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

September 24, 2018

Abstract

Applied research often tolerates misspecification in order to reach informative conclusions. This paper studies theoretically and empirically misspecification associated with structured preferences using a revealed preference approach. We focus on how the the degree of misspecification varies with the level of aggregation of the data. Using scanner data, we find that while *all* individuals are inconsistent with quasilinear utility, we cannot refute the hypothesis that a representative agent is a quasilinear utility maximizer. This provides empirical evidence that deviations from quasilinear utility may average away and that quasilinear utility may not be so bad for aggregated scanner data.

*We thank David Blau, Christopher P. Chambers, Ian Crawford, Paweł Dziewulski, Yuichi Kitamura, Lance Lochner, John Quah, Andres Santos, and Joel Sobel 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]*

A recent literature studies whether general demand models with unrestricted heterogeneity can describe the behavior of a population of individuals. For example, Kitamura and Stoye [forthcoming] 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.

To address this broad question, we focus on the restrictiveness of quasilinear utility at different levels of aggregation using a revealed preference approach. To assess the restrictiveness we construct a new measure of quasilinear misspecification, which can be used in a statistical or deterministic framework. Because policy questions often depend on a given level of aggregation, we pay particular attention to measuring quasilinear rationality at different levels of aggregation. In the empirical application, we examine the Stanford Basket Data and find a sharp contrast between individual and aggregate consumption: while *all* individuals depart from quasilinear utility, aggregate consumption is consistent with quasilinear utility using either a deterministic or statistical framework. This empirical finding supports the hypothesis of Becker [1962] that inconsistencies with a model at the individual level may vanish with aggregation.

The primary reasons we focus on quasilinear utility are that it is widely used, has tractable aggregation properties, and is in units of dollars. The fact that the units are dollars allows us to measure misspecification in interpretable units such as dollars lost. We measure misspecification by considering an individual (or the representative agent) as an approximate maximizer of a quasilinear utility function. 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. Thus, the units of ε^* may be interpreted as the maximal dollar loss due to approximate optimization or limited discernment. This measure is quickly computable by linear programming for a moderate-sized dataset.

Approximate quasilinearity has a natural interpretation directly in terms of observable prices and quantities. When there is a single good, an exactly quasilinear model is characterized by quantity decreasing with price. In an approximately quasilinear model, we show that it is possible to observe a small increase in price and a large increase in demand (Proposition 1). For larger increases in price, however, demand can increase at most by a small amount. Thus, our utility-based approach provides a notion of distance that qualitatively differs from notions based on e.g. Euclidean distance.

The measure of misspecification we propose is theoretically linked for individual and aggregate demands. In particular, if individual behavior is (approximately) quasilinear, then aggregate behavior will also be (approximately) quasilinear (Proposition 3). More specifically, we show that the measure of quasilinear misspecification weakly decreases when we aggregate individual demands, relative to the average of the individual measures of misspecification. This 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 do not have much structure even if *all* individuals act consistently with the model.[1]

One convenient feature of the measure of misspecification is that for aggregated data, it can be used whether one prefers a deterministic or statistical framework. We provide a statistical test of whether a representative agent has a degree of misspecification no greater than a given value ε. The test involves linear moment inequality restrictions involving the mean of quantities given prices. When prices take finitely

[1] 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 (cf. McFadden and Richter [1990] and McFadden [2005]).

many values, there are finitely many unconditional moment inequalities associated with a conjectured value of quasilinear rationality. Testing such inequalities is well-understood (e.g. Chernozhukov et al. [2014], Romano et al. [2014]). We also provide a two-sided confidence interval for the measure of misspecification ε^* for the aggregated population-level data. In our empirical application, despite the presence of many goods and a large number of moment inequalities, linear programming facilitates quick computation.[2]

This paper is interested in measuring the amount of misspecification of quasilinear utility at different levels of aggregation. In contrast, an existing literature is interested in testing whether models are exactly consistent with the data. 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.[3] In contrast, our test statistic measures dollar deviations from quasilinearity. Thus, our test statistic gives an economic measure of misspecification in the spirit of Varian [1990].

While we believe this statistical analysis is of independent interest for other panel or cross-sectional datasets, we did not need to invoke it in our empirical application. Our empirical application uses Stanford Basket Data, which has been used by Echenique et al. [2011]. 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 significant heterogeneity of approximation error at the individual level, the pattern of heterogeneity does not lead to a rejection of quasilinear utility in the aggregate.

Lastly, we study the relationship between individual-level measures of misspecifica-

[2]The number of inequalities in the linear program is the square of the number of time periods. Our empirical application has 375 goods and 26 time periods. In contrast, Kitamura and Stoye [forthcoming] reach the computational boundary with around 5 goods and 8 time periods.

[3]Our analysis is thus more nuanced than simply running a specification test. Our approach is related to work on sensitivity analysis in econometric models, e.g. Conley et al. [2012], Kline and Santos [2013], and Masten and Poirier [2018].

tion and expenditure. To compare with literature in the general consumer problem (e.g. Afriat [1973]), we consider a scaled measure of quasilinear misspecification that divides by total household expenditure, which is in units of percentage of the budget. Examining the cumulative distribution function according to Z-scores of our original measure and the scaled measure, we find these distributions are almost identical. In particular, we do not reject the null hypothesis of equality (p-value .40). This suggests that an analyst may reach the same conclusion about the shape of the distribution of misspecification regardless of whether the analyst divides by expenditure or not. In contrast, when we compare the measures of misspecification with household expenditure we find the proposed misspecification measure is positively correlated with expenditure while the scaled measure is negatively correlated. Thus, the choice to scale by expenditure affects whether a researcher concludes individuals with higher expenditure are "further" from quasilinearity.[4]

The remainder of the paper proceeds as follows. Section 2 provides definitions, characterizes the model of approximate quasilinearity, and provides intuition on which behavior is allowed in approximately quasilinear models. This section also gives an interpretation of approximation error as a measure of misspecification, describes computation of the measure, and shows that the approximation error of the representative agent is less than the average error of individuals. Section 3 details a statistical test for the representative agent and a method to construct confidence intervals. Section 4 provides an empirical assessment of quasilinear utility for individuals and the representative agent. Section 5 provides a review of the related literature. Section 6 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. 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

[4] See Choi et al. [2014] for analysis of the relationship between rationality and expenditure in the general consumer problem. They use a scaled measure of misspecification that is in units of percentage expenditure.

$u : \mathbb{R}^K_+ \to \mathbb{R}$ is a non-satiated function and y is interpreted as the numeraire good.[5] A numeraire good is one which has a price of one. For the remainder of the paper, we interpret the numeraire as wealth and say that utility is measured in dollars.

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

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

We allow y to be negative, 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.

We differ from previous work by studying when demand data is approximately quasilinear. In particular, we say a model is approximately quasilinear if the observed demand data is within ε dollars of a quasilinear utility maximizer. This relaxation of quasilinear utility allows small 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 is 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.

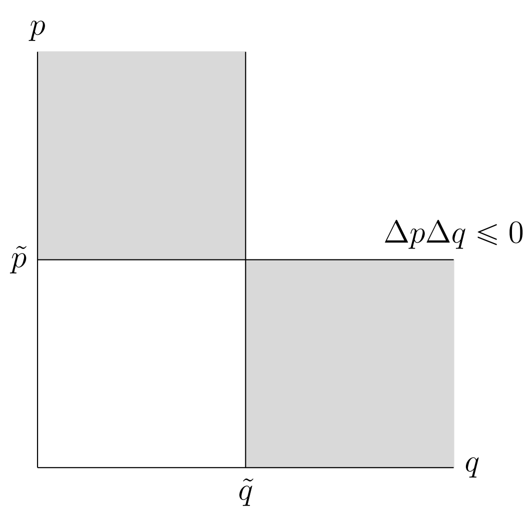
[5] We use $\mathbb{R}_+ = \{x \in \mathbb{R} \mid x \geqslant 0\}$ and $\mathbb{R}_{++} = \{x \in \mathbb{R} \mid x > 0\}$.

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.

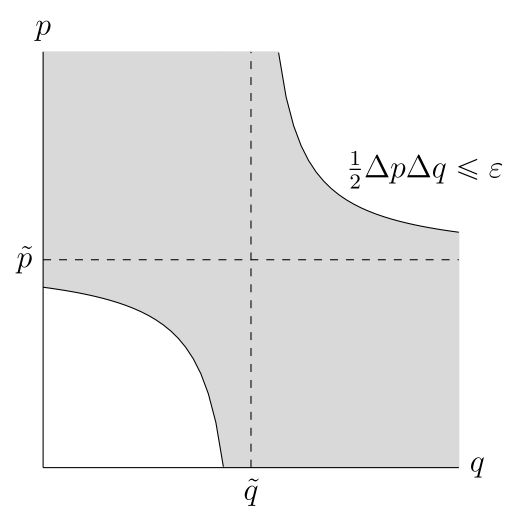
Proposition 1 (Approximate Law of Demand)**.** *If the dataset $\{(x^t, p^t)\}_{t=1}^T$ is ε-rationalized by quasilinear utility, then for any $r, s \in \{1, \ldots, T\}$ it follows that*

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

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



(a) Law of Demand



(b) Approximate Law of Demand

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 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. The inequality of (iii) is a requirement that the average money extracted from a "money pump" be less than ε.[6] 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 considered in Example 1.

Example 1. *Consider the dataset with $(q^1, p^1) = (2, 1)$ and $(q^2, p^2) = (6, 2)$. We*

[6] 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 $(x^{t_m} - x^{t_{m+1}})$ in Theorem 1 represents one such trade. More details on this and the relation to Echenique et al. [2011] are in Appendix B.

wish to check $\varepsilon = 1$ so that the average "money pump" must be less than one dollar. 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 is a natural measure of misspecification. In this paper, we denote ε^* as the smallest ε described above, and call it the measure of quasilinear misspecification. 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]). We provide a formal relationship between our measure of quasilinear misspecification and other measures in Appendix B. The following proposition gives a direct relation between the approximation error interpretation and the "money pump" interpretation.

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:*

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

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

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). Importantly, this proposition shows that the smallest ε that ε-quasilinear rationalizes the data can be found using linear programming.[7] We highlight that this measure differs from 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.[8]

2.2 Aggregation

It is well-known that quasilinear utility induces demand that can be aggregated across individuals.[9] We show a similar aggregation property holds when each individual dataset is ε^i-quasilinear rationalized and individuals face the same prices. Suppose that there are $i = 1, \ldots, n$ 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 ε^{i*} is the measure of misspecification for the dataset for individual i.

This result helps understand why aggregate datasets may rationalize data more often than individual-level data as explored in a related setting by Becker [1962]. Propo-

[7] One 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 $\varepsilon_V = (\varepsilon_V^1, \ldots, \varepsilon_V^T)$, other criterion functions such as the mean $\sum_{t=1}^T \varepsilon_V^t$ are also computable by linear programming. See Theorem 2 in Appendix B.

[8] See Appendix B for more details on the difference from Echenique et al. [2011].

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

sition 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 gap in our application.

Proposition 3(ii) 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. We note that for models that are not preserved under aggregation, such as the general consumer problem, there exist no such convex measures of misspecification.[10]

We give an example when average demand data is rationalized by quasilinear utility even when there are individuals whose datasets cannot be exactly rationalized by quasilinear utility.

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 our statistical analysis in Section 3. 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 $X^i = \left(X^{(i,1)\prime}, \dots, X^{(i,T)\prime}\right)'$. We let p^t be a predetermined (nonrandom) vector of prices for the K goods at observation t.

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

[10] To see this consider a collection of individual datasets that are rational at the individual level, but not at the aggregate. Any measure of misspecification violates convexity if we require that the measure maps to zero if and only if the data is consistent with the model.

(ii) The mapping $\varepsilon^* : \prod_{t=1}^{T} \mathbb{R}^K \to \mathbb{R}_+$ *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.

3 Inference and Testing

The analysis of Section 2.2 establishes that aggregation weakly decreases the measure of quasilinear misspecification, but we are also interested in conducting inference on this measure in a statistical framework. In this section, we describe how to test whether the representative agent is an ε-quasilinear maximizer We also show 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. 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.[11]

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 $j \in \{1,\ldots,J\}$.

[11] 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.

Associated with each moment condition, we form a sample average across individuals,

$$\begin{aligned}\hat{\mu}_j &= \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{M}\sum_{m=1}^{M} p^{t_m}\cdot\left(X^{(i,t_m)} - X^{(i,t_{m+1})}\right)\right]\\ &= \left[\frac{1}{M}\sum_{m=1}^{M} p^{t_m}\cdot\left(\overline{X}^{t_m} - \overline{X}^{t_{m+1}}\right)\right],\end{aligned} \tag{4}$$

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.

Recall that $X^i = \left(X^{(i,1)\prime}, \ldots, X^{(i,T)\prime}\right)'$ denotes the demands for individual i. Let $X = (X^{1\prime}, \ldots, X^{n\prime})'$ 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. In order to compute the term in brackets, Proposition 2 may be applied to the dataset $\left\{\left(\overline{X}^t, p^t\right)\right\}_{t=1}^{T}$, except without imposing the restriction $\varepsilon \geqslant 0$ in the linear program. The statistic S may thus be computed as a linear programming problem with T^2 inequality constraints. 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.

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. The critical value $c_{1-\alpha}$ is constructed in the following manner.[12] Let B be the number of bootstrap draws. In the application, we use $B = 5{,}000$.

i. Draw independent and identically distributed sample of size n uniformly from the individual demands $\{X^i\}_{i=1}^{n}$. Recall $X^i = \left(X^{(i,1)\prime}, \ldots, X^{(i,T)\prime}\right)'$ denotes the consumption vector across all time periods. Let $X^{*i} = \left(X^{*(i,1)\prime}, \ldots, X^{*(i,T)\prime}\right)'$ denote quantities for individual $*i$ in the bootstrap sample.[13]

[12] Appendix A.1 provides intuition for the validity of this test.

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

ii. Compute the sample mean of the moment conditions from the bootstrap draw via Equation 4, 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.

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 (On 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.[14] 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 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.[15]

Remark 2 (Inference on ε^*)*.* The previous arguments describe how to test whether the representative agent is ε-rationalizable by quasilinear utility and provides a one-sided confidence set for ε^*, defined as the maximum of zero and

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

where the maximum is taken over all distinct sequences. A two-sided confidence set for ε^* may be constructed as

$$CI_{1-\alpha} = \left[\max_{j\in\{1,\ldots,J\}} \hat{\mu}_j - c_{1-\alpha/2}, \max_{j\in\{1,\ldots,J\}} \hat{\mu}_j + c_{1-\alpha/2}\right] \cap [0,\infty),$$

where $c_{1-\alpha/2}$ is a bootstrap quantile computed as before, and where division by 2 arises because we now consider a two-sided confidence interval.[16] Refinements of this confidence interval (or our test previously presented) that incorporate a form

[14] 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].

[15] 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 Brown and Calsamiglia [2007] and the present paper, but for empirical implementation Shi et al. [2018] focus on sequences of length 2.

[16] See Appendix A.1 for the motivation behind this confidence interval.

of "moment selection" may be possible by appealing to the framework of Fang and Santos [forthcoming]. Recall that by Proposition 4, the mapping ε^* is convex in quantities, which implies directional differentiability (see Rockafellar [2015], Section 23 for a textbook reference). Thus, ε^* fulfills a key requirement of Fang and Santos [forthcoming].

4 Empirical Application

In this section, we report the results of our empirical application. 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. After brand aggregation there is a total of 375 unique goods. For full details on this dataset, we refer the reader to Echenique et al. [2011].

4.1 Aggregate Data

We implement the statistical test of Section 3 for aggregate data. We find that the average demands are consistent with maximization of quasilinear utility. In particular, if we specify $\varepsilon = 0$ and consider the null hypothesis that the average demands are consistent with quasilinear utility, then the test statistic $S(X)$ does not exceed zero. This means we do not need to compute critical values: we cannot reject the representative agent formulation with *either* a deterministic or statistical model.

The two-sided confidence interval for ε^*, calculated according to Remark 2 with $\alpha = .05$, is given by $[0, 2.69]$. In contrast, all households are inconsistent with quasilinear utility in a deterministic framework. In addition, the upper bound on the confidence interval, 2.69, is smaller than the smallest individual ε^{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 distribution of individual choices.

4.2 Household Data

We now study two measures of deviations from quasilinear utility with household data. The first is the smallest value ε^{i*} such that household i is ε^{i*}-quasilinear. The calculation of each household's ε^{i*} is described by Proposition 2.[17] The units of ε^{i*} are dollars.[18] The second measure we consider is the smallest ε^{i*} just described, but divided by the total expenditure (in the dataset) of household i across all time periods. The units of the latter measure are in terms of percentage of expenditure. We consider this second measure since several measures designed for measuring violations of the generalized axiom of revealed preference contain an inherent division by the level of expenditure. One example is Afriat's Efficiency Index (AEI) [Afriat, 1973]. Such measures are ratio-based, and are thus closer in spirit to scaled ε^{i*}.[19]

We focus on the individual measures of misspecification for three reasons. First, we want to contrast the aggregate and individual-level measures. Second, we examine individual measures to compare our results to those of Echenique et al. [2011]. We find results that are qualitatively similar to the analysis using the money pump. Lastly, we have intuition that how one introduces misspecification error may affect auxillary questions that arise from this analysis. For example, Choi et al. [2014] use the Afriat Efficiency Index to measure who is "more rational." We show that while the scaled and raw measures of misspecification, once standardized, have almost the same distributions, one can reach different conclusions concerning who is "more rational." We find that the raw measure is positively correlated with expenditure, while the scaled measure is negatively correlated with expenditure.

Table 1 reports summary statistics of ε^{i*} as well as its scaled counterpart. As men-

[17] We include observations from all time periods even when individuals have zero consumption. This allows us to abstract from potential endogeneity of when individuals purchase, and also allows us to compare directly with aggregate results.

[18] In our framework, the units are in terms of the price of the numeraire good. Prices were computed by taking a weighted average of shelf prices for the period of expenditure. The prices are in dollars per unit from Echenique et al. [2011] so we interpret this as dollars. We do not adjust for inflation.

[19] We provide additional discussion of ratio-based measures in Appendix B.

tioned previously, *none* of the households can be exactly described by a quasilinear utility model. The average of ε^{i*} is \$31.96. For comparison, the average total expenditure is \$5,557.50. When scaled by total expenditure, the average scaled ε^{i*} is .68%.[20] The highest scaled ε^{i*} is 2.43%.

Table 1: Summary Statistics of ε^{i*} and Scaled ε^{i*}

	Measures	
	ε^{*}	Scaled ε^{*}
Mean	31.96	0.68%
Median	28.59	0.60%
Min	5.73	0.13%
Max	156.42	2.43%
No. Households	494	494

Figure 2 presents the distribution of the smallest ε^{i*} such that a household is ε^{i*}-quasilinear rationalizable, as well as the distribution of its scaled counterpart. To standardize the two distributions, which are in different units, we report the Z-scores. Remarkably, the two distributions are almost identical. A Kolmogorov-Smirnov test fails to reject the null of equality of distribution at conventional significance levels (p-value .40). The shape of the distribution of misspecification thus does not depend on whether the analyst considers misspecification in terms of dollars lost or percentage of expenditure lost. We emphasize that quasilinear utility allows this comparison to be made. In contrast, conventional measures in the standard consumer problem, such as Afriat's Efficiency Index [Afriat, 1973], are in units of percentage of expenditure lost and do not permit an obvious interpretation in terms of (unscaled) dollars lost. Finally, we note the shape of both distributions is broadly consistent with Echenique et al. [2011], who quantify violations of the weak axiom of revealed preference.

We now analyze how the raw and scaled ε^{i*} of households vary with different levels of expenditure. This facilitates further interpretation of ε^{i*} and sheds light on the broad question of whether wealthier individuals are more consistent with economic

[20] We include scaled ε^{i*} because several existing measures of misspecification scale by expenditure. There are alternative ways to scale by expenditure. For example, one could scale ε^{i*} by dividing by a household's average monthly expenditure across the 26 time periods in the dataset. This alternative scaling would result in a scaled ε^{i*} that is 26 times bigger than the summary statistics reported in the second column in Table 1.

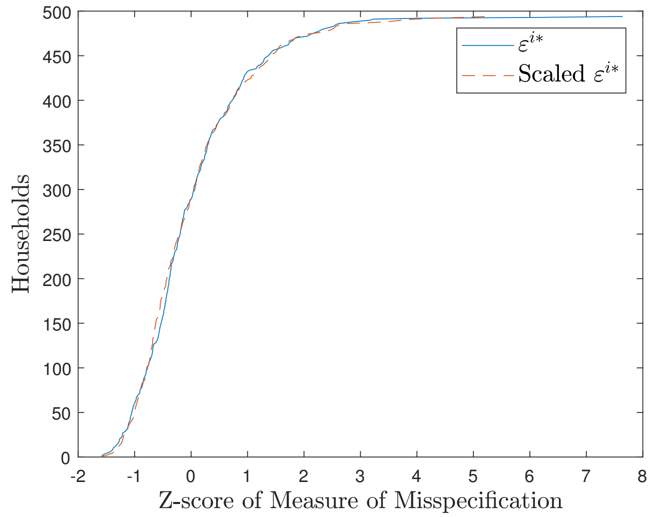


Figure 2: Cumulative Distribution Functions of Household ε^{i*}.

theories. While expenditure in the Stanford Basket dataset is *not* income, wealth, or total expenditure among all goods, we believe this analysis is still relevant for this broad measurement question.

We compare ε^{i*} and scaled ε^{i*} with total expenditure in Figure 3 and Figure 4 to assess how the two measures relate to total expenditure. We find a striking contrast between the two. While ε^{i*} is broadly increasing with expenditure, scaled ε^{i*} is broadly decreasing. Thus, one may reach different conclusions about the broad question of whether wealthier individuals are more consistent with economic theories, depending on whether one rescales by expenditure. This also contrasts with our finding in Figure 2 that the shapes of the (standardized) distributions of ε^{i*} and scaled ε^{i*} are almost identical.

In Figure 3, we plot ε^{i*} against total expenditure for household i. We find that ε^{i*} broadly increases with expenditure, suggesting that wealthier households may make bigger mistakes or be less rational. There are several possible explanations for this phenomenon. One possibility is that individuals with higher expenditure have higher wages, and that the opportunity cost of their time is greater. Thus, optimization may be relatively more costly, and so they are more prone to making mistakes or otherwise acting in a manner inconsistent with quasilinear utility. An alternative possibility is that individuals with higher expenditures simply make more purchases, so that they have more draws of "taste shocks" or chances to make a mistake, and thus have a

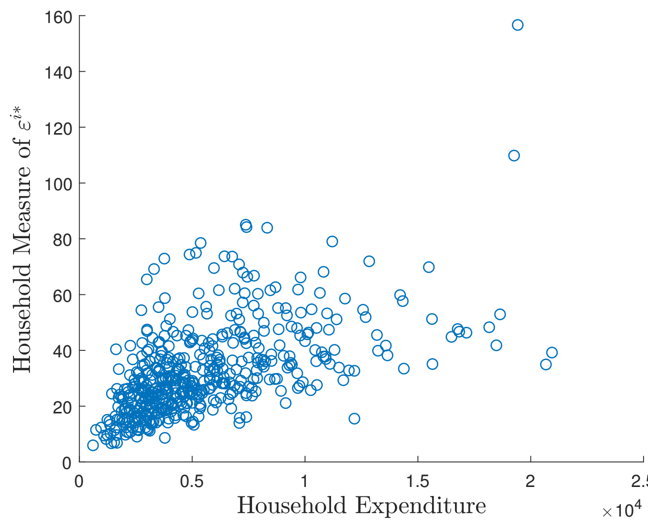


Figure 3: Household ε^{i*} vs. Expenditure.

greater possibility of having a higher ε^{i*}.

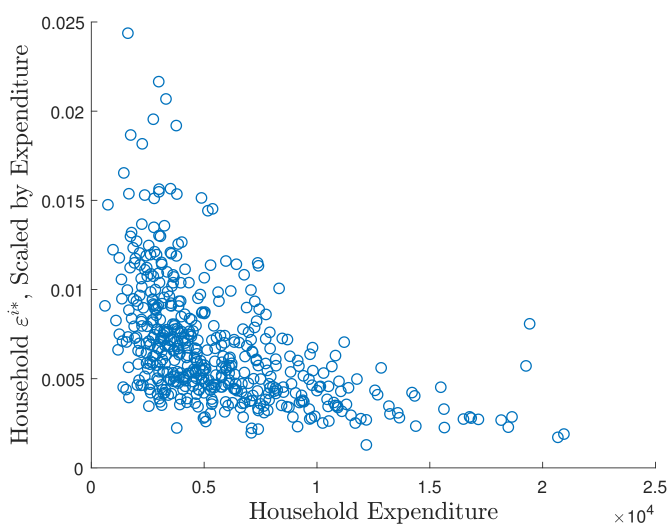


Figure 4: Household Scaled ε^{i*} vs. Expenditure.

In Figure 4, we plot scaled ε^{i*} (which is divided by total expenditure) against total expenditure. The plot is broadly decreasing, in contrast with our previous measure. This produces a finding similar to Choi et al. [2014] who find in experimental data that wealthier individuals have smaller violations of utility maximization according to the Afriat Efficiency Index. This highlights that difference-based measures of misspecification such as ε^{i*} may behave very differently from ratio-based measures such as scaled ε^{i*}.[21]

5 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.[22] 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 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]).[23]

[21] Graphical inspection of ε^{i*} versus scaled ε^{i*} indicates there is no systematic relationship between the two measures. The corresponding figure is available upon request.

[22] See also Grandmont [1992]

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

We provide a detailed relation to the literature on measuring misspecification in Appendix B. One key difference of is that quasilinear utility permits one to compare a level-based measure (units of dollars lost) and a ratio-based measure (units of percentage of total expenditure) of misspecification. We exploit this in Section 4.2, finding that these different scalings suggest very different answers to the question, "Are wealthier individuals more rational," (cf. Choi et al. [2014]).

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 with quasilinear utility with multiple goods.[24] 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 3. The linear inequality restrictions we test have a similar structure to those in Kitamura and Stoye [forthcoming], Deb et al. [2018], and Cattaneo et al. [2017],[25] though our methods are distinct.

[24]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 univariate demands, monotonicity in own-price is the only restriction of quasilinear utility.

[25]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.

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.[26]

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 aggregate datasets. These papers examine conditions for when 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.

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

6 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.

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 a large amount of heterogeneity of 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 in applied work 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. 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.[27] This paper raises natural follow-up questions concerning welfare and prediction that we are investigating in ongoing work.

[27] The most time-intensive component of our empirical application involved calculation of bootstrap quantiles. This took 316 seconds with 5,000 bootstrap draws with an Intel Core i7-7700 processor, without parallelizing.

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

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

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. □

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). Define

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

which is motivated by re-arranging (iii). By construction, $U(x)$ 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 $U(x^t) = \left(x^t - x^{\sigma_t(M_t)}\right) \cdot p^{\sigma_t(M_t)} + \sum_{m=1}^{M_t-1} (x^{\sigma_t(m+1)} - x^{\sigma_t(m)}) \cdot p^{\sigma_t(m)} + M_t\varepsilon$ 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. □

Proof of Proposition 2. For the dataset $\{(x^t, p^t)\}_{t=1}^T$, we first show existence of a solution ε^* to the linear programming problem while dropping the restriction $\varepsilon \geqslant 0$,[28]

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

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

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

[28] 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 $\varepsilon \geqslant 0$.

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 rationalization by Theorem 1 (parts (i) and (iii)). □

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

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

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

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 weighted averages we obtain

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

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. □

A.1 Motivation for Test and Confidence Interval

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

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

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.

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

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. 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

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

Thus, the test (approximately) controls size.

We now provide the motivation for the two-sided confidence interval

$$CI_{1-\alpha} = \left[\max_{j \in \{1,\ldots,J\}} \hat{\mu}_j - c_{1-\alpha/2}, \max_{j \in \{1,\ldots,J\}} \hat{\mu}_j + c_{1+\alpha/2}\right] \cap [0, \infty).$$

By Theorem 1, the measure of misspecification ε^* for the representative agent dataset is equal to the maximum of $\max_{j \in \{1,\ldots,J\}} \mu_j$ and zero. A confidence set for ε^* can thus be constructed by intersecting a confidence set for $\max_{j \in \{1,\ldots,J\}} \mu_j$ with the set of non-negative real numbers.

We assume $c_{1-\alpha/2}$ satisfies the approximation

$$P(\forall j, |\hat{\mu}_j - \mu_j| \leqslant c_{1-\alpha/2}) \approx 1 - \alpha,$$

i.e. this critical value is suitable for a two-sided confidence set for the entire vector $\mu = (\mu_1, \ldots, \mu_J)$.

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

$$\begin{aligned}\left\{\forall j, \hat{\mu}_j - \mu_j \geqslant -c_{1-\alpha/2}\right\} &\subseteq \{\mu_{j*} \leqslant \hat{\mu}_{j*} + c_{1-\alpha/2}\} \\ &\subseteq \left\{\max_{j \in \{1,\ldots,J\}} \mu_j \leqslant \max_{j \in \{1,\ldots,J\}} \hat{\mu}_j + c_{1-\alpha/2}\right\}.\end{aligned}$$

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

$$\begin{aligned}\left\{\forall j, \hat{\mu}_j - \mu_j \leqslant c_{1-\alpha/2}\right\} &\subseteq \left\{\forall j, \hat{\mu}_j - \mu_{j*} \leqslant c_{1-\alpha/2}\right\} \\ &= \left\{\max_{j \in \{1,\ldots,J\}} \mu_j \geqslant \max_{j \in \{1,\ldots,J\}} \hat{\mu}_j - c_{1-\alpha/2}\right\}.\end{aligned}$$

Thus,

$$P\left(\max_{j\in\{1,\ldots,J\}}\hat{\mu}_j - c_{1-\alpha/2} \leqslant \max_{j\in\{1,\ldots,J\}}\mu_j \leqslant \max_{j\in\{1,\ldots,J\}}\hat{\mu}_j + c_{1-\alpha/2}\right)$$
$$\geqslant P(\forall j, |\hat{\mu}_j - \mu_j| \leqslant c_{1-\alpha/2}) \approx 1-\alpha.$$

Appendix B Relation to Other Measures

Here we discuss how the measure of model misspecification we study relates to some existing measures. First, we examine the relation between the standard consumer problem and the quasilinear consumer problem. The standard consumer problem is characterized by the generalized axiom of revealed preference (GARP) which we define later. We then focus the various ways to measure violations of GARP developed in Afriat [1973], Varian [1990], Halevy et al. [2018], and Echenique et al. [2011]. We focus on these measures since they are most closely related to the measure we propose in the paper. Finally, we compare the existing measures to the measure studied in the paper.

B.1 Standard Consumer Problem and Quasilinear Consumer Problem

Analysis of the standard consumer problem without imposing quasilinearity often uses the same dataset of choices $\{(x^t, p^t)\}_{t=1}^T$. In other words, the numeraire is not treated separately. However, there are important distinctions between the standard consumer problem and a quasilinear consumer problem. We present the formal statement of the two models in Table 2.

Table 2: Versions of the Consumer Problem

Standard Consumer Problem	Quasilinear Consumer Problem
$\max_{x\in\mathbb{R}^K_+} U(x)$ s.t. $p\cdot x \leqslant I$	$\max_{x\in\mathbb{R}^K_+, y\in\mathbb{R}} U(x)+y$ s.t. $p\cdot x + y \leqslant I$

An important difference between the models is that in practice, prices in the standard consumer problem are purely relative prices. This is because income is typically constructed from expenditure $p^t \cdot x^t$, so any rescaling of p^t will not affect analysis. Due to this invariance, measures of the misspecification of GARP have centered on those in units of percentage of expenditure. This makes it difficult to quantify the severity of a GARP violation, however research by Echenique et al. [2011] gives one way to measure the severity.

In contrast, prices in the quasilinear problem have a known scale because the price of the numeraire good (y) has been set to 1. Moreover, since the numeraire good is additively separable in utility, it is possible to have a difference-based measure of misspecification such as ε^*, in units of dollars lost (not percentage of expenditure lost). If one is not careful with the labeling of units of prices and normalizations, it is easy to make mistakes in practice. Whether one uses the (static) standard consumer problem or the quasilinear consumer problem, one typically makes an implicit separability assumption considering unobserved goods.

B.2 Measuring Violations of the Standard Consumer Problem

Here we review existing research that measures violations of the standard consumer problem. First, the standard consumer problem is characterized by the general axiom of revealed preference (GARP), which has motivated existing measures. To define GARP, we first introduce the revealed preference relation in the standard consumer problem. For the dataset $\{(x^t, p^t)\}_{t=1}^T$, we say that x^r is *revealed preferred* to x^s if $p^r \cdot x^s \leqslant p^r \cdot x^r$, so that x^s is weakly less expensive than x^r at prices p^r. We denote that x^r is revealed preferred to x^s using the notation $x^r \, R \, x^s$. Similarly, we say x^r is *strictly revealed preferred* to x^s if $p^r \cdot x^s < p^r \cdot x^r$, so that x^s is strictly less expensive than x^r at prices p^r. We denote that x^r is strictly revealed preferred to x^s using the notation $x^r \, P \, x^s$. Using these definitions, GARP states that for every sequence $\{t_m\}_{m=1}^M$ with $t_m \in \{1, \ldots, T\}$, the following implication holds:

$$x^{t_1} \, R \, x^{t_2} \; \ldots \; x^{t_m} \, R \, x^{t_{m+1}} \; \ldots \; x^{t_{M-1}} \, R \, x^{t_M} \implies \text{not } x^{t_M} \, P \, x^{t_1}.$$

We also say that the ordered pair (R, P) is *acyclic* if GARP is satisfied. If GARP is violated, then there is a way to construct a cycle of consumption bundles where each should be revealed preferred to one another with at least one entry strict.

The measures of rationality for the standard consumer problem try to measure the size of GARP violations. We first present Afriat's Efficiency Index (AEI) [Afriat, 1973]. AEI measures the amount one needs to relax the revealed preference relation R to remove violations of GARP. Let $e_{\mathrm{A}} \in [0, 1]$ and define the relaxed revealed preference relation $R(e_{\mathrm{A}})$ as

$$x^r \; R\,(e_{\mathrm{A}}) x^s \quad \text{if and only if} \quad p^r \cdot x^s \leqslant e_{\mathrm{A}} p^r \cdot x^r.$$

We also define a relaxed strict preference $P(e_{\mathrm{A}})$ as

$$x^r \; P\,(e_{\mathrm{A}}) x^s \quad \text{if and only if} \quad p^r \cdot x^s < e_{\mathrm{A}} p^r \cdot x^r.$$

One can then check whether $(R(e_{\mathrm{A}}), P(e_{\mathrm{A}}))$ is acyclic instead of checking acyclicity of (R, P). For an example, if $e_{\mathrm{A}} = 0$, then the revealed preference relation is vacuous and any dataset satisfies acyclicity. Similarly, if $e_{\mathrm{A}} \approx 1$ the violations of GARP can be made arbitrarily small. AEI is defined as

$$\begin{aligned} \mathrm{AEI}\left(\{x^t, p^t\}_{t=1}^T\right) = & \sup_{e_{\mathrm{A}} \in [0,1)} \; e_{\mathrm{A}} \\ & \text{s.t.} \quad (R(e_{\mathrm{A}}), P(e_{\mathrm{A}})) \quad \text{is acyclic.} \end{aligned}$$

In other words, when AEI is one there is (almost) no misspecification error in the standard consumer problem.[29] Since the measure e_A affects the only the budget, a natural interpretation is relaxation of the revealed preference as a percentage of expenditure. Dziewulski [2018] shows how AEI relates to a notion of just-noticable-differences.[30]

A natural criticism of this measure is that e_{A} must give a relaxation for *all* GARP violations, even if only a single choice is inconsistent with GARP. Thus, the AEI is

[29] There are instances where GARP is violated even when $\mathrm{AEI}\left(\{x^t, p^t\}_{t=1}^T\right) = 1$.

[30] A just-noticable-difference model considers utilities when the choice x an individual makes from a budget B must satisfy $u(x) + \delta(x) \geqslant u(y)$ for all $x' \in B$. The δ function here is naturally related to the measure ε^* studied in this paper.

often smaller than the *average* relaxation of the revealed preference relation needed to remove GARP cycles. To account for this, Varian [1990] considers an alternative measure that allows a different relaxation of the revealed preference relation for each choice. To formally define this measure, let $e_{\mathrm{V}} = (e_{\mathrm{V}}^1, \ldots, e_{\mathrm{V}}^T) \in [0,1]^T$. Define the relaxed revealed preference relation $R(e_{\mathrm{V}})$ as

$$x^r \; R\,(e_{\mathrm{V}})x^s \quad \text{if and only if} \quad p^r \cdot x^s \leqslant e_{\mathrm{V}}^r p^r \cdot x^r.$$

Similarly, define the relaxed strict preference $P(e_{\mathrm{V}})$ as

$$x^r \; P\,(e_{\mathrm{V}})x^s \quad \text{if and only if} \quad p^r \cdot x^s < e_{\mathrm{V}}^r p^r \cdot x^r.$$

Varian [1990] suggests using the average quadratic distance (AQD) of the relaxations as a measure of rationality. Formally,

$$\begin{aligned} \mathrm{AQD}\left(\{x^t, p^t\}_{t=1}^T\right) = & \inf_{e_{\mathrm{V}} \in [0,1)^T} \frac{1}{T} \sum_{t=1}^T (1 - e_{\mathrm{V}}^t)^2 \\ & \text{s.t.} \quad (R(e_{\mathrm{V}}), P(e_{\mathrm{V}})) \quad \text{is acyclic.} \end{aligned}$$

The paper by Halevy et al. [2018] extends the ideas of Varian [1990] to show one can use arbitrary aggregators, relates e_{V} to money metric violations, and provides a decomposition for model misspecification from functional form restrictions.

Of the previous two measures, AEI is the most commonly reported in experiments.[31] However, it has been argued that AEI does not adequately capture the "strength" of GARP violations. This critique led to renewed interest in measures of rationality with one of the most well-known being the Money Pump Index (MPI) proposed in Echenique et al. [2011]. For the dataset $\{(x^t, p^t)\}_{t=1}^T$, the MPI is defined for any sequence $\{t_m\}_{m=1}^M$ with $t_m \in \{1, \ldots, T\}$ as

$$\mathrm{MPI}\left(\{(x^{t_m}, p^{t_m})\}_{m=1}^M\right) = \frac{\sum_{m=1}^M p^{t_m} \cdot (x^{t_m} - x^{t_{m+1}})}{\sum_{m=1}^M p^{t_m} \cdot x^{t_m}}$$

where $t_{M+1} = t_1$. This measure is in units of percentage expenditure, similar to AEI.

[31] Choi et al. [2007] report AEI when evaluating decisions under risk and Andreoni and Miller [2002] report AEI when evaluating preferences for altruism.

Importantly, the MPI can be written for any cycle of observations not just those that violate GARP. Thus, even when GARP is satisfied, it is still possible for the money pump to be positive for some other sequence. Echenique et al. [2011] propose examining different statistics of the MPI *conditioning on sequences that violate* GARP. When a sequence violates GARP, it guarantees that the MPI for that sequence is positive. The idea behind the MPI is that if one has a preference cycle, an arbitrager could buy goods at some price and sell them back to the agent to extract profit from the individual. It is important to consider MPI for general cycles because a positive MPI indicates that an arbitrager could extract money from the individual, regardless of whether that cycle violates GARP. Therefore, the interpretation of statistics of MPI while restricting to GARP violations considers the existence of an arbitrager who is restricted to buying goods over GARP cycles.

We briefly note some other research on the MPI. The computational properties of different summary statistics of the MPI were examined in Smeulders et al. [2013]. Dean and Martin [2016] empirically examine a minimal cost index which essentially combines the measure of Echenique et al. [2011] with the idea of looking at minimal consistent subset of data consistent with rationality from Houtman and Maks [1985].

B.3 Measuring Violations of Quasilinear Utility

We now shift focus to quasilinear utility. Quasilinear utility was first characterized using a revealed preference approach in Brown and Calsamiglia [2007]. There are also other characterizations of quasilinear utility that focus on the relation to translation invariance [Chambers et al., 2016, Castillo and Freer, 2016]. We focus on the result from Brown and Calsamiglia [2007], which says that a dataset $\{(x^t, p^t)\}_{t=1}^T$ is rationalized by a quasilinear utility if and only if it satisfies *cyclic monotonicity* (CM). Cyclic monotonicity requires for all finite sequences $\{t_m\}_{m=1}^M$ with $t_m \in \{1, \ldots, T\}$ and with $M \geqslant 2$, the inequality

$$\sum_{m=1}^{M} p^{t_m} \cdot (x^{t_m} - x^{t_{m+1}}) \leqslant 0$$

holds, where $(x^{t_{M+1}}, p^{t_{M+1}}) = (x^{t_1}, p^{t_1})$. Note that this term is the numerator of the MPI for this sequence. In other words, cyclic monotonicity is equivalent to stating that the MPI negative for *any* cycles, not just those that violate GARP. This provides a natural link between cyclic monotonicity and the MPI of Echenique et al. [2011], which we will investigate further.

As shown in Heufer and Hjertstrand [2017] for homothetic preferences, the AEI used to relax revelaed preferences can naturally be adapted when studying preferences with additional structure. We describe how this may be done for quasilinear preferences. Let $e_{A-q} \in [0,1]$ be a quasilinear Afriat efficiency measure and consider $\mathrm{CM}(e_{\text{A-q}})$, the relaxed version of quasilinearity that requires that for all finite sequences $\{t_m\}_{m=1}^{M}$ with $t_m \in \{1,\ldots,T\}$ and with $M \geqslant 2$, the inequality

$$\sum_{m=1}^{M} p^{t_m} \cdot (e_{\text{A-q}} x^{t_m} - x^{t_{m+1}}) \leqslant 0$$

holds, where $(x^{t_{M+1}}, p^{t_{M+1}}) = (x^{t_1}, p^{t_1})$. This inequality is equivalent to

$$e_{\text{A-q}} \sum_{m=1}^{M} p^{t_m} \cdot x^{t_m} \leqslant \sum_{m=1}^{M} p^{t_m} \cdot x^{t_{m+1}}.$$

The left hand side may be interpreted as the total adjusted expenditure for the cycle of choices made. The right hand side may be interpreted as the total expenditure for alternative choices when swapping x^{t_m} for $x^{t_{m+1}}$. Thus, the inequality states that the adjusted actual expenditure is smaller than the total expenditure with these alternative choices.

One can further rearrange this condition to obtain for all finite sequences $\{t_m\}_{m=1}^{M}$ with $t_m \in \{1,\ldots,T\}$ and with $M \geqslant 2$, the inequality

$$\frac{\sum_{m=1}^{M} p^{t_m} \cdot (x^{t_m} - x^{t_{m+1}})}{\sum_{m=1}^{M} p^{t_m} \cdot x^{t_m}} \leqslant (1 - e_{\text{A-q}}) \tag{5}$$

holds, where $(x^{t_{M+1}}, p^{t_{M+1}}) = (x^{t_1}, p^{t_1})$. Thus, we exactly recover the notion of the MPI albeit on more familiar ground. This formally establishes that when the MPI is not restricted to GARP cycles, it measures violations of quasilinear rationality along the lines of AEI. Alternatively, when MPI is restricted to GARP-violating cycles, it can be seen as a "quasilinear" way of measuring violations of GARP.

Similar to finding the AEI from GARP, we look at for the maximal e_{A-q} while imposing CM($e_{\text{A-q}}$). We define the quasilinear efficiency index (QEI):

$$\text{QEI}\left(\{x^t, p^t\}_{t=1}^T\right) = \max_{e_{\text{A-q}} \in [0,1]} e_{\text{A-q}}$$
$$\text{s.t.} \quad \text{CM}(e_{\text{A-q}}) \text{ holds.}$$

A maximum exists by arguments similar to those in Proposition 2.

The above analysis can be adapted to allow a different approximation error for each choice, along the lines of Varian's AQD. To this end, let $e_{\text{V-q}} = (e^1_{\text{V-q}}, \ldots, e^T_{\text{V-q}}) \in [0,1]^T$. Attempting to relate this to the money pump, we obtain the alternative condition CM($e_{\text{V-q}}$), which requires that for all finite sequences $\{t_m\}_{m=1}^M$ with $t_m \in \{1, \ldots, T\}$ and with $M \geqslant 2$, the inequality

$$\sum_{m=1}^{M} p^{t_m} \cdot (x^{t_m} - x^{t_{m+1}}) \leqslant \sum_{m=1}^{M} p^{t_m} \cdot x^{t_m}(1 - e^{t_m}_{\text{V-q}})$$

holds, where $(x^{t_{M+1}}, p^{t_{M+1}}) = (x^{t_1}, p^{t_1})$. In other words, this scales the violations at observation t by the income spent on observed goods in that period. One could then consider minimizing the violations allowed by $e_{\text{V-q}}$ using either the average quadratic distance proposed by Varian [1990] or some other aggregator as in Halevy et al. [2018]. This clarifies the relation of the MPI to the efficiency measures of Afriat and Varian.

We now relate existing measures with the measure of misspecification for quasilinear utility, ε^*, defined in Proposition 2. The natural measure of comparison for ε^* is the QEI. To that end, assume $e^*_{\text{A-q}}$ is the value that maximizes QEI and let $\{t^m\}_{m=1}^{M^*}$ be a binding sequence for the maximized QEI. It follows that

$$M^* \varepsilon^* \geqslant \sum_{m=1}^{M^*} p^{t_m} \cdot x^{t_m}(1 - e^*_{\text{A-q}})$$

for a sequence that binds.[32] Similarly, let $\{t^{\tilde{m}}\}_{\tilde{m}=1}^{\tilde{M}}$ be a sequence such that the

[32] This inequality follows since if $M^*\varepsilon^* < \sum_{m=1}^{M^*} p^{t_m} \cdot x^{t_m}(1 - e^*_{\text{A-q}})$, one obtains the contradiction that

$$\sum_{m=1}^{M^*} p^{t_m} \cdot (x^{t_m} - x^{t_{m+1}}) \leqslant M^*\varepsilon^* < \sum_{m=1}^{M^*} p^{t_m} \cdot x^{t_m}(1 - e^*_{\text{A-q}}) = \sum_{m=1}^{M^*} p^{t_m} \cdot (x^{t_m} - x^{t_{m+1}}),$$

minimum approximation error ε^* from Proposition 2 binds. By construction, ε^* satisfies

$$\tilde{M}\varepsilon^* \leqslant \sum_{\tilde{m}=1}^{\tilde{M}} p^{t_{\tilde{m}}} \cdot x^{t_{\tilde{m}}}(1 - e^*_{\text{A-q}})$$

for a sequence that binds.[33]

Therefore, by rearrangement we have

$$\frac{\sum_{\tilde{m}=1}^{\tilde{M}}(p^{t_{\tilde{m}}} \cdot x^{t_{\tilde{m}}} - \varepsilon^*)}{\sum_{\tilde{m}=1}^{\tilde{M}} p^{t_{\tilde{m}}} \cdot x^{t_{\tilde{m}}}} \leqslant e^*_{\text{A-q}} \leqslant \frac{\sum_{m=1}^{M*}(p^{t_m} \cdot x^{t_m} - \varepsilon^*)}{\sum_{m=1}^{M*} p^{t_m} \cdot x^{t_m}}$$

for any pair of binding sequences. In other words, the quasilinear efficiency index e^*_{A-q} can be bounded by the fraction of expenditure remaining after extracting ε^* each period. The fact that these numbers are not always equivalent can be seen in example datasets. However, if $p^t \cdot x^t = 1$ for each t, then $\varepsilon^* = 1 - e^*_{\text{A - q}}$. A potential wedge emerges because with quasilinear demand, the model is not invariant to rescaling the price vector p^t of the non-numeraire goods. This shows that if one mistakenly normalizes income to one while checking for ε^*, then the QEI and the measure of misspecification for quasilinear utility are exactly related.

B.4 ε_V-quasilinear Rationalizability

As in Varian [1982], we could have proposed a measure $\varepsilon_{\text{V}} = (\varepsilon^1_{\text{V}}, \ldots, \varepsilon^T_{\text{V}}) \in \mathbb{R}^T_+$ 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 $\varepsilon_V = (\varepsilon^1_V, \ldots, \varepsilon^T_V) \in \mathbb{R}^T_+$ if there exists a locally non-satiated utility function $u : \mathbb{R}^K_+ \to \mathbb{R}$*

where the last equality follows from Equation 5 binding.

[33]This inequality follows since if $\tilde{M}\varepsilon^* > \sum_{\tilde{m}=1}^{\tilde{M}} p^{t_{\tilde{m}}} \cdot x^{t_{\tilde{m}}}(1 - e^*_{\text{A-q}})$, one obtains the contradiction that

$$\sum_{\tilde{m}=1}^{\tilde{M}} p^{t_{\tilde{m}}} \cdot (x^{t_{\tilde{m}}} - x^{t_{\tilde{m}+1}}) = \tilde{M}\varepsilon^* > \sum_{\tilde{m}=1}^{\tilde{M}} p^{t_{\tilde{m}}} \cdot x^{t_{\tilde{m}}}(1 - e^*_{\text{A-q}}) \geqslant \sum_{\tilde{m}=1}^{\tilde{M}} p^{t_{\tilde{m}}} \cdot (x^{t_{\tilde{m}}} - x^{t_{\tilde{m}+1}}),$$

where the first equality follows from Equation 2 binding.

such that for all $t \in \{1, \ldots, T\}$ and for all $x \in \mathbb{R}^K_+$, the following inequality holds:

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

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 $\{(x^t, p^t)\}^T_{t=1}$ and $\varepsilon_V \in \mathbb{R}^T_+$, the following are equivalent:*

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

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

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

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

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

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

(iv) $\{(x^t, p^t)\}^T_{t=1}$ 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^t_V in each time period. Moreover, if one considers the smallest ε_V that minimizes the average approximation errors $f(\varepsilon_V) = \frac{1}{T}\sum_{t=1}^{T} \varepsilon^t_V$, 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]. The way in which the smallest ε_{V} and a $e_{\text{V-q}}$ are related will depend on the function used to aggregate errors.

References

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.

James Andreoni and John Miller. Giving according to garp: An experimental test of the consistency of preferences for altruism. *Econometrica*, 70(2):737–753, 2002.

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

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

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

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

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

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.

Christopher P Chambers, Federico Echenique, and Kota Saito. Testing theories of financial decision making. *Proceedings of the National Academy of Sciences*, 113 (15):4003–4008, 2016.

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.

Syngjoo Choi, Raymond Fisman, Douglas Gale, and Shachar Kariv. Consistency and heterogeneity of individual behavior under uncertainty. *American Economic Review*, 97(5):1921–1938, 2007.

Syngjoo Choi, Shachar Kariv, Wieland Müller, and Dan Silverman. Who is (more) rational? *American Economic Review*, 104(6):1518–50, 2014.

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

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

Mark Dean and Daniel Martin. Measuring rationality with the minimum cost of revealed preference violations. *Review of Economics and Statistics*, 98(3):524–534, 2016.

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.

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

Jan Heufer and Per Hjertstrand. Homothetic efficiency: Theory and applications. *Journal of Business & Economic Statistics*, pages 1–13, 2017.

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

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

Martijn Houtman and J Maks. Determining all maximal data subsets consistent with revealed preference. *Kwantitatieve Methoden*, 19(1):89–104, 1985.

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

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*, 2018.

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.

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.

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

Bart Smeulders, Laurens Cherchye, Frits CR Spieksma, and Bram De Rock. The money pump as a measure of revealed preference violations: A comment. *Journal of Political Economy*, 121(6):1248–1258, 2013.

Hal R Varian. The nonparametric approach to demand analysis. *Econometrica*, pages 945–973, 1982.

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.