How Demanding is the Revealed Preference Approach to Demand?

Timothy K.M. Beatty
University of York and IFS. tb526@york.ac.uk

Ian A. Crawford
University of Oxford and IFS. ian.crawford@economics.ox.ac.uk

Abstract

A well known problem with revealed preference methods is that when data are found to satisfy their restrictions it is hard to know whether this should be viewed as a triumph for economic theory, or a warning that these conditions are so undemanding that almost anything goes. This paper allows researchers to make this distinction. Our approach builds on theoretical support in the form of an axiomatic cardinal characterisation of a measure of predictive success due to Selten (1991). We illustrate the idea using a large, nationally representative panel survey of Spanish consumers with broad commodity coverage. The results show that this approach to revealed preference methods can lead us radically to reassess our view of the empirical performance of economic theory.

Acknowledgements: We are very grateful to the Editor and 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 in Brown, Copenhagen, Leuven, LSE, Oxford, Tulane, Queen Mary and Essex for their comments. We are deeply indebted to John D. Hey who brought Selten's Theorem to our attention.

1 Introduction

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 behaviour simply cannot be violated. As Varian (1982) puts it,

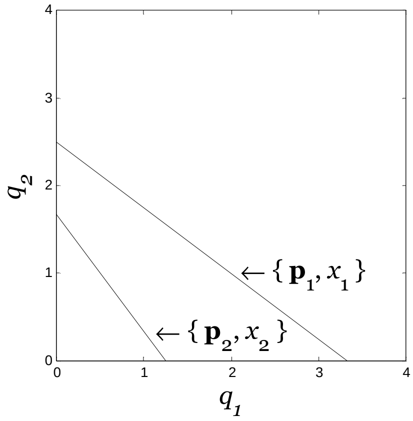
"... lack of variation in the price data limits the power of these methods."
H. Varian, *Econometrica*, 1982, pp 966-7.[2]

[1] Revealed preference tests have found rational behaviour amongst New York dairy farmers (Tauer (1995)), Danish consumers (Blow *et al* (2008)), children (Harbaugh *et al.* (2001)), psychiatric patients (Battalio et al. (1973)), and capuchin monkeys (Chen et al. (2006)).

[2] Note, this is *not* a statement about statistical power. This problem arises in revealed preference analysis conducted with non random variables where the statistical power is, by definition, one. There have been a number of contributions which discuss the statistical power of revealed preference tests on stochastic variables including

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 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]

FIGURE 1: A two good, two choice example of an inability to detect violation.



In the next section, we develop a way to account for the ability (or lack thereof) of revealed preference methods to reject optimising behaviour. Our approach is based on a measure of predictive success proposed by Selten and 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 capture some desirable attributes of such a measure, cardinally identify the proposed measure. Section 3 briefly discusses how the approach in this paper relates to some of the literature on the power of revealed preference tests. Section 4 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.

2 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, two-choice example where prices are $\mathbf{p}_1 = [3,4]'$ $\mathbf{p}_2 = [4,3]'$ and budgets are 10 in each period. If a consumer with concave, monotonic, continuous, non-satiated preferences were to make choices from these two budget sets then those

Varian (1985), Epstein and Yatchew (1985), Bronars (1987), Famulari (1997), Aizcorbe (2001) and Blundell *et al* (2008) who build on the work of Andrews and Guggenburger (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 non-statistical aspects of rejectability in revealed preference tests.

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

[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 maximisation when demands are single-valued and there are only two goods.

[5] This includes revealed preference type approaches to profit maximisation and cost minimisation by perfectly competitive and monopolistic firms (Hanoch and Rothschild, 1972); the strong rational expectations hypothesis (Browning, 1989); expected utility theory (Bar-Shiva, 1992); household sharing models (Cherchye, De Rock and Vermuelen, 2007); firm investment behaviour (Varian, 1983); characteristics models (Blow *et al*, 2008); habits (Crawford, 2009), and so on.

choices must satisfy the Generalised 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]

FIGURE 2: A two good, two choice example with predictive ability.

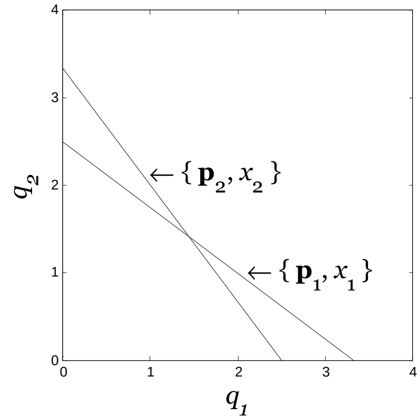
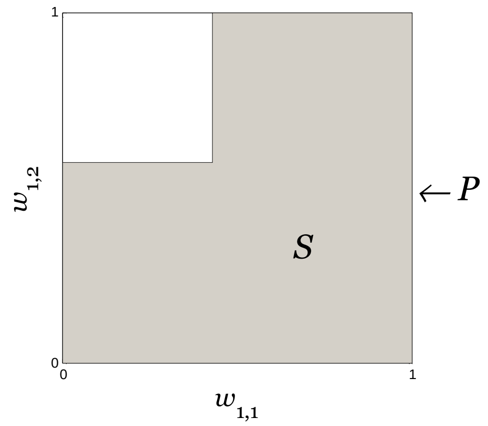


FIGURE 3: The area of the GARP restrictions in Figure 2.



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% on good 1 (and so 0% on good 2) when the consumer faces the prices $\mathbf{p}_1 = [3, 4]'$, and a budget share of 0% on good 1 (and so 100% on good 2) when the consumer faces the prices $\mathbf{p}_2 = [4, 3]'$. This corresponds to demands $\mathbf{q}_1 = \left[3\frac{1}{3}, 0\right]'$ and $\mathbf{q}_2 = \left[0, 3\frac{1}{3}\right]'$ 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

[6] Afriat (1967), Diewert (1972), Varian (1982). $\mathbf{q}_j R \mathbf{q}_i$ implies not $\mathbf{q}_j P^0 \mathbf{q}_i$, where R denotes "is (either directly or indirectly) revealed preferred to" and P^0 denotes "is strictly and directly preferred to".

uninformative about 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% 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 their 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 1 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 Figure 2 but to 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 0 and the relative area of the whole outcome space is 1). If the measure of success - which we denote $m(r, a)$ - should depend on both pass rate and area, the question of the functional form of $m(r, a)$ remains open. To address this, we begin by asking what properties should such a measure 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 fails 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 should 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 $\widetilde{m}(r, a)$ also satisfies these axioms then there exist real numbers $\{\beta, \gamma > 0\}$ such that $\widetilde{m}(r, a) = \beta + \gamma m$.*

Proof: See Appendix A.[8] ■

Selten's Theorem says that not only does the simple difference measure of pass rate minus area satisfy these axioms, but all measures which 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 which satisfy them: the sign of a quantitatively successful model. As m approaches minus one we know that we have

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

[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 which says that two theories should be compared on the basis of the difference in their respective pass rates and areas.

restrictions which allow almost any observed behaviour 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 generalisation 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 at the expense of recognising a qualitative difference between near misses and data that are way off target. A simple way in which to generalise 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 penalising 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.

3 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 $Prob\,(\text{Rejecting } H_0 \mid H_0 \text{ is false})$ so the calculation of any statistical power measure requires an alternative hypothesis to be specified. Bronars' (1987) adopts Becker's (1962) idea of uniform random choices over the outcome space as a general alternative hypothesis to a null of optimising behaviour. The implication is that area may be interpreted as one minus Bronars' (1987) statistical power measure.

A drawback of Bronars' (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 behaviourally 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' (1987) statistical power measure and the non-statistical 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. However, there may be better, more behaviourally relevant sub-sets of the outcome space that might make for more informative comparisons. Again, the specific alternative model one has in mind will dictate precisely which sub-sets those are.

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 aim the better design of experiments. For example, in the context of a lab experiment designed to test revealed preference conditions (e.g. Sippel (1997) and Harbaugh, Krause, and Berry (2001)) the area can be used to optimise the design of the experiment by choosing the price-budget environment to minimise the relative area and thus maximise the sensitivity of the test to non-rational behaviour. More recently Blundell *et al* (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 *et al* (2003) consider the question of how to choose the budget levels at which to evaluate

[10] Firstly 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. Secondly 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.

demands and conduct revealed preference tests with the object of maximising 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 optimise the test conditional on observed behaviour up to that point.

Whilst the approach taken in Blundell *et al* (2003) is quite different in spirit to that taken in this paper, it turns out that it is easy to show in a simple two-good example their method can be interpreted as minimising 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 orderings of the data aimed at minimising the area unconditionally.

4 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 Appendix B, 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% 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 sub-sample are couples with and without children, in which the husband is in full-time employment in a non-agricultural activity and the wife is out of the labour force (this is to minimise 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 non-durable expenditures aggregated into 5 broad commodity groups ("Food, Alcohol and Tobacco","Energy and Services at Home","Non Durables","Travel" and "Personal Services"). The price data are national price indices for the corresponding expenditure categories.

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. However, given the preceding discussion, 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 maximisation out-performed a random number generator - but only by 4.5%. Given this, the unadjusted pass rate of 95.7% 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 1 - 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 who 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 $f_r(1) = 0.957$.

FIGURE 4: Frequency distribution of the areas

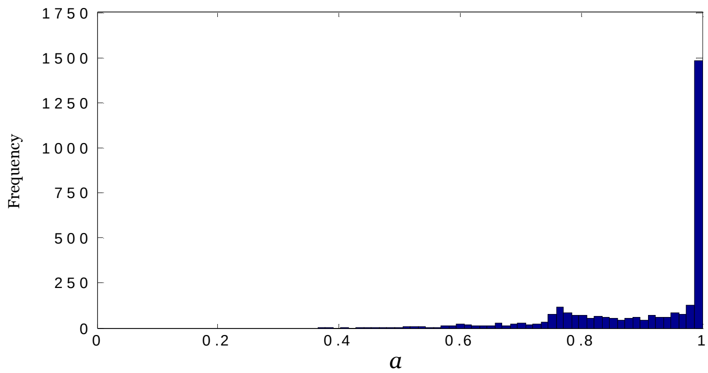
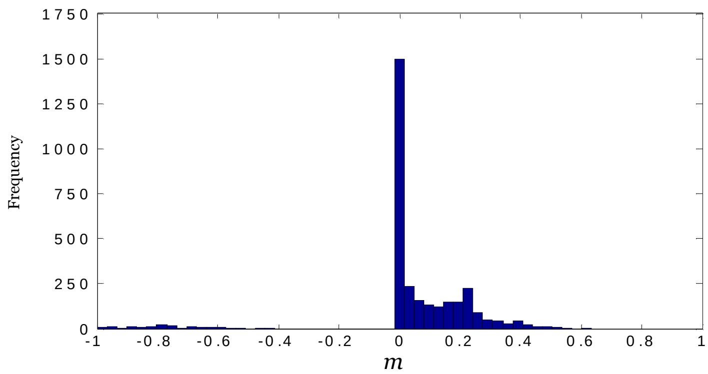


FIGURE 5: Frequency distribution of predictive successes



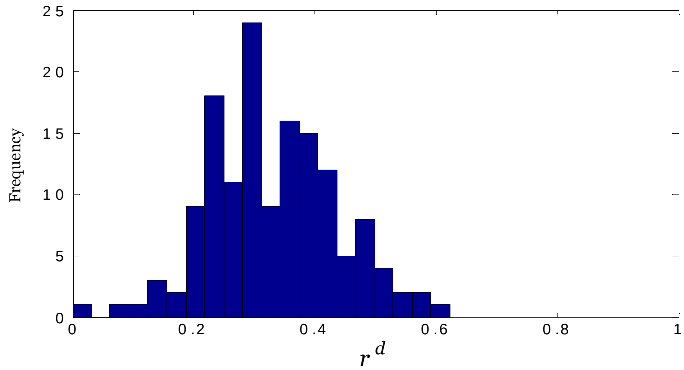
To investigate the question of what might drive these results[13] we looked at how the outcome of the GARP test and size of the relative area were related to household characteristics, the number of times a household is observed and the amount of price variability in the data. Demographic variables[14] do not overall appear to be significant predictors of GARP-consistency nor of the size of the relative area. However, the number of times we observe a household is significantly and negatively related 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 though 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 generalise 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 household 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 who 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 only applies to 4.3% of our data (the percentage that failed) the effect of the generalised 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 $m = 0.045$.

[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 availble form the authors.

Figure 6: The distribution of the modified hit rate



5 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. Whilst we would not defend to the death the particular axioms used in this paper, we would argue that the 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 optimising behaviour, we obtain an unadjusted pass rate of 95.7%. 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 which 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% 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 A - An alternative proof of Selten's Theorem

The aggregability axiom is a Cauchy functional equation which implies that $m(r,a)$ is affine (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 which satisfy these axioms share this form they are all positive affine transformations of each other.■

Appendix B - Model comparison: an illustrative example

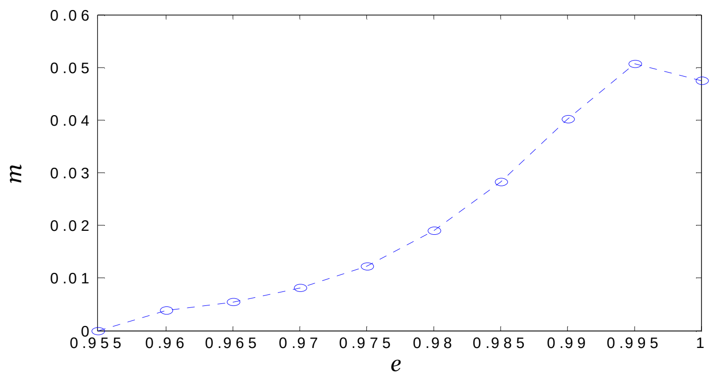
This appendix explores the issue of model comparison and considers two extensions of the basic model of consumer choice. These are: utility maximisation with optimisation error and utility maximisation with measurement error. We ask whether m might provide useful guidance in each case.

B.1 Optimisation Errors

A modification of the revealed preference conditions was developed by Afriat (1967, 1972) and Varian (1985, 1990) to allow for optimisation 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 optimisation 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. However, the difficulty facing researchers 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 maximising 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.

FIGURE 7: Aggregate predictive performance by Afriat efficiency.



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 7, which clearly illustrates the

[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.

effects of the Afriat efficiency index on the performance of the model. Whilst 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%) although it should be noted that even this only raises the performance of the model to $m = 0.051$.

B.2 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 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 which 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 analyse 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 8: Aggregate predictive performance by measurement error.

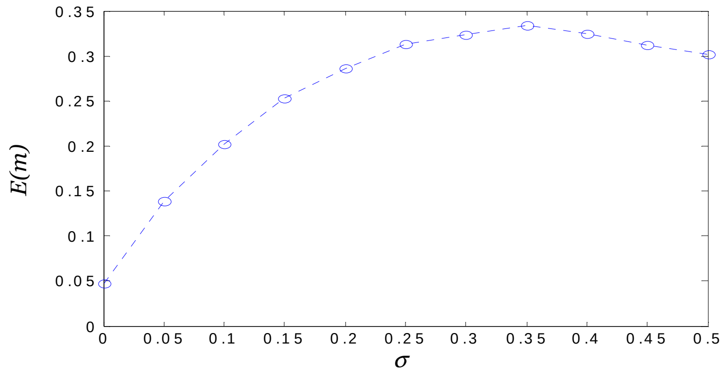


Figure 8 shows the relationship between the standard deviation of the measurement error and the expected performance of the modified theory. With $\sigma = 0$ we have no measurement error and so we have $m = 0.045$ as before. As we gradually increase the measurement error we see that the performance 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. However it is not the case, in this context, that enough measurement allows you to rationalise anything; indeed there is clear evidence that with $\sigma \gtrsim 0.35$ that the predictive performance of the model begins to fall. It would appear that a model of optimising behaviour subject to $N\left(0, 0.35^2\right)$ measurement error in log prices proves the most satisfactory of those considered for this data

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

References

[1] Aczél, J. 1966. *Lectures on Functional Equations and Their Applications*, Dover: New York.

[2] Afriat, S.N. 1967. "The construction of a utility function from expenditure data." *International Economic Review,* 8: 76-77.

[3] Afriat, S.N. 1972. "Efficiency Estimates of Production Functions." *International Economic Review*, 13: 568–598.

[4] Aizcorbe, A. 1991. "A Lower Bound for the Power of Nonparametric Tests." *Journal of Business and Economic Statistics*, 9: 463–467.

[5] Andreoni, J., and W. Harbaugh. 2008. "Power Indices for Revealed Preference Tests." mimeo.

[6] Andrews, D.W.K., and P. Guggenberger. 2007. "Validity of Subsampling and "Plug-In" Asymptotic Inference for Parameters Defined by Moment Inequalities." Cowles Foundation Discussion Paper No. 1620.

[7] Bar-Shira, Z. 1992. "Nonparametric Test of the Expected Utility Hypothesis." *American Journal of Agricultural Economics*, 74(3): 523-533

[8] 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.

[9] Becker, G. S. 1962. "Irrational Behavior and Economic Theory." *Journal of Political Economy*, 70: 1-13.

[10] Blow, L., M. Browning, and I. Crawford. 2008. "Revealed Preference Analysis of Characteristics Models." *Review of Economic Studies*, 75(4):1-19.

[11] Blundell, R., M. Browning, and I. Crawford. 2003. "Nonparametric Engel Curves and Revealed Preference." *Econometrica*, 71(1): 205-240.

[12] Blundell, R., M. Browning, and I. Crawford. 2008. "Best Nonparametric Bounds on on Demand Curves." *Econometrica,* 76(6): 1227–1262.

[13] Bronars, S.G. 1987. "The Power of Nonparametric Tests of Preference Maximization." *Econometrica*, 55(3): 693–698.

[14] Browning, M. 1989. "A nonparametric test of the life-cycle rational expections hypothesis." *International Economic Review*, 30(4): 979-992.

[15] Chen, M.K., V. Lakshminarayanan, and L.R. Santos. 2006. "How basic are behavioral biases? Evidence from Capuchin Monkey Trading Behavior." *Journal of Political Economy*, 114(3): 517-537.

[16] Cherchye, L., B. De Rock, and F. Vermeulen. 2007. "The collective model of household consumption: a nonparametric characterization." *Econometrica*, 75, 553-574.

[17] Crawford, I. 2009. "Habits Revealed." *Review of Economic Studies*, forthcoming.

[18] Diewert, W.E. 1973. "Afriat and revealed preference theory." *Review of Economic Studies*, 40: 419-426.

[19] Epstein, L. G., and A.J. Yatchew. 1985. "Non-parametric hypothesis testing procedures and applications to demand analysis." *Journal of Econometrics*, 30(1-2): 149-169.

[20] Famulari, M. 1995. "A Household-Based, Nonparametric Test of Demand Theory." *Review of Economics and Statistics*, 77: 372–382.

[21] Hanoch, G., and M. Rothschild. 1972. "Testing the Assumptions of Production Theory: A Nonparametric Approach." *Journal of Political Economy*, 80(2): 256-75.

[22] Harbaugh, W.T., K Krause, and T.R. Berry. 2001. "GARP for Kids: On the Development of Rational Choice Behavior." *American Economic Review*, 91(5): 1539-1545.

[23] Selten, R. 1991. "Properties of a measure of predictive success." *Mathematical Social Sciences,* 21:153-167.

[24] Selten R., and S. Krischker. 1983. "Comparison of two theories for characteristic function experiments." In *Aspiration Levels in Bargaining and Economic Decision Making,* ed. R.Tietz, 259-264. Springer.

[25] Sippel, R. 1997, "An Experiment On The Pure Theory Of Consumer's Behaviour," *The Economic Journal*, September 1997, 107(444), 1431-44.

[26] Tauer, L. W. 1995. "Do New York Dairy Farmers Maximize Profits or Minimize Costs?" *American Journal of Agricultural Economics*, 77(2): 421-429.

[27] Varian, H. 1982. "The nonparametric approach to demand analysis." *Econometrica*, 50: 945-973.

[28] Varian, H. 1983. "Nonparametric tests of consumer behaviour." *Review of Economic Studies*, 50: 99-110.

[29] Varian, H. 1984. "The Nonparametric Approach to Production Analysis." *Econometrica,* 52: 579-97.

[30] Varian, H. 1985. "Non-parametric analysis of optimizing behavior with measurement error" *Journal of Econometrics*, 30: 445-458

[31] Varian, H. 1990. "Goodness-of-Fit in Optimizing Models."*Journal of Econometrics*, 46:125–140.