Stochastic choice analysis with mismeasured data*

Henrik Petri†

November 16, 2021

Abstract: Measurement error is an important concern in empirical and experimental work on discrete choice. Ignoring it may lead to an (under) over rejection problem, whereby observed choices appear (in)consistent with a theory of choice due to measurement error rather than underlying choice behavior. Our results are twofold. On the one hand, we show that there are several models of choice that are immune/robust to "small" amounts of measurement error and hence not as exposed to the over/under rejection problem. On the other hand, we identify models that are sensitive to such issues. We discuss methods to study these models and to alleviate the issues of over/under rejection.

Keywords: stochastic choice, measurement error, empirical content

JEL codes: D01, D90

*I thank seminar and conference participants at McGill University, University of Bath and EEA-ESEM 2021 for comments. I am grateful to Rohan Dutta, Larry Epstein, Sean Horan, Jay Lu, Paola Manzini, Marco Mariotti, Shasi Nandeibam, Efe A. Ok and Evan Piermont for valuable suggestions and comments. An earlier version of this manuscript was circulated under the title "Imprecise Data and Stochastic Choice".

†Department of Economics, University of Bath, henrik@petri.se.

1. Introduction

Measurement error is omnipresent in empirical and experimental work on discrete choice. Formally, measurement error is the difference between the measured value of a quantity (observed choice frequencies) and its true but unobserved value (underlying choice frequencies). There are several reasons why measurement error arises in choice data: participants may choose not to respond to a survey, they may have difficulties understanding an experimental design or they may lack the incentives to reveal their true choices. By its very nature, measurement error may (quite drastically) alter the empirical conclusions drawn from choice data (see e.g. Varian (1985), Echenique, Lee, and Shum (2011) and Aguiar and Kashaev (2018) for a discussion of related issues in the revealed preference framework). Understanding the impact of measurement error when evaluating theories of choice is therefore important.

The recent literature has seen a surge in models/theories aiming at explaining choice behavior. A key step in obtaining a better understanding of these models, is the characterization of the behavior consistent with them. The aim is to find a set of properties describing their behavior and the idea is that it should be possible to verify/test these properties by using a *representative* sample of choices. This exercise not only helps in understanding the scope and limitations of the models, but may also guide the empiricist in designing tests of the models.

Due to measurement error, representative samples are not always available, and further, it is not necessarily the case that properties shown to hold using representative data continue to hold when data is subject to measurement error. Failing to take measurement error into account may lead to an (over) under rejection problem whereby observed choices appear (in)consistent with a model due to measurement error rather than underlying choice behavior.[1] Given the prevalence of measurement error, the aim of this paper is thus to develop methods to understand *if* and *how* measurement error impacts the empirical content of models of choice.

The starting point of our empirical analysis is a pair of stochastic choice functions[2] (ρ, ρ^*). In such a pair, ρ^* is not directly observed by the analyst and represents the choices/behavior of a representative sample of the underlying population, whereas ρ represents directly observed (but mismeasured) choice frequencies. The aim of the analyst is to infer properties of ρ^* by using observed choice frequencies ρ. Without any information on how ρ relates to ρ^* (i.e. the magnitude of the distortion) not much can be said about underlying behavior.

[1] We elaborate on these issues in subsection 1.2 of the introduction.

[2] A stochastic choice function encodes choice frequencies of alternatives in menus from a collection of menus.

To study ρ^* using ρ the researcher needs information on how ρ relates to ρ^*. I.e., on measurement error.[3] Our main analysis will be focused on cases where ρ is, in various ways, an ordinal perturbation of ρ^*. This is a relatively mild assumption as it requires no knowledge about the distribution of the measurement error. Further, as we discuss later, it is satisfied by many examples of interest. Our methodology may also be used to study other non parametric assumptions on measurement error (other common restrictions on measurement error are discussed in section 9.1).

As an illustration of the type of situations we are interested in consider the following examples:

Example 1.1. An analyst conducts a survey and asks a representative sample of 120 individuals to pick their favorite brand among chocolates $X = \{x, y, z\}$. Out of the 120 individuals only 100 individuals report their favorite brand and the corresponding choice frequencies are $\{61, 30, 9\}$. Observed choices are encoded by an SCF ρ where $\rho(x, X) = \frac{61}{100}, \rho(y, X) = \frac{30}{100}, \rho(z, X) = \frac{9}{100}$, whereas ρ^* is unknown and depends on the individuals who did not participate in the survey. However, the researcher has information on the number of non participants (20) and using this she can infer that $\rho^*(x, X) > \rho^*(y, X) > \rho^*(z, X)$. This ranking agrees with the observed ranking $\rho(x, X) > \rho(y, X) > \rho(z, X)$. Thus, despite ρ^* not being directly observed by the researcher, she observes the ordinal ranking induced by ρ^*. $\triangleleft$

Example 1.2. A grocery store keeps track of the number of individuals purchasing items from a set $\{x, y, z, u\}$. Suppose that the grocery store registers the following purchases of items: $\{45, 79, 41, 0\}$. Without further information on the number of individuals visiting the store and their respective purchases it is difficult to calculate exact choice probabilities. Some individuals may for instance buy bundles of several items when visiting the store. Underlying choice frequencies ρ^* are unobservable due to the data collection method used by the store. However, an analyst may still distinguish chosen alternatives (such as x, y, z) from never chosen alternatives (such as u). $\triangleleft$

Common to both of these examples is that there is information on observed choices, encoded by ρ, and - crucially - on measurement error (i.e. how ρ relates to ρ^*). We will explore how the researcher can use information on observed choices and on measurement error to infer properties about underlying behavior. In particular, our interest is in understanding the impact of measurement error when testing models of choice. Formally, a model $\mathtt{M}$ is a subset of the set of all stochastic choice functions. These are the set of SCFs conceived as plausible by the researcher. An example is the random utility model, which is the collection

[3] The measurement error can now formally be defined as the difference $\rho - \rho^*$ between observed (but mismeasured) choices ρ and unobserved (but true) choices ρ^*

of stochastic choice functions that are represented by a probability distribution on preferences. The researcher would like to know/test whether underlying behavior ρ^* is explained by a certain model or not (if $\rho^* \in \mathtt{M}$ or not), but she only has access to mismeasured data ρ.

As discussed in the introduction, and as Gillen, Snowberg, and Yariv (2019) judiciously puts it, there are some "inferential dangers posed by measurement error". Observed data may be (in)consistent with a model, despite this not being the case for underlying behavior. A model is subject to *over rejection*, if measurement error makes it appear as if data violates the model ($\rho \notin \mathtt{M}$), despite underlying data being consistent with it ($\rho^* \in \mathtt{M}$). Similarly a model is subject to *under rejection*, if measurement error makes it appear as if data is described by the model ($\rho \in \mathtt{M}$), but underlying data is not ($\rho^* \notin \mathtt{M}$). These issues are closely related to type I/type II errors is statistical decision making. They are referred to as over/under rejection due to the researchers' tendency to incorrectly reject/not reject a model in such circumstances. Generally, the severeness of these issues is directly proportional to the severeness of measurement error.

Due to complications such as these, it would clearly be convenient if there were models that were not exposed to over/under rejection. The researcher could then, in principle, ignore measurement error and use convential tests (developed for traditional stochastic choice data) to study such models. Further, for models that are sensitive to measurement error, it would be desirable to devise tests that reduce issues of over/under rejection. The aim of this paper is to explore these matters further.

Our first contribution is to identify classes of *robust* models that, under suitable assumptions on measurement error, are insensitive to under/over rejection. As explained above, our main focus is on cases where measurement error is such that observed choices are ordinally related to underlying choices. We call such models *ordinal.* But, in principle, our methodology is applicable to study any conceivable restriction on measurement error (see section 2 for further details). Roughly speaking, an ordinal model is robust to ordinal perturbations of data. This means that if we take an SCF in the model and an increasing transformation of this SCF, then the new perturbed SCF has to be explained by the model as well. We will illustrate this concept further in section 1.2 below. There is an abundance of ordinal models: the simple scalability model (Tversky, 1972), the additive perturbed utility model (APU) (Fudenberg, Iijima, and Strzalecki, 2015), the single crossing random utility model (Apesteguia, Ballester, and Lu, 2017) and the dual random utility model (Manzini and Mariotti, 2018) are all such models. We also identify a large class of models, random sequential rules, that are ordinal. As a corollary it follows that stochastic versions of satisficing Simon (1955) and sequential rationalizability Manzini and Mariotti (2007) as well as the gradual pairwise comparison rule of Dutta (2020) are ordinal. Furthermore, some models of stochastic choice require substantially less information in describing their behavior. To accommodate this, we introduce several natural weakenings of ordinality. The weakest notion says that it is possible to fully describe the behavior of a model using deterministic data on choices (such as in example 1.2).

A second contribution of this paper is to discuss and develop methods to study the empirical content of models that are sensitive to measurement error. Again, our focus is on models that are sensitive to ordinal data, but our methodology applies to other restrictions on measurement error as well. Both the random utility model (RUM) and the Luce model are examples of non ordinal models. Measurement error is perhaps of greatest concern for such models, as they are exposed to issues of over/under rejection. To study the empirical content of non ordinal models, section 7 introduces the concepts of *ordinal closure* and *ordinal interior* of a model. Basing a test of a model on its ordinal closure/interior rules out issues of over/under rejection. It is defined as the smallest/largest (in a specific sense) ordinal model contained in/containing a given (non ordinal) model. The ordinal closure/interior of a model thus approximates a given (non ordinal) model with the ordinal model closest to it. Since ordinal models are immune to measurement error, the idea is that the approximated model will inherit some of these properties as well. Roughly speaking, the ordinal closure/interior of a model is the best description possible with mismeasured data. We will continue our discussion of these notions in section 7.

1.1. Outline

The remainder of the paper is devoted to studying versions of these notions and applying them to (well known) models of choice. Our main notions of robustness are introduced and discussed further in section 3 and section 4. We consider several notions of robustness each corresponding to different types of measurement error. Example 1.1 considers one particular type of distorted data. In a sense, the distortion of that data set is "small". We also consider more extreme cases of distortion where less information on choices is available. As a particular example we consider data sets with only deterministic information on choices available (see example 1.2). Arguably, the degree of distortion of any data set "naturally occurring" in practice is in between these two in strength. I.e. it has more information on choices than a deterministic correspondence but less so than the full ranking induced by an SCF. Hence, knowledge of the behavior of a model on these data sets is in many cases sufficient to infer their behavior on general distorted data sets.[4] For illustrative purposes we also consider an example of data set intermediate in strength of distortion.

[4]We elaborate further on this in the accompanying online appendix.

In section 5 and 6, we consider specialized models of choice and their ordinal properties. Section 5 is confined to the classical random utility model and various subclasses of it. The classical random utility model is in general not ordinal. However, two prominent subclasses of the random utility model, the single crossing RUM of Apesteguia et al. (2017) and the dual random utility model of Manzini and Mariotti (2018), are both ordinal. The dual random utility model is not only ordinal, but also deterministic. This means that it is possible to describe the behavior of this model using deterministic data on choices.

Section 5.4 identifies a large class of models, random sequential models (RSMs), that are shown to be ordinal. It follows as a corollary to this result that many subclasses of RSMs, including the GPCR of Dutta (2020) and stochastic versions of satisficing and sequential rationalizability (Manzini and Mariotti (2007)), are ordinal.

Section 6 considers ordinal properties of the Luce model and close relatives to it. More specifically, we consider the classical Luce model, the simple scalability model (SSM) Tversky (1972) and the APU model Fudenberg et al. (2015). Our main finding is that the SSM and the APU model are ordinal, whereas the Luce model is not.

Section 7 introduces the notion of ordinal closure/ interior of a model. We apply these notions in section 8 and study the ordinal closures/interiors of the random utility model and the Luce model. Our first result is a characterization of the ordinal closure of the random utility model. This problem is closely related to that of finding a representation of a subjective probability relation (see Kraft, Pratt, and Seidenberg (1959), Scott (1964), Fishburn (1969, 1986), Insua (1992), Alon and Lehrer (2014)). This result thus links two hitherto largely unconnected literatures: one on stochastic/discrete choice and the other on subjective/imprecise probability.

A second, rather striking result, is that the observable implications of the random utility model are vacuous for a large collection of distorted data sets. More precisely, the random utility model is not testable whenever the data set is such that it only contains information about the relative rankings of alternatives within menus. This result illustrates the rich behavior associated with a RUM and is in stark contrast to for example the Luce model, where the ranking of alternatives within menus needs to agree with a fixed utility function.

Section 9 concludes with a discussion of the framework adopted in the current paper and topics for future research. Subsection 9.1 discusses another type of restriction on measurement error, called moment restricted data. Roughly speaking, the researcher observes the moments of an SCF, but not the entire underlying SCF. This is much like average choice data in Ahn, Echenique, and Saito (2018), except that the researcher also has access to higher order moments (i.e. moments beyond the average). We discuss the relation to Ahn et al. (2018) and obtain a characterization of the Luce rule using moment restricted data. Subsection

9.2 discusses a comparative notion of robustness to address questions such as: when can a model of choice be considered more robust than another model of choice? Which model is the most robust among a collection of models? We also discuss how to apply our framework to classify properties of stochastic choice functions in section 9.3 and further outline how this could be useful in obtaining axiomatic characterizations of models of choice.

1.2. Illustration

The following subsection serves as a further illustration of the issues of over/under rejection in the context of example 1.1. A second purpose is to illustrate that ordinal models are immune to these issues.

The under rejection problem. In example 1.1, suppose that the researcher conjectures a model of choice that stipulates that chocolate x is preferred by more than 60% of the population, i.e. that $\rho^*(x,X) \geq 0.6$. Ignoring the issue of non-participants the researcher would be inclined to conclude that data supports her conjecture since, $\rho(x,X) = \frac{61}{100} > 0.6$. However, suppose that the researcher could elicit the choices of the non participants and it turns out that all of them have y as their favorite brand, then $\rho^*(x,X) = \frac{61}{120} < 0.6$, suggesting that her initial conjecture was false. In such a case, the researcher would end up failing to reject her conjectured model. This would thus be a manifestation of the under rejection problem discussed above.

The over rejection problem. Suppose instead that the researcher conjectures a model of choice that stipulates that choclate x is preferred by more than 65 % of the population, i.e. that $\rho^*(x,X) \geq 0.65$. Ignoring the issue of non-participants the researcher would be inclined to conclude that her initial conjecture was false since $\rho(x,X) = \frac{61}{100} < 0.65$. However, if all non participants have x as their favorite brand, then $\rho^*(x,X) = \frac{81}{120} > 0.65$, suggesting that her initial conjecture was indeed true. In such a case, the researcher would falsely reject her conjectured model, a manifestation of the over rejection problem discussed above.

Ordinal models. Now, consider a model which predicts that x is the most popular brand in X, i.e. that $\rho^*(x,X) \geq \rho^*(y,X)$ and $\rho^*(x,X) \geq \rho^*(z,X)$. This model does not suffer from the issues mentioned above. Irrespectively of the choices of the non participants one can show that x is the most popular choice in X. The stipulated model is not sensitive to ordinal perturbations of data. In other words, ρ^* in example 1.1 is an increasing transformation of ρ, and this implies - independently of the choices of the non participants - that ρ^* is consistent with the model if and only if ρ is. This is hence an example of an ordinal model.

1.3. Related literature

The literature dealing with measurement error in statistical and econometric analysis is vast. To the best of our knowledge this is the first paper to study the impact of measurement error in the context of stochastic choice analysis. There is a tangentially related literature that focuses on similar issues in revealed preference theory. Notable are the papers by Varian (1985), Echenique et al. (2011) and Aguiar and Kashaev (2018) . The framework in Aguiar and Kashaev (2018) is fully nonparametric and in contrast to earlier work they impose no distributional assumptions on measurement error. Our characterization of the random utility model (using data subject to measurement error) in section 7 is also nonparametric. Aguiar and Kashaev (2018) impose a centering (moment) condition on measurement error, whereas we assume that observed (but mismeasured) data is ordinally related to underlying data. In contrast to earlier work we also identify models of choice that are insensitive to measurement error, whereas prior work mainly focuses on providing methods to test models that are known to be affected by measurement error.

Our paper is also related to a literature investigating issues pertaining to limited information and data availability in stochastic/discrete choice. Notable in this regard are the papers by Fishburn (1978), Ok and Tserenjigmid (2020), De Clippel and Rozen (2020) and Ahn, Echenique, and Saito (2018). Ahn et al. (2018) studies an axiomatic characterization of the Luce rule assuming that only data on average choices from menus is available. De Clippel and Rozen (2020) consider issues of limited data availability in testing theories of bounded rationality. Fishburn (1978) and Ok and Tserenjigmid (2020) study deterministic implications of stochastic choice models. I.e. they associate to each stochastic choice function a deterministic choice correspondence and then study the relationship between the properties satisfied by the SCF and those satisfied by the correspondence. Their study mainly considers the rationality properties of the induced correspondences. Our focus is different from Ok and Tserenjigmid (2020). We are interested in the testable implications of stochastic choice models assuming that deterministic data is observed.

Related are also the papers Apesteguia and Ballester (2015) and Apesteguia and Ballester (2020). Apesteguia and Ballester (2015) studies a measure of rationality and welfare. In particular, they study and introduce the SWAPs index, which measures how closely a data set satisfies the utility maximization hypothesis. Apesteguia and Ballester (2020) proposes a goodness of fit measure for stochastic choice models, which is the largest fraction of data consistent with a model. Similarly to us they acknowledge that observed data may differ from true underlying data and they separate observed data into two distinct components: one structured portion that represents underlying behavior and one unstructured portion which represents residual behavior. The residual part in their separation exercise may be interpreted as measurement error. They are interested in the largest fraction of data explained by a certain model, i.e. minimizing the residual behavior, whereas we take residual behavior as given and ask whether it impacts inferences from choice data.

2. Mismeasured data

Before introducing the basic primitive of our analysis, we need some preliminary notation and definitions. Let X be a finite set of alternatives. A nonempty subset $A \subseteq X$ is called a *menu*. We denote by $\mathscr{A}$ a collection of menus, i.e. $\mathscr{A} \subseteq 2^X \setminus \emptyset$. A *stochastic choice function* (SCF) is a function $\rho : X \times \mathscr{A} \mapsto [0,1]$ such that we have $\sum_{a \in A} \rho(a, A) = 1$ for all menus $A \in \mathscr{A}$. Let $\texttt{SCF}$ denote the set of all stochastic choice functions on $X \times \mathscr{A}$.

To study the impact of measurement error, this paper takes as primitive a pair of stochastic choice functions (ρ, ρ^*). The stochastic choice function ρ^* encodes data that is representative of the underlying population. In contrast, ρ reflects mismeasured/distorted data that is not necessarily representative. We assume that the researcher/analyst/outisde observer has access to or "observes" ρ but not ρ^*. [5] The goal of the analyst is then to infer properties of ρ^* by using mismeasured data ρ.

Definition 2.1. A *distorted data set* is a pair (ρ, ρ^*) of stochastic choice functions. $\triangleleft$

The measurement error of a distorted data set (ρ, ρ^*) can now formally be defined as $\varepsilon = \rho - \rho^*$, i.e. the difference between observed and underlying choice frequencies. Knowledge of underlying choice frequencies ρ^* is thus equivalent to knowledge of measurement error ε. Any difference between ρ and ρ^* is by definition due to measurement error ε. To infer properties about underlying behavior ρ^* using ρ, the researcher needs information on measurement errror, or equivalently, on the relation between ρ and ρ^*. Below, we will impose various restrictions on the relation between ρ and ρ^*. Formally, a restriction on measurement error is a subset $\mathscr{R} \subseteq \texttt{SCF} \times \texttt{SCF}$.

Definition 2.2. A distorted data set (ρ, ρ^*) such that $(\rho, \rho^*) \in \mathscr{R}$ is called an $\mathscr{R}$ *data set.* $\triangleleft$

$\mathscr{R}$ can be thought of as representing the type and extent of measurement error that the researcher believes that data is subject to. These restrictions are specific to the particular application at hand and reflect the researchers knowledge about measurement error. If $\mathscr{R} = \texttt{SCF} \times \texttt{SCF}$ then measurement error is unrestricted. The researcher believes that any underlying behavior ρ^* is consistent with observed ρ since $(\rho, \rho^*) \in \mathscr{R}$ for all ρ^*. Thus, given observed ρ nothing can be said about underlying (and unobserved) behavior ρ^*.

[5]In reality, the researcher observes a sample drawn from ρ, which introduces sampling error. In the sequel, we assume a sufficiently large sample to rule out issues of sampling error.

It is clear that the researcher needs to impose restrictions on measurement error, and hence $\mathscr{R}$, to make inferences about underlying behavior. The type of restriction that the researcher imposes depends on the researchers' knowledge of measurement error specific to the studied application. We will next discuss some important examples of such restrictions and situations/applications where these restrictions might be applicable. As a first type of restriction, we consider data sets where measurement error is such that ρ is an increasing/ordinal transformation of ρ^*.

Definition 2.3. A distorted data set (ρ, ρ^*) such that

$$\rho(a, A) \geq \rho(b, B) \Leftrightarrow \rho^*(a, A) \geq \rho^*(b, B)$$

for all $a, b \in X$ and $A \in \mathscr{A}$ is called an *ordinal data set.* ◁

An ordinal data set corresponds to a case where disrtortion of data is "small" in that the complete ordinal ranking induced by some underlying SCF ρ^* is observed. Studying such data could be useful in situations, like example 1.1 where an analyst has some degree of confidence in the ordinal ranking of alternatives, but is perhaps not convinced that observed choice frequencies, as encoded by ρ, reflect exact choice frequencies of the underlying population ρ^*. The following result, which is immediate (the proof is supplied in the appendix), shows that any "generic" pair of SCFs (ρ, ρ^*) is an ordinal data set provided that ρ is measured with sufficiently small error ε. This hence formalizes the intuition that ordinal data sets correspond to situations where measurement error is small.

Proposition 2.1. *For almost all $\rho^* \in \texttt{SCF}$ (in the sense of Lebesgue, see proof for details) there is an $\varepsilon > 0$ such that if $\rho \in \texttt{SCF}$ with $||\rho - \rho^*|| < \varepsilon$ then (ρ, ρ^*) is an ordinal data set.*

Our next example, imposes less stringent assumptions on measurement error than ordinal data sets do.

Definition 2.4. A distorted data set (ρ, ρ^*) such that:

$$\rho(a, A) \geq \rho(b, A) \Leftrightarrow \rho^*(a, A) \geq \rho^*(b, A)$$

for all $a, b \in X$ and $A \in \mathscr{A}$ is called an *within menu data set.* ◁

Intuitively, the observed SCF ρ captures within menu variation of underlying choices ρ^*. Within menu data could be useful in situations where a researcher has reliable information on the relative ranking of choice frequencies within menus, but where choice frequencies across menus are difficult to compare. As we will see later, there are models of stochastic choice whose empirical content could be fully described by using the limited information in an within menu data set. The next example illustrates a scenario where data on choices is naturally encoded by such a data set.

Example 2.1. A researcher thinks that choices are governed by a stochastic choice function ρ^*. She conducts an experiment to test her hypothesis. However, she also believes that with small but positive probability some subjects in her experiment will choose a random option not conforming to the ρ^* process. This could be either because these subjects miss-interpret the choice task, or because they lack the incentives to reveal their true choices. The researcher hence observes an SCF ρ where:

$$\rho(a, A) = (1 - \alpha(A))\rho^*(a, A) + \alpha(A)\frac{1}{|A|}$$

for all $a \in A$ and $A \in \mathscr{A}$, where $\alpha : \mathscr{A} \to [0, 1]$ is an error function. Note that the error term is menu dependent to reflect that subjects' incentives to reveal their true choices differ across menus. Subjects may for instance be less inclined to reveal their true choices when menus are large as opposed to small (as the cognitive costs to do so are higher). Then (ρ, ρ^*) is an example of an within menu data set.[6] $\triangleleft$

The next example reflects a case of quite extreme measurement error. The researcher is only able to distinguish between alternatives based on whether they are chosen or not in a menu.

Definition 2.5. A data set (ρ, ρ^*) such that:

$$\rho(a, A) > 0 \iff \rho^*(a, A) > 0$$

for all $a \in X$ and $A \in \mathscr{A}$ is called a *deterministic data set.* $\triangleleft$

Example 1.2 of the introduction is an example of a deterministic data set.

3. Robust stochastic choice models

We next turn to the issue of testing models of choice in the presence of measurement error. Formally, a *model* is a subset $\mathtt{M} \subseteq \mathtt{SCF}$. I.e. $\mathtt{M}$ is a collection of stochastic choice functions entertained by the researcher. A main concern when testing models of choice using mismeasured data is the problem of over/under rejection. A model is subject to over/under rejection

[6] To see this, note that $\rho(a, A) - \rho(b, A) = (1 - \alpha(A))[\rho^*(a, A) - \rho^*(b, A)]$ for all $a, b \in A$.

if observed data appears (in)consistent with the model despite this not necessarily being the case for underlying behavior. An observed (non-) violation of the model could then be due to measurement error rather than underlying behavior. The aim of this section is to formally introduce the notion of a robust model. Roughly, speaking these are models where distorted choices ρ convey as much about underlying behavior as ρ^* does.

3.1. Over and under rejection problem

Before introducing our main notion of robustness, it is instructive to formally investigate what it means for a model to be subject to over/under rejection. Recall that $\mathscr{R}$ is a subset of the pair of all SCFs and reflects the researcher's knowledge about measurement error. A model is subject to over rejection if there is some conceivable measurement error (consistent with $\mathscr{R}$) such that observed (but mismeasured) choices are inconsistent with the model, whereas underlying choices are not.

Definition 3.1. A model $\mathtt{M}$ is subject to $\mathscr{R}$ over rejection if there is a distorted data set $(\rho, \rho^*) \in \mathscr{R}$ such that

$$\rho \notin \mathtt{M} \quad \text{and} \quad \rho^* \in \mathtt{M}.$$

◁

We refer to this issue as over rejection since if the researcher rejects model $\mathtt{M}$ whenever she observes $\rho \notin \mathtt{M}$, she would sometimes incorrectly reject the true underlying model. In contrast, a model is subject to under rejection, if there is some (conceivable) measurement error such that observed choices are consistent with the model, whereas underlying choices are not.

Definition 3.2. A model $\mathtt{M}$ is subject to $\mathscr{R}$ under rejection if there is a distorted data set $(\rho, \rho^*) \in \mathscr{R}$ such that

$$\rho \in \mathtt{M} \quad \text{and} \quad \rho^* \notin \mathtt{M}.$$

◁

The idea behind the term under rejection, is that if the researcher employs the decision rule: "reject whenever $\rho \notin \mathtt{M}$" she would sometimes fail to reject her hypothesised model $\mathtt{M}$.

3.2. Definition of robustness to measurement error

The discussion above leads us to our main definition of a model being robust to (certain types of) measurement error. Roughly speaking, such a model is not exposed to issues of over/under rejection. Thus, using observed choices ρ, a researcher would be able to tell whether $\rho^* \in \mathtt{M}$ or not.

Definition 3.3. A model $\mathtt{M}$ is $\mathscr{R}$ robust to distorted data if for all data sets $(\rho, \rho^*) \in \mathscr{R}$ we have:

$$\rho \in \mathtt{M} \quad \Leftrightarrow \quad \rho^* \in \mathtt{M}.$$

◁

In other words, a model is robust to $\mathscr{R}$ distorted data if and only if it is not subject to $\mathscr{R}$ over/under rejection. As mentioned before, the researcher needs knowledge on the relation between ρ and ρ^* as encoded by $\mathscr{R}$ in order to deduce anything about underlying behavior ρ^* using observed ρ. The above definition sheds some light on why this is the case. If no restrictions are imposed ($\mathscr{R} = \mathtt{SCF} \times \mathtt{SCF}$), then all models $\mathtt{M}$ (except $\mathtt{SCF}$) will be subject to over/under rejection and it will in general be difficult to use ρ to test whether underlying behavior conforms to a particular model. [7]

4. Ordinal models

We will next focus our discussion around the examples in section 2. We will thus study models that are robust to ordinal/within menu ordinal and deterministic perturbations of data. Arguably, the degree of measurement error of any data set naturally occurring in practice is in between that of a deterministic and ordinal data set.

Definition 4.1. A model $\mathtt{M}$ is ordinal/within menu ordinal/deterministic if for all ordinal/within menu ordinal/deterministic data sets (ρ, ρ^*) we have:

$$\rho \in \mathtt{M} \quad \Leftrightarrow \quad \rho^* \in \mathtt{M}.$$

◁

It is straightforward to verify that any within menu ordinal model is ordinal. Similarly, it follows that any deterministic model is within menu ordinal. The deterministic assumption is in this sense the weakest (fewer models satisfy it), whereas the ordinal assumption is the strongest. We discuss these notions in some greater detail below.

[7] To see this, let $\mathtt{M} \subset \mathtt{SCF}$ and pick $\rho^* \in \mathtt{SCF} \setminus \mathtt{M}$. Then any observed SCF $\rho \in \mathtt{M}$ is consistent with ρ^*, since $(\rho, \rho^*) \in \mathtt{SCF} \times \mathtt{SCF}$. Similarly, if the researcher observes $\rho \notin \mathtt{M}$, then any $\rho^* \in \mathtt{M}$ is consistent with ρ.

4.1. Characterization: ordinal models

A model is ordinal if the information provided by ρ in an ordinal data set (ρ, ρ^*) is sufficient to determine whether $\rho^* \in \mathtt{M}$ or not. Formally, we restrict $\mathscr{R}$ in definition 4.1 to the collection of ordinal data sets. By proposition 2.1 we may view an ordinal model as a model that is robust to small amounts of measurement error. The following proposition characterizes ordinal models, and shows that ordinal models only depend on the ordinal information of choice probabilities.

Proposition 4.1. *The following statements are equivalent:*

1. $\mathtt{M}$ *is an ordinal model.*
2. *If* $\rho' \in \mathtt{M}$ *and* $\rho \in \mathtt{SCF}$ *and*

$$\rho(a, A) > \rho(b, B) \quad \Leftrightarrow \quad \rho'(a, A) > \rho'(b, B) \tag{1}$$

   for all $(a, A), (b, B) \in X \times \mathscr{A}$ *implies that* $\rho \in \mathtt{M}$.

3. *For all* $\rho \in \mathtt{SCF}$ *with* $\rho = v \circ \rho'$ *for some* $\rho' \in \mathtt{M}$ *and strictly increasing (continuous) function* $v : [0, 1] \to [0, 1]$ *with* $v(0) = 0$ *we have* $\rho \in \mathtt{M}$.

The proof of proposition 4.1 is straightforward and follows by unwinding the definition of an ordinal model. We will apply this notion in section 5 and section 6 to show that several well-known models are indeed ordinal.

4.2. Characterization: within menu ordinal models

As explained above, a model is ordinal if it is possible to determine whether an SCF belongs to the model or not by using *the full* ordinal ranking of choice probabilities. If a model is within menu ordinal then it is possible to determine model membership of an SCF by solely using rankings of items within menus. This could be useful in situations such as in example 2.1. The following proposition characterizes within menu ordinal models.

Proposition 4.2. *The following statements are equivalent:*

1. $\mathtt{M}$ *is an within menu ordinal model.*
2. *If* $\rho' \in \mathtt{M}$ *and* $\rho \in \mathtt{SCF}$ *and*

$$\rho(a, A) > \rho(b, A) \quad \Leftrightarrow \quad \rho'(a, A) > \rho'(b, A)$$

   for all $(a, A), (b, A) \in X \times \mathscr{A}$ *implies that* $\rho \in \mathtt{M}$.

3. *If* $\rho' \in \mathtt{M}$ *and* $\rho \in \mathtt{SCF}$*, and if there for each* $A \in \mathscr{A}$ *is a strictly increasing (continuous) function* $v_A : [0,1] \to [0,1]$ *with* $v_A(0) = 0$ *such that* $\rho(\cdot, A) = v_A \circ \rho'(\cdot, A)$ *then* $\rho \in \mathtt{M}$.

4.3. Characterization: deterministic models

The strongest notion (logically) of ordinality that we will consider essentially says that the properties satisfied by a model are deterministic. Differently put, such a model is (very) insensitive to the correct specification of choice probabilities. The following proposition characterizes deterministic models.

Proposition 4.3. *The following statements are equivalent:*

1. $\mathtt{M}$ *is a deterministic model.*

2. *If* $\rho' \in \mathtt{M}$ *and* $\rho \in \mathtt{SCF}$ *and*
$$\rho(a, A) > 0 \quad \Leftrightarrow \quad \rho'(a, A) > 0$$
for all $(a, A) \in X \times \mathscr{A}$ *implies that* $\rho \in \mathtt{M}$.

3. *If* $\rho \in \mathtt{SCF}$ *and* $\rho' \in \mathtt{M}$*, and if there for each* $A \in \mathscr{A}$ *is a (continuous) function* $v_A : [0,1] \to [0,1]$ *that is strictly positive on* $(0,1]$ *and has* $v_A(0) = 0$ *such that* $\rho(\cdot, A) = v_A \circ \rho'(\cdot, A)$ *then* $\rho \in \mathtt{M}$.

Thus, if a model is deterministic and ρ is an SCF of a certain model, then every SCF with support equal to ρ belongs to the model. Deterministic models represent a case of extreme impreciseness, in that it is possible to pin down model membership although data is (very) imprecise. A deterministic *model* should not be confused with a deterministic stochastic choice function. The former refers to a property satisfied by a class of SCFs, whereas the latter refers to a property satisfied by a single SCF. For example, the dual Random Utility Model of Manzini and Mariotti (2018) is a deterministic model. But it is clear that not every SCF with a dRUM representation is a deterministic SCF.

5. Ordinality of random utility models

We first investigate ordinal properties of the random utility model[8] (and close relatives to it). As we will see, a notable property of the random utility model is that it is not ordinal. There are however interesting subclasses of RUMs that are ordinal.

[8] The random utility model has received wide attention in the literature on stochastic choice. Notable contributions are Gul and Pesendorfer (2006), Lu (2016), Apesteguia, Ballester, and Lu (2017), Manzini and Mariotti (2018), Frick, Iijima, and Strzalecki (2019) and Piermont and Teper (2020).

5.1. Classical random utility model

Let $\mathscr{P}$ denote the set of strict linear orders on X. For all $A \subseteq X$ and $a \in A$ denote by $\mathscr{P}(a, A) = \{P \in \mathscr{P} : \max(P, A) = a\}$. A stochastic choice function ρ has a *random utility representation* if there is a measure μ on $\mathscr{P}$ such that $\rho(a, A) = \mu(\mathscr{P}(a, A))$ for all $a \in A$ and $A \in \mathscr{A}$. The *random utility model* is the class of SCFs with a random utility representation and is denoted by `RUM`.

Proposition 5.1. *The random utility model is not an ordinal model.*

The proposition above follows by the next example.[9]

Example 5.1. Let $X = \{a, b, c, d\}$. First note that any RUM must satisfy the following inequality:

$$\rho(a, ab) - \rho(a, abc) \geq \rho(a, abd) - \rho(a, abcd) \tag{2}$$

We construct a RUM ρ and a stochastic choice function ρ' such that (ρ, ρ') is an ordinal data set and such that ρ' violates the previous inequality. Let linear orders P_1, P_2, P_3 be defined as in the table below.

P_1	P_2	P_3
c	c	d
a	d	a
b	b	b
d	a	c

Let p_1, p_2, p_3 be numbers such that $0 < p_1 < p_2 < p_3 < 1$ and $p_1 + p_2 + p_3 = 1$. Define a probability measure μ on $\mathscr{P}$ by $\mu(P_i) = p_i$ for all $i \in \{1, 2, 3\}$. Now we note that:

$$\rho(a, ab) = p_1 + p_3 \tag{3}$$

$$\rho(a, abc) = p_3 \tag{4}$$

$$\rho(a, abd) = p_1 \tag{5}$$

$$\rho(a, abcd) = 0 \tag{6}$$

Thus the inequality in equation (2) is binding. Further, note that there is only one $(b, B) \in X \times \mathscr{A}$ such that $\rho(b, B) = p_1 + p_3$ and this is $(b, B) = (a, ab)$ and similarly there is only one (b, B) such that $\rho(b, B) = p_2$ and this is (b, ab). Let

$$\varepsilon_1 = \min_{(b,B) \in X \times \mathscr{A} \setminus (a,ab)} |\rho(a, ab) - \rho(b, B)|,$$

[9] Example 5.1 requires $|X| > 3$. The random utility model is characterized by monotonicity when $|X| \leq 3$. The set of SCFs satisfying monotonicity is an ordinal model.

$$\varepsilon_2 = \min_{(b,B)\in X\times\mathscr{A}\setminus(b,ab)} |\rho(b,ab) - \rho(b,B)|$$

and let $0 < \varepsilon < \frac{\min\{\varepsilon_1,\varepsilon_2\}}{2}$. Define a new stochastic choice function ρ' by $\rho'(a,A) = \rho(a,A)$ for all $(a,A) \in X\times\mathscr{A}\setminus\{(a,ab),(b,ab)\}$ and $\rho'(a,ab) = \rho(a,ab)-\varepsilon$ and $\rho'(b,ab) = \rho(b,ab)+\varepsilon$. It is now straightforward to check that $\rho(a,A) > \rho(b,B) \Leftrightarrow \rho'(a,A) > \rho'(b,B)$ for all $(a,A),(b,B) \in X\times\mathscr{A}$. But note that

$$\rho'(a,ab) - \rho'(a,abc) = p_1 + p_3 - \varepsilon - p_3 = p_1 - \varepsilon < p_1 = \rho'(a,abd) - \rho'(a,abcd).$$

It follows that ρ' fails to satisfy equation (2) and hence cannot have a random utility representation. ◃

The above example demonstrates that the random utility model is in general sensitive to measurement error. Small perturbations of choice probabilities gives data that is inconsistent with the random utility model. However, the menu collection used to construct example 5.1 is quite specific and contains all nonempty subsets of the grand set X of alternatives. In empirical/experimental work on choice it is common to encounter more restricted menu collections. This begs the question whether there are specific choices of $\mathscr{A}$ such that the random utility model is ordinal when using $\mathscr{A}$ as menu collection. The following proposition shows that the random utility model is ordinal when restricted to the "leave-one-out" menu variation used by Abaluck and Adams (2020) and further analysed in Dardanoni et al. (2020). More, precisely the "leave-one-out" menu variation $\mathscr{A}$ consists of the grand set of alternatives X and the collection of subsets of X obtained by removing one element from X (all sets of the form $X\setminus a$ for some $a \in X$).

Proposition 5.2. *Let $\mathscr{A}$ be the "leave-one-out" menu variation. Then the random utility model is ordinal.*

The proof of proposition 5.2 is in subsection A.2 of the appendix.

5.2. Single crossing random utility models

In a recent, very interesting paper, Apesteguia et al. (2017) introduce the single crossing random utility model (SCRUM). Formally, let $\succ$ be a linear order. A stochastic choice function ρ is a single crossing random utility model, if ρ is a RUM with measure μ on $\mathscr{P}$ and there is an ordering $\{P_1,...,P_T\}$ of the support of μ such that if $x \succ y$ and $s > t$ then xP_ty implies xP_sy. I.e. P_s is more "aligned" with $\succ$ than P_t. Denote by SCRUM the class of SCRUM stochastic choice functions, i.e. the collection of SCFs that are SCRUM w.r.t. some linear order $\succ$. Clearly SCRUM $\subset$ RUM. Since RUM is not an ordinal model, one may expect the same to be true for SCRUM. This is however not the case. As the following proposition shows SCRUM is an ordinal model:

Proposition 5.3. *Let $\mathscr{A}$ contain all binary and ternary subsets then* SCRUM *is an ordinal model.*

Proof. Let $\rho \in$ SCRUM and let $\rho' \in$ SCF with $\rho(a,A) \geq \rho(a,B)$ if and only if $\rho'(a,A) \geq \rho'(a,B)$ for all $a \in X$ and $A,B \in \mathscr{A}$. Apesteguia et al. (2017) characterize SCRUMs using monotonicity/regularity and a property called *centrality*: If $x \succ y \succ z$ and $\rho(y,\{x,y,z\}) > 0$, then $\rho(x,\{x,y\}) = \rho(x,\{x,y,z\})$ and $\rho(z,\{y,z\}) = \rho(z,\{x,y,z\})$. Hence it suffices to show that ρ' satisfies monotonicity and centrality. It is immediate that ρ' satisfies monotonicity. Let $x \succ y \succ z$ and $\rho'(y,\{x,y,z\}) > 0$. Then, since $\rho'(y,\{x,z\}) = 0$ it follows that $\rho'(y,\{x,y,z\}) > \rho'(y,\{x,z\})$. Hence $\rho(y,\{x,y,z\}) > \rho(y,\{x,z\}) = 0$. Since ρ satisfies centrality we have $\rho(x,\{x,y\}) = \rho(x,\{x,y,z\})$ and $\rho(z,\{y,z\}) = \rho(z,\{x,y,z\})$ and using that for all $a \in X$ and $A,B \in \mathscr{A}$: $\rho(a,A) \geq \rho(a,B)$ if and only if $\rho'(a,A) \geq \rho'(a,B)$, it immediately follows that $\rho'(x,\{x,y\}) = \rho'(x,\{x,y,z\})$ and $\rho'(z,\{y,z\}) = \rho'(z,\{x,y,z\})$. □

Thus, in contrast to the random utility model, SCRUM coops quite well with ordinal data. However, SCRUM is not an within menu ordinal model (and hence not a deterministic model).

Proposition 5.4. SCRUM *is not an within menu ordinal model.*

Proposition 5.4 follows by example 5.2 below.

Example 5.2. To see that SCRUM is not within menu ordinal. Let $a \succ b \succ c$. Define P_1 by cP_1bP_1a and P_2 by bP_2aP_2c. Then P_1 and P_2 are single crossing w.r.t. $\succ$. Define a SCRUM ρ by $\rho(a,A) = \frac{1}{3}\mathbf{1}\{a = \max(A,P_1)\} + \frac{2}{3}\mathbf{1}\{a = \max(A,P_2)\}$. Note that $\rho(c,\{a,b,c\}) = \rho(c,\{b,c\}) = \frac{1}{3}$ and $\rho(b,\{a,b,c\}) = \rho(b,\{b,c\}) = \frac{2}{3}$. Define an SCF ρ' by $\rho'(b,B) = \rho(b,B)$ for all $b \in B \subseteq \{a,b,c\}$ with $B \neq \{b,c\}$. Set $1-\rho'(c,bc) = \rho'(b,bc) = \frac{3}{4}$. It is then clear that (ρ,ρ') is an within menu data set. However $\rho'(c,bc) = \frac{1}{4} < \frac{1}{3} = \rho'(c,abc)$, so ρ' is not monotone. Hence $\rho' \notin$ SCRUM. ◁

5.3. Dual random utility model

The dual Random Utility Model (dRUM) is a variation of the classical random utility model, where utility depends on at most two states and where choice probabilities are allowed to be menu dependent. The class of dRUMs was introduced by Manzini and Mariotti (2018).[10] Formally, a stochastic choice function ρ is a dual Random Utility Model (dRUM) if there is a function $\alpha : 2^X \setminus \emptyset \to (0,1)$ and there are linear orders P_1 and P_2 such that for all $a \in A$ and $A \subseteq X$:

$$\rho(a,A) = \alpha(A)\mathbf{1}\{a = \max(A,P_1)\} + (1-\alpha(A))\,\mathbf{1}\{a = \max(A,P_2)\}.$$

[10] see also Manzini, Mariotti, and Petri (2019).

If there is $\alpha \in (0,1)$ with $\alpha(A) = \alpha$ for all $A \in \mathscr{A}$ then ρ is an independent dRUM (idRUM). Every idRUM is a RUM, but the class of dRUMs is not necessarily a subset of the class of RUMs, due to the menu dependent choice probabilities $\alpha(A)$. The following proposition shows that only deterministic (hence ordinal) information is needed in determining whether an SCF is a dRUM or not.

Proposition 5.5. *The dual random utility model is a deterministic model.*

Proof. Let ρ be an SCF s.t. there is a function $\alpha : 2^X \setminus \emptyset \to (0,1)$ and there are linear orders P_1 and P_2 such that for all $a \in A$ and $A \subseteq X$: $\rho(a, A) = \alpha(A)\mathbf{1}\{a = \max(A, P_1)\} + (1 - \alpha(A))\,\mathbf{1}\{a = \max(A, P_2)\}$. Let $\rho' \in \mathtt{SCF}$ be such that (ρ, ρ') is a deterministic data set. Define $\beta(A) = \rho'(\max(P_1, A), A)$ for all $A \in \mathscr{A}$. Then $\rho'(a, A) = \beta(A)\mathbf{1}\{a = \max(A, P_1)\} + \big(1 - \beta(A)\big)\,\mathbf{1}\{a = \max(A, P_2)\}$. Thus $\rho' \in \mathtt{dRUM}$, as we wanted to show. □

Since $\mathtt{dRUM}$ is a deterministic model, it follows as a corollary that $\mathtt{dRUM}$ is an within menu ordinal model as well. In contrast to the situation for dRUMs, much more information about choice probabilities is needed in order to pin down the behavior of idRUMs. As the following proposition shows idRUMs are ordinal, but not necessarily within menu ordinal (and hence not deterministic). The intuition behind this result is quite simple as idRUMs put harsh restrictions on choice probabilities in that they are required to be constant across menus.

Proposition 5.6. *The class of idRUMs is an ordinal model, but not an within menu ordinal model.*

Proof. Let ρ be an SCF s.t. there is an $\alpha \in (0,1)$ and there are linear orders P_1 and P_2 such that for all $a \in A$ and $A \subseteq X$: $\rho(a, A) = \alpha\mathbf{1}\{a = \max(A, P_1)\} + (1 - \alpha)\,\mathbf{1}\{a = \max(A, P_2)\}$. Let $\rho' \in \mathtt{SCF}$ be such that (ρ, ρ') is an ordinal data set. Then there is a strictly increasing function $v : [0,1] \to [0,1]$ s.t. $v(0) = 0$ and $\rho' = v \circ \rho$. By checking a few cases (depending on whether $a = \max(A, P_1)$ or $a = \max(A, P_2)$ or both) it is readily verified that $\rho'(a, A) = v(\alpha)\mathbf{1}\{a = \max(A, P_1)\} + (1 - v(\alpha))\,\mathbf{1}\{a = \max(A, P_2)\}$ for all $a \in A$ and $A \subseteq X$. The proof that $\mathtt{idRUM}$ is not within menu ordinal follows by example 5.2. □

5.4. Random sequential models

This section identifies a large class of within menu ordinal models. The models studied in this section all have a sequential structure. Faced with a menu A, the decision maker narrows down choices sequentially using a sequence of choice rules/correspondences $C_1, ..., C_n$ such that for each $A \in \mathscr{A}$: $C_1(A) \supseteq ... \supseteq C_n(A)$. The set $C_i(A)$ is interpreted as the set of alternatives that remains after step i. We assume that $C_1(A) = A$ for all $A \in \mathscr{A}$, i.e. the decision maker starts out with all alternatives in each menu A. A stopping function is a function $\pi : \{C_1,...,C_n\} \times \mathscr{A} \to [0,1]$ such that $\sum_{i=1}^{n} \pi(C_i, A) = 1$. Thus, the stopping function gives the probability by which the decision maker stops at a particular step of her narrowing down process.

Definition 5.1. A stochastic choice function ρ is a *Random sequential rule* (RSR) if there is a finite set $I = \{1,...,n\}$, choice correspondences $(C_i)_{i\in I}$ such that for each $A \in \mathscr{A}$: $C_1(A) \supseteq ... \supseteq C_n(A)$, and a stopping function π such that

$$\rho(a, A) = \sum_{i\in I: a\in C_i(A)} \frac{\pi(C_i, A)}{|C_i(A)|} \tag{7}$$

for all $a \in A$ and $A \in \mathscr{A}$. ◁

Let $\mathbb{C}$ denote the set of all choice correspondences. Let $\mathscr{S} \subseteq 2^{\mathbb{C}}$ denote the set of all $\{C_1,...,C_n\} \in 2^{\mathbb{C}}$ such that $\{C_1,...,C_n\}$ is sequential (i.e. $C_1(A) \supseteq ... \supseteq C_n(A)$). A subset $\mathscr{H} \subseteq \mathscr{S}$ then corresponds to various restrictions put on the sequences $\{C_1,...C_n\}$. The model composed of all SCFs with an RSM representation using a subset of correspondences $\{C_1,...,C_n\} \in \mathscr{H}$ is denoted by RSM($\mathscr{H}$). By varying $\mathscr{H}$ we obtain various submodels of RSM.

Example 5.3. (Gradual pairwise comparison rule (GPCR)). Let $\mathscr{H} \subseteq 2^{\mathbb{C}}$ be such that $\{C_1,...,C_n\} \in \mathscr{H}$ if and only if there is preference P and an ordered partition $\mathscr{X} = \{X_1,...,X_{n-1}\}$ of P such that $C_1(A) = A$ and for all $1 < k \leq I$: $C_k(A) = \{a \in C_{k-1}(A) : \forall b \in C_{k-1}(A), (b,a) \notin X_{k-1}\}$. This is a version of the GPCR considered in Dutta (2020). ◁

Example 5.4. (Sequentially rationalizable choice) Let $\mathscr{H} \subseteq 2^{\mathbb{C}}$ be such that $\{C_1,...,C_n\} \in \mathscr{H}$ if and only if there is a sequence of asymmetric relations $P_1,...,P_n$ such that $C_0(A) = A$ and $C_k(A) = \{x \in C_{k-1}(A) : yP_kx \text{ for no } y \in C_{k-1}(A)\}$. Clearly, $\mathscr{H} \subseteq \mathscr{S}$. This is a stochastic version of sequentially rationlizable choice in Manzini and Mariotti (2007). ◁

Example 5.5. (Stochastic satisficing Simon (1955)) Let $\mathscr{H} \subseteq 2^{\mathbb{C}}$ be such that $\{C_1,...,C_n\} \in \mathscr{H}$ if and only if there is a utility function $u : X \to \mathbb{R}$ and parameters $\delta_1 < ... < \delta_n$ such that $C_i(A) = \{a \in A : u(a) \geq \delta_i\}$. Such a collection $\{C_1,...,C_n\}$ is clearly sequential, so $\mathscr{H} \subseteq \mathscr{S}$. ◁

The following proposition shows that a large class of models are within menu ordinal. Indeed, any choice of $\mathscr{H}$ gives an within menu ordinal model. As a corollary it follows that the models in example 5.4 - 5.3 are all within menu ordinal. The proof of proposition 5.7 is delegated to subsection A.3 of the appendix.

Proposition 5.7. *Let $\mathscr{H} \subseteq \mathscr{S}$. Then* RSM($\mathscr{H}$) *is an within menu ordinal model (and hence an ordinal model).*

6. Ordinality of the Luce model and related models

In this section we will look at ordinal properties of the Luce model and closely related models, such as the additive perturbed utility model (Fudenberg et al., 2015) and simple scalability (Tversky, 1972).

6.1. Luce model

A cornerstone model of stochastic choice is the eponymous Luce model (Luce, 1959). It postulates a utility function, where the choice probability of each alternative is proportional to its utility. Formally, a positive stochastic choice function[11] ρ has a Luce representation if there is a non-negative utility function $u : X \to \mathbb{R}$ such that:

$$\rho(a, A) = \frac{u(a)}{\sum_{b \in A} u(b)}$$

for all $a \in A$ and $A \in \mathscr{A}$. It is fairly well known that the Luce model is characterized by *Independence of Irrelevant Alternatives*, stating that: $\frac{\rho(a,A)}{\rho(b,A)} = \frac{\rho(a,B)}{\rho(b,B)}$ for all $a, b \in A \cap B$. By this property it is immediate that the Luce model is not ordinal.

Proposition 6.1. *The Luce model is not an ordinal model.*

Proof. We exhibit SCFs ρ and ρ' such that (ρ, ρ') is an ordinal data set, $\rho \in \texttt{Luce}$ and $\rho' \in \texttt{SCF} \setminus \texttt{Luce}$. Let $X = \{a_1, a_2, a_3\}$ and $\mathscr{A} = \{\{a_1, a_2, a_3\}, \{a_1, a_2\}\}$ and let ρ be a Luce model with utility function u defined by $u(a_1) = 1$ and $u(a_{i+1}) = u(a_i) + \frac{1}{10}$ for all $i \in \{1, 2\}$. Then $\rho(a_1, \{a_1, a_2\}) = \frac{10}{21}$. Define a stochastic choice function ρ' such that $\rho'(a_1, \{a_1, a_2\}) = \frac{100}{201}$ and $\rho'(a_i, \{a_1, a_2, a_3\}) = \rho(a_i, \{a_1, a_2, a_3\})$ for all $i \in \{1, 2, 3\}$ then ρ' is ordinally equivalent to ρ. But it is clear that ρ' is not a Luce rule as $\frac{\rho'(a_1,\{a_1,a_2\})}{\rho'(a_2,\{a_1,a_2\})} = \frac{100}{101} > \frac{10}{11} = \frac{\rho(a_1,\{a_1,a_2\})}{\rho(a_2,\{a_1,a_2\})} = \frac{\rho(a_1,\{a_1,a_2,a_3\})}{\rho(a_2,\{a_1,a_2,a_3\})} = \frac{\rho'(a_1,\{a_1,a_2,a_3\})}{\rho'(a_2,\{a_1,a_2,a_3\})}$. □

As a corollary to the above result it follows that the Luce model is neither within menu ordinal nor deterministic. The Luce model does in this sense not coop very well with measurement error.

6.2. Additive perturbed utility model

The additive perturbed utility model studied and introduced by Fudenberg, Iijima, and Strzalecki (2015) can be thought of as a model of deliberate randomization. The decision maker

[11] An SCF ρ is positive if $\rho(a, A) > 0$ for all $a \in A \in \mathscr{A}$.

randomizes by choosing the lottery over alternatives that gives the highest expected utility subject to a perturbation cost. For simplicity of analysis, we will throughout this section assume that $\rho(a, A) > 0$ for all $a \in A$ and $A \in \mathscr{A}$. A function c is a *cost function* if $c : [0,1] \to \mathbb{R} \cup \{\infty\}$ is strictly convex and C^1 over $(0,1)$ and $\lim_{q \to 0} c'(q) = -\infty$. A stochastic function has an APU representation if:

$$\{\rho(a, A)\} = \underset{p \in \Delta(A)}{\arg\max} \sum_{a \in A} \left[u(a) p(a) - c(p(a)) \right].$$

Denote by APU the class of SCFs with an APU representation. From its definition it is not obvious that such a model is ordinal. However, this is immediate once we appeal to the main characterization result in Fudenberg, Iijima, and Strzalecki (2015).They characterize APUs using a property called acyclicity that only impacts the ordinal ranking of choice probabilities.

Proposition 6.2. *The APU model is an ordinal model.*

Proof. Call a pair of sequences $(a_1, A_1), ..., (a_n, A_n)$ and $(b_1, B_1), ..., (b_n, B_n)$ such that $(a_1, ..., a_n)$ is a permutation of $(b_1, ..., b_n)$ and $(A_1, ..., A_n)$ is a permutation of $(B_1, ..., B_n)$ admissible.

> **Acyclicity:** There is no admissible sequence such that $\rho(a_i, A_i) \geq \rho(b_i, B_i)$ for all $i \in \{1, ..., n-1\}$ and $\rho(a_n, A_n) > \rho(b_n, B_n)$.

It is clear that if a stochastic choice function ρ satisfies acyclicity and if ρ' is an increasing transformation of ρ then ρ' satisfies acyclicity. Thus it follows that the APU model is ordinal.

□

6.3. Simple scalability

The simple scalability model (SSM) is formally a generalization of the Luce model. It was first characterized in Tversky (1972). [12] To define it, we need some further notation. Define a relation $\geq_m$ on $\mathbb{R}^m$ by $x \geq_m y$ if $x_1 \geq y_1$ and $x_i \leq y_i$ for all $i \in \{2, ..., m\}$. Similarly, $>_m$ is defined by $x >_m y$ if $x \neq y$ and $x \geq_m y$.

Definition 6.1. A stochastic choice function ρ has a simple scalability representation if there is a scale $u : X \to \mathbb{R}$ and for each $2 \leq k \leq n$ a function F_k such that for each $A = \{a_1, ..., a_k\}$ with $|A| = k$ it holds that

$$\rho(a_1, A) = F_k(u(a_1), ..., u(a_k)). \tag{8}$$

[12] Suppes et al. (1989) corrects an issue with the definition of the simple scalability model presented in Tversky (1972).

Further, each F_k is strictly $\geq_k$ increasing, meaning that: $F_k(x) \geq F_k(y)$ if $x \geq_m y$ and $F_k(x) > F_k(y)$ if $x >_m y$. $\triangleleft$

Intuitively, a stochastic choice model is SSM if the alternatives in A can be scaled by a utility function u such that any choice probability $\rho(a, A)$ is a function $F_{|A|}$ of the scale values. Denote the class of stochastic choice functions with a simple scalability representation by $\texttt{SSM}$. Already from the definition of SSM, due to the functions F_k, it is quite clear that $\texttt{SSM}$ is an ordinal model.

Proposition 6.3. *Simple scalability is an ordinal model.*

Proof. Let (ρ, ρ') be an ordinal data set such that $\rho' \in \texttt{SSM}$. Then it follows by proposition 4.1 that there is a strictly increasing function $v : [0,1] \to [0,1]$ with $v(0) = 0$ and $\rho = v \circ \rho'$. Since $\rho' \in \texttt{SSM}$ there is a scale u and there are strictly $\geq_k$ increasing functions F'_k for $2 \leq k \leq n$ such that equation (8) holds. Thus for each $A = \{a_1, ..., a_k\}$ with $|A| = k$ it holds that

$$\rho(a_1, A) = v \circ \rho'(a_1, A) = v \circ F'_k(u(a_1), ..., u(a_k)).$$

To conclude, it suffices to show that $F_k = v \circ F'_k$ is strictly $\geq_k$ increasing for all $2 \leq k \leq n$. But this is immediate since v is strictly increasing. $\square$

Remark: Many of the popular binary stochastic choice models are ordinal.[13] A binary stochastic choice function is a function $P : X \times X \to [0,1]$ such that $P(a,b) + P(b,a) = 1$ for all $a, b \in X$. A binary stochastic choice function is *Fechnerian* if there is a utility function $u : X \to \mathbb{R}$ and non-decreasing function $F : \mathbb{R} \to \mathbb{R}$ such that $P(a,b) = F(u(a) - u(b))$ for all $a, b \in X$. It follows as a corollary to proposition 4.1 that Fechnerian models are ordinal.

6.4. Random attention models

There is by now a large literature focusing on stochastic choice models where individuals only consider a subset of all available options and then choose a preference maximizing alternative within the subset. Notable in this regard are the papers by Manzini and Mariotti (2014); Brady and Rehbeck (2016); Cattaneo, Ma, Masatlioglu, and Suleymanov (2020). The models in Manzini and Mariotti (2014); Brady and Rehbeck (2016) fail to be ordinal for much the same reason as the Luce model fails to be ordinal. However, the recent random attention model (RAM) by Cattaneo, Ma, Masatlioglu, and Suleymanov (2020) can indeed be shown to be ordinal. [14]

[13] I thank Jay Lu for bringing this to my attention.
[14] A proof is available from the author upon request.

7. Non robust models

Not all models are robust to ordinal perturbations of data. For example, both the Luce and RUM model fail to be ordinal (see section 5 and 6 for further details). Further, some models that are immune to small amounts of mismeasurement fail to be immune to more severe types of mismeasurement. Consider for example the additive perturbed utility model: it is ordinal, but not within menu ordinal. This raises the question of how to test these models using such data. For instance, since the Luce model is not ordinal, we know that an observed violation of IIA is not sufficient to reject it. So the question is, which property is? Which property needs to be violated by observed data to safely conclude that it cannot have been generated by a Luce SCF? I.e. can we devise a test of the Luce model that minimizes the issue of over rejection? Conversely, we know that the Luce model is subject to under rejection: observed data may satisfy IIA, despite this not being the case for underlying behavior. Can we devise a test of the Luce model, which eliminates the issue of under rejection?

In an attempt to address these queries we next introduce the notions of the robust closure and robust interior of a model. Basing a test on the robust closure of a model addresses the over rejection issue, whereas a test based on the robust interior of a model addresses the under rejection issue. As we will see, there is trade-off between these notions, similar to the well known trade-off between type I and type II errors in statistical hypothesis testing.

Robust closures of models. The robust closure of a model is the "smallest" robust model containing a given model. As explained above, basing a test on the robust closure of a model reduces the over rejection issue. A further motivation will follow the formal definition below.

Definition 7.1. The $\mathscr{R}$ *robust closure* of model $\mathtt{M}$, denoted $\overline{\mathscr{R}}[\mathtt{M}]$, is the smallest $\mathscr{R}$ robust model containing $\mathtt{M}$. [15] The *ordinal closure* of model $\mathtt{M}$, denoted $\overline{\mathscr{O}}[\mathtt{M}]$, is the smallest ordinal model containing $\mathtt{M}$. The within menu (deterministic) closure of model $\mathtt{M}$ is defined similarly and denoted $\overline{\mathscr{W}}[\mathtt{M}]$ ($\overline{\mathscr{D}}[\mathtt{M}]$). $\triangleleft$

So why is knowledge about the robust closure of a model useful? Consider the Luce model again. For expositional purposes, assume that measurement error is such that data is ordinal. An observed violation of IIA is then not necessarily a violation of the Luce model. By basing a test of the Luce model on this property, the researcher may end up over rejecting the Luce model (i.e. it may happen that $\rho \notin \mathtt{M}$ and $\rho^* \in \mathtt{M}$). However, suppose that the researcher bases a test of the Luce model on its ordinal closure and rejects the model whenever $\rho \notin \overline{\mathscr{O}}[\mathtt{M}]$.

[15] Formally, $\overline{\mathscr{R}}[\mathtt{M}]$ is the intersection of all robust models that contain $\mathtt{M}$. It is easily verified that robustness is preserved under intersections. Note also that $\mathtt{SCF}$ is an $\mathscr{R}$ robust model, so the robust closure of a model is a well defined notion.

This would eliminate the under rejection issue. Since, measurement error is such that (ρ, ρ^*) is an ordinal data set, if $\rho \notin \overline{\mathscr{O}}[\mathtt{M}]$ then it must be that $\rho^* \notin \mathtt{M}$. Thus it can never be the case that $\rho \notin \overline{\mathscr{O}}[\mathtt{M}]$ and $\rho^* \in \mathtt{M}$.

Robust interiors of models. Under rejection may still be an issue when basing a test of a model on its robust closure. I.e. the researcher may observe $\rho \in \mathscr{R}[\mathtt{M}] \setminus \mathtt{M}$ while $\rho^* \notin \mathtt{M}$. If under rejection is considered a concern one could similarly consider the robust interior of a model $\mathtt{M}$. This is the largest robust model contained in $\mathtt{M}$.

Definition 7.2. The $\mathscr{R}$ *robust interior* of model $\mathtt{M}$, denoted $\mathscr{R}^\circ[\mathtt{M}]$, is the largest $\mathscr{R}$ robust model contained in $\mathtt{M}$.[16] The *ordinal interiror* of model $\mathtt{M}$, denoted $\mathscr{O}^\circ[\mathtt{M}]$, is the largest ordinal model contained in $\mathtt{M}$. The within menu (deterministic) interior of model $\mathtt{M}$ is defined similarly and denoted $\mathscr{W}^\circ[\mathtt{M}]$ ($\mathscr{D}^\circ[\mathtt{M}]$). ◁

Basing a test on the robust interior of a model addresses the under rejection issue. For non-robust models we have that the robust interior of model $\mathtt{M}$ is a strict subset of $\mathtt{M}$. Suppose that the researcher believes that measurement error is such that any pair of SCFs $(\rho, \rho^*) \in \mathscr{R}$ is conceivable. If the researcher bases a test of model $\mathtt{M}$ on its robust interior, she will be inclined to reject model $\mathtt{M}$ whenever $\rho^* \notin \mathtt{M}$ since then $\rho \notin \mathscr{R}^\circ[\mathtt{M}]$ I.e. it may never happen that $\rho^* \notin \mathtt{M}$ and $\rho \in \mathscr{R}^\circ[\mathtt{M}]$.

8. Calculating robust closures/interiors of models

We will focus our discussion in this section around various types of ordinal data. The notions in section 7 will be illustrated by calculating the robust closures/interiors of a few models of choice. Our main focus will be on the random utility model and the Luce model due to their non ordinal nature.

8.1. Random utility model

Let us first consider the random utility model. As described above $\mathtt{RUM}$ is not an ordinal model. Thus, it may be of interest to know exactly what the ordinal properties of it are. I.e. suppose that underlying population behavior ρ^* is described by a RUM, what testable restrictions (if any) does the random utility model impose on observed data ρ in a data set (ρ, ρ^*)? And conversely, observing ρ when is ρ^* consistent with a RUM?

[16] More formally, $\mathscr{R}^\circ[\mathtt{M}]$ is the union of all robust models contained in $\mathtt{M}$. Note that the union of an arbitrary collection of robust models is robust, i.e. this operation is preserved under unions.

Ordinal closure of RUM. We next characterize the ordinal closure of RUM. The main property is called cancellation and is similar to cancellation properties appearing in the literature on subjective probability Kraft, Pratt, and Seidenberg (1959). Call a pair of sequences $(a_1, A_1), ..., (a_n, A_n)$ and $(b_1, B_1), ..., (b_n, B_n)$ *admissible* if there are natural numbers $(k_i)_{i=1}^n$ s.t. $\sum_{i=1}^n k_i \mathbf{1}_{\mathscr{P}(a_i, A_i)}(P) \leq \sum_{i=1}^n k_i \mathbf{1}_{\mathscr{P}(b_i, B_i)}(P)$ for all $P \in \mathscr{P}$.

Cancellation (C): There is no admissible sequence such that $\rho(a_i, A_i) \geq \rho(b_i, B_i)$ for all $i \in \{1, ..., n\}$ and $\rho(a_n, A_n) > \rho(b_n, B_n)$.

To understand cancellation, consider some of its implications. Cancellation implies the well-known monotonicity/regularity property requiring that for all $a \in X$ and $A, B \in \mathscr{A}$: if $A \subseteq B$ then $\rho(a, A) \geq \rho(a, B)$. Monotonicity alone is however not sufficient to characterize $\mathscr{O}[\mathtt{RUM}]$.[17] Finding a "simple" behavioral characterization of ordinal data sets described by random utility models would be desirable. But this is a challenging problem, related to that of finding a behavioral characterization of the random utility model (which to date is open).

Proposition 8.1. *A stochastic choice function ρ satisfies cancellation if and only if $\rho \in \overline{\mathscr{O}}[\mathtt{RUM}]$.*

The proof of proposition 8.1 uses a rational version of Farkas' lemma and is delegated to the appendix. [18] An implication of proposition 8.1 is that the random utility model is "testable" with ordinal data. I.e. there are realizations of choices ρ that are inconsistent with RUM ($\rho \notin \overline{\mathscr{O}}[\mathtt{RUM}]$). Thus, at least in theory, the model can be falsified using ordinal data.

Within menu properties of RUM: Our next result demonstrates that distortion of data could be a great concern for non ordinal models. It shows that the testable implications of the random utility model are vacuous for quite mild forms of measurement error. Thus some care is needed when devising tests of non ordinal models in the presence of measurement error.

We show that there for every stochastic choice function ρ is a RUM μ such that the ranking of alternatives induced by the RUM coincides with the ranking of ρ in every menu A. This type of data is hence not rich enough to test the random utility model. The result also illustrates the rich behavior consistent with a RUM. Indeed, rank the alternatives of any given menu in any way you like and there is a RUM inducing exactly the same ranking of alternatives! As we will see, this is in quite stark contrast to the Luce model, where the ranking of alternatives within menus needs to be consistent with a fixed utility function.

[17] It is straightforward to verify that the following additivity property holds for any such data set: For all $A, B, C \subseteq X$ with $A \supseteq B$ if $\rho(a, A) \geq \rho(a, B)$ then not $\rho(a, B \cup C) > \rho(a, A \cup C)$.

[18] A main technical challenge in proving theorem 8.1 is that the collection of sets of the form $\mathscr{P}(a, A)$ is not an algebra of sets and hence theorem 8.1 does not follow as a direct corollary to similar results in subjective probability.

Proposition 8.2. *For every stochastic choice function ρ there is a RUM μ such that*

$$\rho(a,A) \geq \rho(b,A) \;\Leftrightarrow\; \mu(\mathscr{P}(a,A)) \geq \mu(\mathscr{P}(b,A))$$

for all $a, b \in A$ and $A \in \mathscr{A}$.

The proof of proposition 8.2 is lengthy and deferred to the online appendix. A striking consequence of proposition 8.2 is that there is no relation whatsoever between the rankings that a RUM induces on different menus. For instance, it follows as a corollary to proposition 8.2 that there is a RUM such that $\rho(a,A) > \rho(b,A)$ for all $A \supseteq \{a,b\}$ with $|A| > 2$, but with $\rho(b,ab) > \rho(a,ab)$. As noted above, this behavior distinguishes the RUM from other models of stochastic choice like the Luce model.

Ordinal interior of RUM: The ordinal interior of the random utility model is more difficult to characterize. Note that the single crossing random utility model is ordinal and is a subset of RUM. Thus we already now that the ordinal interior of RUM contains the single crossing random utility model. Are they equal? Unfortunately not, one can show that the ordinal interior of the random utility model is a strict superset of the collection of SCRUMs. To see this, note that the uniform SCF ρ^u defined by $\rho^u(a,A) = \frac{1}{|A|}$ for all $A \subseteq X$ has a random utility representation. Further, one may check that $\{\rho^u\}$ is an ordinal model. Hence $\mathtt{SCRUM} \cup \{\rho^u\}$ is an ordinal model.[19] A full characterization of the ordinal interior of the random utility model is left for future research.

8.2. Luce model

In order to facilitate the discussion of the ordinal closure of the Luce model, we introduce the following ordinal analogue of the Luce model.

Definition 8.1. A stochastic choice function ρ is *ordinally rational* if there is a utility function $u : X \to \mathbb{R}$ such that:

$$u(a) \geq u(b) \;\Leftrightarrow\; \rho(a,A) \geq \rho(b,A)$$

for all $a, b \in A$ and for all $A \in \mathscr{A}$. Denote by $\mathtt{OR}$ the class/model of stochastic choice functions that are ordinally rational. ◁

We do not obtain a full characterization of the of the ordinal closure of the Luce model. However, the following proposition characterizes the within menu closure of the Luce model, and hence provides an "upper bound" on its ordinal closure (note that the ordinal closure of a model is a subset of its within menu closure).

[19] The uniform SCF ρ is not a SCRUM as it fails to satisfy the centrality axiom in Apesteguia et al. (2017).

Proposition 8.3. *The within menu closure of the Luce model equals the collection* $\mathtt{OR}$.

Proof. If ρ is in the within menu closure of the Luce model then there is a Luce rule ρ with utility function $u : X \to \mathbb{R}$ such that:

$$\rho(a, A) \geq \rho(b, A) \Leftrightarrow \frac{u(a)}{\sum_{c \in A} u(c)} \geq \frac{u(b)}{\sum_{c \in A} u(c)} \Leftrightarrow u(a) \geq u(b).$$

By the equivalences above it hence follows that $\rho \in \mathtt{OR}$. If $\rho \in \mathtt{OR}$ then there is a utility function $u : X \to \mathbb{R}$ such that for all $A \in \mathscr{A}$: $(a, A) \succsim (b, A)$ if and only if $u(a) \geq u(b)$. Using the second equivalence above the claim then follows. □

The within menu interior of the Luce model is also fairly straightforward to characterize. It coincides with the singleton set $\{\rho^u\}$ where ρ^u is the uniform SCF defined above.

Proposition 8.4. *The within menu interior of the Luce model equals* $\{\rho^u\}$.

Proof. It is straightforward to check that $\{\rho^u\}$, where ρ^u is the uniform SCF, is an within menu ordinal model. Next, assume that there is some ordinal model $\mathtt{M}$ that is a subset of the Luce model and that is such that $\mathtt{M}$ strictly contains $\{\rho^u\}$. Let $\rho \in \mathtt{M} \setminus \{\rho^u\}$. Since ρ is a non-uniform Luce rule there is a utility function $u : X \to (0, \infty)$ with $u(a) > u(b)$ for some $a, b \in X$ and such that $\rho(a, A) = \frac{u(a)}{\sum_{b \in A} u(b)}$ for all $a \in A$ and $A \subseteq X$. Since $u(a) > u(b)$ it follows that $\rho(a, ab) > \rho(b, ab)$. Let $\varepsilon > 0$ be such that $\rho(a, ab) - \varepsilon > \varepsilon + \rho(b, ab)$. Define a new SCF ρ' such that $\rho'(\cdot, A) = \rho(\cdot, A)$ for all $A \subseteq X$ with $A \neq \{a, b\}$ and set $\rho'(a, ab) = \rho(a, ab) - \varepsilon$ and $\rho'(b, ab) = \rho(b, ab) + \varepsilon$. Then (ρ, ρ') is an within menu data set. However ρ' is not a Luce SCF, since $\frac{\rho'(a,ab)}{\rho'(b,ab)} = \frac{\rho(a,ab)-\varepsilon}{\rho(b,ab)+\varepsilon} \neq \frac{\rho(a,B)}{\rho(b,B)} = \frac{\rho'(a,B)}{\rho'(b,B)}$. □

For testing purposes the properties satisfied by the within menu interior of the Luce model can be considered vacuous. Only a non-generic subset of Luce rules satisfy this property. Thus, when (ρ, ρ^*) is an within menu data set, it is virtually impossible to explain underlying choices using a Luce rule. It is only when $\rho = \rho^u$ that also underlying choices ρ^* are explained by the Luce model. Conversely, the researcher can rule out the Luce model as a proper description of within menu data whenever ρ is not ordinally rational ($\rho \notin \mathtt{OR}$) as underlying choices ρ^* must violate IIA then.

9. Discussion

9.1. Moment restricted data

Our main discussion has so far been focused on restrictions on measurement error that are such that observed choices are ordinally related to underlying choices. As indicated before, the framework developed in this paper may also be of use when considering other common restrictions $\mathscr{R}$ on measurement error. As an illustration, this section discusses a situation, where the researcher believes that data is subject to various moment restrictions on measurement error. Such restrictions are quite extensively discussed in an econometric/statistical literature dealing with measurement error (see e.g. Aguiar and Kashaev (2018)). Formally, a function $g : X \to \mathbb{R}^d$ is called a *moment function.* This function can be thought of as encoding information about the levels of different attributes of alternatives x in X. Roughly speaking, moment restricted data is such that average measurement error within each menu is zero.

Definition 9.1. A data set (ρ, ρ^*) such that:

$$\sum_{a \in A} g(a)\rho(a, A) = \sum_{a \in A} g(a)\rho^*(a, A) \tag{9}$$

for all $A \in \mathscr{A}$ is called a *data set subject to moment condition* g. ◁

Since $\varepsilon(a, A) = \rho(a, A) - \rho^*(a, A)$ for all $a \in A$ and $A \in \mathscr{A}$, equation (9) can equivalently be thought of as a moment condition imposed on measurement error ε. A type of moment restricted data is studied further in Ahn et al. (2018). However, they consider a setting where the grand set of alternatives X is a compact subset of $\mathbb{R}^n$ and where $g(x) = x$ for all $x \in X$. This corresponds to a situation where the researcher only has access to one (the first) moment (average choice data).

Suppose that we have access to data subject to moment condition g on measurement error, what can we say about underlying behavior? Generally such data is quite sparse. The more moments we have access to, the more we can say about underlying behavior. An average choice data set as in Ahn et al. (2018) is generally not as rich as stochastic choice data. But, if we have access to sufficiently many moments then this is the case. We will next try to formalise this intuition. Basically, our next result is a reformulation of the statistical result which says that a distribution with finite support is completely determined by its moments. Before stating it we need to introduce some more terminology. A moment function g is a *spanning moment function* if the vectors $(g(x))_{x \in X}$ are linearly independent (since the range of g is a subset of $\mathbb{R}^d$ this in particular implies that $d \geq n$).

Example 9.1. Let X be a finite set and $u : X \to \mathbb{R}$ a utility/scale function. Define $g(x) = (u(x), u(x)^2, ..., u(x)^d)$ for all $x \in X$, then $\sum_{a \in A} g(a)\rho(a, A)$ gives the first d moments w.r.t. u in menu A. A necessary and sufficient condition for g to be spanning is that u is injective (i.e. $u(x) \neq u(y)$ for all $x, y \in X$). [20] ◁

[20]This can be shown by noticing that the matrix obtained by "stacking" the vectors $g(x)$ =

Proposition 9.1. *Let g be a spanning moment function. If (ρ,ρ^*) is a data set subject to moment condition g then $\rho=\rho^*$.*

Proof. Let (ρ,ρ^*) be subject to moment condition g. Then $\sum_{a\in A} g(a)\rho(a,A)=\sum_{a\in A} g(a)\rho^*(a,A)$, so $\sum_{a\in A} g(a)[\rho(a,A)-\rho^*(a,A)]=0$. By linear independence of the vectors $(g(a))_{a\in A}$ it then follows that $\rho(a,A)=\rho^*(a,A)$ for all $a\in A$. □

In words, knowing the first n moments of ρ^* is equivalent to knowing the entire underlying distribution. As an immediate corollary to the proposition above it follows that any model is robust to measurement error subject to a spanning moment condition g. Formally, given a moment function g let $\mathscr{R}^g$ denote the subset of pairs of SCFs (ρ,ρ^*) with $\sum_{a\in A} g(a)\rho(a,A)=\sum_{a\in A} g(a)\rho^*(a,A)$ for all $A\in\mathscr{A}$.

Corollary 9.2. *Let g be a spanning moment function. Then every model $\mathtt{M}$ is $\mathscr{R}^g$ robust.*

This implies that moment restricted data (w.r.t. some spanning moment function) can be used as an alternative empirical primitive to stochastic choice data. To illustrate this point we next provide a characterization of the Luce rule in terms of moment restricted data. Our characterization is similar to the average choice characterization of the Luce rule in Ahn et al. (2018) and uses a version of their path independence axiom. A notable difference between our characterization and theirs is that we assume a richer data set in the sense that the researcher has access to several moments, not just one. In contrast to moment restricted data, average choice data as in Ahn et al. (2018) does not necessarily uniquely pin down the underlying SCF ρ^*. To state the path independence property characterizing the Luce rule, let $\text{co}(A)$ denote the convex hull of a set $A\subseteq\mathbb{R}^d$ (the intersection of all convex sets containing A). Further, let $E_\rho[g|A]=\sum_{a\in A} g(a)\rho(a,A)$ for all $A\in\mathscr{A}$.

Plott's partial path independence w.r.t. g: Let g be a moment function. A stochastic choice function ρ satisfies *Plott's partial path independence w.r.t. g* if

$$E_\rho[g|A\cup B]\in \text{co}\{E_\rho[g|A],E_\rho[g|B]\}$$

for all disjoint $A,B\in\mathscr{A}$.

The axiom can be seen as a generalization of Plott's path independence axiom, which roughly speaking requires the choice from a union of two disjoint sets to be equal the choice from either one of those two sets. For an extensive discussion of this property we refer the

$(1,u(x),u(x)^2,...,u(x)^d)$ next to each other is an Vandermonde matrix, and such a matrix has non zero determinant if and only if u is injective.

interested reader to Ahn et al. (2018). We are now ready to state our characterization of the Luce rule.

Proposition 9.3. *Let g be a spanning moment function. Then a positive stochastic choice function ρ is a Luce rule if and only if it satisfies Plott's partial path independence w.r.t. g.*

Proof. Let $b \in A$ and $A \subseteq X$. Our aim is to show that $\frac{\rho(a,A)}{\rho(c,A)} = \frac{\rho(a,A\setminus b)}{\rho(c,A\setminus b)}$ for all $c \in A \setminus b$, which in particular (by induction) implies that IIA holds. Define $f_g(A) = E_\rho[g|A]$. Note that

$$f_g(A) = \alpha f_g(b) + (1-\alpha) f_g(A \setminus b) =$$

$$= \alpha g(b) + (1-\alpha) \sum_{c \in A \setminus b} g(c)\rho(c, A \setminus b).$$

Further, we know that $f_g(A) = \sum_{c\in A} g(c)\rho(c,A) = \rho(b,A)g(b) + (1-\alpha)\sum_{c\in A\setminus b} g(c)\rho(c,A)$. Hence, we have that $\rho(b,A)g(b)+(1-\alpha)\sum_{c\in A\setminus b} g(c)\rho(c,A\setminus b) = \alpha g(b)+(1-\alpha)\sum_{c\in A\setminus b} g(c)\rho(c,A)$ and rearranging the later expression gives:

$$(\rho(b,A) - \alpha)g(b) + \sum_{c\in A\setminus b} [(1-\alpha)\rho(c, A\setminus b) - \rho(c,A)]g(c) = 0.$$

By linear independence of the vectors $g(c)$, it then follows that $(1-\alpha)\rho(c,A\setminus b) - \rho(c,A) = 0$ for all $c \in A \setminus b$, which in particular implies that $\frac{\rho(a,A)}{\rho(c,A)} = \frac{\rho(a,A\setminus b)}{\rho(c,A\setminus b)}$ for all $c \in A \setminus b$. The proof that every Luce rule ρ satisfies Plott's partial path independence w.r.t. g is straightforward and omitted. □

9.2. Comparative robustness

Previous sections discussed various notions of robustness, depending on the extent and type of measurement error that data is subject to. From the perspective of a researcher it would be useful if models could be assessed based on their degree of robustness. In other words, *how* sensitive is a particular model to measurement error? Is the SCRUM more robust to measurement error than the random utility model? More generally, is model $\mathtt{M}$ more robust than model $\mathtt{M}'$? Among a colleciton of models $\{\mathtt{M}_1, \ldots, \mathtt{M}_n\}$ considered as plausible by the researcher which one is the most robust, and in this sense most suitable for empirical analysis?

To answer such questions, we next briefly outline a method to compare the robustness of various models of choice. A natural way to define such a comparative notion of robustness is to say that a model $\mathtt{M}$ is more robust than another model $\mathtt{M}'$ if and only if $\mathtt{M}$ is $\mathscr{R}$ robust for all restrictions $\mathscr{R}$ for which $\mathtt{M}'$ is $\mathscr{R}$ robust. However, an undesirable feature of this order is that any pair of models $\mathtt{M}, \mathtt{M}'$ are incomparable according to it.[21] To meaningfully compare pairs of models we need to consider more specific restrictions on measurement error. Given our focus on ordinal models, we will consider an order which compares the degree of ordinality of different models. To this end, define a relation $\gg$ by

$$\mathtt{M} \gg \mathtt{M}' \Leftrightarrow \text{ for all restrictions } \mathscr{R} \text{ with } \mathscr{O} \subseteq \mathscr{R} \subseteq \mathscr{D} : \mathscr{R}[\mathtt{M}'] = \mathtt{M}' \Rightarrow \mathscr{R}[\mathtt{M}] = \mathtt{M}.$$

The relation above thus requires measurement error to be restricted such that $\mathscr{O} \subseteq \mathscr{R} \subseteq \mathscr{D}$. A data set $(\rho, \rho^*) \in \mathscr{R}$ with $\mathscr{O} \subseteq \mathscr{R} \subseteq \mathscr{D}$ contains more information than a deterministic data set, but less so than an ordinal data set, and can in this sense be viewed as a type of ordinal data set. Hence, model $\mathtt{M}$ can arguably be viewed as "more ordinal" than model $\mathtt{M}'$ if $\mathtt{M} \gg \mathtt{M}'$. Using this order we are able to formally compare the degree of ordinality of the different random utility models considered in section 5.

Proposition 9.4. *We have that dRUM $\gg$ SCRUM $\gg$ RUM.*

Proof. First, note that RUM is not $\mathscr{R}$ robust for any $\mathscr{O} \subseteq \mathscr{R} \subseteq \mathscr{D}$. Secondly, we note that if we define $\mathscr{A} = \{(\rho, \rho^*) : \rho(a, A) \geq \rho(a, B) \Leftrightarrow \rho^*(a, A) \geq \rho^*(a, B) \text{ for all } a \in X \text{ and } A, B \in \mathscr{A}\}$ then SCRUM is robust w.r.t. $\mathscr{A}$ (the proof is identical to the proof that SCRUM is ordinal). Hence SCRUM $\gg$ RUM. Finally, we note that since dRUM is deterministic if follows that dRUM is $\mathscr{R}$ robust for all $\mathscr{O} \subseteq \mathscr{R} \subseteq \mathscr{D}$. Thus, since $\gg$ is a transitive relation, it follows that dRUM $\gg$ SCRUM $\gg$ RUM. □

It would also be interesting to compare other models of choice using the above ordering. Is for instance the GPCR model more ordinal than the APU model? Also note that the above notion of comparative ordinality is one out of several plausible ways in which a comparative notion of ordinality/robustness could be defined. We leave a more detailed study of these issues for future research.

9.3. Classifying properties of SCFs

The framework developed in this paper may be used to classify properties of stochastic choice models. We identify with each property P a subset of SCFs satisfying this property. As an example, the well known monotonicity/regularity property, requiring that $\rho(a, A) \geq \rho(a, B)$ for all $A \subseteq B$ and $a \in A$, is identified with the subset of SCFs satisfying monotonicity. In this way, properties and models (as defined in the previous section) are in one-to-one

[21] This is because this relation is an extremely conservative ordering that takes *all* restrictions on measurement error into account.

correspondence with each other. Each property corresponds to a subset $\mathtt{P} \subseteq \mathtt{SCF}$ as follows: $\mathtt{P} := \{\rho \in \mathtt{SCF} : \rho \text{ satisfies property P}\}$.

Definition 9.2. A property P is an ordinal (within menu ordinal, deterministic) property if the model $\mathtt{P}$ is an ordinal (within menu ordinal, deterministic) model. ◁

Viewed in this way, a set of properties characterize a model if and only if their intersection equals the model.

Definition 9.3. We say that a model $\mathtt{M}$ is characterized by using a set of properties $\mathtt{P}_1, \ldots, \mathtt{P}_n$ if

$$\mathtt{M} = \bigcap_{i=1}^{k} \mathtt{P}_i.$$

◁

Every ordinal (within menu ordinal, deterministic) model is in a trivial way characterized using ordinal properties, take $\mathtt{P} = \mathtt{M}$. Conversly, if a model is characterized using a set of ordinal (within menu ordinal, deterministic) properties $\mathtt{P}_1, \ldots, \mathtt{P}_k$ then, since the intersection of ordinal (within menu ordinal, deterministic) properties is ordinal (within menu ordinal, deterministic), it follows that $\mathtt{M}$ is an ordinal model. We have the following observation:

Proposition 9.5. *A model $\mathtt{M}$ is characterized by a collection of ordinal (within menu ordinal, deterministic) properties $\mathtt{P}_1, \ldots, \mathtt{P}_n$ if and only if $\mathtt{M}$ is ordinal (within menu ordinal, deterministic).*

Thus, if an analyst knows that a model is ordinal, then this suggests that there is a characterization of the model entirely in terms of ordinal properties. Say that a researcher wishes to take a model to data and test it. If it is known that a model could be expressed using only ordinal properties, then this is convenient for the analyst, as less precise information about choice probabilities is required in testing the properties of the model. As an example, consider the dual Random Utility Model which is a deterministic model. Due to dRUM being deterministic there ought to be a characterization of it using deterministic properties. This is indeed correct, as the properties characterizing dRUM in Manzini and Mariotti (2018) can all be shown to be deterministic properties.

9.4. What are "reasonable" restrictions on measurement error?

This paper has mainly studied restrictions on measurement error where observed choices are ordinally related to underlying choices. Are these reasonable restrictions? As discussed before, this depends a bit on the application at hand and the researcher's knowledge on measurement error. For instance if measurement error is "small" then the ordinal assumption seems like a quite good approximation, whereas if measurement error is large then deterministic data is a reasonable "lower bound" on the information available to the analyst. An alternative way to approach the issue of deciding on a reasonable restriction on measurement error is through the axiomatic approach. I.e. if $\mathscr{R}$ ought to measure the extent and type of measurement error in data, what properties on $\mathscr{R}$ are then reasonable to assume? Note that $\mathscr{R}$ can be viewed as a binary relation on the set of all SCFs. A similar approach is undertaken in Balakrishnan et al. (2021) to infer choice correspondences (what they call "Fishburn imputations") from stochastic choice data. We leave it for future research to identify reasonable conditions on $\mathscr{R}$ characterizing various restrictions on measurement error (such as ordinal/moment restrictions).

A. Proofs

A.1. Proof of proposition 2.1

Proof. Let $\texttt{SCF}^*$ denote the subset of SCFs such that $\rho(a,A) \neq \rho(b,B)$ for all $a \in A$ and $b \in B$ and $A, B \in \mathscr{A}$. This set has measure one when $\texttt{SCF}$ is equipped with the Lebesgue measure restricted to $\texttt{SCF}$ (and normalized to equal one). Let $\rho \in \texttt{SCF}^*$. Then $\rho(a,A) \neq \rho(b,B)$ for all $a \in A$ and $b \in B$ and $A, B \in \mathscr{A}$. Let

$$\varepsilon = \min_{a \in A, b \in B, A, B \in \mathscr{A}: \rho(a,A) > \rho(b,B)} \left|\rho(a,A) - \rho(b,B)\right|.$$

Let ρ' be an SCF such that $||\rho' - \rho|| < \frac{\varepsilon}{4}$. It is then clear that (ρ, ρ') is an ordinal data set. □

A.2. Proof of proposition 5.2

Proof. It suffices to show that if $\mathscr{A}$ satisfies the premises of the proposition, then an SCF is a random utility model if and only if it satisfies monotonicity, i.e. for all $a \in A \subseteq B$: $\rho(a,A) \geq \rho(a,B)$. The set of all SCFs satisfying monotonicity is clearly an ordinal model. The proof that the random utility model is characterized by monotonicity on this domain is straightforward. Suppose that ρ satisfies monotonicity. Note that the collection $\mathscr{H} = \{\mathscr{P}(b,X) \cap \mathscr{P}(a, X \setminus b) : b \in X, a \in X \setminus b\}$ forms a partition of $\mathscr{P}$. Moreover, $\mathscr{P}(b,X) = \cup_{a \in X \setminus b} \mathscr{P}(b,X) \cap \mathscr{P}(a, X \setminus b)$ and $\mathscr{P}(a, X \setminus b) = [\mathscr{P}(a,X) \cup \mathscr{P}(b,X)] \cap \mathscr{P}(a, X \setminus b) = [\mathscr{P}(a,X) \cap \mathscr{P}(a, X \setminus b)] \cup [\mathscr{P}(b,X) \cap \mathscr{P}(a, X \setminus b)]$, so for each $A \in \mathscr{A}$ we have that $\mathscr{P}(a,A)$ can be written as a disjoint union of members from the partition $\mathscr{H}$.

Define μ on $\mathscr{P}(b,X)\cap\mathscr{P}(a,X\setminus b)$ by setting $\mu(\mathscr{P}(b,X)\cap\mathscr{P}(a,X\setminus b))=\rho(a,X\setminus b)-\rho(a,X)$. To show that μ is a measure (on the algebra generated by the partition $\mathscr{H}$) it suffices to show that μ is positive on $\mathscr{H}$ and that μ sums to one. By monotonicity of ρ it follows that $\mu(\mathscr{P}(b,X)\cap\mathscr{P}(a,X\setminus b))=\rho(a,X\setminus b)-\rho(a,X)\geq 0$, so μ is positive. Next note that

$$\sum_{b\in X}\sum_{a\in X\setminus b}\mu(\mathscr{P}(b,X)\cap\mathscr{P}(a,X\setminus b))=\sum_{b\in X}\sum_{a\in X\setminus b}\rho(a,X\setminus b)-\rho(a,X)=\sum_{b\in X}\rho(b,X)=1.$$

Thus μ is a measure. Finally, we have for all $b\in X$: $\rho(b,X)=\sum_{a\in A\setminus b}\big(\rho(a,X\setminus b)-\rho(a,X)\big)=\sum_{a\in A\setminus b}\mu(\mathscr{P}(b,X)\cap\mathscr{P}(a,X\setminus b))=\mu(\mathscr{P}(b,X))$. Further, since we have $\mathscr{P}(a,X\setminus b)=\mathscr{P}(a,X)\cup(\mathscr{P}(b,X)\cap\mathscr{P}(a,X\setminus b))$ it follows by additivity of μ that $\mu(\mathscr{P}(a,X\setminus b))-\mu(\mathscr{P}(a,X))=\mu(\mathscr{P}(b,X)\cap\mathscr{P}(a,X\setminus b))$. Thus, it follows that $\rho(a,X\setminus b)=\rho(a,X\setminus b)-\rho(a,X)+\rho(a,X)=\mu(\mathscr{P}(b,X)\cap\mathscr{P}(a,X\setminus b))+\mu(\mathscr{P}(a,X))=\mu(\mathscr{P}(a,X\setminus b))-\mu(\mathscr{P}(a,X))+\mu(\mathscr{P}(a,X))=\mu(\mathscr{P}(a,X\setminus b))$. $\square$

A.3. Proof of proposition 5.7

Proof. Let $\rho\in\mathtt{RSM}$ and ρ' be an SCF s.t. (ρ,ρ') is an within menu data set. Since $\rho\in\mathtt{RSM}(\mathscr{H})$ there is a finite set $I=\{1,...,n\}$, a sequential collection of choice correspondences $\{C_1,...,C_n\}\in\mathscr{H}$ and a stopping function π such that $\rho(a,A)=\sum_{i\in I:a\in C_i(A)}\frac{\pi(C_i,A)}{|C_i(A)|}$ for all $a\in A$ and $A\in\mathscr{A}$. We claim that there is a stopping function π' such that ρ' is an RSM with the same sequential collection $\{C_1,...,C_n\}$ as ρ.

Let i_1 be the smallest index such that $C_{i_1}(A)\setminus C_{i_1+1}(A)$ is non empty. By induction, one can show that there are indices $\{i_1,...,i_m\}$ such that i_{k+1} is the smallest index in $\{i_k+1,...,n\}$ with $C_{i_k}(A)\setminus C_{i_{k+1}}(A)\neq\emptyset$ for all $k<m$ and $C_i(A)=C_{i_m}(A)$ for all $i\geq i_m$. Pick a sequence of alternatives $\{a_{i_1},...,a_{i_m}\}\subseteq A$ such that $a_{i_k}\in C_{i_k}(A)\setminus C_{i_{k+1}}(A)$ for all $k<m$ and $a_{i_m}\in C_{i_m}(A)$. Note that $\rho(a_{i_{k+1}},A)\geq\rho(a_{i_k},A)$ for all $k\geq 1$, implying that $\rho'(a_{i_{k+1}},A)\geq\rho'(a_{i_k},A)$ for all $k\geq 1$. Thus, it is possible to define a function π' such that $\pi'(C_i,A)\geq 0$ for all $i\in I$ and such that

$$\sum_{i=1}^{i_1}\frac{\pi'(C_i,A)}{|C_{i_1}(A)|}=\rho'(a_{i_1},A)$$

$$\sum_{i=i_k+1}^{i_{k+1}}\frac{\pi'(C_i,A)}{|C_{i_{k+1}}(A)|}=\rho'(a_{i_{k+1}},A)-\rho'(a_{i_k},A)$$

for all $k<m$. We next check that π' is a stopping function. To see this note that:

$$\sum_{i\in I}\pi'(C_i,A)=|C_{i_1}(A)|\rho'(a_{i_1},A)+\sum_{k=1}^{m-1}|C_{i_{k+1}}(A)|[\rho'(a_{i_{k+1}},A)-\rho'(a_{i_k},A)]=$$

$$= \sum_{k=1}^{m-1} |C_{i_k}(A) \setminus C_{i_{k+1}}(A)| \rho'(a_{i_k}, A) + |C_{i_m}(A)| \rho'(a_{i_m}, A) = \sum_{a \in A} \rho'(a, A) = 1.$$

The last equality follows since $\rho'(a, A) = \rho'(a_{i_k}, A)$ for all $a \in C_{i_k}(A) \setminus C_{i_{k+1}}(A)$. We finally check that ρ' satisfies the defining equation (7) of an RSM. Let $a \in A$. Then either there is an index $k < m$ such that $a \in C_{i_k}(A) \setminus C_{i_k+1}(A)$ or otherwise $a \in C_{i_m}(A)$. Hence, by definition of an RSM, it follows that $\rho(a, A) = \rho(a_{i_k}, A)$ for some $k \leq m$. Thus we have $\rho'(a, A) = \rho'(a_{i_k}, A) = \sum_{i=1}^{i_k} \frac{\pi'(C_i, A)}{|C_i(A)|} = \sum_{i \in I: a \in C_i(A)} \frac{\pi'(C_i, A)}{|C_i(A)|}$. This proves the claim. □

A.4. Rational version of Farkas' lemma

The proof of theorem 8.1 is an application of a rational version of Farkas' Lemma. The lemma below follows as a corollary to theorem 1.6.1. in Stoer and Witzgall (2012) (see also Echenique and Saito (2015) who applies this result in a different context). Alternatively, the result can be derived from theorem 3.2 in Fishburn (1973) (see also Chambers and Echenique (2014) for an application of this result).

Lemma A.1. *Let A be an $n \times m$ matrix, B an $n \times l$ matrix and C be an $n \times k$ matrix all with rational entries. Then exactly one of the following statements is true.*

1. *There are $x \in \mathbb{Q}^m$, $y \in \mathbb{Q}^l$ and $z \in \mathbb{Q}^k$ such that $Ax + By + Cz = 0$, $x \geq 0$, $y \geq 0$ and $z > 0$.*
2. *There is an $x \in \mathbb{Q}^n$ such that $A^T x \geq 0$, $B^T x \geq 0$ and $C^T x \gg 0$.*

A.5. Proof of theorem 8.1

To prove theorem 8.1 we will show a more general result from which it follow as corollary. Call a triple $(X, \mathscr{A}, \succsim)$ where $\succsim$ is a binary relation on $X \times \mathscr{A}$ a data set. Binary relation $\succsim$ is non-trivial, if there are $(a, A), (b, B) \in X \times \mathscr{A}$ with $(a, A) \succ (b, B)$. Let $\mathscr{S}$ denote a subset of the collection of choice functions on $\mathscr{A}$, i.e. $\mathscr{S} \subseteq \mathscr{C}$. For all $A \subseteq X$ and $a \in A$ denote by $\mathscr{S}(a, A) = \{c \in \mathscr{S} : c(A) = a\}$. Call a pair of sequences $(a_1, A_1), \ldots, (a_n, A_n)$ and $(b_1, B_1), \ldots, (b_n, B_n)$ $\mathscr{S}$-*admissible* if (1) $a_i \in A_i$ and $b_i \in B_i$ for all i, (2) there are natural numbers $(k_i)_{i=1}^n$ such that $\sum_{i=1}^n k_i \mathbf{1}_{\mathscr{S}(a_i, A_i)}(c) \leq \sum_{i=1}^n k_i \mathbf{1}_{\mathscr{S}(b_i, B_i)}(c)$ for all $c \in \mathscr{S}$.

> $\mathscr{S}$-**Cancellation (C):** There is no $\mathscr{S}$- admissible sequence such that $(a_i, A_i) \succsim (b_i, B_i)$ for all $i \in \{1, \ldots, n\}$ and $(a_n, A_n) \succ (b_n, B_n)$.

Theorem A.2. *A non-trivial data set $(X, \mathscr{A}, \succsim)$ satisfies $\mathscr{S}$- cancellation if and only if there is a measure μ on $\mathscr{S}$ such that*

$$(a, A) \succ (b, B) \;\Rightarrow\; \mu(\mathscr{S}(a, A)) > \mu(\mathscr{S}(b, B)) \tag{10}$$

$$(a,A) \succsim (b,B) \;\Rightarrow\; \mu(\mathscr{S}(a,A)) \geq \mu(\mathscr{S}(b,B)) \tag{11}$$

for all $(a,A),(b,B) \in X \times \mathscr{A}$.

Proof. Let $(X,\mathscr{A},\succsim)$ be a imprecise data set such that $\mathscr{S}$-cancellation and non-triviality holds. For all $(a,A),(b,B) \in X \times \mathscr{A}$ with $(a,A) \succsim (b,B)$ define a vector $\mathbf{x}[(a,A),(b,B)] = \mathbf{1}_{\mathscr{S}(a,A)} - \mathbf{1}_{\mathscr{S}(b,B)}$. Define a matrix A by letting the columns correspond to the vectors $\mathbf{x}[(a,A),(b,B)]$. Let B be the identity matrix, i.e. $B = I$. Define for all $(a,A),(b,B) \in X \times \mathscr{A}$ with $(a,A) \succ (b,B)$ a vector $\mathbf{y}[(a,A),(b,B)] = \mathbf{1}_{\mathscr{S}(a,A)} - \mathbf{1}_{\mathscr{S}(b,B)}$. Define a matrix C by letting its columns correspond to the vectors $\mathbf{y}[(a,A),(b,B)]$. Since $(X,\mathscr{A},\succsim)$ is non-trivial there is at least one such vector $\mathbf{y}[(a,A),(b,B)]$.

STEP 1: We first claim that there are no $x \in \mathbb{Q}^m$, $y \in \mathbb{Q}^l$ and $z \in \mathbb{Q}^k$ such that $Ax + By + Cz = 0$, $x \geq 0$, $y \geq 0$ and $z > 0$.

Suppose that there are $x \in \mathbb{Q}^m$, $y \in \mathbb{Q}^l$ and $z \in \mathbb{Q}^k$ such that $Ax + By + Cz = 0$, $x \geq 0$, $y \geq 0$ and $z > 0$. Hence there are positive rational numbers $x_1, \ldots x_m$ $y_1, \ldots. y_l$ and positive $z_1, \ldots, z_k$ with $z_j > 0$ for some $j \geq 1$ such that

$$\sum_{i=1}^{n} x_i \mathbf{x}[(a_i,A_i),(b_i,B_i)] + \sum_{i=1}^{l} y_i \mathbf{e}_i + \sum_{i=1}^{k} z_i \mathbf{y}[(c_i,C_i),(d_i,D_i)] = 0,$$

where $(a_i,A_i) \succsim (b_i,B_i)$ for all $i \in \{1,\ldots,m\}$ and $(c_i,C_i) \succ (d_i,D_i)$ for all $i \in \{1,\ldots,k\}$.

Multiplying the expression above with least common denominators it follows that there are integers $k_1,\ldots,k_m$ $s_1,\ldots,s_l$ and $r_1,\ldots,r_k$ such that

$$\sum_{i=1}^{n} k_i \mathbf{x}[(a_i,A_i),(b_i,B_i)] + \sum_{i=1}^{l} s_i \mathbf{e}_i + \sum_{i=1}^{k} r_i \mathbf{y}[(c_i,C_i),(d_i,D_i)] = 0,$$

and this further implies that

$$\sum_{i=1}^{n} k_i \mathbf{x}[(a_i,A_i),(b_i,B_i)] + \sum_{i=1}^{k} r_i \mathbf{y}[(c_i,C_i),(d_i,D_i)] \leq$$

$$\sum_{i=1}^{n} k_i \mathbf{x}[(a_i,A_i),(b_i,B_i)] + \sum_{i=1}^{l} s_i \mathbf{e}_i + \sum_{i=1}^{k} r_i \mathbf{y}[(c_i,C_i),(d_i,D_i)] = 0$$

The latter inequality holds if and only if

$$\sum_{i=1}^{n} k_i \left[\mathbf{1}_{\mathscr{S}(a_i,A_i)} - \mathbf{1}_{\mathscr{S}(b_i,B_i)}\right] + \sum_{i=1}^{k} r_i \left[\mathbf{1}_{\mathscr{S}(c_i,C_i)} - \mathbf{1}_{\mathscr{S}(d_i,D_i)}\right] \leq 0.$$

But $\mathscr{S}$-cancellation now implies that $(d_i, D_i) \succsim (c_i, C_i)$ for all $i \in \{1, ..., k\}$. A contradiction. By lemma A.1 it follows that there is an $x \in \mathbb{Q}^n$ such that $A^T x \geq 0$, $B^T x \geq 0$ and $C^T x \gg 0$.

STEP 2: Define a set function by:

$$\mu(R) = \frac{x \cdot \mathbf{1}_R}{x \cdot \mathbf{1}_{\mathscr{S}}}$$

for all $R \subseteq \mathscr{S}$. We next show that μ is a probability measure. To do this we need to check that $\mu(\cdot)$ is a positive, finitely additive measure and that $\mu(\mathscr{S}) = 1$. It is clear that $\mu(\mathscr{S}) = \frac{x \cdot \mathbf{1}_{\mathscr{S}}}{x \cdot \mathbf{1}_{\mathscr{S}}} = 1$. Additivity follows since for all disjoint $R, S \subseteq \mathscr{S}$ we have $\mu(R \cup S) = \frac{x \cdot \mathbf{1}_{R \cup S}}{x \cdot \mathbf{1}_{\mathscr{S}}} = \frac{x \cdot [\mathbf{1}_R + \mathbf{1}_S]}{x \cdot \mathbf{1}_{\mathscr{S}}} = \frac{x \cdot \mathbf{1}_R}{x \cdot \mathbf{1}_{\mathscr{S}}} + \frac{x \cdot \mathbf{1}_S}{x \cdot \mathbf{1}_{\mathscr{S}}} = \mu(R) + \mu(S)$. Positivity follows by noting that $B^T x \geq 0$ implies that $x \geq 0$. Using this observation and that the inner product of two positive vectors is a positive scalar it follows that $\mu(R) = \frac{x \cdot \mathbf{1}_R}{x \cdot \mathbf{1}_{\mathscr{S}}} \geq 0$ for all $R \subseteq \mathscr{S}$. Thus μ is a probability measure.

STEP 3: For all $(a, A), (b, B) \in X \times \mathscr{A}$: If $(a, A) \succsim (b, B)$ then $\mu(\mathscr{S}(a, A)) \geq \mu(\mathscr{S}(b, B))$ and if $(a, A) \succ (b, B)$ then $\mu(\mathscr{S}(a, A)) > \mu(\mathscr{S}(b, B))$.

If $(a, A) \succsim (b, B)$ then since $A^T x \geq 0$ it follows that $x \cdot \mathbf{x}[(a, A), (b, B)] \geq 0$ and hence $x \cdot (\mathbf{1}_{\mathscr{S}(a, A)} - \mathbf{1}_{\mathscr{S}(b,B)}) \geq 0$ implying that $\mu(\mathscr{S}(a, A)) \geq \mu(\mathscr{S}(b, B))$. If $(a, A) \succ (b, B)$ then since $C^T x \gg 0$ it follows that $x \cdot \mathbf{x}[(a, A), (b, B)] > 0$ and hence $x \cdot (\mathbf{1}_{\mathscr{S}(a,A)} - \mathbf{1}_{\mathscr{S}(b,B)}) > 0$. Thus $\mu(\mathscr{S}(a, A)) > \mu(\mathscr{S}(b, B))$.

Conversely, let $(X, \mathscr{A}, \succsim)$ be a triple such that equation (10)-(11) holds. We show that $\mathscr{S}$-cancellation holds. Let $(a_1, A_1), ..., (a_n, A_n)$ and $(b_1, B_1), ..., (b_n, B_n)$ be such that $(a_i, A_i) \succsim (b_i, B_i)$ for all $i \in \{1, ..., n\}$ and assume that there is a sequence of natural numbers $(k_i)_{i=1}^n$ such that

$$\sum_{i=1}^n k_i \mathbf{1}_{\mathscr{S}(a_i, A_i)}(c) \leq \sum_{i=1}^n k_i \mathbf{1}_{\mathscr{S}(b_i, B_i)}(c)$$

for all $c \in \mathscr{S}$. Then taking expectations it follows that

$$\sum_{i=1}^n k_i \mu(\mathscr{S}(a_i, A_i)) = E_\mu \left[\sum_{i=1}^n k_i \mathbf{1}_{\mathscr{S}(a_i, A_i)} \right] \leq E_\mu \left[\sum_{i=1}^n k_i \mathbf{1}_{\mathscr{S}(b_i, B_i)} \right] = \sum_{i=1}^n k_i \mu(\mathscr{S}(b_i, B_i)). \qquad (12)$$

Note that $(a_i, A_i) \succsim (b_i, B_i)$ implies that $\mu(\mathscr{S}(a_i, A_i)) \geq \mu(\mathscr{S}(b_i, B_i))$ for all $i \in \{1, ..., n\}$. If $(a_n, A_n) \succ (b_n, B_n)$ then $\mu(\mathscr{S}(a_n, A_n)) > \mu(\mathscr{S}(b_n, B_n))$. It then follows that $\sum_{i=1}^n k_i \mu(\mathscr{S}(a_i, A_i)) > \sum_{i=1}^n k_i \mu(\mathscr{S}(b_i, B_i))$. A contradiction to equation (12). □

We now prove theorem 8.1. Let ρ be an SCF satisfying cancellation. Define a binary relation $\succsim$ on $X \times \mathscr{A}$ by $(a, A) \succsim (b, B)$ if and only if $\rho(a, A) \geq \rho(b, B)$. Since ρ satisfies cancellation it follows that $\succsim$ satisfies $\mathscr{P}$-cancellation, where we recall that $\mathscr{P}$ is the set of all linear orders. Hence, by theorem A.2 there is a measure μ on $\mathscr{P}$ such that for all $(a,A),(b,B) \in X \times \mathscr{A}$: $(a,A) \succ (b,B) \;\Rightarrow\; \mu(\mathscr{P}(a,A)) > \mu(\mathscr{P}(b,B))$ and $(a,A) \succsim (b,B) \;\Rightarrow\; \mu(\mathscr{P}(a,A)) \geq \mu(\mathscr{P}(b,B))$. Since, by construction, $\succsim$ is complete it then follows that $(a,A) \succsim (b,B) \;\Leftrightarrow\; \mu(\mathscr{P}(a,A)) \geq \mu(\mathscr{P}(b,B))$ for all $(a,A),(b,B) \in X \times \mathscr{A}$. It thus follows that $\rho \in \mathscr{O}[\mathtt{RUM}]$.

References

J. Abaluck and A. Adams. What do consumers consider before they choose? identification from asymmetric demand responses. *Quarterly Journal of Economics, forthcoming.*, 2020.

V. H. Aguiar and N. Kashaev. Stochastic revealed preferences with measurement error. *Review of Economic Studies, forthcoming.*, 2018.

D. S. Ahn, F. Echenique, and K. Saito. On path independent stochastic choice. *Theoretical Economics*, 13(1):61–85, 2018.

S. Alon and E. Lehrer. Subjective multi-prior probability: A representation of a partial likelihood relation. *Journal of Economic Theory*, 151:476–492, 2014.

J. Apesteguia and M. A. Ballester. A measure of rationality and welfare. *Journal of Political Economy*, 123(6):1278–1310, 2015.

J. Apesteguia and M. A. Ballester. Separating predicted randomness from residual behavior. *Journal of the European Economic Association*, 2020.

J. Apesteguia, M. A. Ballester, and J. Lu. Single-crossing random utility models. *Econometrica*, 85(2):661–674, 2017.

N. Balakrishnan, E. A. Ok, and P. Ortoleva. Inferential choice theory. 2021.

R. L. Brady and J. Rehbeck. Menu-dependent stochastic feasibility. *Econometrica*, 84(3): 1203–1223, 2016.

M. D. Cattaneo, X. Ma, Y. Masatlioglu, and E. Suleymanov. A random attention model. *Journal of Political Economy*, 128(7):000–000, 2020.

C. P. Chambers and F. Echenique. On the consistency of data with bargaining theories. *Theoretical Economics*, 9(1):137–162, 2014.

V. Dardanoni, P. Manzini, M. Mariotti, H. Petri, and C. J. Tyson. Mixture choice data: Revealing preferences and cognition. 2020.

G. De Clippel and K. Rozen. Bounded rationality and limited datasets. *Theoretical Economics, forthcoming*, 2020.

R. Dutta. Gradual pairwise comparison and stochastic choice. *Theoretical Economics*, 15(4): 1335–1364, 2020.

F. Echenique and K. Saito. Savage in the market. *Econometrica*, 83(4):1467–1495, 2015.

F. Echenique, S. Lee, and M. Shum. The money pump as a measure of revealed preference violations. *Journal of Political Economy*, 119(6):1201–1223, 2011.

P. C. Fishburn. Weak qualitative probability on finite sets. *The Annals of Mathematical Statistics*, 40(6):2118–2126, 1969.

P. C. Fishburn. *The theory of social choice.* Princeton University Press, 1973.

P. C. Fishburn. Choice probabilities and choice functions. *Journal of Mathematical Psychology*, 18(3):205–219, 1978.

P. C. Fishburn. The axioms of subjective probability. *Statistical Science*, pages 335–345, 1986.

M. Frick, R. Iijima, and T. Strzalecki. Dynamic random utility. *Econometrica*, 87(6):1941–2002, 2019.

D. Fudenberg, R. Iijima, and T. Strzalecki. Stochastic choice and revealed perturbed utility. *Econometrica*, 83(6):2371–2409, 2015.

B. Gillen, E. Snowberg, and L. Yariv. Experimenting with measurement error: Techniques with applications to the caltech cohort study. *Journal of Political Economy*, 127(4):1826–1863, 2019.

F. Gul and W. Pesendorfer. Random expected utility. *Econometrica*, 74(1):121–146, 2006.

D. R. Insua. On the foundations of decision making under partial information. *Theory and decision*, 33(1):83–100, 1992.

C. H. Kraft, J. W. Pratt, and A. Seidenberg. Intuitive probability on finite sets. *The Annals of Mathematical Statistics*, 30(2):408–419, 1959.

J. Lu. Random choice and private information. *Econometrica*, 84(6):1983–2027, 2016.

R. D. Luce. *Individual choice behavior: A theoretical analysis*. 1959.

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

P. Manzini and M. Mariotti. Stochastic choice and consideration sets. *Econometrica*, 82(3): 1153–1176, 2014.

P. Manzini and M. Mariotti. Dual random utility maximisation. *Journal of Economic Theory*, 177:162–182, 2018.

P. Manzini, M. Mariotti, and H. Petri. Corrigendum to "Dual random utility maximisation"[J. Econ. Theory 177 (2018) 162–182]. *Journal of Economic Theory*, 184:104944, 2019.

E. A. Ok and G. Tserenjigmid. Indifference, indecisiveness, experimentation and stochastic choice. 2020.

E. Piermont and R. Teper. Disentangling strict and weak choice in random expected utility models. 2020.

D. Scott. Measurement structures and linear inequalities. *Journal of mathematical psychology*, 1(2):233–247, 1964.

H. A. Simon. A behavioral model of rational choice. *The quarterly journal of economics*, 69 (1):99–118, 1955.

J. Stoer and C. Witzgall. *Convexity and optimization in finite dimensions I*, volume 163. Springer Science & Business Media, 2012.

P. Suppes, D. H. Krantz, R. D. Luce, and A. Tversky. *Foundations of measurement: Geometrical, threshold, and probabilistic representations*, volume 2. 1989.

A. Tversky. Choice by elimination. *Journal of mathematical psychology*, 9(4):341–367, 1972.

H. R. Varian. Non-parametric analysis of optimizing behavior with measurement error. *Journal of Econometrics*, 30(1-2):445–458, 1985.