REVEALED PRICE PREFERENCE: THEORY AND EMPIRICAL ANALYSIS

RAHUL DEB[◊], YUICHI KITAMURA[†], JOHN K.-H. QUAH[‡], AND JÖRG STOYE[⋆]

JANUARY 10, 2018

ABSTRACT. We develop a model of demand where consumers trade-off the utility of consumption against the disutility of expenditure. This model is appropriate whenever a consumer's demand over a *strict* subset of all available goods is being analyzed. Data sets consistent with this model are characterized by the absence of revealed preference cycles over prices. For the random utility extension of the model, we devise nonparametric statistical procedures for testing and welfare comparisons. The latter requires the development of novel tests of linear hypotheses for partially identified parameters. Our applications on national household expenditure data provide support for the model and yield informative bounds concerning welfare rankings across different prices.

1. INTRODUCTION

Imagine a consumer who is asked what quantity she will purchase of L goods at given prices; in formal terms, she is asked to choose a bundle $x^t \in \mathbb{R}^L_+$ at the price vector $p^t \in \mathbb{R}^L_{++}$. To fix ideas, we could think of these goods as grocery items and x^t as the monthly purchase of groceries if p^t are the prevailing prices. With T such observations, the data set collected is $\mathcal{D} = \{(p^t, x^t)\}_{t=1}^T$. What patterns of choices in $\mathcal{D}$ should we expect to observe?

There are at least two ways to approach this question. If, at observations t and t', we find that $p^{t'} x^t < p^{t'} x^{t'}$, then

the consumer has revealed that she strictly prefers the bundle $x^{t'}$ to the bundle x^t.

[◊]UNIVERSITY OF TORONTO, [†]YALE UNIVERSITY, [‡]JOHNS HOPKINS UNIVERSITY, [⋆]CORNELL UNIVERSITY, UNIVERSITY OF BONN AND HAUSDORFF CENTER FOR MATHEMATICS.

E-mail addresses: rahul.deb@utoronto.ca, yuichi.kitamura@yale.edu, john.quah@economics.ox.ac.uk, stoye@cornell.edu.

We are very grateful for the comments provided by the seminar participants at the Cowles Foundation Conferences (April 2015, June 2015), University of Toronto (March 2016), Heterogeneity in Supply and Demand Conference at Boston College (December 2016), Identification of Rationality in Markets and Games Conference at Warwick University (March 2017), University of Cambridge (March 2017), UCSD (April 2017), UCLA (April 2017), the Canadian Economic Theory Conference (May 2017), "Meeting in Econometrics" at TSE (May 2017), Iowa State University (September 2017) and Boston University (October 2017). We also thank Samuel Norris and Krishna Pendakur for generously sharing their data with us. Deb thanks the SSHRC for their generous financial support and the Cowles foundation for their hospitality. Stoye acknowledges support from NSF grant SES-1260980 and through Cornell University's Economics Computer Cluster Organization, which was partially funded through NSF Grant SES-0922005. We thank Matthew Thirkettle for excellent research assistance.

If this were not true, then either x^t or a nearby bundle would be strictly preferred to $x^{t'}$ and cost less, which means the choice of $x^{t'}$ is not optimal. The standard revealed preference theory of consumer demand is built on requiring that this preference over grocery bundles, as revealed by a data set such as $\mathcal{D}$, is internally consistent. In particular, Afriat (1967)'s Theorem says that so long as the consumer's revealed preferences over bundles do not contain cycles (a property known as the *generalized axiom of revealed preference* or GARP, for short) then there is a strictly increasing utility function $\widetilde{U} : \mathbb{R}^L_+ \to \mathbb{R}$ such that x^s maximizes $\widetilde{U}(x)$ in the budget set $\{x \in \mathbb{R}^L_+ : p^s x \leq p^s x^s\}$, for every observation $s = 1, 2, ..., T$.[1] Notice that this theory implicitly assumes that it makes sense to speak of the consumer's preference over groceries, independently of her consumption of other goods, currently or in the future. In formal terms, this requires that the consumer has a preference over grocery bundles that is *weakly separable* from her consumption of all other goods.[2]

But, presented with the same data set $\mathcal{D} = \{(p^t, x^t)\}_{t=1}^T$, it would be entirely natural for us (as observers) to think along different lines; instead of inferring the consumer's preference over grocery bundles from the observations, we could draw conclusions about the consumer's preference over prices. If at observations t and t', we find that $p^{t'} x^t < p^t x^t$, then

the consumer has revealed that she strictly prefers the price $p^{t'}$ to the price p^t.

This is because, at the price vector $p^{t'}$ the consumer can, if she wishes, buy the bundle bought at the price p^t and she would still have money left for the purchase of other, non-grocery, consumption goods. (Put another way, if p^t and $p^{t'}$ are the prices at two grocery stores and the consumer could choose to go to one store or the other, then the observations in $\mathcal{D}$ will lead us to conclude that she will opt for the store where the prices are $p^{t'}$.) This concept of revealed preference recognizes that there are alternative uses to money besides groceries and that expenditure on groceries has an opportunity cost. Is it possible to build a revealed preference theory of consumer demand based on this alternate requirement that the consumer's preference over grocery prices, as revealed by the data set $\mathcal{D}$, is internally consistent; if so, what would such a theory look like? The objective of this paper is to answer this question and to demonstrate the appealing features of a theory of revealed price preference.

[1] The term GARP is from Varian (1982), which also contains a proof of the result. An extension of Afriat's Theorem to nonlinear budget sets can also be found in Forges and Minelli (2009).

[2] The consumer's overall utility function will have the form $G(\widetilde{U}(x), y)$, where $\widetilde{U}$ is the utility function defined over grocery bundles x, y is the bundle of all other goods consumed by the consumer and G is the overall utility function.

1.1. *The expenditure-augmented utility model*

We show that the absence of revealed preference cycles over prices — a property we call the *generalized axiom of revealed price preference* or GAPP, for short — has a very natural characterization. It is both necessary and sufficient for the existence of a strictly increasing function $U : \mathbb{R}^L_+ \times \mathbb{R}_- \to \mathbb{R}$ such that $x^s \in \arg\max_{x \in \mathbb{R}^L_+} U(x, -p^s x)$ for all $s = 1, 2, ..., T$. The function U should be interpreted as an *expenditure-augmented utility function*, where $U(x, -e)$ is the consumer's utility when she acquires x at the cost of e.[3] Unlike the standard consumer optimization problem, notice that the consumer does not have a budget constraint; instead, she is deterred from choosing an arbitrarily large bundle by the increasing expenditure it incurs, which reduces her utility. This is a reduced form utility function which implicitly holds fixed all other variables that may be relevant to the consumer's preference over $(x, -e)$; these variables could include the consumer's wealth, the prices of other goods which the consumer considers relevant to her consumption decision on these L goods, etc.

Besides being behaviorally compelling in its own right, the expenditure-augmented utility model has two distinctive features that makes it a worthwhile alternative to the standard model. (i) Notice that the marginal rate of substitution between two goods at a given bundle $x \in \mathbb{R}^L_+$ will typically depend on the expenditure incurred in acquiring that bundle; it follows that the marginal rate of substitution can vary with unobserved goods (whose consumption levels could change as e changes). In other words, our model does not require the assumption of weak separability and so it could be appropriate is situations where that assumption is problematic. (ii) Essentially because the standard setting only tries to model the agent's preference among bundles of the L goods, the only price information it requires are the prices of those goods *relative to each other*: it does not require the researcher to have any information on the prices of any other good that the consumer may also purchase. This modest informational requirement is an important advantage but it also means that the model cannot tell us anything about the consumer's *overall* welfare when the prices for the L goods change (as, in particular, the model does not distinguish between scalar multiples of prices). On the other hand our model recognizes that expenditure levels are endogenous (that is, chosen by the consumer at the observed prices) and exploits this to compare the consumer's welfare at different prices for the L goods; in fact, as we have pointed out, the model is *characterized* by this feature.

In the theoretical literature in public economics and industrial organization, it is common to assume individuals have quasilinear utility functions. In addition to tractability, this assumption ensures that the difference in consumer surplus due to price changes is

[3]In the main part of the paper, we allow for the consumption space to be a discrete subset of $\mathbb{R}^L_+$ and for prices to be nonlinear.

an exact measure of the change in welfare because quasilinearity ensures that there are no income effects.[4] This is a special case of our model, with $U(x,e) = H(x) - e$ for some real-valued and increasing function H. Our approach allows greater flexibility in the form of the utility function U and, in particular, does not require a constant marginal disutility of expenditure, which imposes strong and sometimes implausible restrictions on demand.

In the empirical literature, welfare changes are often calculated by explicitly introducing a numeraire good over and above the L goods being examined and then imputing demand for this numeraire good by using data on (for example) annual income; one could then calculate the impact on welfare following prices changes on these L goods at a given income and a given price for the numeraire good.[5] Obviously, compared with this approach, ours is useful when information on income is not available (which is a feature of many commonly used data sets). But more importantly, even when this information is available, by not using it, we are avoiding taking a stand on the precise budget from which the consumer draws her expenditure on these L goods; for example, the consumer could mentally set aside some expenditure for a group of commodities containing these L goods and this 'mental budget' may differ from the annual income (there are large literatures on mental accounting in economics, finance, marketing and psychology that stem from the work described in Thaler (1999)). That said, our approach does rely on the assumption that consumer's utility function U is stable over the period where her demand is observed; presumably large changes to the consumer's wealth (both her current resources and her future prospects) will have an impact on U, so in effect (and in a sense made formal in Proposition 1), we are assuming that these fluctuations are modest.

1.2. *Testing the model and estimating the welfare impact of price changes*

After formulating the expenditure-augmented utility model and exploring its theoretical features, the second part of our paper is devoted to testing it empirically. Our aim is twofold: we want to develop tests that are appropriate for repeated cross-sectional data and we want to design a procedure to determine the welfare impact of price changes. This latter step will require us to develop novel econometric theory which, we will argue, has more general applications.

We could in principle test our model on panel data sets of household or individual demand, for example, from information on purchasing behavior collected in scanner panels. Tests and applications of the standard model using Afriat's Theorem or its extensions are

[4] See, for instance, Varian (1985), Schwartz (1990) and the papers that follow.

[5] For recent work using this approach, see Blundell, Horowitz, and Parey (2012) and Hausman and Newey (2016). In both these papers, the empirical application involves the case where $L = 1$ (specifically, the good examined is gasoline, so it is a two-good demand system when the numeraire is included); when $L = 1$, our approach imposes no meaningful restrictions on data so it does not readily provide a viable alternative way to study the specific empirical issues in those papers.

very common,[6] and our model is just as straightforward to test on a panel data set.[7] Instead, we develop a random (expenditure-augmented) utility extension of our model for which we design an econometric test that we also apply to data. There are at least two reasons why this extension is useful: (i) The random utility model is more general in that it allows for individual preferences in the population to change over time, provided the population distribution of preferences stays the same. (ii) The random utility version can be tested on repeated cross sectional data, makes no restriction on unobserved preference heterogeneity and does not require any demand aggregation.

The random utility extension of the standard Afriat model is due to McFadden and Richter (1991), McFadden (2005). They assume that the econometrician observes the distribution of demand choices on each of a finite number of budget sets and characterize the observable content of this model, under the assumption that the distribution of preferences is stable across observations. There are two main challenges that must be overcome before their model can be taken to data. First, they do not account for sampling uncertainty as they assume that the econometrician observes the population distributions of demand; methodology for this was recently developed by Kitamura and Stoye (2017) who provide a testing procedure for the McFadden-Richter model which incorporates sampling error. Second, their test requires the observation of large samples of consumers who face not only the same prices but also make *identical expenditures*; a feature which is not true of any observational data. By contrast, the random utility version of our model can be tested directly on data where the choice distribution for each price can correspond to different expenditures. This feature of our model allows us to estimate the choice probabilities by simply using sample frequencies. By contrast, to test the McFadden-Richter model, Kitamura and Stoye (2017) need to estimate demand distributions at fixed levels of expenditure and this requires them to implement an instrumental variable technique (with all its attendant assumptions) to adjust for the endogeneity of observed expenditure.

For data sets that do not reject our model, we develop a test to determine whether a given proportion of households are revealed better (or worse) off for any observed price change in the data. The test can be inverted to yield confidence intervals for the proportion of consumers who are better or worse off. While we use existing results to test our model overall, this hypothesis test is a novel contribution. It can be applied generally to linear hypothesis testing for parameter vectors that are partially identified in a class

[6] For a recent survey, see Crawford and De Rock (2014).

[7] It is also common to test the standard model in experimental settings. Note that, our test is *not* appropriate on typical experimental data where the budget is provided exogenously (and is not endogenous as in our model) as part of the experiment.

of models investigated here, in Kitamura and Stoye (2017) and elsewhere. Other possible applications within our setting include bounds on average demand or quantiles of demand at counterfactual prices.

Finally, we test, and find support for, our model on two separate national household expenditure data sets from Canada and the U.K. Additionally, these applications demonstrate that (despite being partially identified) we can extract meaningful welfare implications of price changes in our data. The estimated bounds on the percentage of households who are made better or worse off by a price change are almost always narrower than 10% and more often are substantially narrower.

1.3. *Organization of the paper*

The remainder of this paper is structured as follows. Section 2 lays our the deterministic model and its revealed preference characterization, and Section 3 generalizes it to our analog of a random utility model. Section 4 describes the nonparametric statistical test of our model and the novel econometric theory needed to estimate the welfare bounds. Section 5 illustrates the econometric techniques with an empirical application, and Section 6 concludes. The appendix contains proofs that are not in the main body of the paper.

2. The Deterministic Model

The primitive in the analysis in this section is a data set of a single consumer's purchasing behavior collected by an econometrician. The econometrician observes the consumer's purchasing behavior over L goods and the prices at which those goods were chosen. In formal terms, the bundle is in $\mathbb{R}^L_+$ and the prices are in $\mathbb{R}^L_{++}$ and so an observation at t can be represented as $(p^t, x^t) \in \mathbb{R}^L_{++} \times \mathbb{R}^L_+$. The *data set* collected by the econometrician is $\mathcal{D} := \{(p^t, x^t)\}_{t=1}^T$. We will slightly abuse notation and use T both to refer to the number of observations, which we assume is finite, and the set $\{1, \ldots, T\}$; the meaning will be clear from the context. Similarly, L could denote both the number, and the set, of commodities.

We shall begin with a short description of Afriat's Theorem. This provides the background to our model and, additionally, we employ it to prove some of our results.

2.1. *Afriat's Theorem*

Given a data set $\mathcal{D} := \{(p^t, x^t)\}_{t=1}^T$, a locally nonsatiated[8] utility function $\widetilde{U} : \mathbb{R}^L_+ \to \mathbb{R}$ is said to *rationalize* $\mathcal{D}$ if

$$x^t \in \underset{\{x \in \mathbb{R}^L_+ :\, p^t x \leq p^t x^t\}}{\operatorname{argmax}} \widetilde{U}(x) \qquad \text{for all } t \in T. \tag{1}$$

[8]This means that at any bundle x and open neighborhood of x, there is a bundle y in the neighborhood with strictly higher utility.

This is the standard notion of rationalization and the one addressed by Afriat's Theorem.

A basic concept used in Afriat's Theorem is that of *revealed preference*. This is captured by two binary relations, $\succeq_x$ and $\succ_x$, defined on the chosen bundles observed in $\mathcal{D}$, that is, the set $\mathcal{X} := \{x^t\}_{t\in T}$. These revealed preference relations are defined as follows:

$$x^{t'} \succeq_x (\succ_x) x^t \text{ if } p^{t'} x^{t'} \geq (>) p^{t'} x^t.$$

We say that the bundle $x^{t'}$ is *directly revealed preferred* to x^t if $x^{t'} \succeq_x x^t$, that is, whenever the bundle x^t is cheaper at prices $p^{t'}$ than the bundle $x^{t'}$. If x^t is strictly cheaper, so $x^{t'} \succ_x x^t$, we say that $x^{t'}$ is *directly revealed strictly preferred* to x^t. This terminology is, of course, very intuitive. If the agent is maximizing some locally nonsatiated utility function $\widetilde{U} : \mathbb{R}^L_+ \to \mathbb{R}$, then if $x^{t'} \succeq_x x^t$ ($x^{t'} \succ_x x^t$), it must imply that $\widetilde{U}(x^{t'}) \geq (>) \widetilde{U}(x^t)$.

We denote the transitive closure of $\succeq_x$ by $\succeq_x^*$, that is, for $x^{t'}$ and x^t in $\mathcal{X}$, we have $x^{t'} \succeq_x^* x^t$ if there are t_1, t_2,...,t_N in T such that $x^{t'} \succeq_x x^{t_1}$, $x^{t_1} \succeq_x x^{t_2}, \ldots, x^{t_{N-1}} \succeq_x x^{t_N} \succeq_x x^t$, and $x^{t_N} \succeq_x x^t$; in this case, we say that $x^{t'}$ is *revealed preferred* to x^t. If anywhere along this sequence, it is possible to replace $\succeq_x$ with $\succ_x$ then we say that $x^{t'}$ is *revealed strictly preferred* to x^t and denote that relation by $x^{t'} \succ_x^* x^t$. Once again, this terminology is completely natural since if $\mathcal{D}$ is rationalizable by some locally nonsatiated utility function $\widetilde{U}$, then $x^{t'} \succeq_x^* (\succ_x^*) x^t$ implies that $\widetilde{U}(x^{t'}) \geq (>) \widetilde{U}(x^t)$.

Definition 2.1. A data set $\mathcal{D} = \{(p^t, x^t)\}_{t=1}^T$ satisfies the *Generalized Axiom of Revealed Preference* or *GARP* if there do not exist two observations $t, t' \in T$ such that $x^{t'} \succeq_x^* x^t$ and $x^t \succ_x^* x^{t'}$.

Afriat showed that this condition is necessary and sufficient for rationalization.

Afriat's Theorem (Afriat (1967))**.** *Given a data set $\mathcal{D} = \{(p^t, x^t)\}_{t=1}^T$, the following are equivalent:*

(1) $\mathcal{D}$ can be rationalized by a locally nonsatiated utility function.
(2) $\mathcal{D}$ satisfies GARP.
(3) $\mathcal{D}$ can be rationalized by a strictly increasing, continuous, and concave utility function.

REMARK 1. The discussion preceding the theorem essentially shows that (1) implies (2). The standard argument that shows (2) implies (3) (for instance in Fostel, Scarf, and Todd (2004)) works by showing that a consequence of GARP is that there exist numbers ϕ^t and $\lambda^t > 0$ (for all $t \in T$) that solve the inequalities

$$\phi^{t'} \leq \phi^t + \lambda^t p^t (x^{t'} - x^t) \text{ for all } t' \neq t. \tag{2}$$

It is then straightforward to show that

$$\widetilde{U}(x) = \min_{t \in T} \{\phi^t + \lambda^t p^t (x - x^t)\} \tag{3}$$

rationalizes $\mathcal{D}$, with the utility of the observed consumption bundles satisfying $\widetilde{U}(x^t) = \phi^t$. The function $\widetilde{U}$ is the lower envelope of a finite number of strictly increasing affine functions, and so it is strictly increasing, continuous, and concave. A remarkable feature of this theorem is that while GARP follows simply from local nonsatiation of the utility function, it is sufficient to guarantee that $\mathcal{D}$ is rationalized by a utility function with significantly stronger properties.

REMARK 2. To be precise, GARP guarantees that there is preference $\succsim$ (in other words, a complete, reflexive, and transitive binary relation) on $\mathcal{X}$ that extends the (potentially incomplete) revealed preference relations $\succeq^*_x$ and $\succ^*_x$ in the following sense: if $x^{t'} \succeq^*_x x^t$, then $x^{t'} \succsim x^t$ and if $x^{t'} \succ^*_x x^t$ then $x^{t'} \succ x^t$. One could then proceed to show that, for *any* such preference $\succsim$, there is in turn a utility function $\widetilde{U}$ that rationalizes $\mathcal{D}$ and extends $\succsim$ (from $\mathcal{X}$ to $\mathbb{R}^L_+$) in the sense that $\widetilde{U}(x^{t'}) \geq (>)\widetilde{U}(x^t)$ if $x^{t'} \succsim (\succ)x^t$ (see Quah (2014)). This has implications on the inferences one could draw from the data. If $x^{t'} \not\succeq^*_x x^t$ (or if $x^{t'} \succeq^*_x x^t$ but $x^{t'} \not\succ^*_x x^t$) then it is always possible to find a preference extending the revealed preference relations such that $x^t \succ x^{t'}$ (or $x^{t'} \sim x^t$ respectively).[9] Therefore, $x^{t'} \succeq^*_x (\succ^*_x)x^t$ if and only if every locally nonsatiated utility function rationalizing $\mathcal{D}$ has the property that $\widetilde{U}(x^{t'}) \geq (>)\widetilde{U}(x^t)$. We will similarly argue that the revealed price preference relation contains the most detailed information for welfare comparisons in our model.

REMARK 3. A feature of Afriat's Theorem that is less often remarked upon is that in fact $\widetilde{U}$, as given by (3), is well-defined, strictly increasing, continuous, and concave on the domain $\mathbb{R}^L$, rather than just the positive orthant $\mathbb{R}^L_+$. Furthermore,

$$x^t \in \underset{\{x \in \mathbb{R}^L:\, p^t x \leq p^t x^t\}}{\operatorname{argmax}} \widetilde{U}(x) \qquad \text{for all } t \in T. \tag{4}$$

In other words, x^t is optimal even if $\widetilde{U}$ is extended beyond the positive orthant and x can be chosen from the larger domain. (Compare (4) with (1).) This curiosity will turn out to be useful when we apply Afriat's Theorem in our proofs.

2.2. *Consistent welfare comparisons across prices*

Typically, the L goods whose demand is being monitored by the econometrician constitutes no more than a part of the purchasing decisions made by the consumer. The consumer's true budget (especially when one takes into account the possibility of borrowing and saving) is never observed and the expenditure which she devotes to the L goods is a decision made by the consumer and is dependent on the prices. The consumer's choice

[9]We use $x^{t'} \sim x^t$ to mean that $x^{t'} \succsim x^t$ and $x^t \succsim x^{t'}$.

over the L goods inevitably affects what she could afford on, and therefore her consumption of, other goods not observed by the econometrician. Given this, the rationalization criterion (1) used in Afriat's Theorem will only make sense under an additional assumption of weak separability: the consumer has a sub-utility function $\overline{U}$ defined over the L goods and the utility function of the consumer, defined over all goods, takes the form $H(\overline{U}(x), y)$, where x is the bundle of L goods observed by the econometrician and y is the bundle of unobserved goods. Assuming that the consumer chooses (x, y) to maximize $G(x, y) = H(\overline{U}(x), y)$, subject to a budget constraint $px + qy \leq M$ (where M is her wealth and p and q are the prices of the observed and unobserved goods respectively), then, at the prices p^t, the consumer's choice x^t will obey (1) provided H is strictly increasing in either the first or second argument. This provides the theoretical motivation to test for the existence of a sub-utility function $\widetilde{U}$ that rationalizes a data set $\mathcal{D} = \{(p^t, x^t)\}_{t=1}^T$. Notice that once the weak separability assumption is in place, the background requirements of Afriat's test are very modest in the sense that it is possible for prices of the unobserved goods and the unobserved total wealth to change arbitrarily across observations, without affecting the validity of the test. This is a major advantage in applications but the downside is that the conclusions of this model are correspondingly limited to ranking different bundles among the observed goods via the sub-utility function.

Now suppose that instead of checking for rationalizability in the sense of (1), the econometrician would like to ask a different question: given a data set $\mathcal{D} = \{(p^t, x^t)\}_{t=1}^T$, can he sign the welfare impact of a price change from p^t to $p^{t'}$? Stated differently, this question asks whether he can compare the consumer's *overall* utility after a change in the prices of the L goods from p^t and $p^{t'}$, holding fixed other aspects of the economic environment that may affect the consumer's welfare, such as the prices of unobserved goods and her overall wealth. Perhaps the most basic welfare comparison in this setting can be made as follows: if at prices $p^{t'}$, the econometrician finds that $p^{t'}x^t < p^t x^t$ then he can conclude that the agent is better off at the price vector $p^{t'}$ compared to p^t. This is because, at the price $p^{t'}$ the consumer can, if she wishes, buy the bundle bought at p^t and she would still have money left over to buy other things, so she must be strictly better off at $p^{t'}$. This ranking is eminently sensible, but can it lead to inconsistencies?

Example 1. Consider a two observation data set

$$p^t = (2, 1)\,,\; x^t = (4, 0) \;\text{ and }\; p^{t'} = (1, 2)\,,\; x^{t'} = (0, 1).$$

which is depicted in Figure 1. Given that the budget sets do not even cross, we know that GARP holds. Since $p^{t'}x^t < p^t x^t$, it seems that we may conclude that the consumer is better off at prices $p^{t'}$ than at p^t; however, it is also true that $p^{t'}x^{t'} > p^t x^{t'}$, which gives the opposite conclusion.

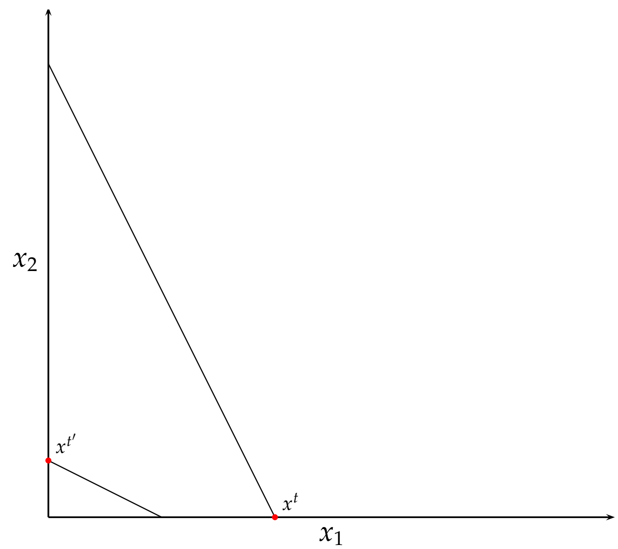


FIGURE 1. Choices that do not allow for consistent welfare predictions.

This example shows that for an econometrician to be able to consistently compare the consumer's welfare at different prices, some restriction (different from GARP) has to be imposed on the data set. To be precise, define the binary relations $\succeq_p$ and $\succ_p$ on $\mathcal{P} := \{p^t\}_{t\in T}$, that is, the set of price vectors observed in $\mathcal{D}$, in the following manner:

$$p^{t'} \succeq_p (\succ_p) p^t \text{ if } p^{t'} x^t \leq (<) p^t x^t.$$

We say that price $p^{t'}$ is *directly revealed preferred* to p^t if $p^{t'} \succeq_p p^t$, that is, whenever the bundle x^t is cheaper at prices $p^{t'}$ than at prices p^t. If it is strictly cheaper, so $p^{t'} \succ_p p^t$, we say that $p^{t'}$ is *directly revealed strictly preferred* to p^t. We denote the transitive closure of $\succeq_p$ by $\succeq_p^*$, that is, for $p^{t'}$ and p^t in $\mathcal{P}$, we have $p^{t'} \succeq_p^* p^t$ if there are t_1, t_2,...,t_N in T such that $p^{t'} \succeq_p p^{t_1}$, $p^{t_1} \succeq_p p^{t_2}$,..., $p^{t_{N-1}} \succeq_p p^{t_N}$, and $p^{t_N} \succeq_p p^t$; in this case we say that $p^{t'}$ is *revealed preferred* to p^t. If anywhere along this sequence, it is possible to replace $\succeq_p$ with $\succ_p$ then we say that $p^{t'}$ is *revealed strictly preferred* to p^t and denote that relation by $p^{t'} \succ_p^* p^t$. Then the following restriction is the bare minimum required to exclude the possibility of circularity in the econometrician's assessment of the consumer's wellbeing at different prices.

Definition 2.2. The data set $\mathcal{D} = \{(p^t, x^t)\}_{t=1}^T$ satisfies the *Generalized Axiom of Price Preference* or *GAPP* if there do not exist two observations $t, t' \in T$ such that $p^{t'} \succeq_p^* p^t$ and $p^t \succ_p^* p^{t'}$.

This in turn leads naturally to the following question: if a consumer's observed demand behavior obeys GAPP, what could we say about her decision making procedure?

2.3. *The Expenditure-Augmented Utility Model*

An *expenditure-augmented utility function* (or simply, an *augmented utility function*) is a utility function that has, as its arguments, both the bundle consumed by the consumer x and the total expense e incurred in acquiring the bundle. Formally, the augmented utility function U has domain $\mathbb{R}^L_+ \times \mathbb{R}_-$, where $U(x, -e)$ is assumed to be strictly increasing in the last argument; in other words, utility is strictly decreasing in expenditure. This second argument captures the opportunity cost of money and is a simple, reduced form way of modeling the tradeoff with all other financial decisions made by the consumer. At a given price p, the consumer chooses a bundle x to maximize $U(x, -px)$. We denote the *indirect utility at price p* by

$$V(p) := \sup_{x \in \mathbb{R}^L_+} U(x, -px). \tag{5}$$

If the consumer's augmented utility maximization problem has a solution at every price vector $p \in \mathbb{R}^L_{++}$, then V is also defined at those prices and this induces a reflexive, transitive, and complete preference over prices in $\mathbb{R}^L_{++}$.

A data set $\mathcal{D} = \{(p^t, x^t)\}_{t=1}^T$ is *rationalized by an augmented utility function* if there exists such a function $U : \mathbb{R}^L_+ \times \mathbb{R}_- \to \mathbb{R}$ with

$$x^t \in \underset{x \in \mathbb{R}^L_+}{\operatorname{argmax}}\, U(x, -p^t x) \qquad \text{for all } t \in T.$$

Notice that unlike the notion of rationalization in Afriat's Theorem, we do not require the bundle x to be chosen from the budget set $\{x \in \mathbb{R}^L_+ : p^t x \le p^t x^t\}$. The consumer can instead choose from the entire consumption space $\mathbb{R}^L_+$, though the utility function penalizes expenditure.

It is straightforward to see that GAPP is necessary for a data set to be rationalized by an augmented utility function. First, notice that if $p^{t'} \succeq_p p^t$, then $p^{t'} x^t \le p^t x^t$, and so

$$V(p^{t'}) \ge U(x^t, -p^{t'} x^t) \ge U(x^t, -p^t x^t) = V(p^t).$$

Furthermore, $U(x^t, -p^{t'} x^t) > U(x^t, -p^t x^t)$ if $p^{t'} \succ_p p^t$, and in that case $V(p^{t'}) > V(p^t)$. Suppose GAPP were not satisfied and there were two observations $t, t' \in T$ such that $p^{t'} \succeq^*_p p^t$ and $p^t \succ^*_p p^{t'}$. Then there would exist $t_1, t_2, \ldots, t_N \in T$ such that

$$V(p^{t'}) \ge V(p^{t_1}) \ge \cdots \ge V(p^{t_N}) \ge V(p^t) > V(p^{t'})$$

which is impossible.

Our main theoretical result, which we state next, also establishes the sufficiency of GAPP for rationalization. Moreover, the result states that whenever $\mathcal{D}$ can be rationalized, it can be rationalized by an augmented utility function U with a list of properties that make it convenient for analysis. In particular, we can guarantee that there is always a solution to $\max_{x \in \mathbb{R}^L_+} U(x, -p \cdot x)$ for any $p \in \mathbb{R}^L_{++}$. This property guarantees that U will generate preferences over all the price vectors in $\mathbb{R}^L_{++}$. Clearly, it is also necessary for making out-of-sample predictions and, indeed, it is important for the internal consistency of the model.[10]

Theorem 1. *Given a data set $\mathcal{D} = \{(p^t, x^t)\}_{t=1}^T$, the following are equivalent:*

(1) $\mathcal{D}$ can be rationalized by an augmented utility function.
(2) $\mathcal{D}$ satisfies GAPP.
(3) $\mathcal{D}$ can be rationalized by an augmented utility function U that is strictly increasing, continuous, and concave. Moreover, U is such that $\max_{x \in \mathbb{R}^L_+} U(x, -p \cdot x)$ has a solution for all $p \in \mathbb{R}^L_{++}$.

REMARK: In order to keep the presentation simple, our maintained assumption throughout this paper is that the consumption space is $\mathbb{R}^L_+$. However, this theorem (and the rest of the theory which builds on it) applies also to the case where the consumption space is some closed subset X of $\mathbb{R}^L_+$; indeed, our proof covers this more general case. For example, if we model all the observed goods as discrete, then $X = \mathbb{N}^L_+$. Another possibility is that the consumer is deciding on buying a subset of objects from a set A with L items. Then each subset could be represented as an element of $X = \{0, 1\}^L$; for $x \in X$, the i entry (x_i) equals 1 if and only if the ith object is chosen.

When the consumption space is X, the notion of a rationalization will obviously have to be altered to require the optimality, in X rather than in $\mathbb{R}^L_+$, of the chosen bundle. Retaining our definitions of the revealed preference relations $\succeq_p$ and $\succ_p$, it is straightforward to check that GAPP is still necessary for the maximization of an augmented utility function. Indeed, the whole theorem remains essentially valid, with statement (3) modified to read: "$\mathcal{D}$ can be rationalized by an augmented utility function $U : X \times \mathbb{R}_- \to \mathbb{R}$ with the following properties: (i) it admits an extension $\widehat{U} : \mathbb{R}^L_+ \times \mathbb{R}_- \to \mathbb{R}$ that is strictly increasing, continuous, and concave; (ii) $\max_{x \in X} U(x, -px)$ has a solution for all $p \in \mathbb{R}^L_{++}$."

PROOF OF THEOREM 1. We will show that (2) $\implies$ (3). We have already argued that (1) $\implies$ (2) and (3) $\implies$ (1) by definition.

Choose a number $M > \max_t p^t x^t$ and define the augmented data set $\widetilde{\mathcal{D}} = \{(p^t, 1), (x^t, M - p^t x^t)\}_{t=1}^T$. This data set augments $\mathcal{D}$ since we have introduced an $L + 1^{\text{th}}$ good, which we have priced at 1 across all observations, with the demand for this good equal to $M - p^t x^t$.

The crucial observation to make here is that

$$(p^t, 1)(x^t, M - p^t x^t) \geq (p^t, 1)(x^{t'}, M - p^{t'} x^{t'}) \text{ if and only if } p^{t'} x^{t'} \geq p^t x^{t'},$$

[10]Suppose the data set is rationalized but only by an augmented utility function for which the existence of an optimum is not generally guaranteed, then it undermines the hypothesis tested since it is not clear why the sample collected should then have the property that an optimum exists.

which means that

$$(x^t, M - p^t x^t) \succeq_x (x^{t'}, M - p^{t'} x^{t'}) \text{ if and only if } p^t \succeq_p p^{t'}.$$

Similarly,

$$(p^t, 1)(x^t, M - p^t x^t) > (p^t, 1)(x^{t'}, M - p^{t'} x^{t'}) \text{ if and only if } p^{t'} x^{t'} > p^t x^{t'},$$

and so

$$(x^t, M - p^t x^t) \succ_x (x^{t'}, M - p^{t'} x^{t'}) \text{ if and only if } p^t \succ_p p^{t'}.$$

Consequently, $\mathcal{D}$ satisfies GAPP if and only if $\widetilde{\mathcal{D}}$ satisfies GARP. Applying Afriat's Theorem, when $\widetilde{\mathcal{D}}$ satisfies GARP, there is $\widetilde{U} : \mathbb{R}^{L+1} \to \mathbb{R}$ such that

$$(x^t, M - p^t x^t) \in \underset{\{x \in \mathbb{R}^L_+ : p^t x + m \leq M\}}{\operatorname{argmax}} \widetilde{U}(x, m) \qquad \text{for all } t \in T.$$

The function $\widetilde{U}$ can be chosen to be strictly increasing, continuous, and concave, and the lower envelope of a finite set of affine functions. Note that augmented utility function $\overline{U} : \mathbb{R}^L_+ \times \mathbb{R}_- \to \mathbb{R}$ defined by $\overline{U}(x, -e) := \widetilde{U}(x, M - e)$ rationalizes $\mathcal{D}$ as x^t solves $\max_{x \in \mathbb{R}^L_+} \overline{U}(x, -p^t x)$ by construction. Furthermore, $\overline{U}$ is strictly increasing in $(x, -e)$, continuous, and concave.

Let $h : \mathbb{R}_+ \to \mathbb{R}$ be a differentiable function with $h(0) = 0$, $h'(k) > 0$, $h''(k) \geq 0$ for $k \in \mathbb{R}_+$, and $\lim_{k \to \infty} h'(k) = \infty$.[11] Define $\widehat{U} : \mathbb{R}^L_+ \times \mathbb{R}_- \to \mathbb{R}$ by

$$\widehat{U}(x, -e) := \overline{U}(x, -e) - h(\max\{0, e - M\}). \tag{6}$$

It is clear that this function is strictly increasing in $(x, -e)$, continuous, and concave. Furthermore, x^t solves $\max_{x \in \mathbb{R}^L_+} \widehat{U}(x, -p^t x)$ and hence solves $\max_{x \in X} \widehat{U}(x, -p^t x)$. This is because $\widehat{U}(x, -e) \leq \overline{U}(x, -e)$ for all $(x, -e)$, and $\widehat{U}(x^t, -p^t x^t) = \overline{U}(x^t, -p^t x^t)$.

Lastly, we claim that at every $p \in \mathbb{R}^L_{++}$, $\operatorname{argmax}_{x \in X} \widehat{U}(x, -px)$ is nonempty. Choose a sequence $x^n \in X$ such that $\widehat{U}(x^n, -px^n)$ tends to $\sup_{x \in X} \widehat{U}(x, -px)$ (which we allow to be infinity). It is impossible for $px^n \to \infty$ because the piecewise linearity of $U(x, -e)$ in x and the assumption that $\lim_{k \to \infty} h'(k) \to \infty$ implies that $\widehat{U}(x^n, -px^n) \to -\infty$. So the sequence px^n is bounded, which in turn means that there is a subsequence of x^n that converges to $x^\star \in X$ (since X is closed). By the continuity of $\widehat{U}$, we obtain $\widehat{U}(x^\star, -px^\star) = \sup_{x \in X} \widehat{U}(x, -px)$. Lastly, note that the restriction of $\widehat{U}$ to $X \times \mathbb{R}_-$ still rationalizes $\mathcal{D}$. ■

From this point onwards, when we refer to 'rationalization' without additional qualifiers, we shall mean rationalization by an augmented utility function, that is, in the sense established by Theorem 1 rather than in the sense established by Afriat's Theorem.

The inclusion of expenditure in the augmented utility function captures the opportunity cost incurred by the consumer when she chooses to buy some bundle of goods. The

[11]For example, $h(k) = k^3$.

proof of Theorem 1 itself provides a particular interpretation of this cost. We could suppose that the consumer is maximizing an overall utility function that depends both on the observed bundle x and on a bundle y of other goods, subject to an overall wealth of M; more formally, the consumer is maximizing the overall utility $G(x,y)$ subject to $px + qy \leq M$, where q are the prices of other goods. Keeping q and M fixed, we can then interpret $U(x,-e)$ as the greatest overall utility the consumer can achieve by choosing y optimally conditional on consuming x, that is,

$$U(x,-e) = \max_{\{y \geq 0 \,:\, qy \leq M-e\}} G(x,y). \tag{7}$$

It is worth emphasizing, however, that a strength of our framework is that we do not have to take a stand on what goods y the consumer trades off with x or the amount M that she allocates to them jointly. This has the advantage of allowing us to account for potentially unobserved 'mental budgeting' (see Thaler (1999)) that the consumer may be engaged in. Indeed, in addition to the large evidence from laboratory data, there is recent empirical evidence from field data that shows that money is not fungible and that consumers often keep separate accounts for different categories of purchases (see, for instance, Feldman (2010), Hastings and Shapiro (2013) and Milkman and Beshears (2009)).

The interpretation of the augmented utility function we have just given is by no means the only natural one. For example, the following is a related but distinct way of motivating the augmented utility function. We suppose a consumer who maximizes discounted expected utility over an infinite horizon, where $\tilde{v} : \mathbb{R}^L_+ \to \mathbb{R}$ is the felicity function. At each period the consumer receives an income m' and faces a vector of prices p', both of which are stochastic and drawn from a stationary distribution Υ. In this case, the opportunity cost of current expenditure consists of foregone savings which lower future consumption. Under fairly standard conditions, it is possible to cast the consumer's purchasing decision in its recursive form, that is, the consumer chooses $x \in \mathbb{R}^L_+$ to maximize

$$\tilde{v}(x) + \delta \int \mathcal{V}\left((1+r)\,(W - px) + m', p'\right)\, d\Upsilon(m',p'),$$

given current (realized) wealth W and price vector p.[12] ($\mathcal{V}$ is the value function, r the interest rate on saving or borrowing, and δ the discount factor.) Keeping W fixed, notice that the consumer's current demand as a function of current prices p arises from the maximization of an augmented utility function (with $e = px$).

These two interpretations notwithstanding, it is worth emphasizing that Theorem 1 simply states that if we require a data set to exhibit consistent welfare comparisons across prices, then the consumer will behave as though she is maximizing an augmented utility

[12]See, for example, the treatment in Ozaki and Streufert (1996), which also emphasizes that neither additivity across time nor additivity across states (expected utility) is crucial to this recursive formulation of the consumer's problem.

function. That, in itself, provides motivation for the augmented utility function. We are not required to subscribe to a specific – or *any* – 'more fundamental' model from which the augmented utility function could be derived. Indeed, it may not be any less plausible simply to interpret the augmented utility function as representing the consumer's preference over bundles of L goods and their associated expenditure, which she has developed as a habit and which guides her purchasing decisions.

The augmented utility function could be thought of as a generalization of the *quasilinear* model which is commonly employed in partial equilibrium analysis, both to model demand and also to carry out a welfare analysis of price changes. In this case, the consumer maximizes utility net of expenditure, that is, she chooses a bundle x that maximizes

$$U(x, -e) := \overline{U}(x) - e, \tag{8}$$

where $\overline{U}(x)$ is the utility of the bundle x, and $e \geq 0$ is the expense incurred when acquiring the bundle. This formulation can be motivated by assuming that the consumer has an overall utility function that is quasilinear, that is, $G(x, y) = \overline{U}(x) + y$, where there is one unobserved representative good consumed at level $y \in \mathbb{R}_+$. If we normalize the price of the outside good at 1 and assume that M is sufficiently large, then the consumer maximizes $G(x, y)$ subject to $px + y \leq M$ if and only if he chooses x to maximize $U(x, -px) = \overline{U}(x) - px$.

The quasilinear utility model imposes a strong restriction on the consumer's demand behavior that is not necessarily desirable.[13] For example, suppose $L = 2$ and the consumer at prices $(p_1, p_2) \in \mathbb{R}^2_{++}$ prefers the bundle $(2, 1)$ to another bundle $(1, 2)$; then it is straightforward to check that this preference is maintained at the prices $(p_1 + k, p_2 + k) \in \mathbb{R}^2_{++}$ for any k. In contrast, for the general augmented utility function, the agent's marginal rate of substitution between any two goods (among the L observed goods) can depend on the expenditure incurred in acquiring the L-good bundle.[14] So we allow for the possibility that the consumer's willingness to trade, say, food for alcohol depends on the overall expenditure incurred in acquiring the bundle; if food and alcohol prices increase and the expenditure incurred in acquiring a given bundle goes up, the consumer will have less to spend on other things such as leisure, which could well have an impact on her marginal rate of substitution between those two goods.

[13]A consumer with a quasilinear utility function will generate a data set obeying both GAPP and GARP. The precise conditions on a data set that characterize rationalization with a quasilinear utility function can be found in Brown and Calsamiglia (2007).

[14]If the augmented utility function is differentiable, then the marginal rate of substitution between goods 1 and 2 is given by $(\partial U/\partial x_1)/(\partial U/\partial x_2)$ evaluated at $(x, -e)$. This will in general depend on e, though not in the quasilinear case.

2.4. *Local robustness of the GAPP test*

The augmented utility function captures the idea that a consumer's choice is guided by the satisfaction she derives from the bundle as well as the opportunity cost of acquiring that bundle, as measured by its expense. Price data may contain measurement error which implies that prices of the observed goods p^t we use for testing may well differ from their true values q^t. Put differently, at time t, the consumer actually maximizes $U(x, -q^t x)$ instead of $U(x, -p^t x)$ which is the hypothesis we are testing.

Another potential source of error is that the agent's unobserved wealth or mental budget may change from one observation to the next and this could manifest itself as a change in the consumer's stock of the representative outside good. If this happens, then at observation t, the consumer would be maximizing $U(x, -p^t x + \delta^t)$, where δ^t is the perturbation in wealth at time t, instead of maximizing $U(x, -p^t x)$. Of course, the errors could potentially enter simultaneously in both prices and wealth.

Statement (2) of Proposition 1 below shows that inference from a GAPP test is unaffected as long as the size of the errors are bounded (equation (9) provides the specific bound). Specifically, so long as $\mathcal{D}$ obeys a mild genericity condition (which is satisfied in our empirical application), the GAPP test is locally robust in the sense that any conclusion obtained through the test remains valid for sufficiently small perturbations of the original hypothesis. For example, a data set that fails GAPP is not consistent with the maximization of an augmented utility function, for all prices sufficiently close to the ones observed and after allowing for small wealth perturbations.

Not surprisingly, if the errors are allowed to be unbounded, then the model ceases to have any content (statement (1) of Proposition 1). Specifically, we can always find wealth perturbations δ^t (while prices p^t are assumed to be measured without error) such that each x^t maximizes $U(x, -p^t x + \delta^t)$, with no restrictions on $\mathcal{D}$. In particular, these perturbations can be chosen to be mean zero.

Proposition 1. *Given a data set $\mathcal{D} = \{(p^t, x^t)\}_{t=1}^T$, the following hold:*

(1) *There exists an augmented utility function U and $\{\delta^t\}_{t=1}^T$, with $\sum_{t=1}^T \delta^t = 0$, such that $x^t \in \text{argmax}_{x \in \mathbb{R}^L_+} U(p^t, -p^t x^t + \delta^t)$ for all $t \in T$.*

(2) *Suppose, $\mathcal{D}$ satisfies $p^t x^t - p^{t'} x^t \neq 0$ for all $t \neq t'$ and let $\{q^t\}_{t=1}^T$ and $\{\delta^t\}_{t=1}^T$ satisfy*

$$2 \max_{t \in T} \{|\delta^t|\} + 2B \max_{t \in T, i \in L} \{|\epsilon_i^t|\} < \min_{t,t' \in T, t \neq t'} |p^t x^t - p^{t'} x^t| \tag{9}$$

where $B = \max_{t \in T} \{\sum_{i=1}^L |x_i^t|\}$ and $\epsilon_i^t := q_i^t - p_i^t$.

If $\mathcal{D}$ obeys GAPP, there is an augmented utility function U such that

$$x^t \in \underset{x \in \mathbb{R}^L_+}{\text{argmax}}\, U(q^t, -q^t x^t + \delta^t) \text{ for all } t \in T, \tag{10}$$

and if $\mathcal{D}$ violates GAPP, then there is no augmented utility function U such that (10) holds.

Additionally, Aguiar and Kashaev (2017) have recently developed a technique to explicitly account for measurement error in revealed preference tests. Their method can, in principle, also be applied to our model (as the test of our model can be equivalently written in the first order condition form they require).

2.5. *Comparing GAPP and GARP*

Recall that Example 1 in Section 2.2 is an example of a data set that obeys GARP but fails GAPP. We now present an example of a data set that obeys GAPP but fails GARP.

Example 2. Consider the data set consisting of the following two choices:

$$p^t = (2,1)\,,\, x^t = (2,1) \text{ and } p^{t'} = (1,4)\,,\, x^{t'} = (0,2).$$

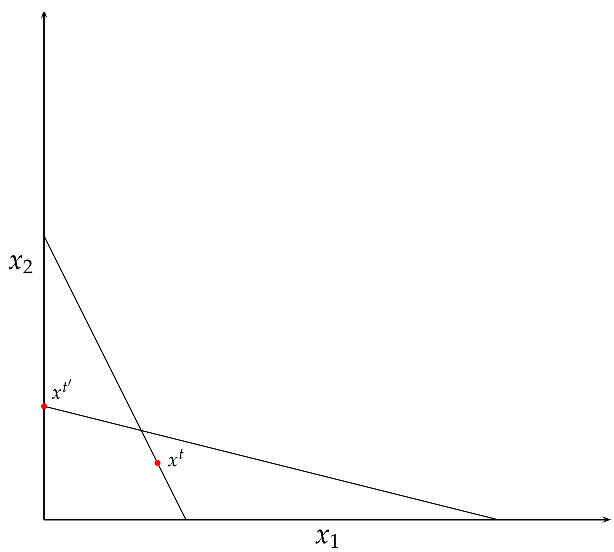


FIGURE 2. Choices that satisfy GAPP but not GARP

These choices are shown in Figure 2. This is a classic GARP violation as $p^t x^t = 5 > 2 = p^t x^{t'}$ ($x^t \succ_x x^{t'}$) and $p^{t'} x^{t'} = 8 > 6 = p^{t'} x^t$ ($x^{t'} \succ_x x^t$). In words, each bundle is strictly cheaper than the other at the budget set corresponding to the latter observation. However, these choices satisfy GAPP as $p^{t'} x^{t'} = 8 > 2 = p^t x^{t'}$ ($p^t \succ_p p^{t'}$) but $p^t x^t = 5 \not> 6 = p^{t'} x^t$ ($p^{t'} \not\succ_p p^t$).

The upshot of this example is that there are data sets that admit rationalization with an augmented utility function that cannot be rationalized in Afriat's sense (as in (1)). If we interpret the augmented utility function in the form $G(x,y)$ (given by (7)), then this means that while the agent's behavior is consistent with the maximization of an overall utility function G, this utility function is not weakly separable in the observed goods x. In particular, this implies that the agent does not have a quasilinear utility function (with $G(x,y) = U(x) + y$), which *is* weakly separable in the observed goods and will generate data sets obeying both GAPP and GARP.

While GAPP and GARP are not comparable conditions, there is a way of converting a GAPP test into a GARP test that will prove to be very convenient. Given a data set $\mathcal{D} = \{(p^t, x^t)\}_{t=1}^T$, we define the *expenditure-normalized* version of $\mathcal{D}$ as another data set $\breve{\mathcal{D}} := \{(p^t, \breve{x}^t)\}_{t=1}^T$, such that $\breve{x}^t = x^t/(p^t x^t)$. This new data set has the feature that $p^t \breve{x}^t = 1$ for all $t \in T$. Notice that the revealed price preference relations $\succeq_p$, $\succ_p$ remain unchanged when consumption bundles are scaled. Put differently, a data set obeys GAPP if and only if its normalized version also obeys GAPP. Therefore, from the perspective of testing for rationalization (by an augmented utility function) this change in the data set is immaterial. The next proposition makes a different and less obvious observation about $\breve{\mathcal{D}}$.

Proposition 2. *Let $\mathcal{D} = \{(p^t, x^t)\}_{t=1}^T$ be a data set and let $\breve{\mathcal{D}} = \{(p^t, \breve{x}^t)\}_{t\in T}$, where $\breve{x}^t = x^t/(p^t x^t)$, be its expenditure-normalized version. Then the revealed preference relations $\succeq_p^*$ and $\succ_p^*$ on $\mathcal{P} = \{p^t\}_{t=1}^T$ and the revealed preference relations $\succeq_x^*$ and $\succ_x^*$ on $\breve{\mathcal{X}} = \{\breve{x}^t\}_{t=1}^T$ are related in the following manner:*

(1) $p^t \succeq_p^* p^{t'}$ if and only if $\breve{x}^t \succeq_x^* \breve{x}^{t'}$.
(2) $p^t \succ_p^* p^{t'}$ if and only if $\breve{x}^t \succ_x^* \breve{x}^{t'}$.

As a consequence, $\mathcal{D}$ obeys GAPP if and only if $\breve{\mathcal{D}}$ obeys GARP.

Proof. Notice that

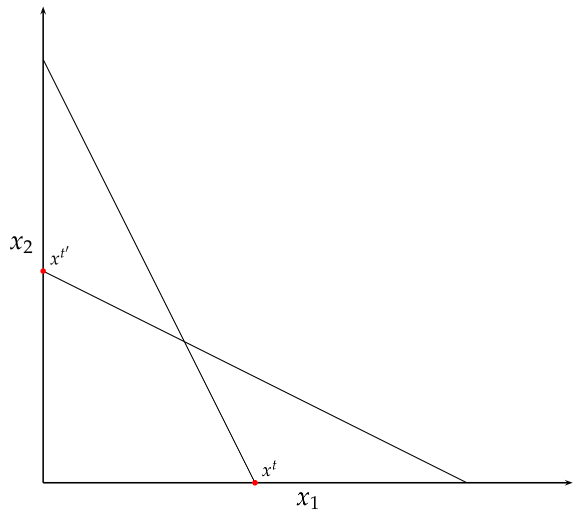
$$p^t \frac{x^t}{p^t x^t} \geq p^t \frac{x^{t'}}{p^{t'} x^{t'}} \iff p^{t'} x^{t'} \geq p^t x^{t'}.$$

The left side of the equivalence says that $\breve{x}^t \succeq_x \breve{x}^{t'}$ while the right side says that $p^t \succeq_p p^{t'}$. This implies (1) since $\succeq_p^*$ and $\succeq_x^*$ are the transitive closures of $\succeq_p$ and $\succeq_x$ respectively. Similarly, it follows from

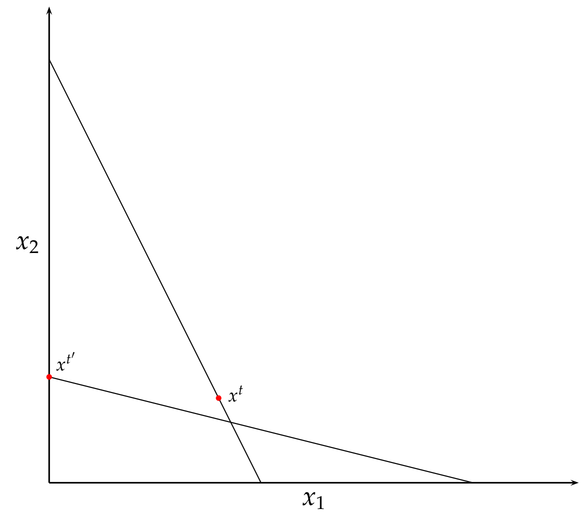
$$p^t \frac{x^t}{p^t x^t} > p^t \frac{x^{t'}}{p^{t'} x^{t'}} \iff p^{t'} x^{t'} > p^t x^{t'}$$

that $\breve{x}^t \succ_x \breve{x}^{t'}$ if and only if $p^t \succ_p p^{t'}$, which leads to (2). The claims (1) and (2) together guarantee that there is a sequence of observations in $\mathcal{D}$ that lead to a GAPP violation if and only if the analogous sequence in $\breve{\mathcal{D}}$ lead to a GARP violation. □

As an illustration, compare the data sets in Figure 1 and Figure 2 to the expenditure-normalized data sets in Figure 3a and Figure 3b. It can be clearly observed that the expenditure-normalized data in Figure 3a contains a GARP violation (which implies it does not satisfy GAPP) whereas the data in Figure 3b does not violate GARP (and, hence, satisfies GAPP).



(a) Example 1



(b) Example 2

FIGURE 3. Expenditure-Normalized Choices

A consequence of Proposition 2 is that the expenditure-augmented utility model can be tested in two ways: (i) we can test GAPP directly, or, (ii) we can normalize the data by expenditure and then test GARP. If we are simply interested in testing GAPP on a data set $\mathcal{D}$ then both methods are computationally straightforward and there is not much to choose between them: they both require the construction of their (respective) revealed preference relations and testing involves checking for acyclicity. However, as we shall see in the next section, the indirect procedure supplied by Proposition 2 will prove to be very useful for testing in the random augmented utility environment.

2.6. *Preference over Prices*

We know from Theorem 1 that if $\mathcal{D}$ obeys GAPP then it can be rationalized by an augmented utility function with an indirect utility that is defined at all price vectors in $\mathbb{R}^L_{++}$. It is straightforward to check that any indirect utility function V as defined by (5) has the following two properties:

(a) it is *nonincreasing in* p, in the sense that if $p' \geq p$ (in the product order) then $V(p') \leq V(p)$, and
(b) it is *quasiconvex in* p, in the sense that if $V(p) = V(p')$, then $V(\beta p + (1-\beta)p') \leq V(p)$ for any $\beta \in [0, 1]$.

Any rationalizable data set $\mathcal{D}$ could potentially be rationalized by many augmented utility functions and each one of them will lead to a different indirect utility function. We denote this set of indirect utility functions by $\mathbf{V}(\mathcal{D})$. We have already observed that if $p^t \succeq_p^* (\succ_p^*) p^{t'}$ then $V(p^t) \geq (>) V(p^{t'})$ for any $V \in \mathbf{V}(\mathcal{D})$; in other words, the conclusion that the consumer prefers the prices p^t to $p^{t'}$ is *fully nonparametric* in the sense that it is independent of the precise augmented utility function used to rationalize $\mathcal{D}$. The next result says that, without further information on the agent's augmented utility function, this is *all* the information on the agent's preference over prices in $\mathcal{P}$ that we could glean from the data. Thus, in our nonparametric setting, the revealed price preference relation contains the most detailed information for welfare comparisons.

Proposition 3. *Suppose $\mathcal{D} = \{(p^t, x^t)\}_{t=1}^T$ is rationalizable by an augmented utility function. Then for any p^t, $p^{t'}$ in $\mathcal{P}$:*

(1) $p^t \succeq_p^* p^{t'}$ if and only if $V(p^t) \geq V(p^{t'})$ for all $V \in \mathbf{V}(\mathcal{D})$.
(2) $p^t \succ_p^* p^{t'}$ if and only if $V(p^t) > V(p^{t'})$ for all $V \in \mathbf{V}(\mathcal{D})$.

2.7. *Extension to nonlinear pricing*

We have so far assumed that prices are linear, but this feature is not crucial to our main result. Given that goods are sometimes priced nonlinearly (for instance, quantity discounts, bundle pricing etc.), and this feature can be important in certain modeling contexts, it would be useful to have a version of Theorem 1 that allows for that possibility.

In the linear case, the good i has price p_i and the cost of a bundle x is px. More generally, we define a price system as a map $\psi : \mathbb{R}_+^L \to \mathbb{R}_+$, where $\psi(x)$ is interpreted as the cost to the consumer of the bundle x. We require ψ to be continuous, strictly increasing (in the sense that $\psi(x'') > \psi(x')$ whenever $x'' > x'$), and that, for any number $M > 0$, $\psi^{-1}(M)$ is a bounded set. The last condition means that the consumer could not acquire an arbitrarily large bundle with finite expenditure. We assume that both the *price system* and the bundle chosen by the consumer are observed, so that a data set is formally a collection $\mathcal{D} = \{(\psi^t, x^t)\}_{t=1}^T$. This data set is rationalized by an expenditure augmented utility function $U : \mathbb{R}_+^L \times \mathbb{R}_- \to \mathbb{R}$ if $x^t \in \arg\max_{x \in \mathbb{R}_+^L} U(x, -\psi^t(x))$ for all $t \in T$.

The notion of revealed preference over prices can be extended to a revealed preference over price systems. Let $\mathcal{P} := \{\psi^t\}_{t \in T}$. We say that price $\psi^{t'}$ is *directly revealed preferred* (*directly revealed strictly preferred*) to ψ^t if $\psi^{t'}(x^t) \leq (<) \psi^t(x^t)$; we denote this by $\psi^{t'} \succeq_p (\succ_p) \psi^t$.

As in the linear case, we denote the transitive closure of $\succeq_p$ by $\succeq_p^*$, that is, $\psi^{t'} \succeq_p^* \psi^t$ if there are t_1, t_2,...,t_N in T such that $\psi^{t'} \succeq_p \psi^{t_1}$, $\psi^{t_1} \succeq_p \psi^{t_2}, \ldots, \psi^{t_{N-1}} \succeq_p \psi^{t_N}$, and $\psi^{t_N} \succeq_p \psi^{t'}$; in this case we say that $\psi^{t'}$ is *revealed preferred* to ψ^t. If anywhere along this sequence, it is possible to replace $\succeq_p$ with $\succ_p$ then we denote that relation by $\psi^{t'} \succ_p^* \psi^t$ and say that $\psi^{t'}$

is *strictly revealed preferred* to ψ^t. It is straightforward to check that if $\mathcal{D}$ can be rationalized by an expenditure augmented utility function then it obeys the following generalization of GAPP to nonlinear price systems:

there do not exist observations $t, t' \in T$ *such that* $\psi^{t'} \succeq_p^* \psi^t$ *and* $\psi^t \succ_p^* \psi^{t'}$.

The following result says that the converse is also true. The proof uses the extension of Afriat's Theorem to nonlinear prices and general consumption spaces[15] found in Nishimura, Ok, and Quah (2017).

Theorem 2. *Given a data set* $\mathcal{D} = \{(\psi^t, x^t)\}_{t=1}^T$, *the following are equivalent:*

(1) $\mathcal{D}$ *can be rationalized by an augmented utility function.*
(2) $\mathcal{D}$ *satisfies nonlinear GAPP.*
(3) $\mathcal{D}$ *can be rationalized by an augmented utility function* U *that is strictly increasing and continuous, such that* $\max_{x \in \mathbb{R}_+^L} U(x, -\psi(x))$ *has a solution for any price system* ψ.

3. The Stochastic Model

In this section, we develop the stochastic version of expenditure-augmented utility model. We begin by explaining the corresponding extension of the Afriat model, as found in McFadden and Richter (1991), McFadden (2005) (henceforth to be referred to as MR).

3.1. *Rationalization by Random Utility*

Suppose that instead of observing single choices on T budget sets, the econometrician observes choice probabilities on each budget set. Our preferred interpretation is that each observation corresponds to the distribution of choices made by a population of consumers and the data set consists of a repeated cross section of such choice probabilities. This interpretation is appropriate for our empirical application which uses repeated cross-sectional data from a population.

We denote the budget set corresponding to observation t by $B^t := \{x \in \mathbb{R}_+^L : p^t x = 1\}$. In this model, only relative prices matter, so we can scale prices and normalize income to 1 without loss of generality. We use $\mathring{\pi}^t$ to denote the probability measure of choices on budget set B^t at observation t. Thus, for any subset $X^t \subset B^t$, $\mathring{\pi}^t(X^t)$ denotes the probability that the choices lie in the subset X^t. Following MR and Kitamura and Stoye (2017) (henceforth, referred to as KS), we assume that the econometrician observes the stochastic data set $\widetilde{\mathscr{D}} := \{(B^t, \mathring{\pi}^t)\}_{t=1}^T$, which consists of a finite collection of budget sets along with the corresponding choice probabilities. Note that, in practice, $\mathring{\pi}^t$ needs to be estimated.

For ease of exposition, we also impose the following assumption on the data.

[15]As with Theorem 1, the proof is for the more general case where the consumption space is a subset $X \subset \mathbb{R}_+^L$.

Assumption 1. For all $t, t' \in T$ with $B^t \neq B^{t'}$, the choice probabilities satisfy $\mathring{\pi}^t \left(\{B^t \cap B^{t'}\} \right) = 0$.

In other words, the choice probability measure $\mathring{\pi}$ has no mass where the budget sets intersect. This is convenient because it simplifies some of the definitions that follow.[16]

A random utility is denoted by a measure $\widetilde{\mu}$ over the set of locally nonsatiated utility functions defined on $\mathbb{R}^L_+$, which we denote by $\widetilde{\mathscr{U}}$. The data set $\widetilde{\mathscr{D}}$ is said to be *rationalized by a random utility model* if there exists a random utility $\tilde{\mu}$ such that for all $X_t \subset B^t$,

$$\mathring{\pi}^t(X^t) = \widetilde{\mu}(\widetilde{\mathscr{U}}(X^t)) \text{ for all } t \in T, \text{ where } \widetilde{\mathscr{U}}(X^t) := \left\{ \widetilde{U} \in \widetilde{\mathscr{U}} : \operatorname*{argmax}_{x \in B^t} \widetilde{U}(x) \in X^t \right\}.$$

In other words, to rationalize $\widetilde{\mathscr{D}}$ we need to find a distribution on the family of utility functions that generates a demand distribution at each budget set B^t corresponding to what was observed.

Crucially, McFadden (2005) observed that this problem can be discretized as follows. Let $\{B^{1,t}, \ldots, B^{I_t,t}\}$ denote the collection of subsets (which we call *patches*) of the budget B^t where each subset has as its boundaries the intersection of B^t with other budget sets and/or the boundary hyperplanes of the positive orthant. Formally, for all $t \in T$ and $i_t \neq i_t'$, each set in $\{B^{1,t}, \ldots, B^{I_t,t}\}$ is closed and convex and, in addition, the following hold:

(i) $\cup_{1 \leq i_t \leq I_t} B^{i_t,t} = B^t$,
(ii) $\text{int}(B^{i_t,t}) \cap B^{t'} = \phi$ for all $t' \neq t$ that satisfy $B^t \neq B^{t'}$,
(iii) $B^{i_t,t} \cap B^{i_t',t} \neq \phi$ implies that $B^{i_t,t} \cap B^{i_t',t} \subset B^{t'}$ for some $t' \neq t$ that satisfies $B^t \neq B^{t'}$,

where $\text{int}(\cdot)$ denotes the relative interior of a set.

We use the vector $\pi^t \in \Delta^{I_t}$ belonging to the I^t dimensional simplex to denote the *discretized choice probabilities* over the collection $\{B^{1,t}, \ldots, B^{I_t,t}\}$. Formally, coordinate i_t of π^t is given by

$$\pi^{i_t,t} = \mathring{\pi}^t \left(B^{i_t,t} \right), \quad \text{for all } B^{i_t,t} \in \left\{ B^{1,t}, \ldots, B^{I_t,t} \right\}.$$

Even though there may be t, i_t, i_t' for which $B^{i_t,t} \cap B^{i_t',t} \neq \phi$ (as these sets may share parts of their boundaries), Assumption 1 guarantees that π^t is still a probability measure since choice probabilities on the boundaries of $B^{i_t,t}$ have measure 0. We denote $\pi := (\pi^1, \ldots, \pi^T)'$.

We call a deterministic data set $\mathcal{D} = \{(p^t, x^t)\}_{t=1}^T$ *typical* if, for all t, there is i^t such that $x^t \in \text{int}(B^{i^t,t})$ for all $t \in T$; in other words, x^t lies in the interior of some patch at each observation.[17] If a typical deterministic data set $\mathcal{D} = \{(p^t, x^t)\}_{t=1}^T$ satisfies GARP, then for

[16] This simplification is not conceptually necessary for the procedure described here or its adaptation to our setting in the next subsection. See the explanations given in KS, all of which also apply here.

[17] The choice of terminology reflects the the fact that most data sets used in the literature, including ours, are 'typical.'

all other $\acute{x}^t \in \text{int}(B^{i^t,t})$, $t \in T$, the data set $\{(p^t, \acute{x}^t)\}_{t=1}^T$ also satisfies GARP. This is because the revealed preference relations $\succeq_x$, $\succ_x$ are determined only by where a choice lies on the budget set relative to its intersection with another budget. Thus, as far as testing rationality is concerned, all choices in a given set $\text{int}(B^{i^t,t})$ are interchangeable. Therefore, we may classify all typical deterministic data sets according to the patch occupied by the bundle x^t at each budget set B^t. In formal terms, we associate to each typical deterministic data set $\mathcal{D}$ that satisfies GARP a vector $a = \left(a^{1,1}, \ldots, a^{I_T,T}\right)$ where

$$a^{i_t,t} = \begin{cases} 1 & \text{if } x^t \in B^{i_t,t}, \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

Notice that we have now partitioned all typical deterministic data sets obeying GARP (of which there are infinitely many) into a *finite* number of distinct classes or types, based on the vector a associated with each data set. We use A to denote the matrix whose columns consist of all the a vectors corresponding to these GARP-obeying types (where the columns can be arranged in any order) and use $|A|$ to denote the number of such columns (which is also the number of types).

The problem of finding a measure $\widetilde{\mu}$ on the family of utility functions to rationalize $\widetilde{\mathscr{D}}$ is essentially one of disaggregating $\widetilde{\mathscr{D}}$ into rationalizable deterministic data sets or, given Afriat's Theorem, into deterministic data sets obeying GARP. If we ignore non-typical deterministic data sets (which is justified because of Assumption 1), this is in turn equivalent to finding weights on the finitely many distinct types represented by the columns of A, so that their aggregation at each observation coincides with the discretized choice probabilities. The following result (which we credit to MR though it does not appear as stated in their work) summarizes these observations.

MR Theorem. *Suppose that the stochastic data set $\widetilde{\mathscr{D}} = \{(B^t, \mathring{\pi}^t)\}_{t=1}^T$ satisfies Assumption 1. Then $\widetilde{\mathscr{D}}$ is rationalized by a random utility model if and only if there exists a $\nu \in \mathbb{R}_+^{|A|}$ such that the discretized choice probabilities π satisfy $A\nu = \pi$.*[18]

Before we turn to the stochastic version of the expenditure-augmented utility model, it is worth highlighting an important shortcoming of the setting envisaged by MR: data sets typically do not take the presumed form $\widetilde{\mathscr{D}} = \{(B^t, \mathring{\pi}^t)\}_{t=1}^T$. This is because even consumers that face the same prices on the L observed goods (as is commonly assumed in applications with repeated cross sectional data, for instance, Blundell, Browning, and Crawford (2008)) will typically spend different amounts on these goods and, therefore, the observed choices of the population will lie on different budget sets. Furthermore, even if one conditions on that part of the population that chooses the same expenditure

[18]Since any ν that satisfies $A\nu = \pi$ must also satisfy $\sum_{j=1}^{|A|} \nu_j = 1$, we do not need to explicitly impose the latter condition.

level at some vector of prices, as these prices change, they will go on to choose different expenditure levels. The theory set out by MR to test the random utility model relies heavily on the stylized environment they assume and it is not at all clear how it could be modified to test the model when expenditure is allowed to be endogenous. Recently, KS operationalized the theory by MR and developed econometric methodologies to implement it. Moreover, KS presented an application of their test for rationalizability to the FES data set (one of the two data sets we used in Section 5), though obtaining $\{\mathring{\pi}\}_{t=1}^{T}$ for this application necessitated estimation of counterfactuals, that is, what the distribution of demand would be if, hypothetically, all consumers were restricted to the same budget set. Inevitably, estimating these distributions will require the use of instrumental variables (to get round the endogeneity of expenditure) and can involve a variety of assumptions on the smoothness of Engel curves, the nature of unobserved heterogeneity across individuals, etc.

3.2. *Rationalization by Random Augmented Utility*

Once again, the starting point of our analysis is a stochastic data set $\mathscr{D} := \{(p^t, \mathring{\pi}^t)\}_{t=1}^{T}$ (of price-probability pairs) which consists of a finite set of distinct prices along with a corresponding distribution over chosen bundles. But there is one important departure from the previous section (where the data consisted of budget-probability pairs): we no longer require the support of $\mathring{\pi}^t$ to lie on the budget set B^t; instead the support could be any set in $\mathbb{R}^L_+$. In other words, we no longer require all consumers to incur the same expenditure at each price observation; each consumer in the population can decide how much she wishes to spend on the L observed goods and this could differ across consumers and across price observations. As we pointed out at the end of Section 3.1, this is the form that data typically takes. Also, as in the previous section, $\mathring{\pi}^t$ needs to be estimated in practice and this is the source of statistical uncertainty that we address in Section 4.

A random expenditure-augmented utility is denoted by a measure μ over the set of augmented utility functions which we denote by $\mathscr{U}$.

Definition 3.1. The data set $\mathscr{D}$ is said to be *rationalized by the random augmented utility model* if there exists a random augmented utility μ such that for all $X_t \subset \mathbb{R}^L_+$,

$$\mathring{\pi}^t(X^t) = \mu(\mathscr{U}(X^t)) \text{ for all } t \in T, \text{ where } \mathscr{U}(X^t) := \left\{ U \in \mathscr{U} \,:\, \operatorname*{argmax}_{x \in \mathbb{R}^L_+} U(x, -p^t x) \in X^t \right\}.$$

In actual empirical applications, observations are typically made over time. Therefore, we are effectively asking whether or not $\mathscr{D}$ is generated by utility maximization over a distribution of augmented utility functions that is stable over the period where observations are taken. This assumption is plausible if (i) there is no change in the prices of the unobserved goods or, more realistically, that these changes could be adequately accounted for by the use of a deflator, and (ii) there is sufficient stability in the way consumers in the population view their long term economic prospects, so that there are only small changes (in the sense of Proposition 1) in their willingness to trade off consumption of a bundle of goods in L with the expenditure it incurs.

The problem of finding a measure μ to rationalize $\mathscr{D}$ is essentially one of disaggregating $\mathscr{D}$ into deterministic data sets rationalizable by augmented utility functions or, given Theorem 1, into deterministic data sets obeying GAPP. Crucially, Proposition 2 tells us that a deterministic data set obeys GAPP if and only if its expenditure-normalized version obeys GARP. It follows that μ exists if and only if the normalized version of the stochastic data set $\mathscr{D}$ obeys the condition identified by MR (as stated in Section 3.1).

To set this out more formally, we first define the *normalized choice probability* $\breve{\pi}^t$ corresponding to $\mathring{\pi}^t$ by scaling observations from the entire orthant onto the budget plane B^t generated by prices p^t and expenditure 1. Formally,

$$\breve{\pi}^t(X^t) = \mathring{\pi}^t\left(\left\{x : \frac{x}{p^t x} \in X^t\right\}\right), \quad \text{for all } X^t \subset B^t \text{ and all } t \in T.$$

We suppose that Assumption 1 holds on the normalized data set $\{(B^t, \breve{\pi}^t)\}_{t=1}^T$; abusing the terminology somewhat, we shall say that $\mathscr{D}$ obeys Assumption 1.[19] We then define the patches on the budgets set B^t (as in Section 3.1) and denote them by $\{B^{1,t}, \ldots, B^{I_t,t}\}$. With these patches in place, we derive the *normalized and discretized choice probabilities* $\pi^t = (\pi^{1,t}, \ldots, \pi^{I_t,t})$ from $\breve{\pi}^t$ by assigning to each $\pi^{i_t,t}$ the normalized choice probability $\breve{\pi}^t(B^{i_t,t})$ corresponding to $B^{i_t,t}$. Finally, we construct the matrix A, whose columns are defined by (11); the columns represent distinct GARP-obeying types, which by Proposition 2 coincides with the distinct GAPP-obeying types. The rationalizability of $\mathscr{D}$ can then be established by checking if there are weights on these types that generate the observed normalized and discretized choice probabilities $\pi = (\pi^1, \ldots, \pi^T)$. The following result summarizes these observations.

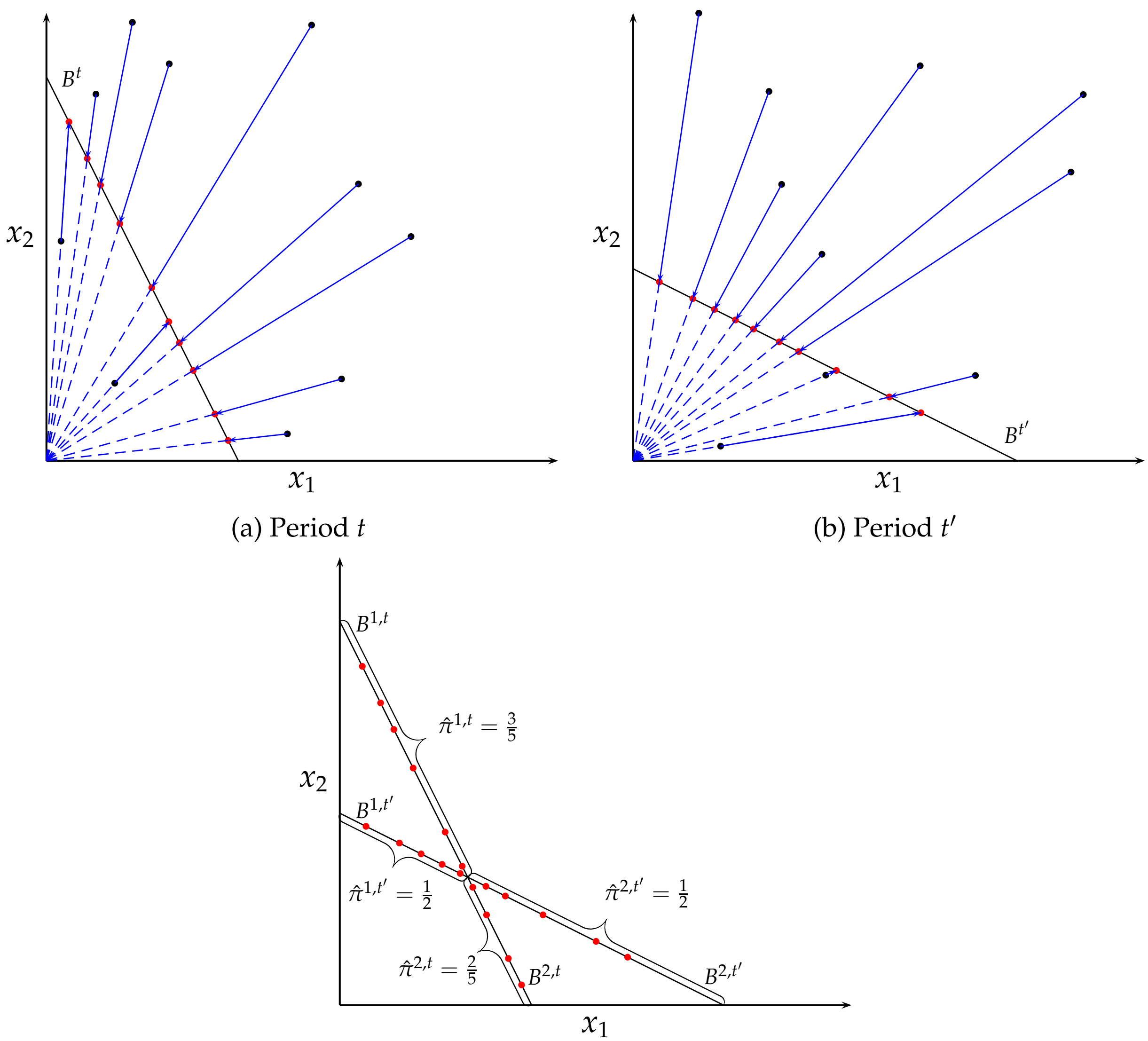
Theorem 3. *Let $\mathscr{D} = \{(p^t, \mathring{\pi}^t)\}_{t=1}^T$ be a stochastic data set obeying Assumption 1. Then $\mathscr{D}$ is rationalized by the random augmented utility model if and only if there exists a $\nu \in \mathbb{R}_+^{|A|}$ such that the normalized and discretized choice probabilities π (derived from $\mathring{\pi}$) satisfy $A\nu = \pi$.*

It is worth emphasizing that this theorem provides us with a very clean procedure for testing the random augmented utility model. If we were testing the random utility model, then the MR test requires a data set where expenditures are common across consumers as

[19] A sufficient condition for Assumption 1 is that $\mathring{\pi}$ assigns 0 probability to sets with a Lebesgue measure of 0. Also, as before, this is merely for ease of exposition: our test does not depend on this assumption and, importantly, the data in our empirical application satisfies Assumption 1.

a starting point; since this is not commonly available it would have to be estimated, which in turn requires an additional econometric procedure with all its attendant assumptions. By contrast, to test the random augmented utility model, all we have to do is apply the MR test to the expenditure-normalized data set $\widetilde{\mathscr{D}} = \{(p^t, \breve{\pi}^t)\}_{t=1}^T$, which is obtained (via a simple transformation purely as a consequence of the theoretical model itself) from the original data set $\mathscr{D} = \{(p^t, \mathring{\pi}^t)\}_{t=1}^T$.

We end this subsection with an example that makes explicit the operationalization of Theorem 3 using data where the normalized and discretized choice probabilities are determined by the sample frequency of choices.



(a) Period t

(b) Period t'

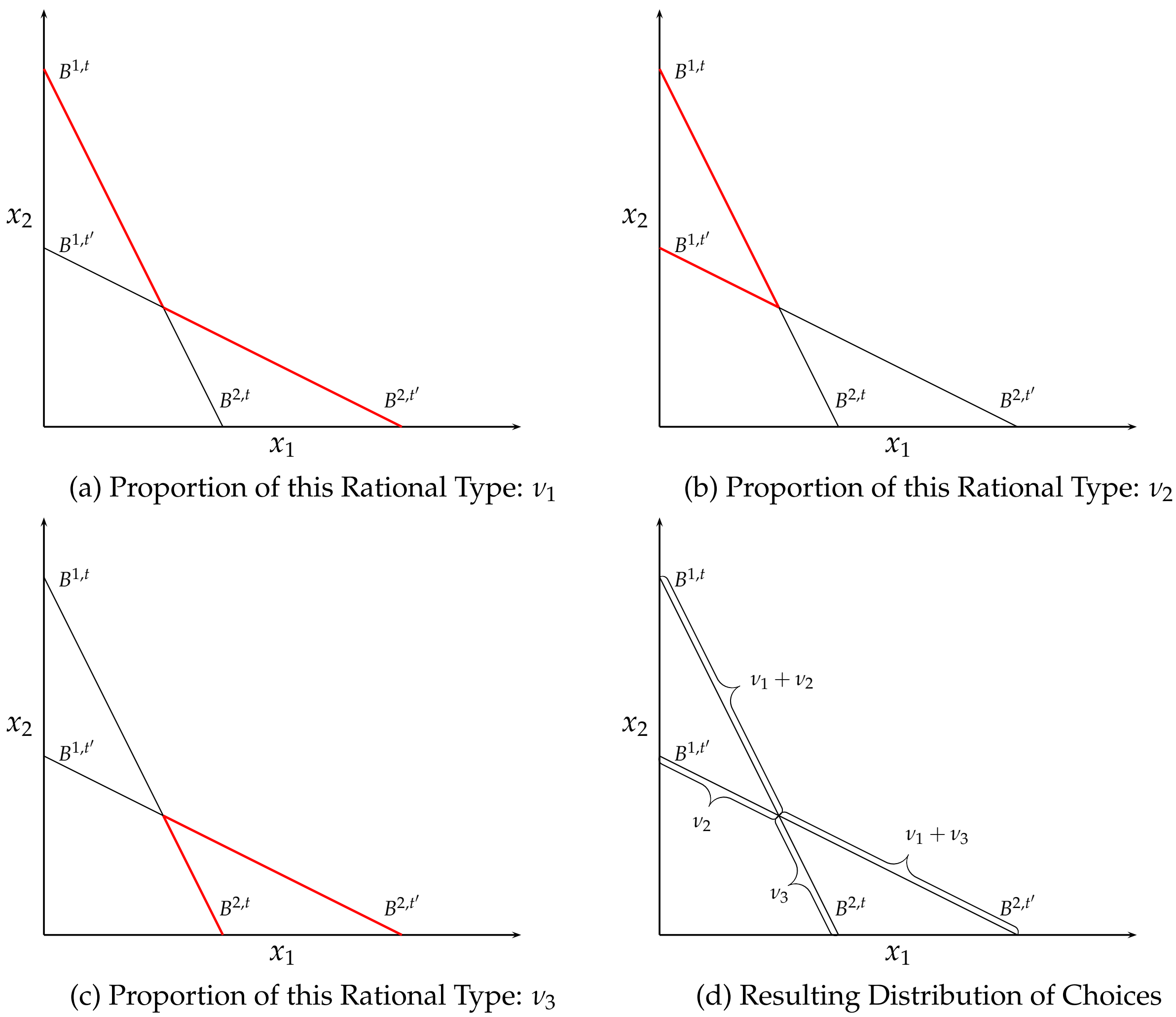
(c) The Normalized and Discretized Empirical Choice Probabilities $\hat{\pi}$

FIGURE 4. Observed and Scaled Choice Data

Example 3. Suppose the econometrician observes the set of ten choices at two price vectors, $p^t = (2,1)$ and $p^{t'} = (1,2)$, given by the black points in Figures 4a and 4b. These figures also demonstrate how the choices are scaled (to the red points) on to the normalized budget sets. Figure 4c then shows that the choice probabilities

$$\hat{\pi} = \left(\frac{3}{5}, \frac{2}{5}, \frac{1}{2}, \frac{1}{2}\right)'$$

(the hat notation refers to the fact that the choice probabilities are derived from the sample choice frequencies which is how we estimate choice probabilities) are determined by the proportion of the normalized choices that lie on each segment of the budget lines. Lastly, Figure 5 illustrates the various rational types for these two budget sets. Note that GARP



(a) Proportion of this Rational Type: ν_1

(b) Proportion of this Rational Type: ν_2

(c) Proportion of this Rational Type: ν_3

(d) Resulting Distribution of Choices

FIGURE 5. Distribution of Rational Types

(equivalently, GAPP) violations only occur when the choices lie on $B^{2,t}$ and $B^{1,t'}$. The resulting A matrix is given by

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \text{ and, additionally, } A\nu = \begin{pmatrix} \nu_1 + \nu_2 \\ \nu_3 \\ \nu_2 \\ \nu_1 + \nu_3 \end{pmatrix}.$$

The first, second and third column of A correspond to the three types of GARP-consistent demand behavior, which are depicted in Figures 5a, 5b and 5c respectively. If the proportion of the three types in the population is ν_1, ν_2 and ν_3, the resulting distribution on the segments of the budget sets are given by $A\nu$, the expression for which is displayed above and depicted in Figure 5d. Theorem 3 says that rationalization is equivalent to the existence of $\nu \in \mathbb{R}^3_+$ such that $A\nu = \hat{\pi}$.

The data from this example can be rationalized by the distribution of rational types

$$\nu = \left(\frac{1}{10}, \frac{1}{2}, \frac{2}{5}\right)'.$$

While the rationalizing distribution ν is unique in this example, this is typically not the case (as the equation $A\nu = \pi$ may have multiple solutions for ν). Notice also that it is not the case that a solution always exists. Indeed, if $\hat{\pi}^{1,t'} > \hat{\pi}^{1,t}$, then the choice probabilities would not be rationalizable as $\nu_2 > \nu_1 + \nu_2$ is not possibile.

3.3. *Welfare Comparisons*

Since the test for rationalizability involves finding a distribution ν over different types, it is possible to use this distribution for welfare comparisons: for any two prices in the data set and given a distribution ν that rationalizes $\mathscr{D}$, we can determine the proportion of types who are revealed better off and the proportion who are revealed worse off. However, since there may be multiple ν that satisfy $A\nu = \pi$, the welfare rankings extractable from the data will typically be in terms of bounds.

To be specific, suppose we would like to determine the welfare effect of a price change from $p^{t'}$ to p^t. Let $\mathbb{1}_{t \succeq^*_p t'}$ denote a vector of length $|A|$ such that the j^{th} element is 1 if $p^t \succeq^*_p p^{t'}$ for the rational type corresponding to column j of A and 0 otherwise. In words, $\mathbb{1}_{t \succeq^*_p t'}$ enumerates the set of rational types for which p^t is revealed preferred to $p^{t'}$. For a rationalizable data set $\mathscr{D}$, the solution to the optimization problem

$$\underline{\mathcal{N}}_{t \succeq^*_p t'} := \begin{array}{c} \min_\nu \ \mathbb{1}_{t \succeq^*_p t'} \, \nu, \\ \text{subject to } \ A\nu = \pi, \end{array} \tag{12}$$

gives the lower bound, while

$$\overline{\mathcal{N}}_{t \succeq^*_p t'} := \begin{array}{c} \max_\nu \ \mathbb{1}_{t \succeq^*_p t'} \, \nu, \\ \text{subject to } \ A\nu = \pi, \end{array} \tag{13}$$

gives the upper bound on the proportion of consumers who are revealed better off at prices p^t compared to $p^{t'}$.

Since (12) and (13) are both linear programming problems (which have solutions if, and only if, $\mathscr{D}$ is rationalizable), they are easy to implement and computationally efficient. Suppose that the solutions are $\underline{\nu}$ and $\overline{\nu}$ respectively; then for any $\beta \in [0,1]$, $\beta\overline{\nu} + (1-\beta)\underline{\nu}$ is also a solution to $A\nu = \pi$ and, in this case, the proportion of consumers who are revealed better off at p^t compared to $p^{t'}$ is exactly $\beta\,\underline{\mathcal{N}}_{t \succeq_p^* t'} + (1-\beta)\,\overline{\mathcal{N}}_{t \succeq_p^* t'}$. In other words, the proportion of consumers who are revealed better off can take any value in the interval $[\underline{\mathcal{N}}_{t \succeq_p^* t'}, \overline{\mathcal{N}}_{t \succeq_p^* t'}]$.

Proposition 3 tells us that the revealed preference relations are tight, in the sense that if, for some consumer, p^t is not revealed preferred to $p^{t'}$ then there exists an augmented utility function which rationalizes her consumption choices and for which she strictly prefers $p^{t'}$ to p^t. Given this, we know that, amongst all rationalizations of $\mathscr{D}$, $\underline{\mathcal{N}}_{t \succeq_p^* t'}$ is also the infimum on the proportion of consumers who are better off at p^t compared to $p^{t'}$. At the other extreme, we know that there is a rationalization for which the proportion of consumers who are revealed better off at $p^{t'}$ compared to p^t is as low as $\underline{\mathcal{N}}_{t' \succeq_p^* t}$.[20] Applying Proposition 3 again, a rationalization could be chosen such that all other consumers prefer p^t to $p^{t'}$. Therefore, across all rationalizations of $\mathscr{D}$, $1 - \underline{\mathcal{N}}_{t' \succeq_p^* t}$ is the supremum on the proportion of consumers who prefer p^t and $p^{t'}$. Lastly, since the set of distributions ν that rationalize $\mathscr{D}$ form a convex set, the true proportion of consumers who are better off can lie anywhere between the two extremes identified.

The following proposition summarizes these observations.

Proposition 4. *Let $\mathscr{D} = \{(p^t, \mathring{\pi}^t)\}_{t=1}^T$ be a stochastic data set that satisfies Assumption 1 and is rationalized by the augmented utility model.*

(1) Then for every $\eta \in [\underline{\mathcal{N}}_{t \succeq_p^* t'}, \overline{\mathcal{N}}_{t \succeq_p^* t'}]$, there is rationalization of $\mathscr{D}$ for which η is the proportion of consumers who are revealed better off at p^t compared to $p^{t'}$.

(2) For any rationalization of $\mathscr{D}$, there is a proportion of consumers who are better off at p^t compared to $p^{t'}$. This proportion can take any value in the interval $[\underline{\mathcal{N}}_{t \succeq_p^* t'}, 1 - \underline{\mathcal{N}}_{t' \succeq_p^* t}]$.

It may be helpful to consider how Proposition 4 applies to Example 3. In that case, there is a unique solution to $A\nu = \pi$ and the proportion of consumers who are revealed better off at p^t compared to $p^{t'}$ is $\nu_2 = 1/2$, while the proportion who are revealed better off at $p^{t'}$ to p^t is $\nu_3 = 2/5$. Those consumers who belong to neither of these two types could be either better or worse at p^t compared to $p^{t'}$. Therefore, across all rationalizations of that data set, the proportion of consumers who are better off at p^t compared to $p^{t'}$ can be as low as $1/2$ and as high as $1 - 2/5 = 3/5$.

[20] Because of Assumption 1, we could assume that this is a strict revealed preference.

4. THE ECONOMETRIC METHODOLOGY

In this section, we develop the econometric methodology required to test our model and conduct welfare analysis. Specifically, we first briefly discuss how Theorem 3 allows us to statistically test the random augmented utility model by using the methodology developed by KS. We then present the novel econometric theory that is required to extract the welfare bounds (characterized in statement (1) of Proposition 4) from the data. As shall become clear, the procedure we develop (for inference on linear transforms of partially identified vectors) has other applications beyond our specific empirical analysis.

4.1. *Testing the Random Augmented Utility Model*

Recall from Theorem 3 that, given a stochastic data set $\mathscr{D} = \{(p^t, \mathring{\pi}^t)\}_{t=1}^T$, a test of the random augmented utility model is a test of the hypothesis

$$H_0 : \exists \nu \in \mathbb{R}_+^{|A|} \text{ such that } A\nu = \pi,$$

where π is the normalized and discretized choice probability vector derived from $\mathring{\pi}$. KS restate this test in an equivalent (and more convenient) form

$$H_0 : \min_{\nu \in \mathbb{R}_+^{|A|}} [\pi - A\nu]'\Omega[\pi - A\nu] = 0, \tag{14}$$

where Ω is a positive definite matrix. The solution of the above minimization problem is the projection of π onto the cone $\left\{A\nu \mid \nu \in \mathbb{R}_+^{|A|}\right\}$ under the weighted norm $\|\zeta\|_\Omega = \sqrt{\zeta'\Omega\zeta}$, where $\zeta \in \mathbb{R}^{|A|}$. The corresponding value of the objective function is the squared length of the projection residual vector and is zero if, and only if, π can be stochastically rationalized.

Of course, in practice, the data corresponding to each price p^t is not in the form of a distribution $\mathring{\pi}^t$ but instead consists of a cross section of choices $\{x_{n_t}^t\}_{n_t=1}^{N_t}$ made by N_t households. We estimate π by its sample analog $\hat{\pi} = (\hat{\pi}^1, \ldots, \hat{\pi}^T)$ where each $\hat{\pi}^t$ is obtained by normalizing and discretizing the empirical distribution of the observed choices $\{x_{n_t}^t\}_{n_t=1}^{N_t}$ (as in Figure 4).[21] To test (14), KS use a sample counterpart

$$J_N := N \min_{\nu \in \mathbb{R}_+^{|A|}} [\hat{\pi} - A\nu]'\Omega[\hat{\pi} - A\nu], \tag{15}$$

where $N = \sum_{t=1}^T N_t$ denotes the total number of observations (the sum of the number of households across years). Here, the minimizing value $\hat{\nu}$ may not be unique but the implied choice probabilities $\hat{\eta} := A\hat{\nu}$ are. Note that $\hat{\eta} = \hat{\pi}$ and $J_N = 0$ iff the sample choice frequencies can be rationalized by a random augmented utility model. In this

[21] In the data we employ for our empirical application, there are no observations that lie on the intersection of any two normalized budget sets so $\hat{\pi}^t$ determines a probability measure on $\{B^{1,t}, \ldots, B^{I_t,t}\}$. In other words, the data in our empirical application satisfies Assumption 1.

case, the null hypothesis is trivially accepted. We determine critical values for the test statistic by employing the modified bootstrap procedure of KS.

4.2. *Bounds for Linear Transforms of a Partially Identified Vector*

Recall that our model allows for welfare comparisons where, for any two prices p^t and $p^{t'}$ in our data, we can determine bounds $[\underline{\mathcal{N}}_{t \succeq_p^* t'}, \overline{\mathcal{N}}_{t \succeq_p^* t'}]$ on the proportion of the population that is better off at the former prices. This is an instance of a more general problem of determining bounds on a linear transform

$$\theta = \rho\nu \quad \text{subject to } A\nu = \pi,$$

where $\rho \in \mathbb{R}^{|A|}$ is a given vector and $\theta \in \mathbb{R}$ is the parameter of interest. In our main application, $\rho = \mathbb{1}_{t \succeq_p^* t'}$ (which, recall, is a vector where the j^{th} element is 1 if $p^t \succeq_p^* p^{t'}$ for the rational type corresponding to column j of A and 0 otherwise) and $\theta = \mathcal{N}_{t \succeq_p^* t'}$. It is worth stressing that the methodology we develop here is valid not just for our model, but for inference in any setting where identification has the same structure (such as KS). In this section, we provide a high-level description of our testing procedure, the formal treatment can be found in Appendix B.

The goal is to learn about the identified set Θ_I, which is given by

$$\Theta_I := \{\theta \mid \pi \in \mathcal{S}(\theta)\}, \text{ where } \mathcal{S}(\theta) := \left\{ A\nu \mid \rho\nu = \theta,\ \nu \in \Delta^{|A|-1} \right\}.$$

We do this by inverting a test of

$$\pi \in \mathcal{S}(\theta) \tag{16}$$

or, equivalently,

$$\min_{\nu \in \Delta^{|A|-1}, \theta = \rho\nu} [\pi - A\nu]'\Omega[\pi - A\nu] = 0.$$

We use a sample counterpart of the above display as the test statistic:

$$\begin{aligned} J_N(\theta) &= \min_{\nu \in \Delta^{|A|-1}, \theta = \rho\nu} [\hat{\pi} - A\nu]'\Omega[\hat{\pi} - A\nu] \\ &= \min_{\eta \in \mathcal{S}(\theta)} N[\hat{\pi} - \eta]'\Omega[\hat{\pi} - \eta]. \end{aligned}$$

The challenge is to compute an appropriate critical value for this test statistic which accounts for the fact that its limiting distribution depends discontinuously on nuisance parameters in a very complex manner.[22] This issue is similar to that faced in the canonical moment inequality testing problem (see, for instance, Andrews and Soares (2010)) which has a tight theoretical link, and is in fact the dual, to our problem. Formally, it can be

[22] It is, in principle, possible to derive critical values by ($m < n-$)subsampling. While asymptotically valid, critical values derived in this way can be quite conservative, as has been forcefully argued in Andrews and Soares (2010), page 137.

shown that $\pi \in \mathcal{S}(\theta)$ if, and only if,

$$\pi \text{ satisfies } B\pi \leq 0, \; \tilde{B}\pi = d(\theta) \text{ and } \mathbf{1}'\pi = 1, \tag{17}$$

for some fixed matrices B, $\tilde{B}$ and a vector d that depends on θ (see the Appendix B for details), where $\mathbf{1}$ is the $|A|$-vector of ones. Importantly, for fixed θ and prices $\{p_t\}_{t=1}^{T}$, these objects are known to exist and are nonstochastic.

That said, the reason we cannot directly use testing procedures from the moment inequality literature (although we do draw on this literature for the proof of the asymptotic validity of our method) is that numerically deriving the dual representation (17) is computationally prohibitive for problems of empirically relevant scale (such as our empirical application). This computational bottleneck requires us to work directly with (16) and so we instead follow a *tightening* procedure.

The tightening procedure we employ is similar in spirit to that used in KS but has to account for the additional features of our testing problem. In particular, establishing the validity of the tightening method is more intricate as it depends on certain geometric properties of the set $\mathcal{S}(\theta)$. Moreover, unlike the hypothesis considered in KS, our hypothesis (16) depends on a parameter θ that lives in a compact space $\Theta := \left\{\rho\nu : \nu \in \Delta^{|A|-1}\right\}$ and may in practice be close to the boundary of that space. This requires a delicate *restriction-dependent* tightening, which now we describe.

Let τ_N be a tuning parameter chosen such that $\tau_N \downarrow 0$ and $\sqrt{N}\tau_N \uparrow \infty$ and, for $\theta \in \Theta_I$, define the τ_N-tightened version $\mathcal{S}_{\tau_N}(\theta)$ of $\mathcal{S}(\theta)$, that is,

$$\mathcal{S}_{\tau_N}(\theta) := \{A\nu \mid \rho\nu = \theta, \nu \in \mathcal{V}_{\tau_N}(\theta)\},$$

where $\mathcal{V}_{\tau_N}(\theta)$ is obtained by appropriately constricting (tightening) the set $\Delta^{|A|-1}$, the parameter space for ν in the definition of $\mathcal{S}$. (Appendix B contains the formal definition of $\mathcal{V}_{\tau_N}(\theta)$). As the notation suggests, the way we tighten $\mathcal{V}_{\tau_N}(\theta)$ is dependent on the hypothesized value θ in order to deal with the aforementioned issue associated with the boundary of the parameter space Θ. Note that $\mathcal{V}_{\tau_N}(\theta)$ converges to $\Delta^{|A|-1}$, hence $\mathcal{S}_{\tau_N}(\theta)$ to $\mathcal{S}$ as $\tau_N \to 0$, for every value of θ. Though the definition of $\mathcal{V}_{\tau_N}(\theta)$ provided in Appendix B for a general (linear) counterfactural is rather involved, we show that it simplifies for the welfare comparison application considered in this paper.

The set $\mathcal{S}_{\tau_N}(\theta)$ replaces $\mathcal{S}(\theta)$ in the bootstrap population, inducing locally conservative distortions, which guard against discontinuities in the test statistic's limiting distribution as a function of nuisance parameters. In Appendix B, we show that tightening procedure turns inequality constraints into binding ones in the bootstrap DGP if their degrees of slackness are small, though inequalities with large degrees of slackness remain non-binding after tightening. This prevents the bootstrap procedure from over-rejection when the sampling DGP is near the boundary of inequality constraints implied by (16).

Our bootstrap algorithm proceeds as follows. For each $\theta \in \Theta$

(i) Compute the τ_N-*tightened restricted* estimator of the empirical choice distribution

$$\hat{\eta}_{\tau_N} := \underset{\eta \in \mathcal{S}_{\tau_N}(\theta)}{\operatorname{argmin}} N[\hat{\pi} - \eta]'\Omega[\hat{\pi} - \eta].$$

(ii) Define the τ_N-*tightened recentered* bootstrap estimators

$$\hat{\pi}^{*(r)}_{\tau_N} := \hat{\pi}^{*(r)} - \hat{\pi} + \hat{\eta}_{\tau_N}, \quad r = 1, ..., R,$$

where $\hat{\pi}^{*(r)}$ is a straightforward bootstrap analog of $\hat{\pi}$ and R is the number of bootstrap samples. For instance, in our application, $\hat{\pi}^{*(r)}$ is generated by the simple nonparametric bootstrap of choice frequencies. This is the step that makes constraints with small slack binding.

(iii) For each $r = 1, ..., R$, compute the value of the bootstrap test statistic

$$J^{*(r)}_{N,\tau_N}(\theta) = \min_{\eta \in \mathcal{S}_{\tau_N}(\theta)} N[\hat{\pi}^{*(r)}_{\tau_N} - \eta]'\Omega[\hat{\pi}^{*(r)}_{\tau_N} - \eta].$$

(iv) Use the empirical distribution of $J^{*(r)}_{N,\tau_N}(\theta), r = 1, ..., R$ to obtain the critical value for $J_N(\theta)$.

We obtain a confidence interval for θ by collecting the values of θ that are not rejected by the bootstrap implemented with the algorithm (i)-(iv).

The theorem below establishes the asymptotic validity of the above procedure. Let

$$\mathcal{F} := \{(\theta, \pi) \mid \theta \in \Theta, \pi \in \mathcal{S}(\theta) \cup \mathcal{P}\}$$

where $\mathcal{P}$ denote the set of all π's that satisfy Condition 1 in Appendix for some (common) value of (c_1, c_2).

Theorem 4. *Choose τ_N so that $\tau_N \downarrow 0$ and $\sqrt{N}\tau_N \uparrow \infty$. Also, let Ω be diagonal, where all the diagonal elements are positive. Then under Assumptions 2 and 3,*

$$\liminf_{N\to\infty} \inf_{(\theta,\pi)\in\mathcal{F}} \Pr\{J_N(\theta) \le \hat{c}_{1-\alpha}\} = 1 - \alpha,$$

where $0 \le \alpha \le \frac{1}{2}$ and $\hat{c}_{1-\alpha}$ is the $1-\alpha$ quantile of J^*_{N,τ_N}.

5. Empirical Application

In this section, we test our model and conduct welfare analysis on two separate data sets: the U.K. Family Expenditure Survey (FES) and the Canadian Surveys of Household Spending (SHS). The aim of the empirical analysis is to show that the data supports the model and to demonstrate that the estimated welfare bounds are informatively tight.

We first present the analysis for the FES as it is widely used in the nonparametric demand estimation literature (for instance, by Blundell, Browning, and Crawford (2008), KS, Hoderlein and Stoye (2014), Adams (2016), Kawaguchi (2017)). In the FES, about 7000 households are interviewed each year and they report their consumption expenditures in a variety of different commodity groups. In addition, the data contains a variety of other demographic information (most of which we do not need nor exploit). Consumption bundles are derived from the expenditures by using the annual (real) price indices for these commodity groups taken from the annual Retail Prices Index. To directly contrast our analysis with that of Blundell, Browning, and Crawford (2008), we follow them and restrict attention to households with cars and children, leaving us with roughly 25% of the original data.

We implement tests for 3, 4, and 5 composite goods. The coarsest partition of 3 goods–food, services, and nondurables–is precisely what is examined by Blundell, Browning, and Crawford (2008) (and we use their replication files). As in KS, we increase the dimension of commodity space by first separating out clothing and then alcoholic beverages from the nondurables. Recall that the test of our model involves projecting all the consumption bundles for a given year onto a single budget set corresponding to expenditure 1 (as in Figure 5d), and choice probabilities are then estimated by the corresponding sample frequencies. It is worth reiterating this strength of our model; since our test can be directly applied to the data, it avoids the layers of smoothing, endogeneity adjustment, and, in some cases, aggregation conditions that characterize other analyses of these data.

While it is, in principle, possible to conduct a single test on the entire data from 1975 to 1999, we instead implement the test in blocks of 6 years. We do so mainly for two reasons. The first is practical; the test involves finding a distribution (ν) over the set of possible rational types and the computational complexity of the test depends the size of this set (given by the size of the A matrix). This in turn is determined by the number of distinct patches formed by the intersection of the budget sets and this grows exponentially in the in the number of years.[23] The second reason is that the main assumption of the model—a time-invariant distribution of augmented utility functions—is more plausible over shorter time horizons. Indeed, we might not expect this to hold over the entire time horizon spanned by the data because there are first-order changes to the U.K. income distribution over this period (Jenkins, 2016).

Table 1 (columns correspond to different blocks of 6 years and rows contain the values of the test statistic and the corresponding p-values) shows that our model is not rejected by the FES data.[24] A similar result is also obtained by KS (the MR model is not rejected on these data either) but p-values are hard to compare because their analysis incurs two

[23]This computational constraint prevents us from testing more than 8 years at a time.

[24]The proof of the asymptotic validity of the test of our model presumes that $\alpha \leq 0.5$ and so p-values above 0.5 should not be interpreted to indicate anything other than nonsignificance at interesting test sizes.

		Year Blocks									
		75-80	76-81	77-82	78-83	79-84	80-85	81-86	82-87	83-88	84-89
3 Goods	Test Statistic (J_N)	0.337	0.917	0.899	0.522	0.018	0.082	0.088	0.095	0.481	0.556
	p-value	0.04	0.34	0.55	0.59	0.99	0.67	0.81	0.91	0.61	0.48
4 Goods	Test Statistic (J_N)	0.4	0.698	0.651	0.236	0.056	0.036	0.037	0.043	0.043	0.232
	p-value	0.25	0.58	0.63	0.91	0.96	0.99	0.96	0.95	0.99	0.68
5 Goods	Test Statistic (J_N)	0.4	0.687	0.705	0.329	0.003	0.082	0.088	0.104	0.103	0.144
	p-value	0.3	0.66	0.68	0.88	0.999	0.96	0.79	0.85	0.9	0.83

		Year Blocks									
		85-90	86-91	87-92	88-93	89-94	90-95	91-96	92-97	93-98	94-99
3 Goods	Test Statistic (J_N)	0.027	1.42	2.94	1.51	1.72	0	0.313	0.7	0.676	0.26
	p-value	0.69	0.3	0.18	0.24	0.21	1	0.59	0.48	0.6	0.83
4 Goods	Test Statistic (J_N)	0.227	0.025	0.157	0.154	0.004	1.01	0.802	0.872	0.904	0.604
	p-value	0.48	0.96	0.8	0.73	0.97	0.21	0.31	0.57	0.65	0.74
5 Goods	Test Statistic (J_N)	0.031	0.019	0.018	0.019	0.023	0.734	0.612	0.643	0.634	0.488
	p-value	0.85	0.98	0.97	0.91	0.83	0.22	0.4	0.72	0.78	0.79

TABLE 1. Test Statistics and p-values for sequences of 6 budgets of the FES. Bootstrap size is $R = 1000$.

Comparison	Estimated Bounds	Confidence Interval
$p^{1976} \succ_p^* p^{1977}$	$[.150, .155]$	$[.13, .183]$
$p^{1977} \succ_p^* p^{1976}$	$\{.803\}$	$[.784, .831]$
$p^{1979} \succ_p^* p^{1980}$	$[.517, .530]$	$[.487, .56]$
$p^{1980} \succ_p^* p^{1979}$	$\{.463\}$	$[.436, .497]$

TABLE 2. Estimated bounds and confidence intervals for the proportion of consumers who reveal prefer one price to another one in the FES data. Data used are for 1975-1980. Bootstrap size is $R = 1000$.

additional layers of statistical noise by smoothing over expenditure (done by series estimation) and by adjusting for endogeneity. A consequence of this additional noise is that (ceteris paribus) the test of our model is expected to have higher power, as it has nontrivial power against deviations of the order $N^{-1/2}$ from the null, whereas smoothing-based tests requires deviations of the order of a nonparametric rate to have nontrivial power.

We also estimated the bounds $[\underline{\mathcal{N}}_{t \succeq_p^* t'}, \overline{\mathcal{N}}_{t \succeq_p^* t'}]$ on the proportion of households that are better off at prices p^t than at prices $p^{t'}$. For brevity, we present a few representative estimates using data for the years 1975-1980 in Table 2. The column 'Estimated Bounds' are the bounds obtained by calculating $\mathbb{1}_{t \succeq_p^* t'} \nu$ from the (not necessarily unique) values of ν that minimize the test statistic (15). Note that in two cases this estimate is unique while it is not in the other two cases. In either scenario, the procedure set out for calculating confidence intervals in Section 4.2 is applicable and they give the intervals displayed (which must necessarily contain the estimated bounds). It is worth noting that the width of these intervals is less than .1 throughout, so they are quite informative.

For our second empirical application using Canadian data, we use the replication kit of Norris and Pendakur (2013, 2015). Like the FES, the SHS is a publicly available, annual data set of household expenditures in a variety of different categories. It also contains rich demographic data (most of which we once again do not need nor use). We study annual expenditure in 5 categories that constitute a large share of the overall expenditure on nondurables: food purchased (at home and in restaurants), clothing and footwear, health and personal care, recreation, and alcohol and tobacco. The SHS data is rich enough to allow us to analyze the data separately for the nine most populous provinces: Alberta, British Columbia, Manitoba, New Brunswick, Newfoundland, Nova Scotia, Ontario, Quebec, and Saskatchewan. The number of households in each province-year range from 291 (Manitoba, 1997) to 2515 (Ontario, 1997). We use province-year prices indices (as constructed by Norris and Pendakur (2015)) and deflate them using province-year CPI data from Statistics Canada to get real price indices.

Table 4 displays the test statistics and associated p-value for each province and every 6 year block. Compared to the FES data, there are two notable differences. The first is that many more test statistics are exactly zero; that is, the observed choice frequencies are rationalized by the random augmented utility model. The second is that, for a small proportion of year blocks, there are statistically significant positive test statistics (see, for instance, the last three columns for British Columbia). These low p-values suggest that the conservative distortion which we incur to guarantee uniform validity is modest. Additionally, the p-values taken together do not reject the model if multiple testing is taken into account; for example, step-down procedures would terminate at the first step (that is, Bonferroni adjustment). Finally, we also provide some sample welfare bounds in Table 4. The bounds for the SHS are qualitatively similar to those computed for the FES: there are instances of point identification and in the case of partial identification, the bounds are informative and no wider than .1.

We end this section with two remarks. First, while there is an indirect effect through the potential size of the matrix A, neither the computational complexity of our test nor its statistical power depend directly on the dimension of commodity space. In particular, we avoid a statistical curse of dimensionality and most of our tests for a five dimensional commodity space can be replicated on everyday equipment. Second, we note that finite sample power does not necessarily increase in the number of years being tested or the

Province		Year Blocks							
		97-02	98-03	99-04	00-05	01-06	02-07	03-08	04-09
Alberta	Test Statistic (J_N)	.07	0	0	0	0	0	.003	4.65
	p-value	.94	1	1	1	1	1	.98	.04
British Columbia	Test Statistic (J_N)	.89	.56	.48	.07	.05	6.23	8.87	8.71
	p-value	.47	.47	.98	.96	.97	.05	.02	.01
Manitoba	Test Statistic (J_N)	0	0	0	0	0	0	.01	.01
	p-value	1	1	1	1	1	1	1	1
New Brunswick	Test Statistic (J_N)	.08	.05	0	0	0	.60	.58	.57
	p-value	.94	.94	1	1	1	.58	.79	.68
Newfoundland	Test Statistic (J_N)	.10	.32	.29	.29	.38	3.08	2.30	2.08
	p-value	.85	.90	.91	.87	.81	.21	.35	.27
Nova Scotia	Test Statistic (J_N)	.05	.03	0	0	0	0	.93	1.02
	p-value	.97	.98	1	1	1	1	.69	.58
Ontario	Test Statistic (J_N)	.064	.040	.035	0	0	0	0	0
	p-value	.98	.95	.91	1	1	1	1	1
Quebec	Test Statistic (J_N)	.11	0	0	0	0	.51	.54	.49
	p-value	.88	1	1	1	1	.67	.67	.65
Saskatchewan	Test Statistic (J_N)	0	0	0	0	0	.02	.02	0
	p-value	1	1	1	1	1	1	1	1

TABLE 3. Test Statistics and p-values for sequences of 6 budgets of the SHS. Bootstrap size is $R = 1000$.

Comparison	Estimated Bounds	Confidence Interval
$p^{1998} \succ_p^* p^{2001}$	$\{.099\}$	$[.073, .125]$
$p^{2001} \succ_p^* p^{1998}$	$\{.901\}$	$[.875, .927]$
$p^{1999} \succ_p^* p^{2002}$	$[.299, .341]$	$[.272, .385]$
$p^{2002} \succ_p^* p^{1999}$	$[.624, .701]$	$[.594, .728]$

TABLE 4. Estimated bounds and confidence intervals for the proportion of consumers who reveal prefer one price to another one in the SHS data. Data used are for 1997-2002 in British Columbia. Bootstrap size is $R = 1000$.

dimension of the commodity space. Either could increase the set of revealed preference relations to be tested which will increase power if one of them is in fact violated; but this has to be traded off against increased noise.

6. CONCLUSION

We developed a revealed preference analysis of a model of consumption in which the consumer maximizes utility over an observed set of purchases, taking into account a disutility of expenditure, but is not subjected to a hard budget constraint. Our model naturally generalizes to a random utility context which is suitable for demand analysis using repeated cross-section data. We show how to statistically test the model and develop novel econometric theory for inference on the proportion of the population that benefits from a price change (and, more generally, for determining bounds on linear transforms of partially identified vectors). The model is supported by empirical evidence from Canadian and U.K. data and these applications show that meaningful welfare comparisons can be extracted from the data.

REFERENCES

ADAMS, A. (2016): "Mutually consistent revealed preference bounds," Manuscript, University of Oxford.

AFRIAT, S. N. (1967): "The Construction of Utility Functions from Expenditure Data," *International Economic Review*, 8(1), 67–77.

AGUIAR, V. H., AND N. KASHAEV (2017): "Stochastic Revealed Preferences with Measurement Error: Testing for Exponential Discounting in Survey Data," Discussion paper, Mimeo.

ANDREWS, D. W., AND G. SOARES (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78, 119–157.

BLUNDELL, R., M. BROWNING, AND I. CRAWFORD (2008): "Best Nonparametric Bounds on Demand Responses," *Econometrica*, 76(6), 1227–1262.

BLUNDELL, R., J. L. HOROWITZ, AND M. PAREY (2012): "Measuring the Price Responsiveness of Gasoline Demand: Economic Shape Restrictions and Nonparametric Demand Estimation," *Quantitative Economics*, 3(1), 29–51.

BROWN, D. J., AND C. CALSAMIGLIA (2007): "The Nonparametric Approach to Applied Welfare Analysis," *Economic Theory*, 31(1), 183–188.

CRAWFORD, I., AND B. DE ROCK (2014): "Empirical Revealed Preference," *Annual Review of Economics*, 6(1), 503–524.

FELDMAN, N. E. (2010): "Mental Accounting Effects of Income Tax Shifting," *Review of Economics and Statistics*, 92(1), 70–86.

FORGES, F., AND E. MINELLI (2009): "Afriat's Theorem for General Budget Sets," *Journal of Economic Theory*, 144(1), 135–145.

FOSTEL, A., H. E. SCARF, AND M. J. TODD (2004): "Two New Proofs of Afriats Theorem," *Economic Theory*, 24(1), 211–219.

HASTINGS, J. S., AND J. M. SHAPIRO (2013): "Fungibility and Consumer Choice: Evidence from Commodity Price Shocks," *Quarterly Journal of Economics*, 128(4), 1449–1498.

HAUSMAN, J. A., AND W. K. NEWEY (2016): "Individual Heterogeneity and Average Welfare," *Econometrica*, 84(3), 1225–1248.

HODERLEIN, S., AND J. STOYE (2014): "Revealed Preferences in a Heterogeneous Population," *The Review of Economics and Statistics*, 96(2), 197–213.

JENKINS, S. P. (2016): "The Income Distribution in the UK," in *Social Advantage and Disadvantage*, ed. by H. Dean, and L. Platt, pp. 135–160. Oxford University Press.

KAWAGUCHI, K. (2017): "Testing rationality without restricting heterogeneity," *Journal of Econometrics*, 197(1), 153 – 171.

KITAMURA, Y., AND J. STOYE (2017): "Nonparametric Analysis of Random Utility Models," *Cowles Foundation Discussion Paper*.

McFadden, D., and K. Richter (1991): "Stochastic rationality and revealed stochastic preference," in *Preferences, Uncertainty and Rationality*, ed. by J. Chipman, D. McFadden, and K. Richter, pp. 161–186. Westview Press, Boulder.

McFadden, D. L. (2005): "Revealed Stochastic Preference: A Synthesis," *Economic Theory*, 26(2), 245–264.

Milkman, K. L., and J. Beshears (2009): "Mental Accounting and Small Windfalls: Evidence from an Online Grocer," *Journal of Economic Behavior and Organization*, 71(2), 384–394.

Nishimura, H., E. A. Ok, and J. K.-H. Quah (2017): "A Comprehensive Approach to Revealed Preference Theory," *American Economic Review*, 107(4), 1239–1263.

Norris, S., and K. Pendakur (2013): "Imputing Rent in Consumption Mmeasures, with an Application to Consumption Poverty in Canada, 1997–2009," *Canadian Journal of Economics*, 46(4), 1537–1570.

——— (2015): "Consumption Inequality in Canada, 1997 to 2009," *Canadian Journal of Economics*, 48(2), 773–792.

Ozaki, H., and P. A. Streufert (1996): "Dynamic Programming for Non-additive Stochastic Objectives," *Journal of Mathematical Economics*, 25(4), 391–442.

Quah, J. K.-H. (2014): "A Test for Weakly Separable Preferences," *Oxford University Discussion Paper*.

Schwartz, M. (1990): "Third-Degree Price Discrimination and Output: Generalizing a Welfare Result," *American Economic Review*, 80(5), 1259–1262.

Thaler, R. H. (1999): "Mental Accounting Matters," *Journal of Behavioral Decision Making*, 12(3), 183–206.

Varian, H. R. (1982): "The Nonparametric Approach to Demand Analysis," *Econometrica*, 50(4), 945–973.

——— (1985): "Price Discrimination and Social Welfare," *American Economic Review*, 75(4), 870–875.

Varian, H. R. (1988): "Revealed Preference with a Subset of Goods," *Journal of Economic Theory*, 46(1), 179–185.

Ziegler, G. M. (1995): *Lectures on polytopes*, Graduate texts in mathematics. Springer, New York.

Appendix A. Omitted Proofs from Section 2

Proof of Proposition 1. Our proof of part (1) relies on a result of Varian (1988), which says that given any data set $\mathcal{D}$, there always exist $\{z^t\}_{t=1}^T$ such that the augemented data set $\{((p^t,1),(x^t,z^t))\}_{t=1}^T$ obeys GARP. By Afriat's Theorem, there is a utility function $\widetilde{U}:\mathbb{R}^{L+1}\to\mathbb{R}$ such that (x^t,z^t) is optimal in the budget set $\{(x,z)\in\mathbb{R}^L\times\mathbb{R} : p^t x+z\leq M^t\}$, where $M^t = z^t+p^t x^t$. Afriat's Theorem also guarantees that this utility function has various nice properties and, in particular, $\widetilde{U}$ can be chosen to be strictly increasing. Let $\bar{M}=\sum_{t=1}^T M^t/T$ and define the augmented utility function $U:\mathbb{R}_+^L\times\mathbb{R}_-\to\mathbb{R}$ by $U(x,-e):=\widetilde{U}(x,\bar{M}-e)$. Then since $\widetilde{U}$ is strictly increasing, x^t solves $\max_{x\in\mathbb{R}_+^L} U(x,\delta^t - p^t x)$ where $\delta^t = M^t-\bar{M}$.

There are two claims in (2). We first consider the case where $\mathcal{D}$ obeys GAPP. Note that whenever $p^{t'}x^{t'}-p^t x^{t'}<0$, then for any $\{q^t\}_{t=1}^T$ and $\{\delta^t\}_{t=1}^T$ such that (9) holds, we obtain $p^{t'}x^{t'}-p^t x^{t'}<\delta^{t'}-\epsilon^{t'}x^{t'}-\delta^t+\epsilon^t x^{t'}$. This inequality can be re-written as

$$-q^t x^{t'}+\delta^t < -q^{t'}x^{t'}+\delta^{t'}. \tag{18}$$

Choose a number $M>\max_t(q^t x^t-\delta^t)$ and define the data set $\widetilde{\mathcal{D}}=\{(q^t,1),(x^t,M+\delta^t - q^t x^t)\}_{t=1}^T$. Since (18) holds whenever $p^{t'}x^{t'}-p^t x^{t'}<0$,

$$(q^t,1)(x^t,M+\delta^t-q^t x^t) > (q^t,1)(x^{t'},M+\delta^{t'}-q^{t'}x^{t'}) \text{ only if } p^{t'}x^{t'}>p^t x^{t'}.$$

This guarantees that $\widetilde{\mathcal{D}}$ obeys GARP since $\mathcal{D}$ obeys GAPP. By Afriat's Theorem, there is a strictly increasing utility function $\widetilde{U}:\mathbb{R}^L\to\mathbb{R}$ such that $(x^t,M+\delta^t-q^t x^t)$ is optimal in the budget set $\{(x,z)\in\mathbb{R}^{L+1}: q^t x+z\leq M+\delta^t\}$. Define the augmented utility function $U:\mathbb{R}_+^L\times\mathbb{R}_-\to\mathbb{R}$ by $U(x,-e):=\widetilde{U}(x,M-e)$. Since $\widetilde{U}$ is strictly increasing, x^t solves $\max_{x\in\mathbb{R}_+^L} U(x,\delta^t-q^t x)$.

Now consider the case where $\mathcal{D}$ violates GAPP. Observe that whenever $p^{t'}x^{t'}-p^t x^{t'}>0$, then for any $\{q^t\}_{t=1}^T$ and $\{\delta^t\}_{t=1}^T$ such that (9) holds, we obtain $p^{t'}x^{t'}-p^t x^{t'}>\delta^{t'}-\epsilon^{t'}x^{t'}-\delta^t+\epsilon^t x^{t'}$. This inequality can be re-written as

$$-q^t x^{t'}+\delta^t > -q^{t'}x^{t'}+\delta^{t'}. \tag{19}$$

By way of contradiction, suppose there is an augmented utility function U such that x^t maximizes $U(x,-q^t x+\delta^t)$ for all t. For any observation t, we write $V^t:=\max_{x\in X} U(x,-q^t x+\delta^t)$. Since (19) holds, we obtain

$$V^t\geq U(x^{t'},-q^t x^{t'}+\delta^t) > U(x^{t'},-q^{t'}x^{t'}+\delta^{t'}) = V^{t'}.$$

Thus, we have shown that $V^t>V^{t'}$ whenever $p^{t'}x^{t'}-p^t x^{t'}>0$. Since $\mathcal{D}$ violates GAPP there is a finite sequence $\{(p^{t_1},x^{t_1}),\ldots,(p^{t_N},x^{t_N})\}$ of distinct elements in $\mathcal{D}$, such that $p^{t_i}x^{t_{i+1}}<p^{t_{i+1}}x^{t_{i+1}}$ for all $i\in\{1,\ldots,N-1\}$ and $p^{t_N}x^{t_1}<p^{t_1}x^{t_1}$. By the observation we have just made, we obtain $V^{t_1}>V^{t_2}>...V^{t_N}>V^{t_1}$, which is impossible. ■

PROOF OF PROPOSITION 3.

(1) We have already shown the 'only if' part of this claim, so we need to show the 'if' part holds. From the proof of Theorem 1, we know that for a large M, it is the case that $p^t \succeq_p p^{t'}$ if and only if $(x^t, M - p^t x^t) \succeq_x (x^{t'}, M - p^{t'} x^{t'})$ and hence $p^t \succeq_p^* p^{t'}$ if and only if $(x^t, M - p^t x^t) \succeq_x^* (x^{t'}, M - p^{t'} x^{t'})$. If $p^t \not\succeq_p^* p^{t'}$, then $(x^t, M - p^t x^t) \not\succeq_x^* (x^{t'}, M - p^{t'} x^{t'})$ and hence there is a utility function $\widetilde{U} : \mathbb{R}_+^{L+1} \to \mathbb{R}$ rationalizing the augmented data set $\widetilde{D}$ such that $\widetilde{U}(x^t, M - p^t x^t) < \widetilde{U}(x^{t'}, M - p^{t'} x^{t'})$ (see Remark 2 in Section 2.1). This in turn implies that the augmented utility function U (as defined by (6)), has the property that $U(x^t, -p^t x^t) < U(x^{t'}, -p^{t'} x^{t'})$ or, equivalently, $V(p^t) < V(p^{t'})$.

(2) Given part (1), we need only show that if $p^t \succeq_p^* p^{t'}$ but $p^t \not\succ_p^* p^{t'}$, then there is some augmented utility function U such that $U(x^t, -p^t x^t) = U(x^{t'}, -p^{t'} x^{t'})$. To see that this holds, note that if $p^t \succeq_p^* p^{t'}$ but $p^t \not\succ_p^* p^{t'}$, then $(x^t, M - p^t x^t) \succeq_x^* (x^{t'}, M - p^{t'} x^{t'})$ but $(x^t, M - p^t x^t) \not\succ_x^* (x^{t'}, M - p^{t'} x^{t'})$. In this case there is a utility function $\widetilde{U} : \mathbb{R}_+^{L+1} \to \mathbb{R}$ rationalizing the augmented data set $\widetilde{D}$ such that $\widetilde{U}(x^t, M - p^t x^t) = \widetilde{U}(x^{t'}, M - p^{t'} x^{t'})$. This in turn implies that the augmented utility function U (as defined by (6)) satisfies $U(x^t, -p^t x^t) = U(x^{t'}, -p^{t'} x^{t'})$ and so $V(p^t) = V(p^{t'})$. ■

PROOF OF THEOREM 2. We prove this result for the more general case where the consumption space is a closed subset X of $\mathbb{R}_+^L$. It is obvious that (3) implies (1). We first show that (1) implies (2), i.e., nonlinear GAPP is necessary for rationalization. Given a price system $\psi : X \to \mathbb{R}$, we define the indirect utility $V(\psi) = \max_{x \in X} U(x, -\psi(x))$. If $\psi^{t'} \succeq_p \psi^t$, then $\psi^{t'}(x^t) \leq \psi^t(x^t)$, and so

$$V(\psi^{t'}) \geq U(x^t, -\psi^{t'}(x^t)) \geq U(x^t, -\psi^t(x^t)) = V(\psi^t).$$

Furthermore, $U(x^t, -\psi^{t'}(x^t)) > U(x^t, -\psi^t(x^t))$ if $\psi^{t'} \succ_p \psi^t$, and in that case $V(\psi^{t'}) > V(\psi^t)$. It follows that if $p^{t'} \succeq_p^* p^t$, then $V(\psi^{t'}) \geq V(\psi^t)$, which mean that it is not possible for $p^t \succ_p^* p^{t'}$.

It remains for us to show that (2) implies (3). Choose a number $M > \max_{t \in T} \psi^t(x^t)$. We associate to each observation t the constraint set

$$C^t = \{(x, z) \in X \times \mathbb{R}_+ : \psi^t(x) + z \leq M\}$$

and define the augmented data set $\widetilde{\mathcal{D}} = \{(C^t, (x^t, z^t)\}$, where $z^t = M - \psi^t(x^t)$. Our assumptions on ψ^t guarantee that C^t is a compact set and that it is downward comprehensive, i.e., if $(x, z) \in C^t$ then so is $(x', z') \in X \times \mathbb{R}_+$ such that $(x', z') \leq (x, z)$.

Notice that $(x^t, z^t) \in C^{t'}$ if and only if $\psi^{t'} \succeq_p \psi^t$. This is because $(x^t, z^t) \in C^{t'}$ if and only if

$$\psi^{t'}(x^t) + z^t \leq M = \psi^t(x^t) + z^t.$$

When this occurs, we write $(x^{t'}, z^{t'}) \succeq_x (x^t, z^t)$. Since ψ^t is increasing, if $(\hat{x}, \hat{z}) > (x, z)$, then $\psi(\hat{x}) + \hat{z} > \psi(x) + z$. This guarantees that $(x^t, z^t) \in \underline{C}^{t'}$ if and only if $\psi^{t'} \succ_p \psi^t$, where $\underline{C}^{t'}$ consists of those elements in $C^{t'}$ which are dominated by some element in $C^{t'}$, i.e., $(x, z) \in \underline{C}^{t'}$ if it is in $C^{t'}$ and there is $(\hat{x}, \hat{z}) \in C^{t'}$ such that $(\hat{x}, \hat{z}) > (x, z)$; in this case, we write $(x^{t'}, z^{t'}) \succ_x (x^t, z^t)$. Given $\succeq_x$ and $\succ_x$, we can then define GARP in the usual way. It is clear that $\widetilde{\mathcal{D}}$ satisfies GARP if and only $\mathcal{D}$ satisfies nonlinear GAPP.

Since $\mathcal{D}$ obeys nonlinear GAPP, $\widetilde{\mathcal{D}}$ obeys GARP and by Theorem 2 in Nishimura, Ok, and Quah (2017), there is a continuous and strictly increasing function $\widetilde{U} : X \times \mathbb{R}_+ \to \mathbb{R}$ for which (x^t, z^t) maximizes $\widetilde{U}$ in C^t, for all $t \in T$. Furthermore, $\widetilde{U}$ can be chosen to have the following property: if $(x, z) \in C^t$ and $\widetilde{U}(x, z) = \widetilde{U}(x^t, z^t)$ then, for some $s \in T$, $(x, z) = (x^s, z^s)$ and $(x^s, z^s) \succeq_x^* (x^t, z^t)$. In other words, any bundle (x, z) in C^t has strictly lower utility than (x^t, z^t), unless it is revealed preferred to (x^t, z^t) (in which case the two bundles will have the same utility).

Note that we can always choose $\widetilde{U}$ to be bounded above on $X \times \mathbb{R}_+$. Indeed, if $\widetilde{U}$ is not bounded, we can define another utility function $\widehat{U}$ on $X \times \mathbb{R}_+$ by

$$\widehat{U}(x, z) = \min\{\widetilde{U}(x, z), K + h(x, z)\},$$

where $K > \max_{t \in T}\{\widetilde{U}(x^t, z^t)\}$ and h is a positive, continuous, strictly increasing, and bounded real-valued function on $X \times \mathbb{R}_+$. Then $\widehat{U}$ is continuous, strictly increasing, and bounded above. Furthermore, for all $(x, z) \in C^t$, $\widetilde{U}(x, z) < K$ and so $\hat{U}(x, z) = \tilde{U}(x, z)$. Therefore, (x^t, z^t) maximizes $\widehat{U}(x, z)$ in C^t and $\widehat{U}(x^t, z^t)$ will be strictly greater $\widehat{U}(x, z)$ for any $(x, z) \in C^t$ that is not revealed preferred to (x^t, z^t).

We define the expenditure-augmented utility function

$$U(x, -e) = \widetilde{U}(x, \max\{M - e, 0\}) - B \min\{e - M, 0\},$$

where $B > 0$. This function is continuous, strictly increasing, and bounded above. We claim that for B sufficiently large, $x^t \in \arg\max_{x \in X} U(x, -\psi^t(x))$. Clearly, if x satisfies $M - \psi^t(x) \geq 0$, in other words $(x, M - \psi^t(x)) \in C^t$, then

$$U(x, -\psi(x)) = \widetilde{U}(x, M - \psi(x)) \leq \widetilde{U}(x^t, M - \psi(x^t)) = U(x^t, -\psi^t(x^t))$$

for any $B > 0$. So if there is no B obeying the specified condition, there must be a sequence $B_n \to \infty$, and x_n with $M - \psi^t(x) < 0$ such that, for all n,

$$U(x_n, -\psi^t(x_n)) = \widetilde{U}(x_n, 0) - B_n(\psi^t(x_n) - M) > U(x^t, -\psi^t(x^t)) = \widetilde{U}(x^t, z^t). \tag{20}$$

Since $\widetilde{U}$ is bounded above and $B_n \to \infty$, (20) implies that $\psi^t(x_n) \to M$. By our assumption on ψ^t, x_n is contained in a bounded set; by taking subsequences if necessary we can assume that x_n has a limit $\underline{x}$, with $\psi^t(\underline{x}) = M$. Then, it follows from (20) again that $\widetilde{U}(\underline{x}, 0) \geq \widetilde{U}(x^t, z^t)$. But this is impossible because since $(\underline{x}, 0) \neq (x^s, z^s)$

for any $s \in T$ and thus $\widetilde{U}(\underline{x}, 0) < \widetilde{U}(x^t, z^t)$. We conclude that for B sufficiently large $x^t \in \arg\max_{x \in X} U(x, -\psi^t(x))$. Since the number of observations is finite, we could find a B for which this holds for all $t \in T$ (in other words, U rationalizes $\mathcal{D}$).

Lastly, we need to show that $\max_{x \in X} U(x, -\psi(x))$ has a solution for any price system ψ. Since U is bounded above, $u^* = \sup_{x \in X} U(x, -\psi(x))$ exists and there is a sequence x_n such that $U(x_n, -\psi(x_n)) \to u^*$. If x_n is unbounded, $\psi(x_n) \to \infty$, which (given the form of U) means that $U(x_n, -\psi(x_n)) \to -\infty$. This is impossible and so x_n is bounded and will have a subsequence converging to some $x^* \in X$. By the continuity of U and ψ, we obtain $U(x^*, -\psi(x^*)) = u^*$. ■

APPENDIX B. OMITTED MATERIAL FROM SECTION 4

In this section, we formally develop our bootstrap procedure from Section 4.2. We begin by describing Weyl-Minkowski duality[25] which is used for the equivalent (dual) restatement (17) of our test (16). As we mentioned earlier, we will also appeal to this duality in the proof of the asymptotic validity of our testing procedure.

Theorem 5. *(Weyl-Minkowski Theorem for Cones) A subset $\mathcal{C}$ of $\mathbb{R}^I$ is a finitely generated cone*

$$\mathcal{C} = \{\nu_1 a_1 + \ldots + \nu_{|A|} a_{|A|} : \nu_h \geq 0\} \text{ for some } A = [a_1, \ldots, a_H] \in \mathbb{R}^{I \times |A|} \tag{21}$$

if, and only if, it is a finite intersection of closed half spaces

$$\mathcal{C} = \{t \in \mathbb{R}^I | Bt \leq 0\} \text{ for some } B \in \mathbb{R}^{m \times I}. \tag{22}$$

The expressions in (21) and (22) are called a $\mathcal{V}$-representation (as in "vertices") and a $\mathcal{H}$-representation (as in "half spaces") of $\mathcal{C}$, respectively. In what follows, we use an $\mathcal{H}$-representation of cone(A) corresponding to a $m \times I$ matrix B as implied by Theorem 5.

We are now in a position to show that the bootstrap procedure defined in Section 4.2 is asymptotically valid. Note first that $\Theta = [\underline{\theta}, \bar{\theta}]$, where

$$\bar{\theta} = \max_{\nu \in \Delta^{|A|-1}} \rho\nu = \max_{1 \leq j \leq |A|} \rho_j \tag{23}$$

$$\underline{\theta} = \min_{\nu \in \Delta^{|A|-1}} \rho\nu = \min_{1 \leq j \leq |A|} \rho_j, \tag{24}$$

where ρ_j denotes the jth component of ρ. We normalize (ρ, θ) such that $\Theta = [\underline{\theta}, \underline{\theta} + 1]$. Next, define

$$\mathcal{H} := \{1, 2, \ldots, |A|\} \tag{25}$$

$$\overline{\mathcal{H}} := \{j \in \mathcal{H} \mid \rho_j = \bar{\theta}\} \tag{26}$$

$$\underline{\mathcal{H}} := \{j \in \mathcal{H} \mid \rho_j = \underline{\theta}\} \tag{27}$$

[25]See, for example, Theorem 1.3 in Ziegler (1995).

$$\mathcal{H}_0 \quad := \quad \mathcal{H} \setminus (\overline{\mathcal{H}} \cup \underline{\mathcal{H}}). \tag{28}$$

Recall that τ_N is a tuning parameter chosen such that $\tau_N \downarrow 0$ and $\sqrt{N}\tau_N \uparrow \infty$. For $\theta \in \Theta_I$, we now formally define the τ_N-tightened version of $\mathcal{S}$ as

$$\mathcal{S}_{\tau_N}(\theta) := \{A\nu \mid \rho\nu = \theta, \nu \in \mathcal{V}_{\tau_N}(\theta)\},$$

where

$$\mathcal{V}_{\tau_N}(\theta) := \left\{ \nu \in \Delta^{|A|-1} \;\middle|\; \begin{array}{l} \nu_j \geq \frac{(\bar{\theta}-\theta)\tau_N}{|\underline{\mathcal{H}} \cup \mathcal{H}_0|}, j \in \underline{\mathcal{H}},\; \nu_{j'} \geq \frac{(\theta-\underline{\theta})\tau_N}{|\overline{\mathcal{H}} \cup \mathcal{H}_0|},\; j' \in \overline{\mathcal{H}}, \\ \\ \nu_{j''} \geq \left[1 - \frac{(\bar{\theta}-\theta)|\underline{\mathcal{H}}|}{|\underline{\mathcal{H}} \cup \mathcal{H}_0|} - \frac{(\theta-\underline{\theta})|\overline{\mathcal{H}}|}{|\overline{\mathcal{H}} \cup \mathcal{H}_0|}\right] \frac{\tau_N}{|\mathcal{H}_0|},\; j'' \in \mathcal{H}_0 \end{array} \right\}.$$

In applications where ρ is binary, the above notation simplifies. Specifically, in our empirical application on deriving the welfare bounds, $\rho = \mathbb{1}_{t \succeq_p^* t'}$ and $\theta = \mathcal{N}_{t \succeq_p^* t'}$. Here, $\bar{\theta} = 1$, $\underline{\theta} = 0$, and $\bar{\theta} - \underline{\theta} = 1$ holds without any normalization. Also, $\overline{\mathcal{H}}$ ($\underline{\mathcal{H}}$) is just the set of indices for the types that (do not) prefer price p^t compared to $p^{t'}$, while $\mathcal{H}_0$ is empty. We then have:

$$\mathcal{S}_{\tau_N}(\mathcal{N}_{t \succeq_p^* t'}) = \left\{ A\nu \;\middle|\; \mathbb{1}'_{t \succeq_p^* t'}\nu = \mathcal{N}_{t \succeq_p^* t'}, \nu \in \mathcal{V}_{\tau_N}(\mathcal{N}_{t \succeq_p^* t'}) \right\},$$

where

$$\mathcal{V}_{\tau_N}(\mathcal{N}_{t \succeq_p^* t'}) = \left\{ \nu \in \Delta^{|A|-1} \;\middle|\; \nu_j \geq \frac{(1 - \mathcal{N}_{t \succeq_p^* t'})\tau_N}{|\underline{\mathcal{H}}|}, j \in \underline{\mathcal{H}},\; \nu_{j'} \geq \frac{\mathcal{N}_{t \succeq_p^* t'}\tau_N}{|\overline{\mathcal{H}}|},\; j' \in \overline{\mathcal{H}} \right\}.$$

We now state the mild data assumptions.

Assumption 2. For all $t = 1, ..., T$, $\frac{N_t}{N} \to \kappa_t$ as $N \to \infty$, where $\kappa_t > 0$, $1 \leq t \leq T$.

Assumption 3. The econometrican observes T independent cross-sections of i.i.d. samples $\left\{x^t_{n(t)}\right\}_{n(t)=1}^{N_t}$, $t = 1, ..., T$ of consumers' choices corresponding to the known price vectors $\{p_t\}_{t=1}^T$.

Next, let $\mathbf{d}^{i,t}_{n(t)} := \mathbf{1}\{x^t_{n(t)} \in B^{i,t}\}$, $\mathbf{d}^t_{n(t)} = [\mathbf{d}^{1,t}_{n(t)}, ..., \mathbf{d}^{I_t,t}_{n(t)}]$, and $\mathbf{d}^t_n = [\mathbf{d}^{1,t}_n, ..., \mathbf{d}^{I_t,t}_n]$. Let $\mathbf{d}_t$ denote the choice vector of a consumer facing price p^t (we can, for example, let $\mathbf{d}_t = \mathbf{d}^t_1$). Define $\mathbf{d} = [\mathbf{d}'_1, ..., \mathbf{d}'_T]'$: note, $\mathrm{E}[\mathbf{d}] = \pi$ holds by definition. Among the rows of B some of them correspond to constraints that hold trivially by definition, whereas some are for non-trivial constraints. Let $\mathcal{K}^R$ be the index set for the latter. Finally, let

$$\begin{aligned} g &= B\mathbf{d} \\ &= [g_1, ..., g_m]'. \end{aligned}$$

With these definitions, consider the following requirement:

Condition 1. For each $k \in \mathcal{K}^R$, $\mathrm{var}(g_k) > 0$ and $\mathrm{E}[|g_k/\sqrt{\mathrm{var}(g_k)}|^{2+c_1}] < c_2$ hold, where c_1 and c_2 are positive constants.

This guarantees the Lyapunov condition for the triangular array CLT used in establishing asymptotic uniform validity. This type of condition has been used widely in the literature of moment inequalities; see Andrews and Soares (2010).

PROOF OF THEOREM 4.

Define
$$\mathcal{C} = \mathrm{cone}(A)$$
and
$$\mathcal{T}(\theta) = \{\pi = A\nu : \rho'\nu = \theta, \nu \in \mathbb{R}^{|A|}\},$$
an affine subspace of $\mathbb{R}^I$. It is convenient to rewrite $\mathcal{T}(\theta)$ as $\mathcal{T}(\theta) = \{t \in \mathbb{R}^I : \tilde{B}t = d(\theta)\}$ where $\tilde{B} \in \tilde{m} \times \mathbb{R}^I$, $d(\cdot) \in \tilde{m} \times 1$, and $\tilde{m}$ all depend on (ρ, A). We let $\tilde{b}_j$ denote the j-th row of $\tilde{B}$. Then
$$\mathcal{S}(\theta) = \mathcal{C} \cap \Delta^{|A|-1} \cap \mathcal{T}(\theta).$$
By Theorem 5, $\mathcal{C} = \{\pi : B\pi \leq 0\}$, therefore
$$\mathcal{S}(\theta) \quad = \quad \{t \in \mathbb{R}^{|A|} : Bt \leq 0, \tilde{B}t = d(\theta), \mathbf{1}_H' t = 1\}. \tag{29}$$

Let
$$\psi(\theta) = [\psi_1(\theta), ..., \psi_H(\theta)]' \quad \theta \in \Theta$$
with
$$\psi_j(\theta) = \begin{cases} \frac{(\bar{\theta}-\theta)}{|\underline{\mathcal{H}} \cup \mathcal{H}_0|} & \text{if } j \in \underline{\mathcal{H}} \\ \frac{(\theta-\bar{\theta})}{|\overline{\mathcal{H}} \cup \mathcal{H}_0|} & \text{if } j \in \overline{\mathcal{H}} \\ \left[1 - \frac{(\bar{\theta}-\theta)|\underline{\mathcal{H}}|}{|\underline{\mathcal{H}} \cup \mathcal{H}_0|} - \frac{(\theta-\underline{\theta})|\overline{\mathcal{H}}|}{|\overline{\mathcal{H}} \cup \mathcal{H}_0|}\right] \frac{1}{|\mathcal{H}_0|} & \text{if } j \in \mathcal{H}_0 \end{cases},$$
where terms are defined in (25)-(28). Then
$$\mathcal{S}_{\tau_N}(\theta) = \{\pi = A\nu : \nu \geq \tau_N \psi(\theta), \nu \in \Delta^{|A|-1}, \rho'\nu = \theta\}.$$
Finally, let
$$\mathcal{C}_{\tau_N} = \{\pi = A\nu : \nu \geq \tau_N \psi(\theta)\}.$$
Then
$$\mathcal{S}_{\tau_N}(\theta) = \mathcal{C}_{\tau_N} \cap \Delta^{|A|-1} \cap \mathcal{T}(\theta).$$
Proceeding as in the proof of Lemma 4.1 in KS, we can express the set $\mathcal{C}_{\tau_N}$ as
$$\mathcal{C}_{\tau_N} = \{t : Bt \leq -\tau_N \phi(\theta)\}$$
where
$$\phi(\theta) = -BA\psi(\theta).$$

As in Lemma 4.1 in KS, let the first $\bar{m}$ rows of B represent inequality constraints and the rest equalities, and also let Φ_{kh} the (k,h)-element of the matrix $-BA$. We have

$$\phi_k = \sum_{h=1}^{|A|} \Phi_{kh}\psi_h(\theta)$$

where, for each $k \leq \bar{m}$, $\{\Phi_{kh}\}_{h=1}^{|A|}$ are all nonnegative, with at least some of them being strictly positive, and $\Phi_{kh} = 0$ for all h if $\bar{m} < k \leq m$. Since $\psi_h(\theta) > 0, 1 \leq h \leq |A|$ for every $\theta \in \Theta$ by definition, we have $\phi_j(\theta) \geq C, 1 \leq j \leq \bar{m}$ for some positive constant C, and $\phi_j(\theta) = 0, \bar{m} < j \leq m$ for every $\theta \in \Theta$. Putting these together, we have

$$\mathcal{S}_{\tau_N}(\theta) \quad = \quad \{t \in \mathbb{R}^{|A|} : Bt \leq -\tau_N \phi(\theta), \tilde{B}t = d(\theta), \mathbf{1}_H' t = 1\}$$

where $\mathbf{1}_H$ denotes the $|A|$-vector of ones. Define the $\mathbb{R}^I$-valued random vector

$$\pi^*_{\tau_N} := \frac{1}{\sqrt{N}}\zeta + \hat{\eta}_{\tau_N},\ \zeta \sim \mathrm{N}(0, \hat{S})$$

where $\hat{S}$ is a consistent estimator for the asymptotic covariance matrix of $\sqrt{N}(\hat{\pi} - \pi)$. Then (conditional on the data) the distribution of

$$\delta^*(\theta) := N \min_{\eta \in \mathcal{S}_{\tau_N}(\theta)} [\pi^*_{\tau_N} - \eta]' \Omega [\pi^*_{\tau_N} - \eta]$$

corresponds to that of the bootstrap test statistics. Let

$$B_* := \begin{bmatrix} B \\ \tilde{B} \\ \mathbf{1}_H' \end{bmatrix}$$

Define $\ell = \mathrm{rank}(B_*)$ for the augmented matrix B_* instead of B in KS, and let the $\ell \times m$-matrix K be such that KB_* is a matrix whose rows consist of a basis of the row space row (B_*). Also let M be an $(I - \ell) \times I$ matrix whose rows form an orthonormal basis of $\ker B_* = \ker(KB_*)$, and define $P = \binom{KB_*}{M}$. Finally, let $\hat{g} = B_* \hat{\pi}$.

Define

$$\begin{aligned} T(x,y) &:= \begin{pmatrix} x \\ y \end{pmatrix}' P^{-1\prime} \Omega P^{-1} \begin{pmatrix} x \\ y \end{pmatrix}, \quad x \in \mathbb{R}^\ell, y \in \mathbb{R}^{I-\ell} \\ t(x) &:= \min_{y \in \mathbb{R}^{I-\ell}} T(x,y) \\ s(g) &:= \min_{\gamma = [\gamma^{\leq\prime}, \gamma^{=\prime}]', \gamma^{\leq} \leq 0, \gamma' \in \mathrm{col}(B)} t(K[g - \gamma]) \end{aligned}$$

with

$$\gamma^{=} = \begin{bmatrix} \mathbf{0}_{m-\bar{m}} \\ d(\theta) \\ 1 \end{bmatrix}$$

where $\mathbf{0}_{m-\bar{m}}$ denotes the $(m-\bar{m})$-vector of zeros. It is easy to see that $t : \mathbb{R}^{\ell} \to \mathbb{R}_+$ is a positive definite quadratic form. By (29), we can write

$$\delta_N(\theta) = Ns(\hat{g}) = s(\sqrt{N}\hat{g}).$$

Likewise, for the bootstrapped version of δ we have

$$\begin{aligned} \delta^*(\theta) &= N \min_{\eta \in \mathcal{S}_{\tau_N}(\theta)} [\pi^*_{\tau_N} - \eta]' \Omega [\pi^*_{\tau_N} - \eta] \\ &= s(\varphi_N(\hat{\xi}) + \zeta), \end{aligned}$$

where $\hat{\xi} = B_* \hat{\pi}/\tau_N$. Note the function $\varphi_N(\xi) = [\varphi_N^1(\xi), ..., \varphi_N^m(\xi)]$ for $\xi = (\xi_1, ..., \xi_m)' \in \mathrm{col}(B_*)$. Moreover, its k-th element φ_N^k for $k \leq \bar{m}$ satisfies

$$\varphi_N^k(\xi) = 0$$

if $|\xi^k| \leq \delta$ and $\xi^j \leq \delta, 1 \leq j \leq m$, $\delta > 0$, for large enough N and $\varphi_N^k(\xi) = 0$ for $k > \bar{m}$. This follows (we use some notation in the proof of Theorem 4.2 in KS, which the reader is referred to) by first noting that it suffices to show that for small enough $\delta > 0$, every element x^* that fulfills equation (9.2) in KS with its RHS intersected with $\cap_{j=1}^{\bar{m}} \tilde{S}_j(\delta), \tilde{S}_j(\delta) = \{x : |\tilde{b}_j' x - d_j(\theta)| \leq \tau\delta\}$ satisfies

$$x^* | \mathcal{S}(\theta) \in \cap_{j=1}^{q} H_j^{\tau} \cap L \cap \mathcal{T}(\theta).$$

If not, then there exists $(\tilde{a}, \tilde{x}) \in F \cap \mathcal{T}(\theta) \times \cap_{j=1}^{q} H_j \cap L \cap \mathcal{T}(\theta)$ such that

$$(\tilde{a} - \tilde{x})'(\tilde{x}|\mathcal{S}_\tau(\theta) - \tilde{x}) = 0,$$

where $\tilde{x}|\mathcal{S}_\tau(\theta)$ denotes the orthogonal projection of $\tilde{x}$ on $\mathcal{S}_\tau(\theta)$. This, in turn, implies that there exists a triplet $(a_0, a_1, a_2) \in \mathcal{A} \times \mathcal{A} \times \mathcal{A}$ such that $(a_1 - a_0)'(a_2 - a_0) < 0$. But as shown in the proof of Theorem 4.2 in KS, this cannot happen. The conclusion then follows by Theorem 1 of Andrews and Soares (2010). ■