Empirical Revealed Preference

Ian Crawford[1,2] and Bram De Rock[3]

[1]Institute for Fiscal Studies, London WC1E 7AE, United Kingdom

[2]Department of Economics, University of Oxford, Oxford OX1 3UQ, United Kingdom; email: ian.crawford@economics.ox.ac.uk

[3]European Center for Advanced Research in Economics and Statistics, Solvay Brussels School of Economics and Management, and Université Libre de Bruxelles, 1050 Brussels, Belgium; email: bderock@ulb.ac.be

Annu. Rev. Econ. 2014. 6:503–24

First published online as a Review in Advance on March 19, 2014

The *Annual Review of Economics* is online at economics.annualreviews.org

This article's doi:
10.1146/annurev-economics-080213-041238

JEL codes: D11, D12, D13, C14

Keywords

consumption behavior, revealed preferences, nonparametric analysis

Abstract

This article aims to provide an introduction to empirical revealed preference (RP) and an overview of the current state of the field. We hope to give a sense of how RP methods work and the types of questions they can address and to assess the strengths and drawbacks of the approach. After briefly recapping the basics of RP theory, we review and critically assess the literature in two main areas representing the principal fields in which recent research has made significant advances: broadening the scope of RP methods and dealing with empirical issues related to bringing RP to the data. We conclude with a discussion of some future directions.

1. INTRODUCTION

Empirical revealed preference (RP) is a structural approach to the analysis and interpretation of data by means of economic theory. Despite sharing the theory-driven focus, it is somewhat distinct from traditional structural econometrics, and this distinction is a useful starting point for discussing empirical RP. The structural approach to econometrics is very familiar: It proceeds by using economic theory to develop formal mathematical statements concerning causes and effects. The causes (explanatory variables), which may be observed (x) or unobserved (η), and the effects (endogenous variables, y) are linked by these theory-derived statements through structural equations, $y = f(x, \eta, \theta)$, where θ represents a set of unknown parameters or functions. Econometricians always then append a statistical structure to the economic model to account for the fact that the economic theory as expressed through the structural equations f does not perfectly explain the data. This extra structure entails statistical assumptions regarding the joint distribution of (x, η) and other unobservables (ε) introduced by the econometrician. When combined, these economic and statistical assumptions deliver an empirical model that is capable of rationalizing any set of observables. The art of structural modeling thus mainly lies in getting this statistical aspect right, because the source and the properties of these econometric errors (ε) can have a critical impact on the estimation results. Unfortunately, this is far from straightforward, as economic theories, which are by and large completely deterministic, generally have little to say about the statistical model, and the data generally have little to say about unobservables.

Similar to structural econometrics, empirical RP begins from economic theory, but the description of the theory's implications is entirely different from the $y = f(x, \eta, \theta)$ type of framework. Rather than describing the theory's implications in terms of parameterized structural equations, empirical RP uses systems of inequalities that depend neither on the form of structural functions nor on unobservables. Statistical error terms and assumptions about the functional structure of the economic model may be added if they are believed to be part of behavior (see Section 4), but it is not an essential requirement of the approach. In a sense, empirical RP is concerned with what one can learn simply by combining theory with the features of the world that can be observed.

The aim of this article is to provide an introduction to empirical RP and an overview of the current state of the field. We hope to give a sense of how empirical RP methods work and the types of questions they can address and to assess the strengths and drawbacks of the approach. We begin by briefly recapping the basics of RP theory—namely, Afriat's theorem and how it can be used to check data for consistency with the canonical utility maximization model and, granted this, to make predictions and allow the recovery of features of the model. We then review and critically assess the literature in two main areas representing the principal fields in which current and recent research has significantly advanced. These fields relate to efforts that have broadened the scope of RP methods to allow the exploration of a richer variety of economic models and work that has sought to address some empirical and statistical challenges involved in applying RP methods to the data. As such, we demonstrate the versatility and attractiveness of empirical RPs, as well as areas in which further work is clearly needed. We conclude with a discussion of some future directions.

2. THE BASIC MODEL OF RATIONAL DEMAND

This section sets the stage for our discussion below by considering the basic RP tools for the most simple case of utility maximization. Section 2.1 presents the most fundamental result on this topic, Afriat's theorem, and shows how it can be used to check whether a given data set with observed consumption choices and prices is consistent with utility maximization. Subsequently, Section 2.2 focuses on recovering the underlying preferences and on forecasting behavior in new situations.

2.1. Afriat's Theorem

We consider a setting with N goods and a finite data set $S = \{\mathbf{p}_t, \mathbf{q}_t\}_{t\in T}$ consisting of N-dimensional price vectors $\mathbf{p}_t \in \mathbb{R}^N_{++}$ and N-dimensional quantity vectors $\mathbf{q}_t \in \mathbb{R}^N_+$. The set $T = \{1, \ldots, |T|\}$ corresponds to the set of observations. A utility function $u: \mathbb{R}^N_+ \rightarrow \mathbb{R}$ is well behaved if it is concave, continuous, and strictly monotone. The following definition is standard.

Definition 1: A data set $\{\mathbf{p}_t, \mathbf{q}_t\}_{t\in T}$ is rationalizable by a well-behaved utility function u if for all $t \in T$,

$$\mathbf{q}_t \in \arg\max\ u(\mathbf{q}) \quad \text{s.t.} \quad \mathbf{p}_t\mathbf{q} \leq \mathbf{p}_t\mathbf{q}_t.$$

In what follows, we present several ways to verify if a data set S is rationalizable. The first one is the generalized axiom of revealed preference (GARP) introduced by Varian (1982).[1]

Definition 2: A data set $\{\mathbf{p}_t, \mathbf{q}_t\}_{t\in T}$ satisfies GARP if and only if we can construct relations R_0 and R such that (a) for all $t, s \in T$, if $\mathbf{p}_t\mathbf{q}_t \geq \mathbf{p}_t\mathbf{q}_s$, then $\mathbf{q}_t\, R_0\, \mathbf{q}_s$; ($b$) for all $t, s, u, \ldots, r, v \in T$, if $\mathbf{q}_t\, R_0\, \mathbf{q}_s, \mathbf{q}_s\, R_0\, \mathbf{q}_u, \ldots,$ and $\mathbf{q}_r\, R_0\, \mathbf{q}_v$, then $\mathbf{q}_t\, R\, \mathbf{q}_v$; and ($c$) for all $t, s \in T$, if $\mathbf{q}_t\, R\, \mathbf{q}_s$, then $\mathbf{p}_s\mathbf{q}_s \leq \mathbf{p}_s\mathbf{q}_t$.

Condition (a) states that the quantities $\mathbf{q}_t$ are directly revealed preferred over $\mathbf{q}_s$ if $\mathbf{q}_t$ was chosen when $\mathbf{q}_s$ was equally attainable. Next, condition (b) imposes transitivity on the RP relation R. Finally, condition (c) states that if a consumption bundle $\mathbf{q}_t$ is revealed preferred to a consumption bundle $\mathbf{q}_s$, then $\mathbf{q}_s$ cannot be more expensive than $\mathbf{q}_t$.

Cherchye et al. (2011c) demonstrate that satisfying GARP is equivalent to having a solution for the integer programming (IP) problem IP-GARP.

Definition 3: Data $\{\mathbf{p}_t, \mathbf{q}_t\}_{t\in T}$ satisfy IP-GARP if and only if there exist, for all $s, t \in T$ binary variables $x_{s,t} \in \{0, 1\}$ such that (a) for all $t, s \in T$, $\mathbf{p}_t\mathbf{q}_t - \mathbf{p}_t\mathbf{q}_s < x_{t,s}\mathbf{p}_t\mathbf{q}_t$; ($b$) for all $t, s, v \in T$, $x_{t,s} + x_{s,v} \leq x_{t,v}$; and ($c$) for all $t, s \in T$, $(x_{t,s}-1)\mathbf{p}_s\mathbf{q}_s \leq \mathbf{p}_s\mathbf{q}_t - \mathbf{p}_s\mathbf{q}_s$.

When we interpret $x_{t,s} = 1$ as $\mathbf{q}_t\, R_0\, \mathbf{q}_s$, we easily observe the similarity between the conditions in Definitions 2 and 3.

The following theorem extends the well-known theorem introduced by Afriat (1967) and Varian (1982) by adding IP-GARP to it.

Theorem 1: Let $S = \{\mathbf{p}_t, \mathbf{q}_t\}_{t\in T}$ be a set of observations. Then the following statements are equivalent. (a) There exists a nonsatiated utility function that rationalizes S. (b) There exists a well-behaved utility function that rationalizes S. (c) S satisfies GARP. (d) For all $t \in T$, there exist $U_t \in \mathbb{R}_+$ and $\lambda_t \in \mathbb{R}_{++}$ such that for all $t, s \in T$,

$$U_t - U_s \leq \lambda_s(\mathbf{p}_s\mathbf{q}_t - \mathbf{p}_s\mathbf{q}_s).$$

(e) S satisfies IP-GARP.

[1]In the literature, there are several RP axioms. Samuelson (1938, 1948) introduced the weak axiom of revealed preference (WARP). This axiom does not take the transitivity of preferences into account. Therefore, Houthakker (1950) introduced the strong axiom of revealed preference (SARP). SARP does not allow for indifference curves with flat parts, which is taken into account by GARP. The above axioms ignore the differentiability of the underlying utility function. To take this consideration into account, Chiappori & Rochet (1987) introduced strong SARP. In this article, we abstract from all these different RP axioms, and we restrict our attention to GARP.

The equivalence between the first two statements indicates that if the data are rationalizable by any utility function, then they are also rationalizable by a well-behaved utility function. Inter alia, this implies that concavity does not have testable implications. Statements (*c*)–(*e*) present three alternative ways to verify whether the data are rationalizable.

The first method is a combinatorial one and was originally suggested by Varian (1982). The method consists of three steps, which comply with the three conditions in Definition 2 of GARP. The first step constructs the relation R_0 from the data set $S = \{\mathbf{p}_t, \mathbf{q}_t\}_{t \in T}$. In particular, one obtains $\mathbf{q}_t\ R_0\ \mathbf{q}_s$ if and only if $\mathbf{p}_t\mathbf{q}_t \geq \mathbf{p}_t\mathbf{q}_s$. The second step computes the transitive closure of R_0 (i.e., the relation R). Varian (1982) suggests using Warshall's (1962) algorithm, which is an efficient algorithm for computing transitive closures. The third step verifies $\mathbf{p}_t\mathbf{q}_t \leq \mathbf{p}_t\mathbf{q}_s$ whenever $\mathbf{q}_s\ R\ \mathbf{q}_t$. If this is the case, the data set satisfies GARP and is therefore rationalizable.

The second method verifies the rationalizability conditions by testing the feasibility of the corresponding Afriat inequalities. These inequalities are linear in the unknowns U_t and λ_t, which implies that their feasibility can be verified using simple linear programming methods. We refer readers to Afriat (1967) and Diewert (1973) for discussions of this method. An advantage of this method is that it provides not only an efficient way to verify the rationalizability conditions, but also, via the computed values of U_t and λ_t, an estimate for the associated utility levels.

Finally, the third method verifies the rationalizability conditions via the conditions in Definition 3. These conditions are linear in the unknown binary variables $x_{s,t}$. Therefore, feasibility can be verified by standard IP methods (e.g., branch and bound, cutting plane). Compared to the other methods, it is very inefficient and is not recommended for applied work for the basic model developed in this section. However, in contrast to the other two methods, the IP method is quite useful when studying RP characterizations of more complex alternative models (see Section 3).

2.2. Recoverability and Forecasting

Recoverability aims at identifying the underlying preferences of the behavioral model under study. In parametric studies, this is mostly equivalent to uniquely identifying the structural model parameters of the (in)direct utility function (representing the preferences). However, such an exercise is not feasible on the basis of RP theory because there are usually many types of preferences that rationalize data. So the recoverability question we have in mind focuses on identifying the set of preferences (or set of utility functions representing the preferences) that are consistent with a given data set.

The recoverability question basically aims at constructing inner and outer bounds for the indifference curves passing through an arbitrary, not necessarily observed, quantity bundle. This construction is primarily based on restrictions on behavior imposed by GARP. Let us illustrate the approach by means of **Figure 1**; the interested reader is referred to Varian (1982, 2006) for more details.

The figure shows a simple data set with only one observation, $\mathbf{q}_1$, and one unobserved bundle, $\mathbf{q}$, for which we want to do recovery. The relative prices are represented by the slope of the budget line through $\mathbf{q}_1$. As shown by Varian (1982), the set $RP(\mathbf{q})$ represents the set of all bundles that are revealed preferred to $\mathbf{q}$, and the set $RW(\mathbf{q})$ contains all the bundles that are revealed worse to $\mathbf{q}$. These sets are independent of the prices associated with $\mathbf{q}$. As such, the boundaries of these two sets form the inner and outer bounds for all indifference curves passing through $\mathbf{q}$, which are consistent with the observed choices and the preferences revealed by those choices.

It is clear from our example that the inner and outer bounds are not necessarily close to each other. This may serve as an illustration of the critique that an RP approach does not have bite: In this particular case, indifference curves can be very different from each other and still be consistent with observed behavior. However, the inner and outer bounds may be much closer

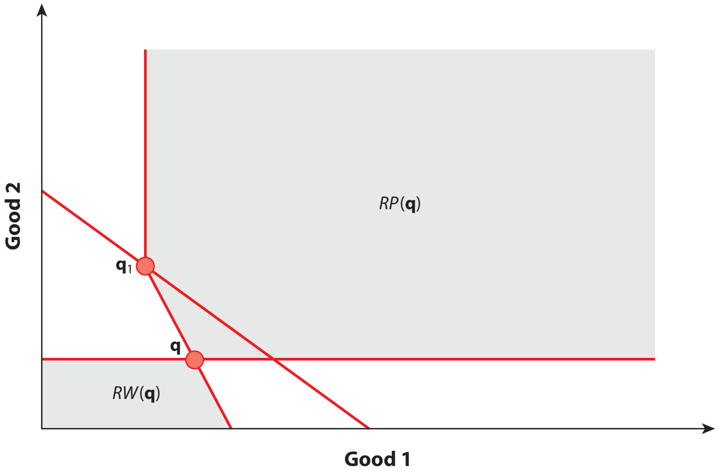

Figure 1

Recovery of the indifference curve. The set $RP(\mathbf{q})$ represents the set of all bundles that are revealed preferred to the unobserved bundle, $\mathbf{q}$, and the set $RW(\mathbf{q})$ contains all the bundles that are revealed worse to $\mathbf{q}$.

together if more observations are available and indeed are uniquely determined as the price-quantity data become completely dense. Moreover, recent research by Blundell et al. (2003, 2008) shows that one can dramatically tighten these bounds by combining RP theory and the nonparametric estimation of Engel curves (see also our discussion in Section 4).

Characterizing indifference curves is not the only thing we can do on the basis of RP theory. We can also make predictions of consumer behavior in new situations, that is, situations in which the consumer is faced with a new budget set. **Figure 2** shows a data set with two observations. Suppose now that the consumer is faced with a new budget line, indicated by the dashed line. All bundles that exhaust this budget clearly are within the reach of the consumer. However, not all these bundles are consistent with GARP. Actually, only the bundles on the blue line segment are consistent with GARP. The other bundles on the dashed line generate inconsistencies with rationality in the sense that they are not cost minimizing with respect to their revealed preferred set. Once again, the set of possible rational outcomes (weakly) clearly shrinks if more observations are available.

3. ALTERNATIVE MODELS OF RATIONAL DEMAND

One of the main focuses of recent research has been the development of RP characterizations of an increasing variety of economic models. Chambers et al. (2014) show that in principle there exists a set of RP-type conditions for any optimizing model that can be expressed as a series of universal statements. Thus, many models of interest to economists can be given an RP characterization. In this section, we do not aim at formally stating all the RP results available in the literature. Instead, we review some of the more fundamental results to provide a good starting point and orientation for the interested reader. This overview is structured around three topics. Section 3.1 focuses on special functional form restrictions that are frequently used to add some more structure to the basic model of rational demand. Section 3.2 discusses extensions of the basic model by relaxing some of the underlying assumptions. Finally, Section 3.3 deals with multiperson behavior.

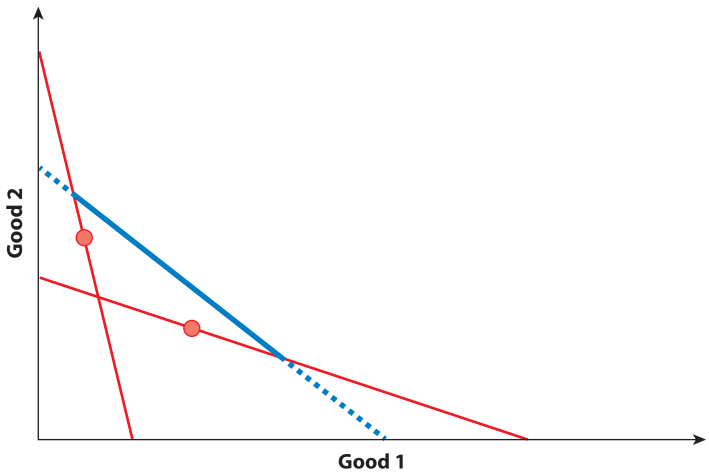

Figure 2

Forecasting new quantity bundles. The blue line segment represents the bundles that are consistent with GARP, and the dashed line represents a new budget set for the consumer. The red circles represent two chosen quantity bundles.

3.1. Investigating Functional Forms

In the basic model of rational demand discussed above, we are considering any type of (well-behaved) utility function. In other words, consistency with GARP, for instance, implies that there exists at least one utility function that allows a description of the data in terms of the behavioral model. However, in general, the class of utility functions is often restricted to simplify the (empirical) analysis.

RP theory allows researchers to investigate these extra assumptions. That is, if the data satisfy GARP, for instance, but not the RP characterization corresponding to the specific class of utility functions, then we can conclude that the problem is not the rationality of preferences per se, but is rather the further restriction on the form of preferences. The results discussed below allow for such tests.

3.1.1. Homotheticity. A utility function is homothetic if it is the positive monotonic transformation of a function that is homogeneous of degree one. This class of functions includes well-known types of utility functions, such as Cobb-Douglas utility functions and constant elasticity of substitution utility functions. Working with this class of utility functions implies, for instance, that Engel curves are straight lines through the origin, meaning that it is straightforward to model income effects for given prices, which in turn is convenient for extrapolating demand behavior.

Based on Afriat (1972) and Diewert (1973), Varian (1983) presents an Afriat-type theorem for characterizing rational demand in terms of homothetic utility functions. Essentially, his characterization combines the Afriat inequalities presented in Theorem 1—we still require that the data are rationalizable—with the extra assumption of a constant income effect for given prices. This results in a system of linear inequalities that the data need to satisfy to be rationalizable by a homothetic utility function. More precisely, Varian (1983) shows that a solution for the Afriat inequalities should exist for which λ_t equals U_t. He also provides a combinatorial reformulation of this system, which he labeled HARP (homothetic axiom of revealed preference), and he shows that the well-behavedness of the utility function is again not testable.

Cherchye et al. (2013a) extend this discussion to the class of quasi-homothetic utility functions. That is, the corresponding Engel curves are straight lines but not through the origin. Remarkably, these authors show that, in the absence of proportional prices, RP characterization boils down to GARP. In other words, assuming quasi-homotheticity is not restrictive at all.

3.1.2. Separability. Separability implies that the goods can be divided into groups and that for each group, there is a subutility function capturing the preferences for those goods, which is independent of the consumption of goods outside of that group. In addition, there is a macro utility function that aggregates the preferences over the groups. If this macro function can be any well-behaved utility function, then we are assuming weak separability. If this macro function is additive in terms of the subutility functions, then we are considering additive separability. Finally, latent separability means that goods can be part of several groups.

Separability is a very strong but useful assumption in applied work. For example, it allows researchers to focus on individual markets for related goods, and combined with (quasi-) homotheticity, it also implies two-stage budgeting, which simplifies the analysis of consumer behavior.

Varian (1983) presents the RP characterizations of both weak and additive separability (see also Afriat 1969 and Diewert & Parkan 1985 for related results). Crawford (2004) provides the RP characterization of latent separability. These characterizations state that we need two types of utility functions. First, for every group, we need a subutility function capturing the preferences for these goods. As such, the observed data for each group need to satisfy the RP conditions discussed in Section 2. Second, we also need a macro utility function aggregating the preferences over the groups. Again this reduces down to the usual RP conditions, but this time in terms of unobservable information (i.e., the unobserved utility levels and marginal utilities of income that solve the Afriat inequalities for each group).

All this implies that the data need to satisfy a system of nonlinear inequalities, which is not attractive from an empirical point of view. The only exception involves the RP conditions of additive separability, as in that case the marginal utility of income is assumed to be constant. This implies that the macro utility function is simply the sum of the subutility functions, and as such, we do not need extra conditions to reconstruct this function (i.e., the RP conditions in terms of unobservable information are redundant). This nontestability of the characterization of weak separability led to several papers focusing on either necessary or (separate) sufficient conditions for testing weak separability (see, e.g., Swofford & Whitney 1987, 1988, 1994; Barnett & Choi 1989; Fleissig & Whitney 2003, 2007, 2008). Finally, in a recent working paper, Cherchye et al. (2012) present an IP formulation for the setting with two subgroups. Attractively, this makes the RP test easy to apply for this special case of weak separability.

3.1.3. (Generalized) quasi-linear utility functions. A final class of often-used utility functions involves (generalized) quasi-linear utility functions. These are utility functions that are linear in at least one good, usually called the numéraire. This has strong implications (e.g., the absence of income effects for all but a single good, risk neutrality) that simplify the empirical analysis substantially. Generalized quasi-linear utility functions slightly relax the linearity assumption by allowing that the numéraire is multiplied by a function defined in terms of a subset of goods, say X. Bergstrom & Cornes (1981, 1983) and Bergstrom (1989) show that this type of preference is equivalent to assuming that utility is transferable among consumers, as long as the subset of goods X are public goods to all these consumers. Transferable utility in turn is a popular assumption in matching models to define the stability of the matchings.

Brown & Calsamiglia (2007) present an Afriat-type theorem for quasi-linear utility functions that essentially adds to Theorem 1 that the marginal utility of income should be constant (i.e., $\lambda_t = \lambda$ for all $t \in T$). Cherchye et al. (2014) extend their results toward generalized quasi-linear utility functions. In both cases, the tests are easy to apply, although in the latter case, it is via IP (which can be time-consuming).

3.2. Investigating Richer Models

In the previous subsection, we focus on extra functional assumptions that restrict the class of utility functions to simplify the (empirical) analysis. In this subsection, we take a different stance by relaxing the assumptions underlying the basic rationality model. That is, until now, we had a consumer in mind who does not take intertemporal issues into account, who faces linear budget sets, and who is consuming a set of nondiscrete goods. Below we review some seminal contributions that focus on relaxing these assumptions to obtain a more realistic model. Importantly, all the results we present are fairly easy to apply, which again makes RP theory more attractive in applied work.

3.2.1. Intertemporal behavior. The model studied in Section 2 does not consider the problem of intertemporal allocations. Implicitly, while taking a decision in some observation t, the consumer does not take decisions from the past or for the future into account. There are, of course, many reasons to argue that this is a naive assumption. But at the same time, these dynamic or intertemporal models are also much more complicated. Indeed, as the future is uncertain, one should ideally also study risk attitudes or work with expected utility. Readers are referred to Varian (1988) for RP results related to risk aversion and Green & Osband (1991) for an RP analysis focusing on expected utility.

The following papers abstract from this uncertainty to derive some benchmark results. Browning (1989) was the first to present an RP characterization of a life-cycle model. In this model, the consumer decides at the beginning of his or her total life consumption plan to smooth consumption over all the periods, which obviously takes future decisions into account. This smoothing implies that the marginal utility of income should be constant over the whole time horizon. Moreover, Browning assumes that the decisions for some period are not influenced by consumption in other periods. Given all this, the RP conditions are equivalent to the linear system of Afriat inequalities discussed in Theorem 1, but this time using discounted prices and, crucially, a constant marginal utility of income. That is, as in the case of quasi-linear utility functions, we need $\lambda_t = \lambda$ for all $t \in T$. As a bibliographical note, we mention that Browning (1989) does not present this precise set of linear inequalities but rather presents an alternative form; readers are referred to Adams et al. (2012) for more discussion on how to derive the linear inequalities. Additionally, it is worth mentioning that Browning (1989) does not (explicitly) include time discounting. This has been added by subsequent authors but at the cost of making the inequalities nonlinear in unknowns—the discount rate interacts with the Afriat numbers. This makes the resulting test much less computationally convenient as it is necessary to grid search over the discount rate, investigating the conditionally linear inequalities at each node. Finally, Crawford (2010) and Demuynck & Verriest (2013) extend Browning's (1989) model by providing the RP conditions for models that allow habit formation or addiction (i.e., consumption in some period depends on consumption in other periods).

3.2.2. Nonlinear budget sets. The results stated in Afriat's theorem crucially depend on the linearity of the budget set. To test GARP, one should check the cost-minimization condition, and this condition can easily be operationalized owing to the linearity of the budget set. This is also clear from the equivalent linear program stated in condition (*d*). Another important implication of linear budget sets is that concavity does not have testable implications, essentially because choices in regions of nonconvexity could never be observed (see Forges & Minelli 2009 for more discussion).

Moreover, there is also an empirical motivation to consider nonlinear budget sets. Indeed, because of tax systems, most labor supply applications of RP theory have to deal with nonlinear budget sets. Or richer intertemporal models that try to incorporate the imperfect workings of financial markets can also lead to nonlinear budget sets.

Matzkin (1991) presents the first RP results in the setting of nonlinear budget sets and concave utility functions. Her results are extended by Forges & Minelli (2009), who drop the concavity of the utility function. As mentioned above, this allows them to show that concavity has testable implications. Finally, Cherchye et al. (2013c) combine the two previous papers by deriving the RP characterizations for very general budget sets and concave utility functions. These authors also provide linear programming formulations of their results, which makes these results attractive for applied work. Essentially, all these papers present Afriat-type theorems in which the linear budgets are replaced by a convenient representation of the nonlinear budget sets.

3.2.3. Discrete goods and characteristics. The results presented by Blow et al. (2008) and Polisson & Quah (2013) allow researchers to relax the assumptions related to the consumed goods. Blow et al. (2008) focus on the setting in which consumers are interested in the characteristics of the goods (and not in the goods themselves). Polisson & Quah (2013) deal with a setting in which goods can be discrete in nature. Both assumptions are crucial for the realistic nature of empirical applications, but they also make the (theoretical) analysis more complex.

Indeed, models of preferences over characteristics instead of preferences over goods imply that the empirical analyst no longer directly observes the willingness to pay. That is, the price paid for the good needs to be decomposed into prices that the consumers are willing to pay for the characteristics. As such, in the RP conditions discussed in Section 2, we should replace the market prices p_t by unobserved shadow prices. Next, if goods are discrete in nature, then this implies that the (nonsatiated) consumers can no longer exhaust their budgets. As such, one needs to deal with the possibility that some of the budget is left over—but the extent of this is not usually known.

3.3. Investigating Multiperson Behavior

The above two subsections focus on alternative versions of our basic model. However, in the end, all these models still have the consumer maximizing some utility function subject to a budget constraint. In this subsection, we go one step further by presenting RP results for multiperson behavior. This is important because empirical applications of RP theory are generally applied to household consumption data. There is a lot of empirical evidence that such applications should take into account the existence of multiple decision makers in multimember households (see, e.g., Vermeulen 2005 for an overview and Cherchye & Vermeulen 2008 and Cherchye et al. 2009 for evidence based on RP tests).

Models of multiperson behavior therefore use a different starting point. Instead of assuming that there is a utility function representing the preferences of the group (or household), these models take into account that each individual has his or her own utility function and that individuals enter into a decision process with the other individuals for deciding how to spend the common budget. The outcome of this decision process should not necessarily lead to a transitive preference ordering. That is, there does not need to be a macro utility function that aggregates the individual preferences.

We start by reviewing the classical answer to these types of questions, namely general equilibrium theory and aggregation. These models take a societal viewpoint. Subsequently, we discuss the recent RP literature on household models. Although our discussion below is formally related to our discussion of separability, we argue that the absence of the macro utility function results in quite different RP characterizations.

3.3.1. Modeling society. The RP analysis of multiperson behavior began with Brown & Matzkin (1996), who focus on a simple exchange equilibrium in which market prices, individual incomes, and aggregate endowments are observed. To obtain testable implications for this setup, these authors derive the conditions that guarantee that the observed data lie on the so-called equilibrium manifold. That is, they show that individual rationality and market clearing restrict the response of endogenous aggregate variables to perturbations on individual endowments, which in turn allows them to state their Afriat inequalities for this setting. More precisely, Brown & Matzkin show that the (observed) aggregate endowments should be decomposed into (unobserved) individual quantity bundles that exhaust the (observed) individual incomes at (observed) market prices. Given individual rationality, the individual quantity bundles and the market prices should then satisfy the RP conditions discussed in Section 2.

Unfortunately, the system of Afriat inequalities is nonlinear in nature owing to the unobservedness of the individual quantity bundles. To obtain RP conditions that are empirically attractive, Brown & Matzkin (1996) subsequently apply the Tarski-Seidenberg algorithm to obtain testable implications for a data set with two observations. All in all, this is a quite surprising result that contrasts with the conclusions obtained by Sonnenschein, Mantel, and Debreu, who basically state that general equilibrium models do not generate testable implications, and it has generated a lot of follow-up research (see Carvajal et al. 2004 for a survey and Cherchye et al. 2011c for a recent contribution).

Another question related to modeling the society is the aggregation problem. That is, does there exist a social welfare function representing the preferences of the society, and if so, what is its relation to the preferences of the individuals in that society? As discussed by Varian (1984), the answer to the first question is equivalent to requiring that the aggregate data (i.e., the sum of the individual demands and the common price) satisfy GARP. However, this social welfare function cannot be used to make normative conclusions, simply because there is no relation to the individuals in the society. We refer readers to Cherchye et al. (2013a) for RP characterizations that do allow for aggregating the preferences of the individuals in the society. Because of the need for a macro utility function, this problem is formally related to our discussion of separability and its corresponding empirical issues.

3.3.2. Modeling household behavior. As discussed above, to model household consumption decisions, one should take into account the individual preferences of the household members and the decision process. There is a wide variety of possibilities for modeling this decision process, with the so-called collective model the most popular one (see Chiappori 1988 for a seminal contribution, which also contains some RP theory). Collective models assume that individuals are rational and allow for any kind of decision process, as long as the outcome is Pareto efficient. This implies that there does not necessarily need to be a macro utility function that aggregates the individual preferences, which distinguishes this model from the separability ones.

Cherchye et al. (2007, 2010, 2011a) present the RP characterizations of collective models. The difference between these characterizations depends on the nature of the goods in the analysis. More precisely, goods can be consumed privately by one of the household members, publicly by all household members, or even both. For instance, a car can be used privately by the husband to drive to work on weekdays and publicly by the household on the weekend to go on a family trip.

In all the RP characterizations, the starting point is that in almost all expenditure surveys, there are price-quantity data available only at the household level, and not at the level of each individual household member. Similar to the exchange equilibrium model discussed above, the RP characterizations therefore consist of several steps. First, we need to construct individual-specific (unobserved) quantity bundles that add up to the (observed) household quantity bundle. Second, to capture the (partly) public nature of the quantities, we also need to construct individual (unobserved) shadow prices that add up to the (observed) market price. Finally, given individual rationality, the individual shadow prices and the individual quantity bundles for each household member should satisfy RP conditions, such as the ones discussed in Section 2. As before, the RP conditions are not directly empirically useful, as they are defined in terms of unobservables. To deal with this problem, Cherchye et al. (2008, 2011a) develop IP formulations similar to IP-GARP that can easily be applied. That is, the conditions are linear in the unknowns, but some of the unknowns are binary variables.

As a final remark, we note that there are also RP characterizations available for alternative forms of modeling the household decision process. First, one could replace the Pareto efficiency assumption by the assumption that the outcome of the decision process should be a Nash equilibrium. Such a model again puts minimal structure on the decision process, but it also takes into account that individuals can behave strategically (see Cherchye et al. 2011b for a more in-depth discussion and for the corresponding RP characterization). Second, one could put more structure on the decision process by assuming that the households take decisions according to some specific bargaining protocol. The RP theory of some of the most popular bargaining models, such as Nash bargaining, is presented by Cherchye et al. (2013b), Carvajal & González (2014), and Chambers & Echenique (2014).

4. BRINGING REVEALED PREFERENCE THEORY TO THE DATA

As discussed above, there has been a significant broadening of the scope of RP methods since the foundational work by Samuelson (1938, 1948), Houthakker (1950), and Afriat (1967). The practical empirical application of RP methods has arguably lagged somewhat. This may be a result of the relative unfamiliarity of RP methods and the fact that empirical RP often requires researchers to address and find practical solutions to some difficult combinatorial problems. Moreover, empirical RP work with sample data presents a number of important challenges. Consider, for example, the question of a straightforward GARP test. To begin, there is the matter of interpreting the test's outcome for a single economic agent—what should we make of it if the subject passes/fails? That problem is made more difficult (and indeed the test itself may be hard to conduct) if there are problems with the data, such as measurement error or missing data. When we have data on a number of different individuals, the question of the pattern and nature of preference heterogeneity arises. Finally, there is the problem of going beyond the data at hand and making inferences about some population of interest. In the following subsections, we discuss these issues in the context of the basic model of rational demand.

4.1. Interpreting RP Tests

Consider the $\{\mathbf{p}_t, \mathbf{q}_t\}_{t\in T}$ data for a single individual and suppose that everything is measured perfectly. Whether this individual's behavior is rationalizable by the theory is completely deterministic: If the data satisfy GARP, then they are consistent with utility maximization; otherwise, they are not. Despite this disarming simplicity, it can still be hard to know what to make of the result.

4.1.1. Sensitivity for detecting alternative behavior. For example, suppose that utility maximization was not the data-generating process (DGP). Will the RP conditions be sensitive enough to detect it? In statistical hypothesis testing, this question concerns the power of the test of a probabilistic model against a probabilistic alternative. In RP tests in this kind of nonstochastic environment, the statistical notion of power is not strictly relevant, yet there is clearly a need to consider the same sort of question, and many of the same considerations apply. In particular, as is the case with statistical power calculations, the answer will depend on the alternative DGP considered: The RP test might be quite successful at detecting violations of GARP under some alternative DGPs but may be less successful given others. The difficulty is that there are many alternatives to rational choice models but no obvious benchmark.

One important, nonrational alternative considered by Becker (1962) is a probabilistic DGP: uniform random choice on the budget constraint. Bronars (1987) applies this in an RP context by calculating the probability of observing a violation of GARP with this DGP operating on the observed constraints. Bronars's approach remains the most popular method, but more recent contributions (notably Andreoni et al. 2013) consider more data-driven alternatives to uniform random choice, while sticking with the idea of a probabilistic alternative DGP. They suggest drawing from the empirical distribution of observed choices to allow for a more realistic alternative. Work on this topic is ongoing, but the leading approaches, which use probability models to frame alternative choice models, are principally variations on Bronars's method.

4.1.2. Economically significant rejections of rationality. A different approach is to try to avoid the problem of having to specify the alternative DGP and, instead of asking whether the outcome of an empirical RP test represents a statistically significant departure from a probabilistic DGP, ask whether the results of the test represent an economically significant departure from rational choice. The key to this approach is to see that when a consumer violates RP conditions, this consumer appears to waste money by buying a consumption bundle when a cheaper bundle is available and also revealed preferred to it. The cost-efficiency measure suggested by Afriat (1973) is the smallest amount of this wastage (as a fraction of the overall budget) consistent with the given demand data. This index provides a simple way of measuring the size of a GARP violation and does so in units that are easy to understand and interpret economically. Andreoni et al. (2013) propose the converse of the Afriat cost-efficiency index as a way of interpreting GARP successes: Given a data set in which no RP violations are detected, the Afriat power index measures the amount the consumer's budget would have to be adjusted to induce a violation. If the required adjustment is small, then the test is considered to be sensitive; if it is high, then the test is not sensitive.

A related approach builds on the ideas of de Finetti (1992 [1937]) concerning Dutch books or money pumps. The idea is that individuals who violate RP conditions have preferences that contain cycles, which means that they are open to being exploited as a money pump by an unscrupulous trader who simply buys goods from them at a price they are willing to accept and then sells the goods back to them again at a (higher) price they are willing to pay. Given an RP cycle of length J with $\mathbf{q}_j R^0 \mathbf{q}_{j-1}$, the intransitivity means that the consumer would also prefer $\mathbf{q}_{j-1}$ to $\mathbf{q}_j$, so $\mathbf{p}_j\mathbf{q}_j - \mathbf{p}_j\mathbf{q}_{j-1}$ can be extracted at each point in the cycle and $\sum_{j=1}^{J}\mathbf{p}_j\mathbf{q}_j - \mathbf{p}_j\mathbf{q}_{j-1}$ in total. Echenique et al. (2011) suggest the money pump (expressed as a proportion of the consumer's total expenditure) as an aid to interpretation when GARP fails: the more money that can be extracted from an individual in this manner, the worse the violation of RP theory. These authors also show how to address the considerable combinatorial/computational challenges involved in calculating the money pump index, as the number of potential cycles that need to be investigated can be huge, even when the data set itself is not large.

4.1.3. Trade-off between sensitivity and pass rate. A last alternative approach, which Beatty & Crawford (2011) apply to RP tests, comes from the literature on experimental game theory and is due to Selten & Krischker (1982) and Selten (1991). The key insight is that in their RP guise, shorn of special functional form assumptions, economic models generally generate restrictions in the form of well-defined sets of choices that are consistent with the model of interest. In the context of models that predict sets, it is useful to consider the feasible outcome space (e.g., P) and the model's prediction as the subset ($S \subseteq P$). It is then important to acknowledge the relative size of the predicted/theoretically consistent subset. The essential idea is that if the set of observations explainable by the model (S) is very large relative to the set of behaviors that the consumer could possibly display (P), then simply noting that many observed choices lie in S is not a demanding requirement—the observed choices could hardly have done otherwise, and the test is therefore not very sensitive. This means that fit alone (the proportion of the sample that passes the relevant test) is not a sufficient basis for assessing the outcome of an RP test. A better approach would be to consider the trade-off between the pass rate and some measure of the test's sensitivity. Let a denote the size of the theory-consistent subset relative to the outcome space for the model of interest. The relative area of the empty set is zero, and the relative area of all outcomes is one, so $a \in [0, 1]$. Now suppose that we have some choice/outcome data. Let r denote the pass rate; this is simply the proportion of the data that satisfies the restrictions of the model of interest. Selten (1991) provides an axiomatic argument that the trade-off between the ability to fit the data and the restrictiveness of the theory should be the difference measure: $r - a$. Other axiomatizations would produce different forms for the measure of the outcome of an RP test, but the basic idea that the measure should combine both the pass rate and some measure of sensitivity remains an important and promising area for further work.

4.1.4. Bayes' theorem. Returning to the question with which we started, in the light of the foregoing discussion, what are we to make of the fact that the $\{\mathbf{p}_t, \mathbf{q}_t\}_{t \in T}$ data for a single individual satisfy GARP, for example? How justified might we be in thinking that this individual is, heuristically at any rate, really a utility maximizer? Clearly our assessment of this depends on many of the issues discussed above: If the number of observations is small, or if we suspect that the ability of the GARP condition to detect nonrational behavior is weak, then the evidence is probably weak, and we may be unwilling to conclude simply that this person must be a utility maximizer because he or she has passed GARP. Bayes' theorem is, as usual, a useful framework for thinking about this issue.

We are interested in whether the individual is a utility maximizer (denoted U), given that the data satisfy GARP (denoted G). Retaining our assumption of no optimizing or measurement error, we have

$$P(U|G) = \frac{P(U)}{P(U) + P(G|\neg U)[1 - P(U)]},$$

where $P(U)$ is the prior. The key term is $P(G|\neg U)$, the probability of passing GARP if the individual is not a utility maximizer. If the GARP test is not able to detect nonrational behavior very well, then $P(G|\neg U)$ is close to one, and $P(U|G) \to P(U)$, which means that the evidence of the successful GARP test should do little to shift our prior beliefs. If, however, the GARP test is very sensitive and $P(G|\neg U)$ is close to zero, then $P(U|G) \to 1$, and consequently the GARP test gives us rational grounds to become very confident that the individual is in fact a utility maximizer. The term

$P(G|\neg U)$ is therefore centrally important. As discussed above, its value depends on the alternative DGP. Bronars's approach, or a variation on it along the lines of Andreoni et al. (2013), would seem to be the natural way to calculate $P(G|\neg U)$ in practice. Combining their approaches with a Bayesian perspective on the question of interest therefore seems a sensible way to make progress.

4.2. Missing Data

Suppose now that the data are less than perfect. In particular, consider the case of missing data. It would seem that missing data are fatal to the empirical implementation of RP methods. In some cases, this is true, but in others, progress can still be made. With a full set of observations, an RP test asks whether there exists a well-behaved utility function that rationalizes these data. When some of the data are missing, we can ask a slightly different question: Are there feasible values for the missing observations such that there exists a well-behaved utility function? In some cases, the answer may always be yes, implying that the utility maximization hypothesis cannot be falsified, and the test collapses. One such situation occurs when all the price or quantity data for a particular good are missing; in this situation, Varian (1988) shows that it is not possible to test RP conditions because it is always possible to find values for the missing price or quantity series such that the data pass the RP conditions. However, the situation is not always so bleak. A common example of missing data in consumer surveys concerns prices that are recorded when the consumer makes a purchase but that are not recorded when the consumer does not make a transaction. Thus, we observe prices when the quantity is positive but not when the quantity is zero. In these situations, there often are restrictions on what the missing prices can possibly be, and by the same token, RP conditions can be violated if these conditions are not met. Blow et al. (2008) describe how to formulate RP tests with this type of missing data in the context of linear characteristics models, but because these models, in which consumers have preferences for characteristics instead of goods, can be easily transformed into the standard preference-for-goods model, the method they describe also works perfectly for the canonical RP test.

4.3. Statistical Errors

As emphasized in Section 1, an important difference between structural econometrics and empirical RP lies in the absence of an error term in the latter. Certainly, error terms rarely appear in RP theory: There is no mention of an error term in Afriat's theorem or in any of the other RP characterizations of the various models discussed in Section 3. But as soon as we attempt to take those RP conditions to data, errors can no longer necessarily be ignored. The most obvious situation arises when we consider measurement errors, but identical issues arise when RPs are applied to statistical objects (e.g., estimates of aggregate consumption as in Browning 1989 or nonparametric Engel curves as in Blundell et al. 2003, 2008). In these cases, the price-quantity data we observe are a function of a random variable. This introduces a statistical element to empirical RP and forms an important link between RP and structural econometrics, which, as discussed in Section 5, appears to be an important future direction for research.

To illustrate the case for classical additive measurement error, consider the model

$$\mathbf{q}_t = \mathbf{q}_t^* + \mathbf{e}_t,$$

where $\mathbf{q}_t^*$ denotes the true values of demands, and $\mathbf{e}_t$ is a vector of classical measurement errors. Suppose we are interested in the null hypothesis that the true data $\{\mathbf{p}_t, \mathbf{q}_t^*\}_{t\in T}$ satisfy GARP. Blundell et al. (2008), building on Varian (1985), construct a statistical test for violations of the RP conditions by supposing that the observed demands are known functions of a finite set of parameters θ_t so that $\mathbf{q}_t = \mathbf{f}(\theta_t)$ for known $\mathbf{f}(\cdot)$. The RP restrictions in the null can be represented by a set of moment inequality restrictions involving θ_t. Blundell et al. (2008) then show that it is possible to appeal to results by Manski (2003), Chernozhukov et al. (2007), and Andrews & Guggenberger (2009) for moment inequality estimators of this type. They establish that there always exist values θ_t that satisfy the moment inequality restrictions as long as the support of the θ_t values allows for any positive demands that satisfy adding up. Generally, there will be a set of values for θ_t that satisfy the RP restrictions, and testing consistency with these conditions is a matter of verifying if this set includes the observed demands. If the RP conditions fail for the observed demands $\mathbf{q}_t$, it is possible to generate a restricted estimator, $\hat{\mathbf{q}}_t$, using the following Gaussian quasi-likelihood ratio or minimum distance criterion function:

$$L = \min_{\{\hat{\mathbf{q}}_t\}_{t\in T}} \sum_{t=1}^{T} \left(\mathbf{q}_t - \hat{\mathbf{q}}_t\right)' \Omega_t^{-1} \left(\mathbf{q}_t - \hat{\mathbf{q}}_t\right),$$

subject to the restriction that $\{\mathbf{p}_t, \hat{\mathbf{q}}_t\}_{t\in T}$ satisfies GARP, and the weight matrix Ω_t^{-1} is the inverse of the covariance matrix of the demands. The solution to this problem leads to demands $\hat{\mathbf{q}}_t$, which satisfy the RP restrictions and are unique almost everywhere. Evaluated at the restricted demands, Blundell et al. (2008) show that the above distance function also provides a test statistic for the RP conditions and that this test falls within the general class of misspecification tests investigated by Andrews & Guggenberger (2009, section 7).

4.4. Heterogeneity

For anyone who has ever looked at consumer micro data, the great variety of behavior on display among consumers and households who are, in most observable respects, very similar is striking. Ideally, the researcher would try to model each household individually, but most consumer panels are small-T, large-N affairs. This makes it impossible to estimate sufficiently flexible and reliable econometric models at the individual level. The standard structural econometric approach is therefore to pool data across consumers and to model the behavior of individuals as a combination of a common component and an idiosyncratic component that reflects unobserved heterogeneity. Of course, this immediately requires a combination of often strong assumptions regarding the form of the statistical model and the joint distribution of unobserved heterogeneity with the observables (see, e.g., Brown & Walker 1989 and Lewbel 2001 for more discussion).

Because RP approaches can be applied to very short panels (e.g., one needs only two observations on a consumer to test GARP), it is generally possible to proceed individual by individual, even when the T dimension is far too small even to contemplate a statistical approach. This one-at-a-time approach of course allows for the maximal amount of heterogeneity—consumers can differ with respect to whether they behave in accordance with the theory, and if they are theory consistent, then they can differ with respect to choices and preferences.

However, in some circumstances (e.g., a pure cross-sectional data set in which individuals are observed only once), heterogeneity cannot be usefully preserved, and indeed, sometimes heterogeneity itself is the object of interest. When this is the case, RP methods can still be used. Instead of applying them to longitudinal data on individual consumers and checking for the existence and stability of well-behaved preferences, one can apply them to cross-sectional data on many different consumers; RP restrictions are then interpretable as a check for the commonality of well-behaved preferences.

Gross (1995) applies RP tests to cross-sectional consumer data to look at the evidence for and against the assumption of homogeneous tastes and concludes that, in a sample drawn from the Panel Study of Income Dynamics, wave IX (1976), individuals did not share a common utility function. The idea that the choices of all the consumers in a large microeconomic data set could be explained perfectly by a single common utility function is probably, as Lewbel (2001) points out, "implausibly restrictive." Dean & Martin (2010) and Crawford & Pendakur (2013) recognize that tests like this one will reject as soon as one consumer has tastes different enough from the rest to be detected by the test. The possibility then that the rest of the data are rationalizable by a single utility function would be masked by the rejection caused by the presence of this single consumer. Investigating this further would in principle require the researcher to look at all possible subsets of the data and to conduct RP tests in all of them to detect the true pattern of preference heterogeneity. This is too computationally demanding as there will be 2^N subsets to check, so this is another example in which researchers have had to take an algorithmic approach.

Dean & Martin (2010) suggest looking for the largest single subset that is consistent with common preferences—this is then a nice summary of an aspect of preference heterogeneity. To do this, they develop a new algorithm, which is much more efficient than existing algorithms, that exploits an analogy between the RP problem at hand and the minimum set covering problem, which is a well-studied problem in the computer sciences and operations research literature. Although the minimum set covering problem is NP hard, there is a wide variety of algorithms that are extremely efficient, so by cleverly translating the RP problem into this form, the authors are able to apply these solution methods. Crawford & Pendakur (2013) take a slightly different approach. Given a result like the one in Gross (1995), the researcher clearly needs more than one utility function to model the data. The question then of course is, How many different utility functions are needed? Crawford & Pendakur (2013) consider the problem of how to find the minimum number of utility functions necessary to fully explain all observed choices in a data set. This is a computationally demanding partitioning problem, and they design an algorithm that is able to place tight, two-sided bounds on this minimum number.

4.5. Inference

If the data involved are a random panel sample of households, and demands are measured without error, then inference about objects such as the proportion of households that satisfy RP restrictions in the population is straightforward. A sample proportion can be viewed as the fraction of successes in N independent Bernoulli trials with the same success probability p. The central limit theorem implies that for large N, the sample proportion $\hat{p} = \sum_{i=1}^{N} I$ (consumer i passes RP) is normally distributed with mean p and standard deviation $\sqrt{p(1-p)N}$, so the statistic $z = (\hat{p} - p)/\sqrt{p(1-p)N}$ follows the standard normal distribution. This serves as the basis for statistical inference regarding population proportions.

Inference with repeated cross sections from a heterogeneous population is more difficult. The issue here is that we do not see the same consumer twice, so we cannot proceed on a consumer-by-consumer basis, checking the RP conditions for each one as before. The object of interest remains the population proportion of consumers who satisfy the RP conditions. However, this parameter depends on the joint distribution of choices over different budget sets, and repeated cross-sectional data do not reveal this: Only its marginal distributions can be observed. Thus, the population parameter of interest is not point identified. Hoderlein & Stoye (2014) show that in the context of WARP, it can be partially identified (i.e., bounded). They describe the problem as a copula problem and use copula techniques to analyze it. They also show that inference on the bounds is an application of partial identification through moment inequalities. This approach is somewhat in the tradition of the literature on the partial identification of treatment effects, and it emphasizes the conceptual value of clearly understanding how much might be learned from the data without identifying assumptions. Consequently, the approach is careful to impose no or very weak homogeneity assumptions, and as a result it seems that WARP may be hard to reject. However, it is important to note that WARP does not exploit transitivity of preferences, a much stronger assumption, so it remains to be seen how this approach might be fruitfully extended to RP conditions, which are more demanding.

5. CONCLUSION

This review focuses on the present state of empirical RP in two general respects: work that extends the scope of these methods to a variety of richer models of behavior and work that seeks to apply these methods to data. To conclude, we briefly consider how each of these areas might develop in the future. In terms of scope, we note that all the RP characterizations of models discussed above have remained firmly embedded within the neoclassical tradition in which the whole literature began. It therefore seems to be an interesting and open question to ask whether these methods might be applied to nonstandard behavioral models.

As far as empirical applications are concerned, above we describe how empirical RP differs from structural econometrics and the benefits and drawbacks of the approach. One of the principal drawbacks of RP methods is that, compared to structural econometric methods, they are relatively ungainly. In other words, they produce characterizations of preferences, for example, that are difficult to represent succinctly and awkward to interpret (e.g., piecewise linear bounds computed on individual indifference curves). This compares unfavorably with the traditional econometric approach, which focuses on simple functional forms with parameters that have useful and straightforward economic interpretations. Conversely, RP methods seem to make fewer maintained assumptions than standard methods. An important area for research may therefore be to investigate whether it is to blend empirical RP and econometric methods and preserve the strengths of both approaches.

5.1. Behavioral Models

Recently, there has been renewed academic interest in economic models that move somewhat away from the neoclassical tradition of treating people as always-rational decision makers. This behavioral economics approach focuses on models that combine conventional neoclassical microeconomic methods with behavioral and modeling assumptions that have more plausible sociological and psychological foundations. For instance, they allow for situations in which people are influenced by others, may make mistakes, or may come to regret their choices. Behavioral economics promises much in terms of its potential to help us understand choices that could otherwise prove resistant to straightforward explanation by standard rational choice models. At present, it remains something of an open question as to whether these models might be amenable to an RP characterization.

Neoclassical models in economics, for all the (often justifiable) criticisms they attract, are at least falsifiable in an RP sense—it is possible in principle to detect when the data and the model are not rationalizable. There is, as far as we know, nothing like an Afriat's theorem for behavioral economic models. This is of interest because it is important to know whether, without auxiliary hypotheses, observational data are able to tell us when behavioral models are unable to rationalize behavior. In this respect, we want to make some concluding remarks on two particularly interesting classes of behavioral models: reference-dependent preferences and time-inconsistent choices. As discussed below, these classes are sufficiently close to existing neoclassical models that do have RP characterizations. So it might be possible to investigate whether they are characterizable by an Afriat-type theorem.

5.1.1. Reference-dependent preferences. Reference-dependent preferences incorporate ideas from prospect theory. Tversky & Kahneman (1991) posit that individuals understand their options in decision problems as gains or losses relative to a reference point. The reference point is not generally observable (to the researcher): Sometimes it is modeled as the current position (i.e., the status quo) of the individual, but it might also depend on past consumption, expectations, social comparisons, social norms, etc. A feature of prospect theory, which reference dependence inherits, is that the value function exhibits loss aversion so that negative departures from one's reference consumption level decrease utility by a greater amount than positive departures increase it. Another feature of prospect theory is that the value function exhibits diminishing sensitivity for both gains and losses, which means that the value function is concave over gains and convex over losses. Taken together, this implies that changes in an unobservable reference point are capable of altering an individual's preferences. It would therefore seem that giving this model an RP characterization with empirical content is going to be far from straightforward, yet the model is deterministic and indeed rational in the sense that, conditional on the reference point, preferences are well behaved. Although this appealing setup is deterministic and indeed rational in the sense that, conditional on the reference point, preferences are well behaved, it seems that giving this model an RP characterization with empirical content is going to be far from straightforward owing to the unobserved reference point.

5.1.2. Time-inconsistent choices. Models of time-inconsistent choice relax the standard assumption that all the disparate motives underlying intertemporal allocations can be condensed into a single parameter, the discount rate, which is constant. Constant discounting entails an evenhandedness in the way a person evaluates time. It implies that a person's intertemporal preferences are time consistent, which means that later preferences confirm earlier preferences. Browning (1989) exploits this consistency in his derivation of an RP characterization of the strong rational expectations hypothesis. However, whereas the standard model assumes constant discounting, the leading behavioral alternative suggests that discounting is hyperbolic—that a person has a declining rate of time preference. This implies that when subjects are asked to compare a smaller-sooner reward to a larger-later reward, the implicit discount rate over longer time horizons is lower than the implicit discount rate over shorter time horizons (see, e.g., Thaler & Shefrin 1981). Once again, the model is perfectly rational and deterministic, but whether it has an RP characterization akin to Browning (1989) is the subject of ongoing work.

5.2. Empirical Revealed Preference and Structural Econometrics

At the beginning of this article, we emphasize a key difference between empirical RP and structural econometrics: Whereas empirical RP focuses on observables, the introduction of unobservables (error terms) is an essential aspect of structural econometrics. These error terms are there in part because the structural functions alone generally do not rationalize the data. The empirical RP approach, being based on the weaker requirements of inequality restrictions, generally has no need to resort to error terms. The great advantage, however, of structural econometrics is that it generally seeks to recover the structural functions of interest uniquely. Empirical RP, by contrast, can typically only place bounds on these. Moreover, if the bounds are wide, then the RP approach is arguably of little utility.

An important area of future research therefore lies at the boundary between econometrics and RP. In particular, there is the question of whether the inequality restrictions from RP arguments can be used to help guide the estimation of structural econometric models. In a sense, this may be simply a question of augmenting traditional econometric loss functions (e.g., sum of squared residuals, least absolute deviations) with loss functions motivated by RP theory. Blundell et al. (2008) provide an initial step in this direction. The authors estimate a system of Engel curves and impose RP restrictions with the finding that the resulting estimated Engel curves minimize least-squares losses subject to WARP. However, the RP restrictions are applied only locally (i.e., at a particular point in the income distribution), meaning that the entire Engel curve is not necessarily constrained to be consistent with a single set of well-behaved preferences. Imposing global consistency in an easily interpretable way is more challenging. In a recent development in this direction, Halevy et al. (2013) aim to fit a single, simple parametric utility function to choice data subject to RP conditions. The loss function in this case is based on the Afriat efficiency index applied at each observation. The objective of the approach is to select a simple, tractable representation of preferences that minimizes the inconsistency between the empirical RP information contained in the choices and the ranking information contained in the recovered preferences. Of course, one could just compute a piecewise linear, perfectly rationalizing utility function as in Afriat (1967), but that method requires recovering twice the number of parameters as there are observations, and the behavioral content of the utility function is almost impossible to interpret. The tradition in econometrics is to work with the simplest model that allows the researcher to adequately fit it to the data and also to interpret it (e.g., to have a simple characterization of concepts such as elasticity of demand, risk aversion, or time preference). Drawing on this econometric approach, the authors' method rather neatly trades off the inevitable misspecification of the necessarily overly simple rationalizing utility function against parsimony/interpretability and represents what might turn out to be an important first step in this broad research program.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

Ian Crawford gratefully acknowledges the support of the Leverhulme Trust (reference F/08 519/E) and the ESRC Centre for Microeconomic Analysis of Public Policy (CPP; reference RES-544-28-5001). Bram De Rock gratefully acknowledges the European Research Council for his Starting Grant.

LITERATURE CITED

Adams A, Cherchye L, De Rock B, Verriest E. 2012. *Commitment in intertemporal household consumption: a revealed preference analysis*. Cent. Econ. Stud. Discuss. Pap. 12.12, KU Leuven

Afriat SN. 1967. The construction of utility functions from expenditure data. *Int. Econ. Rev.* 8:67–77

Afriat SN. 1969. *The construction of separable utility functions from expenditure data*. Tech. Rep., Univ. North Carolina

Afriat SN. 1972. Efficiency estimation of production functions. *Int. Econ. Rev.* 13:568–98

Afriat SN. 1973. On a system of inequalities in demand analysis: an extension of the classical method. *Int. Econ. Rev.* 14:460–72

Andreoni J, Gillen BJ, Harbaugh WT. 2013. *The power of revealed preference tests: ex-post evaluation of experimental design*. Tech. Rep., Univ. Calif., San Diego

Andrews DWK, Guggenberger P. 2009. Validity of subsampling and "plug-in asymptotic" inference for parameters defined by moment inequalities. *Econ. Theory* 25:669–709

Barnett WA, Choi S. 1989. A Monte Carlo study of tests of blockwise weak separability. *J. Bus. Econ. Stat.* 7:363–77

Beatty TKM, Crawford I. 2011. How demanding is the revealed preference approach to demand? *Am. Econ. Rev.* 101:2782–95

Becker GS. 1962. Irrational behavior and economic theory. *J. Polit. Econ.* 70:1–13

Bergstrom T. 1989. A fresh look at the rotten-kid theorem and other household mysteries. *J. Polit. Econ.* 97:1138–59

Bergstrom TC, Cornes RC. 1981. Gorman and Musgrave are dual: an antipodean theorem on public goods. *Econ. Lett.* 7:371–78

Bergstrom TC, Cornes RC. 1983. Independence of allocative efficiency from distribution in the theory of public goods. *Econometrica* 51:1753–65

Blow L, Browning M, Crawford I. 2008. Revealed preference analysis of characteristics models. *Rev. Econ. Stud.* 75:371–89

Blundell R, Browning M, Crawford I. 2003. Nonparametric Engel curves and revealed preference. *Econometrica* 71:205–40

Blundell R, Browning M, Crawford I. 2008. Best nonparametric bounds on demand responses. *Econometrica* 76:1227–62

Bronars SG. 1987. The power of nonparametric tests of preference maximization. *Econometrica* 55:693–98

Brown B, Walker MB. 1989. The random utility hypothesis and inference in demand systems. *Econometrica* 57:815–29

Brown DJ, Calsamiglia C. 2007. The nonparametric approach to applied welfare analysis. *Econ. Theory* 31:183–88

Brown DJ, Matzkin RL. 1996. Testable restrictions on the equilibrium manifold. *Econometrica* 64:1249–62

Browning M. 1989. A nonparametric test of the life-cycle rational expectations hypothesis. *Int. Econ. Rev.* 30:979–92

Carvajal A, González N. 2014. On refutability of the Nash bargaining solution. *J. Math. Econ.* 50:177–86

Carvajal A, Ray I, Snyder S. 2004. Equilibrium behavior in markets and games: testable restrictions and identification. *J. Math. Econ.* 40:1–40

Chambers CP, Echenique F. 2014. On the consistency of data with bargaining theories. *Theor. Econ.* 9:137–62

Chambers CP, Echenique F, Shamaya S. 2014. The axiomatic structure of empirical content. *Am. Econ. Rev.* In press

Cherchye L, Crawford I, De Rock B, Vermeulen F. 2013a. *Gorman revisited: nonparametric conditions for exact linear aggregation*. Cent. Econ. Stud. Discuss. Pap. 13.05, KU Leuven

Cherchye L, De Rock B, Sabbe J, Vermeulen F. 2008. Nonparametric tests of collective rational consumption behavior: an integer programming procedure. *J. Econom.* 147:258–65

Cherchye L, De Rock B, Vermeulen F. 2007. The collective model of household consumption: a nonparametric characterization. *Econometrica* 75:553–74

Cherchye L, De Rock B, Vermeulen F. 2009. Opening the black box of intra-household decision-making. *J. Polit. Econ.* 117:1074–104

Cherchye L, De Rock B, Vermeulen F. 2010. An Afriat theorem for the collective model of household consumption. *J. Econ. Theory* 145:1142–63

Cherchye L, De Rock B, Vermeulen F. 2011a. The revealed preference approach to collective consumption behavior: testing and sharing rule recovery. *Rev. Econ. Stud.* 78:176–98

Cherchye L, Demuynck T, De Rock B. 2014. Is utility transferable? A revealed preference analysis. *Theor. Econ.* In press

Cherchye L, Demuynck T, De Rock B. 2011b. Revealed preference analysis of noncooperative household consumption. *Econ. J.* 121:1073–96

Cherchye L, Demuynck T, De Rock B. 2011c. Testable implications of general equilibrium models: an integer programming approach. *J. Math. Econ.* 47:564–75

Cherchye L, Demuynck T, De Rock B. 2013b. Nash bargained consumption decisions: a revealed preference analysis. *Econ. J.* 123:195–235

Cherchye L, Demuynck T, De Rock B. 2013c. *Revealed preference analysis for convex rationalizations on nonlinear budget sets*. Work. Pap., KU Leuven

Cherchye L, Demuynck T, De Rock B, Hjertstrand P. 2012. *Revealed preference tests for weak separability: an integer programming approach*. Cent. Econ. Stud. Discuss. Pap. 12.15, KU Leuven

Cherchye L, Vermeulen F. 2008. Nonparametric analysis of household labor supply: goodness-of-fit and power of the unitary and the collective model. *Rev. Econ. Stat.* 90:267–74

Chernozhukov V, Hong H, Tamer E. 2007. Estimation and confidence regions for parameter sets in econometric models. *Econometrica* 75:1243–83

Chiappori P. 1988. Rational household labor supply. *Econometrica* 56:63–89

Chiappori P, Rochet J. 1987. Revealed preferences and differentiable demand. *Econometrica* 55:687–91

Crawford I. 2004. *Necessary and sufficient conditions for latent separability*. Tech. Rep. 02/04, Cent. Microdata Methods Practice, London

Crawford I. 2010. Habits revealed. *Rev. Econ. Stud.* 77:1382–402

Crawford I, Pendakur K. 2013. How many types are there? *Econ. J.* 123:77–95

de Finetti B. 1992 (1937). Foresight: its logical laws, its subjective sources. In *Breakthroughs in Statistics*, ed. S Kotz, NL Johnson, pp. 134–74. New York: Wiley

Dean M, Martin D. 2010. *How rational are your choice data?* Tech. Rep., Brown Univ., Providence, RI

Demuynck T, Verriest E. 2013. I'll never forget my first cigarette: a revealed preference analysis of the habits as durables model. *Int. Econ. Rev.* 54:717–38

Diewert WE. 1973. Afriat and revealed preference theory. *Rev. Econ. Stud.* 40:419–25

Diewert WE, Parkan C. 1985. Tests for the consistency of consumer data. *J. Econom.* 30:127–47

Echenique F, Lee S, Shum M. 2011. The money pump as a measure of revealed preference violations. *J. Polit. Econ.* 119:1201–23

Fleissig A, Whitney GA. 2003. A new PC-based test for Varian's weak separability condition. *J. Bus. Econ. Stat.* 21:133–45

Fleissig A, Whitney GA. 2007. Testing additive separability. *Econ. Lett.* 96:215–20

Fleissig A, Whitney GA. 2008. A nonparametric test of weak separability and consumer preferences. *J. Econom.* 147:275–81

Forges F, Minelli E. 2009. Afriat's theorem for general budget sets. *J. Econ. Theory* 144:135–45

Green EJ, Osband K. 1991. A revealed preference theory for expected utility. *Rev. Econ. Stud.* 58:677–95

Gross J. 1995. Testing data for consistency with revealed preference. *Rev. Econ. Stat.* 77:701–10

Halevy Y, Persitz D, Zrill L. 2013. *Parametric recoverability of preferences*. Tech. Rep., Univ. Toronto

Hoderlein S, Stoye J. 2014. Revealed preference in a heterogeneous population. *Rev. Econ. Stat.* In press

Houthakker HS. 1950. Revealed preference and the utility function. *Economica* 17:159–74

Lewbel A. 2001. Demand systems with and without errors. *Am. Econ. Rev.* 91:611–18

Manski C. 2003. *Partial Identification of Probability Distributions*. New York: Springer-Verlag

Matzkin RL. 1991. Axioms of revealed preference for nonlinear choice sets. *Econometrica* 59:1779–86

Polisson M, Quah J. 2013. Revealed preference in a discrete consumption space. *Am. Econ. J. Microecon.* 5:28–34

Samuelson PA. 1938. A note on the pure theory of consumer's behavior. *Economica* 5:61–71

Samuelson PA. 1948. Consumption theory in terms of revealed preference. *Economica* 15:243–53

Selten R. 1991. Properties of a measure of predictive success. *Math. Soc. Sci.* 21:153–67

Selten R, Krischker S. 1982. Comparison of two theories for characteristic function experiments. In *Aspiration Levels in Bargaining and Economic Decision Making*, ed. R Tietz, pp. 259–64. New York: Springer

Swofford JL, Whitney GA. 1987. Nonparametric tests of utility maximization and weak separability for consumption, leisure and money. *Rev. Econ. Stat.* 69:458–64

Swofford JL, Whitney GA. 1988. A comparison of nonparametric tests of weak separability for annual and quarterly data on consumption, leisure, and money. *J. Bus. Econ. Stat.* 6:241–46

Swofford JL, Whitney GA. 1994. A revealed preference test for weakly separable utility maximization with incomplete adjustment. *J. Econom.* 60:235–49

Thaler R, Shefrin HM. 1981. An economic theory of self-control. *J. Polit. Econ.* 89:392–406

Tversky A, Kahneman D. 1991. Loss aversion in riskless choice: a reference-dependent model. *Q. J. Econ.* 106:1039–61

Varian HR. 1982. The nonparametric approach to demand analysis. *Econometrica* 50:945–74

Varian HR. 1983. Non-parametric tests of consumer behavior. *Rev. Econ. Stud.* 50:99–110

Varian HR. 1984. The nonparametric approach to production analysis. *Econometrica* 52:579–97

Varian HR. 1985. Non-parametric analysis of optimizing behavior with measurement error. *J. Econom.* 30:445–58

Varian HR. 1988. Estimating risk aversion from Arrow-Debreu portfolio choice. *Econometrica* 56:973–79

Varian HR. 2006. Revealed preference. In *Samuelson Economics and the Twenty-First Century*, ed. M Szenberg, L Ramrattan, AA Gottesman, pp. 99–115. New York: Oxford Univ. Press

Vermeulen F. 2005. And the winner is . . . an empirical evaluation of unitary and collective labour supply models. *Empir. Econ.* 30:711–34

Warshall S. 1962. A theorem of Boolean matrices. *J. Am. Assoc. Comput. Mach.* 9:11–12

Annual Review of Economics

Volume 6, 2014

Contents

Symposium: The Institutional Underpinnings of Long-Run Income Differences

Indexes

Errata

An online log of corrections to *Annual Review of Economics* articles may be found at http://www.annualreviews.org/errata/economics