DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

REVEALED PREFERENCE TESTS USING SUPERMARKET DATA: THE MONEY PUMP

Federico Echenique

SangMok Lee

Matthew Shum

SOCIAL SCIENCE WORKING PAPER 1328

June 2010

Revealed Preference Tests using Supermarket Data: the Money Pump

Federico Echenique   SangMok Lee   Matthew Shum

Abstract

We use a money pump argument to measure deviations from the revealed preference axioms. Using a panel data set of food expenditures, we find a large number of violations of the weak axiom of revealed preference. The money pump costs are small, which indicate that the violations of revealed preference are not severe. While most households' behavior deviates from rationality, by our measure they are close to being rational.

JEL classification numbers: D11,D12

Key words: Revealed preference; GARP; Money pump; Demand theory.

Revealed Preference Tests using Supermarket Data: the Money Pump

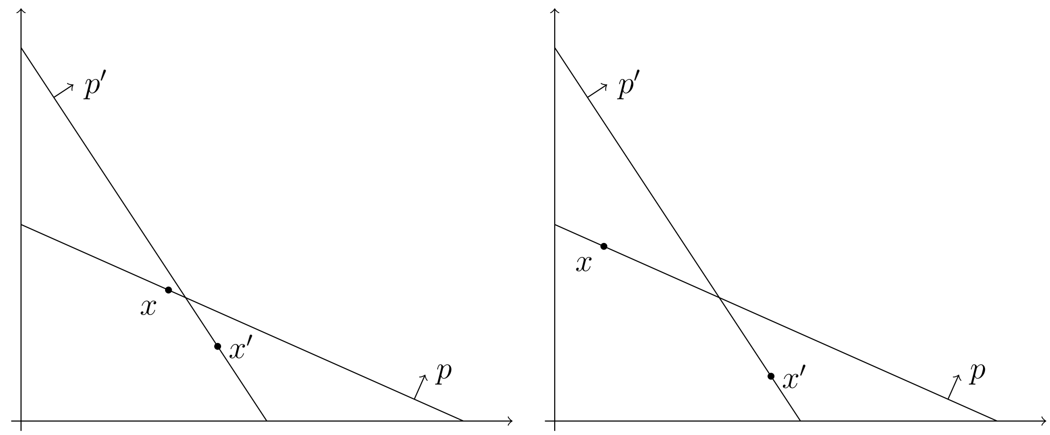
Federico Echenique SangMok Lee Matthew Shum

1 Introduction

The assumption that consumers are rational is one of the oldest and most controversial assumptions in economics. Conceptually, the empirical content of the rationality assumption has been very well understood since the works of Samuelson (1938), Afriat (1967), Richter (1966) and Varian (1982): revealed preference theory captures the empirical content of rational consumption behavior.

As a *practical* matter, however, revealed preference analysis presents two serious problems. One is that a data set either satisfies the generalized axiom of revealed preference (GARP) or it does not; there is no room for judging how severe a violation of the axioms are. We would like to measure the *extent* to which a violation of GARP indicates that a consumer is irrational. The second problem is that GARP very often lacks power as a test of rationality. Consumption data tend to exhibit less variability in prices than in expenditure. As a consequence, it becomes very difficult to reject that consumers are rational.

Our paper presents a new approach to practical revealed preference analysis. We present a measure of the severity of a violation of revealed preference, and we use a data set that seems likely to alleviate the problem of power. Our analysis revealed a substantial number of violations of GARP, but the violations are not severe. Specifically, 396 out of the 494 households in our data set violate GARP at some point. When they violate GARP, our money pump measure (which we define below) of the violation is around 6% of expenditures; which we view as a small number.



(a) (x, p) and (x', p') violate GARP (in fact WARP).

(b) A more "severe" violation of GARP.

Figure 1: Two observations: (x, p) and (x', p').

We correlate our measure with demographic variables. Some results are intuitive: Less educated, poorer, and older households make more severe violations of GARP than do highly educated, richer, and younger households. On the other hand, households with small families make more severe violations of GARP.

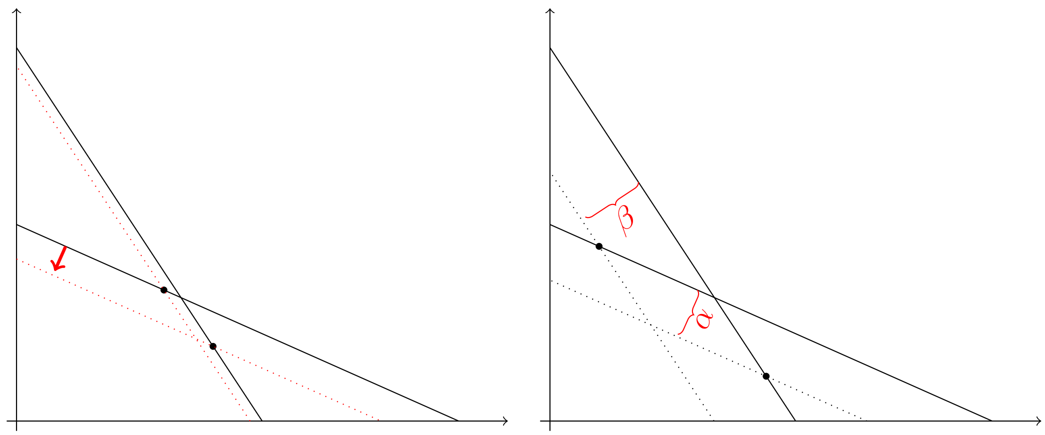
1.1 Money pump

A violation of GARP exposes a consumer to being manipulated as a "money pump." For example, consider the situation in Figure 1(a). A consumer buys bundle x at prices p and x' at prices p'. Evidently, there is a violation of GARP (actually of WARP, the weak axiom of revealed preference) because x was purchased when x' was affordable, and vice versa. Knowing these choices, a devious "arbitrager" who follows the opposite purchasing strategy (buy bundle x at prices p', and bundle x' at prices p), could profitably resell x to the consumer at prices p, and x' at prices p'. The total profit the arbitrager would make equals

$$mp = p \cdot (x - x') + p' \cdot (x' - x),$$

where mp stands for "money pump cost," which we use to measure the severity of the violation of GARP.

The idea that arbitragers can "pump money" from irrational consumers is not new,



(a) Two observations: (x, p) and (x', p'). (b) $mp = \alpha+\beta; \alpha = p\cdot(x-x'), \beta = p'\cdot(x'-x)$.

Figure 2: Money pump costs for Figure 1.

and it has been used as a reason for why one should not observe irrational behavior. Therefore, the money pump measure defined above appears to be an intuitive measure of the severity of a violation of GARP. Consider the situations in Figures 1(a) and 1(b). Each figure presents a violation of GARP, but intuitively the violation in 1(b) is more severe than the one in 1(a). The money pump cost reflects this difference. Figures 2(a) and 2(b) represent the money pump cost: it is the sum of the translation of the p-budget line (from crossing x to crossing x'), and the translation of the p'-budget line (from crossing x' to crossing x). The money pump reflects the severity of the violations, and it is expressed in monetary terms, so the numerical value of a violation has a clear interpretation.

For our purposes, the devious arbitrager is a fictional character. There is a debate on whether irrational consumers would be driven out of the market, because of the actions of arbitragers; see, for example, Mulligan (1996), Rabin (2002) and Laibson and Yariv (2007). We do not take a stand on this debate: Our use of money pump cost is solely a pragmatic application of the idea developed in Figure 1, showing how the pump cost captures intuitively the severity of a violation of GARP.

1.2 Power of GARP

That GARP tends to have low power is well known. The vast majority of empirical studies of revealed preference find very few violations of GARP. The basic issue is the following: In cross-sectional data, income variability is much higher than price variability. As a result, budget sets tend to be nested, and no choices can violate GARP.

To address the problem of low power, we consider a subset of goods, food purchases; and we consider household-level panel data, where the same household is observed making repeated choices over time. We restrict attention to food because we do not expect consumers to change their food expenditure very dramatically in response to changes in income; we should see relatively more variation in the price rather than levels of expenditure. Most food is a basic necessity, and the role for "luxurious" spending on food is much more limited than for other types of good. We use a scanner panel dataset of 494 families' consumption purchases of frequently-purchased gropcery items over a two-year period (see Section 4.1 for details). By using panel data, we can abstract away from household-level unobserved heterogeneity, which may be driving the excessive variability in expenditure levels observed in typical cross-sectional datasets, and which result in low power of GARP tests. Indeed, we find many more violations of GARP than the rest of the empirical literature, which suggests that power may not be a concern in our panel data.

2 Related literature

The literature on testing the revealed preference axioms is large, and contains both classical papers as well as more recent contributions. Afriat (1967) and Varian (1982) are seminal contributions to the methodology of revealed preferences tests; Varian (2006) provides a survey. Empirical applications of revealed preference tests have employed both field as well as experimental data.

Many of the empirical applications using field data employ data from household-level surveys (such as the Consumer Expenditure Survey in the US, and the Family Expenditure Survey in the UK). Since tests of WARP/GARP require repeated observations of a decision-making unit (individual or household) across different pricing regimes, thus well-suited to panel data, an important challenge addressed in these papers is how to "match" households across different time periods to form a synthetic panel. Blundell, Browning, and Crawford (2003) and Blundell, Chen, and Kristensen (2007) address this issue by estimating an "Engel curve" relating a household's consumption to prices, expenditure and household demographics, and test GARP by comparing the predicted consumption behavior of households with similar demographics and expenditure levels across different pricing regimes. Hoderlein and Stoye (2009) take a more agnostic approach, and use results from the copula literature to obtain bounds on the percentage of households which violate WARP in two separate cross-sections of survey data. In the present paper, we abstract away from these difficulties by using a long household-level scanner panel dataset, where the purchase decisions of given households over a two-year period are observed. To our knowledge, empirical tests of the revealed preference axioms using scanner data is new in the literature.

At the same time, a large literature testing revealed preference using experimental data has also developed. These have employed both laboratory experiments (recent contributions include Andreoni and Miller (2002), Sippel (1997) and Fevrier and Visser (2004)), as well as field experiments utilizing unique subject pools (psychiatric patients in Battalio, Kagel, Winkler, Fischer, Basmann, and Krasner (1973), and children in Harbaugh, Krause, and Berry (2001)).

It is fair to say that most of the empirical literature, using both field and (field and laboratory) experimental data, finds relatively few violations of the GARP. Therefore, the power of GARP as a test of rationality is a real concern; these issues have been discussed in, *inter alia*, Bronars (1987), Blundell, Browning, and Crawford (2003), Andreoni and Harbaugh (2008).

At the same time, revealed preference tests are quite stark, allowing for either rational or irrational consumers. In practice, one would like to accommodate a grey area where "small" violations of GARP may not indicate a worrying degree of irrationality (or may indicate imperfections in the data). In the existing literature, various researchers have proposed ways to quantify the degree of violations from GARP, including Afriat (1967), Varian (1985, 1990), and Gross (1995).[1] In terms of assessing the severity of violations of GARP, our proposal is closest to *Afriat's efficiency index,* and to the modification of Afriat's efficiency index proposed by Varian (see Afriat (1967) and Varian (1990)).

Given $e \in [0,1]$, let R_e and P_e be the binary relations defined by $x^k \, R_e \, x^l$ if $ep^k \cdot x^k \geq p^k \cdot x^l$, and $x^k \, P_e \, x^l$ if $ep^k \cdot x^k > p^k \cdot x^l$. Clearly, if $e = 1$, then R_e is the original revealed

[1] Apesteguia and Ballester (2010) axiomatize a measure of deviations from rationality. It applies in general choice environments with finitely many choices. It does not use the special structure of Walrasian budgets.

preference relation, so if R_e satisfies GARP then the data are consistent with rationality. At the other extreme, when $e = 0$ then R_e satisfies GARP trivially. Afriat's efficiency index is defined as the supremum over all the numbers e such that (R_e, P_e) satisfies GARP.

Varian modifies Afriat's index by allowing e to vary across the different price vectors. Consider a vector $e = (e_k)_{k=1}^K$ of numbers in $[0, 1]$, one for each observation. Define the binary relation R as $x^k \, R \, x^l$ if $e_k p^k \cdot x^k \geq p^k \cdot x^l$. Define the strict relation P analogously. There is a set of vectors e such that the corresponding R satisfies GARP. Varian proposes as his measure the closest vector e to the unit vector ($e_k = 1$), among those vectors for which R satisfies GARP.

Here, we can interpret e as the amount of measurement error in prices which is required in order to rationalize the observed consumption choices via GARP. That is, how much larger than $p^k \cdot x^l$ does $p^k \cdot x^k$ need to be before we can conclude that x^k is revealed preferred to x^l? If each observation is measured with error, for example, then we can conclude that x^k is revealed preferred to x^l while accommodating an error of $(1 - e_k)$.

Both Afriat's and Varian's efficiency indices capture ideas that are similar to our money pump measure. The measures differ in their interpretation. The efficiency indices reflect a tolerance to measurement error. We can interpret our measure directly from its monetary value.

3 Definitions

Suppose that we observe the purchases of a single consumer when she faces different prices. Observation k ($k = 1, \ldots, K$) consists of a consumption bundle $x^k \in \mathbb{R}^l_+$ that the consumer bought at prices $p^k \in \mathbb{R}^l_{++}$.

Let X be the set of all observed consumption bundles. That is, $X = \left\{x^k : k = 1, \ldots, K\right\}$. The revealed preference relation on X is the binary relation R defined as $x^k \, R \, x^l$ if $p^k \cdot x^k \geq p^k \cdot x^l$. The strict revealed preference relation is the binary relation P defined as $x^k \, P \, x^l$ if $p^k \cdot x^k > p^k \cdot x^l$.

The data satisfy the *weak axiom of revealed preference* (WARP) if whenever $x^k \, R \, x^l$ it is false that $x^l \, P \, x^k$.

The data satisfy the *generalized axiom of revealed preference* (GARP) if there is no sequence $x^{k_1}, x^{k_2}, \ldots, x^{k_n}$ such that

$$x^{k_1} \, R \, x^{k_2} R, \ldots, R x^{k_n} \text{ while } x^{k_n} \, P \, x^{k_1}. \tag{1}$$

A violation of GARP is identified with a sequence $x^{k_1}, x^{k_2}, \ldots, x^{k_n}$. We say that n is the *length* of the sequence.

Given a sequence $x^{k_1}, x^{k_2}, \ldots, x^{k_n}$ for which (1) holds, we can compute the *money pump cost* associated to this sequence as

$$\sum_{l=1}^{n} p^{k_l} \cdot (x^{k_l} - x^{k_{l+1}}),$$

where we interpret $n+1$ as 1.

Remark 1. *Testing for GARP, and calculating money pump costs, can be a huge computational task. For the data we present in Section 4, $K = 26$; so there are*

$$\sum_{k=2}^{26} \binom{26}{k} (k-1)! \;\approx\; 4.39239 \times 10^{25}$$

potential cycles, which are unique up to rotations.

One can check for violations of GARP involving sequences of limited length. There is, unfortunately, loss of generality in doing so (see Rose (1958) and Shafer (1977)). In general, even if every subset of $K-1$ observations out of K price-consumption data satisfies GARP, the entire K observations may violate GARP.

Our money pump cost is measured in dollars. We normalize the cost to make it comparable with today's dollars, and to compare across consumers with different budgets. Specifically, we present money pump cost as the proportion of total expenditure. If (1) holds for the sequence $x^{k_1}, x^{k_2}, \ldots, x^{k_n}$, we compute the *relative money pump cost* of the sequence as

$$\frac{\sum_{l=1}^{n} p^{k_l} \cdot (x^{k_l} - x^{k_{l+1}})}{\sum_{l=1}^{n} p^{k_l} \cdot x^{k_l}}, \tag{2}$$

where we interpret $n+1$ as 1.

4 Main results: incidence and severity of GARP violations

4.1 Data Description

In this paper, we use the so-called "Stanford Basket Dataset", which is a household-level scanner panel dataset, which contains grocery expenditure data for 494 households from four grocery stores in an urban area of a large US midwestern city, between June 1991 and June 1993 (104 weeks). This dataset was collected by Information Resources, Inc. (IRI), and has also been used in, among others, Bell and Lattin (1998), Shum (2004), and Hendel and Nevo (2006b,a). We focus in this paper on households' expenditures on food categories, of which there are fourteen: bacon, barbecue, butter, cereal, coffee, cracker, eggs, ice-cream, nuts, analgesics, pizza, snack, and sugar.[2]

We observe 103,345 transactions of 4,082 unique items: i.e unique Universal Product Codes (UPC). Each transaction records the consumer (household) identity, store identification number, UPC, transaction week, consumed units, price per unit (cent), and the item's relative scale.[3]

In order to obtain consistent consumption data over goods and time, we aggregate transactions by brand name and category: when distinct items have a common brand name, their transactions are aggregated. Hence, each "product" in the sample is a food product with a distinct brand name, and we aggregate across all sizes/presentational forms of each product. Analogously, aggregate prices at the product level are obtained by averaging the prices of each size, weighted by the amount consumed. To minimize stockpiling and inventory issues, we also aggregate households' expenditures for each good over time, to a four-week period.[4]

[2]We proceed the section including analgesics as food. Indeed, when we process data as the subsequent paragraphs, amongst 375 only 4 items are categorized as analgesics, and including them or not only changes the empirical results marginally.

[3]For each category, the relative scale is the weight of an item compared to the standard weight. 16oz is the standard weight of a coffee item, so the relative scale of a coffee product of 24oz is 1.5.

[4]By focusing on food expenditures, our approach requires an assumption that food items are separable in households' preferences, so that purchases of non-foods affect food consumption only through the income left over from such purchases. Hence, our test is implicitly a joint test of rationality and separability for food. However, separability is ubiquitous as an assumption in applied demand analysis, and has been universally assumed in applied work to reduce the dimensionality of demand system (a

Maximum Cycle Length	2	3	4
Total Numb. Households	494	494	494
Households Violating GARP	395	396	396
Median Money Pump	5.97%	5.95%	5.91%
Mean Money Pump	6.22%	6.12%	6.09%
Possible Cycles	325	5525	95225
Median Numb. Violations	2	3	3
Mean Numb. Violations	2.421	3.874	4.815

Table 1: Money Pump: calculated by Equation 2, averaged over households violating GARP

Even after this aggregation, not all brands are consumed for every time period; some are newly launched, taken off the market, or simply not popular. Since GARP requires price observations over every time period, we use only brands for which price data are available for every time period. For this reason, we drop 12,976 (or 12.5 %) of the purchasesfrom the dataset.

4.2 GARP and Money Pump Cost

Table 1 presents a summary of our results. Out of 494 households, 395 (roughly 80%) of them violate WARP (GARP for sequences of length 2) for at least some pairs of observations. Hence, a significant proportion of households do exhibit violations of WARP, in contrast to much of the previous empirical literature, which fail to find many violations. Given Remark 1, we only check for violations of GARP that involve cycles of limited length: lengths 2, 3 and 4. In Table 1, each column corresponds to the maximum length of the cycles in the test for GARP. When we include cycles of length 3 and 4, thereby searching a substantially larger number of possible cycles (5,525 and 95,225, compared to 325), the overall number of households violating GARP increases only by 1. As we mentioned in Section 3, in theory GARP may be violated when WARP is satisfied. However, Table 1 shows that WARP closely approximates GARP in practice. Only one household satisfies WARP while violating GARP with a sequence of length 3. Moreover, the median and mean level of money pump costs change marginally as we search over longer cycles.

point emphasized by Deaton and Muellbauer (1980) and Blundell (1988), among others).

On the other hand, the severity of the violations, in terms of money pump cost, is moderate or small: the mean and median money pump costs, taken across all households, are only about 6% of total expenditure.[5]

To break this down further, we calculated, for each household, the relative money pump cost (see Equation 2) of each violation of WARP, and obtained the *household-specific* median and mean level of the cost, across all the cycles for this household, which violated WARP. In Figure 4.2, we plot the cumulative distribution function of this household-specific median money pump cost, across the 395 households which exhibit some violation of WARP. Clearly, this function rises very steeply for values of the money pump $< 10\%$, but is largely flat thereafter. This indicates that a large majority of households have very small violations of WARP, and only a few handfuls of households have larger violations, exceeding 10% of expenditures. Thus, large violations do occur, but they are infrequent.

4.3 Demographic Variables

We study the demographic determinants of rational (or irrational) consumption behavior. We consider the following demographic dummy variables:

1. Family Size: Middle and Large. A mid-sized household has 3 or 4 members, and a large household has more than 4 members. Small families are the baseline.

2. Income: Middle and High. A household with a middle income earns more than 20,000 and less than 45,000; a household with a high income earns more than 45,000. Low income is the baseline.

3. Age: Middle and Old. The age variable reflects the average of the spouses' ages. In cases of either a missing husband or wife, the age of a household is the surviving member's age. [6]

[5]Without including analgesics, 400 households violate WARP, and their average median money pump cost is also about 6%.

[6]The original data contains the age levels of each household's husband and wife: (1) 18 to 29, (2) 30 to 34, (3) 35 to 44, (4) 45 to 54, (5) 55 to 64, and (6) 65 to 99. A household with an average age index between 4 and 6 is categorized as middle aged, and a household with the average equals to 6 is categorized as old.

Variables		Households
Family Size	Mid Size	187
	Large Size	65
Income	Mid Income	200
	High Income	141
Age	Mid Age	201
	Old Age	157
Education	High School	197
	College	255
Total Households		480

Table 2: Demographic Variables

4. Education: High school and College. The household's education level is the average level the spouses' education. Depending on the average level, we categorize households as high school graduates, college graduates, and others. In cases of a missing husband or wife, the education level reflects the education of the sole head of the household. [7]

Table 4.3 shows the population distributions of the demographic variables. The demographic data is missing for 14 households. We drop these from our data set, and work with 480 households. The panelists are generally older, and their education levels are higher than the general U.S. population.

Since the money pump cost has a positive value only when consumptions violate WARP, we consider censored Tobit regressions of the money pump costs on demographic variables. Table 4.3 shows the regression results with 156,000 (=480 $\times$ 325) observations: 480 households with $\binom{26}{2}$ possible cycles.

[7]The original data contains the education levels for each household's husband and wife: (1) some grade school, (2) completed grade school, (3) some high school, (4) completed high school, (5) some college, (6) completed college, (7) post graduate school, and (8) technical school. We take each couple's averaged education level. A household with an average between 3 to 5 is categorized as high school leveled education, and a household with an average above 5 is categorized as college leveled education.

MidAge			0.0114*
			(1.97)
OldAge			0.0118
			(1.63)
MidFamily	-0.0161***	-0.0112*	-0.00811
	(-3.55)	(-2.36)	(-1.59)
LargeFamily	-0.0281***	-0.0239***	-0.0188*
	(-4.04)	(-3.39)	(-2.49)
MidIncome		-0.0153**	-0.0146*
		(-2.67)	(-2.43)
HighIncome		-0.0186***	-0.0178**
		(-3.40)	(-2.93)
highsch	-0.0154	-0.00856	-0.00771
	(-1.84)	(-1.00)	(-0.89)
college	-0.0162	-0.00589	-0.00397
	(-1.95)	(-0.67)	(-0.44)
_cons	-0.452***	-0.450***	-0.463***
	(-30.63)	(-30.58)	(-27.55)
sigma			
_cons	0.195***	0.195***	0.195***
	(36.65)	(36.66)	(36.66)
N	156000	156000	156000

t statistics in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 3: Money pump violations of GARP, explained by demographics.

The money pump cost of a violation is higher for older, poorer and less educated households, than for younger, richer and more educated households. These correlations are intuitive and easy to explain. The money pump cost is also higher for smaller households, which is somewhat puzzling. The regressors are highly correlated, so some of the specifications in Table 4.3 show the variables not being significant.

5 Further results

5.1 Seasonality and Stability of preferences.

A consumer may fail GARP because his preferences change: they are not "stable." Given two observations, (x, p) and (x', p'), it is possible that x was a rational choice for a different utility function than x'. We argue that unstable preferences would be reflected in large money pump costs. Therefore, our empirical findings support the hypothesis that preferences are stable.

Consider a consumer who uses one utility for some purchases, and another utility for other purchases. We argue that the money pump cost is positive, for arbitrarily small changes in prices. In fact, the money pump cost is larger when the difference in demands under both utilities is larger, thus implying that when preferences are unstable, the money pump cost can be interpreted as a measure of this instability.

Specifically, consider Figure 4(a). Suppose that a household follows two distinct utility functions. These two utility functions give rise to two different demand functions: $d_1(p, I)$ and $d_2(p, I)$. Fix prices p, and suppose that, at p, we observe $x = d_1(p, I)$; see Figure 4(a). The second utility, on the other hand, would give demand $\hat{x} = d_2(p, I)$. Now, by continuity of demand, if we choose prices p' close to p (as in the figure) then $x' = d_2(p', I)$ is close to $\hat{x}$. But this implies a violation of WARP.

The money pump of the violation of WARP in Figure 4(a) is $(p' - p) \cdot (x' - x)$. We can look at the money pump cost for an arbitrarily small change in prices. In particular, fix a direction of change in price ∇p, and consider an infinitesimal price change in p in the direction of ∇. So $p' = p + \varepsilon \nabla p$; for $\varepsilon > 0$. As ε shrinks to 0, x' converges to $\hat{x}$, so the money pump cost approaches $(\hat{x} - x) \cdot \varepsilon \nabla p$; see Figure 4(b). So a small price change gives an increase in pump cost, as long as the change in prices forms an acute angle with the difference in the demand functions. Note also that a larger difference in demands results in a larger pump cost, for a given direction of change of prices.

Given this interpretation of the money pump cost as a measure of an agent's changes in preferences, we next look and see whether the money pump cost reflects seasonal trends in demand for certain types of grocery items, because these trends may be attributable to changes in preferences over time. Specifically, we focus on the case of ice cream demand, for which the seasonal peak in demand is during the summer months, and the seasonal trough is during the winter months. If this seasonality is in fact due to changing preferences, then we should expect to see larger money pump costs in cycles involving peak and non-peak periods, than in cycles involving only non-peak periods.

	Spring	Summer	Fall	Winter
Spring	1.3333	0.6944	1.0556	0.9375
	(1.90%)	(1.84%)	(2.00%)	(2.16%)
Summer	·	1.2000	0.8333	1.2292
		(2.13%)	(1.65%)	(1.52%)
Fall	·	·	1.0667	0.9583
			(1.82%)	(1.78%)
Winter	·	·	·	1.6429
				(1.87%)

Table 4: No evidence of changing preferences: Ice-cream vs. Other foods.

Such evidence is presented in Table 4. We aggregate data up to 'Ice-creams' or 'all other foods', and for every 4-week (1-period). For each pair of periods, we count the number of households violating WARP, and compute the average of their median money pump costs. Numbers (or parenthesized numbers) in the table are the numbers of households (or average levels of their median money pump costs), which are averaged over the pairs of periods falling into a corresponding pair of seasons.[8]

Surprisingly, we find no evidence of seasonality. For instance, the money pump costs are 1.52% between summer and winter months (a peak/non-peak comparison), versus 2.16% between the winter and spring months (two non-peak periods). This suggests that, while seasonality may indeed be present, prices may also be moving in a fashion such that agents' resulting consumption choices do not violate rationality.[9]

[8] In this table, the data is aggregated further than the previous section. There must be no confusion between 395 households violating WARP in the previous section versus on average 1.x households in Table 4.

[9] Indeed, Chevalier, Kashyap, and Rossi (2003) provide evidence prices on grocery items tend to be lower during peak demand periods for these items (see also Nevo and Hatzitaskos (2005) for further discussion). Such "countercyclical" price variation may mask any seasonal variation in preferences, and lead to no violations of revealed preference.

5.2 Bronars index.

It is customary, in applied revealed preference analysis, to compute an index of the power of the test, the Bronars (1987) index. We are in a somewhat different situation compared to most studies, because we find a large number of violations of GARP. So the power of GARP is not a concern for us, probably because of the nature of our data set, as we outline in the introduction.

Nevertheless, we computed the Bronars index. We find, surprisingly, that the Bronars index indicates low power. Specifically, the Bronars index consists of measuring the number of violations of GARP if behavior on the observed budget sets were purely random. Using budget lines computed from the real data, we generate 100 samples of consumption data sets where each household choose a consumption bundle purely at random on the budget line. We find that amongst 494 there are on average 3-4 households violating GARP for each generated panel data set, while the actual choice data show 396 households violating GARP.

We attribute the phenomenon to a basic flaw in the Bronars index. Purely random behavior is not so irrational: a point made originally by Becker (1962). The irrationality of our consumers is due to some kind of systematic tendency or bias; it is not due to purely random behavior. On the other hand, their irrationality is not severe enough to cause high levels of money pump cost.

6 Conclusion

We present a new measure of the severity of a violation of GARP, and an application to scanner data. We find that the vast majority of the households in our data set violate GARP at some point, but that the median violation is usually rather mild. Our findings contrast with the extant empirical literature, which tends to find very few violations of GARP.

The money pump measure is intuitive and easy to interpret. It rests on ideas similar to Afriat's and Varian's efficiency indices, but it is grounded in an economic "story," and a specific value for the index has a direct interpretation, as a measure of either the severity of a GARP violation, or the instability of household preferences.

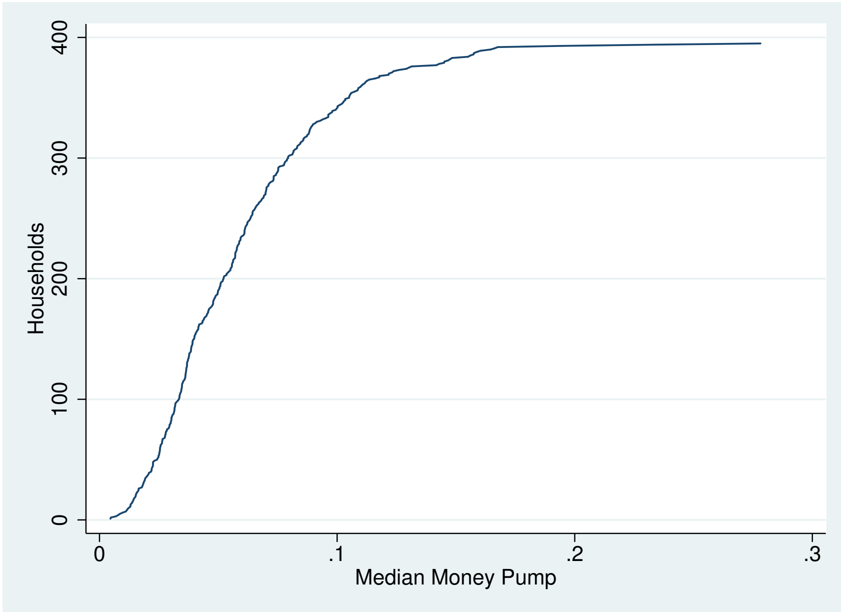
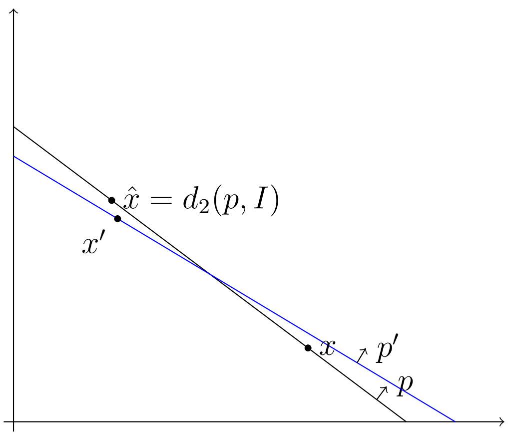
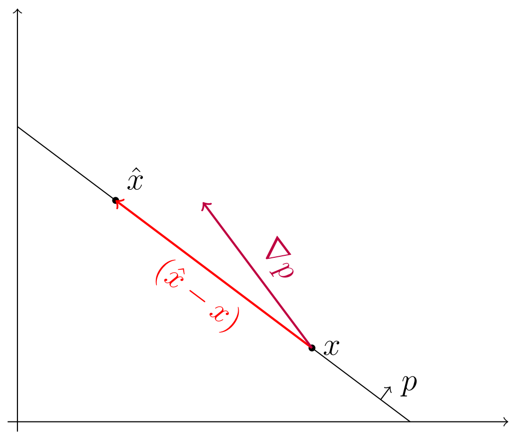


Figure 3: Cumulative distribution of households' median money pump costs



(a) $x = d_1(p, I)$, and $x' = d_2(p', I)$.



(b) $mp = \nabla p \cdot (d_2(p, I) - d_2(p, I))$.

Figure 4: Unstable preferences.

Appendix A An algorithm to calculate the money pump cost

As we remarked in Section 3, calculating the money pump is computationally very heavy. Here we present an algorithm that will approximate the money pump cost, and runs in polynomial time. We did not need to use the approximation in our paper, because the pump costs from violations of WARP capture almost all the pump cost; we include the algorithm because it may be useful in applications to other data sets.

The revealed preference graph is the graph that has the consumption vectors x^k as vertexes, and a (directed) edge pointing from x^k to x^l if $p^k \cdot x^k \geq p^k x^l$.

The adjacency matrix of the graph has K columns and one row for each edge. The edge pointing from x^k to x^l corresponds to the row $e_l - e_k$. We denote by e_k the vector in $\mathbb{R}^K$ with all zeroes except for a 1 in position k.

A basic fact from graph theory (see Theorem 14.5 in Berge (2001)) says that a graph can be decomposed into cycles if and only if the sum of the rows of its adjacency matrix is 0 (the null vector in $\mathbb{R}^K$).

Note that $c(x_1, \ldots x_K)$ is the following sum

$$\sum_C \sum_{e \in C} p^{e_o} \cdot (x^{e_o} - x^{e_d}),$$

where C is the set of cycles of the revealed preference graph, e is a generic edge of a cycle, and e_o is the origin and e_d the destination of edge e.

Consider the following algorithm.

1. Initialize a $1 \times K$ matrix A to be identically zero. Initialize M to 0.
2. Repeat the following for each $k, l = 1, \ldots, K$ with $k \neq l$:

   If $p^l \cdot x^k \leq p^k \cdot x^k$ then let M be

   $$M + p^k \cdot (x^k - x^l);$$

   and add a row $e_l - e_k$ to A.
3. Output A and M.

Note:

1. We run step (2) $K \times (K-1)$ times, so this algorithm is polynomial.
2. A is the adjacency matrix of the revealed-preference graph.
3. If the revealed preference graph can be decomposed into cycles, then the rows of A will add to zero (to the K dimensional null vector).
4. The algorithm adds up the pump cost of each edge in the revealed preference graph. If the rows of A add to zero, then M will be the correct money-pump measure. In general, M exceeds the money pump measure.

Let v be the vector that is the sum of all the rows of A. So $v = \mathbf{1} \cdot A$. We can now use v to correct any excess edges $p^k \cdot (x^k - x^l)$ we should not have added to M because they are not part of a cycle.

Let I be the set of all possible starting (negative) vertexes in v and O the set of all destiny (positive) vertexes in v, with repetitions. That is: Let I be the (multi) set that has $|-v_k|$ copies of k, and O be the (multi) set that has $|v_l|$ copies of l. So the I is the set of sources of an edge and O is the set of destinies, with repetitions.

The "corrected" money pump measure is:

$$M - \sum_{k \in I, l \in O} \left| p^k \cdot (x^k - x^l) \right|$$

In principle, the problem is that v "forgets" the edges in A. It only records who was the origin of a non-cycle edge, and who was a destiny (and how many times). For example, if we have $K = 4$ and

$$v = (1, -1, 1, -1)$$

we don't know if we should correct for $(1, -1, 0, 0)$ and $(0, 0, 1, -1)$, i.e. subtract $p^2 \cdot x^1 + p^4 \cdot x^3$; or if we should correct for $(1, 0, 0, -1)$ and $(0, -1, 1, 0)$, i.e. subtract $p^4 \cdot x^1 + p^2 \cdot x^3$

The corrected measure may still not be right. For example, if $v = (1, -1, 1, -1)$ we may have that both $(1, -1, 0, 0)$ and $(0, -1, 1, 0)$ are edges, but the latter is part of a cycle. Then the corrected measure would have subtracted something it should not.

However, if this happens we would know that something is wrong because we used x^2 as the source of an edge *twice* in the correction.

Our sense is that the algorithm will provide a good approximation to the true pump cost, but an experimental verification is probably needed before one uses the algorithm.

References

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

Andreoni, J., and W. Harbaugh (2008): "Power indices for revealed preference tests," *University of Wisconsin-Madison Department od Economics Working Paper*, 10.

Andreoni, J., and J. Miller (2002): "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism," *Econometrica*, 70(2), 737–753.

Apesteguia, J., and M. A. Ballester (2010): "A Measure of Rationality and Welfare," Mimeo, Universitat Pompeu Fabra.

Battalio, R., J. Kagel, W. Winkler, E. Fischer, R. Basmann, and L. Krasner (1973): "A Test of Consumer Demand Theory using Observations of Individal Consumer Purchases," *Western Economic Journal*, 11(4), 411–428.

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

Bell, D., and J. Lattin (1998): "Shopping Behavior and Consumer Preference for Store Price Format: Why 'Large Basket' Shoppers Prefer EDLP," *Marketing Science*, 17, 68–88.

Berge, C. (2001): *Theory of Graphs*. Dover.

Blundell, R. (1988): "Consumer Behaviour: Theory and Empirical Evidence–A Survey," *The Economic Journal*, 98(389), 16–65.

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

Blundell, R., X. Chen, and D. Kristensen (2007): "Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves," *Econometrica*, 75(6), 1613–1669.

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

CHEVALIER, J. A., A. K. KASHYAP, AND P. E. ROSSI (2003): "Why Don't Prices Rise During Periods of Peak Demand? Evidence from Scanner Data," *American Economic Review*, 93(1), 15–37.

DEATON, A. S., AND J. MUELLBAUER (1980): *Economics and Consumer Behavior*. Cambridge University Press, Cambridge, UK.

FEVRIER, P., AND M. VISSER (2004): "A Study of Consumer Behavior using Laboratory Data," *Experimental Economics*, 7, 93–114.

GROSS, J. (1995): "Testing data for consistency with revealed preference," *The Review of Economics and Statistics*, 77(4), 701–710.

HARBAUGH, W., K. KRAUSE, AND T. BERRY (2001): "GARP for kids: On the development of rational choice behavior," *American Economic Review*, 91(5), 1539–1545.

HENDEL, I., AND A. NEVO (2006a): "Measuring the Implications of Sales and Consumer Stockpiling Behavior," *Econometrica*, 74, 1637–1674.

——— (2006b): "Sales and Consumer Inventory," *RAND Journal of Economics*, 37, 543–561.

HODERLEIN, S., AND J. STOYE (2009): "Revealed Preferences in a Heterogeneous Population," Mimeo, Dep. of Economics, Brown University.

LAIBSON, D., AND L. YARIV (2007): "Safety in Markets: An Impossibility Theorem for Dutch Books," Mimeo, California Institute of Technology.

MULLIGAN, C. (1996): "A Logical Economist's Argument Against Hyperbolic Discounting," University of Chicago, mimeo.

NEVO, A., AND K. HATZITASKOS (2005): "Why Does the Average Price of Tuna Fall During Lent?," NBER Working Papers 11572, National Bureau of Economic Research, Inc.

RABIN, M. (2002): "A perspective on psychology and economics* 1," *European Economic Review*, 46(4-5), 657–685.

RICHTER, M. K. (1966): "Revealed Preference Theory," *Econometrica*, 34(3), 635–645.

Rose, H. (1958): "Consistency of Preference: The Two-Commodity Case," *The Review of Economic Studies*, 25(2), 124–125.

Samuelson, P. (1938): "A note on the pure theory of consumer's behaviour," *Economica*, 5(17), 61–71.

Shafer, W. (1977): "Revealed preference cycles and the Slutsky matrix," *Journal of Economic Theory*, 16(2), 293–309.

Shum, M. (2004): "Does Advertising Overcome Brand Loyalty? Evidence from Breakfast Cereals," *Journal of Economics and Management Strategy*, 13, 241–272.

Sippel, R. (1997): "An Experiment on the Pure Theory of Consumer's Behavior," *Economic Journal*, 107, 1431–1444.

Varian, H. (2006): "Revealed preference," *Samuelsonian economics and the twenty-first century*, pp. 99–115.

Varian, H. R. (1982): "The Nonparametric Approach to Demand Analysis," *Econometrica*, 50(4), 945–974.

——— (1985): "Non-Parametric Analysis of Optimizing Behavior with Measurement Error," *Journal of Econometrics*, 30, 445–458.

——— (1990): "Goodness-of-fit in optimizing models," *Journal of Econometrics*, 46(1-2), 125 – 140.