American Economic Review 101 (October 2011): 2782–2795
http://www.aeaweb.org/articles.php?doi=10.1257/aer.101.6.2782

How Demanding Is the Revealed Preference Approach to Demand?†

By Timothy K. M. Beatty and Ian A. Crawford*

Revealed preference conditions offer simple, intuitive, and direct means of assessing the empirical implications of a wide range of basic economic models. Indeed, when revealed preference conditions are checked, it is often found that the models perform reasonably well.[1] But is this a triumph for economics, or a warning that revealed preference conditions are so undemanding that almost anything goes? The contribution of this paper is to provide a systematic way in which we might, for the first time, be able to tell.

To illustrate the difficulty, consider the classical two-good consumer choice problem illustrated in Figure 1. It shows two budget constraints where prices are $\mathbf{p}_1 = \{3,4\}'$ and $\mathbf{p}_2 = \{4,3\}'$, and budgets are $x_1 = 10$ and $x_2 = 5$. This environment is one in which there is a modest change in relative prices in conjunction with a large change in income. As a result, regardless of where a nonsatiated consumer's choices fall, revealed preference restrictions on their behavior simply cannot be violated. As Hal R. Varian (1982, p. 966) puts it, "... lack of variation in the price data limits the power of these methods."[2]

This issue is well known, and a number of ways of accounting for it have been suggested.[3] The problem is that existing approaches lack a sound theoretical grounding, and this creates two difficulties. First, there is no basis for choosing

* Beatty: Department of Applied Economics, University of Minnesota, 317E Classroom Office Building, 1994 Buford Ave., St. Paul, MN 55108, and Institute for Fiscal Studies (e-mail: tbeatty@umn.edu); Crawford: Department of Economics, University of Oxford, Manor Road Building, Manor Road, Oxford, OX1 3UQ, and Institute for Fiscal Studies/cemmap (e-mail: ian.crawford@economics.ox.ac.uk). We are very grateful to three anonymous referees for their advice and comments. We are also grateful to Richard Blundell, Martin Browning, Jerry Hausman, Clare Leaver, Peter Neary, and seminar audiences at Brown University, University of Copenhagen, University of Leuven, LSE, University of Oxford, Tulane Univerisity, Queen Mary-University of London, and University of Essex for their comments. We are deeply indebted to John D. Hey who brought Selten's Theorem to our attention. Funding for this paper from the ESRC grant RES-000-22-3770 is gratefully acknowledged.

† To view additional materials, visit the article page at http://www.aeaweb.org/articles.php?doi=10.1257/aer.101.6.2782.

[1] Revealed preference tests have found rational behavior among New York dairy farmers (Loren W. Tauer 1995), Danish consumers (Laura E. Blow, Martin J. Browning, and Crawford 2008), children (William T. Harbaugh, Kate Krause, and Timothy R. Berry 2001), psychiatric patients (Raymond C. Battalio et al. 1973), and capuchin monkeys (M. Keith Chen, Venkat Lakshminarayanan, and Laurie R. Santos 2006).

[2] Note, this is *not* a statement about statistical power. This problem arises in revealed preference analysis conducted with nonrandom variables where the statistical power is, by definition, one. There have been a number of contributions that discuss the statistical power of revealed preference tests on stochastic variables, including Varian (1985), Larry G. Epstein and Adonis J. Yatchew (1985), Stephen G. Bronars (1987), Melissa Famulari (1995), Ana M. Aizcorbe (1991), and Richard W. Blundell, Browning, and Crawford (2008), who build on the work of Donald W. K. Andrews and Patrik Guggenberger (2007). In the future, the Andrews and Guggenberger (2007) approach might be usefully combined with the methods developed here to deal with both the statistical and nonstatistical aspects of rejectability in revealed preference tests.

[3] See James Andreoni and Harbaugh (2008) for a recent discussion of the issue, a review of the various measures that have been proposed, suggestions for a number of novel approaches, and a comparative empirical study of the performance for all of the indices.

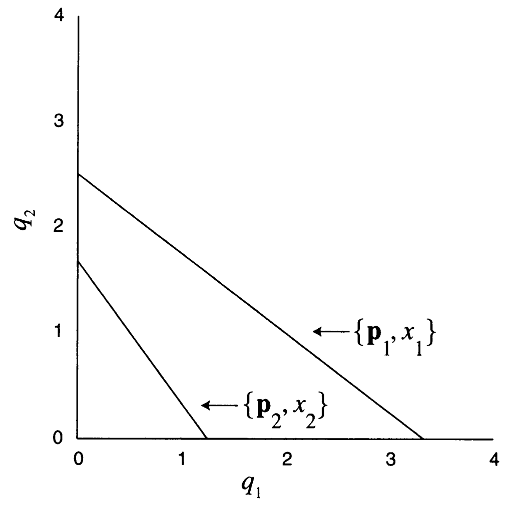

FIGURE 1. A TWO-GOOD, TWO-CHOICE EXAMPLE OF AN INABILITY TO DETECT VIOLATION

among competing proposals, all of which may be plausible. Second, it is unclear how existing methods, which generally rely on the geometric intuition of the weak axiom of revealed preference[4], might extend to other more complex restrictions in the broad revealed preference family.[5]

In the next section, we develop a way to account for the ability (or lack thereof) of revealed preference methods to reject optimizing behavior. Our approach is based on a measure of predictive success proposed by Reinhard Selten and Wilhelm Krischker (1983) and Selten (1991) in the context of experimental game theory. A key feature of the proposed measure is that it has transparent theoretical underpinnings. We show that a set of axioms, which captures some desirable attributes of such a measure, cardinally identifies the proposed measure. Section II briefly discusses how the approach in this paper relates to some of the literature on the power of revealed preference tests. Section III is an empirical illustration showing that this approach is not just theoretically based but is also useful; we show that reporting revealed preference results using our proposed methods is far more informative than the usual approach of simply reporting pass rates.

I. Predictive Success in Revealed Preference Tests

Revealed preference restrictions confine a consumer's observed choices to lie in a specific, well-defined set. To illustrate, consider Figure 2, which shows a two-good,

[4]The weak axiom of revealed preference says that if bundle $\mathbf{q}_j$ is chosen when bundle $\mathbf{q}_i$ was available, and the bundles are distinct, we will never observe $\mathbf{q}_i$ chosen when $\mathbf{q}_j$ is available. The weak axiom involves only direct comparisons between bundles and is a necessary and sufficient condition for utility maximization when demands are single-valued and there are only two goods.

[5]This includes revealed preference–type approaches to profit maximization and cost minimization by perfectly competitive and monopolistic firms (Giora Hanoch and Michael Rothschild 1972); the strong rational expectations hypothesis (Browning 1989); expected utility theory (Zvi Bar-Shira 1992); household sharing models (Laurens Cherchye, Bram De Rock, and Frederic Vermeulen 2007); firm investment behavior (Varian 1983); characteristics models (Blow, Browning, and Crawford 2008); habits (Crawford 2010), and so on.

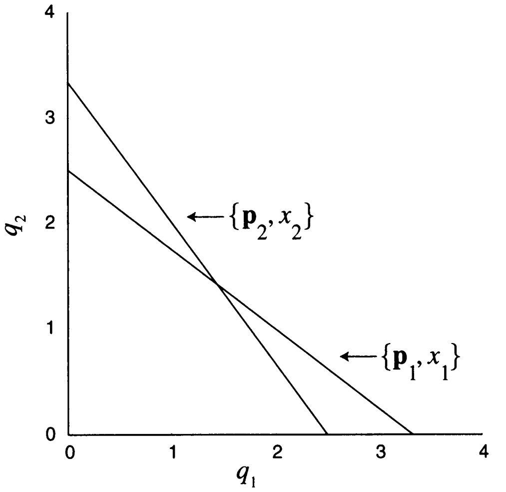

Figure 2. A Two-Good, Two-Choice Example with Predictive Ability

two-choice example, where prices are $\mathbf{p}_1 = \{3,4\}'$ $\mathbf{p}_2 = \{4,3\}'$ and budgets are ten in each period. If a consumer with concave, monotonic, continuous, nonsatiated preferences were to make choices from these two budget sets, then those choices must satisfy the generalized axiom of revealed preference (GARP): $\mathbf{q}_j$ is revealed preferred to $\mathbf{q}_i$, implies that $\mathbf{q}_i$ is not strictly and directly preferred to $\mathbf{q}_j$.[6]

A simple two-dimensional way of representing the restrictions on choices implied by GARP is to illustrate the set of GARP-consistent budget shares for one of the goods ($w_{1,t}$ denotes the budget share of good 1 on budget constraint t) in a unit square—where the budget share of the other good is implied by adding up. This is illustrated in Figure 3, where the shaded area S shows the set of all budget shares for good 1 that are consistent with GARP, and the unit square P is the set of all possible budget shares for this good. For example, the point $(1,0)$ in Figure 3 shows a budget share of 100 percent on good 1 (and so 0 percent on good 2) when the consumer faces the prices $\mathbf{p}_1 = \{3,4\}'$, and a budget share of 0 percent on good 1 (and so 100 percent on good 2) when the consumer faces the prices $\mathbf{p}_2 = \{4,3\}'$. This corresponds to demands $\mathbf{q}_1 = \{3(1/3),0\}'$ and $\mathbf{q}_2 = \{0,3(1/3)\}'$, which satisfy GARP and therefore $(1,0) \in S$.

When we check GARP on observed choices, we are essentially looking to see if the observed shares lie in the predicted/allowed set. A useful analogy is that the set of demands admissible under the theory defines a target for the choice data, and we then check to see if the consumer's choices have hit the target.

Figures 1, 2, and 3 are instructive. They suggest that merely recording the pass rate of revealed preference tests in a consumer panel survey may not, on its own, be a very good guide as to the success or otherwise of the model. To the extent that the constraints imposed by the revealed preference restrictions may represent "unmissable targets," the simple pass rate may be entirely uninformative about

[6] Sydney N. Afriat (1967), Erwin W. Diewert (1973), Varian (1982); $\mathbf{q}_j R \mathbf{q}_i$ implies not $\mathbf{q}_i P^0 \mathbf{q}_j$, where R denotes "is (either directly or indirectly) revealed preferred to" and P^0 denotes "is strictly and directly preferred to."

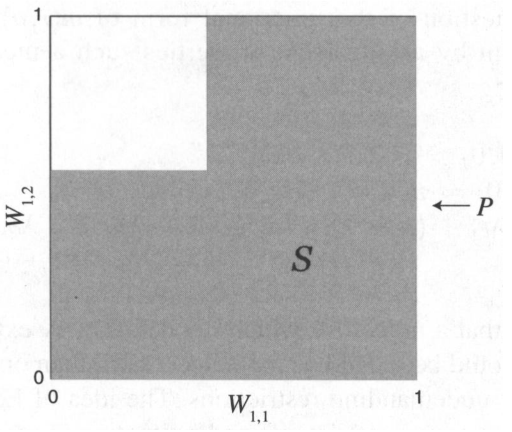

FIGURE 3. THE AREA OF THE GARP RESTRICTIONS IN FIGURE 2

the performance of the model. It would seem to be important to find a way of accounting for this. Figure 3 suggests a possible solution. The size of the set defined by the revealed preference restrictions (S) *relative* to the size of the set of all possible outcomes (P) is a natural measure of the discipline imposed by the restrictions. In Figure 2, the relative size of the predicted set as a proportion of the outcome space is $40/49 \approx 0.816$. In this case, 19.4 percent of possible outcomes are ruled out by the revealed preference restrictions—it is at least possible to miss the target. It therefore seems that we should take the size of the target area as well as the pass/fail indicator into account when evaluating the outcome of a revealed preference test: a model should be counted as more successful in situations in which we observe *both* good pass rates and demanding restrictions.

It is important to appreciate that the relative size of the predicted set of demands depends crucially on the price-budget environment in which the consumer makes choices.[7] As we have seen, the price-budget combination in Figure 2 is such that this relative size is 0.816. By contrast, if we did the same exercise and plotted the revealed preference–consistent budget shares corresponding to Figure 1 (where the prices are the same as those in Figure 2 but the budgets are 10 and 5), the *whole* of the unit square would be shaded. In that case, the relative size of the predicted set is one and the set of outcomes predicted by the theory is also the set of all possible outcomes; the theory rules nothing out, and as a result it is impossible for observed choices to reject the restrictions. As a further example, it is straightforward to show that if we were to keep the budgets the same as in Figure 2 but change the prices to $\mathbf{p}_1 = \{2.5, 5\}'$ and $\mathbf{p}_2 = \{5, 2.5\}'$, the area would be $8/9 \approx 0.889$.

In what follows we denote the pass/fail indicator by $r \in \{0,1\}$ and the relative area of the target $a \in \{0,1\}$ (i.e., the size of S relative to P where the relative area of the empty set is zero and the relative area of the whole outcome space is one). If the measure of success—which we denote $m(r,a)$—should depend on both pass

[7] We are grateful to an anonymous referee for suggesting the following examples.

rate and area, the question of the functional form of $m(r,a)$ remains open. To address this, we begin by asking what properties such a measure should have. Consider the following:

Monotonicity: $m(1,0) > m(0,1)$.
Equivalence: $m(0,0) = m(1,1)$.
Aggregability: $m(\lambda r_1 + (1-\lambda)r_2, \lambda a_1 + (1-\lambda)a_2) = \lambda m(r_1,a_1) + (1-\lambda)m(r_2,a_2)$.

Monotonicity says that a model for which the data satisfy extremely demanding (point) restrictions should be judged as more successful than one in which the data fail to satisfy entirely undemanding restrictions. The idea of Equivalence is that a situation in which there are no restrictions and a situation in which nothing is ruled out are equally (un)informative about the performance of a model. Aggregability says that it is desirable that the measure be additive over heterogeneous consumers. This make it straightforward to calculate a sample average performance measure and to make inferences about the expected value of m in the population. Given these axioms, we have the following result:

SELTEN'S THEOREM: *The function $m = r - a$ satisfies monotonicity, equivalence, and aggregability. If the function $\tilde{m}(r,a)$ also satisfies these axioms, then there exist real numbers $\{\beta, \gamma > 0\}$ such that $\tilde{m}(r,a) = \beta + \gamma m$.*

PROOF:
See Appendix.[8]

Selten's Theorem says that not only does the simple difference measure of pass rate minus area satisfy these axioms, but all measures that satisfy these axioms are positive linear transformations of this difference. The implication is that we might as well use the simple difference.[9] The resulting measure $m \in \{-1,1\}$ can be viewed as a pass/fail indicator, corrected for the ability to find rejections. The interpretation of m is very straightforward. As m approaches one, we know that we have a situation in which the restrictions are extremely demanding, coupled with data that satisfy them: the sign of a quantitatively successful model. As m approaches minus one we know that we have restrictions that allow almost any observed behavior, and yet the data fail to conform: the sign of an almost pathologically unsuccessful model. As m approaches zero we know we have a situation in which the apparent accuracy of the data simply mirrors the size of the target.

To conclude this section, we propose a generalization of the ideas discussed above. Revealed preference methods (somewhat notoriously) give rather hit/miss results; the outcome for an individual consumer is $r = 1$ if they pass and $r = 0$ if they fail. Even though this has the benefit of clarity, it might be argued that it comes

[8] The Theorem is proved in Selten (1991). The proof in this paper is a simpler alternative using standard results on functional equations.

[9] Selten also provides an ordinal characterisation of $m = r - a$ which replaces aggregability with a continuity axiom and an axiom that says that two theories should be compared on the basis of the difference in their respective pass rates and areas.

at the expense of recognizing a qualitative difference between near misses and data that are way off target. A simple way to generalize the binary pass/fail result is to compute the Euclidean distance (d) between the observed data and the target area and use this in place of r. Unfortunately, such a measure is unsuitable for several reasons.[10] A better alternative is to measure the extent of the miss proportionally to the maximum possible distance (denoted $d^{\max}$) between a feasible outcome and the target area (this would be at $(0,1)$ in Figure 3, for example). The new hit rate $r^d = 1 - d/d^{\max}$ lies in the interval zero-one and takes the value one if the data satisfy the revealed preference restrictions, and zero if it misses by the maximum possible amount. This way of measuring hits and misses smooths out a binary result by penalizing close shaves and wild misses differently and, since it lies in the unit interval, the overall measure of predictive success $m^d = r^d - a$ continues to satisfy Selten's Theorem.

II. Connections with the Literature

The relative area is not a probability measure. Nevertheless, it does have all of the necessary properties of a probability.[11] Therefore, if one wished to interpret the relative area as a probability, then one interpretation of $m \approx 0$ is that the theory performs about as well as a uniform random number generator. This interpretation provides a link between the area measure proposed here and the investigation of *statistical* power conducted by Bronars (1987). Statistical power is, of course, a measure of $\Pr(\text{Rejecting } H_0 \,|\, H_0 \text{ is false})$ so the calculation of any statistical power measure requires an alternative hypothesis to be specified. Bronars (1987) adopts Gary S. Becker's (1962) idea of uniform random choices over the outcome space as a general alternative hypothesis to a null of optimizing behavior. The implication is that area may be interpreted as one minus Bronars's (1987) statistical power measure.

A drawback of Bronars's (1987) use of uniform-random choice as the alternative hypothesis is that it treats all bundles as equally likely. Uniform-random choice may be implausible, and, better, more behaviorally relevant alternative hypotheses might place more probability weight on some bundles than others. The specific alternative model one has in mind will dictate precisely what those weights are. The link between Bronars's (1987) statistical power measure and the nonstatistical relative area proposed in this paper shows that the area measure suffers from essentially the same shortcoming.[12] The relative area compares the size of the predicted set to the size of the set of all possible outcomes. There may be better, more behaviorally relevant subsets of the outcome space, however, that might make for more informative comparisons. Again, the specific alternative model one has in mind will dictate precisely which subsets those are.

[10] First, it is unit-dependent and not constrained to lie in the unit interval. Consequently, the resulting measure of predictive success would not satisfy Selten's Theorem. Second, this distance measure will necessarily be inversely related to the area. (If the predicted area almost fills the outcome space, then it will be impossible to miss by much.)

[11] It is nonnegative, the relative area of the whole outcome space is one and the total relative area of two disjoint subsets of the outcome space is the sum of the areas.

[12] We are grateful to a referee for bringing this point to our attention.

The original intent of the ideas developed by Selten (1991) was to find a way of measuring predictive success in experimental game theory. Likewise the area can be thought of as a tool to aid the better design of experiments. For example, in the context of a lab experiment designed to test revealed preference conditions (e.g., Reinhard Sippel 1997; Harbaugh, Krause, and Berry 2001), the area can be used to optimize the design of the experiment by choosing the price-budget environment to minimize the relative area and thus maximize the sensitivity of the test to nonrational behavior. More recently, Blundell, Browning, and Crawford (2003) consider the design of revealed preference tests in the context of observational data when the investigator observes prices and Engel curves. The Engel curves allow the investigator to construct budget expansion paths for demands at the observed prices, and Blundell, Browning, and Crawford (2003) consider the question of how to choose the budget levels at which to evaluate demands and conduct revealed preference tests with the object of maximizing the sensitivity of the test. Their solution—the *sequential maximum power* path—takes an initial price-quantity observation and then sequentially sets the budget for the next choice such that the original choice is exactly affordable, and no more. In this way, they seek sequentially to optimize the test conditional on observed behavior up to that point.

While the approach taken in Blundell, Browning, and Crawford (2003) is quite different in spirit from that taken in this paper, it turns out that it is easy to show in a simple two-good example that their method can be interpreted as minimizing the relative area conditional on the sequential ordering of the path that they choose. This connection also suggests how the ideas developed here could be used to improve their method further by considering alternative ordering of the data aimed at minimizing the area unconditionally.

III. An Illustrative Application

We now turn to a practical application of these ideas. We begin by showing how the proposed measure is useful in interpreting a revealed preference analysis of a heterogeneous sample. We then show how using the smoothed hit rate provides information on the nature of the failures of the theory. In the Appendix we show how our approach can be used to compare alternative models.

We use data from the Spanish Continuous Family Expenditure Survey (the *Encuesta Continua de Presupuestos Familiares* (ECPF)). The ECPF is a quarterly budget survey of Spanish households, which interviews about 3,200 households every quarter. Households are randomly rotated at a rate of 12.5 percent per quarter. It is possible to follow a participating household for up to eight consecutive periods. The data cover the years 1985 to 1997 and the selected subsample are couples with and without children, in which the husband is in full-time employment in a nonagricultural activity and the wife is out of the labor force (this is to minimize the effects of nonseparabilities between consumption demands and leisure for which the empirical application does not otherwise allow). The dataset consists of 21,866 observations on 3,134 households. It records household nondurable expenditures aggregated into five broad commodity groups ("food, alcohol, and tobacco," "energy and services at home," "nondurables," "travel," and "personal services"). The price data are national price indices for the corresponding expenditure categories.

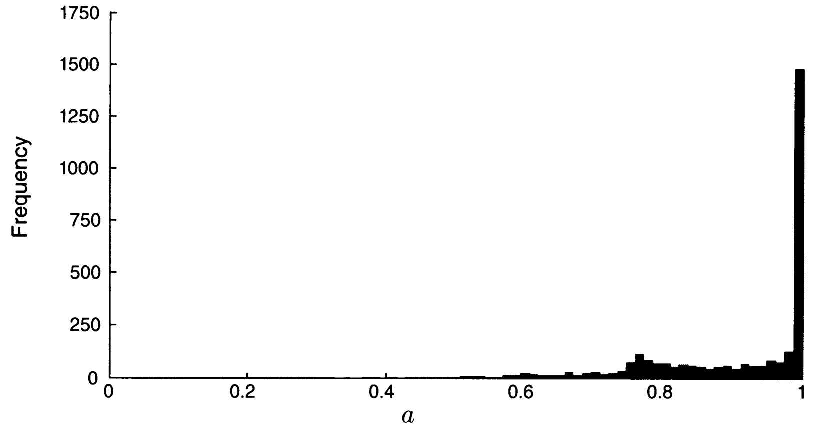

FIGURE 4. FREQUENCY DISTRIBUTION OF THE AREAS

We checked GARP and calculate the area independently for each individual household in our data. The aggregate pass rate for GARP is impressively high, $r = 0.957$. The vast majority of households in the data pass; we can conclude that they behave in a manner consistent with the canonical economic model. Given the preceding discussion, however, we are compelled to ask the question, "How demanding was the test?" We find that the aggregate area is $a = 0.912$. This leads to an aggregate measure of predictive success of $m = 0.045$. The implication is that the standard economic model of utility maximization outperformed a random number generator—but only by 4.5 percent. Given this, the unadjusted pass rate of 95.7 percent seems a great deal less impressive, and even somewhat misleading, regarding the success of the model.

Figure 4 plots the frequency distribution of the household-level areas, and Figure 5 plots the distribution of the household-level measures of predictive success. A key feature of the results highlighted in Figure 4 is that for many households the relative area of the target is equal to one—the theory *cannot* fail. As a consequence, as illustrated in Figure 5, for most of our sample the model has a measure of predictive success equal to zero because the households' observed choices have simply succeeded in hitting an unmissable target. Figure 5 also shows that, while the restrictions of the model provide a modicum of discipline for some households, there are also a small number of households in the left tail that have missed relatively large target areas. The distribution of the individual pass/fail measures r_i (not illustrated) simply has two mass points: $f_r(0) = 0.043$ and $f_r(1) = 0.957$.

To investigate the question of what might drive these results,[13] we looked at how the outcome of the GARP test and size of the relative area were related to household characteristics, the number of times a household is observed, and the amount of price variability in the data. Overall, demographic variables[14] do not appear to

[13] We are very grateful to a referee who suggested this exercise.

[14] In the regression we used the age of the head of household, the age of the spouse, the number and age distribution of children, tenure indicators, and dummies for whether the head of household completed high school and completed university. Details of the regression results are available from the authors.

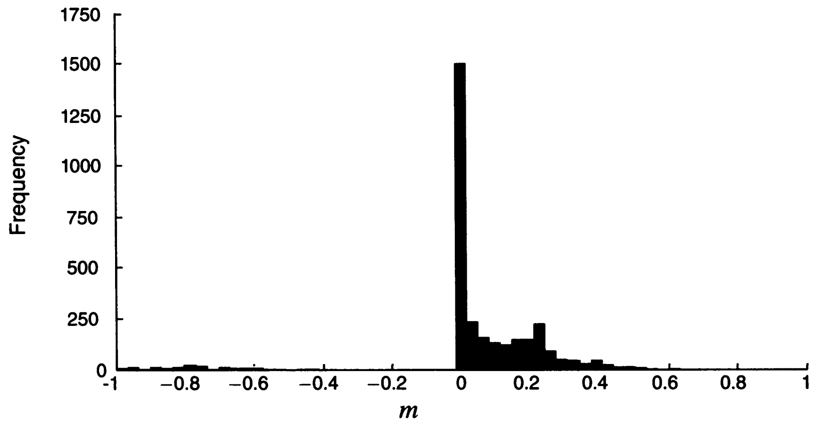

FIGURE 5. FREQUENCY DISTRIBUTION OF PREDICTIVE SUCCESSES

be significant predictors of GARP-consistency nor of the size of the relative area. The number of times we observe a household is significantly and negatively related, however, both to the probability of passing GARP and the relative area. This is entirely as one would expect—more observations make RP tests more demanding. We also find, again as expected, that price variability is important: relative price variability decreases the relative area, whereas absolute price variability increases it. The effects on the probability of satisfying GARP were in line with this, although the effects were statistically insignificant. Finally, the number of commodity groups observed in the household's bundle decreases the probability of passing GARP and also decreases the relative area.

We now generalize the measure of predictive success to distinguish between a near miss and a wild miss. Figure 6 shows the distribution of the modified pass/fail measure for the 133 households in our sample that miss the theoretical target. The distribution is skewed somewhat to the left of its theoretical zero-one range, indicating that most households that fail GARP do so by less than half the extent to which they might, but in general the distances would be hard to describe as being massed close to zero. We might conclude that, in these data, consumers who violate GARP do not do so narrowly. Since this calculation applies only to 4.3 percent of our data (the percentage that failed), the effect of the generalized pass/fail measure on the aggregate performance index is modest: we find that $r^d = 0.97$ compared to $r = 0.957$, and the measure of predicted success is equal to $m^d = 0.058$ compared to $m = 0.045$.

IV. Conclusions

This paper solves two long-standing problems in the revealed preference literature. First, it provides a simple and intuitive approach to accounting for the fact that, sometimes, revealed preference tests just cannot miss. Second, it can be applied to all of the members of the broad family of revealed preference–type methods for which an outcome space can be defined. While we would not defend to the death the particular axioms used in this paper, we would argue that the

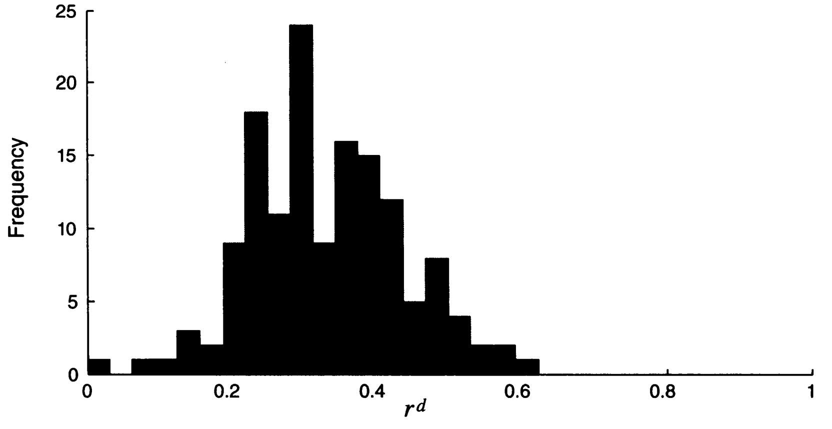

FIGURE 6. THE DISTRIBUTION OF THE MODIFIED HIT RATE

general axiomatic approach based on pass rates and relative area is the right way to make progress on this issue. If these axioms seem unpalatable, then investigators are free to choose others, more to their liking, which may identify another functional form for $m(r,a)$.

Our empirical example demonstrates the potential importance of making these allowances when interpreting the results of revealed preference analyses. In our examination of optimizing behavior, we obtain an unadjusted pass rate of 95.7 percent. At first glance, this seems like a notable validation of a fundamental economic model. But when we account for the quite undemanding nature of the restrictions that theory places on these data, we see that the performance of the model is far less impressive. Put a different way, in our sample, the economic model is revealed to perform about 4.5 percent better than a random number generator. This should reverse our conclusions about the strength of the empirical support for the model. Of course, we are not claiming that these particular results apply more widely than the dataset studied here. But we are claiming that presenting results in this way sheds a great deal more light on the success, or otherwise, of economic theory than does the uncorrected aggregate pass rate, which is uniformly reported in the applied literature. We conclude that the methods developed in this paper provide a more revealing look at revealed preference.

APPENDIX

An Alternative Proof of Selten's Theorem

The aggregability axiom is a Cauchy functional equation which implies that $m(r,a)$ is affine (János D. Azcél 1966) so let $m = \beta_0 + \beta_r r + \beta_a a$. Equivalence then implies that $\beta_0 = \beta_0 + \beta_r + \beta_a$; hence $\beta_r = -\beta_a$. Denote $\beta_r = \beta$ and $\beta_a = -\beta$. Monotonicity then implies that $\beta_0 + \beta > \beta_0 - \beta$; hence $\beta > 0$. Thus, $m = \beta_0 + \beta(r - a)$ where $\beta > 0$. Since all functions that satisfy these axioms share this form, they are all positive affine transformations of each other.

Model Comparison: An Illustrative Example

This Appendix explores the issue of model comparison and considers two extensions of the basic model of consumer choice. These are: utility maximization with optimization error, and utility maximization with measurement error. We ask whether m might provide useful guidance in each case.

Optimization Errors.—A modification of the revealed preference conditions was developed by Afriat (1967, 1972) and Varian (1985, 1990) to allow for optimization errors. This modification introduces a free parameter into the restrictions called the Afriat efficiency parameter (denoted by e), which lies in the interval zero-one.[15] One minus the Afriat efficiency parameter can be interpreted as the proportion of the household's budget that they are allowed to waste through optimization errors. Fixing the Afriat efficiency at one requires perfect cost efficiency and is equivalent to a standard GARP test. Setting it equal to zero allows complete inefficiency, in which case all feasible demand data are consistent with the theory. Values in between one and zero weaken the revealed preference restrictions monotonically.

The Afriat efficiency approach is simple to apply and widely used. The difficulty facing researchers, however, is determining the appropriate level for e.[16] We know that if we set the efficiency parameter low enough, we can always get the data to pass and, in fact, lowering the efficiency parameter just enough to get the data to pass is exactly what is done in much of the literature.[17] But given the preceding discussion, we also know that simply maximizing the pass rate is not the right thing to do if it also increases the area, which is precisely what lowering the Afriat efficiency does. The optimal choice of the efficiency parameter must depend on the balance between pass rate and area.

To investigate the issue, we vary the Afriat efficiency and track the predictive performance of the modified GARP conditions in our data. This is shown in Figure A1, which clearly illustrates the effects of the Afriat efficiency index on the performance of the model. While setting the required efficiency to 0.95 sounds fairly demanding and indeed is sufficient to guarantee that everyone will pass, in fact, doing so enlarges the target area so as to be unmissable. The optimal level for efficiency is much higher (0.995 percent), although it should be noted that even this only raises the performance of the model to $m = 0.051$.

Measurement Errors.—As discussed, the data are composed of expenditures by households on commodity groups collected in the ECPF, and corresponding national price indices published by the *Instituto Nacional de Estadística*. Since the expenditures are recorded in the survey, but the prices are national time series data, it seems highly likely that, if there is measurement error, most of it will be found in the price data. To this end, we consider an extension of the basic model discussed in Varian (1985) which allows for classical, mean zero, measurement errors in log

[15] Briefly, $\mathbf{q}_t R^0(e) \mathbf{q}_s \Leftrightarrow e\mathbf{p}_t'\mathbf{q}_t \geq \mathbf{p}_t'\mathbf{q}_s$ and $R(e)$ denotes the transitive closure of $R^0(e)$. The modified version of GARP is then $\mathbf{q}_i R(e) \mathbf{q}_k \Rightarrow e\mathbf{p}_k'\mathbf{q}_k \leq \mathbf{p}_k'\mathbf{q}_i$.

[16] Varian's (1990) tongue-in-cheek suggestion was $e = 0.95$.

[17] See Andreoni and Harbaugh (2008) and references therein.

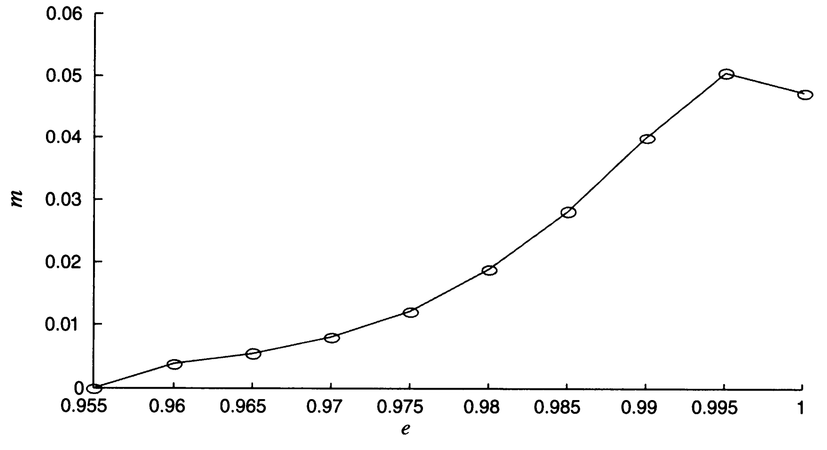

Figure A1. Aggregate Predictive Performance by Afriat Efficiency

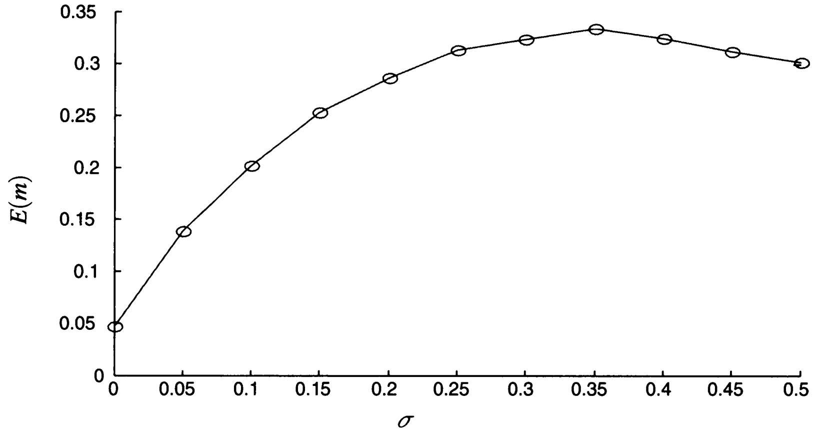

Figure A2. Aggregate Predictive Performance by Measurement Error

prices.[18] The error variance (which for illustrative purposes we assume is common across commodity groups) is of course unknown, so once again it represents a free parameter in the model.

The effects of increasing the error variance, unlike those of the Afriat efficiency parameter in the previous example or the case of attenuation bias in statistical tests, can go either way: households that previously passed (failed) may now fail (pass) once measurement error is allowed for, and the effects on the area could also go in either direction. To analyze the effects on the predictive performance of the theory, we simulate the measurement error by drawing from a multivariate $N(0, \sigma)$ and compute the expected value of m for different values of σ.

Figure A2 shows the relationship between the standard deviation of the measurement error and the expected performance of the modified theory. With $\sigma = 0$ we

[18] We opt for the log specification to avoid the possibility that true prices are ever negative.

have no measurement error and so we have $m = 0.045$ as before. As we gradually increase the measurement error, we see that the performance of the augmented model improves. This is mainly due to the fact that, even though pass rates are dropping over the early part of this range, the area is falling faster as the increased variance of the prices makes budget lines cross to a greater extent. In this context, however, it is not the case that enough measurement allows you to rationalize anything; indeed, there is clear evidence that, with $\sigma \gtrsim 0.35$, the predictive performance of the model begins to fall. It would appear that a model of optimizing behavior subject to $N(0, 0.35^2)$ measurement error in log prices proves the most satisfactory of those considered for these data.

REFERENCES

Aczél, János D. 1966. *Lectures on Functional Equations and Their Applications*. New York: Dover.

Afriat, Sydney, N. 1967. "The Construction of a Utility Function from Expenditure Data." *International Economic Review*, 8(1): 76–77.

Afriat, Sidney N. 1972. "Efficiency Estimation of Production Function." *International Economic Review*, 13(3): 568–98.

Aizcorbe, Ana M. 1991. "A Lower Bound for the Power of Nonparametric Tests." *Journal of Business and Economic Statistics*, 9(4): 463–67.

Andreoni, James and William T. Harbaugh. 2008. "Power Indices for Revealed Preference Tests." Unpublished.

Andrews, Donald W. K., and Patrik Guggenberger. 2007. "Validity of Subsampling and 'Plug–in Asymptotic' Inference for Parameters Defined by Moment Inequalities." Cowles Foundation Discussion Paper 1620

Bar-Shira, Ziv. 1992. "Nonparametric Test of the Expected Utility Hypothesis." *American Journal of Agricultural Economics*, 74(3): 523–33.

Battalio, Raymond C., John H. Kagel, Robin C. Winkler, Edwin B. Fisher, Robert L. Basmann, and Leonard Krasner. 1973. "A Test of Consumer Demand Theory Using Observations of Individual Consumer Purchases." *Western Economic Journal*, 11(4): 411–28.

Beatty, Timothy K. M., and Ian Crawford. 2011. "How Demanding is the Revealed Preference Approach to Demand?: Dataset." *American Economic Review*. http://www.aeaweb.org/articles.php?doi=10.1257/aer.101.6.2782.

Becker, G. S. 1962. "Irrational Behavior and Economic Theory." *Journal of Political Economy*, 70(1): 1–13.

Blow, Laura, Martin Browning, and Ian Crawford. 2008. "Revealed Preference Analysis of Characteristics Models." *Review of Economic Studies*, 75(2): 371–89.

Blundell, Richard W., Martin J. Browning, and Ian A. Crawford. 2003. "Nonparametric Engel Curves and Revealed Preference." *Econometrica*, 71(1): 205–40.

Blundell, Richard W., Martin J. Browning, and Ian A. Crawford. 2008. "Best Nonparametric Bounds on Demand Responses." *Econometrica*, 76(6): 1227–62.

Bronars, Stephen G. 1987. "The Power of Nonparametric Tests of Preference Maximization [the Nonparametric Approach to Demand Analysis]." *Econometrica*, 55(3): 693–98.

Browning, Martin. 1989. "A Nonparametric Test of the Life-Cycle Rational Expectations Hypothesis." *International Economic Review*, 30(4): 979–92.

Chen, M. Keith, Venkat Lakshminarayanan, and Laurie R. Santos. 2006. "How Basic Are Behavioral Biases? Evidence from Capuchin Monkey Trading Behavior." *Journal of Political Economy*, 114(3): 517–37.

Cherchye, Laurens, Bram De Rock, and Frederic Vermeulen. 2007. "The Collective Model of Household Consumption: A Nonparametric Characterization." *Econometrica*, 75(2): 553–74.

Crawford, Ian. 2010. "Habits Revealed." *Review of Economic Studies*, 77(4): 1382–1402.

Diewert, Erwin W. 1973. "Afriat and Revealed Preference Theory." *Review of Economic Studies*, 40(3): 419–25.

Diewert, Erwin W. 1973. "Afriat and Revealed Preference Theory." *Review of Economic Studies*, 40(3): 419–26.

Epstein, Larry G., and Adonis J. Yatchew. 1985. "Non-Parametric Hypothesis Testing Procedures and Applications to Demand Analysis." *Journal of Econometrics*, 30(1–2): 149–69.

Famulari, Melissa. 1995. "A Household-Based, Nonparametric Test of Demand Theory." *Review of Economics and Statistics*, 77(2): 372–82.

Hanoch, Giora, and Michael Rothschild. 1972. "Testing the Assumptions of Production Theory: A Nonparametric Approach." *Journal of Political Economy*, 80(2): 256–75.

Harbaugh, William T., Kate Krause, and Timothy R. Berry. 2001. "Garp for Kids: On the Development of Rational Choice Behavior." *American Economic Review*, 91(5): 1539–45.

Selten, Reinhard. 1991. "Properties of a Measure of Predictive Success." *Mathematical Social Sciences*, 21(2): 153–67.

Selten, Reinhard, and Wilhelm Krischker. 1983. "Comparison of Two Theories for Characteristic Function Experiments." In *Aspiration Levels in Bargaining and Economic Decision Making*, ed. R.Tietz, 259–64. Berlin: Springer.

Sippel, Reinhard. 1997. "An Experiment on the Pure Theory of Consumer's Behaviour." *Economic Journal*, 107(444): 1431–44.

Tauer, Loren W. 1995. "Do New York Dairy Farmers Maximize Profits or Minimize Costs?" *American Journal of Agricultural Economics*, 77(2): 421–29.

Varian, Hal R. 1982. "The Nonparametric Approach to Demand Analysis." *Econometrica*, 50(4): 945–73.

Varian, Hal R. 1983. "Non-Parametric Tests of Consumer Behaviour." *Review of Economic Studies*, 50(1): 99–110.

Varian, Hal R. 1985. "Non-Parametric Analysis of Optimizing Behavior with Measurement Error." *Journal of Econometrics*, 30(1-2): 445–58.

Varian, Hal R. 1990. "Goodness-of-Fit in Optimizing Models." *Journal of Econometrics*, 46(1–2): 125–40.