the limited attention sense.

presented below as the marginal utilities are always (weakly) higher than the standard case. As will be seen, this is due to this fixed cost of consideration. The main distinction is that this chapter thinks of there always being a fixed cost of consideration whereas Demuynck and Seel (2018) only model goods not being in the consideration set with subjective prices.[5]

2.2.2 Revealed Preferences

This subsection familiarises the reader with the conventions of revealed preference.
Let $\mathbf{T}$ denote time periods such that $\mathbf{T} = \{1, 2, ..., T\}$
Let $\mathbf{p}_t \in \mathbb{R}^n_{++}$ denote prices in period $t \in \mathbf{T}$
Let $\mathbf{q}_t \in \mathbb{R}^n_{+}$ denote quantities in period $t \in \mathbf{T}$
where n denotes the *total* number of available goods.
Define a finite dataset, $\mathbf{D}$, as a collection of all prices and quantities i.e. $\mathbf{D} = \{\mathbf{p}_t, \mathbf{q}_t\}_{t\in\mathbf{T}}$. This dataset is a collection of observed consumption behaviour, $\mathbf{q}_t$, for a consumer facing prices, $\mathbf{p}_t$, at observation, t.

A dataset $\mathbf{D} = \{\mathbf{p}_t, \mathbf{q}_t\}_{t\in\mathbf{T}}$ is rationalisable if there exists a utility function $u : \mathbb{R}^n \to \mathbb{R}$ and for all observations $t \in \mathbf{T}$, there exists a weakly positive income level y_t such that:

$$\mathbf{q}_t \in \underset{\mathbf{q}}{\operatorname{argmax}}\, u(\mathbf{q}) \text{ subject to } \mathbf{p}_t'\mathbf{q} \leq y_t$$

This says that a dataset is rationalisable if the bundles, $\{\mathbf{q}_t\}_{t\in\mathbf{T}}$, are consistent with utility maximisation with a linear budget constraint.

A bundle $\mathbf{q}_i$ is directly revealed preferred to $\mathbf{q}_j$ if $\mathbf{p}_i'\mathbf{q}_i \geq \mathbf{p}_i'\mathbf{q}_j$. In words, the bundle $\mathbf{q}_i$ was at least as affordable as bundle $\mathbf{q}_j$ at observation i. Let R denote the *directly revealed preferred* binary relation i.e. $\mathbf{q}_i$ is directly revealed preferred to $\mathbf{q}_j$ if $\mathbf{q}_i R \mathbf{q}_j$.

A bundle $\mathbf{q}_i$ is strictly directly revealed preferred to $\mathbf{q}_j$ if $\mathbf{p}_i'\mathbf{q}_i > \mathbf{p}_i'\mathbf{q}_j$. In

[5] In fact, one of the notions of subjective price is that the price cannot be "too far away" from the true price. So the notion of a distorted price could be incorporated as a special case, if the cost of consideration only exists for goods outside the consideration set.

words, the bundle $\mathbf{q}_i$ was costlier than bundle $\mathbf{q}_j$ at observation i. Let R_S denote the *strictly directly revealed preferred* binary relation i.e. $\mathbf{q}_i$ is strictly directly revealed preferred to $\mathbf{q}_j$ if $\mathbf{q}_i R_S \mathbf{q}_j$.

A bundle $\mathbf{q}_i$ is indirectly revealed preferred to $\mathbf{q}_j$ if there exists a sequence of observations $x, y, ..., z$ in $\mathbf{T}$ such that $\mathbf{q}_i R \mathbf{q}_x, \mathbf{q}_x R \mathbf{q}_y, ..., \mathbf{q}_z R \mathbf{q}_j$. Let P denote the *indirectly revealed preferred* binary relation i.e. $\mathbf{q}_i$ is indirectly revealed preferred to $\mathbf{q}_j$ if $\mathbf{q}_i P \mathbf{q}_j$.[6] The number of bundles in a chain of directly revealed preferred bundles is called the *sequence length*.[7]

The Generalised Axiom of Revealed Preference (GARP) is satisfied by $\mathbf{D} = \{\mathbf{p}_t, \mathbf{q}_t\}_{t \in \mathbf{T}}$, if, for all $\mathbf{q}_i P \mathbf{q}_j$, then it cannot be that $\mathbf{q_j} R_S \mathbf{q_i}$; if $\mathbf{q}_i$ is indirectly preferred to $\mathbf{q}_j$, it cannot be the case that $\mathbf{q}_j$ was purchased even when $\mathbf{q}_i$ is cheaper.

The seminal contribution of Afriat (1967) showed that GARP is both necessary and sufficient for a dataset to be rationalisable.[8] Given a dataset $\mathbf{D} = \{\mathbf{p}_t, \mathbf{q}_t\}_{t \in \mathbf{T}}$ the following statements are equivalent:

(i) $\mathbf{D}$ is rationalisable by a locally non-satiated utility function

(ii) $\mathbf{D}$ satisfies GARP

(iii) For all observations $t \in \mathbf{T}$, there exists $u_t, \lambda \in \mathbb{R}$ and $u_t, \lambda_t \in \mathbb{R}_{++}$ such that for all pairs of observations $i, j \in \mathbf{T}$

$$u_i - u_j \leq \lambda_j \mathbf{p_j}'(\mathbf{q_i} - \mathbf{q_j})$$

(iv) $\mathbf{D}$ is rationalisable by a strictly monotone and concave utility function.

where the inequalities of (iii) are called the *Afriat inequalities*. A neat interpretation of these come from the first order conditions of a constrained maximisation problem using KKT conditions, where λ_t is the Lagrange multiplier for the budget constraint at observation t (assuming differentiability and concavity).

[6] The indirectly revealed preferred binary relation is the transitive closure of R. It is the binary relation that is transitive and minimal with respect to the set it is on.

[7] For example, $\mathbf{q}_a R \mathbf{q}_b, \mathbf{q}_b R \mathbf{q}_c, \mathbf{q}_c R \mathbf{q}_d$ has a sequence length of 4.

[8] See Afriat (1967), Diewert (1973), Varian (1982b), Fostel et al. (2003) for detailed proofs of Afriat's Theorem.

Following Demuynck and Seel (2018), suppose a consumer has access to n total number of goods i.e. she chooses a consumption bundle $\mathbf{q}_t$ from a set of goods $G = \{1, ..., n\}$. If the consumer does not necessarily take into account all n goods when purchasing, the consumer is said to have used a consideration set $I_t \subseteq G$. This means that consumption of any goods that lie outside of her consideration set must be equal to zero (but does not exclude the option that goods within the consideration set are also zero).

This leads to a natural definition of rationalisability with a consideration set. A dataset $\mathbf{D} = \{\mathbf{p}_t, \mathbf{q}_t\}_{t\in\mathbf{T}}$ is rationalisable with consideration set if there exists a utility function $u : \mathbb{R}^n \to \mathbb{R}$ and for all observations $t \in \mathbf{T}$, there exists a weakly positive income level y_t and consideration set $I_t \subseteq G$ such that:

$$\mathbf{q}_t \in \operatorname*{argmax}_{\mathbf{q}} u(\mathbf{q}) \text{ subject to } \mathbf{p}_t'\mathbf{q} \leq y_t,$$
$$q^i = 0 \text{ for all } i \notin I_t$$

This says that a dataset is rationalisable if the bundles, $\{\mathbf{q}_t\}_{t\in\mathbf{T}}$, are consistent with utility maximisation with a linear budget constraint and additional constraints where the consumption of the i^{th} good in $\mathbf{q}_t$ is zero.

2.3 Theoretical Framework

2.3.1 Model

In order to incorporate a consideration set cost into the standard utility maximisation process widely studied in economics, I present a simple modification that involves consumers maximising their utility, subject to, what is essentially, a specific non-linear budget constraint. For every time period, all consumers solve the following optimisation problem[9]:

$$\max_{\mathbf{q}} U(\mathbf{q}) \tag{2.1}$$

[9] I include the period subscript only when necessary for complete clarification.

subject to

$$\mathbf{p}'\mathbf{q} + c(\mathbf{q}) \leq y \tag{2.2}$$

Clearly, the choice of $c(\mathbf{q})$ is important in determining the process in which consideration sets affect the bundle choice. If $c(\mathbf{q})$ is 0 for all $\mathbf{q}$, then the standard practice holds. This is non-trivial and is discussed in much further detail below. In particular, how this cost function will incorporate a cost of consideration.

In keeping with the notion of consideration sets, an "ideal" optimisation problem would be as follows:

$$\max_{\mathbf{q}} U(\mathbf{q}) \tag{2.3}$$

subject to

$$\mathbf{p}'\mathbf{q} \leq y \tag{2.4}$$

$$c(\mathbf{q}) \leq F \tag{2.5}$$

where equation (2.5) is some form of cognitive constraint arising from the use of consideration sets. If the cost of consideration is strictly less than the "stock of cognition", then the consumer is not constrained by consideration and the optimisation reverts to standard theory. However, there are many issues associated with this modification despite the clarity of the concept. What restrictions should one put on c($\mathbf{q}$)? How does one measure F? What interpretation does F hold? Is this decision completely separable from income?

Firstly, to address the issue of interpretability, I use an idea from the revealed preference theory literature which relates the notion of irrationality with (what is essentially) a loss of income. As was seen in the previous chapter, there are many indices that seek to measure irrationality as forms of wasting income. In essence, buying affordable bundles that were previously not chosen can be rationalised if one supposes the consumer has behaved as if they had less income. With this in mind, the cost of consideration can be thought of as being linked to a monetary value. Suppose we solve the above "ideal" optimisation problem, we get the following lagrangian:

$$\mathcal{L} = u(\mathbf{q}) - \lambda_1 \left[\mathbf{p}'\mathbf{q} - y\right] - \lambda_2 \left[c(\mathbf{q}) - F\right]$$

Making use of the arbitrariness of the scale of the cognitive constraint leads to:

$$\begin{aligned}
\mathcal{L} &= u(\mathbf{q}) - \lambda_1 \left[\mathbf{p}'\mathbf{q} - y\right] - \lambda_1 \frac{\lambda_2}{\lambda_1} \left[c(\mathbf{q}) - F\right] \\
&= u(\mathbf{q}) - \lambda_1 \left[\mathbf{p}'\mathbf{q} - y\right] - \lambda_1 \left[\frac{\lambda_2}{\lambda_1} c(\mathbf{q}) - \frac{\lambda_2}{\lambda_1} F\right] \\
&= u(\mathbf{q}) - \lambda_1 \left[\mathbf{p}'\mathbf{q} - y\right] - \lambda_1 \left[\frac{\lambda_2}{\lambda_1} c(\mathbf{q}) - \tilde{F}\right] \\
&= u(\mathbf{q}) - \lambda_1 \left[\mathbf{p}'\mathbf{q} + \frac{\lambda_2}{\lambda_1} c(\mathbf{q}) - y - \tilde{F}\right] \\
&= u(\mathbf{q}) - \lambda_1 \left[\tilde{\mathbf{p}}'\mathbf{q} + \tilde{c}(\mathbf{q}) - 1\right]
\end{aligned}$$

where income and prices have been normalised.[10]

Thus, when it is binding, it is possible to transform the cognitive constraint such that it is incorporated into the standard budget constraint giving it an interpretation that is directly related to (a proportion of) expenditure. If it is not binding, then the standard case holds. In fact, in what will follow, it can be seen as a price distortion. In effect, in order to rationalise consumer behaviour, it is as if consumers make bundle choices that are consistent with higher prices.

Secondly, what kind of restriction on behaviour should the constraint represent? Specifically, what could be the functional form of the constraint? Conceptually, the following cognitive constraint is appealing from an intuitive point of view:

$$a_1 \mathbb{1}[q^1 > 0] + ... + a_n \mathbb{1}[q^n > 0] \leq F \tag{2.6}$$

which says that there is some *fixed* cognitive cost associated with a positive purchase of any q^i good (the i^{th} good in vector $\mathbf{q}$). If the cost of additional goods in the consideration set is too high, then a subset of the goods will have zero consumption, due to the cap on cognition. This is not trivial as a consumer is

[10] In essence, by scaling the cognitive constraint such that the shadow price of cognition is the same as the marginal utility of income.

potentially faced with 1000s upon 1000s of potential goods that they can buy and so there must be some limit as to how many they can consider. What equation (2.6) also highlights is that there is a one-off cost associated with introducing a good into the consideration set. In other words, for any positive purchase, there is a consideration cost. Intuitively, this cost should not be increasing in the amount purchased; the cost of consideration takes into account the price difference required to be paid to bring a good into the set of considered goods. Simply, it is costly to evaluate every single possible alternative due to limited cognition. Consider the following basic example:

$$\max_{q_1, q_2 \in \mathbb{R}} U(q_1, q_2) \tag{2.7}$$

subject to

$$p_1 q_1 + p_2 q_2 \leq y \tag{2.8}$$

$$a_1 \mathbb{1}[q_1 > 0] + a_2 \mathbb{1}[q_2 > 0] \leq F \tag{2.9}$$

Without equation 2.9, if the optimal q_1, q_2 are positive and $a_1 + a_2 < F$, then this is an agent who is not cognitively constrained by their choices of the two goods.[11] However, if $a_1 > J > a_2$, then it becomes optimal to set $q_1 = 0$ and only consume q_2 with positive amount. It is as if the agent has to pay too much attention to the 1st good, so much so, that it is not worth doing so; the cost of considering the 1st good is too high. This exemplifies the literature's view on consideration sets insofar as it is not possible to consider all possible alternatives as there is some cost of doing so. In the example above, the cost of considering each good is a_1 and a_2, and if these costs are too high, then it is possible for the consumer not to consider those good/s, even if physically affordable.

Consider a basic example of choosing from different yoghurts (brands, flavours, etc...). It is not reasonable (nor feasible) to believe that a consumer is willing to make a choice from every single possible combination of yoghurts. The consumer

[11] This form of constraint is also seen in machine learning and econometrics, namely Lasso regression.

uses some heuristic or previous knowledge to eliminate certain brands from their choice set in order to make their decision-making process easier. In this sense, there is a cost for investigating the numerous different alternatives, and at some point, it does not make sense to consider some alternatives, hence a consideration set arises. What is essential here is that the cost of considering an alternative is a simple fixed cost for any positive purchase of a good. Once a good is considered, the cost of doing so is essentially sunk and does not pose further constraint once a good is chosen with positive consumption.

Penultimately, given the specific function above for the constraint, and the combining of constraints, I propose the following consideration set constraint, allowing the function to be differentiable.

$$\mathbf{p}'\mathbf{q} + \beta_1 \tan^{-1}[(q_1+1)^f - 1] + ... + \beta_n \tan^{-1}[(q_n+1)^f - 1] \leq y \tag{2.10}$$

$$\mathbf{p}'\mathbf{q} + \boldsymbol{\beta}' \tan^{-1}[(\mathbf{q}+1)^f - 1] \leq y \tag{2.11}$$

At first glance, the above equation looks arbitrary in the use of trigonometry. However, in order to replicate the step-wise nature of equation (2.6), the asymptotes of the inverse tan function are abused, as such, the value of f determines the speed at which the function approaches its asymptote. Note that as f tends to infinity, that the trigonometric part of the constraint replicates equation (2.6). The below figures explains these points graphically. Thus, by using this function, I am able to keep differentiability whilst incorporating a cost function which is step-wise in nature.

As mentioned in section 2.2.2, Demuynck and Seel (2018) provide a parsimonious definition of rationalisability with limited consideration. In keeping with this, I incorporate the constraint on consideration in order to identify and estimate a cost of consideration. Consider the following definition of rationalisability with consideration set costs:

A dataset $\mathbf{D} = \{\mathbf{p}_t, \mathbf{q}_t\}_{t\in\mathbf{T}}$ is *fully* rationalisable with *complete* consideration set costs if there exists a concave utility function $u : \mathbb{R}^n \to \mathbb{R}$, and cost function

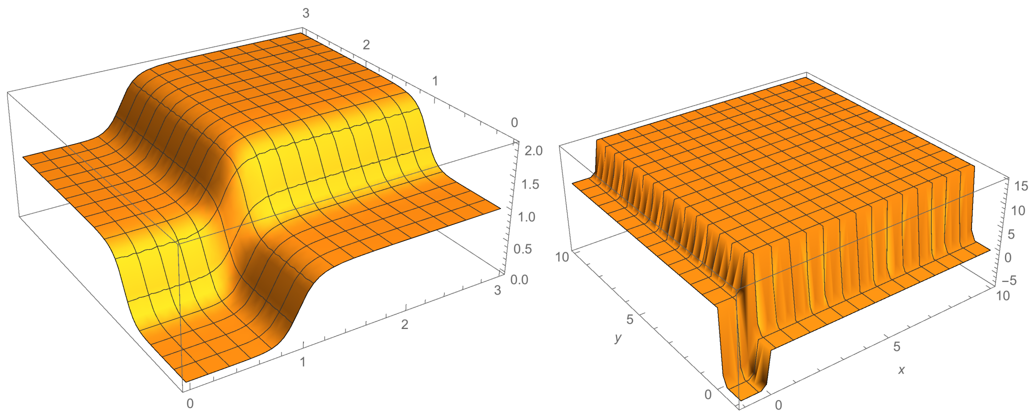

Figure 2.1: When f is large, the function is virtually step-wise.

$c : \mathbb{R}^n \to \mathbb{R}$ and for all observations $t \in \mathbf{T}$, there exists a weakly positive income level y_t and consideration set $I_t \subseteq G$ such that:

$$\mathbf{q}_t \in \operatorname*{argmax}_{\mathbf{q}} u(\mathbf{q}) \text{ subject to } \mathbf{p}_t'\mathbf{q} + c(\mathbf{q}) \leq y_t,$$
$$q^i = 0 \text{ for all } i \notin I_t$$
$$c(\mathbf{q}) = \sum_{i \in I_t} \beta_i \mathbb{1}[q_i > 0]$$

Now consider the set of goods with strictly positive consumption J_t, where clearly $J_t \subseteq I_t$ meaning that there can some goods that are considered but rationed to zero. Given the above definition of rationalisability with consideration set costs, it is not possible to identify the cost of consideration for goods that have zero consumption, as any value of β would be rationalisable for goods with zero consumption. Hence, the focus must be on goods with positive consumption. This leads to the following definition and lemma:

A dataset $\mathbf{D} = \{\mathbf{p}_t, \mathbf{q}_t\}_{t \in \mathbf{T}}$ is rationalisable with consideration set costs if there exists a concave utility function $u : \mathbb{R}^n \to \mathbb{R}$, and cost function $c : \mathbb{R}^n \to \mathbb{R}$ and for all observations $t \in \mathbf{T}$, there exists a weakly positive income level y_t and set of goods $J_t \subseteq G$ such that:

$$\begin{aligned} \mathbf{q}_t \in \underset{\mathbf{q}}{\operatorname{argmax}}\, u(\mathbf{q}) \text{ subject to } \mathbf{p}_t'\mathbf{q} + c(\mathbf{q}) \leq y_t, \\ q^i = 0 \text{ for all } i \notin J_t \\ c(\mathbf{q}) = \sum_{i \in J_t} \beta_i \mathbb{1}[q_i > 0] \end{aligned}$$

Thus, it is possible to derive costs for goods for which there is positive consumption and so the set of goods with positive consumption can be thought of as the consideration set. In the spirit of Lemma 1 of Demuynck and Seel (2018), the equivalent lemma holds for rationalisability with consideration set costs:

Lemma 0.1. *A dataset $\boldsymbol{D} = \{\mathbf{p}_t, \mathbf{q}_t\}_{t \in \mathbf{T}}$ is fully rationalisable with complete consideration set costs if it is fully rationalisable using the consideration set $I_t = J_t$.*

Proof. Recall that $J_t \subseteq I_t$ and $q^j = 0$ for all $j \notin J_t$. In particular, this means that $q^j = 0$ for all $j \in I_t \setminus J_t$ i.e. there are goods in the consideration set that can have zero consumption. This is exactly equivalent to being fully rationalisable with consideration set costs when the consideration set is J_t.[12] $\square$

The implication of this lemma is that it is possible to think of the consideration set as the set of goods with positive consumption. However, this lemma also means that it limits the estimation of the cost of consideration to those goods with positive consumption. In essence, it is not possible to identify whether someone really considered a good and rationed it to zero consumption, or whether their cost of consideration was too high for it to be considered. Therefore, the cost of consideration in this setting is the cost associated with the goods purchased with positive consumption.

Given the concavity of the utility function and the first order conditions from the utility maximisation process, the following Afriat-style inequalities are derived for every time period combination:

[12] Consider a simple case of $I_t = \{1, 2, 3, 4\}$ and $J_t = \{1, 2\}$. As these sets are always finite, the proof holds without loss of generality.

$$u(\mathbf{q_j}) \leq u(\mathbf{q_i}) + u'(\mathbf{q_i}) * (\mathbf{q_j} - \mathbf{q_i}) \tag{2.12}$$

$$u(\mathbf{q_j}) \leq u(\mathbf{q_i}) + \lambda \left(\mathbf{p}' + \boldsymbol{\beta}' \frac{f * (\mathbf{q_i} + 1)^{f-1}}{1 + ((\mathbf{q_i} + 1)^f - 1)^2} \right) * (\mathbf{q_j} - \mathbf{q_i}) \tag{2.13}$$

It is important to note that this is a non-convex optimisation problem. This means that it is not possible to guarantee (without stricter assumptions), that this problem has a maximiser. However, as it stands, the above system only provides sufficient conditions for a solution. The following subsection discuss the convexification of this problem which will lead to a theorem that guarantees that the system will have an optimal solution and that the above system will provide the necessary and sufficient conditions required for rationality in the sense of utility maximisation.

2.3.2 Convex Optimisation

As described above, there are certain properties of the CSC that are appealing from a behavioural point of view. However, the issue is that non-convex optimisation is considered non-trivial and does not have a guarantee for optimality. One method to overcome this is issue is to 'convexify' the problem whilst maintaining the main qualities of the original problem. Recall that the current problem is:

$$\max_{\mathbf{q}} U(\mathbf{q}) \tag{2.14}$$

subject to

$$\mathbf{p}'\mathbf{q} + c(\mathbf{q}) \leq y \tag{2.15}$$

where $\mathbf{p}'\mathbf{q} + c(\mathbf{q}) \leq y$ is concave given the current choice of $c(\mathbf{q})$ being concave. Hence, this is a non-convex problem. With parameterisation of the function given by θ, the following desirable properties of $c(\mathbf{q}; \theta)$ are:

- $c(\mathbf{q}; \theta)$ convex

- $c(\mathbf{0};\theta) \approx 0$

- $c(\boldsymbol{\epsilon};\theta) \approx 1$, for ϵ sufficiently small

- $c'(\mathbf{q};\theta) \approx 0$

I propose the following as a cost function that encompasses the above restrictions:

$$c(q;\theta_1,\theta_2,\theta_3) = \left[\theta_1 tan^{-1}\left(\frac{1}{q-\theta_2}\right)\right]^{\theta_3} + 1 \tag{2.16}$$

for θ_2 sufficiently small, and θ_3 sufficiently large (and odd). For $q > \theta_2$, $c(q;\theta_1,\theta_2,\theta_3)$ is convex and θ_1 is set to $\frac{2}{\pi}$ so that the asymptote of $c(q;\theta_1,\theta_2,\theta_3)$ is 1. The cost function is normalised to 1 so that coefficient on the cost function (β) is easily interpretable as a direct price distortion.

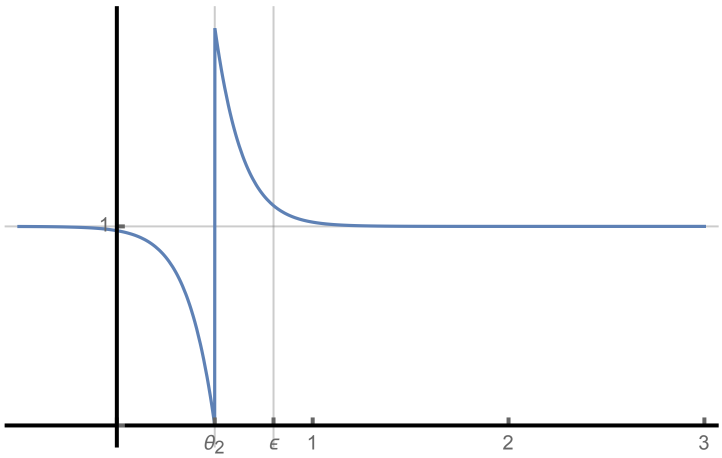

Figure 2.2: $c(q;\theta_1,\theta_2,\theta_3)$ with high θ_2 and low θ_3.

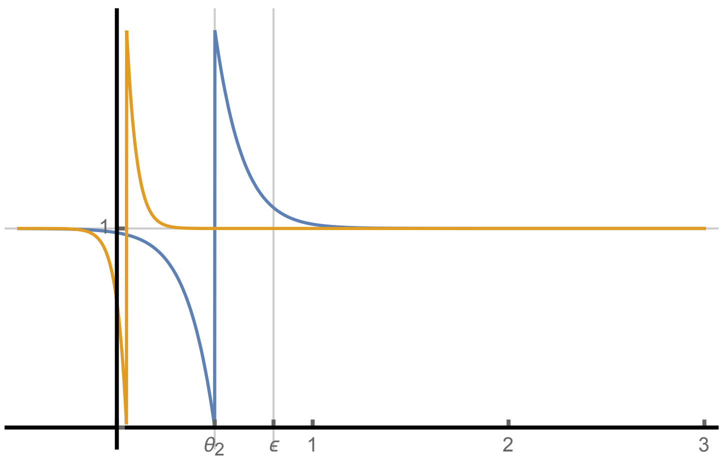

Figure 2.3: With lower θ_2 and higher θ_3, $c(q;\theta_1,\theta_2,\theta_3)$ achieves the desirable properties as described above.

Hence, the convexified utility maximisation problem now involves maximising a standard concave utility function subject to a consideration set constraint(s) that takes the form of:

$$\mathbf{p}'\mathbf{q} + \boldsymbol{\beta}' \left[\theta_1 tan^{-1}\left(\frac{1}{\mathbf{q}-\theta_2}\right)\right]^{\theta_3} + 1 \leq y \tag{2.17}$$

$$\mathbf{q} \neq \theta_2 \tag{2.18}$$

Given the convexity of $c(\mathbf{q}; \theta_1, \theta_2, \theta_3)$, I can now proof Theorem 1, which provides necessary and sufficient conditions for optimality for the previously given consumer maximisation problem.

Theorem 1. *A given dataset, $\boldsymbol{D} = \{\mathbf{p_t}, \mathbf{q_t}\}_{t\in\mathbf{T}}$, is said to be consistent with the rationalisable model of consideration set costs if and only if there exist numbers $u_i \in \mathbb{R}$, $\lambda_i \in \mathbb{R}^{++}$, and $\beta_i \in \mathbb{R}$, for all $i \in \boldsymbol{T}$, such that:*

$$u_j \leq u_i + \lambda_i \left[\boldsymbol{p}_i' + \boldsymbol{\beta}_i' c'(\mathbf{q_i})\right] * (q_j - q_i) \tag{2.19}$$

$$\boldsymbol{\beta}_i \leq \boldsymbol{p}_i \tag{2.20}$$

$$\boldsymbol{q} \neq \theta_2 \tag{2.21}$$

where $c'(\boldsymbol{q})$ is given by:

$$\frac{-\,\theta_3 {\theta_1}^{\theta_3} \left[tan^{-1}\left(\dfrac{1}{\mathbf{q}-\theta_2}\right)\right]^{\theta_3 - 1}}{\left[1 + \left(\dfrac{1}{\boldsymbol{q}-\theta_2}\right)^2\right] [\boldsymbol{q}-\theta_2]^2} \tag{2.22}$$

Theorem 1 provides a parsimonious way of estimating consideration set costs with familiar Afriat inequalities. For goods with positive consumption, it is possible to estimate the cost of consideration which is consistent with the aforementioned rationalisable model of consideration set costs, where the consideration set is the set of goods with positive consumption. If there is cost of consideration, then it is as if the consumer chooses their bundle with a distorted price which is higher than the observed price; this is the source of irrationality that the model implies. If the agent does not have a cost of consideration, then they are rational and their $\boldsymbol{\beta}$ vector is zero.

It is important to note that there are specific restrictions on the $\boldsymbol{\beta}$ vector. In a sense, the costs of consideration can provide a range on how irrational a consumer can be. Without putting additional restrictions on the cost of consideration, it would always be possible to rationalise all behaviour as allowing consideration sets to grow arbitrarily large ensures the Afriat inequalities hold in the correct direction. In some sense, by putting these additional restrictions in place, it is possible to classify agents as completely rational, partially irrational, and completely irrational, where being completely irrational is an agent whose choices cannot be made consistent with the model of consideration set costs. This is discussed further in the estimation and implementation sections.

2.3.3 Estimation

Given a dataset of prices and bundle choices, for each individual, the solution to the following linear program[13] gives estimates for utility levels (u_t), the marginal utility of income (λ), and the costs of consideration (β).

$$\min \mathbf{S} \tag{2.23}$$

subject to

$$u_j = u_i + \lambda_i \left[\mathbf{p}_i' + \boldsymbol{\beta}_i' c'(\mathbf{q})\right] * (q_j - q_i) - S_{ij} + \mathbf{S}^{14} \tag{2.24}$$

$$\mathbf{S}, S_{ij} > 0 \tag{2.25}$$

$$\mathbf{q} > \theta_2 \tag{2.26}$$

[13]The main idea for the linear program comes directly from Fleissig and Whitney (2005)

[14]If $\mathbf{S} = 0$ and $S_{ij} \geq 0$, the inequalities hold in the correct direction. However, if $\mathbf{S} > 0$, this is a definite violation of rationality as the inequality, for sure, will not hold in the correct direction.

$$\boldsymbol{\beta}_i \leq \mathbf{p}_i \tag{2.27}$$

$$\lambda_i \geq 1 \tag{2.28}$$

where $c'(\mathbf{q})$ is given by:

$$\frac{-\theta_3 {\theta_1}^{\theta_3} \left[tan^{-1}\left(\dfrac{1}{x+\theta_2} \right) \right]^{\theta_3 - 1}}{\left[1 + \left(\dfrac{1}{\mathbf{q}+\theta_2} \right)^2 \right] [\mathbf{q}+\theta_2]^2} \tag{2.29}$$

Hence, this is a linear programming problem with $(T+1)^2 + n * T$ variables.

For the system above, it is only *relative* utility is what is identified by the program, in the sense that it would be possible to set u_t to an arbitrary constant and all other utility levels would adjust. Also, the constraint on $\lambda \geq 1$ is required as linear systems require weak inequalities to work, however, the KKT-conditions would require $\lambda > 0$. Given that the Afriat inequalities with consideration are still homogenous in U and λ, this constraint on λ is sufficient to resolve this issue of strict inequality.

Additionally, note that the specific restrictions on the $\boldsymbol{\beta}$ vector are, in some sense, restrictions on the rationality of the consumer. If the cost of consideration were to exceed the actual cost of a good, it would definitely not be rational for the consumer to include that good in the consideration set. When it comes to implementing the linear program, I solve the linear programs for all levels of β via a grid search and then reject rationality if any of the βs are too high, as according to equation (2.27). Using this method allows the program to find the maximum β at which the consumer can be just considered rational for any good. The maximum β serves as the lower bound on the set of identified β's for which the consumer is rational, with prices being the upper bound.

2.4 Implementation

As in chapter 1, the Stanford Basket Dataset is used for the analysis in chapter 2; specifically, a description can be found in section 1.4.1

2.4.1 Results

2.4.1.1 Cost of Consideration

The following analysis of the consideration set cost are in terms of a lower bound. The reason for this is due to the rationality constraint. If there are $\beta_k > p_k$, then this suggests there should be zero consumption for that good. As a result, that particular $\boldsymbol{\beta}$ is not identified by the linear program. As such, the estimated cost captures the lowest cost of consideration given the goods with positive consumption. Additionally, for computational ease, in the following analysis, it is assumed that $\boldsymbol{\beta}$ is constant across good and time (still heterogeneous). In terms of robustness checks, under all different scenarios[15], this assumption of estimating an 'average' β did not make any substantial difference to the estimation. The results were also robust to suitable parameterisation of θ_2 and θ_3, as well as different choices of smoothing functions[16], and different variations of the programming routine[17]. In practice, as long as θ_2 is smaller than the smallest quantity purchased of any good, there are no issues with this parameter. This is because Theorem 1 will hold as the cost function is convex for quantities larger thanθ_2). After some calibration, the changes in the results were minimal (and not noticeable) even when θ_3 was as low as 9. This is because the cost function drops back down to the asymptote sufficiently quickly to make immaterial differences in the estimation.

Recall from Theorem 1, that the cost of consideration can be seen as a price distortion, given a positive quantity of a good consumed. As the cost function is normalised appropriately, the coefficient on the coefficient measures the price

[15] Varying $\boldsymbol{\beta}$ across good and time and varying $\boldsymbol{\beta}$ only across time/good keeping it constant across good/time.

[16] Both convex and non-convex.

[17] Convex optimisation routines, mixed-integer programming, and different forms of linear program

distortion directly as a result of consideration. Figure 2.4 presents an average cost over goods and observations for each individual is shown. For example, if an individual has an estimated average cost of 2 ($\bar{\boldsymbol{\beta}}$=2), if the average price they faced was \$5, then the individual behaved as if they were faced with a price of \$7 (in order to rationalise their choices), in essence, it cost them an additional \$2 to consider the goods they bought. In figure 2.5, instead of plotting simply the average cost, the average cost as a proportion of expenditure is plotted. So in the above example, if the expenditure on a bundle of 10 goods was \$300, and the average cost of consideration was 2, then the total average cost of consideration would be $\$2 * 10$, hence as a proportion of observed expenditure, this would be $\frac{20}{300}$.

From figure 2.4, in ascending order, for roughly 100 of these households, the average cost of consideration is virtually 0 suggesting that those households are not constrained by consideration.[18] However, the remaining households appear to exhibit some use of consideration sets with positive $\bar{\boldsymbol{\beta}}$'s. There were only 2 households who failed to satisfy the rationality inequalities and for those households, their average $\bar{\boldsymbol{\beta}}$'s were extremely high suggesting some other forces of irrationality contrasting with around 80% of households who were found to be pure GARP violators.[19]

Figure 2.5 shows the cost of consideration as a proportion of expenditure. Simply put, it is the percentage of consideration cost relative to the amount actually spent. As per the figure, the highest proportional consideration cost is approximately 9%. This suggests that for this individual household, the cost of consideration (on average) is equivalent to 9% of what they spent.

2.4.1.2 Power and Prediction

The most obvious dimension of performance of revealed preference tests is the pass rate i.e. the proportion of households that can be rationalised within the specified

[18]Note that the graph looks stepwise in nature due to the grid search used in the estimation.

[19]In the sense that, when it comes to the revealed preference approach, there is (usually) only 1 contextual way to be rational, but this can be violated in many fashions.

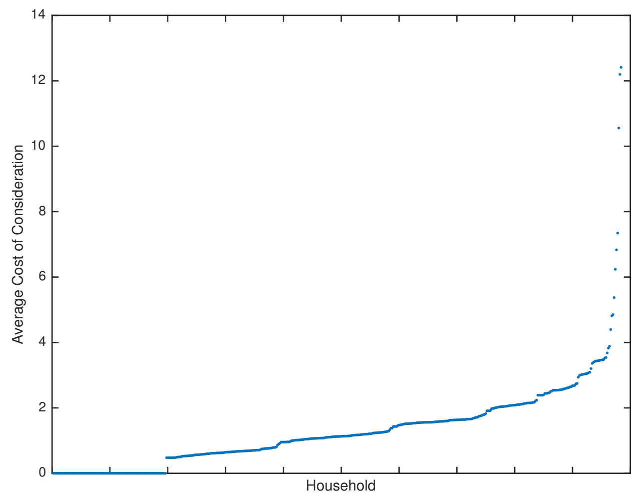

Figure 2.4: Average cost of consideration (β)

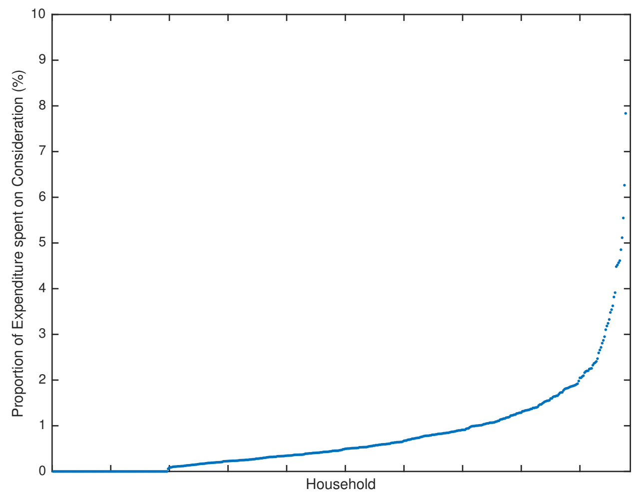

Figure 2.5: Average cost of consideration as a proportion of expenditure

model. The results here show that only 2 out of 494 households failed to be rational which suggests an almost perfect pass rate. Although this can be seen as a positive result from an empirical point of view, it is important to know how 'high' is a 'high' pass rate. This is where the concept of power needs to be introduced, in the sense of Bronars (1987). In order to check that the revealed preference test is actually able to discern between random and irrational behaviour, it would be wise to check whether the pass rate for random budget sets is low. This is the concept of a power measure. If this is the case, then this is good evidence that the revealed preference test is not simply 'passing' everyone but is able to tell irrational behaviour from random behaviour. For this particular setting and dataset, the power measure was approximately 90% which suggests roughly only 10% of random behaviour was deemed rational.[20]

The metric that is most commonly used combines the power measure and the pass rate, namely, predictive success. Although it is simply the pass rate minus (1 - power measure), this measure has an axiomatic backing (Selten (1991), Beatty and Crawford (2011)) as so is considered to be one of the most informative measures of success. In this case, the predictive success is around 90% which suggests that, with these data, there is strong evidence to suggest the use of consideration sets in the way that is explained in previous sections.

An online appendix shows that these results are extremely robust to the parameterisation of the consideration set constraint, the choice of smoothing functions, as well as different programming methods such as convex as well as non-convex optimisation routines.

2.4.1.3 Correlation with CCEI

One might argue that there could be a link between rationality as measured by the CCEI and the cost of consideration. Based on a simple correlation between the average βs and the average CCEIs for each household, there does appear to be a negative relationship. In this specific case, the correlation is approximately -0.32, somewhat suggesting that those who are more rational may be more likely to have lower consideration set costs. A correlation of -0.4 is achieved when using the proportional consideration set cost again providing more promising evidence for the potential link between more classically defined rationality and consideration sets. Additionally, for this particular dataset, the consumers are typically more rational than consumer from other scanner datasets which further elaborates the point that it would be perhaps unwise to separate consideration sets and rationality.[21]

[20]In the sense that random behaviour is considered completely irrational, this implies a 'perfect' model should have a power measure of 100%. This idea of randomness vs. irrationality originates from Becker (1962).

[21]This could be a topic of further interest but with only the use of the average β, future work will need to see how this correlation changes with a more elaborate estimation of β.

In a sense, this correlation is not altogether surprising given the interpretation of the costs of consideration and the rationality indices used. The cost of considering additional goods can be interpreted as a price distortion meaning that consumers are making bundle choices as if they are facing higher prices. From the economist's point of view, in order to rationalise the choices made, it is as if the consumer wastes income because they are choosing bundles as if they are more expensive than they actually are. Given that the 1-CCEI is a measure of how much income is wasted, this correlation corroborates the idea that consideration set costs could be a source of irrationality. In a sense, this price distortion leads to bundle choices that are not completely rational, but can be rationalised by the cost of consideration set model, but only up to a certain point.

2.4.1.4 Empirical Results

Table 1 in Section B.2 reports a set of regressions of the consideration set cost, both absolute and as a proportion of expenditure, on a selection of demographic variables from the Stanford Basket Dataset.[22] The demographic data collected were dummy variables for age[23], income[24], family size[25], and education[26].

In terms of the the rationality indices, both the CCEI and (1-Money Pump Index) (see Echenique et. al (2011) for more details), appear to be negatively associated with consideration set cost.[27] This was expected given the unconditional correlation calculated previously. The regressions seem to suggest that more irrational agents are typically those with higher consideration set costs which is consistent with the idea that rational agents are better at making decisions.[28]

The regressions seem to suggest that those with larger families, higher income, and more education are more susceptible to higher consideration cost. On one hand, this direction make sense if those types of agents are more constrained in

[22]Households with missing demographics were omitted from the regressions.
[23]OldAge > 65, MidAge $\in [30,65]$
[24]HighIncome $>$ \$45000, MidIncome $\in$ [\$20000,\$45000]
[25]LargeFamily > 4, MidFamily $\in [3,4]$
[26]High School, College; Average education across partners.
[27]The larger is 1-CCEI/MPI, the more irrational the agent.
[28]At least in terms of the rationality paradigm as defined by revealed preference theory.

terms of time. For example, the main shopper from the larger family may have less time to 'shop around' which may lead to a relative efficiency loss in terms of information resulting in a higher consideration set cost. Those with higher incomes (potentially as a result of more education) may also be time constrained in similar way, and simply choose their goods in a 'satisficing' manner, yet, resulting in a cost of consideration. On the other hand, it would not have been surprising to see negative coefficients on the education variable as it would be reasonable to believe that more educated individuals are more capable of making better decisions.

The coefficients on the age variables seem to suggest that older agents tend to make better decisions in terms of consideration cost. This may be because they are more experienced and have had more opportunities to improve their information set. Also, the oldest cohort may have more time to shop and thus make better decisions (as mentioned in the previous paragraph). This may be the case on average, but it is not unlikely that this does not apply in several specific cases.

2.5 Concluding remarks

Revealed preference analysis incorporating consideration sets naturally requires a less strict form of rationality than otherwise prescribed owing to the additional cost of consideration. With a unique contribution to the literature, consumers may only consider a subset of the goods available to them as the cost of consideration is too high. This is modelled via a modified budget constraint which incorporates, what is essentially, a fixed cost of consideration. By modifying the existing budget constraint, the cost of consideration can be interpreted in terms of expenditure rather than an arbitrary measure of cognitive ability. Alongside papers such as Demuynck and Seel (2018), this chapter attempts to bring forward the case of a broader and deeper theory of consideration sets.

Via a modified linear programming algorithm, I am able to semi-parametrically estimate the costs associated with consideration sets using revealed preference theory. With an application to a scanner dataset, for the vast majority of individuals, the model cannot reject the use of consideration sets in the presence of suitable restrictions with an estimated average consideration set cost of 2% of monthly expenditure. Interestingly, there also appears to be a strong link between consumer rationality and limited consideration.

There are clearly many more opportunities for research in this area such as applying the model to other datasets in which the consumers are (on average) less rational than the ones presented in this chapter. This would allow for a greater variation in consideration costs which would be ideal for some form of empirical work in terms of demographic factors. This could go some way towards explaining exactly how and why consideration costs can differ across consumers. Another interesting question would be to examine how the above analysis changes when decisions are made on a discrete basis rather than a continuous one, in the style of Polisson and Quah (2013). This would also facilitate the use of the originally proposed and most intuitive consideration set constraint.

Chapter 3

Choices with just-noticeable differences or status quo - Is there a (noticeable) difference?

3.1 Introduction

[1] Reference-dependent utility models of human behaviour have received ample notice from both theoretical and applied economists. One particular pattern of such behaviour is that of status quo bias (SQB), a term originally coined by Samuelson and Zeckhauser (1988), and a concept seemingly at odds with textbook definitions of economic rationality.[2] This is what provides the main motivation behind this chapter; I propose a choice theory that explains the presence of SQB, and its effects on choice behaviour, with the fundamental principle of just-noticable differences (JNDs).

First introduced into the economics literature by Luce (1956), a JND is the

[1]For this specific chapter, I am grateful to Ran Spiegler, Richard Blundell, Fernanda Senra de Moura, Rubén Poblete Cazenave, Ryan Kendall, and Terri Kneeland for critical stimulating discussions. I would also like to extend my gratitude, non-exhaustively and alphabetically, to Roland Benabou, Chris Chambers, Syngjoo Choi, Inga Deimen, Paweł Dziewulski, Kfir Eliaz, Andrew Ellis, William Fuchs, Gilat Levy, Yusufcan Masatlioglu, Stefania Minardi, Pietro Ortoleva, Thomas Pugh, Joao Ramos, Silvia Sarpietro, Roberto Serrano, Tomasz Strzalecki, Severine Toussaert, Christopher Tyson, Stephanie Wang, and seminar participants at FUR 2018 for invaluable suggestions and comments.

[2]In the sense that an agent can continually purchase an alternative/set of alternatives, that, from the economist's point of view, appears worse for the agent.

minimal stimulus required to be able to compare/perceive change.[3] A common and prevalent example comes from the marketing literature where adjustments in price are considered the most salient, but changes in, for example, the weight or size of a product are not as noticeable.[4] Firms, in turn, tend to exploit the fact that consumers are seemingly not perceiving these changes in order to further maximise profits. From the economics literature, the famous example that comes from Luce (1956), involving coffee and sugar: *If there are 401 cups of coffee in a row, each with an additional grain of sugar than the previous cup. In general, it will be impossible to tell the difference between 2 consecutive cups. However, cups that are not near each other will be easily comparable.* As is clear from this example, JNDs were introduced in order to explain intransitive indifference. However, more critically, these examples highlight that people are imperfect in their ability to compare with absolute precision. In utility terms, this is to say that x is preferred to y if $u(x) > u(y) + \epsilon$. So x is only preferred if it has higher utility than y *and* by the JND, ϵ.[5] In this chapter, I show that SQB is a major implication of imprecise perceptibility.

In this setting, SQB comes from an inclination to choose a default option/current choice when decision-making. By relaxing completeness[6] of the preferences, Masatlioglu and Ok (2005) are able to represent utility as a vector.[7] Using this, they establish a choice procedure such that the status quo initially eliminates any alternatives that are worse than it (in any number of dimension of utility), otherwise, the status quo remains chosen. Note that, without this vector representation in the first stage (and with perfect discrimination), there is no role for the status quo, as the choice problem simply becomes standard. In other words, in the case where there is no JND utility and just single-valued utility, then the status quo and the choice procedure become trivial as the status quo will always/never be chosen if it is the best/worst alternative. Thus, in the setting of Masatlioglu

[3]Introduced by psychophysicists, Weber (1834), Fechner (1860).

[4]http://www.bbc.co.uk/news/business-40703866 provide a whole list of goods, for example, Toblerone chocolate, McVities biscuits etc..

[5]If $|u(x) - u(y)| < \epsilon$, then x and y are said to be indistinguishable from each other.

[6]A binary relation, R, on a set A, is complete if either $a \; R \; b$ or $b \; R \; a$, for all a, b in A

[7]The different dimensions can be thought of as different characteristics of an alternative.

and Ok (2005) choice behaviour without multi-valued utility would be observationally equivalent to standard utility and thus would not explain SQB. It is the relaxation of completeness in the first stage that is vital in providing the vector representation, and thus ensuring the status quo is able to actually effect choice behaviour.[8] Masatlioglu and Ok (2005) is one of the most important and influential papers in the literature of axiomatising utility functions that are consistent with behaviours that are considered explainable by classical reference-dependence utility (e.g. Tversky and Kahneman (1991, 1992), Kőszegi and Rabin (2006)).

In parallel with Masatlioglu and Ok (2005), the chapter provides a comprehensive understanding of SQB with an axiomatic approach to its choice procedure and utility representation. The axioms presented here will be descriptive of choice behaviour that should exhibit SQB, but will lead to a familiar utility representation. It is vital to point out that the representation offered by Masatlioglu and Ok (2005) should be regarded as a specific example of status quo bias, and how a status quo can effect choice behaviour. As such, I view their representation as a special case of a more general notion of status quo bias, as according to the literature in behavioural economics. I achieve this generalisation with the JND utility representation. This is illustrated further with examples below, involving both examples of different choice behaviours, and stylised utility representations. The work I present in this chapter should be seen as an accompaniment to the literature by offering a generalised representation of the idea of SQB with the classic JND utility representation, thus explaining a behavioural phenomenon with a longstanding adage of economic theory.

Conceptually, SQB is an interesting behavioural phenomenon as it cannot be typically explained by standard economic arguments. This is to say that, an economist might observe (choice) behaviour that is not consistent with standard utility models. For example, an individual at a supermarket may simply overlook other products in favour of a good that they have already purchased, even if those products may be better (in some, or all dimensions). A reasonable story (of many)

[8]Common examples: Current job, average consumption, previously purchased bundle, default choice (opt-out/in), etc...

for this would be that the individual is biased toward that which they already know, hence the term, status quo bias. An alternative story, that would lead to very similar (if not, the same) choice behaviour is one in which the individual does consider the other products but does not shift away from the choice that they are comfortable with unless there is another alternative which is substantially better than their status quo (again, potentially, in some, or all dimensions). It is the latter interpretation of this choice behaviour that further motivates this chapter and will be what drives the following JND utility representation. Another common example comes from a longstanding phenomenon in the UK where people seem to stick with their current accounts despite the fact it is (essentially) costless and even incentivised (free gifts etc...). It may be that agents do not perceive the additional benefits from switching, and therefore decide to stick with their current choice of bank.[9]

Furthermore, this kind of utility representation may also explain many other kinds of behaviour (e.g. intransitive indifference) that are not necessarily identifiable as consistent with SQB, but may be potentially rationalised by a JND representation. Equivalently, choice behaviour that is consistent with SQB may always be represented by JND utility, but, not all cases of choices with JND representations can be applied to situations with SQB. In what follows, the SQB utility representation of Masatlioglu and Ok (2005) can be derived from the theorem that yields JND utility representation, but not always conversely. This comes as a natural consequence of the way that the main SQB axioms have been formulated.

An additional interesting result can be shown. Suppose further that a good or product can be described completely as a vector of characteristics. This means that comparisons between alternatives are done pair-wise for each characteristic. In this case, it is also possible to explain SQB behaviour with JNDs. As per the story above, if an individual does not move away from their current choice, it may be because there are no other alternatives that are noticeably better in all the characteristics. This gives rise to an interesting question of how to make choices

[9]https://www.bbc.co.uk/news/business-44522630 - gives a brief overview of the analysis done by the Competition and Markets Authority.

over goods when comparisons over characteristics are done with imperceptibility. Provided the agents aggregate the characteristics consistently, it is possible to show that agents can have either perfect or imperfect perceptibility over their final choice of goods. This is explained in much further detail in Appendix C.2.

The chapter is organised in the following way: the remainder of Section 3.1 provides examples of how SQB behaviour can be explained by a JND representation. Section 3.2 introduces the main axioms and representation of the paper.[10] Section 3.3 contains a brief literature review with Section 3.4 concluding. All proofs can be found in the accompanying appendix.

3.1.1 Example Representations

The following examples describe what can be predicted of choice behaviours in a framework with status quo in terms of the Masatlioglu and Ok (2005) representation and contrasts those with what is prescribed by the representation of this chapter vis-à-vis JND utility.

3.1.1.1 Comparisons with representation from Masatlioglu and Ok (2005)

The following two examples of SQB using Masatlioglu and Ok (2005) appeal to a vector representation of utility. An interpretation of this is that each component of the utility vector represents a characteristic/feature of an alternative.[11]

Example 1a Suppose $X \equiv \{a, b, c, d, x\}$, $\mathbf{u}(X) \subset \mathbb{R}^2$, with $\mathbf{u}(x) = (2, 1)$, $\mathbf{u}(a) = (4, 8)$, $\mathbf{u}(b) = (7, 3)$, $\mathbf{u}(c) = (1, 4)$, $\mathbf{u}(d) = (4, 0.5)$, with x being the status quo.[12]

[10] Section C.2 presents intuition and a plausible explanation on how to aggregate preferences over (a finite number of) characteristics of goods, if those characteristics are governed by JND preferences.

[11] In this sense, in order to explain SQB, there is an appeal to relaxing completeness to get the vector utility representation.

[12] Note that the utilities are vectors to draw an easier comparison with the examples from Masatlioglu and Ok (2005). However, in these examples, what is actually being compared with JND utilities is the post-aggregation utilities of Masatlioglu and Ok (2005).

Under Masatlioglu and Ok (2005), the status quo, x, would do the job of eliminating c and d as x is better than c and d in some dimension of utility. The agent is then posed with the choice problem with just a and b which the agent can easily evaluate with some aggregator function mapping the multi-valued utility back to $\mathbb{R}$. For example, suppose the aggregator is the product of the elements i.e. $\mathbf{g}(\mathbf{u}(x)) = 2$, $\mathbf{g}(\mathbf{u}(a)) = 32$, and $\mathbf{g}(\mathbf{u}(b)) = 21$, so a is chosen.[13]

Example 1b Now suppose that $X \equiv \{e, f, g, h, x\}$ with $\mathbf{u}(x) = (4, 4)$, $\mathbf{u}(e) = (1, 6)$, $\mathbf{u}(f) = (5, 5)$, $\mathbf{u}(g) = (1, 2)$, $\mathbf{u}(h) = (4.6, 3.6)$, with x being the status quo. From Masatlioglu and Ok (2005), e, h, and g are eliminated by the status quo, and, with the same aggregator as before, f is chosen. This highlights an interesting feature of their model. Alternative h is not too dissimilar to the status quo, in fact, it is only marginally worse (and better) in one dimension of utility. However, their choice procedure is such that the status quo eliminates any alternative worse than it in any dimension of utility, even if it is better in all others. In essence, even if $\mathbf{u}(h) = (500, 3.9)$, it would still be eliminated.

3.1.1.2 Example JND representations of SQB

Example 2a Suppose $X \equiv \{a, b, c, d, x\}$, with $\mathbf{v}(x) = 2$, $\mathbf{v}(a) = 32$, $\mathbf{v}(b) = 21$, $\mathbf{v}(c) = 4$, $\mathbf{v}(d) = 2$, with x being the status quo.

Recall that JND utility is such that alternative p is (strictly) preferred to alternative q when $v(p) > v(q) + \epsilon$. Even in the case where ϵ is large and $\epsilon < 30$, the status quo is eliminated by a as it would be noticeably better. If $\epsilon \in [0, 11)$, then alternative a is the unique choice. This example mirrors Example 1a above, however it did not appeal to a vector utility representation with the status quo eliminating alternatives in a first stage of a choice procedure. Rather, the JND utility simply choses the maximal alternative, in a noticeable sense, without the need to decompose the utility into characteristics.

For clarity of comparison, the aggregator function used in Example 1a corroborates exactly with the utilities in Example 2a. Thus, both Example 1a and

[13] The final choice of alternative is dependent on the aggregator function.

Example 2a show that the Masatlioglu and Ok (2005) representation and JND utility representation can rationalise the same choices.

Example 2b Now suppose that $X \equiv \{e, f, g, h, x\}$ with $\mathbf{v}(x) = 16$, $\mathbf{v}(e) = 6$, $\mathbf{u}(f) = 25$, $\mathbf{u}(g) = 2$, $\mathbf{u}(h) = 16.56$, with x being the status quo.
Again, in direct comparison with Example 1b, instead of dealing with vector utilities, this example deals directly with the aggregated utilities. There are two interesting choice behaviours here that are different from Masatlioglu and Ok (2005). Firstly, alternative h is not eliminated by the status quo as the status quo can never be the strictly preferred option. In fact, h can eliminate the status quo with a small enough JND. Secondly, if $\epsilon \in (8.44, 9)$, not only is h not eliminated by it can also be chosen as it would not be noticeably worse than (the objectively best) alternative f.

This example shows that there are sensible choice behaviours that can never be explained by Masatlioglu and Ok (2005) but are rationalisable by a JND utility function. From the economist's point of view, the JND utility representation allows agents to choose an objectively worse alternative. The explanation for this being that agents may not be particularly concerned by marginal improvements to their current choice.

3.2 Model

3.2.1 Primer

To ensure that the following representations are clear, this section will provide a brief introduction to semiorders and interval orders, and their respective representations. I refer the interested reader to Luce (1956) and Fishburn (1975/85) for further details.

Let R denote a generic binary relation on a finite set A, meaning that R is a subset of $A \times A$. In essence, the binary relation is a set of ordered pairs of elements from A. If an ordered pair, say, (x, y) belongs to R, this is written as xRy. If

x, y do not belong to R, this is written as $x \neg R y$ The relevant properties of binary relations that are required for this chapter are:

Reflexivity: $a \; R \; a$

Irreflexivity: $a \; \neg R \; a$

Transitivity: $[a \; R \; b$ and $b \; R \; c] \implies a \; R \; c$

Semitransitivity: $[a \; R \; b$ and $b \; R \; c] \implies [a \; R \; d$ or $d \; R \; c]$

Interval Order Condition: $[a \; R \; b$ and $c \; R \; d] \implies [a \; R \; d$ or $c \; R \; b]$

for all $a, b, c, d \in A$.

A binary relation, $\succ_I$, is an interval order if it satisfies irreflexivity, and the interval order condition. It is a standard result that if $\succ_I$ is an interval order on A, then there exists a function $g : A \to \mathbb{R}$, and a positive threshold function $\delta : A \to \mathbb{R}^+$ such that, for all $x, y \in A$, then $x \succ_I y \iff g(x) \geq g(y) + \delta(y)$. The threshold function $\delta(.)$ is what determines the (utility) level beyond which alternatives can be compared. Similarly, in the language of preferences, we can say that x is preferred to y if it is "substantially better" than y, where the threshold for being "substantially better" is a function of the alternatives. As above, in the absence of a strict preference, i.e. neither $x \succ_I y$ nor $y \succ_I x$, then x and y are not perceptibly different from each other.

A binary relation, $\succ_S$, is a semiorder if it satisfies irreflexivity, semitransitivity, and the interval order condition, thus implying that a semiorder is a semitransitive interval order. It has been shown that if $\succ_S$ is semiorder on A, then there exists a function $g : A \to \mathbb{R}$, and a constant $\epsilon > 0$ such that, for all $x, y \in A$, then $x \succ_S y \iff g(x) \geq g(y) + \epsilon$. This constant ϵ is what is commonly referred to as the just-noticeable difference. In the language of preferences, we can say that x is strictly preferred to y if it "substantially better" than y. In the absence of a strict preference, i.e. neither $x \succ_S y$ nor $y \succ_S x$, then x and y are considered indistinguishable from each other.

Both of these notions and representations will play a large role in the representation theorems to follow.

3.2.2 Axioms

Let X be a non-empty finite set where each element of X is an alternative i.e. the space of all alternatives. For later exposition, let $\diamond$ denote an element such that $\diamond \notin X$. $x \in \{X \cup \diamond\}$ is then a generic element of $X \cup \diamond$. Let $\bar{S} \subset X$ denote a generic non-empty subset of X which is augmented with x such that S is $\bar{S} \cup \{x\}$. So S is the choice set; this is simply a subset of X which also includes x. A choice problem is denoted as (S,x) such that when $x = \diamond$, then this is a problem without a status quo, and when $x \in S$, this is a choice problem with status quo.

Define a choice correspondence $c(\cdot, \cdot)$ such that, for all (S, x), $c(S, x) \subseteq S$ i.e. the choice correspondence is a mapping from the choice problem back to some (non-empty) subset of itself.[14] The following are conditions on the choice correspondence that are consistent with status quo bias and a JND representation.

Axiom $\boldsymbol{\alpha}$ For any $(A, x), (B, x)$, if $y \in B \subseteq A$ and $y \in c(A, x)$
$\implies y \in c(B, x)$

Axiom $\boldsymbol{\delta}$ For any (A, x), if $z, y \in c(A, x)$, and $A \subseteq B$
$\implies \{y\} \neq c(B, x)$

These are simply Sen's (1971) α and δ properties with the added notation of status quo.[15] Axiom α requires that an alternative remains chosen even after other alternatives are removed. Axiom δ requires that alternatives chosen before a choice set is expanded are not uniquely chosen after the choice set is expanded.

Consider the following scenario (which will be carried throughout the chapter) involving Mr. Smith, who is a food critic. When Mr. Smith is travelling, his status quo cuisine depends on the country he is visiting, but when Mr. Smith is at home, he does not have a status quo cuisine.

To illustrate Axiom α, suppose Mr. Smith has to choose from *German, Italian* and *French* cuisine and he chooses *Italian*. If *German or French* cuisine become

[14] The 2nd argument just makes explicit whether it is a status quo choice problem or otherwise

[15] Sen's α property is also referred to as the *contraction* axioms, referring to the concept of the independence of irrelevant alternatives.

unavailable, then Mr. Smith continues choosing *Italian*. So, in the presence of fewer options of cuisine, Mr. Smith must still continue choosing the cuisine he chose before, given they are still available. This is true for whatever his status quo is.

To illustrate Axiom δ, suppose Mr. Smith has to choose from *German* and *Italian* cuisine and he chooses both *German* and *Italian*. If *French* cuisine also becomes available, it cannot be the case that *German* food is uniquely chosen. Again, this is true for regardless of his status quo.

Axiom D (Dominance) For any (A, x), if $\{y\} = c(A, x)$, $A \subseteq B$ and $y \in c(B, \diamond)$
$\implies y \in c(B, x)$

Axiom D and Axiom SQI come directly from Masatlioglu and Ok (2005). Axiom D is referred to as the 'Dominance' axiom. The intuition being that if an alternative is considered (strictly) better than the status quo, then, in a bigger choice problem without status quo where that alternative is still chosen, it must remain chosen when there is a status quo.

To illustrate Axiom D, suppose Mr. Smith is visiting Germany, where *German* cuisine is his status quo. However, as *Italian* cuisine is also available, he decides not to eat *German* food in favour of the *Italian*. When back home (i.e. no status quo cuisine), if Mr. Smith is presented with options of *Italian*, *German*, and *French* cuisine and chooses *Italian*, then when he returns to Germany, it would seem sensible that he would still choose *Italian* even if *French* cuisine is also available.

Axiom SQI (Status Quo Irrelevance) For any (A, x), if $y \in c(A, x)$ and there does not exist a $B \subseteq A$ with $B \neq \{x\}$ and $x \in c(B, x)$
$\implies y \in c(A, \diamond)$

Axiom SQI is an 'inapplicability' property. If an alternative is chosen, and there is not a subset of that choice problem where the status quo is chosen, then the status quo played could not have played a role in the original choice problem.

To illustrate Axiom SQI, suppose again that Mr. Smith is visiting Germany (status quo cuisine is *German*). Mr. Smith has a choice from *German*, *French*, *Italian*, and *Spanish* cuisine, and he chooses *Italian*. However, on other days, although *German* cuisine is always available, some of the cuisines are not. On those days, if Mr. Smith never chooses *German* cuisine (unless it is the only one available), then it is as if Mr. Smith was choosing cuisine at home i.e. with no status quo cuisine.

Axiom SQB For any (A, x), if $y \in c(A, x)$

$$\implies \{y\} = c(A, y)$$

Also from Masatlioglu and Ok (2005), Axiom SQB says that if an alternative is chosen when it was not the status quo, then it is the only alternative when it becomes the status quo. The idea being that there is absolutely no need to move away from the status quo given it was already shown to be a preferred choice.

To illustrate Axiom SQB, suppose again that Mr. Smith is visiting Germany (status quo cuisine is *German*). Mr. Smith decides to eat *Italian* food when his choices are *Italian*, *French*, and *German* cuisine. When Mr. Smith travels to Italy, and is offered the same choices, he refuses all other cuisine apart from *Italian*. So, when the status quo cuisine changed to one that Mr. Smith had previously chosen, he decides to stick with that and not try any other cuisine.

The following two axioms are important generalisations of Axiom SQB. By using the following two axioms instead of Axiom SQB, it will be possible to derive JND utility as a representation of SQB.

Axiom ASQB (Augmented Status Quo Bias) For any (A, x), there exists a B (with $A \cap B = \emptyset$) such that,

if $y \in c(A, x)$

then, $y \in c(A \cup B, y)$ and $a \notin c(A \cup B, y)$, $\forall a \in A \setminus \{y\}$

Axiom ASQB describes the role that a status quo can play when a choice set becomes larger. This axiom is what yields higher regard to the status quo. If y is revealed preferred in a choice set with status quo x, and y is also revealed preferred in a “choice superset”, then when y is the status quo, only y and the additional elements from the superset are choosable. It is in this sense that there is some disposition for y which gives it a higher regard as y is the only element from the original choice problem that remains chosen. However, there is not always necessarily a complete tendency towards y given it is not the uniquely chosen element. Unlike the previous axioms, Axiom ASQB introduces a rationality concession in order to accommodate status quo bias behaviour; although a restrained form of rationality, it is still permissible, nonetheless.

To illustrate Axiom ASQB, suppose again that Mr. Smith is visiting Germany (status quo cuisine is *German*). Mr. Smith decides to go to one particular street to eat where he can choose from *German*, *Italian*, and *French* cuisines, and he decides on *Italian*. When Mr. Smith travels to Italy (where the status quo cuisine is *Italian*) and he can eat the same cuisine as before, as well as *Spanish* and *Greek* cuisine, then Mr. Smith continues choosing *Italian* food. Additionally, Mr. Smith will not try any other cuisine except *Spanish* and/or *Greek* cuisine as they are newly available.[16]

From this example, Axiom SQB is violated if *Italian* food is not uniquely chosen. If *Spanish* and/or *Greek* cuisines are also chosen, then Axiom SQB is violated as it requires that *Italian* food is chosen uniquely. So, in some sense, Axiom ASQB potentially allows cuisines that are as good as/similar to *Italian* to be chosen. Note that if *Italian* food is not chosen after it becomes the status quo, this does not describe behaviour that is consistent with Mr. Smith being status quo biased. In essence, Mr. Smith has not shown an inclination towards the status

[16]To see how this is a generalisation of Axiom SQB, if B is empty, then the only choosable element is $\{y\}$ when $\{y\}$ becomes the status quo. In the example, this means that when Mr. Smith was in Italy, he had the same choice as when he was in Germany, and therefore, only chooses *Italian* food.

quo cuisine and as such, both Axiom SQB and Axiom ASQB are violated.

Axiom DSQB (Distinct Status Quo Bias) For any (A, x) (B, w), (with $A \cap B = \emptyset$),

if $y \in c(A, x)$ and $v \in c(B, w)$,

then

$$\begin{cases} y \in c(A \cup B, y) \text{ and } v \notin c(A \cup B, y) \text{ or} \\ v \in c(A \cup B, v) \text{ and } y \notin c(A \cup B, v) \end{cases}$$

Axiom DSQB describes the role of status quos when combining distinct choice problems. Axiom DSQB says that, when two different choice problems are combined, only choices from one of the individual choice problems can have the higher status quo regard. Suppose y & v are revealed preferred in their respective choice sets with status quos x & w. The axiom posits that y is also revealed preferred in the combination of the choice sets when y becomes the status quo. However, the choices that were originally made from B are no longer chosen, giving a higher regard for the status quo from only one choice problem (or vice versa, if v becomes the status quo). As with Axiom ASQB, this is an axiom that describes a higher regard for the status quo but in a conservatively rational way

To illustrate Axiom DSQB, suppose Mr. Smith travels to France (status quo cuisine is *French*) where he chooses *Italian* cuisine from a choice of *Italian* and *French* cuisine. The following day, Mr. Smith goes to the *Spain* (status quo cuisine is *Spanish*) to buy *Greek* cuisine when only *Spanish* and *Greek* food are available. The following day, Mr. Smith travels to *Italy* where he has a choice from *Italian*, *French*, *Spanish*, and *Greek* cuisines. Mr. Smith continues choosing *Italian* cuisine when everything is available, but decides against having *Greek* food.[17,18]

Again, from this example, Axiom SQB is violated if *Italian* food is not uniquely chosen. If *Greek* cuisine is also chosen, then Axiom SQB is violated as it requires that *Italian* food is chosen uniquely. Thus, Axiom DSQB can allow for cuisines that are as good as/similar to *Italian* to be chosen, when it becomes the status quo. If *Italian* food is not chosen after it becomes the status quo, or, *Greek* food is chosen, this does not describe behaviour that is consistent with having a status quo. For instance, if Mr. Smith really enjoys both *Italian* and *Greek* cuisines, but still continues to choose *Greek* food when *Italian* is the status quo, then Mr. Smith has not shown an inclination towards *Italian* food over the previously chosen *Greek* cuisine i.e. he has not been biased by the status quo cuisine. As such, both Axiom SQB and Axiom DSQB would be violated if *Italian* food is not chosen, and Axiom SQB violated if anything but *Italian* food is chosen.

3.2.3 Main Representation

For some generic nonempty finite set, W, and status quo, w, define the following:

$\Gamma_{v,\epsilon}(W,w) \equiv \{p \in W : v(p) > v(w) + \epsilon\}$

$\Pi_{v,\epsilon}(W,w) \equiv \{p \in W : \big|v(p) - v(w)\big| \leq \epsilon\}$

$\Gamma_{v,\epsilon}(W,w)$ is the set of alternatives that are noticeably better than the status quo whereas $\Pi_{v,\epsilon}(W,w)$ contains the alternatives that are not noticeably better or worse than the status quo.

$\Omega_{v,\epsilon}(W) \equiv \{q \in W : v(p) > v(q) + \epsilon \text{ for no } p \in W\}$

$\Omega_{v,\epsilon}(W)$ is the set of maximal in elements insofar as this is the set of alternatives for which there are no other alternatives that are noticeably better.

[17] Note that, if Mr. Smith had decided to go Greece (status quo cuisine is *Greek*), and had to choose from *Italian*, *French*, *Spanish*, and *Greek* cuisines, he would choose *Greek* cuisine and would decide against *Italian*.

[18] To see how this is a generalisation of Axiom SQB, if B is empty, then there is no other choice problem and $\{y\}$ is still chosen from the original choice set, when $\{y\}$ becomes the status quo. If $\{y\}$ is uniquely chosen, then it reduces to Axiom SQB. To guarantee this, both Axiom DSQB and Axiom ASQB combined with empty B imply Axiom SQB. In the current example, this means that Mr. Smith continues choosing *Italian* cuisine, when *Italian* cuisine becomes the status quo.

Theorem 2. *Let S be a nonempty finite set. A choice correspondence $c(\cdot, \cdot)$ satisfies Axiom α, Axiom δ, Axiom D, Axiom SQI, Axiom ASQB, and Axiom DSQB, if, and only if, there exists a $\epsilon \in \mathbb{R}$, a function $v : S \to \mathbb{R}$, such that*

$$c(S, \diamond) = \Omega_{v,\epsilon}(S)$$

and

$$c(S, x) = \begin{cases} \Pi_{v,\epsilon}(S, x) & \Gamma_{v,\epsilon}(S, x) = \emptyset \\ \Omega_{v,\epsilon}(\Gamma_{v,\epsilon}(S, x)) & \textit{otherwise} \end{cases}$$

The theorem says that an agent chooses the best noticeable element or the set of elements that are not noticeably different from the best element. This applies to all elements including the status quo, but, there may exist some element/s which is/are strictly but not noticeably better than the status quo, hence the status quo bias.

Consider the choice problem where there is a status quo where the status quo is the current alternative of the agent. If there do not exist any noticeably better alternatives, the current choice remains chosen among the other alternatives that are not noticeably better/worse than the status quo. For example, let us consider the ubiquitous problem of buying a mobile phone.[19] Suppose a Huawei phone user is faced with a new choice between Huawei, Samsung, HTC, and Sony phones, all of which operate using the Android operating system. If the user deems the HTC and Sony phones absolutely superior to her current choice of Huawei and the Samsung phone, she will decide over HTC and Sony phones. However, if the HTC and Sony phones are noticeably worse, but the Samsung is not not noticeably worse or better, she can continue choosing the Huawei or even switch to the Samsung phone. So the user definitely moves away from her current choice if there is a phone substantively better, otherwise, she continues using her current phone (or something similar enough).

[19]Recall that brand loyalty is a form of status quo bias. See Samuelson and Zeckhauser (1988) for further examples and details.

Now consider the choice problem where there is seemingly no status quo. In this case, the agent simply chooses whichever alternative is noticeably best. In essence, the agent simply chooses from a set of alternatives that are not noticeably worse than any other alternative. As a representation of choice behaviour, the JND utility has strong appeal from a psychological perspective, as such, it forms the basis of behaviour even without a status quo alternative. The only real difference between this and a choice problem with status quo is that the status quo is a current choice, in essence, a reference point. However, without this reference point, the agent still maintains their basic choice behaviour as represented by JND utility. Given this, essentially, the representation above always gives rise to the risk of there being SQB behaviour.

Taking this from another perspective, a natural question could be, why does the agent not have a standard utility function in the absence of a status quo? As written above, the representation of behaviour that allows for some level of inexact perception allows a much greater degree of realism. More specifically, in the standard utility case, there is no real role for a status quo, as explained in the introduction. From a modelling point of view, it makes little sense to have the agent's decision making represented by JND utility simply because they have a status quo, and have a standard utility function when there is no reference point. In a sense, the agent can always be susceptible to status quo bias and thus should have JND utility with and without a defined status quo. Given that perfect discrimination can be extremely tricky even in the simplest situations, as supported by the psychology literature, and physics literature, it is not unreasonable to believe that simple choice behaviour can be represented by JND utility. However, as is the purpose of this chapter, having such a representation does facilitate in explaining behavioural phenomenon such as SQB, especially given that the axioms are descriptive of SQB.

From above, it is clear that $c(S, x) \subseteq c(S, \diamond)$ i.e. choice problems with a status quo can always be derived from a choice problem without a status quo due to JND utility. However, from observable choice data, it might not be the case that all choice behaviour consistent with JND utility can be necessarily rationalised by status quo bias. With this representation, the agents still appear to be status quo biased if there does not exist any alternatives that are noticeably better than the status quo i.e. the agent continues to choose the status quo unless another alternative is noticeably better. From the above representation, it is clear that agents may choose an alternative that might be worse for them, but not noticeably so.[20] This is corroborated by the following:

Implication 2.1. *Let X be a nonempty finite set. Theorem 1 of Masatlioglu and Ok (2005) holds if a choice correspondence $c(\cdot,\ \cdot)$ satisfies Axiom α, Axiom δ, Axiom D, Axiom SQI, Axiom ASQB, and Axiom DSQB.*

Interestingly, this implication essentially says that, under a slightly more general choice structure, the strong representation of SQB given by Masatlioglu and Ok (2005) can be seen as a special case of a JND representation.[21] In essence, it is possible to get the representation of Masatlioglu and Ok (2005), but it is not possible to get the above axioms from their representation.

To illustrate this point, returning to Examples 1b and 2b, suppose that $X \equiv \{e, f, g, h, x\}$ with $\mathbf{u}(x) = (4, 4)$, $\mathbf{u}(f) = (5, 5)$, $\mathbf{u}(g) = (1, 2)$, $\mathbf{u}(h) = (4.6, 1)$, with x being the status quo. In this case, the choice procedure of Masatlioglu and Ok (2005) eliminates g and h, and f is chosen. As long as these vectors are aggregated with a strictly increasing function, irrespective of any level of imperception (i.e. for any JND value), f will always remain in the maximal set, and thus is always chosen. Bringing this to the cuisines example, if Mr. Smith decides to go to one particular street to eat where he can choose from *German*, *Italian*, and *French* cuisines, and he decides on *Italian*, he must continue choosing *Italian* if it becomes the status quo cuisine. If no other cuisines become available when the status quo

[20] A stark example: if $v(c) = 4, v(d) = 6, v(e) = 6.5, v(f) = 7$, with $\epsilon = 1$, then a ordinary choice problem would choose e and f. In a status quo choice problem, if the status quo is e, then the agent can continue to choose e as there is nothing noticeably better. This is also true if the status quo is f. Conditional on observing e or f, this observed choice could have come from a choice problem without status quo. Note that, if the status quo was c, it would not remain the status quo as there are alternatives noticeably better.

[21] As far as is known, no link between incompleteness and intransitive indifference has been made in the literature. It may be the case that there is some link between these properties that has yet to be discovered, at least, beyond SQB.

changes, then *Italian* cuisine still continues being chosen.

To further illustrate this implication, suppose that a new alternative becomes available such that $\mathbf{u}(y) = (5 + \epsilon_1, 5 - \hat{\epsilon}_2)$, with the new status quo being f. The choice procedure of Masatlioglu and Ok (2005) eliminates x, g, h, and y, with f being chosen as there is no alternative that beats the status quo in both dimensions. So here, y is eliminated even if barely differs from the status quo. However, there must exist a (class of) strictly increasing function(s), $v : \mathbb{R}^2 \to \mathbb{R}$ such that $v(\mathbf{u}(y)) > v(\mathbf{u}(f))$. If it is the case that $v(\mathbf{u}(y)) > v(\mathbf{u}(f)) + \epsilon$, i.e. the new alternative is noticeably better than the status quo, then y is chosen, and the previous status quo is no longer chosen. This represents an individual who made a choice as if they did not have a status quo, and simply just chose the maximal alternative. However, if $|v(\mathbf{u}(y)) - v(\mathbf{u}(f))| < \epsilon$, although the new alternative is technically better, it is not noticeably better, thus, both f and y are chosen. In essence, it is possible for new alternatives to be similar enough to the current status quo such that they are choosable. Bringing this back to the cuisine example, if Mr. Smith travels to Italy where *Italian* is the status quo cuisine, and *Spanish* cuisine is newly available, if *Spanish* is similar enough to the *Italian* cuisine, then it can also be chosen. This highlights that both representations predict that the status quo cuisine is chosen, however, Masatlioglu and Ok (2005) would not be able to predict that anything similar is chosen. This is a result of their first stage representation of SQB which would eliminate any alternative even if it is marginally worse in any dimension. However, the representation of this chapter does allow for other alternatives to be chosen, as long as they are similar enough to the status quo.[22]

The implication describes the fact that choice behaviour that may have been labelled as SQB as defined by Masatlioglu and Ok (2005), may be observationally equivalent to JND-type behaviour. However, the converse is not necessarily true; there may be many instances where JND representations are suitable and in

[22]Note that, if an alternative that is status quo is not chosen, this violates all the axioms of SQB i.e. an individual that simply chooses irrespective of status quo would be naturally unaffected by SQB.

which a status quo essentially plays no role in the choice procedure. In this sense, JNDs are able to explain different types of behaviour, one of which is SQB. The simplest example comes directly from Luce (1956) where intransitive indifference is explained, which is exactly what the JND representation is able to manage, however, SQB as in Masatlioglu and Ok (2005), in general, would not be able to explain the phenomenon of imperceptibility.[23]

3.2.4 Interval Order Representation

The previous theorem shows that JND utility is sufficient in explaining SQB behaviour. This relies upon a semiorder binary relation, which is what is required for a JND utility. However, by construction, there is another interesting feature of the previous theorem in that it is also capable of dealing with an interval order binary relation; this gives rise to its own representation, but is similar to that of Theorem 2.

As a reminder, a semiorder is a binary relation that obeys irreflexitivity, semitransitivity[24], and the interval order condition[25]. An interval order is a binary relation that is irreflexive and satisfies the interval order condition i.e. a semiorder is simply a semitransitive interval order. Given the structure of the axioms, it is also possible to achieve an interval order representation as given by the corollary below. In essence, as can be seen from Section C.1, Axiom ASQB is what yields semitransitivity, and Axiom DSQB is what yields the interval order condition. The interpretation is similar to that of JND utility insofar as the representation looks near identical except now that the JND is also a function of the alternatives. For some generic nonempty finite set, W, and status quo, w, define the following:

$\Gamma_{v,\epsilon(\cdot)}(W, w) \equiv \{p \in W : v(p) > v(w) + \epsilon(w)\}$

$\Pi_{v,\epsilon(\cdot)}(W, w) \equiv \{p \in W : \left|v(p) - v(w)\right| \leq \epsilon(w)\}$

$\Omega_{v,\epsilon(\cdot)}(W) \equiv \{q \in W : v(p) > v(q) + \epsilon(q) \text{ for no } p \in W\}$

[23] Naturally, there would be no expectation for their representation to explain phenomena beyond SQB.

[24] $\succ$ is semitransitive if $\forall a, b, c, d \in X$, $a \succ b \wedge b \succ c \implies a \succ d \vee d \succ c$.

[25] $\succ$ satisfies the IOC if $\forall a, b, c, d \in X$, $a \succ b \wedge c \succ d \implies a \succ d \vee c \succ b$.

Corollary 2.1. *Let S be a nonempty finite set. A choice correspondence $c(\cdot,\ \cdot)$ satisfies Axiom α, Axiom δ, Axiom D, Axiom SQI, and Axiom DSQB, if, and only if, there exists a threshold function $\epsilon : S \to \mathbb{R}_+$, a function $v : S \to \mathbb{R}$, such that*

$$c(S, \diamond) = \Omega_I(S)$$

and

$$c(S,x) = \begin{cases} \Pi_{v,\epsilon(\cdot)}(S,x) & \Gamma_{v,\epsilon(\cdot)}(S,x) = \emptyset \\ \Omega_{v,\epsilon(\cdot)}(\Gamma_{v,\epsilon(\cdot)}(S,x)) & otherwise \end{cases}$$

As can be seen by the definitions of $\Gamma_{v,\epsilon(\cdot)}(\cdot,\cdot)$, $\Pi_{v,\epsilon(\cdot)}(\cdot,\cdot)$, and $\Omega_{v,\epsilon(\cdot)}(\cdot,\cdot)$, the interpretation is as in Theorem 2 except that the comparison between alternatives depends not just on the utility but on a JND function of the alternatives. In terms of explaining SQB, the interpretation has remained essentially unchanged. However, given some specification for the JND function, it might be possible to observe different behaviour that is consistent with Corollary 2.1 but not necessarily with Theorem 2. This is simply due to the fact that the semiorder required Axiom ASQB and Axiom DSQB, whereas Corollary 2.1 needed only Axiom DSQB to show that the binary relation is an interval order.

In actuality, this particular representation may have a somewhat different explanation for SQB as alternatives that are being compared to the status quo are specifically given a higher regard. However, this is simply because the JND is a function of the alternative itself. In pure comparison terms, it is almost as if those alternatives that are not the status quo receive some utility boost, and yet, this representation can still explain SQB given that there is still this utility threshold.

One interpretation of a JND that is a function of the alternatives could be that of a switching cost. The idea being that the agent must incur some cost of switching from their status quo (or any other alternative), or perhaps that it is not costless for the agent to compare alternatives. In essence, if there is some constant switching cost across alternatives, then this is the interpretation of given by Theorem 2, as the JND is a constant. However, if the switching cost depends on the comparison being made, then this is the interpretation of Corollary 2.1.

Another interesting interpretation is that of the endowment effect, a common cognitive bias often studied in behavioural economics. It is mostly attributed to the fact that agents are status quo biased. This is seen as a discrepancy in the willingness-to-accept (WTA) and willingness-to-pay (WTP). In standard models, the WTA should always be exactly equal to the WTP, as the WTA/WTP is simply a minimum/maximum threshold value at which an agent is willing to sell/buy.[26] This difference is seen as a result of an *"endowment effect"* given the observation that people tend to require a higher selling value when in possession of the item they are going to sell. The representation of Corollary 2.1 is able to provide an explanation of this.

Consider Axiom DSQB which says that an SQB agent would decide against choosing an alternative that was chosen previously from a different choice problem (and with different status quo). Recall from Axiom ASQB that an SQB agent would decide against anything chosen from the same choice problem before the choice set was augmented. This means that with Axiom DSQB, there is a higher regard for a particular status quo as that status quo is specifically chosen over an alternative from another choice problem.

For example, suppose one choice set involved choosing between three large mugs coloured red, blue, and green with a separate choice problem involving three small mugs of the same colours. If an agent chooses a large blue mug when the large green mug was the status quo, and similarly for the small mugs, when the large blue mug becomes the status quo, the agent never chooses the small blue mug. In some sense, the agent has a strong affinity to the large blue mug as it is the status quo, especially as the previously chosen small blue mug is not chosen. The same would be true, vice versa, if the status quo was the small blue mug. So, there is something specific about the status quo mug that means that

[26]Typically, the endowment effect is explained in terms of prices with respect to WTA and WTP.

something that was chosen in a completely separate choice problem, is no longer chosen.[27] As such, the interval order representation is able to offer an explanation for endowment-type effects.

3.3 Related Literature

As mentioned previously, Samuelson and Zeckhauser (1988) introduced the idea and term of status quo bias into the economics literature. Although behavioural economics provides natural deviations from standard models *(for example, non-exhaustively, Tversky and Kahneman (1991, 1992), Rabin (1998), Kőszegi and Rabin (2006))* to explain such phenomena, there is a large growing literature in decision theory *(for example, Masatlioglu et al. (2012), Manzini and Mariotti (2012), Ok et al. (2015), Masatlioglu and Ok (2005, 2014), Frick (2016), Argenziano and Gilboa (2017), and many others)* with the aim of providing axiomatic foundations to their utility representations for (what is normally considered) non-standard decision making. I contribute to this literature by explaining SQB behaviour with an existing utility representation i.e. JND utility, with the familiar choice procedure of Simon (1955), Bewley (2002), Masatlioglu and Ok (2005).

Luce (1956), Scott and Suppes (1958), Jamison and Lau (1973), Fishburn (1975) are key papers in the study of semiorders. Luce (1956) introduced the semiorder binary relation as the primitive for an axiomatisation of JND utility. Both Fishburn (1975), Jamison and Lau (1973) derive an equivalent set of axioms with a choice function/correspondence primitive. These papers do not make specific mention of any specific behavioural phenomenon beyond imperceptibility as a way of rationalising intransitive indifference. However, understanding the interplay of axioms was essential in devising axioms that allowed incorporation of a status quo with a JND interpretation.

[27]It is in combining these axioms is what allows the JND to be constant over all comparisons, as Axiom ASQB does not preclude any new alternative from being chosen.

3.4 Concluding remarks

The choice theory presented in this chapter is able to incorporate a status quo effect into a standard decision-making procedure with a classic just noticeable difference representation. The channel for status quo bias is such that, conditional on some current choice, the agent does not move away from this unless there is an alternative that is necessarily and noticeably better. This representation also gives rise to several plausible explanations for other behaviours that are relatable to status quo bias.

An interesting parallel can be drawn with the literature on consideration sets. Given the observability of status quo bias behaviour, and the fact that this can be represented by JND utility, this gives further rise to the notion that consumers choose within a consideration set, and that this set need not necessarily coincide with what economists believe to be their choice set.[28] Although JND utility is derived from axioms of standard rationality combined with status quo, it is not necessarily obvious why agents are not able/willing to notice some subset of alternatives even when there can exist better choices.[29] Is there anything fundamental or intrinsic about individuals that results in this behaviour? Is there a cost of search/consideration? Do agents suffer from choice overload?[30] Is there an optimal level of attention? These are questions have been tackled successfully in the growing literature on limited consideration.[31] These papers may provide more fundamental answers as to what are the *observable* qualities of agents that give rise to JND utility and, by extension, status quo bias.

[28]Horowitz and Louviere (1995), Spiegler and Eliaz (2011), Manzini and Mariotti (2014) etc...

[29]In a status quo bias framework, if an alternative is not noticeably better/worse, is it still in the consideration set?

[30]The increased difficulty of choosing optimally for larger choice sets.

[31]For example, Frick (2016)

Conclusion

This thesis comprises three chapters that use revealed preference theory and decision theory to provide behavioural insights into consumer rationality, consideration sets, and status quo bias.

The objective of Chapter 1 was to present analyses that combines the role of consideration sets with economic rationality in the decision-making process using revealed preference theory. Using a scanner panel dataset and a simulated dataset, and applying two well-known measures of rationality and introducing a new index of rationality, there is strong evidence that the consumer decision-making process is more rational in the presence of consideration sets, defined exogenously as goods with positive consumption. The role of consideration sets is reflected through increasingly irrational consumption choices that can exacerbated by the 'size' of the consideration set, as well as longer sequence lengths. From the regression analysis, these measures of consideration set complexity, as well as certain demographic factors, are correlated with rationality; specifically, when the consideration set is larger and/or the sequence length is longer, there appears to be a negative correlation with rationality. Overall, there is a clear relationship between the role of consideration sets and revealed preference theory in the decision-making process of the household. In this chapter, I have provided an initial insight into how these concepts can be combined and how that this can improve our ability to study economic behaviour, leading naturally on to Chapter 2.

Chapter 2 incorporates a natural extension to the canonical model of utility which requires a less strict form of rationality than otherwise prescribed owing to the additional cost of consideration. As a unique contribution to the literature, consumers may only consider a subset of the goods available to them if the cost of consideration is too high. To model this, a modified budget constraint is used that incorporates, what is essentially, a fixed cost of consideration. Through modification of the existing standard budget constraint, the cost of consideration can be interpreted in terms of expenditure as a means of a price distortion, rather than any arbitrary measure of cognitive ability. This chapter attempts to bring forward the case of a broader and deeper theory of consideration sets.

A natural extension to the work of Chapter 2 is applications to other price-quantity datasets that have seemingly less rational individuals than the Stanford Basket Dataset. This could allow for larger variation in consideration costs which would be ideal for the reduced-form analysis in terms of the demographic factors. Another interesting avenue that revealed preference frameworks can tend to avoid is related to the analysis when decisions are made on a discrete basis rather than a continuous one, as in the style of Polisson and Quah (2013). As the original formation of the consideration set cost is that of a fixed cost, incorporating the use of discrete choices, as opposed to choices from a continuum, could lead to further interesting results and insights.

Chapter 3 presents a choice theory that incorporates a status quo effect into a standard decision-making procedure with a just noticeable difference utility representation. The contribution towards the understanding of status quo is such that, conditional on some current choice, the agent does not move away from this unless there is an alternative that is necessarily and noticeably better. This chapter also shows there are consistent ways of aggregating finite preferences that are represented by JND utility, with aggregators that can also be a JND utility function or a standard utility function.

One natural extension to this chapter would be related to dynamic choice. An interesting idea that arises is whether time inconsistent decisions can be explained from having some form dynamic JND utility i.e. the current period is considered the 'status-quo' period, and hence there is an inclination toward it. A further extension to this chapter could include incorporation of uncertainty, which would require some form of independence axiom with respect to quality of alternatives. In terms of the empirical side, a future project would involve relating the JND utility representation to that of a class of random utility models, as is seen in applied econometrics and industrial organisation. The basic premise of random utility models is that utility can suffer from random shocks and with that there are probabilities associated to making choices. If there does exist some relationship between these random shocks in random utility models and the JND parameter, this may shed new light on how economists think of JND utility, both econometrically and theoretically speaking.

Bibliography

Afriat, S. (1967). The Construction of a Utility Function From Expenditure Data. *International Economic Review*, 8:67–77.

Afriat, S. (1972). Efficiency Estimates of Production Functions. *International Economic Review*, 13:568–598.

Andreoni, J., Gillen, B., and Harbaugh, W. T. (2011). The Power of Revealed Preference Tests: Ex-Post Evaluation of Experimental Design. *Tech. rep.*

Andreoni, J. and Harbaugh, W. (2002). Power Indices for Revealed Preference Tests. *Mimeo.*

Andreoni, J. and Harbaugh, W. (2006). Power indices for revealed preference tests. Levine's bibliography, UCLA Department of Economics.

Andreoni, J. and Miller, J. H. (2002). Giving According to GARP: An Experimetal Test of the Consistency of Preference for Altruism. *Econometrica*, 70(2):737–753.

Andrews, R. and Srinivasa, T. (1995). Studying Consideration Effects in Empirical Choice Models using Scanner Panel Data. *American Marketing Association*, 32:30–41.

Apesteguia, J. and Ballester, M. A. (2010). A measure of rationality and welfare. Economics Working Papers 1220, Department of Economics and Business, Universitat Pompeu Fabra.

Argenziano, R. and Gilboa, I. (2017). Psychophysical foundations of the cobb–douglas utility function. *Economics Letters*, 157:21 – 23.

Argenziano, R. and Gilboa, I. (2018). Perception-theoretic foundations of weighted utilitarianism. *The Economic Journal*, 0(0).

Beatty, T. K. and Crawford, I. A. (2011). How Demanding is the Revealed Preference Approach to Demand. *American Economic Review*, 101(6):2782–2795.

Becker, G. S. (1962). Irrational behavior and economic theory. *Journal of Political Economy*, 70(1):1–13.

Bewley, T. F. (2002). Knightian decision theory. part i. *Decisions in Economics and Finance*, 25(2):79–110.

Blundell, R. (1988). Consumer Behaviour: Theory and Empirical Evidence - a Survey. *Economic Journal*, 98(389):16–65.

Blundell, R. (2005). How revealing is revealed preference? *Journal of the European Economic Association*, 3(23):211–235.

Blundell, R., Browning, M., and Crawford, I. (2007). Improving Revealed Preference Bounds on Demand Responses. *International Economic Review*, 48:1227–1262.

Blundell, R. and Crawford, I. A. (2003). Nonparametric Engel Curves and Revealed Preference. *Econometrica*, 71:208–240.

Bronars, S. G. (1987). The Power of Nonparametric Tests of Preference Maximisation. *Econometrica*, 55(3):693–698.

Chevalier, J. A., Kashyap, A. K., and Rossi, P. E. (2003). Why don't prices rise during periods of peak demand? evidence from scanner data. *The American Economic Review*, 93(1):15–37.

Chiang, J. S., Chib, S., and Narasimhan, C. (1999). Markov Chain Monte Carlo and Models of Consideration Set and Parameter Heterogeneity. *Journal of Econometrics*, 89:223–248.

Choi, S., Fisman, R., Gale, D., and Kariv, S. (2003). Revealing Preferences Graphically: An Old Methods Gets a New Tool Kit. *American Economic Review*, 93(1):153–158.

Choi, S., Fisman, R., Gale, D., and Kariv, S. (2007). Consistency and Heterogeneity of Individual Behaviour under Uncertainty. *American Economic Review*, 97:1921–1938.

Choi, S., Kariv, S., Muller, W., and Silverman, D. (2011). Who is (more) Rational? *American Economic Review*.

Deaton, A. S. and Muellbauer, J. (1980). Economics and Consumer Behavior. *Cambridge University Press*.

Demuynck, T. and Seel, C. (2018). Revealed preference with limited consideration. *American Economic Journal: Microeconomics*, 10(1):102–31.

Diewert, W. E. (1973). Afriat and Revealed Preference Theory. *Review of Economic Studies*, 40(3):419–25.

Dziewulski, P. (2017). Eliciting the just-noticeable difference. *Economics Series Working Papers 798, University of Oxford, Department of Economics*.

Dziewulski, P. (2018). Just-noticeable difference as a behavioural foundation of the critical cost-efficiency. *Economics Series Working Papers 848, University of Oxford, Department of Economics*.

Echenique, F., Lee, S., and Shum, M. (2011). The Money Pump as a Measure of Revealed Preference Violations. *Journal of Political Economy*, 119:1201–1223.

Featherstone, A. M. and Rahman, M. H. (1996). Nonparametric analysis of the optimizing behavior of midwestern cooperatives. *Review of Agricultural Economics*, 18(2):265–273.

Fechner, G. (1860). Elemente der psychophysik. *British Journal of Statistical Psychology*, 13(1):1–10.

Fishburn, P. C. (1975). Semiorders and choice functions. *Econometrica*, 43(5/6):975–977.

Fishburn, P. C. (1985). Interval graphs and interval orders. *Discrete Mathematics*, 55(2):135 – 149.

Fleissig, A. R. and Whitney, G. A. (2005). Testing for the Significance of Vioations of Afriat's Inequalities. *Journal of Business and Economic Statistics*, 23:355–62.

Fostel, A., Scarf, H. E., and Todd, M. J. (2003). Two New Proofs of Afriat's Theorem. Cowles Foundation Discussion Papers 1415, Cowles Foundation for Research in Economics, Yale University.

Frick, M. (2016). Monotone threshold representations. *Theoretical Economics*, 11(3):757–772.

Gabaix, X. (2014). A Sparsity-Based Model of Bounded Rationality. *NBER Working Paper*.

Goeree, M. S. (2008). Limited Information and Advertising in the U.S. Personal Computer Industry. *Econometrica*, 76:1017–1074.

Gross, J. (1995). Testing data for consistency with revealed preference. *The Review of Economics and Statistics*, 77(4):701–710.

Harbaugh, W. T., Krause, K., and Berry, T. R. (2001). Garp for kids: On the development of rational choice behavior. *American Economic Review*, 91(5):1539–1545.

Horowitz, J. and Louviere, J. (1995). What is the role of consideration sets in choice modeling? *International Journal of Research in Marketing*, 12:39–54.

Jamison, D. T. and Lau, L. J. (1973). Semiorders and the theory of choice. *Econometrica*, 41(5):901–912.

Kahneman, D., Knetsch, J. L., and Thaler, R. H. (1990). Experimental tests of the endowment effect and the coase theorem. *Journal of Political Economy*, 98(6):1325–1348.

Knetsch, J. L. (1989). The endowment effect and evidence of nonreversible indifference curves. *The American Economic Review*, 79(5):1277–1284.

Kőszegi, B. and Rabin, M. (2006). A model of reference-dependent preferences*. *The Quarterly Journal of Economics*, 121(4):1133.

Kreps, D. (1988). *Notes on the Theory of Choice*. Westview Press.

Luce, R. D. (1956). Semiorders and a theory of utility discrimination. *Econometrica*, 24(2):178–191.

Manzini, P. and Mariotti, M. (2007). Sequentially rationalizable choice. *American Economic Review*, 97(5):1824–1839.

Manzini, P. and Mariotti, M. (2012). Choice by lexicographic semiorders. *Theoretical Economics*, 7(1):1–23.

Manzini, P. and Mariotti, M. (2014). Stochastic Choice and Consideration Sets. *Econometrica*, 82:1153–1176.

Masatlioglu, Y. and Nakajima, D. (2013). Choice by Iterative Search. *Theoretical Economics*, 8(3):701–728.

Masatlioglu, Y., Nakajima, D., and Ozbay, E. Y. (2012). Revealed Attention. *American Economic Review*, 102(5):2183–2205.

Masatlioglu, Y. and Ok, E. A. (2005). Rational choice with status quo bias. *Journal of Economic Theory*, 121(1):1 – 29.

Masatlioglu, Y. and Ok, E. A. (2014). A canonical model of choice with initial endowments. *The Review of Economic Studies*, 81(2):851–883.

Mehta, N., Rajiv, S., and Srinivasan, K. (2003). Price Uncertainty and Consumer Search: A Structural Model of Consideration Set Formation. *Marketing Science*, 22:58–84.

Ok, E. A. (2002). Utility representation of an incomplete preference relation. *Journal of Economic Theory*, 104(2):429 – 449.

Ok, E. A., Ortoleva, P., and Riella, G. (2015). Revealed (p)reference theory. *American Economic Review*, 105(1):299–321.

Polisson, M. and Quah, J. K.-H. (2013). Revealed Preference in a Discrete Consumption Space. *American Economic Journal: Microeconomics*, 5(1):28–34.

Rabin, M. (1998). Psychology and economics. *Journal of Economic Literature*, 36(1):11–46.

Richter, M. (1966). Revealed Preference Theory. *Econometrica*, 34:635–645.

Rubinstein, A. (1988). Similarity and decision-making under risk (is there a utility theory resolution to the allais paradox?). *Journal of Economic Theory*, 46(1):145 – 153.

Samuelson, P. (1948). Consumption Theory in Terms of Revealed Preference. *Economica*, 15:243–253.

Samuelson, W. and Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty*, 1(1):7–59.

Scott, D. and Suppes, P. (1958). Foundational aspects of theories of measurement. *Journal of Symbolic Logic*, 23(2):113–128.

Sen, A. K. (1971). Choice functions and revealed preference. *The Review of Economic Studies*, 38(3):307–317.

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, 69(1):99–118.

Sippel, R. (1997). An experiment on the pure theory of consumer's behaviour. *The Economic Journal*, 107(444):1431–1444.

Spiegler, R. (2014). Competitive framing. *American Economic Journal: Microeconomics*, 6(3):35–58.

Spiegler, R. and Eliaz, K. (2011). Consideration Sets and Competitive Marketing. *Review of Economic Studies*, 78:235–262.

Szpilrajn, E. (1930). Sur l'extension de l'ordre partiel. *Fundamenta Mathematicae*, 16(1):386–389.

Tversky, A. and Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics*, 106(4):1039–1061.

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323.

Varian, H. (1982a). Goodness-of-fit in Optimising Models. *Journal of Econometrics*, 46:125–140.

Varian, H. (1985). Non-Parametric Analysis of Optimizing Behaviour with Measurement Error. *Journal of Econometrics*, 30:445–458.

Varian, H. (2006). Revealed Preference. *Samuelsonian Economics and the 21st Century*, pages 99–115.

Varian, H. R. (1982b). The nonparametric approach to demand analysis. *Econometrica*, 50(4):945–973.

Varian, H. R. (1994). Goodness-of-Fit for Revealed Preference Tests. Econometrics 9401001, University Library of Munich, Germany.

Weber (1834). *De Tactu.* ("Concerning Touch").

Wright, P. and Barbour, F. (1977). Phased decision strategies: Sequels to an initial screening. *Studies in Management Sciences, Multiple Criteria Decision Making*, pages 91–109.

Appendix A

Appendix for Chapter 1

A.1 Tables

Table 1.1 - Summary Statistics

Number of equivalence classes	24.47
	(2.04)
Maximum	26
Minimum	14
Number of distinct goods per households	56
	(21.65)

Sample standard deviation in parentheses.

Table 1.2 - Frequency of Equivalence Classes

Number of equivalence classes	Longest possible partition sequence length	Frequency
14	13	1
16	11	2
17	10	4
18	9	4
19	8	7
20	7	12
21	6	14
22	5	24
23	4	46
24	3	70
25	2	93
26	1	217

Table 1.3 - Rationality indices summary statistics without consideration sets

Cycle Lengths	2	2,3	2,3,4	2,3,4,5
Total number of households	494	494	494	494
Households violating GARP	395	396	396	396
Average MPI	0.0622	0.0612	0.0609	0.0604
Average GAV Index	0.0651	0.0649	0.0642	0.0642
Average (1-CCEI)	0.0263	0.0238	0.0231	0.0230
Median MPI	0.0597	0.0595	0.0591	0.0590
Median GAV Index	0.0646	0.0640	0.0635	0.0635
Median (1-CCEI)	0.0240	0.0209	0.0203	0.0202

Table 1.4 - Rationality indices summary statistics with consideration sets

Cycle Lengths	2	2,3	2,3,4	2,3,4,5
Number of households	184	184	114	68
Households violating GARP	83	50	28	12
Average MPI	0.0443	0.0442	0.0412	0.0358
Average GAV Index	0.0464	0.0464	0.0433	0.0382
Average (1-CCEI)	0.0187	0.0175	0.0137	0.0105
Median MPI	0.0419	0.0417	0.0385	0.0334
Median GAV Index	0.0428	0.0425	0.0395	0.0364
Median (1-CCEI)	0.0146	0.0138	0.0106	0.0095

Table 1.5 - Rationality indices summary statistics with consideration sets with 2 to 7 goods

Consideration set size	2	3	4	5	6	7
Number of households	118	178	184	184	184	184
Households violating GARP	80	82	83	83	83	83
Average sequence length	1.67	1.67	1.60	1.60	1.60	1.60
Average MPI	0.0423	0.0443	0.0443	0.0443	0.0443	0.0443
Average GAV Index	0.0459	0.0464	0.0464	0.0464	0.0464	0.0464
Average (1-CCEI)	0.0184	0.0187	0.0187	0.0187	0.0187	0.0187
Median MPI	0.0421	0.0418	0.0419	0.0419	0.0419	0.0419
Median GAV Index	0.0431	0.0428	0.0428	0.0428	0.0428	0.0428
Median (1-CCEI)	0.0150	0.0146	0.0146	0.0146	0.0146	0.0146

Table 1.6 - Rationality indices summary statistics with consideration sets with 8 to 13 goods

Consideration set size	8	9	10	11	12	13
Number of households	157	130	96	63	18	10
Households violating GARP	78	67	45	28	6	6
Average sequence length	1.58	1.52	1.43	1.44	1.42	1.42
Average MPI	0.0444	0.0444	0.0511	0.0531	0.0513	0.0513
Average GAV Index	0.0464	0.0465	0.0541	0.0552	0.0537	0.0537
Average (1-CCEI)	0.0187	0.0188	0.0222	0.0232	0.0216	0.0216
Median MPI	0.0418	0.0419	0.0422	0.0424	0.0423	0.0423
Median GAV Index	0.0431	0.0430	0.0430	0.0455	0.0441	0.0441
Median (1-CCEI)	0.0146	0.0150	0.0151	0.0154	0.0146	0.0146

Table 1.7 - Number of households with sequence length 2-6 and consideration set size 3-10

Consideration set size

Sequence length	3	4	5	6	7	8	9	10
1	65	101	106	106	106	116	90	67
2	35	48	54	54	54	20	23	17
3	12	20	19	19	19	16	10	12
4	5	4	3	3	3	9	4	0
5	1	5	1	1	1	5	2	0
6	0	0	1	1	1	1	1	0

Table 1.8 - Average MPI with sequence length 2-6 and consideration set size 3-10

Consideration set size

Sequence length	3	4	5	6	7	8	9	10
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.0444	0.0446	0.0446	0.0446	0.0446	0.0446	0.0446	0.0495
3	0.0455	0.0506	0.0506	0.0506	0.0506	0.0506	0.0535	0.0634
4	0.0368	0.0581	0.0581	0.0581	0.0581	0.0581	0.0612	0.0671
5	0.0342	0.0586	0.0586	0.0586	0.0586	0.0586	0.0585	0.0729
6	0.0411	0.0598	0.0598	0.0598	0.0598	0.0598	0.0634	0.0838

Table 1.9 - Average (1-CCEI) with sequence length 2-6 and consideration set size 3-10

Consideration set size

Sequence length	3	4	5	6	7	8	9	10
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.0188	0.0189	0.0189	0.0189	0.0189	0.0188	0.0188	0.0224
3	0.0176	0.0267	0.0267	0.0267	0.0267	0.0265	0.0232	0.0280
4	0.2000	0.0318	0.0318	0.0318	0.0318	0.0282	0.0203	0.0356
5	0.0249	0.0419	0.0419	0.0419	0.0419	0.0385	0.0273	0.0406
6	0.0179	0.0368	0.0368	0.0368	0.0368	0.0502	0.0320	0.0474

A.2 Regressions

A.2.1 Table 7

	(1)	(2)	(3)	(4)	(5)
	1-CCEI	1-CCEI	1-CCEI	1-CCEI	1-CCEI
LarFMSize	-0.000151***	-0.000134***	-0.000108**	-0.000098*	-0.000097*
	(-4.32)	(-3.77)	(-2.79)	(-1.98)	(-1.99)
MidFMSize	-0.0000713**	-0.0000533*	-0.0000386	-0.0000385	-0.0000385
	(-2.91)	(-2.07)	(-1.39)	(-1.38)	(-1.37)
Highsch	-0.0000824	-0.0000490	-0.0000472	-0.0000512	-0.0000519
	(-1.65)	(-0.95)	(-0.91)	(-0.81)	(-0.83)
College	-0.0000917	-0.0000466	-0.0000407	-0.0000333	-0.0000310
	(-1.85)	(-0.89)	(-0.76)	(-0.66)	(-0.67)
HighIncome		-0.0000660*	-0.0000697	-0.0000678*	-0.0000677*
		(-1.99)	(-1.92)	(-2.04)	(-2.10)
MidIncome		-0.0000871**	-0.0000879**	-0.0000878**	-0.0000849**
		(-2.96)	(-2.81)	(-2.79)	(-2.78)
OldAge			0.0000518	0.0000314	0.0000412
			(1.34)	(1.22)	(1.30)
MidAge			0.0000670*	0.0000671*	0.0000671*
			(2.25)	(2.25)	(2.25)
ASL				0.0000655**	0.0000095
				(2.79)	(1.10)
ACSS					0.0000712*
					(2.05)
_cons	0.000384***	0.000393***	0.000336***	0.000316***	0.000310***
	(8.23)	(8.40)	(5.41)	(4.44)	(4.41)
N	156000	156000	156000	156000	156000

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

A.3 Simulations

A.3.1 Simulated Dataset

The reason for simulating households[1] is so that I can have separate datasets which represent set levels of rationality for different consumers. This is so that I am able to do some sensitivity analysis across CCEI for more irrational decisions and see if any of the patterns still exist/ are exacerbated by individuals with lower CCEIs. One of the issues with the empirical data is that individuals tend to be fairly rational (score CCEIs greater than 0.95), and as such, it may be tricky to discern increases in rationality. By constructing datasets with individuals that start with lower CCEIs, it may be easier to detect potential patterns.

For example, suppose there is an individual who has an average CCEI level of 0.5. So, in order to rationalise this individual's data, their income needs to be relaxed by 50%. However, suppose this was calculated in the standard way, without taking into account any limited consideration. What we want to see is how rationality changes if we partition the dataset properly, according to their consideration sets.

Using the same structure as the Stanford Basket Dataset, I reverse engineer random consumption choices (given random prices) such that the data yield a certain level of CCEI. Using the definition of the CCEI, I solve a large linear system of equations in order to choose consumption levels of a certain number of goods for 2 bundles/weeks[2] such that for each household, on average, the consumption and price levels simulated provide a specifically chosen CCEI level. It is important to note that the simulations for a particular CCEI are themselves subject to Monte-Carlo simulations. This is to ensure that the simulated datasets for a particular CCEI were not simply yielding special-case results. In essence, I simulate several datasets with a fixed CCEI, collect the relevant numbers from each dataset and take an average of these; any results explained using simulated data are based on the initial CCEI simulation and the subsequent Monte-Carlo simulation. For the

[1] To be as consistent as possible with the Stanford Basket Dataset, I simulated 494 households.
[2] I solve for a sequence length of 2 for computational ease.

sake of the simulations, I chose an average CCEI of 0.85.[3]

A.3.2 Algorithm

A. Choose number of goods (G) and number of bundles (B) based on identifiability

i.e. the number of equations has to be at least as large as number of variables. In this instance, the number of variables is the number of actual goods multiplied by the number of weeks(bundles) and the number of equations is based on the CCEI definition for a sequence length of 2 and income feasibility.

$$\underbrace{G * B}_{\text{No. of Variables}} \leq \overbrace{2 * \binom{B}{2}}^{\text{No. of CCEI equations}} + \underbrace{B}_{\text{No. of Feasibility Constraints}} \tag{A.3.1}$$

$$G * B \leq 2 * \frac{B!}{(B-2)! * 2!} + B \tag{A.3.2}$$

$$G * B \leq 2 * \frac{B * (B-1)!}{(B-2)! * 2} + B \tag{A.3.3}$$

$$G \leq \frac{(B-1)!}{(B-2)!} + 1 \tag{A.3.4}$$

$$G \leq (B-1) + 1 \tag{A.3.5}$$

$$G \leq B \tag{A.3.6}$$

i.e. Number of goods has to be less than number of bundles.[4]

In the simulations completed for this chapter, I chose 4 to 12 consideration

[3]This is a fairly low score to get in applied revealed preference work. However, it is low enough so that any discernible patterns are much more noticeable, but is high enough as not to be unrealistic.

[4]If this were not the case, then it would be necessary to parameterise the system of equations, which would mean setting quantities to some random numbers, which would artificially worsen rationality, or setting them to zero as to not interfere with the measuring of rationality.

sets from 12 goods for a total of 12 bundles.[5]

B. Given a fixed CCEI for the whole dataset, randomly choose CCEIs on the household level. After this, randomly choose CCEIs for every sequence of bundles. As I require a distribution such that the random draws have a mean of the fixed CCEI with a low chance of outliers on the irrational side, I chose a (truncated) log-normal distribution for 1-CCEI[6] to exploit the desired skew of the distribution. This choice seemed to be without loss of generality as many distributions yielded the same result as long as I had the same moment restrictions.[7]

Let X denote 1-CCEI $\sim \ln \mathcal{N}(\mu, \sigma)$

The moment conditions are as follows:

$$E[X] = e^{\mu + \frac{\sigma^2}{2}} \tag{A.3.7}$$

$$V[X] = (e^{\sigma^2} - 1) \cdot e^{2\mu + \sigma^2} \tag{A.3.8}$$

Choose $E[X]$ equal to the fixed 1-CCEI and set $V[X]$ equal to 0.0015[8] and solve simultaneously for μ and σ. Once, these parameters of the log-normal distribution have been calculated, create a truncated log-normal distribution[9] based on the derived moment conditions of the log-normal distribution above. Once this is done, take random draws from the truncated distribution. This is then repeated for each household given their specific CCEI for each GARP sequence.

Prices were drawn randomly from a uniform distribution[10]

[5] The number of consideration sets is chosen randomly and roughly follows the Stanford Basket Dataset in terms of frequency.

[6] Recall that CCEI=0/1 signifies perfection irrationality/rationality.

[7] 1st moment being the fixed CCEI and the 2nd moment being based on the 2nd moment of the original sequence length 2 CCEIs from the Stanford Basket Dataset.

[8] Variance of CCEIs for the Stanford Basket Dataset (sequence length = 2).

[9] Truncated between 0 and 1 inclusive.

[10] Using same range as the prices in the Stanford Basket Dataset Random prices. Drawing from any of the common distributions and any positive range did not affect any of the simulations.

C. Construct feasibility constraints for each bundle.[11]

Given the randomly drawn CCEIs, construct CCEI equations based on the definition of the CCEI. Recall that the CCEI for sequence length 2 is the maximum between 2 numbers. To ensure that the highest CCEIs are chosen, another set of equations with lower CCEIs must be constructed.

Solve the above system of equations (feasibility constraints, both sets of CCEI equations and non-negativity) in order to get a dataset of goods purchased per week.[12]

D. Run a Monte-Carlo Simulation.

i.e. run the above steps 250 times whilst collecting all the data needed for constructing the tables in section A.2, then take averages over the collected data to get the required numbers.

e.g. calculate the MPI[13] (sequence length 3) 250 times and take an average.

Two inconveniences arise when doing these data simulations. As mentioned previously, the number of GARP cycles that need to be checked can be extremely large. For a sequence length of L, the number of calculations is equal to

$$\begin{pmatrix} 26 \\ L \end{pmatrix} * (L-1)! \tag{A.3.9}$$

As can be imagined, this is computationally heavy, as the calculations are also done per household (494). For example, sequence lengths 5, 6 and, 7 involve 502,822,320 calculations per household(!).

Secondly, there is an associated 'curse of dimensionality' when simulating consumption choices for specific CCEIs. For example, for the 20 goods (G)and 20 bundles (B) case, this implies 400 variables, and at least 400 equations. Solving such a large system is not a trivial computational task as, again, this must be done

[11]Based on a uniform random draw from 1 to 100, again, simulations were not affected by distribution or range.

[12]As a check, I calculated the CCEIs for the dataset and compare to the randomly drawn CCEIs.

[13]Average MPI across all households.

for each household. Also, as to improve the robustness of the results, this was done 250 times for the Monte Carlo simulations. This implies 123,500 systems of equations involving 49,400,000 equations and variables. The issue of dimensionality arises mostly from the number of households and the Monte Carlo simulations. In fact, reducing the number of households and Monte Carlo simulations did not significantly alter the results.

A.3.3 Results

Table A - Frequency of Equivalence Classes

Number of equivalence classes	Longest possible partition sequence length	Frequency
4	9	1
5	8	2
6	7	3
7	6	7
8	5	15
9	4	34
10	3	80
11	2	100
12	1	252

Table B - Rationality indices summary statistics without consideration sets

Cycle Lengths	2	2,3	2,3,4	2,3,4,5
Total number of households	494	494	494	494
Households violating GARP	405	405	405	405
Average MPI	0.1234	0.1200	0.1197	0.1166
Average GAV Index	0.1540	0.1541	0.1433	0.1422
Average (1-CCEI)	0.1543	0.1321	0.1306	0.1265
Median MPI	0.1187	0.1045	0.1044	0.1044
Median GAV Index	0.1345	0.1348	0.1301	0.1388
Median (1-CCEI)	0.1078	0.1084	0.1023	0.1020

Table C - Rationality indices summary statistics with consideration sets

Cycle Lengths	2	2,3	2,3,4	2,3,4,5
Number of households	142	142	62	28
Households violating GARP	130	130	50	25
Average MPI	0.1321	0.1245	0.1144	0.1133
Average GAV Index	0.1299	0.1233	0.1111	0.1096
Average (1-CCEI)	0.1081	0.1053	0.1051	0.1040
Median MPI	0.1354	0.1377	0.1387	0.1354
Median GAV Index	0.1367	0.1300	0.1255	0.1199
Median (1-CCEI)	0.1054	0.1044	0.1047	0.1089

Table D - Rationality indices summary statistics with consideration sets with 4 to 12 goods

Consideration set size	4-7	8	9	10	11	12
Number of households	142	100	99	40	41	22
Households violating GARP	130	90	90	29	24	14
Average sequence length	1.88	1.78	1.77	1.78	1.64	1.66
Average MPI	0.1321	0.1333	0.1376	0.1578	0.1823	0.1938
Average GAV Index	0.1299	0.1426	0.1789	0.1866	0.2098	0.2107
Average (1-CCEI)	0.1081	0.1254	0.1289	0.1356	0.1401	0.1444
Median MPI	0.1354	0.1061	0.1062	0.1211	0.1465	0.1499
Median GAV Index	0.1367	0.1421	0.1533	0.1699	0.1976	0.2033
Median (1-CCEI)	0.1054	0.1176	0.1211	0.1289	0.1343	0.1349

The results for the simulated data follow the patterns found in the empirical dataset, hence I refer the reader to the main text. However, the only additional observation is that the results seem to be exacerbated for agents that start off as being more irrational for smaller consideration sets. This again corroborates the idea that larger consideration sets are not necessarily facilitating the making of more rational choices.

A.4 Statistical Test for GAV Index

Let $\mathbf{p}_i = \mathbf{K}_i + \mu_i$ denote mis-measured prices.

Let θ_i denote income shares.

Assume that is $\frac{\mu_i}{\theta_i}$ i.i.d and normally distributed with mean 0, and variance $\frac{\sigma^2}{\theta_i^2}$.

Take $\mathbf{q}_{N+1} = \mathbf{q}_1$, $\mathbf{P}_{N+1} = \mathbf{p}_1$, $\theta_{N+1} = \theta_1$ as given.

Let $H(N)$ denote the statistical distribution for the GAV index of length N.

$$H(N) = \max\left[\frac{1}{N}\sum_{i=1}^{N}\frac{\mathbf{p}_i * (\mathbf{q}_i - \mathbf{q}_{i+1})}{\theta_i}, 0\right] \tag{A.4.1}$$

Note the max operator is required as GARP is violated when the GAV index is negative i.e. the law of demand is violated. Let $\frac{\mathbf{p}_i}{\theta_i} = \frac{\mathbf{K}_i + \mu_i}{\theta_i}$ be denoted as $X_i = Y_i + \omega_i$

$$H(N) = \max\left[\frac{1}{N}\sum_{i=1}^{N} X_i * (\mathbf{q}_i - \mathbf{q}_{i+1}), 0\right] \tag{A.4.2}$$

Let G(N) denote the statistical distribution of the GAV Index, under the null hypothesis of rationality.

$$\begin{aligned} G(N) &= \frac{1}{N}\sum_{i=1}^{N} X_i * (\mathbf{q}_i - \mathbf{q}_{i+1}) && \text{(A.4.3)} \\ &= \frac{1}{N}\sum_{i=1}^{N}(Y_i + \omega_i) * (\mathbf{q}_i - \mathbf{q}_{i+1}) && \text{(A.4.4)} \\ &= \frac{1}{N}\sum_{i=1}^{N} Y_i * (\mathbf{q}_i - \mathbf{q}_{i+1}) + \sum_{i=1}^{N}\omega_i * (\mathbf{q}_i - \mathbf{q}_{i+1}) && \text{(A.4.5)} \end{aligned}$$

By Afriat's Theorem, the law of negativity must hold which implies that for every bundle, $(Y_i - Y_{i+1})(\mathbf{q}_i - \mathbf{q}_{i+1}) \leq 0$. Combining this with the fact that ω_i is assumed to be mean 0 for all time periods, then all that is left to calculate is the variance of $G(N)$, conditional on the observed bundles. Under the i.i.d assumption, it is clear that

$$Var(G(N)) = Var\left[\frac{1}{N}\sum_{i=1}^{N}\omega_i * (\mathbf{q}_i - \mathbf{q}_{i+1})\right] \tag{A.4.6}$$

$$= \frac{1}{N^2}\sum_{i=1}^{N} Var[\omega_i] * (\mathbf{q}_i - \mathbf{q}_{i+1}) \tag{A.4.7}$$

$$= \frac{1}{N^2}\sum_{i=1}^{N} \frac{\sigma^2}{\theta_i^2} * (\mathbf{q}_i - \mathbf{q}_{i+1}) \tag{A.4.8}$$

as stated in the main text.

Appendix B

Appendix for Chapter 2

B.1 Proof of Theorem 1

Sufficiency. Suppose the inequalities of Theorem 1 holds. Define the following utility function as follows:

$$U(\mathbf{q}) = \min_{t\in\mathbf{T}} \left\{u_t + \lambda_t \left[\mathbf{p}'_t + \boldsymbol{\beta}'_t c'(\mathbf{q_t})\right] * (\mathbf{q} - \mathbf{q_t})\right\} \tag{B.1.1}$$

Given that $\mathbf{q} \neq \theta_2$ and $\lambda > 0$, if $\mathbf{p}'_t + \boldsymbol{\beta}'_t c'(\mathbf{q_t}) > 0$, then $U(\mathbf{q})$ is increasing in $\mathbf{q}$.[1] Note that as $c'(\mathbf{q})$ is almost flat, and $\boldsymbol{\beta} \leq \mathbf{p}$, this is not a very restrictive technical assumption. In light of this, define $\tilde{\mathbf{p}}_\mathbf{t} = \mathbf{p}'_t + \boldsymbol{\beta}'_t c'(\mathbf{q_t})$ which yields:

$$U(\mathbf{q}) = \min_{t\in\mathbf{T}} \left\{u_t + \lambda_t \tilde{\mathbf{p}}_\mathbf{t} * (\mathbf{q} - \mathbf{q_t})\right\} \tag{B.1.2}$$

If the inequalities from Theorem 1 are to hold, it must be that they are consistent with a (concave) utility function that rationalises $\mathbf{D}$. This is to say that, it must be shown that $U(\mathbf{q_s}) = u_s$ and $U(\mathbf{q_k}) \leq u_s$, $\forall k \in \mathbf{T}$ when prices are $\mathbf{p_s}$.

$U(\mathbf{q_s}) = \min_{t\in\mathbf{T}} \left\{u_t + \lambda_t \tilde{\mathbf{p}} * (\mathbf{q_s} - \mathbf{q_t})\right\}$ which is clearly minimised at $t = s$. This implies that

$U(\mathbf{q_s}) = u_s$

Now note that $\mathbf{p_s}'\mathbf{q_k} \leq \mathbf{p_s}'\mathbf{q_s} \implies U(\mathbf{q_k}) \leq U(\mathbf{q_s})$

[1] Another desirable property of $c(\mathbf{q})$ is that $c'(\mathbf{q})$ is concave.

By definition $U(\mathbf{q_k}) \leq u_s + \lambda_s \tilde{\mathbf{p_s}} * (\mathbf{q_k} - \mathbf{q_s})$

As $\lambda_s \tilde{\mathbf{p_s}} > 0$ and $\mathbf{p_s}'\mathbf{q_k} \leq \mathbf{p_s}'\mathbf{q_s} \implies$

$U(\mathbf{q_k}) \leq u_s + \lambda_s \tilde{\mathbf{p_s}} * (\mathbf{q_k} - \mathbf{q_s}) \leq u_s \implies$

$U(\mathbf{q_k}) \leq U(\mathbf{q_s})$ which is true $\forall k \in \mathbf{T}$

Necessity. Given the concavity of the utility function, convexity of the constraint, and the first order conditions from the utility maximisation process, the following Afriat-style inequalities can be derived using an appropriate Taylor-series expansion of the utility function.

$$U(\mathbf{q_j}) \leq U(\mathbf{q_i}) + U'(\mathbf{q_i})(\mathbf{q_j} - \mathbf{q_i}) \tag{B.1.3}$$

$$u_j \leq u_i + \lambda_i \left[\mathbf{p}_i' + \boldsymbol{\beta}_i' c'(\mathbf{q})\right] * (\mathbf{q_j} - \mathbf{q_i}) \tag{B.1.4}$$

It is clear that the Karush-Kuhn-Tucker (KKT) conditions have been used to get the above expression for the inequality. The question remains as to whether the KKT conditions are themselves necessary and sufficient for optimality of the consumer maximisation problem. If this is the case, then the inequalities must be necessary for optimality. It is well known that the KKT conditions are necessary conditions for constrained optimisation problems. The problem lies in trying to establish whether they are also sufficient. In the case of a convex optimisation problem, I have to ensure that Slater's condition holds. If this is the case, then the KKT conditions are necessary and sufficient for optimality.[2] Slater's condition says:

If $\exists$ $\mathbf{x} \in \mathbb{R}^n$ such that $\mathbf{p}'\mathbf{x} + c(\mathbf{x}) < y$, there there must exist a solution that maximises the objective function such that the KKT conditions hold. Clearly, it is possible to set $\mathbf{x} >> \theta_2$ but also sufficiently small such that the CSC holds with strict inequality. $\square$

[2]Note that it is not strictly enough that we have convexity of the constraints as the utility function is defined over the entire real line, whereas the constraints are not continuously differentiable at all points. If this were the case, it would again be trivial to proof necessity.

B.2 Table 1

	(1)	(2)	(3)	(4)
	beta	beta	beta_exp	beta_exp
MidFMSize	-0.000925	-0.000925	-0.164	-0.164
	(-13.14)	(-13.13)	(-13.51)	(-13.50)
LarFMSize	0.00236	0.00236	0.247	0.247
	(24.04)	(24.06)	(14.60)	(14.62)
MidIncome	-0.00108	-0.00108	-0.267	-0.267
	(-13.61)	(-13.60)	(-19.47)	(-19.45)
HighIncome	0.000905	0.000906	0.00135	0.00155
	(9.83)	(9.84)	(0.08)	(0.10)
MidAge	-0.00170	-0.00170	-0.301	-0.301
	(-22.43)	(-22.44)	(-23.03)	(-23.04)
OldAge	-0.00112	-0.00112	-0.267	-0.267
	(-11.45)	(-11.46)	(-15.74)	(-15.75)
Highsch	0.000382	0.000383	0.0265	0.0266
	(2.90)	(2.90)	(1.16)	(1.17)
college	0.000935	0.000935	0.147	0.147
	(6.85)	(6.85)	(6.26)	(6.26)
CCEI		-0.00738		-2.004
		(0.54)		(0.84)
1-MPI		-0.00289		-0.467
		(0.34)		(0.32)
Constant	0.00423	0.00423	0.861	0.860
	(26.81)	(26.79)	(31.63)	(31.59)
Observations	156000	156000	156000	156000

t statistics in parentheses

B.3 Dimensionality Issues

B.3.1 Dimensionality Issue

The following subsection below highlights the issue of modelling and deducing consideration sets in a revealed preference approach.

Define a finite dataset, $\mathbf{D} = \{\mathbf{p_t}, \mathbf{q_t}\}_{t \in \mathbf{T}}$

Let D_k be the k^{th} partition of D, $k = 1, ...j$

K_T is the T^{th} Bell number such $K_T = \sum\limits_{z=0}^{T-1} \binom{T-1}{Z} K_z$, $K_0 = 1$.

For example, if T = 4 (i.e. 4 observations), the maximum number of ways of partitioning the data is 52. This means that for a dataset with only 4 observations, there are 15 potential combinations of consideration sets. As T grows large, this grows exponentially; for T = 6/7/8, the maximum number of partitions is 877/4140/21147.

Denoting $card(D_k) = j$ implies there are j consideration sets for dataset, D i.e. the number of partitions denotes the number of consideration sets. Clearly, $\sup_k\{card(D_k)\} = T$.

For illustration, suppose T=3, the partitions of D are:

$S1 = \{\{\mathbf{p_1}, \mathbf{q_1}\}, \{\mathbf{p_2}, \mathbf{q_2}\}, \{\mathbf{p_3}, \mathbf{q_3}\}\}$

$S2 = \{\{\mathbf{p_1}, \mathbf{q_1}, \mathbf{p_2}, \mathbf{q_2}\}, \{\mathbf{p_3}, \mathbf{q_3}\}\}$

$S3 = \{\{\mathbf{p_1}, \mathbf{q_1}, \mathbf{p_3}, \mathbf{q_3}\}, \{\mathbf{p_2}, \mathbf{q_2}\}\}$

$S4 = \{\{\mathbf{p_1}, \mathbf{q_1}\}, \{\mathbf{p_2}, \mathbf{q_2}, \mathbf{p_3}, \mathbf{q_3}\}\}$

$S5 = \{\mathbf{p_1}, \mathbf{q_1}, \mathbf{p_2}, \mathbf{q_2}, \mathbf{p_3}, \mathbf{q_3}\}$

Here, $S1, S2, S4, S5$ make intuitive sense as potential partitions (in terms of consideration sets). In general, the number of possible consideration set configurations will be far fewer than K_T. A sensible assumption would be that once a good is in the consideration set, it remains a part of all future consideration sets. This is also commonly made assumption in the literature. Here, $S3$ implies decisions in $t = 1, 3$ came from one consideration set, but decisions in $t = 2$ are somehow separate. Given the above assumption, choices from $t = 1, 3$ could only have come from a consideration set that involves all periods in-between as well i.e. $S5$ would be the only accommodating partition.

This implies that the new maximum number of ways of partitioning the data is $\hat{K}_T = 2^{T-1}$, $\hat{K}_0 = 1$ which is clearly much lower than K_T.

Appendix C

Appendix for Chapter 3

C.1 Proofs

C.1.1 Theorem 2

As an outline of the proof, it starts with showing that there is a binary relation on the alternatives space which is a semiorder. This essential initial step is required in order to get the JND utility function representation such as in the definitions of the sets above. Then, it must be shown that the choice correspondence only chooses the "near-maximal" elements according to the JND utility function. The final step is then to show that a choice problem with a status quo is equivalent to a choice problem without a status quo, but as if the choice set was reduced, in the sense that only elements that have beaten the status quo remain.

Proof. $\Rightarrow$ Sufficiency

The first step is to prove that the binary relation is a semiorder by showing it satisfies irreflexivity, semitransitivity, and the interval order condition (IOC). These properties will be defined precisely below.

Recall that $y \succ x$ *iff* $y \in c(\{x, y\}, x)$ and $x \notin c(\{x, y\}, y)$.[1]

[1] See Sen (1971) for a comprehensive survey of binary relations and choice functions

$\succ$ is irreflexive if $a \succ a$ is false for all $a \in X$.

Suppose $a, b \in X$ and $a \succ b$. By definition, we have $a \in c(\{a, b\}, b)$ and $b \notin c(\{a, b\}, a)$. As the choice correspondence cannot be empty, by definition, this can never hold true when $a = b$, this implies that $\succ$ is irreflexive.

$\succ$ is semitransitive if $\forall a, b, c, d \in X$, $a \succ b \wedge b \succ c \implies a \succ d \vee d \succ c$.

Starting with $a \in c(\{a, b\}, b)$, $b \notin c(\{a, b\}, a)$, $b \in c(\{b, c\}, c)$, and $c \notin c(\{b, c\}, b)$, by Axiom α, $b \notin c(\{a, b, c, d\}, a)$ and $c \notin c(\{a, b, c, d\}, b)$. This first step will be to show that $a \succ d$ can be obtained. Using Axiom ASQB, initially set $A \cup B = \{a, b, c, d\}(= A)$. Suppose $a \in c(\{a, b, c, d\}, a)$. By Axiom ASQB, $a = c(\{a, b, c, d\}, a)$ which implies $b, c, d \notin c(\{a, b, c, d\}, a)$. Again, by Axiom ASQB, $b, c, d \notin c(\{a, b, c, d\}, b)$. By iterative applications of Axiom ASQB, $b, c, d \notin c(\{a, b, c, d\}, c)$ and $b, c, d \notin c(\{a, b, c, d\}, d)$. This implies that $a \in c(\{a, b, c, d\}, a)$, $a \in c(\{a, b, c, d\}, b)$, $a \in c(\{a, b, c, d\}, c)$, and $a \in c(\{a, b, c, d\}, d)$. By Axiom α, $a \in c(\{a, d\}, d)$. Recall that $a = c(\{a, b, c, d\}, a)$ and Axiom α implies that $a = c(\{a, d\}, a) \implies d \notin c(\{a, d\}, a)$. Hence $a \succ d$ is obtained.

Now, to show that $d \succ c$ can be obtained, suppose that $A \cup B = \{a, b, c, d\}$ with $A = \{a, b, c\}$ and $B = \{d\}$. Again, $b \notin c(\{a, b, c, d\}, a)$ and $c \notin c(\{a, b, c, d\}, b)$. To show $d \succ c$, using Axiom ASQB, $a \in c(\{A \cup B\}, a)$ but now only elements in B can also belong to this set when a becomes the status quo. This means that $b \notin c(\{A \cup B\}, a)$ and $c \notin c(\{A \cup B\}, a)$. By Axiom ASQB, there is also $d \in c(\{A \cup B\}, a)$. However, as $c \notin c(\{A \cup B\}, a)$ this yields $c \notin c(\{A \cup B\}, d)$. By Axiom ASQB, $b, c \notin c(\{A \cup B\}, c)$ which means $a, d \in c(\{A \cup B\}, c)$. So, by Axiom α, $d \in c(\{d, c\}, c)$. To complete the proof for semitransitivity, $c \notin c(\{c, d\}, d)$ needs to be shown. Suppose that $c \in c(\{c, d\}, d)$. As $d \in c(\{c, d\}, d)$ and $d \in c(\{A \cup B\}, d)$, by Axiom δ, $c \in c(\{A \cup B\}, d)$, which is a contradiction. This must mean that $c \notin c(\{c, d\}, d)$. Thus $d \succ c$ is obtained and hence $\succ$ is semitransitive.

$\succ$ satisfies the IOC if $\forall a, b, c, d \in X$, $a \succ b \wedge c \succ d \implies a \succ d \vee c \succ b$.

Starting with $a \in c(\{a,b\},b)$, $b \notin c(\{a,b\},a)$, $c \in c(\{c,d\},d)$, and $d \notin c(\{c,d\},c)$, by Axiom α, $b \notin c(\{a,b,c,d\},a)$ and $d \notin c(\{a,b,c,d\},c)$. By Axiom DSQB, there are two cases to consider. The case where only $\{a,b\}$ or $\{c,d\}$ have (potential) status quo effects.

Consider the first case. By Axiom DSQB, $a \in c(\{a,b,c,d\},a)$ and $c \notin c(\{a,b,c,d\},a)$. By Axiom DSQB or Axiom ASQB, $b \notin c(\{a,b,c,d\},a)$ and $d \notin c(\{a,b,c,d\},a)$ as $b \notin c(\{a,b,c,d\},a)$ and $d \notin c(\{a,b,c,d\},c)$. However, this implies that $a = c(\{a,b,c,d\},a)$, which by Axiom αimplies that $d \notin c(\{a,d\},a)$. Now suppose that $a \notin c(\{a,b,c,d\},d)$, but this cannot be true by Axiom DSQB or Axiom ASQB, as $a \in c(\{a,b,c,d\},a)$. So, $a \in c(\{a,b,c,d\},d)$ and by Axiom α, $a \in c(\{a,d\},d)$. Hence $a \succ d$ is obtained.

In the other case, using Axiom DSQB yields $c \in c(\{a,b,c,d\},c)$ and $a \notin c(\{a,b,c,d\},c)$. By symmetry of the above argument, $c \succ b$ is obtained.

Combined with the first case, this shows that $\succ$ satisfies the interval order condition.

As per above, $\succ$ is an irreflexive semitransitive binary relation that satisfies the interval order condition, hence, $\succ$ is a semiorder on X.

The rest of the proof follows that of Lemma 1 and Theorem 1 of Masatlioglu and Ok (2005). Define $D(x) \equiv \{y \in S | y \succ x\}$ and $d(x) \equiv \{y \in S : \neg y \succ x \wedge \neg x \succ y\}$. Let $\bowtie_i$ denote the binary relation in the absence of $\succ_i$ or $\prec_i$.

First assume $D(x) = \emptyset$. Suppose $y \in c(S,x)$, by Axiom α, $y \in c(\{x,y\},x)$. By Axiom ASQB/Axiom DSQB, $y = c(\{x,y\},y)$ which implies that $x \notin c(\{x,y\},y)$ which yields the contradiction that $D(x) \neq \emptyset$. This means that $y \succ x$ is not possible for any $y \in S$. If $\neg(y \succ x)$, this means that either $x \succ y$ or $x \bowtie y$. If it is the case that all y that $x \succ y$, then clearly $c(S,x) = \{x\}$. However, as it is possible that there are some $y \in S$ such that $x \bowtie y$, then $x \in c(S,x) = d(x)$.

Now assume that $D(x) \neq \emptyset$. Suppose $y \in c(S,x)$, by Axiom α, $y \in c(\{x,y\},x)$. By Axiom ASQB/Axiom DSQB, $y = c(\{x,y\},y)$ which implies that $x \notin c(\{x,y\},y)$ which implies that $y \in D(x)$. If $y = x$ then $y \notin c(\{x,y\},x)$, $\forall y \in S$ implying

$D(x) = \emptyset$. By contradiction, it must be that $c(S, x) \subseteq D(x)$ as no y can be equal to x.

To complete the sufficiency part of the proof, it must be shown that a choice problem with a status quo is equivalent to a choice problem without a status quo but with goods only superior to the status quo, i.e. when $D(x) \neq \emptyset$ then $c(S, x) = c(D(x), \diamond)$. Starting with $c(S, x)$, as $D(x) \neq \emptyset$, this implies that there exists $y \in c(S, x)$. By Axiom α , $y \in c(D(x) \cup \{x\}, x)$. As $x \notin c(D(x) \cup \{x\}, x)$ for any subset of $D(x) \cup \{x\}$, applying Axiom SQI , it must be that $y \in c(D(x) \cup \{x\}, \diamond)$. Applying Axiom α yields $y \in c(D(x), \diamond)$ as required.
Now suppose that $y, z \in c(D(x), \diamond)$. By Axiom δ, if $z \in c(D(x) \cup \{x\}, \diamond)$, then it must be that $y \in c(D(x) \cup \{x\}, \diamond)$ as $z \neq c(D(x) \cup \{x\}, \diamond)$. Also, as $D(x) \neq \emptyset$, this implies that $x \notin c(\{x, y\}, x)$ implying $y = c(\{x, y\}, x)$. By Axiom D, $y \in c(D(x) \cup \{x\}, x)$, which by Axiom α gives $y \in c(D(x), x)$. For an arbitrary $w \in c(S, x)$, by Axiom α, $w \in c(D(x), x)$. Applying Axiom δ yields $y \in c(S, x)$ as required.

Taking the choice correspondence that satisfies Axiom α, Axiom δ, Axiom D, Axiom SQI, Axiom ASQB, and Axiom DSQB, there exists a $u : S \to \mathbb{R}$ and a ϵ such that $y \succ x \iff u(y) > u(x) + \epsilon$. As the choice correspondence satisfies Axiom α, Axiom δ, Axiom ASQB, and Axiom DSQB, and $|S| < \infty$, it must be that there exists a subset of the choice set that has only maximal elements according to a just noticeable difference utility function i.e. $c(S, \diamond) = \Omega(S)$, $\forall S$. This is also trivially true for when there is a status quo i.e. $c(S, x) = \Omega(\Gamma(S, x))$. Additionally, as shown above, when $D(x) = \emptyset$, then $x \in c(S, x) = d(x) \equiv \{y \in S : \neg y \succ x \wedge \neg x \succ y\}$. In combination with the above axioms, it must be that when $D(x) = \emptyset$, then $c(S, x) = \Pi(S, x)$.

$\Leftarrow$ Necessity

Given the choice correspondence and representation above, Axiom α, Axiom δ, Axiom D, Axiom SQI, Axiom ASQB, and Axiom DSQB, need to be proven.[2]

Axiom α: $y \in c(A, x) = \Omega(\Gamma(A, x))$. As $y \in B \subseteq A$, $y \in \Omega(\Gamma(B, x)) \implies y \in c(B, x)$.

Axiom δ: $\{z, y\} \in \Omega(\Gamma(S, x))$. As $y \in S \subseteq T$, for any ϵ, it can never be that y is the unique choice as z is sufficiently similar i.e. it cannot be the case that $\{y\} \in \Omega(\Gamma(T, x))$

Axiom D: $\{y\} = c(S, x) = \Omega(\Gamma(S, x))$, $S \subseteq T$ and $y \in c(T, \diamond) = \Omega(\Gamma(T, \diamond))$. As $x \in S \subseteq T$ and $x \in T$, then, $y \in \Gamma(T, x)$. As $y \in \Omega(\Gamma(T, \diamond))$ and $y \in \Gamma(T, x)$, then $y \in \Omega(\Gamma(T, x)) \implies y \in c(T, x)$.

Axiom SQI: $y \in \Omega(\Gamma(S, x))$. As $\nexists$ any $T \subseteq S$ such that $x \in c(T, x)$, this means that $\Omega(\Gamma(S, x)) = \Omega(\Gamma(S, \diamond))$. Hence, $y \in \Omega(\Gamma(S, \diamond)) \implies y \in c(S, \diamond)$

Axiom ASQB: $y \in \Omega(\Gamma(A, x))$ and $y \in \Omega(\Gamma(A \cup B, x))$. By definition, $y \in \Gamma(A \cup B, x) = \{p \in A \cup B : v(p) > v(x) + \epsilon\}$. As y is part of the set which beats x in $A \cup B$, there does not exist any element that outright beats y as $y \in c(A \cup B, x)$. This implies that $\{p \in A \cup B : v(p) > v(y) + \epsilon\} = \emptyset$. By definition, $y \in c(A \cup B, y) = \Pi(A \cup B, y)$. Note that $y \in c(A, y) = \Pi(A, y)$ is also true as $y \in \Omega(\Gamma(A, x))$. As $y \in \Pi(A \cup B, y)$ must also hold, when the choice set is augmented from A to $A \cup B$, then $y \cup B' \in \Pi(A \cup B, y)$, for some potentially nonempty $B' \subseteq B$.

Axiom DSQB: $y \in \Omega(\Gamma(A, x))$ and $v \in \Omega(\Gamma(B, w))$. These imply that $y \in c(A, y) = \Pi(A, y)$ and $v \in c(B, v) = \Pi(B, v)$ as above. When choices become $A \cup B$ then there can exist some $B' \subseteq B$ such that $y \cup B' \subseteq \Gamma(A \cup B, y) = c(A \cup B, y)$. Alternatively, the other possibility is that $v \cup A' \subseteq \Gamma(A \cup B, v) = c(A \cup B, v)$ for

[2]Showing the axioms hold is a fairly trivial task, but is included for constructiveness and exposition.

some $A' \subseteq A$. Considering the case when $y \in c(A \cup B, y)$, then as $y \in \Omega(\Gamma(A, x))$ for any $x \in A$, in particular, when $v = x$. By definition of $\Gamma(\cdot,\cdot)$, it must be that y is not dominated by v, and, by definition, $v \notin c(A \cup B, y)$. □

C.1.2 Corollary 2.1

Proof. Virtually identical to the proof of Theorem 2 except in the last step where an interval order utility exists such that $y \succ_{IO} x \iff u(y) > u(x) + \epsilon(x)$. Note that this is one form of the representation; as by the nomenclature, it is possible to assign utility as closed intervals of $\mathbb{R}$ to alternatives such that $y \succ_{\hat{IO}} x \iff a_y > b_x$ where $u_x = [a_x, b_x]$ and $u_y = [a_y, b_y]$.[3] □

C.1.3 Implication 2.1

Proof. Apply the proof of Theorem 2, and use the fact that 1) Axiom SQB can be formed as a special case of combining both Axiom ASQB and Axiom DSQB and setting $B = \emptyset$. Recall the definition for Axiom β, for any (B, x), if $z, y \in c(B, x)$, $B \subseteq A$, and $z \in c(A, x)$. Clearly, if this holds, then Axiom δ holds. □

C.1.4 Lemma 2.1

As a brief outline of the theorem below, what needs to be shown is how a change in a single characteristic can effect the overall choice of the good. This needs to be true for all of the characteristics in order to establish logical consistency with the aggregation. The proof closely follows the literature on social welfare functions

Proof. $\Rightarrow$ Sufficiency

Let $\{z_1, ..., z_J\}$ be denoted by Z and $\{z_1, ..., x_j, ..., z_J\}$ denoted by (z_{-j}, x_j).

What ultimately needs to be proven is that that when $u_j(x_j) = u_j(y_j)$ for all $j \leq J$, that $U(X) = U(Y)$.

In order for this aggregation to represent all the characteristics, the following need to be shown:

[3] It is easy to go from one form of representation to the other, but the chosen initial representation is easier given how Theorem 2 is proven.

if $u_j(x_j) \geq u_j(y_j) \implies U(z_{-j}, x_j) \geq U(z_{-j}, y_j)$ and $u_j(x_j) > u_j(y_j) \implies U(z_{-j}, x_j) > U(z_{-j}, y_j)$.

Firstly, assume that $u_j(x_j) \geq u_j(y_j) \implies U(x) < U(y)$. By assumption, there exists a w_i such that $U(z_{-j}, y_j) > U(z_{-j}, w_j) > U(z_{-j}, x_j)$. By SC, there are 4 possible chains of preference that can occur per characteristic i.e. A $y_j \succ_j w_j \succ_j x_j$, B $y_j \succ_j w_j \bowtie_j^+ x_j$, C $y_j \bowtie_j^+ w_j \succ_j x_j$, and D $y_j \bowtie_j^+ w_j \bowtie_j^+ x_j$.

Starting with A, $u_j(y_j)$ - $u_j(w_j) > \epsilon_j$ and $u_j(w_j)$ - $u_j(x_j) > \epsilon_j$ imply by addition that $u_j(y_j)$ - $u_j(x_j) > 2\epsilon_j$. This means that $u_j(y_j)$ - $u_j(x_j) > 0$ (as $\epsilon_j > 0$) and $u_j(y_j) > u_j(x_j)$ which is a contradiction.

B yields $u_j(y_j)$ - $u_j(w_j) > \epsilon_j$ and $0 < u_j(w_j)$ - $u_j(x_j) \leq \epsilon_j$

$$0 < u_j(w_j) - u_j(x_j) \leq \epsilon_j < u_j(y_j) - u_j(w_j)$$
$$2u_j(w_j) - 2u_j(x_j) < u_j(y_j) - u_j(x_j)$$

as $u_j(w_j) - u_j(x_j) > 0$

$$0 < u_j(y_j) - u_j(x_j)$$
$$u_j(y_j) > u_j(x_j)$$

which is a contradiction.

Similarly C yields $u_j(w_j)$ - $u_j(x_j) > \epsilon_j$ and $0 < u_j(y_j)$ - $u_j(w_j) \leq \epsilon_j$

$$0 < u_j(y_j) - u_j(w_j) \leq \epsilon_j < u_j(w_j) - u_j(x_j)$$
$$2u_j(y_j) - 2u_j(w_j) < u_j(y_j) - u_j(x_j)$$
$$0 < u_j(y_j) - u_j(x_j)$$
$$u_j(y_j) > u_j(x_j)$$

which again is a contradiction as $u_j(y_j) - u_j(w_j) > 0$.

D gives $0 < u_j(y_j) - u_j(w_j) \leq \epsilon_j$ and $0 < u_j(w_j) - u_j(x_j) \leq \epsilon_j$. By addition, $0 < u_j(y_j) - u_j(x_j) \leq 2\epsilon_j$ which leads to $u_j(y_j) - u_j(x_j) > 0$ which is the final contradiction. So if $u_j(x_j) \geq u_j(y_j) \implies U(z_{-j}, x_j) \geq U(z_{-j}, y_j)$ holds for any $j \leq J$.

To show $u_j(x_j) > u_j(y_j) \implies U(z_{-j}, x_j) > U(z_{-j}, y_j)$, given the range of $u_i(\cdot)$, there exists a w_i such that $u_j(x_j) > u_j(w_j) + \epsilon_j > u_j(y_j) + \epsilon_j > u_j(w_j) > u_j(y_j)$. By definition, $x_j \succ_j w_j$ and $\neg[y_j \succ_j w_j] \wedge y_j \bowtie_j^- w_j$.[4] By SC', $(z_{-j}, x_j)\mathbb{W}(z_{-j}, w_j)$ and $\neg[(z_{-j}, y_j)\mathbb{W}(z_{-j}, w_j)]$. By definition,

$$U(z_{-j}, x_j) > U(z_{-j}, w_j) \geq U(z_{-j}, y_j)$$
$$U(z_{-j}, x_j) > U(z_{-j}, y_j)$$

as required.

To finish the proof, note that $u_j(x_j) \geq u_j(y_j) \implies U(z_{-j}, x_j) \geq U(z_{-j}, y_j)$ holds for any $j \leq J$. As this holds for any $j \leq J$ and for arbitrary (z_{-j}, z_j), then by an inductive argument, it must be the case that if $u_j(x_j) \geq u_j(y_j)$ for all $j \leq J$, then $U(X) \geq U(Y)$. Note that if $u_j(x_j) \leq u_j(y_j)$ for all $j \leq J$, then $U(X) \leq U(Y)$ which implies that if $u_j(x_j) = u_j(y_j)$ for all $j \leq J$, then $U(X) = U(Y)$ where $U(X) = g(u_1(x_1), ..., u_J(x_J))$. Given the above constructive proof, $g(\cdot)$ is clearly strictly monotonic.

$\Leftarrow$ Necessity

Assume there exists a strictly monotone function $g : \mathbb{R}^n \to \mathbb{R}$ such that $U(X) = g(u_1(x_1), ..., u_J(x_J))$. Starting with SC, assume $(z_{-j}, x_j)\mathbb{W}(z_{-j}, y_j)$ which means that $U(z_{-j}, x_j) > U(z_{-j}, y_j)$. By definition:

$$g(u_1(x_1), ..., u_j(x_j), ..., u_J(x_J)) > g(u_1(x_1), ..., u_j(y_j), ..., u_J(x_J)).$$

As $g(\cdot)$ is strictly monotonic, this implies $u_j(x_j) > u_j(y_j)$. This means that either $u_j(x_j) > u_j(y_j) + \epsilon_j > u_j(y_j)$ or $u_j(y_j) > u_j(x_j) > u_j(y_j)$ which yield $x_j \succ_j y_j$ and $x_j \bowtie_j^+ y_j$, respectively. Now, assume that $x_j \succ_j y_j$ which means $u_j(x_j) > u_j(y_j) + \epsilon_j$. This implies that $u_j(x_j) > u_j(y_j)$. It must be true that when comparing

[4]Note that $\neg[w_j \succ_j y_j]$ is obtained but it is not true that $w_j \bowtie_j^- y_j$. In fact, this simply corroborates the fact that y_j and w_j are indistinguishable with y_j worse than w_j in pure utility.

all the characteristics such that $g(u_{-j}(x_{-j}), u_j(x_j))$ with $g(u_{-j}(x_{-j}), u_j(y_j))$ that $g(u_{-j}(x_{-j}), u_j(x_j)) > g(u_{-j}(x_{-j}), u_j(y_j))$. By definition, $(z_{-j}, x_j)\mathbb{W}(z_{-j}, y_j)$. As this holds for any arbitrary (z_{-j}, z_j), this completes the necessity of SC.[5] $\square$

C.2 Aggregation

Analogously to Masatlioglu and Ok (2005), suppose now that the domain of options is expanded such that there are set of finite alternatives which are characterised by a set of finite features/characteristics i.e. an alternative is defined by a list of characteristics. The status quo could then potentially eliminate alternatives that are worse than it in *any* dimension/characteristic. The agent, in principle, can have a choice correspondence *for each characteristic* of the alternatives. This can be achieved simply by applying Theorem 2 to each of the characteristics of the alternatives. This means that the agent has semi-ordered preferences represented by JND utility for each characteristic of the alternatives.[6]

The aim for this section is to tackle the issue of how an agent makes a decision after analysing just the characteristics of the goods. In essence, is there a way to consistently aggregate such semi-ordered preferences? How does an agent make the final decision over alternatives from choices over characteristics? Returning to our common example of buying a mobile phone, with the options being an iPhone, Samsung, and Huawei. Suppose they are characterised by camera quality, screen quality, and battery life. Conditional on each characteristic, the agent is able to make a choice. The agent may not be able to compare each characteristic perfectly

[5]The proof of the necessity of SC' is shown for exposition although virtually identical to SC given it is its natural negation. For SC', assume that $\neg[x_j \succ_j y_j] \wedge x_i \bowtie_i^+ y_i$. By definition, $\neg[u_j(x_j) > u_j(y_j) + \epsilon_j] \implies u_j(x_j) \leq u_j(y_j) + \epsilon_j$ and $0 < u_j(x_j) - u_j(y_j) \leq \epsilon_j$. Re-arranging gives $u_j(y_j) < u_j(x_j) \leq u_j(y_j) + \epsilon_j$. As $u_j(y_j) < u_j(x_j)$ has to hold, by monotonicity of $g(\cdot)$, it must be that, for any arbitrary (z_{-j}, z_j) that $(z_{-j}, x_j)\mathbb{W}(z_{-j}, y_j)$. For the second part of SC' , $\neg[x_j \succ_j y_j] \wedge x_i \bowtie_i^- y_i$. Similarly, $u_j(x_j) \leq u_j(y_j) + \epsilon_j$ and $0 < u_j(y_j) - u_j(x_j) \leq \epsilon_j$ which after re-arranging gives $u_j(x_j) < u_j(y_j) \leq u_j(x_j) + \epsilon_j$. Again, by monotonicity of $g(\cdot)$, it must be that $(z_{-j}, x_j)\mathbb{W}(z_{-j}, y_j)$.

[6]Previously, the derived utility function was only single-valued. However, it may be desirable to have a multi-valued utility functions as a way of representing an alternative, as the alternatives can now be described by characteristics. Before, the role of the status quo was moderate as it is always easy to compare utility numbers. However, if there were a vector-valued utility, then the status quo could eliminate on the basis of any of these characteristics.

as it may be difficult to perceive technological differences and/or characteristics may be approximately similar to human perception, etc... Given each of those conditional choices, is there a way to make a final decision between an iPhone, Samsung, and Huawei? The purpose of this section is to answer this type of question i.e. given the characteristic choices, what restrictions do we need on the overall preferences to get a consistent choice?

As alluded to, the ensuing consistency restrictions for aggregation will use preferences over alternatives as the premise for the following analysis. One reason for this is to bring it in line with social choice theory where individuals' preferences are aggregated by a social planner. Analogously, an individual may wish to aggregate characteristics of alternatives, in order to establish a ranking over the alternatives themselves. For example, if someone prefers the iPhone camera over a Samsung camera over a Huawei camera, but prefers the Huawei screen over the iPhone screen over the Samsung screen, is there a way to rank the iPhone, Samsung, and Huawei phones in a consistent way? So, if we observe the semi-ordered preferences of the agent over characteristics, what consistency restrictions are required on semi-ordered/weak-ordered preferences of the alternatives? This is what is addressed below; given that semi-ordered preferences over characteristics are observed, what restrictions do we need to get consistent preferences over the alternatives?

The interesting contribution here is that the semi-ordered preferences over characteristics are not just simply assumed, but rather, have an axiomatic foundation in status quo bias. In other words, it is assumed that there is a semi-ordering in preferences over characteristics, but the behavioural explanation for this is status quo bias. Thus, for the following analysis, the full preference ranking over characteristics is needed to get the preference ranking over the alternatives.

Typically, in order to derive a (finite) vector-valued utility function, it is common to use Szpilrajn's Theorem (Szpilrajn (1930)).[7] In brief summary, by taking all of the linear orders $C_{\{1,\ldots,i,\ldots,n\}}$ such that $\succ \subseteq C_i \ \forall i$, where $\succ$ is a weak partial

[7]Masatlioglu and Ok (2005) and Ok (2002) for further details.

order (WPO), standard results give us that these linear orders are representable by regular utility functions.[8] Stacking these standard utility functions is then what gives the vector-valued utility function.[9] However, as described previously, it may be the case that agents have semi-ordered preferences over characteristics. This means that this type of approach will not be appropriate as what would be required is stacking/aggregating semiorders rather than standard weak orders. Before showing an adequate method of aggregation, Example 4 illustrates what it means to aggregate semi-ordered preferences.

Example 3 Suppose that $X \equiv \{i, j, k, l, m, x\}$ with $\mathbf{u}(x) = (4, 4, 4)$, $\mathbf{u}(i) = (0, 0, 0)$, $\mathbf{u}(j) = (4, 5, 7)$, $\mathbf{u}(k) = (4, 2, 4)$, $\mathbf{u}(l) = (7, 6, 5)$, and $\mathbf{u}(m) = (8, 7, 6)$. If the JND for each dimension of the vector utility is $\epsilon_1 = \epsilon_2 = \epsilon_3 = 1$, the status quo, x, eliminates i and k as x is noticeably better than k in the $2nd$ dimension of utility, and noticeably better than i in all dimensions of utility. However, what should be chosen from j, l, and m, given the semi-ordered preferences in each dimension of utility? What is clear is that m is better (although not noticeably) than l in all dimensions. It would make sense, ex-ante, that l would effectively be eliminated by m. Indeed, the final choice of good will always depend on the intrinsic preferences of the agent, however, the aggregation should exhibit a form of monotonicity (in vectors), as described with l and m. Suppose the aggregator is simply the sum of elements, then m is clearly the final good chosen. Even though the characteristics are not noticeably different, when the agent is judging a product as an entire set of characteristics, then the agent can take into account that a product is inferior even if it is marginally worse.

This example illustrates the fact that, although characteristics of certain products are not so easy to compare, the inferiority/superiority of a characteristic

[8] A WPO is a binary relation that is reflexive, transitive, and antisymmetric.

[9] Note that a semiorder is just a strict partial order (SPO). A WPO can be converted to an SPO by removing its reflexive components i.e. the SPO is the corresponding irreflexive kernel such that *only* different elements can be compared. By working with the SPO and putting structure on it such that it yields a semiorder, and then adjoining the semiorder with its reflexive component gets back the WPO. This is required as Szpilrajn's Theorem requires working with the WPO. This is important detail in showing how the representation of Masatlioglu and Ok (2005) is also obtainable from the above representation. This is guaranteed by the fact that Axiom ASQB and Axiom DSQB imply Axiom SQB but not conversely.

becomes more salient when comparing products as described by their features. Consider another basic example of choosing a phone as described by its camera, screen, and price. Conditional on the prices being similar, it may not be clear which camera / screen is best suited for the agent, however, when buying their phone, ultimately, the agent may realise that it is better to buy the phone which is marginally "better", even if the individual characteristics themselves are difficult to compare.

Another reason for allowing such an aggregation is to ensure that incompleteness is not a feature of this model i.e. the agent should always be able to compare/make a choice between alternatives. Here, incompleteness arises due to the assumed vector-valued utility.[10] Aggregation of the vectors is an obvious way to allow the agent to make easy comparisons.[11] The following axiom is introduced in order to be able to characterise such behaviour.

Denote $\{z_1, ..., x_j, ..., z_J\}$ by (z_{-j}, x_j). Let $\mathbb{W}$ denote the weak order for aggregate semi-ordered preferences i.e. $(x_{-j}, x_j)\mathbb{W}(z_{-j}, x_j)$ when $U(x_{-j}, x_j) - U(z_{-j}, x_j) > 0$. Let $\succ_j$ denote a semiorder for characteristic j. Assume that j is finite such that $j \in \{1, ..., J\}$.

Let $\bowtie_i$ denote the binary relation when neither $\succ_i$ or $\prec_i$ hold i.e. $x_i \bowtie_i y_i$ if, and only if, $|u_i(x_i) - u_i(y_i)| \leq \epsilon_i$. For what is about to become clear, let $\bowtie_i^+$ denote the "positive" part of $\bowtie_i$ such that $x_i \bowtie_i^+ y_i$ if, and only if, $0 < u_i(x_i) - u_i(y_i) \leq \epsilon_i$ and let $\bowtie_i^-$ denote the "negative" part of $\bowtie_i$ such that $x_i \bowtie_i^- y_i$ if, and only if, $0 < u_i(y_i) - u_i(x_i) \leq \epsilon_i$.[12]

Strong Consistency (SC) If $(x_{-j}, x_j)\mathbb{W}(x_{-j}, y_j) \iff x_j \succ_j y_j$ or $x_j \bowtie_j^+ y_j$

For expositional purposes, let **SC'** be the natural negation of **SC**.

[10]The $\geq/\leq$ binary relation in the vector space is not complete.

[11]Manzini and Mariotti (2012) have a lexicographic choice procedure as a way of decision making over an ordered list of semiorders.

[12]Both $\bowtie_i^+$ and $\bowtie_i^-$ are not required as they are logically dependent, however, they are defined purely for clarity of the axiom.

$$\textbf{SC'}\quad \begin{cases} \text{If } \neg[x_j \succ_j y_j] \wedge x_j \bowtie_j^{+} y_j \iff (x_{-j}, x_j)\mathbb{W}(x_{-j}, y_j) \\ \text{or} \\ \text{If } \neg[x_j \succ_j y_j] \wedge x_j \bowtie_j^{-} y_j \iff \neg[(x_{-j}, x_j)\mathbb{W}(x_{-j}, y_j)] \end{cases}$$

SC is a natural form of monotonicity. In one direction, it says that if the agent prefers a good (as completely described by its characteristics) due to one particular characteristic difference, in isolation, the agent must either noticeably or "not noticeably" prefer that particular characteristic of the good. The converse simply says that the choice of the characteristic defines the choice of the good, ceteris paribus. This axiom is saying that if an agent chooses a good, at the very least, this choice was made because the good was strictly better in one particular characteristic, noticeably or otherwise.

This axiom highlights an interesting issue of aggregating the semi-ordered preferences with a weak order. In order to achieve this form of consistency between the weak order and the semiorders, what is required is that the agent may not have noticed that a particular characteristic is better. Overall, the agent is still able to make the best choice, even with imprecision in comparability of this dominating characteristic. However, the agent had to know, at the very least, that the good was definitely not worse. In this sense, the agent is able to better judge the good as a whole set of characteristics than just the characteristics in isolation. An equivalent interpretation could be that the agent may know a characteristic is better or worse, but, in the specific choice of that characteristic, the agent is not concerned if the characteristics are sufficiently similar.

SC' is the natural negation of SC . It requires that if the agent does not prefer a characteristic of a good, then the agent does not prefer the overall good as a result of the characteristic, noticeably or otherwise. Note that the negation of $x_j \succ_j y_j$ means that the agent may actually just prefer y_j over x_j or that y_j and x_j are indistinguishable. This means that the agent may actually get higher utility from x_j but not noticeably more than y_j. As the weak order is meant to represent noticeable choices, this axiom rules out behaviour where the agent may choose a good even if it is worse, but not noticeably, in some characteristic. As with SC, this is a form of logical consistency in choices in order to aggregate the semiorders with a weak order.

So, under SC (or SC'), and the assumption that the agent has weak-ordered preferences over the goods, the aggregator for the semi-ordered preferences is also representable by a weak order.

Lemma 2.1. *Suppose there are semiorders* $\{\succ_j\}_{j\in\{1,...,J\}}$ *represented by* $\{u_j, \epsilon_j\}_{j\in\{1,...,J\}}$ *with weak order (over goods)* $\mathbb{W}$ *and corresponding* U *as above. Assume that* $\{u_j\}_{j\in\{1,...,J\}}$ *and* U *are continuous with range* $\mathbb{R}$*. Strong Consistency (or SC') holds, if, and only if, there exists a strictly monotone aggregator* $g : \mathbb{R}^J \to \mathbb{R}$ *such that* $U(x_1, ..., x_j, ..., x_J) = g(u_1(x_1), ..., u_J(x_J))$

With Lemma 2.1, it is possible to aggregate the utilities of each characteristic of a good with an aggregator that is strictly monotonic; moreover, this aggregator can represent a weak order. This means that an agent is able to make standard choices over goods even if there are degrees of imperceptibility in their characteristics, where this difficulty to perceive may be due to status quo bias.

Building on this, the main representation theorem gave rise to a JND utility representation of choices. As suggested, choices can be made per characteristic of each alternative (in the phone example, screen quality, camera, etc...), rather than simply the alternatives themselves. If we observe this full ranking of preferences per characteristic, the above lemma suggests that a ranking over the alternatives can be made in a way that is consistent with the semi-ordering in the preferences of the characteristics. Specifically, the ranking over alternatives can be governed by a weak order.

Consider the example above of buying a mobile phone, with the options being an iPhone, Samsung, and Huawei, characterised by camera quality, screen quality, and battery life. Suppose the consumer is comparing phones that do not vary in camera quality or screen quality. In the battery life characteristic, the agent is still indifferent but only within their JND i.e. the iPhone lasts for 12 hours, Samsung for 11.5 hours, and Huawei for 11 hours but for the consumer, these are all similar enough to be indifferent. The above lemma says that, in the final choice of product, the overall preferred option is the iPhone as it is slightly better in the battery life characteristic. In some sense, if there is no change in any of the characteristics but one, the tie break is decided by that one characteristic. So, in the end, the agent just chooses the iPhone, even if they are not particularly fussed by the improved battery life, given the identicalness of the other characteristics.

However, suppose that, even after aggregation of product characteristics, the agent is still not able to perfectly perceive whether a good is preferred to another. In essence, after aggregation, the overall choice of good may also be semi-ordered. The following axiom comes directly from Argenziano and Gilboa (2018) and deals perfectly with how to aggregate semi-ordered preferences with a semiorder.[13]
Let $\mathbb{S}$ denote the semiorder for aggregate semi-ordered preferences i.e. $(x_{-j}, x_j)\mathbb{S}(x_{-j}, y_j)$ when $U_s(x_{-j}, x_j) - U_s(x_{-j}, y_j) > \epsilon_s$

Weak Consistency (WC) $(x_1, ..., x_j, ..., x_J)\mathbb{S}(x_1, ..., y_j, ..., x_J) \iff x_j \succ_j y_j$

This axiom says that, if all but one of the characteristics of a good differ, it must be that this ordering is preserved when looking at that particular characteristic in isolation. The other direction is simpler in that it says if the individual has preferences in one characteristic and the other characteristics do not change, then the aggregation is consistent with this ordering. Alongside WC, with the assumption that preferences over the goods are semi-ordered, the aggregator for the semi-ordered preferences is also representable by a semiorder. This is because the sensitivity of the characteristic semiorders is represented by the aggregate semiorder.

From above, it is clear why this axiom is not restrictive enough if the aggregation is to be governed by a weak order. In essence, WC does not deal with the imperceptibility across characteristics, as it does not need to given that the aggregation is also in accordance with imperceptibility. Hence, SC is a stronger condition on behaviour but also leads to a stricter representation of that behaviour.

[13] Argenziano and Gilboa (2018) name this simply *Consistency*, I re-name it weak consistency for clarity of exposition.

Lemma 2.2. *Suppose there are semiorders* $\{\succ_j\}_{j\in\{1,...,J\}}$ *represented by* $\{u_j, \epsilon_j\}_{j\in\{1,...,J\}}$ *as in Theorem 2, with semiorder (over goods)* $\mathbb{S}$ *and corresponding* U_s *as above. Assume that* $\{u_j\}_{j\in\{1,...,J\}}$ *and* U *are continuous with range* $\mathbb{R}$. *Weak Consistency holds, if, and only if, there exists a strictly monotone aggregator* $g : \mathbb{R}^J \to \mathbb{R}$ *such that* $U_s(x_1, ..., x_j, ..., x_J) = g(u_1(x_1), ..., u_J(x_J))$

Proof. Theorem 1 of Argenziano and Gilboa (2018)[14] □

This lemma suggests that aggregation of the product characteristics can also be governed by a semiorder as well. Consider the same example above of buying a mobile phone, with the options being an iPhone, Samsung, and Huawei, characterised by camera quality, screen quality, and battery life. Suppose the consumer is comparing phones that do not vary in camera quality or screen quality. In the battery life characteristic, the agent exhibits a strict preference for the iPhone's battery life (e.g. iPhone lasts for 12 hours as opposed to Samsung for 10 hours, and Huawei for 8 hours). As this is the only characteristic that exhibits a strict preference, then the agent chooses the iPhone. Thus, conditional on all the other characteristics being the same, the strict preference from the remaining characteristic carries through to the end choice.

[14] Argenziano and Gilboa (2018) take the representation a step further for their context. They show there is also an additive structure possible if they enforce choices on a JND-grid (a subset of choices where utility differences are multiples of the JND). In the current context of this chapter, this would not be entirely appropriate, but, what is required and still holds, is that such an aggregator does exist.