THE MONEY PUMP AS A MEASURE OF REVEALED PREFERENCE VIOLATIONS
(WITH AN APPLICATION TO SUPERMARKET DATA)

FEDERICO ECHENIQUE, SANGMOK LEE, AND MATTHEW SHUM

. We introduce a measure of the severity of violations of the revealed preference axioms, which we call the *money pump index (MPI)*. The measure is based on the idea that a consumer who violates the axioms is subject to being exploited as a money pump. The MPI has a simple interpretation as the certain dollar amount that can be extracted from a consumer who behaves irrationally.

We carry out an empirical application, using a panel data set of food expenditures. We find a large number of violations of the revealed preference axioms. On average, however, the MPI calculated for these violations is small, suggesting that the violations of revealed preference are not severe. Formal statistical testing indicates that the hypothesis of consumer rationality cannot be rejected.

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 is problematic due to the "all or nothing" nature of the exercise: a data set either satisfies the generalized axiom of revealed preference (GARP) or it does not. In practice, however, it is useful to gauge how *severely* consumers violate the axiom. Our paper presents a new measure of the severity of a violation of GARP. The measure is based on the idea that a consumer who violates GARP is subject to being exploited as a "money pump." We propose that the severity of a violation be measured by the amount

---

Division of the Humanities and Social Sciences, California Institute of Technology, Pasadena CA 91125. Emails: {fede,sangmok,mshum}@hss.caltech.edu. We thank the editor and two anonymous referees for useful comments.

of money that could be extracted from the consumer; we call this the ***money pump index (MPI)***.

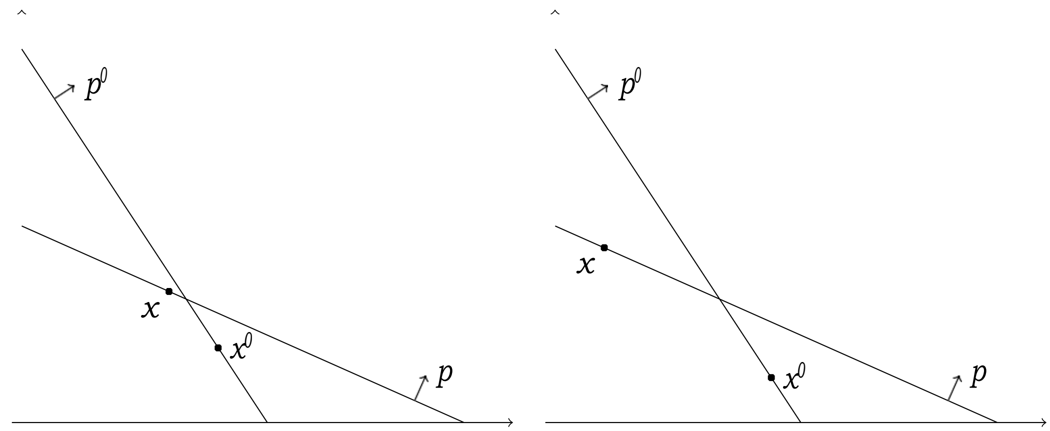
We present an empirical application to household-level "scanner" panel data containing time-series of household-level food grocery purchases collected at checkout scanners in supermarkets. In contrast to many earlier studies, using more aggregate, or cross-sectional consumption data, our analysis revealed a substantial number of violations of GARP. Specifically, 396 out of the 494 households in our data set violate GARP at some point. However, most of these violations are not severe: our MPI is centered around 6% of a household's food expenditures, or about $12.80 when evaluated at the average monthly food expenditure of $213.

The magnitudes of the MPI appear intuitively very small (in both dollar and percentage terms). We proceed to formally use the MPI to test the hypothesis of consumer rationality. We test whether the MPI could be accounted for simply by measurement errors in the variables. In our empirical application, we are unable to reject the null hypothesis that the observed MPIs are consistent with rational behavior and measurement errors. In other words, the apparently small 6% MPI is also small in a statistical sense.

We correlate our measure with demographic variables. Most 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, smaller households make more severe violations of GARP. Moreover, because the demand for many food grocery items reflect seasonal trends, we also check whether GARP violations are more severe when comparing consumption between peak vs. non-peak seasonal periods. We find no evidence of this, implying that consumption in our data can be modeled by stable preferences which exhibit no seasonal component.

Money pump. Our measure of the severity of a GARP violation is motivated by the idea that 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^0 at prices p^0. Evidently, there is a violation of GARP (actually of WARP, the weak axiom of revealed preference) because x was purchased when x^0 was affordable, and vice versa. Knowing these choices, a devious "arbitrager" who follows the opposite purchasing strategy (buying bundle x at prices p^0, and bundle x^0 at prices p), could profitably resell x to the consumer at prices p,



(a) (x, p) and (x^0, p^0) violate GARP (in fact WARP).

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

Figure 1. Two observations: (x, p) and (x^0, p^0).

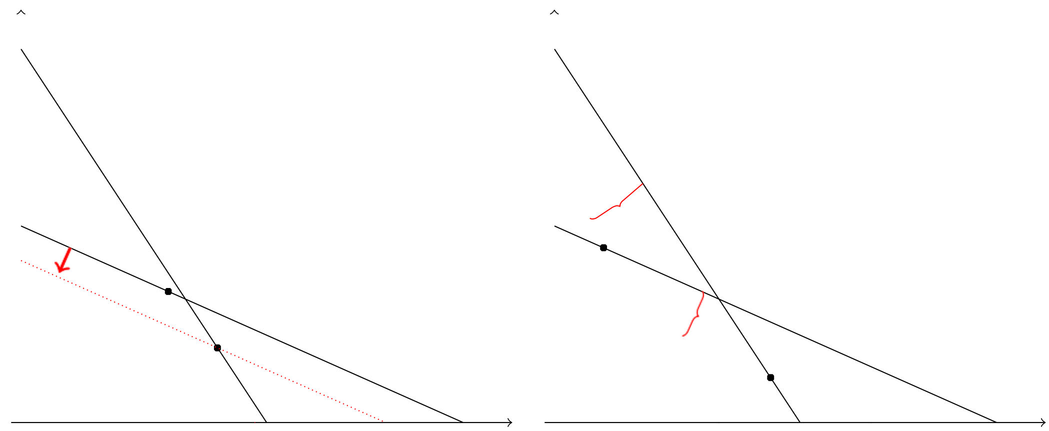
and x^0 at prices p^0. The total profit the arbitrager would make equals

$$mp = p \cdot (x - x^0) + p^0 \cdot (x^0 - x).$$

We use the magnitude mp, "money pump cost," to measure the severity of the violation of GARP. Specifically, our MPI is the money pump cost expressed as a percentage of expenditure.

The MPI is 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^0), and the translation of the p^0-budget line (from crossing x^0 to crossing x). The money pump represents the severity of the violations, and it is expressed in monetary terms, so the numerical value of a violation has a clear interpretation.

The idea that arbitragers can "pump money" from irrational consumers is not new, and it has been used as a reason for why one should not observe irrational behavior. For our purposes, however, the devious arbitrager is a fictional character. There is a debate on whether GARP violators are "irrational consumers" who 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



(a) Two observations: (x, p) and (x^0, p^0). (b) $mp = \alpha + \beta$; $\alpha = p \cdot (x - x^0), \beta = p^0 \cdot (x^0 - x)$.

Figure 2. Money pump costs for Figure 1.

stand on these issues: our use of the money pump is purely an application of the idea captured by Figure 1. The money pump cost represents intuitively the severity of a violation of GARP.

Panel data. To the best of our knowledge, ours is the first (field) study of GARP using a panel of household purchases. Many of the recent studies of GARP employ repeated cross-sectional data; tests of GARP implemented on such data require some sort of aggregation or "matching" of similar households across different cross- sections. In contrast, our panel data allows us to study how household purchases vary with prices over time, without the need to aggregate or "match" consumers. By focusing on supermarket purchases, we also observe a higher frequency of price changes relative to expenditure, compared to standard cross-sectional consumption data sets. As is well-known (see e.g. Blundell, Browning, and Crawford (2003a)), a large variability in expenditure relative to prices can result in GARP having low power.

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 preference tests; Varian (2006) provides a survey. Empirical applications of revealed preference tests have employed both field as well as experimental data.

In principle, tests of WARP/GARP require repeated observations of a decision-making unit (individual or household) across different pricing regimes. However, many of the empirical investigations of GARP using field data employ data from cross-sectional household-level surveys (such as the Consumer Expenditure Survey in the US, and the Family Expenditure Survey in the UK). Thus, 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 (2003b) and Blundell, Chen, and Kristensen (2007) address this issue by estimating an "En- gel 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 avoid 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, testing 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), children in Harbaugh, Krause, and Berry (2001), tufted capuchin monkeys in Chen, Lakshminarayanan, and Santos (2006)).

It is fair to say that most of the empirical literature, using both field and experimental data, finds relatively few violations of 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 (2003b), Andreoni and Harbaugh (2008), Beatty and Crawford (2010). Experiments suffer less from this problem because they are often carefully designed to avoid power issues (see e.g. Andreoni and Miller (2002)).

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, MPI is closest in spirit to the *efficiency index* originally proposed by Afriat (1967) and subsequently modified by Varian (1990). We review these developments in Section 3.2 below.

3. Methodology

3.1. **Money Pump Index.** 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 $x^k \in \mathbf{R}^l_+$ that the consumer bought at prices $p^k \in \mathbf{R}^l_{++}$.

Let X be the set of all observed consumption bundles. That is, $X = \{x^k : k = 1, \ldots, K\}$. 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

$$\sum_{l=1}^{n} p^{k_l} \cdot (x^{k_l} - x^{k_{l+1}}), \quad (\text{taking } k_{n+1} = k_1).$$

Our money pump cost is measured in dollars. In order to compare this cost across consumers with different budgets, we normalize the cost by each household's total expenditure. Specifically, the money pump index MPI equals the money pump cost

[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.

as a proportion of total expenditure: if (1) holds for the sequence $x^{k_1}, x^{k_2}, \ldots, x^{k_n}$, we compute the *MPI* of the sequence as

$$(2) \qquad MPI_{\langle (x^{k_1},p^{k_1}),\ldots,(x^{k_n},p^{k_n})\rangle} = \frac{\sum_{l=1}^{n} p^{k_l} \cdot (x^{k_l} - x^{k_{l+1}})}{\sum_{l=1}^{n} p^{k_l} \cdot x^{k_l}}, \quad (\text{taking } k_{n+1} = k_1).$$

Remark 1. *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. There are fast algorithms for checking if GARP has been violated (see Varian (1982)), but they do not suffice to calculate MPI.[2]

3.2. Comparison with Afriat's efficiency index.

We briefly review and compare our approach to the "efficiency indices" proposed by Afriat (1967) and Varian (1990) to quantify violations from GARP. 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 preference relation, so if R_1 satisfies GARP then the data are consistent with rationality. At the other extreme, R_0 satisfies GARP trivially. *Afriat's efficiency index (AEI)* is defined as the supremum over all the numbers e such that (R_e, P_e) satisfies GARP.

The ideas behind AEI are similar to our MPI (perhaps unavoidably so, as they try to measure the same phenomenon); but AEI and MPI differ in their interpretations. MPI is the monetary magnitude that can be extracted from a consumer that violates GARP. AEI can be interpreted as a "margin of error" (Varian, 1990) that we allow the agent to make in his consumption choices, or as a tolerance for wasted expenditure.

The behaviors of the MPI and AEI can be quite different, and they can give opposite conclusions on the same data. We present two simple examples to illustrate this point. The first example is in Figure 3, which shows two pairs of observations $\{(z,p),(x',p')\}$ and $\{(x,p),(x',p')\}$ which both violate WARP. The MPI for these

[2] Warshall's algorithm, suggested by Varian for checking GARP, can be used to calculate an approximation of MPI.