Assessing Misspecification and Aggregation for Structured Preferences*

Roy Allen
Department of Economics
University of Western Ontario
rallen46@uwo.ca

John Rehbeck
Department of Economics
The Ohio State University
rehbeck.7@osu.edu

June 21, 2019

Abstract

Applied research often tolerates misspecification in order to reach informative conclusions. We focus on how the degree of misspecification varies with the level of aggregation of data for quasilinear utility models. We present aggregation results formalizing that the model cannot get worse when aggregating. Using scanner data, we find that while *all* individuals are inconsistent with a quasilinear utility model, we cannot refute the hypothesis that a representative agent is a quasilinear utility maximizer. This provides evidence that deviations from a quasilinear model may average away.

*We thank Víctor Aguiar, Fernando Alvarez (discussant), Don Andrews, David Blau, Stephane Bonhomme, Christopher P. Chambers, Tim Conley, Ian Crawford, Paweł Dziewulski, Nail Kashaev, Yuichi Kitamura, Arthur Lewbel, Lance Lochner, Nirav Mehta, Matt Masten, Salvador Navarro, Alexandre Poirier, John Quah, Bernard Salanié, Andres Santos, Joel Sobel, conference participants at the Robustness in Economics and Econometrics conference at the University of Chicago, and seminar participants at Ohio State, Waterloo, and Western for helpful comments.

1 Introduction

> *But economics is not, in the end, much interested in the behavior of single individuals. Its concern is with the behavior of groups. A study of individual demand is only a means to the study of market demand.*
>
> — John Hicks *[1946]*

An extensive literature has studied whether general demand models are consistent with data. For example, Kitamura and Stoye [2018] are unable to refute utility maximization with general preference heterogeneity using repeated cross-sections. However, there is a tension between generality and certain pragmatic goals of applied research. Indeed, many applied papers impose stronger restrictions on preferences to reach stronger conclusions even though the preferences might be misspecified. A natural question is whether more-structured preferences could describe data while allowing a limited amount of misspecification.

In this paper, we study the consistency of quasilinear utility with data allowing misspecification/misoptimization. We focus on the relationship between individual and aggregate data.[1] A distinguishing feature of quasilinear utility is if individuals are exactly consistent with the model, then a representative agent is as well [Gorman, 1953]. We provide a conceptual and mathematical generalization of this: if individuals are approximately consistent, then aggregation will maintain this feature. The important possibility remains that data are (approximately) consistent with a representative agent but *not* with individuals. Indeed, Becker [1962] theoretically shows individual demands may be erratic, yet market demand may have economic regularities. We empirically investigate this possibility, finding support for quasilinear models to describe market demand of grocery store purchases.

To accomplish this, we provide a simple way to check whether individual or aggregate data are consistent with quasilinear utility while allowing up to ε dollars loss due to misspecification/misoptimization. We also propose a method of statistical inference for market demand that accounts for sampling variability. Specifically, we provide a

[1] Chetty [2012] studies a related problem involving parameter recoverability, misoptimization, and the level of aggregation.

statistical test of whether a representative agent (population mean demands) has a degree of misspecification no greater than a given value ε. The test involves linear moment inequality restrictions involving the mean of quantities in different time periods, which is a well-understood problem (e.g. Chernozhukov et al. [2014], Romano et al. [2014]).

In an empirical application using grocery store panel data on purchases, we examine the amount of misspecification needed to explain individual and aggregate data.[2] We find that all households are inconsistent with quasilinear utility in a deterministic framework. In contrast, we find that average demands are consistent with quasilinear utility using either a statistical or deterministic framework. Thus, while there is approximation error for every individual, the pattern of heterogeneity does not lead to a rejection of quasilinear utility in the aggregate. We thus reach a different conclusion from Russell and Thaler [1985], who hypothesize that systematic violations at the individual level would aggregate. Indeed, if demand curves are upward sloping *for everyone* this would be preserved under aggregation and would violate the model, but we do not find evidence of this. We note that in the empirical application there are many goods, yet linear programming facilitates quick computation.[3] Thus, the methods we present can readily be applied to other datasets.

We now discuss in more detail the focus on quasilinear utility and the results in the paper. First, this paper is focused on the relationship between misspecification, amount of modelling structure imposed, and the level of aggregation. While these concepts can be studied in other settings, we focus on quasilinear utility because it has been widely used, including in mechanism design [Myerson, 1981], discrete choice [McFadden, 1981], testing rational expectations [Browning, 1989], and testing the generalized axiom of revealed preference [Echenique et al., 2011].[4] It forms the foundation for some random coefficients models of demand in the spirit of Berry et al.

[2] The dataset is the Stanford Basket Data, which was previously used in Echenique et al. [2011].

[3] The number of inequalities to check in the linear program is the square of the number of time periods. Our empirical application has 375 goods and 26 time periods. For comparison, Kitamura and Stoye [2018] test the general model of utility maximization and reach the computational boundary with around 5 goods and 8 time periods.

[4] Echenique et al. [2018] provide a statistical test of the generalized axiom of revealed preference. In order to compute a critical value, they assume that the marginal utility of income is constant, stating, "It seems a reasonable assumption for our application to supermarket purchases." Constant marginal utility of income is observationally equivalent to quasilinear utility.

[1995].[5] A closely related structure also shows up for a large class of econometric models with observable characteristics, where an unknown utility index for each good plays the role of a (negative) price [Allen and Rehbeck, 2019a].

While quasilinear utility is prized for its tractability, it is usually viewed either explicitly or implicitly as an approximation.[6] The model assumes no income effects, which for many economists is "obviously" wrong. For some, however, general utility maximization is also "obviously" wrong. Rather than check whether data are *exactly* consistent with a theory, we provide a method to check whether the data are consistent with the theory, allowing a fixed level of misoptimization. If the data retute the model allowing a level of misspecification, we can still measure the minimum amount of misspecification needed.[7] Specifically, we calculate the smallest ε^* such that for some quasilinear utility function, all of an agent's choices are within ε^* dollars of the maximum utility possible.

While there are many ways to enlarge a baseline model, the approach we consider preserves certain important convexity properties of the original quasilinear model. Specifically, we show if individual behavior is (approximately) quasilinear, then aggregate behavior will also be (approximately) quasilinear (Proposition 3). The aggregation property for approximately quasilinear models has no such analogue for general preferences. This is because for demand models that allow general income effects and heterogeneity, aggregate demands may not have much structure even if *all* individuals act consistently with the model.[8]

This paper is meant to complement the standard paradigm in economics that posits a correctly specified model that includes individual shocks or measurement error. Instead the paper focuses on a framework that has misspecification built into the model.

[5] Equation (2.3) in Berry et al. [1995] describes a quasilinear demand system combined with characteristics entering the utility additively, while (2.7a) in that paper is not quasilinear in terms of prices.

[6] Weyl [forthcoming] discusses quasilinear utility as just one type of approximation used in economics. Weyl [forthcoming] highlights with notable exceptions (e.g. Willig [1976], Chetty [2012]), the approximation of quasilinear utility is not treated explicitly.

[7] We are not the first to advocate for this. See for example Hansen et al. [1995], Hansen and Jagannathan [1997], and Masten and Poirier [2018b].

[8] The Sonnenschein-Mantel-Debreu Theorem clarifies that average demands have only trivial restrictions if general income effects and heterogeneity are allowed. See Rizvi [2006] and references therein. This is not true when one studies the entire distribution of demand [McFadden and Richter, 1990, McFadden, 2005, Kitamura and Stoye, 2018]

One distinction is that "errors" arising due to misspecification/misoptimization are not necessarily welfare-relevant. This contrasts with standard application of latent utility models, where all shocks enter the indirect utility function [McFadden, 1981, Small and Rosen, 1981]. We further discuss in Section 3.2 how one could interpret misspecification when there are also stochastic components, according to the framework in this paper. In addition, the empirical question "What is an appropriate level of aggregation for a model?" is relevant since stochastic components are typically used to augment a model that *already* exhibits " 'laws' of behavior" [Haavelmo, 1944]. It is not *a priori* obvious at what level of aggregation such laws might hold.

Finally, we further contrast measuring misspecification with testing. A byproduct of our statistical analysis is that we can test whether aggregate demands are exactly rationalized by quasilinear utility, which is a novel contribution in a statistical framework. Even when we refute this hypothesis, we obtain an informative measure of the size of deviations from quasilinear utility by providing a confidence set for ε^*. Typically, when one rejects a model using a specification test, it is not obvious how to interpret how far away the particular specification is from the dataset.[9] In contrast, a lower one-sided confidence set provides a lower bound on the minimal misspecification needed to explain the data. The bound is in units of dollars per time period and is an economic measure of misspecification in the spirit of Varian [1990].[10]

The remainder of the paper proceeds as follows. Section 2 provides definitions, characterizes the model of approximate quasilinearity, and provides intuition on the behavior allowed in approximately quasilinear models. This section also gives an interpretation of approximation error as a measure of misspecification and describes computation of the measure. Section 3 shows that the approximation error of the representative agent is less than the average error of individuals, describes how our analysis relates to the general model of utility maximization, and discusses interpretation of the measure when there are also shocks. Section 4 details a statistical test for the representative agent and a method to construct confidence intervals. Section 5 provides an em-

[9] This is related to work on sensitivity analysis in econometric models. Recent work includes Conley et al. [2012], Kline and Santos [2013], Masten and Poirier [2018a], Andrews et al. [2017], Andrews et al. [2018], Bonhomme and Weidner [2018], Armstrong and Kolesár [2019], Christensen and Connault [2019].

[10] In a different setting, Hansen and Jagannathan [1997] provide a measure of misspecification and study large sample properties of it. Their measure is related to misspricing and can also be interpreted in units of dollars.

pirical assessment of quasilinear utility for individuals and the representative agent. Section 6 conducts simulations to better understand the results of the empirical analysis. Section 7 provides a review of the related literature. Section 8 contains our final remarks.

2 Definitions and Model

We consider a notion of approximate rationality for the consumer problem when an "individual" has quasilinear utility. In this paper we interpret "individual" as either a single person, a household, or a sum of individual demands. Quasilinear utility specifies that an individual values the consumption bundle $(x, y) \in \mathbb{R}^K_+ \times \mathbb{R}$ using the function $u(x) + y$ where $u : \mathbb{R}^K_+ \to \mathbb{R}$ is a locally non-satiated function and y is interpreted as the numeraire good.[11] A numeraire good is one which has a price of one. This means utility is measured in dollars per time period.

With quasilinear utility, maximization of the utility function is well-defined when an individual faces prices $p \in \mathbb{R}^K_{++}$. A consumer facing prices $p \in \mathbb{R}^K_{++}$ and with income $I \in \mathbb{R}$ solves

$$\max_{x \in \mathbb{R}^K_+, y \in \mathbb{R}} \quad u(x) + y \quad \iff \quad \max_{x \in \mathbb{R}^K_+} \quad u(x) + I - p \cdot x$$
$$\text{s.t. } p \cdot x + y \leqslant I.$$

We allow y to be negative (e.g. borrowing), otherwise this equivalence may not hold for low levels of income. We consider datasets of the form $\{(x^t, p^t)\}_{t=1}^T$ where each $x^t \in \mathbb{R}^K_+$, $p^t \in \mathbb{R}^K_{++}$, and T is an integer greater than or equal to one. We treat consumption x^t as an abstract object for the theoretical analysis of this section, which can accommodate several distinct settings. Each observation x^t may be interpreted as the quantities chosen at the specified prices p^t for a single individual. Alternatively, x^t may be interpreted as the sum or average of individual demands. Finally, in a statistical framework one may interpret x^t as (population) mean demands at prices p^t.

[11] We use $\mathbb{R}_+ = \{x \in \mathbb{R} \mid x \geqslant 0\}$ and $\mathbb{R}_{++} = \{x \in \mathbb{R} \mid x > 0\}$.

We differ from previous work by studying when demand data are approximately quasilinear. In particular, we say a model is approximately quasilinear if the observed demand data are within ε dollars of a quasilinear utility maximizer. This relaxation of quasilinear utility allows income effects and other violations of quasilinear utility as long as they are less than the prespecified amount ε. We provide a formal definition of when data are ε-rationalized by a quasilinear utility model.

Definition 1. *A dataset $\{(x^t, p^t)\}_{t=1}^T$ is ε-rationalized by quasilinear utility for $\varepsilon \geqslant 0$ if there exists a locally non-satiated utility function $u : \mathbb{R}_+^K \to \mathbb{R}$ such that for all $t \in \{1, \ldots, T\}$ and for all $x \in \mathbb{R}_+^K$, the following inequality holds:*

$$u(x^t) - p^t \cdot x^t + \varepsilon \geqslant u(x) - p^t \cdot x.$$

We also refer to the above by saying a dataset is ε-quasilinear rationalized. When ε equals zero, it is convenient to say the dataset is quasilinear rationalized.

It is helpful to contrast this with additive random utility models, which include an additive disturbance to the desirability of each possible choice (e.g. McFadden [1981], Dagsvik [1994]). Here, while ε is an additive disturbance, it only shows up on one side of the inequality and is a disturbance to optimization. In addition, it is a single value across all decision problems and is *not* random.[12] We note that McFadden [1981] states additive disturbances in the additive random utility model may be interpreted as "errors in judgment." This paper provides a different formalism of this idea. We further discuss how the value ε is related to utility shocks in Section 3.2.

Before characterizing ε-quasilinearity, it is useful to understand how this notion of approximate rationality relates to standard results on quasilinear utility. Recall that if an individual has quasilinear utility then they satisfy the law of demand for prices. Similarly, if an individual is ε-rationalized by quasilinear utility, then they satisfy the ε-law of demand.

Proposition 1 (Approximate Law of Demand)**.** *If the dataset $\{(x^t, p^t)\}_{t=1}^T$ is ε-rationalized by quasilinear utility, then for any $r, s \in \{1, \ldots, T\}$ it follows that*

$$\frac{1}{2}(p^s - p^r) \cdot (x^s - x^r) \leqslant \varepsilon.$$

[12] It is straightforward to extend the results in this paper to the case where ε is time-specific. See Appendix C.

To see the restrictions this places on data, consider univariate demand ($K = 1$), where we use the notation x to represent the quantity demanded of the good that is not the numeraire. Consider demand restrictions about a point $(\tilde{x}, \tilde{p})$ in Figure 1. When ε equals zero, this requires that demand be downward sloping through $(\tilde{x}, \tilde{p})$. This is illustrated by the dark area in Figure 1. However, when $\varepsilon > 0$, one could observe consumption of x that increases as its price increases. Thus, the law of demand must only hold approximately.

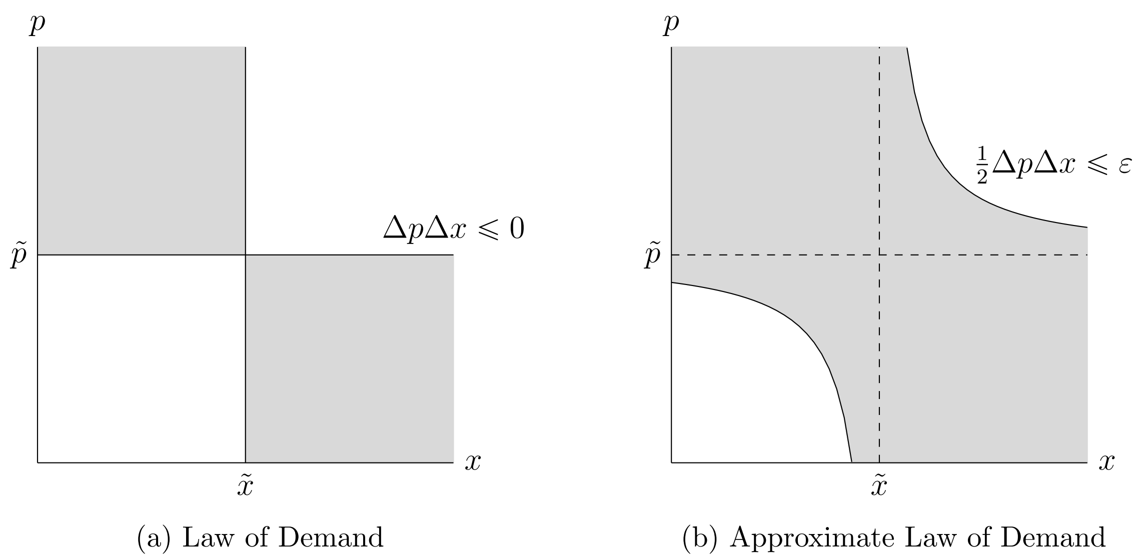

Figure 1: Versions of the Law of Demand

Perhaps surprisingly, a small ε can allow a great deal of flexibility for small price changes due to the hyperbolic nature of the approximate law of demand. Indeed, even when ε is small, a small increase in price could allow a large *increase* in quantities. This theoretically-motivated enlargement of the set in Figure 1(a) is thus distinct from alternative topological enlargements. We now provide a complete characterization when a dataset is ε-rationalized by quasilinear utility.

Theorem 1. *For any dataset* $\{(x^t, p^t)\}_{t=1}^T$ *and* $\varepsilon \geqslant 0$, *the following are equivalent:*

(i) $\{(x^t, p^t)\}_{t=1}^T$ *is* ε*-rationalized by quasilinear utility.*

(ii) There exist numbers $\{u^t\}_{t=1}^T$ *that satisfy the following inequalities for all* $r, s \in \{1, \ldots, T\}$*:*

$$u^s \leqslant u^r + p^r \cdot (x^s - x^r) + \varepsilon.$$

(iii) For all finite sequences $\{t_m\}_{m=1}^M$ with $t_m \in \{1, \ldots, T\}$ and $M \geqslant 2$, the inequality

$$\frac{1}{M}\sum_{m=1}^{M} p^{t_m} \cdot (x^{t_m} - x^{t_{m+1}}) \leqslant \varepsilon$$

holds, where $(x^{t_{M+1}}, p^{t_{M+1}}) = (x^{t_1}, p^{t_1})$.

(iv) $\{(x^t, p^t)\}_{t=1}^T$ is ε-rationalized by a quasilinear utility function that is continuous, monotonic increasing, and concave.

This is a generalization of Browning [1989] and Brown and Calsamiglia [2007]), who study the case $\varepsilon = 0$ and impose concavity. Part (ii) resembles the Afriat inequalities, except it allows model approximation error of ε. For computational purposes, part (ii) is the most useful. Indeed, one can check (ii) by a linear program with T variables (the u^t), and T^2 inequality constraints. The inequality of (iii) is a requirement that the average money extracted from a "money pump" be less than ε.[13] The equivalence between (i) and (iv) shows that continuity, monotonicity, and concavity place no additional empirical restrictions on data. We note that there is always some ε that will rationalize the model. In particular, $\varepsilon = \max_{t \in \{1,\ldots,T\}}\{p^t \cdot x^t\}$ suffices. However, there may be values $\varepsilon > 0$ such that a dataset cannot be described by the model. One example is presented below.

Example 1. *Consider the dataset with $(q^1, p^1) = (2, 1)$ and $(q^2, p^2) = (6, 2)$. We check whether $\varepsilon = 1$ can ε-quasilinear rationalize the data so that the average "money pump" must be less than one dollar per time period. We obtain*

$$\frac{1}{2}\left(p^1(q^1 - q^2) + p^2(q^2 - q^1)\right) = 2 > 1 = \varepsilon.$$

Thus, this dataset cannot be ε-rationalized by quasilinear utility when $\varepsilon = 1$.

2.1 Measure of Misspecification

The smallest ε that ε-quasilinear rationalizes the data are a natural measure of misspecification. We denote this as ε^* and call it the measure of quasilinear misspeci-

[13] A "money pump" assumes the existence of an arbitrager who will buy goods and re-sell them to the consumer for a profit. Each difference $p^{t_m} \cdot (x^{t_m} - x^{t_{m+1}})$ in Theorem 1 represents one such trade. This idea is explored in the model with income effects in Echenique et al. [2011].

fication. We interpret it as the least approximation error induced by a quasilinear model. However, there are many other ways to interpret this measure. It could be interpreted as the least average money extracted through a "money pump" (Echenique et al. [2011]). Alternatively, one can think of this measure as an additively separable version of the Afriat Efficiency Index (AEI) (Afriat [1973]) for quasilinear utility. Lastly, one could interpret this measure as measuring the width of thick indifference curves for an individual with approximately quasilinear utility (cf. Dziewulski [2018]).

The following result describes two equivalent ways to compute this measure. The first way involves unknown utility numbers and is a tractable linear program. The second condition only involves the observables, and motivates the approach to statistical inference described in Section 4.

Proposition 2. *Given a dataset $\{(x^t, p^t)\}_{t=1}^T$, there exists a smallest non-negative value ε^* such that the dataset is ε-quasilinear rationalized. Moreover, for all $\varepsilon \geqslant \varepsilon^*$, the dataset is ε-quasilinear rationalized. This value may be computed by either of the following equivalent linear programs:*

$$\min_{\substack{\varepsilon \in \mathbb{R}_+ \\ u^1,\ldots,u^T \in \mathbb{R}_+}} \varepsilon \quad s.t. \quad u^s \leqslant u^r + p^r \cdot (x^s - x^r) + \varepsilon \quad \textit{for all } r, s \in \{1, \ldots, T\}, \tag{1}$$

$$\min_{\varepsilon \in \mathbb{R}_+} \varepsilon \quad s.t. \quad \frac{1}{M} \sum_{m=1}^{M} p^{t_m} \cdot (x^{t_m} - x^{t_{m+1}}) \leqslant \varepsilon, \tag{2}$$

where the minimum of (2) is taken with respect to all finite sequences $\{t_m\}_{m=1}^M$ with $t_m \in \{1, \ldots, T\}$, $M \geqslant 2$, and $(x^{t_{M+1}}, p^{t_{M+1}}) = (x^{t_1}, p^{t_1})$.

This proposition follows essentially from Theorem 1(ii)-(iii).[14] We highlight that this measure differs from the money pump index of Echenique et al. [2011] because it does not scale by expenditure of the cycle and does not restrict attention to violations of the generalized axiom of revealed preference.

[14] One straightforward generalization of our setup is to allow a separate wedge for each time period, as in Varian [1990]. If we consider a separate ε_V^t for each time period t and consider the vector $\varepsilon_V = (\varepsilon_V^1, \ldots, \varepsilon_V^T)$, then other objective functions such as the mean $\frac{1}{T}\sum_{t=1}^T \varepsilon_V^t$ are also computable by linear programming. See Theorem 2 in Appendix C.

3 Aggregation

A classic result is that quasilinear utility induces demand that can be aggregated across individuals.[15] We show a similar aggregation property holds when each individual dataset is ε^i-quasilinear rationalized and individuals face the same prices. These results illustrate that if a given level of aggregation is "acceptable" in the sense that the misspecification is bounded by some prespecified value, then aggregating any further will maintain this feature. A special case of this is the classic version stating that if individuals are exactly consistent with quasilinear utility, then aggregation will preserve this.

To formalize this, suppose that there are $i = 1, \dots, n$ individuals, where each individual has a dataset $\{(x^{(i,t)}, p^t)\}_{t=1}^T$. Let $\overline{x}^t = \frac{1}{n}\sum_{i=1}^n x^{(i,t)}$ denote the average demand at the price p^t. We now define the *aggregate dataset* as $\{(\bar{x}^t, p^t)\}_{t=1}^T$. We make use of Theorem 1(ii) to show how the measure of misspecification behaves under aggregation.

Proposition 3. *(i) If each individual dataset* $\{(x^{(i,t)}, p^t)\}_{t=1}^T$ *is* ε^i*-rationalized by quasilinear utility, then the* aggregate dataset $\{(\bar{x}^t, p^t)\}_{t=1}^T$ *is* $\overline{\varepsilon}$*-rationalized by quasilinear utility, where* $\overline{\varepsilon} = \frac{1}{n}\sum_{i=1}^n \varepsilon^i$.

(ii) The measures of misspecification satisfy

$$\overline{\varepsilon}^* \leqslant \frac{1}{n}\sum_{i=1}^n \varepsilon^{i,*}, \tag{3}$$

where $\overline{\varepsilon}^*$ *is the measure of misspecification for the aggregate dataset and* $\varepsilon^{i,*}$ *is the measure of misspecification for the dataset for individual* i.

Proposition 3(i) provides an alternative way to show the classical result that quasilinear utility is preserved under aggregation. Indeed, if all individuals are exactly quasilinear rationalizable ($\varepsilon^{i,*} = 0$) then $\overline{\varepsilon}^* = 0$, i.e. the representative agent is exactly quasilinear rationalizable. It also describes linear aggregation properties of the class of ε-quasilinear models. Proposition 3(ii) shows that the measure of misspecification *weakly* decreases with aggregation, but does not say how much. If all individual data sets are identical, then (3) holds with equality. The magnitude of the gap in (3) depends on the particular distribution of individual demands. We investigate this

[15] See e.g. Varian [1992], Section 10.6 for a textbook reference.

gap in our application.

We illustrate (3) in a simple example with one good and a dataset consisting of two prices.[16] Since there are only two prices, we need only check a single cycle. We obtain

$$\varepsilon^{i,*} = \max\left\{0, \frac{1}{2}\Delta p \Delta x^i\right\},$$

where Δ takes a difference. We consider three cases: (1) all individuals have downward sloping demands; (2) all individuals have upward sloping demands; and (3) some individuals have upward sloping demands, and some have downward sloping demands. In cases (1) or (2) we have

$$\overline{\varepsilon}^* = \frac{1}{n}\sum_{i=1}^{n} \varepsilon^{i,*}.$$

However, in case (3) we instead have

$$\overline{\varepsilon}^* < \frac{1}{n}\sum_{i=1}^{n} \varepsilon^{i,*}.$$

Thus, the measure of misspecification strictly decreases upon aggregation provided some individuals are consistent with downward sloping demands, and some individuals are not. We illustrate this further in the following example, in which average demand data are rationalized by quasilinear utility even when there are individuals whose datasets cannot be exactly rationalized.

Example 2. *Consider univariate demand* ($K = 1$) *with two individuals. Individual one has a dataset with* $(x^{(1,1)}, p^1) = (2, 1)$ *and* $(x^{(1,2)}, p^2) = (6, 2)$ *and individual two has a dataset with* $(x^{(2,1)}, p^1) = (6, 1)$ *and* $(x^{(2,2)}, p^2) = (2, 2)$*. The aggregate dataset is given by* $(\bar{x}^1, p^1) = (4, 1)$ *and* $(\bar{x}^2, p^2) = (4, 2)$*. The first individual has minimal* $\varepsilon^{1*} = 2$*. The second individual has minimal* $\varepsilon^{2*} = 0$*. However, the aggregated demand is also rationalized by quasilinear utility since*

$$1(4 - 4) + 2(4 - 4) = 0.$$

We provide a probabilistic counterpart of Proposition 3 to lay the foundation for

[16] We thank Fernando Alvarez for this example.

our statistical analysis in Section 4. To state the result, we now consider $X^{(i,t)}$ as a random vector of quantities for K goods for individual i at observation t. We stack quantities in the vector $X^i = \left(X^{(i,1)\prime}, \ldots, X^{(i,T)\prime}\right)'$. We let p^t be a predetermined (nonrandom) vector of prices for the K goods at observation t. Fixing prices, we may then consider the mapping $\varepsilon^*(X^i)$ that finds the smallest ε such that the dataset $\{(X^{(i,t)}, p^t)\}_{t=1}^T$ is ε-rationalized by quasilinear utility.

Proposition 4. *Assume X^i is identically distributed and $\mathbb{E}[X^i]$ exists.*

(i) $\{(\mathbb{E}[X^{(i,t)}], p^t)\}_{t=1}^T$ is $\overline{\varepsilon}$-rationalized by quasilinear utility, where $\overline{\varepsilon} = \mathbb{E}[\varepsilon^*(X^i)]$.

(ii) The mapping $\varepsilon^* : \mathbb{R}^{K \times T} \to \mathbb{R}_+$ is convex. In particular,

$$\varepsilon^*(\mathbb{E}[X^i]) \leqslant \mathbb{E}[\varepsilon^*(X^i)].$$

Proposition 4(i) is a probabilistic counterpart to Proposition 3 over the minimal $\varepsilon^*(X^i)$. A similar result holds for any random variable ε^i that satisfies $\varepsilon^i \geqslant \varepsilon^*(X^i)$ almost surely. The inequality in Proposition 4(ii) follows from convexity and Jensen's inequality.

We can further interpret these aggregation results in the context of a latent utility model. To that end, suppose individual i satisfies

$$u(X^{(i,t)}, \eta^i) - p^t \cdot X^{(i,t)} + \varepsilon^i \geqslant \sup_{x \in \mathbb{R}_+^K} u(x, \eta^i) - p^t \cdot x \qquad \forall t \in \{1, \ldots, T\}, \tag{4}$$

where η^i represents unobservable heterogeneity in tastes. Then the dataset $\{(X^i, p^t)\}_{t=1}^T$ is ε^i-rationalized by quasilinear utility. Proposition 4 shows that data generated in this manner will satisfy

$$\overline{u}(\mathbb{E}[X^{(i,t)}]) - p^t \cdot \mathbb{E}[X^{(i,t)}] + \mathbb{E}[\varepsilon^i] \geqslant \sup_{x \in \mathbb{R}_+^K} \overline{u}(x) - p^t \cdot x \qquad \forall t \in \{1, \ldots, T\}$$

for some utility function $\overline{u}$.

3.1 Relation to General Utility Functions

Proposition 3 is an aggregation result that holds because ε^* is convex when viewed as a function of the quantities, fixing prices. Thus, it is a measurement result that does not rely on specific assumptions concerning how the data are generated. We now relate our aggregation result to the more general model of utility maximization subject to a budget constraint, without the assumption that utility is also quasilinear.

First, we recall that if we aggregate individuals who each satisfy the general model of utility maximization, then the average demands may not be consistent with the general model. Aggregation for utility maximization is known to require more structure [Gorman, 1953]. Thus, without more structure the previous aggregation results have no analogues for the general model.

Becker [1962] provides a theoretical result that flips this: even "irrational" individuals may aggregate in a manner that is consistent with general utility maximization.[17] Becker's analysis relies on the presence of the budget constraint in an essential way. A version of Becker's result holds for quasilinear utility, which can be written without a budget constraint, but it is trivial. Suppose $X^{(i,t)}$ is independent and identically distributed across time and individuals, with prices treated as nonrandom parameters. In other words, choices are unrelated to prices. This means the vector of mean demands $\mathbb{E}[X^{(i,t)}]$ is constant across time. From Proposition 1(iii) it is easy to see that with this data generating process, $\{(\mathbb{E}[X^{(i,t)}], p^t)\}_{t=1}^T$ is consistent with quasilinear utility, where we interpret $\mathbb{E}[X^{(i,t)}]$ as a population-level mean demand.

This example is close to Becker's in the sense that we integrate to form a (nonrandom) population mean. A natural question is whether a similar result might hold when there is sampling variability. The following result suggests the answer is no, suggesting instead that rationalizations for *quasilinear utility* are unlikely if individuals are choosing in a manner unrelated to prices. To state the result, consider a sample of size n and let $\overline{X}^t = \frac{1}{n}\sum_{i=1}^n X^{(i,t)}$ be the vector of average demands for the sample for each time period t.

Proposition 5. *Treat prices as nonrandom parameters and assume $p^r \neq p^s$ for $r \neq s$. Assume $X^{(i,t)}$ has a density with respect to Lebesgue measure and is independent and*

[17] See also Grandmont [1992], Kirman [1992], Birchenall [2016], and references therein.

identically distributed across time and individuals. The probability $\left\{\left(\overline{X}^t, p^t\right)\right\}_{t=1}^T$ *is rationalized by quasilinear utility is given by*

$$P\left(\varepsilon^*\left(\left\{\overline{X}^t\right\}_{t=1}^T\right) = 0\right) = \frac{1}{T!}.$$

In contrast, $\{(\mathbb{E}[X^{(i,t)}], p^t)\}_{t=1}^T$ *is exactly consistent with quasilinear utility.*

We are not aware of a related analytical result for general utility maximization; usually the probability of a dataset satisfying the restrictions of general utility maximization is estimated by simulation [Bronars, 1987]. A noteworthy feature of Proposition 5 is that the sample size n does not alter the probability of a rationalization. Thus, in finite samples aggregation need not "mechanically" alter the probability of a rationalization. This clarifies that our previous results (e.g. Proposition 3) do *not* imply that we should necessarily expect a representative agent to be consistent with quasilinear utility with an arbitrary dataset. We note that if $X^{(i,t)}$ has a discrete distribution, in general this probability needs to be adapted to depend on the number of individuals and the distribution of quantities.

3.2 Shocks

This paper studies the measurement of misspecification relative to a baseline deterministic model without random preference shocks. In this section we describe how the measure of misspecification in this paper can be interpreted when there are random preference shocks *and* individuals are not exact optimizers. First, however, we emphasize there is a core difference in interpretation between the present framework and a typical shocks framework. Following McFadden [1981] and Small and Rosen [1981], standard welfare analysis of latent utility models interprets all shocks as structural, welfare-relevant unobservables. In contrast, deviations from a baseline model that are due to misspecification/misoptimization need not be welfare-relevant.

We now turn to analysis of the measure of misspecification when there are shocks. To formalize this, let $\eta^{(i,t)}$ be an individual-time specific shock, which enters the utility function. Similar to Proposition 1, we obtain an approximate law of demand, but now with shocks:

$$\frac{1}{2}(p^s - p^r) \cdot (X^{(i,s)} - X^{(i,r)}) \leqslant \varepsilon^i + \frac{u(X^{(i,s)}, \eta^{(i,s)}) - u(X^{(i,s)}, \eta^{(i,r)})}{2} + \frac{u(X^{(i,r)}, \eta^{(i,r)}) - u(X^{(i,r)}, \eta^{(i,s)})}{2}.$$

This bound relates the data, the degree of misspecification due to misoptimization (ε^i), and the unobservable shocks. We can repeat this argument with longer cycles, similar to the statement of Theorem 1(iii). For all finite sequences $\{t_m\}_{m=1}^M$, $t_m \in \{1, \ldots, T\}$ and $M \geqslant 2$, the inequality

$$\frac{1}{M}\sum_{m=1}^{M} p^{t_m} \cdot (X^{(i,t_m)} - X^{(i,t_{m+1})}) \leqslant \varepsilon^i + \frac{1}{M}\sum_{m=1}^{M} u(X^{(i,t_m)}, \eta^{(i,t_m)}) - u(X^{(i,t_m)}, \eta^{(i,t_{m+1})}) \tag{5}$$

holds, where $(X^{(i,t_{M+1})}, p^{t_{M+1}}) = (X^{(i,t_1)}, p^{t_1})$. The sum of utility terms differences out when $\eta^{(i,t)}$ does not vary, but in general is nonzero.

If we place bounds on the shocks, then we can relate $\varepsilon^{i,*}$ and ε^i.

Proposition 6. *Let*

$$\sup_{r,s \in \{1,\ldots,T\}} \left| u(X^{(i,r)}, \eta^{(i,r)}) - u(X^{(i,r)}, \eta^{(i,s)}) \right| \leqslant \delta^i.$$

It follows that

$$\varepsilon^{i,*} \leqslant \varepsilon^i + \delta^i.$$

This is immediate from (5) and Proposition 2. If we interpret ε^i as misspecification for a model allowing taste shocks, and we instead use $\varepsilon^{i,*}$ to measure misspecification (as in Proposition 2 without taste shocks), then $\varepsilon^{i,*}$ can be higher than ε^i. However, if δ^i is small then $\varepsilon^{i,*}$ cannot be much higher.

Note that the bound relating the measure of misspecification and shocks in Proposition 6 holds for every realization of the shocks. If the shocks are not systematically related to prices, we obtain another aggregation property when we take expectations.

Proposition 7. *Assume $(\eta^{(i,t)}, \varepsilon^i)$ has the same distribution for each individual and time period, $X^{(i,t)}$ is $(\eta^{(i,t)}, \varepsilon^i)$-measurable, and $\mathbb{E}[X^{(i,t)}]$ exists for each t. Then $\{(\mathbb{E}[X^{(i,t)}], p^t)\}_{t=1}^T$ is $\mathbb{E}[\varepsilon^i]$-rationalized by quasilinear utility.*

Note that Proposition 7 holds regardless of the relationship between ε^i and the vector of taste shocks $(\eta^{(i,1)}, \ldots, \eta^{(i,T)})$. In addition, we do not need to assume taste shocks are independent across time for the result to hold.

This result provides a synthesis of a shocks and a misspecification/misoptimization approach. It differs from the previous aggregation results, which allow arbitrary individual heterogeneity provided it is fixed across time at the individual level. Here, even at the individual level the values $\eta^{i,t}$ can change with time.

To interpret this result, consider first the case $\varepsilon^i = 0$ for each individual. We may find that if we calculate the measure of misspecification for each individual we obtain $\varepsilon^{i,*} > 0$, since this measure does not accomodate shocks. However, if the distribution of shocks is the same across time periods, the model admits a representative agent in the sense that mean demands exactly maximize a quasilinear utility function.[18] Thus, even if there are shocks that do not "cancel out" at the individual level, they can cancel out upon aggregating to population mean demands. Proposition 7 is a generalization allowing ε^i to be nonzero. It shows that misspecification cannot get worse upon forming expectations, provided the distrubtion of taste shocks is the same across time.

4 Statistical Inference

Proposition 5 describes a class of data generating processes in which the population mean quantities are consistent with quasilinear utility, but it is unlikely the sample averages are exactly consistent with the model. Thus, there can be stark differences between deterministic and statistical tests of quasilinear utility. We note that a number of deterministic tests have been proposed in the nonparametric revealed preference literature, beginning with the seminal work of Samuelson [1938], Houthakker [1950], and Afriat [1967]. There has been less work on statistical inference in such settings. Notable exceptions that study measurement error include Varian [1985], Echenique et al. [2011], and Aguiar and Kashaev [2018]. This paper instead considers inference when randomness occurs to sampling variability, and is thus conceptually closer to

[18]Related aggregation results have appeared in McFadden [1981], Hofbauer and Sandholm [2002], and Allen and Rehbeck [2019a,b]. Here we differ because we allow misoptimization ($\varepsilon^i > 0$).

Kitamura and Stoye [2018].

We first describe how to test whether the representative agent is an ε-quasilinear maximizer. We differ conceptually from the previous literature by conducting inference on an approximate model, without restricting attention to the exact case ($\varepsilon = 0$). Then we describe how to construct a confidence interval for the smallest ε^* such that the representative agent is an ε-quasilinear maximizer.

In our application we have panel data, so we provide a statistical test for such data. It is straightforward to adapt these ideas to repeated cross-sections. Each observation consists of the pair $(X^{(i,t)}, p^t)$, where i denotes the individual and t denotes the time period. The vector p^t is the price vector at time t, which is common across individuals, and is treated as predetermined. The vector $X^{(i,t)}$ encodes the quantities of K goods for individual i at time t, which is treated as a random variable.

Sampling uncertainty arises because we interpret the particular sample of individuals as a random draw from a population. We allow arbitrary statistical dependence across time for each person, but require independence between individuals. We formalize this as follows.

Assumption 1. *The quantity vector for individual i, $X^i = \left(X^{(i,1)\prime}, \ldots, X^{(i,T)\prime}\right)'$, is independent and identically distributed across individuals.*

Recall we treat prices as predetermined. To relate this to our previous setup, one may interpret $\mathbb{E}[X^{(i,t)}]$ as the demand vector for a representative agent at price vector p^t. The representative agent population-level dataset is then given by $\{(\mathbb{E}[X^{(i,t)}], p^t)\}_{t=1}^T$. Our null hypothesis is formulated by using condition (iii) in Theorem 1, applied to this representive agent dataset. In particular, we formalize the statement that a representative agent is ε-rationalized by a quasilinear utility function for a given $\varepsilon \geqslant 0$ by the following null hypothesis:

$$H_0 : \text{For all finite sequences } \{t_m\}_{m=1}^M \text{ with } t_m \in \{1, \ldots, T\} \text{ and with } M \geqslant 2,$$
$$\frac{1}{M} \sum_{m=1}^{M} p^{t_m} \cdot \left(\mathbb{E}\left[X^{(i,t_m)}\right] - \mathbb{E}\left[X^{(i,t_{m+1})}\right]\right) \leqslant \varepsilon.$$

Our preferred interpretation of this null hypothesis is that one directly hypothesizes that the representative agent is ε-quasilinear rationalized for a pre-specified value of ε. Alternatively, one can view this as an implication of the assumption that *all* individuals act consistently with ε-quasilinearity. This null hypothesis arises as an implication due to Proposition 4.[19] Another interpretation is that there are utility shocks at the individual level as in Proposition 7. Then the inequalities in the null arise if the mean of the individual level measures of misspecification is no greater than ε.

Making use of the fact that individuals each face the same prices, we may test the null hypothesis by drawing on the literature on testing finitely many unconditional moment inequalities. To test H_0, we follow the methodology of Chernozhukov et al. [2014]. We note that our setup has *more* structure than general moment inequalities problems. The reason is that the null hypothesis can be written as a set of linear inequality restrictions on means of the quantities, $\mathbb{E}[X^i] = (\mathbb{E}[X^{(i,1)}]', \ldots, \mathbb{E}[X^{(i,T)}]')'$. We use this fact for numerical purposes, but present the testing procedure without exploiting this structure to better connect it to the existing literature.

To describe the test, we introduce some additional notation. First, let J denote the number of distinct sequences $\{t_m\}_{m=1}^M$ in the sense of Theorem 1(iii). Thus, under H_0 we have J moment inequalities, each of the form

$$\mathbb{E}\left[\frac{1}{M}\sum_{m=1}^{M} p^{t_m} \cdot \left(X^{(i,t_m)} - X^{(i,t_{m+1})}\right)\right] \leqslant \varepsilon,$$

where $(X^{(i,t_{M+1})}, p^{t_{M+1}}) = (X^{i,t_1}, p^{t_1})$. We index each such sequence by $j \in \{1, \ldots, J\}$. Associated with each moment condition, we form a sample average across individuals,

$$\begin{aligned} \hat{\mu}_j &= \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{M}\sum_{m=1}^{M} p^{t_m} \cdot \left(X^{(i,t_m)} - X^{(i,t_{m+1})}\right)\right] \\ &= \left[\frac{1}{M}\sum_{m=1}^{M} p^{t_m} \cdot \left(\overline{X}^{t_m} - \overline{X}^{t_{m+1}}\right)\right], \end{aligned} \tag{6}$$

where $\overline{X}^t = \frac{1}{n}\sum_{i=1}^{n} X^{(i,t)}$ is the K-dimensional vector of average demands in period t.

[19] With panel data, one can directly calculate each individual measure of misspecification and check whether $\varepsilon^*(X^i) \leqslant \varepsilon$ almost surely. We note that with repeated cross-sections, this is not possible and an analyst may use the moment conditions in H_0 for such a setting.

Recall that $X^i = \left(X^{(i,1)\prime}, \ldots, X^{(i,T)\prime}\right)^\prime$ denotes the demands for individual i. Let $X = (X^{1\prime}, \ldots, X^{n\prime})^\prime$ collect the quantities across all n individuals. The test statistic is given by

$$S(X) := \max_{j\in\{1,\ldots,J\}} \hat{\mu}_j - \varepsilon = \max_{\{t_m\}_{m=1}^M} \left[\frac{1}{M}\sum_{m=1}^{M} p^{t_m} \cdot \left(\overline{X}^{t_m} - \overline{X}^{t_{m+1}}\right)\right] - \varepsilon$$

where each j indexes a sequence. The statistic S may thus be computed as a linear programming problem with T^2 inequality constraints using Proposition 2.[20] This is true despite the fact that the number of cycles ($J = \sum_{\ell=2}^{T} \frac{T!}{(T-\ell)!\ell}$) itself grows quickly with T. Note that the term in brackets is $\varepsilon^*\left(\left\{\overline{X}^t\right\}_{t=1}^T\right)$ whenever it is non-negative.

We reject the null hypothesis at nominal size α if

$$S(X) > c_{1-\alpha},$$

where $c_{1-\alpha}$ is the $(1-\alpha)$-quantile of S from a bootstrap distribution with B draws. In the application, we use $B = 5{,}000$. Appendix A.1 provides intuition for the validity of this test. The critical value $c_{1-\alpha}$ is constructed in the following manner:

i. Draw an independent and identically distributed sample of size n uniformly from the individual demands $\{X^i\}_{i=1}^n$. Recall $X^i = \left(X^{(i,1)\prime}, \ldots, X^{(i,T)\prime}\right)^\prime$ denotes the consumption vector across all time periods. Let $X^{*i} = \left(X^{*(i,1)\prime}, \ldots, X^{*(i,T)\prime}\right)^\prime$ denote quantities for individual $*i$ in the bootstrap sample.[21]

ii. Compute the sample mean of the moment conditions from the bootstrap draw via Equation 6, denoted $\hat{\mu}^* = (\hat{\mu}_1^*, \ldots, \hat{\mu}_J^*)$. Specifically, each $\hat{\mu}_j^*$ is given by

$$\hat{\mu}_j^* = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{M}\sum_{m=1}^{M} p^{t_m} \cdot \left(X^{*(i,t_m)} - X^{*(i,t_{m+1})}\right)\right].$$

iii. Repeat (i) and (ii) a total of B times.

[20] One can apply the first linear program in that proposition to the dataset $\left\{\left(\overline{X}^t, p^t\right)\right\}_{t=1}^T$, omitting the inequality constraint on the degree of misspecification, to calculate $S(X)$.

[21] Following convention, we use $*$ here but note that this usage is distinct from the measure of rationality.

iv. Define

$$c_{1-\alpha} = \inf \left\{ c \mid P^* \left(\max_{j \in \{1,\ldots,J\}} (\hat{\mu}_j^* - \hat{\mu}_j) \geqslant c \right) \leqslant \alpha \right\},$$

where P^* is the simulated distribution of B bootstrap draws, described by steps (i)-(iii).

Recall that after rearranging,

$$\hat{\mu}_j^* - \hat{\mu}_j = \left[\frac{1}{M} \sum_{m=1}^{M} p^{t_m} \cdot \left(\left(\overline{X}^{*t_m} - \overline{X}^{t_m} \right) - \left(\overline{X}^{*t_{m+1}} - \overline{X}^{t_{m+1}} \right) \right) \right],$$

and so computation of $\max_{j \in \{1,\ldots,J\}} (\hat{\mu}_j^* - \hat{\mu}_j)$ in step (iv) is again facilitated by Proposition 2. One simply applies the linear program in Proposition 2 to the dataset $\left\{ \left(\overline{X}^{*t} - \overline{X}^{t}, p^t \right) \right\}_{t=1}^{T}$, except one drops the inequality restriction $\varepsilon \geqslant 0$ in the linear program. In addition, note that ε does not show up in the calculation of the critical value. Thus, the critical value may be computed once, even if the analyst wishes to test multiple values of ε or construct a confidence interval via test inversion.

Remark 1 (Studentization)*.* An alternative procedure would consider the studentized test statistic

$$\tilde{S}(X) = \max_{j \in \{1,\ldots,J\}} \frac{\hat{\mu}_j - \varepsilon}{\hat{\sigma}_j}$$

(or a version multiplied by $\sqrt{n}$), where

$$\hat{\sigma}_j = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (g_j(X^i) - \hat{\mu}_j)^2}$$

and

$$g_j(X^i) = \frac{1}{M} \sum_{m=1}^{M} p^{t_m} \cdot \left(X^{(i,t_m)} - X^{(i,t_{m+1})} \right).$$

We do not studentize the test statistic for interpretation, computational reasons, and to direct power against certain alternatives.[22] Multiplying by $\sqrt{n}$ or dividing by $\hat{\sigma}_j$ means we can no longer interpret the statistic as measuring dollars lost. In light of Proposition 2, the unstudentized statistic is easy to compute. In contrast, Proposition 2 cannot directly be used to simplify computation of a studentized statistic. To

[22]The unstudentized test statistic is not formally covered by Chernozhukov et al. [2014]. For a general result on bootstrap validity that applies to S, see Proposition 4.2 in Chernozhukov et al. [2017].

understand the power differences between S and the studentized counterpart, note that the variance of $\hat{\mu}_j$ will typically be smaller for sequences of longer length. This is because $\hat{\mu}_j$ is formed as an average over M terms, where M is the sequence length. Thus we anticipate that the unstudentized statistic S directs power toward violations of the moment inequalities that occur for shorter sequences. This choice is consistent with a common intuition in the revealed preference literature that for theories that are characterized by acyclicity conditions, violations typically occur for sequences of shorter length.[23]

Remark 2 (Moment Selection). Refinements of the test are possible by applying a "moment selection" approach, which originates in the literature on testing moment inequalities, e.g. Andrews and Soares [2010], Chernozhukov et al. [2014].[24] This literature has focused on models with no particular relationship between moment conditions, and directly applying the ideas would involve dropping certain cycles j such that $\hat{\mu}_j$ is sufficiently negative, i.e. consistent with the null hypothesis.[25] An alternative approach would be to drop values t such that $\overline{X}^t$ is sufficiently "far" from altering the test statistic $S(X)$. This would amount to dropping all cycles j with the t-th time period in them, and thus provides an alternative methodology to the moment inequalities literature. Thus, this alternative approach involves dropping groups of moment conditions using the fact that there is a theoretical relationship between these means. Dropping certain moments could lead to computational issues since one would need to account for all $J = \sum_{\ell=2}^{T} \frac{T!}{(T-\ell)!\ell}$ moment conditions. Note that when all moment conditions are included as we propose, we can utilize the equivalent linear programming condition which is only of size T^2 where T is the number of time periods.

We leave a formal study of power refinements and potential computational issues to future work. However, we note that since the mapping ε^* is convex in quantities, it is directional differentiable (see Rockafellar [2015], Section 23 for a textbook reference). The test statistic S is thus also convex and directly differentiabile in quantities. Thus,

[23] Echenique et al. [2011] consider violations of the generalized axiom of revealed preference, which is an acyclicity condition involving sequences of arbitrary length. In their empirical implementation, the money pump they compute involves only certain sequences of short length. In the discrete choice literature, Shi et al. [2018] study an acyclicity condition closely related to that of Browning [1989], Brown and Calsamiglia [2007], and the present paper, but for empirical implementation Shi et al. [2018] focus on sequences of length 2.

[24] See also Kitamura and Stoye [2018].

[25] Romano et al. [2014] present an alternative procedure that bootstraps a recentered test statistic.

S fulfills a key requirement of Fang and Santos [forthcoming].

4.1 Inference on ε^*

The previous arguments describe how to test whether the representative agent is ε-rationalizable by quasilinear utility. By inverting the test, one can construct a one-sided confidence set for the representative agent's measure of misspecification ε^*, which is defined as

$$\varepsilon^* = \max\left\{0, \max_{\{t_m\}_{t=1}^T} \mathbb{E}\left[\frac{1}{M}\sum_{m=1}^{M} p^{t_m} \cdot \left(X^{(i,t_m)} - X^{(i,t_{m+1})}\right)\right]\right\}$$

where the maximum is taken over all distinct sequences. Recall the second part of Proposition 2 describes the measure of misspecification as a maximum across sequences. A lower one-sided confidence interval is given by

$$\left[\varepsilon^*\left(\left\{\overline{X}^t\right\}_{t=1}^T\right) - c_{1-\alpha}, \infty\right) \cap [0,\infty).$$

In Appendix A.1 we provide the motivation behind the confidence intervals described in this section.

Building a confidence interval for ε^* goes beyond just running a specification test. Indeed, a test of exact quasilinear maximization amounts to checking whether 0 is in the confidence interval. Here, when we reject a hypothesis that $\varepsilon = 0$, we obtain a bound on the minimal misspecification needed to explain the data holding fixed nominal size α. By measuring violations in dollars, one can assess whether violations of the model are economically significant, rather than only statistically significant.

It is possible to construct other confidence sets for ε^*. For example, a two-sided confidence set for ε^* may be constructed as

$$\left[\varepsilon^*\left(\left\{\overline{X}^t\right\}_{t=1}^T\right) - \tilde{c}_{1-\alpha}, \varepsilon^*\left(\left\{\overline{X}^t\right\}_{t=1}^T\right) + \tilde{c}_{1-\alpha}\right] \cap [0,\infty),$$

where $\tilde{c}_{1-\alpha}$ is a bootstrap quantile computed the same as $c_{1-\alpha}$, except in step (iv), we set

$$\tilde{c}_{1-\alpha} = \inf \left\{ c \mid P^* \left(\max_{j \in \{1,\ldots,J\}} |\hat{\mu}_j^* - \hat{\mu}_j| \geqslant c \right) \leqslant \alpha \right\}.$$

The difference is that $\tilde{c}$ involves the maximum of absolute values, whereas c does not include the absolute values.[26] This confidence interval may be used both for testing and to provide an upper bound the minimal amount of misspecification needed to rationalize the data.

5 Empirical Application

We now apply our methodological tools to a dataset previously used in Echenique et al. [2011]. We use the Stanford Basket Dataset, which is a household-level panel dataset on grocery store purchases. This dataset consists of 494 households from June 1991 to June 1993. We use the transformed dataset of Echenique et al. [2011], which restricts attention to food purchases and aggregates to the brand level for 4 week periods. There are a total of 26 four week periods. After brand aggregation there are 375 unique goods. For full details on this dataset, we refer the reader to Echenique et al. [2011].

5.1 Household Data

We first study the measure of misspecification with household data. Table 1 reports summary statistics of the measure of misspecification. None of the households can be exactly described by a quasilinear utility model. The average of $\varepsilon^{i,*}$ is \$31.96 per (four week) time period. For comparison, the average total expenditure is \$213.75 per period.

Figure 2 presents the distribution of the smallest $\varepsilon^{i,*}$ such that a household is $\varepsilon^{i,*}$-quasilinear rationalizable. Most of the mass lies between 20 and 60, though there are some outliers that are larger. We note the shape of both distributions is broadly

[26] We thank Don Andrews for pointing out a mistake in a previous draft. Appendix A.1 provides motivation for this confidence interval. See also Andrews et al. [2019] and references therein.

Table 1: Summary Statistics of $\varepsilon^{i,*}$

Mean	31.96
Median	28.59
Min	5.73
Max	156.42
No. Households	494

consistent with a different measure of misspecification considered in Echenique et al. [2011], who quantify violations of the weak axiom of revealed preference.

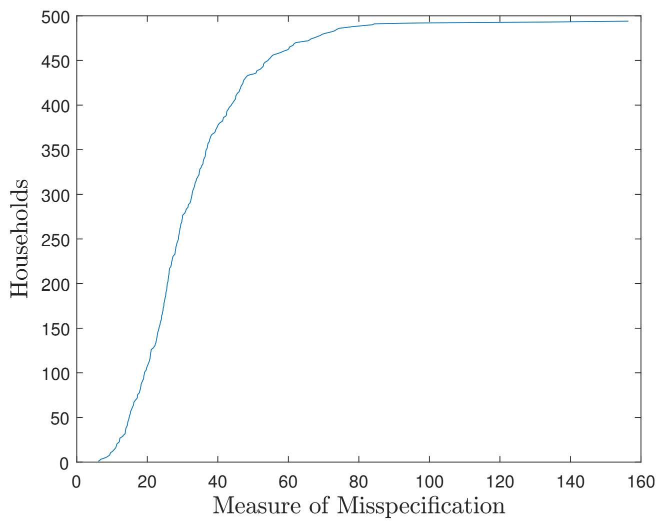

Figure 2: Cumulative Distribution Function of Household $\varepsilon^{i,*}$.

5.2 Aggregate Data

In contrast with the individual data, we find that the average demands are exactly consistent with maximization of quasilinear utility. This also implies that the statistical test of Section 4 fails to reject.

To further analyze the aggregate data, we construct a two-sided confidence interval for ε^*, calculated as described in Section 4.1 with $\alpha = .05$. This interval is given by $[0, 3.17]$. We can contrast these bounds with the household data, where all households are inconsistent with quasilinear utility. The upper bound on the confidence interval, 3.17, is smaller than the smallest individual $\varepsilon^{i,*}$, 5.73, which is reported in Table 1.

This is a demonstration that the inequality of the measure ε^* on aggregate data from Proposition 4(ii) can greatly reduce the magnitude of misspecification relative to the average of the individual measures. Again, we mention that the magnitude of this gap due to aggregation depends on the particular configuration of individual choices.

There are several reasons why we may obtain a sharp contrast between the individual level and aggregate level measures of misspecifcation. One possibility is that individual misoptimization is not systematic, so that when we aggregate it averages away. Another possibility is individuals exactly optimize but receive taste shocks. Then aggregation across individuals can serve to cancel out these taste shocks as in Proposition 7. We note that this paper does not try to separate these two stories, though the inequalities in Section 3.2 help interpret the measure of misspecification in the presence of shocks. Ultimately, a formal framework that attempts to separate the stories would involve a number of subjective judgments on behavior and measurement error. Finally, Proposition 5 suggests the contrast is not mechanical: individuals whose consumption is unrelated to prices would not tend to deterministically satisfy the restrictions of quasilinear utility upon aggregating to form sample averages.

6 Simulations

In the previous section, we obtain the contrasting result that while all households are inconsistent with quasilinear utility, a representative agent with 494 households is exactly consistent. We now present simulations designed to better understand how aggregation leads to this contrast.[27]

These simulations also contrast quasilinear utility with the more general model of utility maximization. We find several qualitative differences between these models in the simulations. Overall, quasilinear utility appears to be a considerably more restrictive model than general utility. This suggests the finding that aggregate data are consistent with quasilinear utility is qualitatively more surprising than the finding that the aggregate data are also consistent with general utility [Echenique et al., 2011].

[27] We also provide simulations to investigate how the length of the dataset affects the analysis. Moreover, we examine whether we might obtain data consistent with a model even if choices come from pure noise. These simulations are presented in Appendix B.

In order to further examine the role of aggregation, we study intermediate cases in which we do not average over all 494 households, but over a smaller number of households. We do this by a simulation design involving different (simulated) sample sizes n_b that range from 1 to 494. In the simulations, we examine how likely a simulated dataset can be described by quasilinear and general utility models. Keeping with the vocabulary from the revealed preference literature, we examine the pass rate of the simulated data. The pass rate is the percentage of simulations that can be described by a given model.

First, we compare pass rates of quasilinear utility with the larger model of general utility maximization. We study how pass rates depend on the sample size. For each sample size $1 \leqslant n_b \leqslant 494$, we construct a dataset in the following manner:

1. Draw a simple random sample (with replacement) of size n_b from the collection of all individual datasets $\{X^i\}_{i=1}^n$. Each individual dataset describes quantities for all goods at all time periods for individual i.
2. Form the sample average $\overline{X}^*_{n_b} = \left(\overline{X}^{*1\prime}_{n_b}, \ldots, \overline{X}^{*T\prime}_{n_b}\right)'$ obtained from the previous step. Here,
$$\overline{X}^{*t\prime}_{n_b} = \frac{1}{n_b} \sum_{i=1}^{n_b} X^{*(i,t)},$$
where $X^{*(i,t)}$ is obtained from the i-th draw in the previous step.

We do this 1,000 times for each n_b. With these simulated quantities datasets $\overline{X}^*_{n_b}$ and the original prices, we then report the simulated probability of being consistent with quasilinear utility. For comparison, we also report the simulated probability that $\overline{X}^*_{n_b}$ is consistent with the more general model of utility maximization subject to a budget constraint, i.e. we check the generalized axiom of revealed preference (GARP) [Afriat, 1967].

We report the results in Figure 3. With quasilinear utility, we find that with small sample sizes there is almost no chance of the aggregate dataset $\overline{X}^*_{n_b}$ being consistent with the model. The first simulation in which an aggregate dataset is consistent with quasilinear utility occurs at $n_b = 11$. The pass rate increases most rapidly between about 25 and 125 individuals, and then only slow increases, reaching about 63%.[28]

[28] We note that if our simulation instead drew datasets without replacement, this probability would peak at 100% with $n_b = 494$ since the average demands are exactly consistent with quasilinear

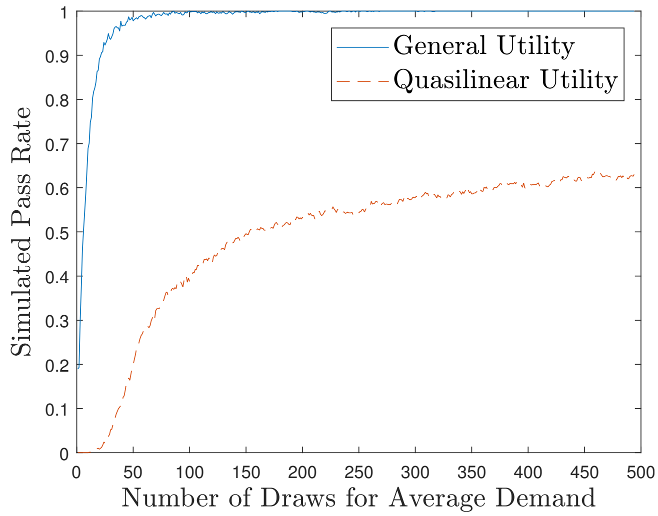

Figure 3: Simulated Pass Rate for General Utility and Quasilinear Utility

Our simulation exhibits different behavior when checking GARP. First, we note that Echenique et al. [2011] have previously reported that about 20% of individuals satisfy GARP with this dataset, but that the aggregated data are consistent with GARP. Figure 3 indicates that aggregating over a few individuals rapidly increases the probability of being consistent with GARP. While about 20% of simulations with one household are consistent with GARP, when we simulate with 10 households this number jumps to 70%. With 50 households it is 98%. It is important to note that this is an empirical *inverse* of the fact that GARP itself does not aggregate, i.e. even if all individuals are consistent with GARP it may not be consistent with a representative agent. In contrast, the simulation result for GARP indicates that even though relatively few households are consistent with GARP, when we aggregate the model admits a representative agent with high probability.

We summarize the main differences between quasilinear utility and general utility in this simulation design. First, aggregating over only a few individuals leads to an aggregate dataset passing GARP with high probability, while the aggregation effect with quasilinear utility is slower and never reaches 100%.[29]

utility when all individuals are used to form the average.

[29] The probability of an aggregate dataset passing GARP is not 1 with this design at the largest sample size $n_b = 494$, even though all of our 1,000 simulations pass GARP. This is because by drawing with replacement, there is a positive probability that we could draw the same individual who does not satisfy GARP all 494 times. An aggregate dataset constructed from these draws would not satisfy GARP.

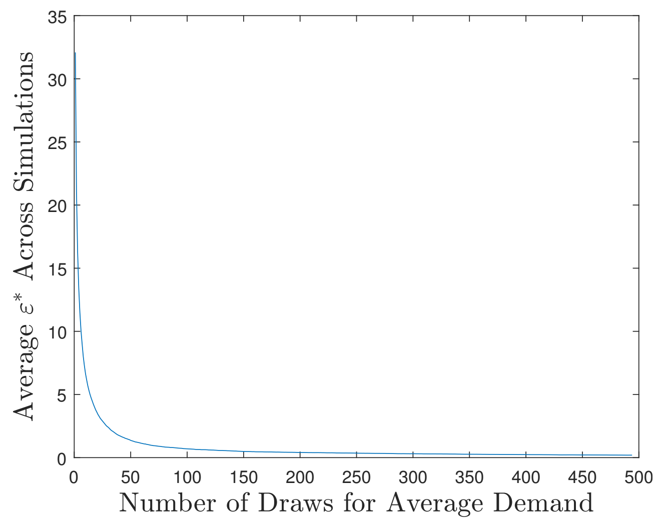

Figure 4: Average of Measure of Misspecification Across Simulations

Next, we study how the average measure of misspecification for quasilinear utility varies with the simulated sample size n_b, when applied to the aggregated dataset $\overline{X}^*_{n_b}$. Figure 4 reports the average measure of misspecification across 1,000 simulation draws, against the simulated sample size. We find that the measure rapidly decreases when observations are added. This suggests that while quasilinear utility may be inappropriate for modelling a single individual, it may be appropriate for modelling a small group. For example, with 10 draws of the average demand the average is about \$7, and with 25 draws it falls to \$3. When we consider larger groups, the average measure of misspecification becomes negligible. For example, with $n_b = 494$ draws, it is \$.20. This contrasts with the finding in Figure 3 that with $n_b = 494$, only 63% of simulation draws lead to a representative agent that is consistent with quasilinear utility.[30]

We note that some features of Figure 4 are specific to our particular dataset. For example, if there were no heterogeneity the measure of misspecification would be flat in the number of draws of average demand. In addition, because the original aggregate dataset is exactly consistent with quasilinear utility, we can conclude that as $n_b \to \infty$, the average measure of misspecification shrinks to 0.[31]

[30] We do not report a related result for the general model of utility maximization. The measure of misspecification in this paper is designed for quasilinear utility, which is in units of dollars, and does not immediately apply to general utility.

[31] This follows from the law of large numbers and the continuous mapping theorem since ε^* is convex, hence continuous.

7 Related Literature

At a high level, this paper is related to a research agenda set out by Becker [1962], who emphasizes assessing the useful implications of a theory, rather than all of the implications. In particular, Becker [1962] focuses on aggregates rather than individual behavior since aggregate behavior is often the object of interest (e.g. Hicks [1946]). Apart from these broad conceptual points, the main conclusions of Becker [1962] are different from ours. Becker [1962] shows that under certain assumptions on individual heterogeneity, budget constraints may *mechanically* induce aggregates to satisfy a version of the law of demand, even if individuals behave irrationally.[32] Our model of approximate quasilinearity does not have a budget constraint, so our aggregation result (Proposition 3) does not appear to have any formal relation to the analysis of Becker [1962].

Throughout the paper, we focus on quasilinearity due to its theoretical tractability and use in applications. Aggregation in an exact quasilinear model is well-known and a revealed preference characterization of quasilinear utility has been established by Browning [1989] and Brown and Calsamiglia [2007]. We generalize these results by allowing model approximation error and dropping the assumption of concavity. There has been relatively little empirical work assessing the ability of nonparametric quasilinear preferences to describe data. One exception is Castillo and Freer [2016], who examine the extent to which quasilinear preferences can describe individual-level data. They find evidence against quasilinear preferences at the individual level.

To assess the amount of misspecification of quasilinearity, we introduce a new additive measure of misspecification. This is the first measure of misspecification devoted to quasilinear utility, though several measures exist for the standard consumer problem (e.g. Afriat [1973], Varian [1990], Echenique et al. [2011], Dziewulski [2018]).[33]

While the focus of this paper is on assessing the ability of quasilinear utility to describe data at different levels of aggregation, to our knowledge this paper is the first to provide a nonparametric statistical specification test for a representative agent

[32] See also Grandmont [1992].

[33] Other measures include Apesteguia and Ballester [2015], Echenique et al. [2018], and de Clippel and Rozen [2018].

with quasilinear utility when there are multiple goods.[34] The approach of Aguiar and Kashaev [2018] may be used with panel data to test certain models including quasilinear utility, but they do not provide a test for a representative agent. To our knowledge, the closest precedent for our statistical testing approach appears to be Melo et al. [2017]. They test a model of strategic behavior in a game-theoretic setting. The moment inequalities they test are related to ours, though our test statistic is a supremum-type (without studentization) while theirs is an inverse-variance weighted quadratic form.

Our statistical test of the hypothesis that a representative agent is an ε-quasilinear maximizer involves linear inequality restrictions on means. There is a large literature on testing unconditional moment inequalities, including Andrews and Soares [2010], Chernozhukov et al. [2014], and Romano et al. [2014]. Relative to this literature, which has focused on general models with no particular relationship between the moment conditions, our setup has more structure. This has computational implications that we exploit, and also informs our deliberate choice not to studentize our test statistic, as discussed in detail in Section 4. The linear inequality restrictions we test have a similar structure to those in Kitamura and Stoye [2018], Deb et al. [2018], and Cattaneo et al. [2017],[35] though our methods are distinct.

The analysis of individual and aggregate data helps interpret Chetty [2012], who uses a representative agent model to place bounds on certain elasticities of demand. Chetty [2012] argues that "small" deviations from exact optimization can rationalize a variety of elasticities. Our theoretical and empirical analysis suggest that whether deviations are viewed as "small" depends on the level of aggregation. We note that we differ from Chetty [2012] because our preferred measure of deviations from quasilinearity does not divide by expenditure, whereas the measure of Chetty [2012] is a budget-weighted average measure and is designed for parametric utility functions while ours is nonparametric.[36]

[34] When there is a single good (plus an unobserved numeraire good), the literature on testing regression monotonicity (e.g. Chetverikov [forthcoming]) can be applied to conduct a specification test for a representative agent with quasilinear utility. This is because for this case, monotonicity in own-price is the only restriction of quasilinear utility. See also Härdle et al. [1991].

[35] Cattaneo et al. [2017] considers a studentized max statistic when testing whether a class of stochastic choice models is consistent with data. The unstudentized version of their statistic may have a natural interpretation of "economic" violations of their model, much like the statistic S we use.

[36] It is possible to adapt ε-rationalizability for quasilinear utility to parametric quasilinear utility.

Our empirical analysis is most closely related to Cherchye et al. [2016], who provide a revealed preference test of exact aggregation using the Gorman polar form [Gorman, 1953]. In contrast to the results here, Cherchye et al. [2016] find that individual-level data from the Spanish Continuous Family Expenditure Survey (SCFES) can be rationalized, but there is no common scale for the population of choices, which is necessary to admit a representative agent. The SCFES aggregates data into a small number of categories (15 goods) from disparate sources of total expenditure. In contrast, the Stanford Basket data we use is aggregated at the level of brand (375 goods) for grocery store products. It is possible that the difference in types of goods studied and the pre-processing of data into goods drives this difference, but more research is required.

This paper is also related to theoretical work by Hildenbrand [1983] and Quah [1997], who examine when the law of demand holds in the aggregate. These papers examine conditions in which rational individual demands generate aggregate demand that exactly [Hildenbrand, 1983] or approximately [Quah, 1997] satisfies the law of demand. Thematically, these papers differ from our research because they study aggregation of rational individuals. One interesting experiment looking at the role of aggregation in markets is the work of Crockett et al. [2018]. Finally, a large literature on aggregation of different demand systems is reviewed in Stoker [1993].

8 Discussion

This paper examines the ability of structured utility maximization to describe data at different levels of aggregation focusing on restrictions imposed by quasilinear utility. To conduct this analysis, we provide a measure of misspecification for quasilinear models and establish an aggregation result for this measure. We provide a statistical test for the hypothesis that a representative agent is ε-rationalized by quasilinear utility for panel data. We then analyze the aggregate implications of quasilinear utility using panel data from the Stanford Basket Dataset.

For a conjectured parametric utility function, a natural adaptation of our setup would be to label an individual ε-rationalizable if the utility obtained is always within ε of the indirect utility of the conjectured parametric function.

We find that while all individuals are inconsistent with quasilinear utility, the representative agent is consistent using a deterministic (or statistical) framework. This shows that while there is approximation error from quasilinearity in individual data, it is not systematic enough that the model of aggregate quasilinear behavior is invalidated. This is consistent with a view discussed in Becker [1962] that while certain models may be misspecified at the individual level, they may nonetheless provide useful information on market demand. The results here support this hypothesis for quasilinear utility and provide a way to check this hypothesis for other panel datasets. This paper also provides evidence against the hypothesis of Russell and Thaler [1985] that individual errors are systematic and will not average away in markets.

It is possible that the quality of the quasilinear model may degrade as the length of a panel dataset increases, since time may allow income effects to manifest more strongly. We note that the number of goods we analyze in the application is 375 and it appears this test can be applied to large datasets with little difficulty.[37] This paper raises natural follow-up questions concerning welfare and prediction that we are investigating in ongoing work.

Appendix A Proofs of Main Results

Proof of Proposition 1. Since the dataset is ε-rationalized by quasilinear utility, it follows that there exists a function $u : \mathbb{R}^K_+ \to \mathbb{R}$ such that

$$\begin{aligned} u(x^r) - p^r \cdot x^r + \varepsilon &\geqslant u(x^s) - p^r \cdot x^s \\ u(x^s) - p^s \cdot x^s + \varepsilon &\geqslant u(x^r) - p^s \cdot x^r. \end{aligned}$$

Adding the inequalities gives

$$2\varepsilon - p^r \cdot x^r - p^s \cdot x^s \geqslant -p^r \cdot x^s - p^s \cdot x^r.$$

The result follows by rearrangement. □

[37] For the statistical test of quasilinear utility and confidence interval construction, the most time-intensive component involved calculation of bootstrap quantiles. This took 661 seconds with 5,000 bootstrap draws with an Intel Core i7-7700 processor, without parallelizing.

Proof of Theorem 1. The implications (i) $\implies$ (ii) $\implies$ (iii) and (iv) $\implies$ (i) are straightforward. We now show that (iii) $\implies$ (iv).

Fix $(x^0, p^0) \in \{(x^t, p^t)\}_{t=1}^T$ and let Σ define the set of finite sequences of $t \in \{1, \ldots, T\}$ with no cycles that begins at (x^0, p^0).[38] Define

$$U(x) = \min_{\sigma \in \Sigma} \left\{ \left(x - x^{\sigma(M)}\right) \cdot p^{\sigma(M)} + \sum_{m=1}^{M-1} \left(x^{\sigma(m+1)} - x^{\sigma(m)}\right) \cdot p^{\sigma(m)} + M\varepsilon \right\},$$

which is motivated by re-arranging (iii). $U(x)$ is the minimum of finitely many affine functions, and thus is continuous, monotonic increasing, and concave. We now show that $U(x)$ provides an ε-quasilinear rationalization of the dataset. Consider $x \in \mathbb{R}_+^K$ such that $x \neq x^t$ and let $\sigma_t \in \Sigma$ be a minimizing sequence, i.e. a sequence such that $U(x^t) = (x^t - x^{\sigma_t(M_t)}) \cdot p^{\sigma_t(M_t)} + \sum_{m=1}^{M_t-1} (x^{\sigma_t(m+1)} - x^{\sigma_t(m)}) \cdot p^{\sigma_t(m)} + M_t\varepsilon$ where M_t is the length of sequence σ_t. It follows that

$$\begin{aligned} U(x) - p^t \cdot x \leqslant\ & (x - x^t) \cdot p^t + (x^t - x^{\sigma_t(M_t)}) \cdot p^{\sigma_t(M_t)} \\ & + \sum_{m=1}^{M-1} (x^{\sigma_t(m+1)} - x^{\sigma_t(m)}) \cdot p^{\sigma_t(m)} + (M_t + 1)\varepsilon - p^t \cdot x \\ =\ & -p^t \cdot x^t + \varepsilon + (x^t - x^{\sigma_t(M_t)}) \cdot p^{\sigma_t(M_t)} + \sum_{m=1}^{M-1} (x^{\sigma_t(m+1)} - x^{\sigma_t(m)}) \cdot p^{\sigma_t(m)} + M_t\varepsilon \\ =\ & U(x^t) - p^t \cdot x^t + \varepsilon \end{aligned}$$

where the first inequality follows since $U(x)$ by construction is the smallest term for all sequences, the second equality follows by rearrangement and canceling out the money spent on x at prices from period t, and the final equality holds by invoking the choice of σ_t. Thus, $U(x)$ ε-quasilinear rationalizes the dataset. $\square$

Proof of Proposition 2. For the dataset $\{(x^t, p^t)\}_{t=1}^T$, we first show existence of a so-

[38] It is without loss to consider such sequences since if there is a cycle, then $\sum_{m=1}^{M} p^{t_m} \cdot (x^{t_{m+1}} - x^{t_m}) + M\varepsilon \geqslant 0$ by assumption so that this will not be a minimizing sequence.

lution ε^* to the linear programming problem while dropping the restriction $\varepsilon \geqslant 0$,[39]

$$\min_{\substack{\varepsilon \in \mathbb{R} \\ u^1,\ldots,u^T \in \mathbb{R}_+}} \varepsilon \quad \text{s.t.} \quad u^s \leqslant u^r + p^r \cdot (x^s - x^r) + \varepsilon \quad \text{for all } r, s \in \{1, \ldots, T\}$$

where $u = (u_1, \ldots, u_T) \in \mathbb{R}_+^T$ and $\varepsilon \in \mathbb{R}$. Let $\alpha = \max_{t=1,\ldots,T}\{p^t \cdot x^t\}$ and note that $u = (0, \ldots, 0)$ and $\varepsilon = \alpha$ is feasible for the primal problem. Using standard duality (see Boyd and Vandenberghe [2004] Chapter 5), we obtain the dual maximization problem

$$\begin{aligned} \max_{\substack{(\lambda_{s,r})_{s,r} \in \mathbb{R}_+^{T^2} \\ (\lambda_s)_{s=1}^T \in \mathbb{R}_+^T}} \quad & -\sum_{s=1}^T \sum_{r=1}^T \lambda_{s,r} p^r \cdot (x^s - x^r) \\ \text{s.t.} \quad & \sum_{r=1}^T \lambda_{s,r} - \sum_{r=1}^T \lambda_{r,s} - \lambda_s = 0 \\ & \sum_{s=1}^T \sum_{r=1}^T \lambda_{s,r} = 1. \end{aligned}$$

Any set of λ terms that correspond to a cycle is feasible. For example, $\lambda_{1,2} = 1/2$, $\lambda_{2,1} = 1/2$, $\lambda_{s,r} = 0$ otherwise, and $\lambda_s = 0$ for all s is feasible. Since the primal and dual problems are both feasible, we can invoke the fundamental duality theorem of linear programming to ensure existence of a minimizer (see Gale [1989] Theorem 3.1). This shows the existence of a minimal ε regardless of a lower bound on ε.

That the two minimization problems in the main text have the same minimum follows from Theorem 1 (parts (ii) and (iii)). This equivalence says if ε is in the feasible set of one program, then it is in the feasible set of the other program. Lastly, for any $\varepsilon > \varepsilon^*$, the dataset is ε-quasilinear rationalized. This follows since for all finite sequences $\{t_m\}_{m=1}^M$ with $t_m \in \{1, \ldots, T\}$ and with $M \geqslant 2$, the inequality

$$\frac{1}{M} \sum_{m=1}^M p^{t_m} \cdot (x^{t_m} - x^{t_{m+1}}) \leqslant \varepsilon^* < \varepsilon$$

holds, where $(x^{t_{M+1}}, p^{t_{M+1}}) = (x^{t_1}, p^{t_1})$. This is equivalent to an ε-quasilinear ratio-

[39] This establishes existence when this bound is imposed, as a special case. We cover the case without bound to cover our test statistic S and bootstrap, which do not impose an inequality restriction analogous to $\varepsilon \geqslant 0$.

nalization by Theorem 1 (parts (i) and (iii)). $\square$

Proof of Proposition 3. We first prove part (i) of the proposition. For every $i = 1, \ldots, n$, suppose that $\{(x^{(i,t)}, p^t)\}_{t=1}^T$ is ε^i-rationalizabled by quasilinear utility. Let the *aggregate dataset* $\{(\bar{x}^t, p^t)\}_{t=1}^T$ be defined as in the main text. Since each individual dataset is ε^i-quasilinear rationalized, using Theorem 1 (parts (i) and (ii)) we see that for every $i = 1, \ldots, n$ there exist numbers $\{u^{(i,t)}\}_{t=1}^T$ such that

$$u^{(i,s)} \leqslant u^{(i,r)} + p^r \cdot (x^{(i,s)} - x^{(i,r)}) + \varepsilon^i$$

for all $s, r \in \{1, \ldots, T\}$. Summing up the inequalities across individuals and dividing by n, we obtain

$$\frac{1}{n}\sum_{i=1}^{n} u^{(i,s)} \leqslant \frac{1}{n}\sum_{i=1}^{n} u^{(i,r)} + p^r \cdot \left(\frac{1}{n}\sum_{i=1}^{n} x^{(i,s)} - \frac{1}{n}\sum_{i=1}^{n} x^{(i,r)}\right) + \frac{1}{n}\sum_{i=1}^{n} \varepsilon^i$$

for all $r, s \in \{1, \ldots, T\}$. Letting $\bar{u}^t = \frac{1}{n}\sum_{i=1}^n u^{(i,t)}$ and $\bar{\varepsilon} = \frac{1}{n}\sum_{i=1}^n \varepsilon^i$, we see that

$$\bar{u}^s \leqslant \bar{u}^r + p^r \cdot (\bar{x}^s - \bar{x}^r) + \bar{\varepsilon}$$

for all $s, r \in \{1, \ldots, T\}$. By Theorem 1 (parts (i) and (iii)), the aggregate dataset is $\bar{\varepsilon}$-quasilinear rationalized.

Part (ii) of the proposition follows since by part (i), the aggregate dataset is $\frac{1}{n}\sum_{i=1}^n \varepsilon^{i,*}$-quasilinear rationalized. Proposition 2 shows that a minimal $\bar{\varepsilon}^*$ exists and that it must be less than or equal to $\frac{1}{n}\sum_{i=1}^n \varepsilon^{i,*}$. $\square$

Proof of Proposition 4. We first prove part (i). For each sequence as in Theorem 1(iii), we obtain

$$\frac{1}{M}\sum_{m=1}^{M} p^{t_m} \cdot (X^{(i,t_m)} - X^{(i,t_{m+1})}) \leqslant \varepsilon^*(X^i).$$

Since $\mathbb{E}[X^i]$ exists and $\varepsilon^*(X^i)$ is non-negative and satisfies

$$\varepsilon^*(X^i) \leqslant \max_{t \in \{1,\ldots,T\}} \left\{p^t \cdot X^{(i,t)}\right\},$$

we obtain that $\mathbb{E}[\varepsilon^*(X^i)]$ exists because integrability is preserved under (finite) maxima and affine transformations. Taking expectations, we obtain

$$\frac{1}{M}\sum_{m=1}^{M} p^{t_m} \cdot (\mathbb{E}[X^{(i,t_m)}] - \mathbb{E}[X^{(i,t_{m+1})}]) \leqslant \mathbb{E}[\varepsilon^*(X^i)].$$

Since this is true for every cycle, by Theorem 1 the representative agent dataset $\{(\mathbb{E}[X^{(i,t)}], p^t)\}_{t=1}^T$ is $\mathbb{E}[\varepsilon^*(X^i)]$-rationalized by quasilinear utility.

To prove part (ii), let x^1 and x^2 be arbitrary quantities datasets. We want to show convexity, i.e. for any $\alpha \in [0,1]$, $\varepsilon^*(\alpha x^1 + (1-\alpha)x^2) \leqslant \alpha\varepsilon^*(x^1) + (1-\alpha)\varepsilon^*(x^2)$. Recall that

$$\frac{1}{M}\sum_{m=1}^{M} p^{t_m} \cdot (x^{(i,t_m)} - x^{(i,t_{m+1})}) \leqslant \varepsilon^*(x^i)$$

for each $i \in \{1,2\}$. Since this inequality is preserved under positive weighted averages we obtain

$$\frac{1}{M}\sum_{m=1}^{M} p^{t_m} \cdot \Big(\big(\alpha x^{(1,t_m)} + (1-\alpha)x^{(2,t_m)}\big) - \big(\alpha x^{(1,t_{m+1})} + (1-\alpha)x^{(2,t_{m+1})}\big)\Big)$$
$$\leqslant \alpha\varepsilon^*(x^1) + (1-\alpha)\varepsilon^*(x^2)$$

Since this is true for arbitrary cycles, the aggregate dataset $\{(\alpha x^{(1,t)} + (1-\alpha)x^{(2,t)}, p^t)\}_{t=1}^T$ is $(\alpha\varepsilon^*(x^1) + (1-\alpha)\varepsilon^*(x^2))$-rationalized by quasilinear utility. Thus, $\varepsilon^*(\alpha x^1 + (1-\alpha)x^2) \leqslant \alpha\varepsilon^*(x^1) + (1-\alpha)\varepsilon^*(x^2)$. Since ε^* is convex, the inequality in part (ii) follows by Jensen's inequality. □

Proof of Proposition 5. When $\varepsilon = 0$, Theorem 1(iii) is equivalent to the statement that for every permutation π of $\{1,\ldots,T\}$,

$$\sum_{t=1}^{T} p^t \cdot \overline{X}^t \leqslant \sum_{t=1}^{T} p^{\pi(t)} \cdot \overline{X}^t. \tag{7}$$

In other words, there is no permutation of prices that decreases total expenditure. We will write this in vector form. To that end, let

$$p = \left(p^{1\prime}, \ldots, p^{T\prime}\right)' \qquad \overline{X} = \left(\overline{X}^{1\prime}, \ldots, \overline{X}^{T\prime}\right)'.$$

Write the permutation π as a $KT \times KT$ block matrix Π. In more detail, Π is comprised of T blocks.[40] Thus, (7) in vector form reads

$$p \cdot \overline{X} \leqslant \Pi p \cdot \overline{X}.$$

Let E_Π denote the event that the minimum of $\tilde{\Pi} p \cdot \overline{X}$ across block permutation matrices $\tilde{\Pi}$ is attained at Π. The probability that Π^1 and Π^2 are both minimizers satisfies

$$P\left(E_{\Pi^1} \cap E_{\Pi^2}\right) \leqslant P\left((\Pi^1 p - \Pi^2 p) \cdot \overline{X} = 0\right).$$

Recall that $p^t \neq 0$ for each t, and $p^r \neq p^s$ for $r \neq s$. This implies that whenever $\Pi^1 \neq \Pi^2$, $(\Pi^1 p - \Pi^2 p)$ is not identically zero. Because $\left\{\overline{X}^t\right\}_{t=1}^T$ has a density with Lebesgue measure, we obtain that

$$P\left((\Pi^1 p - \Pi^2 p) \cdot \overline{X} = 0\right) = 0$$

whenever $\Pi^1 \neq \Pi^2$. Note that this is the only step in which we use the assumption of a density. This probability can be zero even when $\left\{\overline{X}^t\right\}_{t=1}^T$ does not have a density, and thus the assumption of a density is not crucial.

Because $X^{(i,t)}$ is independent and identically distributed across time and individuals, we have that $P(E_\Pi)$ is the same across permutations. Thus,

$$1 = P(\cup_\Pi E_\Pi) = \sum_\Pi P(E_\Pi) = T! P(E_\Pi). \tag{8}$$

Note that $T!$ is the cardinality of the set of block permutation matrices Π described above; equivalently this is the cardinality of the set of permutations of $\{1, \ldots, T\}$.

Note that as described in (7),

$$\varepsilon^*\left(\left\{\overline{X}^t\right\}_{t=1}^T\right) = 0$$

if and only if the event E_Π obtains with Π equal to the identity mapping. Thus, from

[40] More formally, Π is generated by turning π into a $T \times T$ permutation matrix. Then one takes the Kroenecker product of this matrix with the $K \times K$ identity matrix. This results in a $KT \times KT$ permutation matrix Π comprised of blocks.

(8) we have

$$P\left(\varepsilon^*\left(\left\{\overline{X}^t\right\}_{t=1}^T\right)=0\right)=P(E_\Pi)=\frac{1}{T!}.$$

□

Proof of Proposition 6. This is proven in the text. □

Proof of Proposition 7. For the proof we remove superscripts that depend on the i-th individual since the result will hold for any individual. The demand for each time period is assumed to be measurable with regard to the disturbances, and so there is a function X^t mapping tastes (η) and the degree of misoptimization (ε) to choices at period t. Here, η is a period-specific shock, not the full vector of shocks across periods. Recall that prices themselves are nonrandom in each time period and the same across individuals. Thus, their role is absorbed into X^t already because it depends on t. From the definition of an ε-maximizer, we note that for every (η, ε) that for all finite sequences $\{t_m\}_{m=1}^M$ with $t_m \in \{1, \ldots, T\}$ and $M \geqslant 2$, the inequality

$$\frac{1}{M}\sum_{m=1}^{M} p^{t_m}\cdot\left(X^{t_m}(\eta,\varepsilon)-X^{t_{m+1}}(\eta,\varepsilon)\right)\leqslant\varepsilon$$

holds, where $(X^{t_{M+1}}(\eta,\varepsilon), p^{t_{M+1}}) = (X^{t_1}(\eta,\varepsilon), p^{t_1})$ as in Theorem 1. To see this, note that fixing the value of η and ε we have for observations $s, r \in \{1, \ldots, T\}$

$$\begin{aligned}
u(X^r(\eta,\varepsilon),\eta)-p^r\cdot X^r(\eta,\varepsilon)+\varepsilon &\geqslant u(X^s(\eta,\varepsilon),\eta)-p^r\cdot X^s(\eta,\varepsilon)\\
u(X^s(\eta,\varepsilon),\eta)-p^s\cdot X^s(\eta,\varepsilon)+\varepsilon &\geqslant u(X^r(\eta,\varepsilon),\eta)-p^s\cdot X^r(\eta,\varepsilon),
\end{aligned}$$

which recovers the approximate law of demand

$$\frac{1}{2}(p^s-p^r)\cdot\left(X^s(\eta,\varepsilon)-X^r(\eta,\varepsilon)\right)\leqslant\varepsilon$$

by adding the inequalities. The cycles condition follows by similar reasoning.

Let μ denote the marginal distribution over (η, ε). Note that μ is the same each time period. We obtain that for all finite sequences $\{t_m\}_{m=1}^M$ with $t_m \in \{1, \ldots, T\}$ and

$M \geqslant 2$, the inequality

$$\int \frac{1}{M} \sum_{m=1}^{M} p^{t_m} \cdot (X^{t_m}(\eta, \varepsilon) - X^{t_{m+1}}(\eta, \varepsilon)) d\mu \leqslant \int \varepsilon d\mu$$

holds. The inequality above is equivalent to

$$\frac{1}{M} \sum_{m=1}^{M} p^{t_m} \cdot (\mathbb{E}[X^{t_m}(\eta, \varepsilon)] - \mathbb{E}[X^{t_{m+1}}(\eta, \varepsilon)]) \leqslant \mathbb{E}[\varepsilon],$$

where $(X^{t_{M+1}}(\eta, \varepsilon), p^{t_{M+1}}) = (X^{t_1}(\eta, \varepsilon), p^{t_1})$. From Theorem 1, this shows that $\{(\mathbb{E}[X^t(\eta, \varepsilon)], p^t)\}_{t=1}^{T}$ is $\mathbb{E}[\varepsilon]$-quasilinear rationalized, where the expectations are over μ. Since we assumed X^t is (η^t, ε)-measurable (recall we supress individual superscripts for simplicity), we have $\mathbb{E}[X^t] = \mathbb{E}[X^t(\eta, \varepsilon)]$. Note that we use the assumption that the distribution over (η, ε) is the same each time period in this last step, to link the mean of the observable quantities to the mean of the structural function. □

A.1 Motivation for Test and Confidence Intervals

This section describes the motivation for our test and confidence intervals. We introduce some additional notation. Recall each j indexes a sequence. For each such sequence define

$$\mu_j = \mathbb{E}\left[\frac{1}{M} \sum_{m=1}^{M} p^{t_m} \cdot \left(X^{(i,t_m)} - X^{(i,t_{m+1})}\right)\right].$$

We assume $c_{1-\alpha}$ approximates the $(1-\alpha)$-quantile of the distribution of $\max_{j \in \{1,\ldots,J\}}(\hat{\mu}_j - \mu_j)$, i.e.

$$P\left(\max_{j \in \{1,\ldots,J\}} (\hat{\mu}_j - \mu_j) \geqslant c_{1-\alpha}\right) \approx \alpha.$$

Note that $c_{1-\alpha}$ is constructed as a boostrap analogue of this probability. See Chernozhukov et al. [2017] for a theoretical study of this approximation.

The motivation for the test is standard in the moment inequalities literature, but we include the intuition for completeness. Similar logic has been used in the revealed preference literature [Varian, 1985, Echenique et al., 2011]. Under H_0, the test statistic S satisfies

$$\begin{aligned} S(X) &= \max_{j\in\{1,\ldots,J\}} \hat{\mu}_j - \varepsilon \\ &= \max_{j\in\{1,\ldots,J\}} \left(\hat{\mu}_j - \mu_j + (\mu_j - \varepsilon)\right) \\ &\leqslant \max_{j\in\{1,\ldots,J\}} \left(\hat{\mu}_j - \mu_j\right). \end{aligned}$$

The final inequality uses the fact that H_0 may be written $\mu_j \leqslant \varepsilon$ for each μ_j. With these inequalities, we obtain

$$P\left(S(X) > c_{1-\alpha}\right) \leqslant P\left(\max_{j\in\{1,\ldots,J\}} \left(\hat{\mu}_j - \mu_j\right) > c_{1-\alpha}\right) \approx \alpha.$$

Thus, the test (approximately) controls size.

To construct a confidence interval for the representative agent's ε^*, we can invert this test. First recall

$$S(X) = \max\left\{0, \varepsilon^*\left(\left\{\overline{X}^t\right\}_{t=1}^T\right)\right\} - \varepsilon.$$

Thus, provided $c_{1-\alpha} \geqslant 0$, $S(X) > c_{1-\alpha}$ if and only if $\varepsilon^*\left(\left\{\overline{X}^t\right\}_{t=1}^T\right) > c_{1-\alpha} + \varepsilon$. By inverting this test we obtain a lower one-sided confidence interval

$$\left[\varepsilon^*\left(\left\{\overline{X}^t\right\}_{t=1}^T\right) - c_{1-\alpha}, \infty\right) \cap [0, \infty).$$

We now turn to the motivation for the two-sided confidence interval. To that end, let j^* be a value such that $\mu_{j*} = \max_{j\in\{1,\ldots,J\}} \mu_j$, i.e. j^* corresponds to a worst-case cycle. We obtain

$$\begin{aligned} \{\forall j, \hat{\mu}_j - \mu_j \geqslant -a\} &\subseteq \{\mu_{j*} \leqslant \hat{\mu}_{j*} + a\} \\ &\subseteq \left\{\max_{j\in\{1,\ldots,J\}} \mu_j \leqslant \max_{j\in\{1,\ldots,J\}} \hat{\mu}_j + a\right\}. \end{aligned}$$

The first inclusion is obvious, since the inequality holding for each j implies it holds for j^*. The second inclusion follows from the fact that $\hat{\mu}_{j*} \leqslant \max_{j\in\{1,\ldots,J\}} \hat{\mu}_j$. In addition, we obtain

$$\{\forall j, \hat{\mu}_j - \mu_j \leqslant b\} \subseteq \{\forall j, \hat{\mu}_j - \mu_{j*} \leqslant b\}$$
$$= \left\{ \max_{j \in \{1,\ldots,J\}} \mu_j \geqslant \max_{j \in \{1,\ldots,J\}} \hat{\mu}_j - b \right\}.$$

Thus,

$$P\left(\max_{j \in \{1,\ldots,J\}} \hat{\mu}_j - b \leqslant \max_{j \in \{1,\ldots,J\}} \mu_j \leqslant \max_{j \in \{1,\ldots,J\}} \hat{\mu}_j + a \right)$$
$$\geqslant P(\forall j, -a \leqslant \hat{\mu}_j - \mu_j \leqslant b).$$

Setting $a = b = \tilde{q}_{1-\alpha}$ yields the two-sided confidence interval described in the main text. Andrews et al. [2019] call this a profiled confidence interval and provide further references. Note that $\tilde{q}_{1-\alpha}$ is a bootstrap quantile designed so that the approximation

$$P(\forall j, -\tilde{q}_{1-\alpha} \leqslant \hat{\mu}_j - \mu_j \leqslant \tilde{q}_{1-\alpha}) \approx 1 - \alpha$$

holds.

Appendix B Other Simulations

B.1 Rationalizing Noise

One concern with tests of nonparametric theories is that they may be too "large" in some sense. Thus, one may wonder how surprising it is that aggregate data are consistent with quasilinear utility. In this section, we attempt to understand how likely certain forms of random data would either be consistent with quasilinear utility or the more general model of utility maximization. This is usually referred to as a power calculation in the revealed preference literature [Bronars, 1987].

To that end, we consider a new design in which quantities are completely unrelated to prices. We present pass rates for quasilinear utility and GARP, i.e. the fraction of simulations in which each theory is deterministically consistent with the data. The design is conceptually similar to the theoretical analysis of Becker [1962]. As mentioned in Section 3.1, there is an important distinction between population-level means being exactly consistent with a model, and their sample counterparts. Recall Proposition 5 presents an example in which the probability of a deterministic rationalization is $1/T!$, even though the mean demands are exactly consistent with quasilinear utility.

As before, we study how the sample size alters the probability of a rationalization by each model. This design is related to a simulation design of Andreoni et al. [2013]. For each sample size n_b, we construct a dataset as follows:

1. Pick the time period t.
2. Draw a simple random sample (with replacement) of size n_b from the collection of all individual quantity tuples across all time periods $\left\{\left\{X^{(i,t)}\right\}_{i=1}^{n}\right\}_{t=1}^{T}$. Each individual dataset $X^{*(i,t)}$ is a K-dimensional vector of simulated quantities for period i in time t.
3. Form the sample average

$$\overline{X}_{n_b}^{*t\prime} = \frac{1}{n_b}\sum_{i=1}^{n_b} X^{*(i,t)},$$

   where $X^{*(i,t)}$ is obtained from the i-th draw in the previous step.
4. Repeat steps 1-3 for all time periods $t = 1, \ldots, 26$.
5. Form the average dataset across all time periods $\overline{X}_{n_b}^{*} = \left(\overline{X}_{n_b}^{*1\prime}, \ldots, \overline{X}_{n_b}^{*T\prime}\right)'$.

After $\overline{X}_{n_b}^{*}$ is generated, we check whether it is exactly consistent with quasilinear utility and GARP. We construct the pass rates across 1,000 simulations. We find that none of these datasets is exactly consistent with quasilinear utility. The pass rate for GARP is reported in Figure 6. It is highest for $n_b = 1$ individual, at 3.9% of simulations. The pass rate is small for all simulated sample sizes, and is typically around 1.5%. Thus there is a qualitative difference between GARP and quasilinear utility since one model passes sometimes while the other never does.[41] Intuitively,

[41]This statement holds for our simulations, and does not mean that the probability of passing quasilinear utility is 0 for this design. In fact, it can be shown that it must be positive, i.e. $P\left(\varepsilon^*\left(\overline{X}_{n_b}^{*}\right) = 0\right) > 0$.

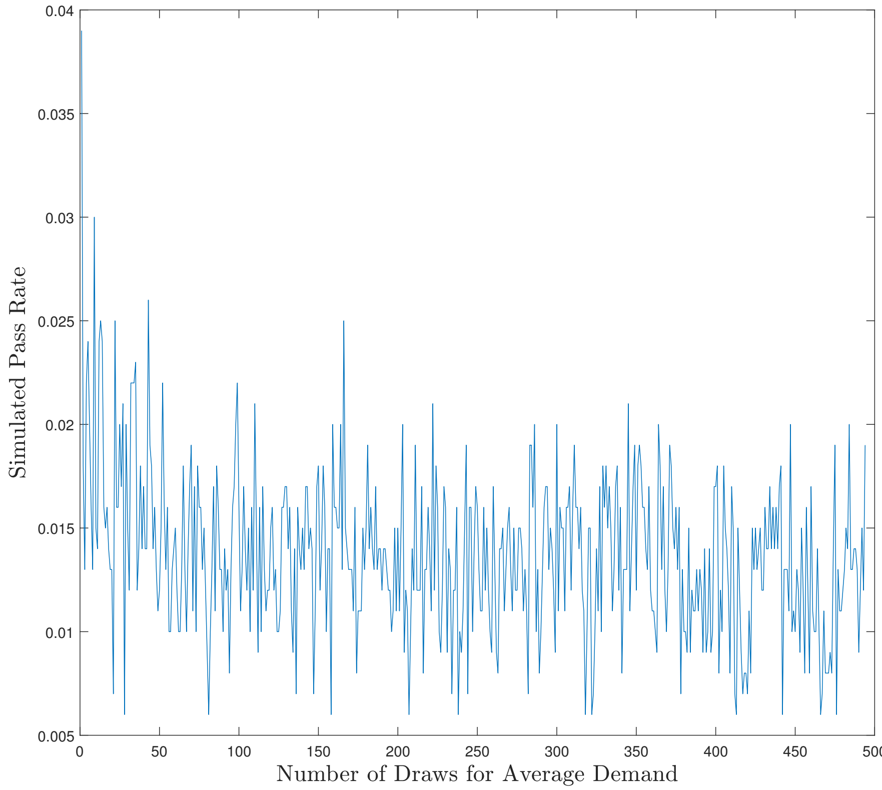

Figure 5: Simulated Pass Rate for General Utility with Noise

the simulations show that quasilinear utility is a much "smaller" model than general utility maximization.

B.2 Role of Number of Time Periods

The previous simulations in Section B.1 suggest that an aggregate dataset generated from pure noise would rarely be consistent with general utility or quasilinear utility. These simulations involved all 26 time periods, and a natural question is what would happen if we had fewer time periods.

First recall that Proposition 5 provides the theoretical guidance that when quantities are unrelated to prices, the probability of passing quasilinear utility rapidly decreases as more time periods are added. In this section, we further examine this by simulation both for quasilinear utility and GARP. We note that this simulation design is not formally covered by the proposition and thus is not redundant given the analytical result.

In this section we simulate aggregate data with a different number of time periods. The design is similar to the previous one in Section B.1, and only differs because we only consider $\overline{T} \leqslant T = 26$ time periods at a time. We simulate the pass rate of general utility and quasilinear utility for datasets of different lengths, i.e. $\overline{T} \in \{1, \ldots, T\}$. For each $\overline{T}$, we simulate data from the first $\overline{T}$ periods. The results are displayed in Figure 6.

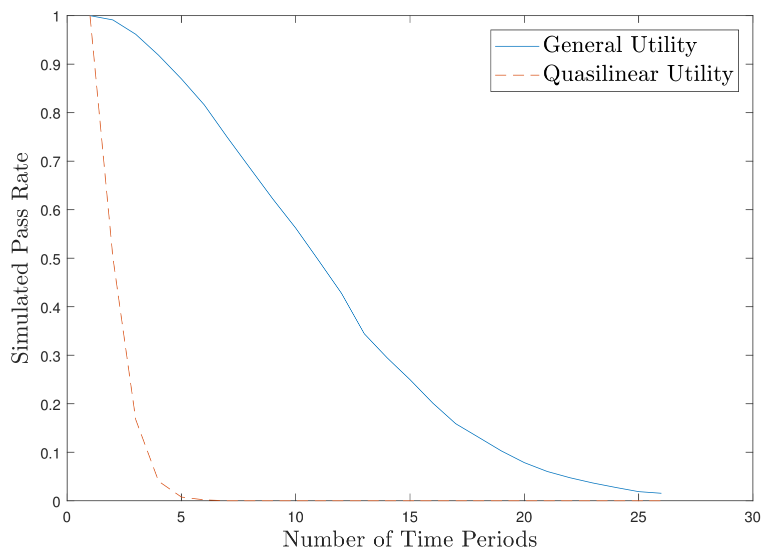

Figure 6: Simulated Pass Rate by Number of Time Periods

We find that while the pass rate with quasilinear utility drops rapidly with simulated datasets involving more time periods, the pass rate with general utility decreases at a much slower pace.[42] This provides intuition that if quantities are completely unrelated to prices, even with a few time periods quasilinear utility is likely to be refuted, whereas general utility needs more time periods. This suggests that measuring misspecification of quasilinear utility is a sensible exercise with a few time periods (e.g. five periods), whereas measuring misspecification or testing general utility requires a fairly large number of time periods (e.g. greater than 25).

Finally, we plot the average measure of misspecification for quasilinear utility for each time period. The average measure is small overall. It necessarily increases as we add

[42]We note that though this setup is not formally covered by Proposition 5, the pass rate of quasilinear utility closely follows the analytical rate $1/T!$.

time periods. The increase is fastest when we go from 1 to about 5 time periods, and then increases more slowly as additional time periods are added. The maximum average measure of misspecification is about 2 dollars. There is a theoretical reason this number is small. Note that since ε^* convex, it is also continuous as a function of the data. Applying the continuous mapping theorem we have that conditional on the original dataset, as $n_b \to \infty$,

$$\varepsilon^* \left(\left\{ \overline{X}_{n_b}^{*t\prime} \right\}_{t=1}^{\overline{T}} \right) \xrightarrow{p} \varepsilon^* \left(\left\{ \mathbb{E} \left[\overline{X}_{n_b}^{*t\prime} \right] \right\}_{t=1}^{\overline{T}} \right). \tag{9}$$

This holds for each $\overline{T}$. The limit is zero, since demands $\left\{ \mathbb{E} \left[\overline{X}_{n_b}^{*t\prime} \right] \right\}_{t=1}^{\overline{T}}$ are constant in time t, and therefore consistent with quasilinear utility. Thus, while the probability of data being exactly consistent with quasilinear utility is small, the expectation of the measure of misspecification ($\mathbb{E}[\varepsilon^*]$) is also small when the number of draws n_b is big.

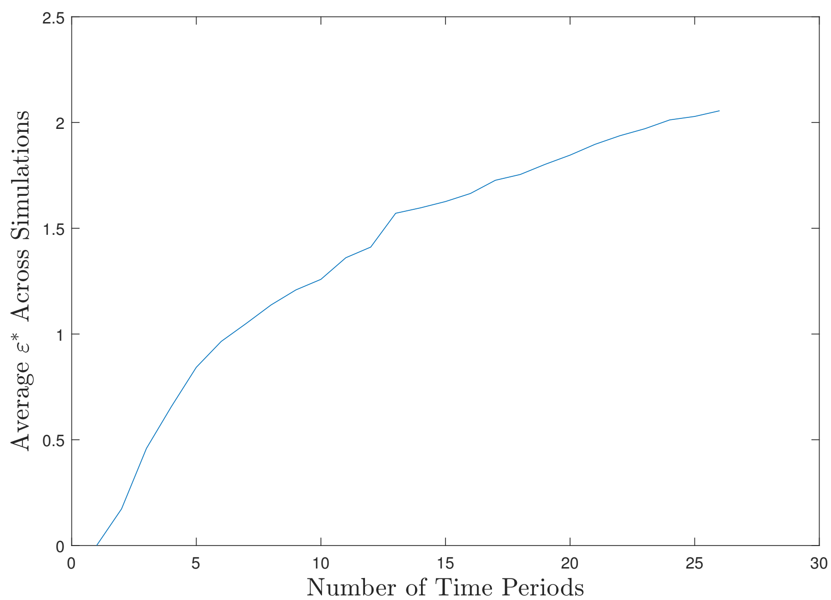

Figure 7: Average Measure of Misspecification by Time Periods

B.3 Role of Fraction of Noise

We have previously considered simulations in which either all demands are linked with prices as in the original data (Section 6) or none of them are (Section B.1).

We now study a simulation design with a mix between these two. This design helps understand how measuring misspecification for quasilinear utility for aggregate data can depend on the fraction of individuals who are optimizing relative to the fraction who are choosing in ways unrelated to prices.

We now study pass rates by a simulation design in which a fraction α of individual datasets are simulated from "noise." Specifically, they are constructed by randomly sampling quantities from the collection of all individual quantity tuples $\{\{X^{(i,t)}\}_{i=1}^{n}\}_{t=1}^{T}$. A fraction $1-\alpha$ of individual datasets are sampled from $\{X^{i}\}_{i=1}^{n}$ (with replacement). Here we examine when $T = 26$ so we use all time periods from the main analysis. Thus when $\alpha = 0$ the design follows Section 6, in which all choices are generated in a manner unrelated to prices, while when $\alpha = 1$ the design is the same as Section B.1. We study simulation results in which all time periods are included and the number of individual datasets we simulate is the same as the original sample, $n_b = 494$.

Simulated pass rates are reported in Figure 8. Quasilinear utility is qualitatively more responsive to the introduction of noise than general utility. Indeed, as the fraction of individual datasets simulated from noise increases, general utility passes in almost all simulations up until about 50% of choices are generated from noise. Even with 80% of data generated from noise, about 80% of simulations still are consistent with general utility when we aggregate. In contrast, pass rates for quasilinear utility are more sensitive to the introduction of noise. For example, with a 50% chance of each individual dataset being generated from noise, about 40% of simulations are consistent with quasilinear utility when we aggregate. When each dataset has an 80% chance of being generated from noise, less than 1% of simulations are consistent with quasilinear utility when we aggregate.

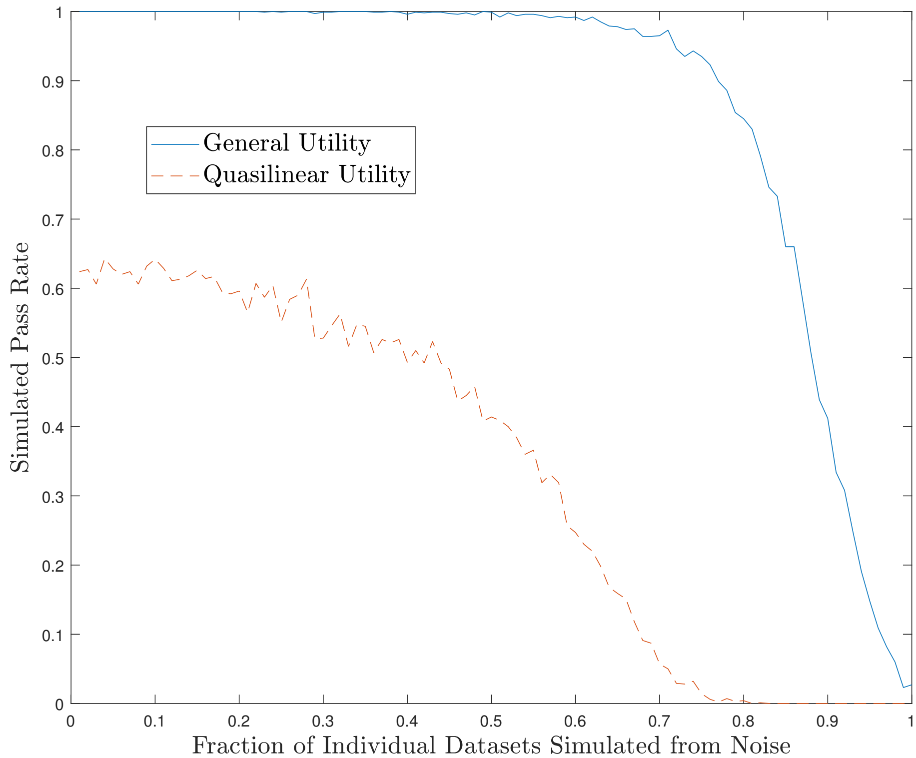

Figure 8: Pass Rates by Fraction of Noise

Finally, we we report how the average measure of misspecification for quasilinear utility relates to the fraction of noise. Results are displayed in Figure 9. The average measure of misspecification slowly rises up until about 60% of simulated datasets are sampled from noise. Then, the average measure of misspecification increases more rapidly. We note that this sample simulates $n_b = 494$ individuals to form each aggregate dataset. If we send $n_b \to \infty$, from arguments similar to the discussion around Equation 9, we would expect ε^* to typically be close to 0.

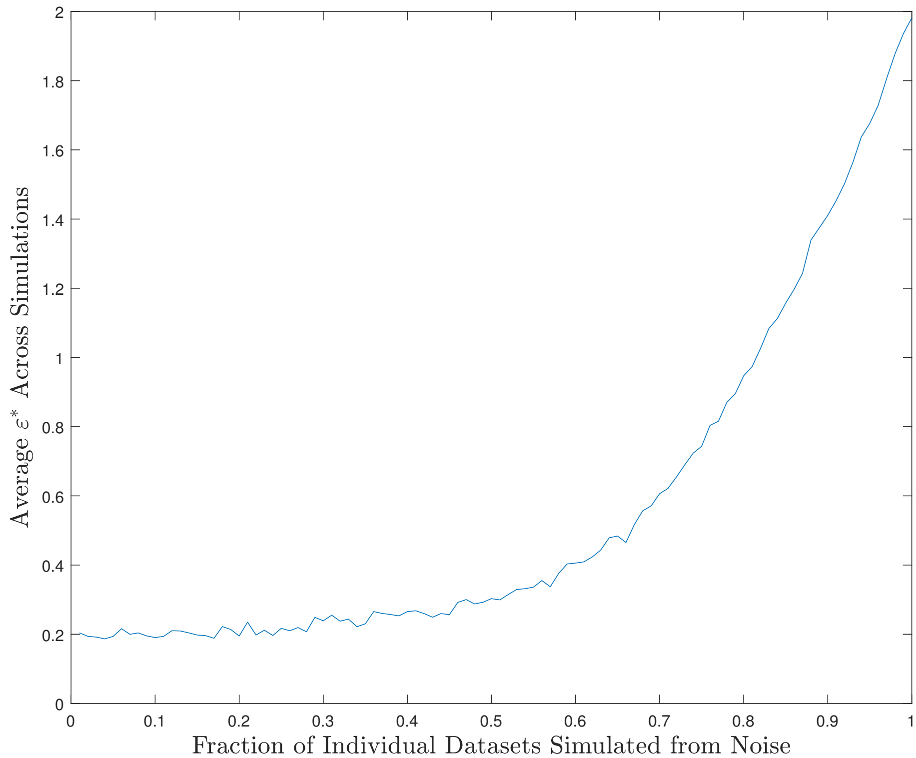

Figure 9: Average Measure of Misspecification by Fraction of Noise

Appendix C ε_V-quasilinear Rationalizability

As in Varian [1990], we could have proposed a measure $\varepsilon_{\mathrm{V}} = (\varepsilon_{\mathrm{V}}^1, \ldots, \varepsilon_{\mathrm{V}}^T) \in \mathbb{R}_+^T$ that allows decision-specific misspecification from quasilinear utility. We define an analogous notion of ε_{V}-quasilinear rational.

Definition 2. *A dataset $\{(x^t, p^t)\}_{t=1}^T$ is ε_V-rationalized by quasilinear utility for $\varepsilon_V = (\varepsilon_V^1, \ldots, \varepsilon_V^T) \in \mathbb{R}_+^T$ if there exists a locally non-satiated utility function $u : \mathbb{R}_+^K \to \mathbb{R}$ such that for all $t \in \{1, \ldots, T\}$ and for all $x \in \mathbb{R}_+^K$, the following inequality holds:*

$$u(x^t) - p^t \cdot x^t + \varepsilon_V^t \geqslant u(x) - p^t \cdot x.$$

We also refer to the above by saying a dataset is ε_V-quasilinear rationalized. When ε_V equals zero, it is convenient to say the dataset is quasilinear rationalized.

In particular, the following theorem follows from the same arguments as Theorem 1.

Theorem 2. *For any dataset* $\{(x^t, p^t)\}_{t=1}^T$ *and* $\varepsilon_V \in \mathbb{R}_+^T$*, the following are equivalent:*

(i) $\{(x^t, p^t)\}_{t=1}^T$ *is* ε_V*-rationalized by quasilinear utility.*

(ii) There exist numbers $\{u^t\}_{t=1}^T$ *that satisfy the following inequalities for all* $r, s \in \{1, \ldots, T\}$*:*

$$u^s \leqslant u^r + p^r \cdot (x^s - x^r) + \varepsilon_V^r.$$

(iii) For all finite sequences with $\{t_m\}_{m=1}^M$*,* $t_m \in \{1, \ldots, T\}$ *and* $M \geqslant 2$*, the inequality*

$$\sum_{m=1}^{M} p^{t_m} \cdot (x^{t_m} - x^{t_{m+1}}) \leqslant \sum_{m=1}^{M} \varepsilon_V^{t_m}$$

holds, where $(x^{t_{M+1}}, p^{t_{M+1}}) = (x^{t_1}, p^{t_1})$*.*

(iv) $\{(x^t, p^t)\}_{t=1}^T$ *is* ε_V*-rationalized by a quasilinear utility function that is continuous, monotonic increasing, and concave.*

Using the results from the main text, it is possible to search over ε_{V} that minimize deviations according to some criterion function with a constraint set of weak linear inequalities. An aggregation property analogous to Proposition 3 holds for ε_V^t in each time period. Moreover, if one considers the smallest ε_V that minimizes the average approximation errors $f(\varepsilon_V) = \frac{1}{T}\sum_{t=1}^T \varepsilon_V^t$, then the minimum for the aggregate dataset is weakly less than the average of the individual minimums. Following the ideas of Proposition 3, these properties are immediate from conditions (ii) and (iii) of Theorem 2. Other aggregator functions of ε_V may be considered, in the spirit of Varian [1990] and Halevy et al. [2018].

References

Sydney N Afriat. The construction of utility functions from expenditure data. *International economic review*, 8(1):67–77, 1967.

Sydney N Afriat. On a system of inequalities in demand analysis: an extension of the classical method. *International Economic Review*, pages 460–472, 1973.

Victor H. Aguiar and Nail Kashaev. Stochastic revealed preferences with measurement error. 2018. Working Paper.

Roy Allen and John Rehbeck. Identification with additively separable heterogeneity. *Econometrica*, 87(3):1021–1054, 2019a.

Roy Allen and John Rehbeck. Revealed stochastic choice with attributes. *Available at SSRN 2818041*, 2019b.

James Andreoni, Benjamin J Gillen, and William T Harbaugh. The power of revealed preference tests: Ex-post evaluation of experimental design. Technical report, 2013.

Donald WK Andrews and Gustavo Soares. Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica*, 78(1):119–157, 2010.

Isaiah Andrews, Matthew Gentzkow, and Jesse M Shapiro. Measuring the sensitivity of parameter estimates to estimation moments. *The Quarterly Journal of Economics*, 132(4):1553–1592, 2017.

Isaiah Andrews, Matthew Gentzkow, and Jesse M Shapiro. On the informativeness of descriptive statistics for structural estimates. Technical report, National Bureau of Economic Research, 2018.

Isaiah Andrews, Toru Kitagawa, and Adam McCloskey. Inference on winners. Technical report, National Bureau of Economic Research, 2019.

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

Timothy Armstrong and Michal Kolesár. Sensitivity analysis using approximate moment condition models. 2019.

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

Steven Berry, James Levinsohn, and Ariel Pakes. Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, pages 841–890, 1995.

Javier A Birchenall. Random choice and market demand. Technical report, 2016.

Stéphane Bonhomme and Martin Weidner. Minimizing sensitivity to model misspecification. *arXiv preprint arXiv:1807.02161*, 2018.

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge university press, 2004.

Stephen G Bronars. The power of nonparametric tests of preference maximization. *Econometrica*, pages 693–698, 1987.

Donald J Brown and Caterina Calsamiglia. The nonparametric approach to applied welfare analysis. *Economic Theory*, 31(1):183–188, 2007.

Martin Browning. A nonparametric test of the life-cycle rational expections hypothesis. *International Economic Review*, pages 979–992, 1989.

Marco Castillo and Mikhail Freer. A revealed preference test of quasi-linear preferences. Technical report, Working Paper, 2016.

Matias D Cattaneo, Xinwei Ma, Yusufcan Masatlioglu, and Elchin Suleymanov. A random attention model. *arXiv preprint arXiv:1712.03448*, 2017.

Laurens Cherchye, Ian Crawford, Bram De Rock, and Frederic Vermeulen. Gorman revisited: nonparametric conditions for exact linear aggregation. *SERIEs*, 7(2): 203–220, 2016.

Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. Testing many moment inequalities. *arXiv preprint arXiv:1312.7614*, 2014.

Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. Central limit theorems and bootstrap in high dimensions. *The Annals of Probability*, 45(4):2309–2352, 2017.

Raj Chetty. Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply. *Econometrica*, 80(3):969–1018, 2012.

Denis Chetverikov. Testing regression monotonicity in econometric models. *Econometric Theory*, forthcoming.

Timothy Christensen and Benjamin Connault. Counterfactual sensitivity and robustness. *arXiv preprint arXiv:1904.00989*, 2019.

Timothy G Conley, Christian B Hansen, and Peter E Rossi. Plausibly exogenous. *Review of Economics and Statistics*, 94(1):260–272, 2012.

Sean Crockett, Daniel Friedman, and Ryan Oprea. Revealed preferences and general equilibrium:a laboratory study. 2018. Working Paper.

John Dagsvik. Discrete and continuous choice, max-stable processes, and independence from irrelevant attributes. *Econometrica*, 62(5):1179–1205, 1994.

Geoffroy de Clippel and Kareen Rozen. Consumer theory with misperceived tastes. Technical report, 2018.

Rahul Deb, Yuichi Kitamura, John Quah, and Jörg Stoye. Revealed price preference: Theory and stochastic testing. 2018. Working Paper.

Paweł Dziewulski. Just-noticeable difference as a behavioural foundation of the critical cost-efficiency index. 2018. Working Paper.

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

Federico Echenique, Taisuke Imai, Kota Saito, et al. Approximate expected utility rationalization. Technical report, CRC TRR 190 Rationality and Competition, 2018.

Zheng Fang and Andres Santos. Inference on directionally differentiable functions. *Review of Economic Studies*, forthcoming.

David Gale. *The theory of linear economic models*. University of Chicago press, 1989.

William M Gorman. Community preference fields. *Econometrica*, pages 63–80, 1953.

Jean-Michel Grandmont. Transformations of the commodity space, behavioral heterogeneity, and the aggregation problem. *Journal of Economic Theory*, 57(1):1–35, 1992.

Trygve Haavelmo. The probability approach in econometrics. *Econometrica*, 12: iii–vi+1–115, 1944.

Yoram Halevy, Dotan Persitz, and Lanny Zrill. Parametric recoverability of preferences. *Journal of Political Economy*, 126(4):1558–1593, 2018.

Lars Peter Hansen and Ravi Jagannathan. Assessing specification errors in stochastic discount factor models. *The Journal of Finance*, 52(2):557–590, 1997.

Lars Peter Hansen, John Heaton, and Erzo GJ Luttmer. Econometric evaluation of asset pricing models. *The Review of Financial Studies*, 8(2):237–274, 1995.

Wolfgang Härdle, Werner Hildenbrand, and Michael Jerison. Empirical evidence on the law of demand. *Econometrica: Journal of the Econometric Society*, pages 1525–1549, 1991.

John Richard Hicks. *Value and Capital*. The Clarendon Press, 1946.

Werner Hildenbrand. On the "law of demand". *Econometrica*, pages 997–1019, 1983.

Josef Hofbauer and William H Sandholm. On the global convergence of stochastic fictitious play. *Econometrica*, 70(6):2265–2294, 2002.

Hendrik S Houthakker. Revealed preference and the utility function. *Economica*, 17 (66):159–174, 1950.

Alan P Kirman. Whom or what does the representative individual represent? *Journal of economic perspectives*, 6(2):117–136, 1992.

Yuichi Kitamura and Jörg Stoye. Nonparametric analysis of random utility models: testing, 2018.

Patrick Kline and Andres Santos. Sensitivity to missing data assumptions: Theory and an evaluation of the us wage structure. *Quantitative Economics*, 4(2):231–267, 2013.

Matthew A Masten and Alexandre Poirier. Identification of treatment effects under conditional partial independence. *Econometrica*, 2018a.

Matthew A Masten and Alexandre Poirier. Salvaging falsified instrumental variable models. *arXiv preprint arXiv:1812.11598*, 2018b.

Daniel McFadden. Econometric models of probabilistic choice. *Structural analysis of discrete data with econometric applications*, 198272, 1981.

Daniel McFadden and Marcel K Richter. Stochastic rationality and revealed stochastic preference. *Preferences, Uncertainty, and Optimality, Essays in Honor of Leo Hurwicz, Westview Press: Boulder, CO*, pages 161–186, 1990.

Daniel L McFadden. Revealed stochastic preference: a synthesis. *Economic Theory*, 26(2):245–264, 2005.

Emerson Melo, Kirill Pogorelskiy, and Matthew Shum. Testing the quantal response hypothesis. 2017.

Roger B Myerson. Optimal auction design. *Mathematics of operations research*, 6(1): 58–73, 1981.

John K-H Quah. The law of demand when income is price dependent. *Econometrica*, pages 1421–1442, 1997.

S Abu Turab Rizvi. The sonnenschein-mantel-debreu results after thirty years. *History of Political Economy*, 38(1):228–245, 2006.

Ralph Tyrell Rockafellar. *Convex analysis*. Princeton university press, 2015.

Joseph P Romano, Azeem M Shaikh, and Michael Wolf. A practical two-step method for testing moment inequalities. *Econometrica*, 82(5):1979–2002, 2014.

Thomas Russell and Richard Thaler. The relevance of quasi rationality in competitive markets. *The American Economic Review*, 75(5):1071–1082, 1985.

Paul A Samuelson. A note on the pure theory of consumer's behaviour. *Economica*, 5(17):61–71, 1938.

Xiaoxia Shi, Matthew Shum, and Wei Song. Estimating semi-parametric panel multinomial choice models using cyclic monotonicity. *Econometrica*, 86(2):737–761, 2018.

Kenneth A Small and Harvey S Rosen. Applied welfare economics with discrete choice models. *Econometrica*, pages 105–130, 1981.

Thomas M Stoker. Empirical approaches to the problem of aggregation over individuals. *Journal of Economic literature*, 31(4):1827–1874, 1993.

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

Hal R Varian. Goodness-of-fit in optimizing models. *Journal of Econometrics*, 46 (1-2):125–140, 1990.

Hal R Varian. *Microeconomic analysis*. WW Norton, 1992.

E Glen Weyl. Price theory. *Journal of Economic Literature*, forthcoming.

Robert D Willig. Consumer's surplus without apology. *The American Economic Review*, 66(4):589–597, 1976.