CHAPTER 5

Practical Issues in Revealed Preference Analysis

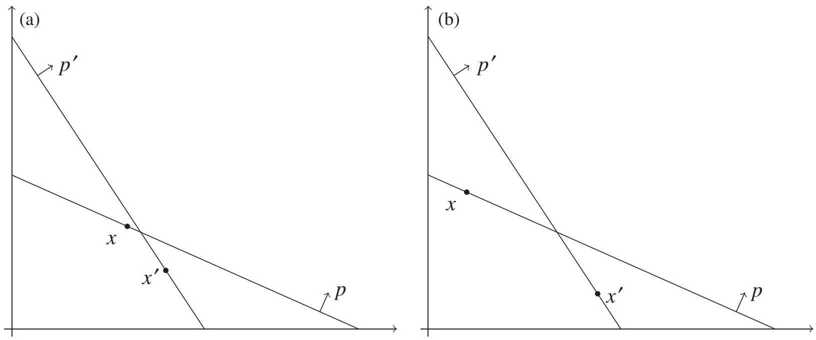
The tests in Chapters 3 and 4 are meant to be applicable to actual datasets, and many researchers have investigated these applications using experiments, consumption surveys, and other sources of data. Naturally, there are complications that arise when one tries to carry out the tests we have described. We shall focus on the basic application of GARP (or SARP) to data on consumption expenditures. The difficulties in applying GARP can be summarized as follows:

First, GARP is an "all or nothing" notion. A dataset either falsifies the theory of a rational consumer or it does not. One may, however, want to distinguish a grayscale of degrees of violation of the theory. It is possible that some violations can be attributed to simple mistakes on the part of a fully rational consumer. We develop concepts along these lines in 5.1.

Second, the nature of budget sets introduces problems with the power of testing for GARP. When two observed budget sets are nested, then there are no choices that can indicate a violation of GARP (actually of WARP in that case). More generally, any dataset in which budget sets have substantial overlap is biased towards the satisfaction of GARP. The problem of budget overlap is very real because often data contain more individual-level variation in expenditure levels than variation in relative prices. As we explain below (Section 5.2), these features cause budget sets to have substantial overlap.

Third, many studies do not track the identities of individual consumers. With such *cross-sectional* datasets, two observations (x^1, p^1) and (x^2, p^2) actually correspond to different individuals (or households), but they are identified as having the same preferences based on their observable characteristics. The procedure of identifying individuals based on their observable characteristics is called "matching" in statistics and econometrics. The basic problem is how to carry out this identification, or matching: when can we treat two individuals as the same for the purposes of revealed preference tests.

Moreover, certain cross-sectional datasets exacerbate the problem of power. The observations (x^1, p^1) and (x^2, p^2) in the data are of two different individuals (treated as the same agent) at similar points in time. Prices p^1 and p^2 are then



(x, p) and (x', p') violate GARP (in fact WARP). A more "severe" violation of GARP.

Fig. 5.1 Two observations: (x, p) and (x', p').

bound to be similar, because even prices in different locations are similar at the same point in time. The main source of variation in the data must then be in expenditure levels $p^k \cdot x^k$. As a consequence budget sets tend to be nested; the power of testing for GARP is thus diminished.

5.1 MEASURES OF THE SEVERITY OF A VIOLATION OF GARP

The revealed preference tests we have seen are dichotomous: either the data satisfy the test or they do not. But we are probably also interested in the *degree* to which a test is violated. Specifically, suppose that a dataset violates GARP, but that we somehow judge the violation to be mild. We may not be willing to conclude that the agents involved behaved irrationally.

5.1.1 Afriat's efficiency index

Afriat observes that if expenditures at each observation are "deflated" by some number $e \in [0,1]$, then the violation of GARP will disappear. Afriat proposes to measure the severity of a violation by how much expenditure needs to be deflated for the data to satisfy GARP.

Formally, define a modified revealed preference relation R_e by $x^k \; R_e \; y$ iff $ep^k \cdot x^k \geq p^k \cdot y$; define P_e similarly. For $e \in [0,1]$ small enough, the pair $\langle R_e, P_e \rangle$ will be acyclic. *Afriat's efficiency index (AEI)* is defined as the supremum over all the numbers e such that $\langle R_e, P_e \rangle$ is acyclic:

$$AEI = \sup\{e \in [0,1] : \langle R_e, P_e \rangle \text{ is acyclic.}\}$$

AEI is an intuitive measure of a violation of GARP. Consider, for example, the violation in Figure 5.1. The violation represented in Figure 5.1(b) is more severe than the one in 5.1(a). The difference is reflected in the AEI because a large deflation of expenditure (a smaller e) is needed to account for the violation.

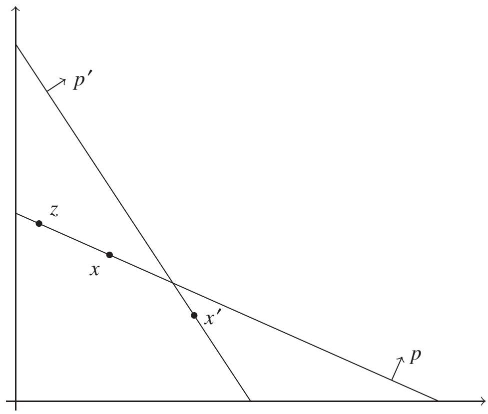


Fig. 5.2 Two violations of WARP: (x, p), (x', p') and (z, p), (x', p').

Afriat's index is also called the *critical cost efficiency index*, or CCEI in the literature.

5.1.2 Varian's version of AEI

Varian modifies Afriat's index by allowing e to vary across the different price vectors. Consider a vector $\theta = (e^k)_{k=1}^K$ of numbers in $[0,1]$, one for each observation. Define the binary relation R_θ as $x^k \, R_\theta \, x^l$ if $e^k p^k \cdot x^k \geq p^k \cdot x^l$. Define the strict relation P_θ analogously. There is a set Θ of vectors θ such that the corresponding $\langle R_\theta, P_\theta \rangle$ satisfies GARP. *Varian's efficiency index (VEI)* is the closest distance of a vector θ to the unit vector ($e^k = 1$ for all k), among those θ for which $\langle R_\theta, P_\theta \rangle$ is acyclic. Formally,

$$VEI = \inf\{\|\mathbf{1} - \theta\| : \langle R_\theta, P_\theta \rangle \text{ is acyclic.}\}$$

5.1.3 The Money Pump Index

Both Afriat's and Varian's indices have a problem. Consider the example in Figure 5.2. The AEI is the same for the two violations in the figure: the data $\{(x,p),(x',p')\}$ and the data $\{(z,p),(x',p')\}$ have the same AEI. This is counterintuitive. Arguably $\{(z,p),(x',p')\}$ presents a worse violation than $\{(x,p),(x',p')\}$. The source of the problem is that every violation of GARP involves at least one cycle, and AEI (or VEI) try to break the cycle at its "weakest link." In the example in Figure 5.2, this requires deflating expenditure to the point that x' is on the budget set for x.

If one instead treats each "link" equally, then the problem goes away. Let the sequence $x^{k_1}, x^{k_2}, \ldots, x^{k_n}$ define a violation of GARP, a cycle of $\langle \succeq^R, \succ^R \rangle$.

The *money pump index (MPI)* of this violation is defined by

$$MPI_{\{(x^{k_1},p^{k_1}),\ldots,(x^{k_n},p^{k_n})\}} = \frac{\sum_{l=1}^{n} p^{k_l} \cdot (x^{k_l} - x^{k_{l+1}})}{\sum_{l=1}^{n} p^{k_l} \cdot x^{k_l}} \quad \text{(taking } k_{n+1} = k_1\text{)}. \quad (5.1)$$

It is easy to see that, in Figure 5.2, $\{(z,p),(x',p')\}$ has a higher level of MPI than $\{(x,p),(x',p')\}$.

The MPI is named for the "money pump" that one can obtain from a consumer who violates GARP. Arguably, what is wrong about the situation in Figures 5.1 and 5.2 is that an outsider could take advantage of the consumer that exhibits the behaviors in the figures. With data (x,p), (x',p') violating WARP, for example, the outsider could trade the consumer x for x' at prices p, thereby getting an amount of money $p \cdot (x - x') > 0$, and then trade x' back for x at prices p', thereby getting an amount of money $p' \cdot (x' - x) > 0$. The total amount obtained by manipulating the consumer in such a way is $p \cdot (x - x') + p' \cdot (x' - x)$. The MPI simply expresses this magnitude as a fraction of total expenditure.

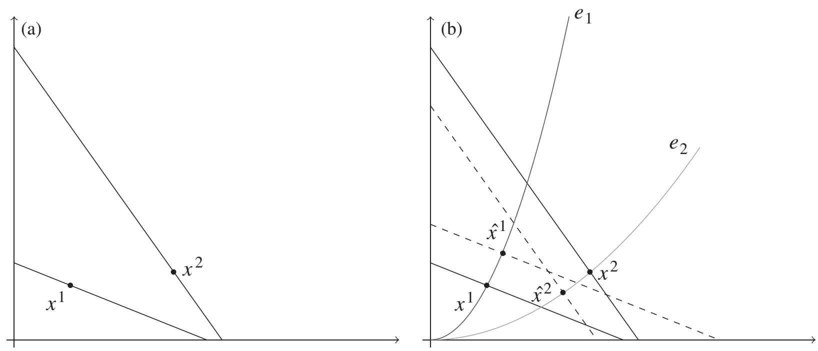
Under certain assumptions, the MPI is the basis for a statistical test of rational consumer behavior. If one assumes a source of statistical errors, one can construct a critical region, so that rational consumer behavior is rejected (beyond what can be explained as an error) if the MPI lies in this critical region.

5.2 POWER OF TESTING GARP

Failure to reject GARP can sometimes be caused by how budgets vary in the data. In fact, certain datasets are problematic in the sense that it is intrinsically hard to observe violations of GARP in them. Consider the example in Figure 5.3(a). No matter what the consumer chooses at these budgets, there will be no violation of GARP.

The figure is a stark example of a common phenomenon: when incomes vary more than prices, it is harder to detect a violation of GARP. In cross-sectional consumption data one has many observations of consumption chosen by different households (individuals) at similar prices. The econometrician tries to identify choices by different individuals as coming from the same (or similar) preference relations, but since prices vary little, and individual incomes can vary widely, GARP is easily satisfied. The situation is similar for aggregate time-series data, meaning time-series data for a whole economy. Such data often exhibit large year-on-year changes in aggregate expenditure (which equals aggregate incomes), and comparatively small changes in relative prices.

One might argue that the issue of power does not apply to the realm of revealed preference analysis. The theory only claims that agents' behavior is *as if* they were maximizing a utility function. So what if the budgets are not set up in a way that might detect a violation of GARP? There is no sense in which utility maximization could be a true model other than when behavior satisfies GARP: this is in contrast to statistical models in which one affirms



Nested budget sets. Choices adjusted by Engel curves violate WARP.

Fig. 5.3 Power of GARP.

the existence of an unknown true parameter. Having mentioned this point of view, we shall nevertheless proceed with an exploration of the role of power in revealed preference tests.

5.2.1 Bronars' index

Recognizing the problems with the power of GARP, Stephen Bronars proposes an index that tries to measure the power of using any particular collection of budgets for testing GARP. His idea is: Given a collection of budget sets $B^1,\dots,B^K$, simulate an "irrational" consumer choosing from these sets and see how frequently the consumer violates GARP. Bronars assumes a consumer who chooses a consumption bundle fully at random from the budget set.

The *Bronars' index* of a collection of budget sets is the probability that randomly chosen consumption bundles $x^1,\dots,x^K$ violate GARP, where x^k is chosen uniformly at random on the boundary (budget line) of the budget B^k.

The index is implemented using Monte Carlo simulation. If there are L goods, one method draws $s^k_1,\dots,s^k_L$ at random and chooses $x^k \in B^k$ such that the budget share of good l is s^k_l. This is done in such a way that the distribution over the budget line is uniform. A second method targets the actual budget shares in the data: If θ^k_l is the actual observed budget share of good l in observation k, let $s^k_l = \theta^k_l z^k_l / \sum_h \theta^k_h z^k_h$, where the variables $z^k_1,\dots,z^k_L$ are drawn uniformly and independently from $[0,1]$.

Bronars' index is now routinely used in experimental studies of GARP. When designing an experiment, researchers choose budgets so as to exhibit a high Bronars' index. An alternative measure of power, introduced by Andreoni, Gillen and Miller, is based on having data from many agents choosing from the same budget sets. One draws one choice at random, among the observed choices, from each budget set. Then one tests if the resulting "bootstrapped" choices pass GARP.

5.2.2 Engel curve correction

One solution to the problem of power is to use Engel curves to adjust the data. Let d be a demand function. For given p, the *Engel curve* associated to p is the function $m \mapsto d(p,m)$.

Given is a consumption dataset $D = \{(x^k, p^k)\}_{k=1}^K$. Suppose that we have available an Engel curve $e_k : \mathbf{R}_+ \to \mathbf{R}^l_+$ associated to price vector p^k; that is, $e_k(y)$ is the bundle demanded at prices p^k when income is y (Engel curves e_k would be fitted from the data). Note that $y = p^k \cdot e_k(y)$. Now we can correct the problem by adjusting income so that the budget lines are no longer nested. Figure 5.3(b) illustrates the idea. The choices and budgets from Figure 5.3(a) exhibit no violation of GARP, but if we adjust the incomes for the two budget lines (keeping the prices the same) to the two dashed lines, then the budget sets are no longer nested and there is scope for a violation of GARP. Indeed, at these hypothetical budgets, we can find the choices consistent with Engel curves e_1 and e_2. The choices induced by the Engel curves present a violation of WARP.

The idea of using Engel curves was developed by Blundell, Browning, and Crawford. They propose to consider a fixed sequence of choices and prices

$$x^{k_1}, \ldots, x^{k_l}, \ldots x^{k_L},$$

with a distinguished choice x^{k_l}. Transform the sequence as follows. For each k, we let $e_k : \mathbf{R}_+ \to \mathbf{R}^n_+$ be an Engel curve for prices p^k. That is, e_k satisfies the following properties:

- For all k, $e_k(p^k \cdot x^k) = x^k$.
- For all k and all $y \le y'$, $e_k(y) \le e_k(y')$.
- For all k and all y, $p^k \cdot e_k(y) = y$.

So, these Engel curves need not come from preference maximization, but importantly, the second criterion requires that the Engel curves are associated with normal demand.

We proceed to describe the construction. For $x^{k_{l-1}}$, the term preceding x^{k_l} in the sequence, we use the Engel curve for prices $p^{k_{l-1}}$ to find a hypothetical choice $\hat{x}^{k_{l-1}}$ on a budget line such that $\hat{x}^{k_{l-1}}$ is revealed preferred to x^{k_l}. Specifically, let $\hat{m}^{k_{l-1}} = p^{k_{l-1}} \cdot x^{k_l}$ and $\hat{x}^{k_{l-1}} = e_{k_{l-1}}(\hat{m}^{k_{l-1}})$. Given $\hat{x}^{k_{l-1}}$, we can use the same idea to construct $\hat{x}^{k_{l-2}}$ from $\hat{x}^{k_{l-1}}$. That is, let $\hat{m}^{k_{l-2}} = p^{k_{l-2}} \cdot \hat{x}^{k_{l-1}}$ and $\hat{x}^{k_{l-2}} = e_{k_{l-2}}(\hat{m}^{k_{l-2}})$. Continuing in this fashion, we construct $\hat{x}^{k_1}, \ldots, \hat{x}^{k_{l-1}}$ with the property that

$$\hat{x}^{k_1} \succeq^R \hat{x}^{k_2} \succeq^R \cdots \succeq^R \hat{x}^{k_{l-1}} \succeq^R x^{k_l}.$$

The construction of $\hat{x}^{k_{l+1}}, \ldots, \hat{x}^{k_L}$ is similar. Let $\hat{m}^{k_{l+1}}$ be the solution to the equation

$$p^{k_l} \cdot x^{k_l} = p^{k_l} \cdot e_{k_{l+1}}(\hat{m}^{k_{l+1}})$$

(such a solution exists and is unique under our assumptions on Engel curves; namely, continuity is implied by normality). Let $\hat{x}^{k_{l+1}} = e_{k_{l+1}}(\hat{m}^{k_{l+1}})$ and note that $x^{k_l} \succeq^R \hat{x}^{k_{l+1}}$ by construction. Given, $\hat{x}^{k_{l+1}}$ and $\hat{m}^{k_{l+1}}$, let $\hat{m}^{k_{l+2}}$ be the solution to the equation $\hat{m}^{k_{l+1}} = p^{k_{l+1}} \cdot e_{k_{l+2}}(\hat{m}^{k_{l+2}})$; and set $\hat{x}^{k_{l+2}} = e_{k_{l+2}}(\hat{m}^{k_{l+2}})$. Continuing in this fashion we obtain a sequence $x^{k+l} \succeq^R \hat{x}^{k_{l+1}} \succeq^R \cdots \succeq^R \hat{x}^{k_L}$.

The usefulness of the construction follows from the following proposition. The construction never eliminates a violation of GARP. Thus every sequence exhibiting a violation of GARP remains a violation after being adjusted by Engel curves. In other words, transforming a sequence in this way can only increase the chances of observing a violation of GARP.

Proposition 5.1 *If $x^{k_1}, \ldots, x^{k_L}$ is a sequence with $x^{k_1} \succeq^R \cdots \succeq^R x^{k_L} \succ^R x^{k_1}$ and $\hat{x}^{k_1}, \ldots, \hat{x}^{k_L}$ is the Engel-curve adjusted sequence when x^{k_L} is the distinguished element, then $\hat{x}^{k_1} \succeq^R \cdots \succeq^R \hat{x}^{k_L} \succ^R \hat{x}^{k_1}$.*

Proof. Let $m^{k_l} = p^{k_l} \cdot x^{k_l}$. It is easy to see by induction that $m^{k_l} \geq \hat{m}^{k_l}$ for $l = 1, \ldots, L$. By construction,

$$\hat{m}^{k_{L-1}} = p^{k_{L-1}} \cdot x^{k_L} \leq p^{k_{L-1}} \cdot x^{k_{L-1}} = m^{k_{L-1}},$$

where the inequality follows from $x^{k_{L-1}} \succeq^R x^{k_L}$. Then, the normality of demand implies that

$$x^{k_{L-1}} = e_{k_{L-1}}(m^{k_{L-1}}) \geq e_{k_{L-1}}(\hat{m}^{k_{L-1}}) = \hat{x}^{k_{L-1}},$$

so we obtain that

$$m^{k_{L-2}} = p^{k_{L-2}} \cdot x^{k_{L-2}} \geq p^{k_{L-2}} \cdot x^{k_{L-1}} \geq p^{k_{L-2}} \cdot \hat{x}^{k_{L-1}} = \hat{m}^{k_{L-2}}.$$

By continuing in this fashion, we show that $m^{k_l} \geq \hat{m}^{k_l}$ for $l = 1, \ldots, L$. Then, $x^{k_L} \succ^R x^{k_1}$ implies that

$$p^{k_L} \cdot x^{k_L} > p^{k_L} \cdot x^{k_1} \geq p^{k_L} \cdot \hat{x}^{k_1},$$

as $x^{k_1} = e_{k_1}(m^{k_1}) \geq e_{k_1}(\hat{m}^{k_1}) = \hat{x}^{k_1}$. Hence, $x^{k_L} \succ^R \hat{x}^{k_1}$; so the fact that $\hat{x}^{k_1} \succeq^R \cdots \succeq^R x^{k_L}$ establishes the result, using the fact that $\hat{x}^{k_L} = x^{k_L}$.

5.3 AN OVERVIEW OF EMPIRICAL STUDIES

We briefly describe some of the best-known empirical studies of GARP, classified by the kind of datasets that they use.

5.3.1 Panel data

Koo (1963) may be the first empirical study of revealed preference in consumption. He uses a panel dataset of consumption choices, and finds that relatively few households (panelists) in his sample satisfy GARP fully. It is interesting that such an early study uses panel data, a kind of data that corresponds closely to the theory, and presents fewer complications than some of the later uses of cross-sectional and time-series data. Koo introduces a measure of the degree of satisfaction of GARP, based on the idea that some subset of the observed choices could be consistent with GARP. His measure is flawed, however, as pointed out by Dobell (1965).

5.3.2 Cross-sectional data

Famulari (1995) tests for violations of GARP using a cross-sectional dataset obtained from a consumer expenditure survey. She focuses on budget sets that involve similar levels of expenditure, as a way of addressing the standard problems of the power of GARP. She assumes that bundles with similar costs, and where goods have similar expenditure shares, may be difficult to compare for consumers. Hence she focuses on comparing observations where the consumption bundles have similar costs, but dissimilar expenditure shares. Famulari also uses Afriat's efficiency index to account for possible measurement errors that could lead her to find spurious violations of GARP.

Based on their economic and demographic characteristics, Famulari groups her sample of some four thousand "consumer units" into 43 different types of households (this methodology is called "matching" in statistics and econometrics). She defines a rate of violation of GARP as follows: a pair of observations x^k and x^l constitute a violation if x^k stands in the transitive closure of the revealed preference relation to x^l (x^k is indirectly revealed preferred to x^l), while x^l is strictly revealed preferred to x^k. Famulari's *rate of violation of GARP* is the number of pairs that induce a violation divided by the total number of pairs in the data. Her results are that violation rates are quite small, and they are attributable to measurement errors (using Afriat's efficiency index).

Dowrick and Quiggin (1994) use a revealed preference approach to cross-country welfare comparisons. As a first step, they test for GARP using cross-country aggregate consumption data. Specifically, they obtain one observation (p^k, x^k) for each country in a sample of 60 countries (using data from 1980). For country k, x^k is a per-capita aggregate consumption bundle, and p^k is a country-specific price index. They treat the cross-country dataset as a consumption dataset, the idea being that a common representative consumer could exist across these countries. They use revealed preference comparisons as a comparison in the standard of living. They find almost no violations of GARP (and that GARP is very close to WARP). The results are consistent with the existence of seven categories of countries, such that the bundle consumed by a high-category country is always revealed preferred to a lower-category country; and such that most countries in the same category are not comparable according to revealed preference.

The study of Blundell, Browning, and Crawford (2003) is also worth mentioning here. They use cross-sectional data and introduce the technique described in Section 5.2.2.

5.3.3 Time-series data

Varian (1982) studies yearly aggregate data on consumption in the period 1947–1978. A test for GARP in this case can be interpreted as a test for the existence of an aggregate representative consumer. Varian finds no violation of GARP. The test, however, has very low power for yearly aggregate data. In the post-war period, the US aggregate income was increasing markedly year after year, while relative prices were relatively stable. This fact is emphasized by Varian in his paper: an observation that prompted later researchers to be concerned with the power of testing for GARP. Landsburg (1981) conducts a similar study on UK consumption data, and also notes that increasing income is problematic, attributing the observation to Gary Becker. See also Chalfant and Alston (1988). Swofford and Whitney (1986, 1987) are early related studies on the demand for financial assets.

Browning (1989) is another well-known study using time-series data. Instead of testing for GARP, Browning is interested in the maximization of a utility function of a particular form: an additively time-separable utility function, with no discounting and a per-period utility index that is constant over time. The test for rationalizability by this kind of utility turns out to be cyclic monotonicity: see Proposition 4.5.

Using aggregate time-series data from the UK and the US, Browning finds that GARP can never be rejected, which is in line with Varian's findings and the comments on the power of GARP we made above. Cyclic monotonicity is rejected, but there are fairly long periods of time for which cyclic monotonicity holds. Given that the data satisfy GARP, we know that there are multipliers for the prices for which the data with "adjusted prices" pass cyclic monotonicity (see the discussion after Afriat's axiom on Page 41). Browning proceeds to calculate such multipliers, and uses them to inquire about the reasons behind the falsification of the additively separable model.

5.3.4 Experimental data

Battalio, Kagel, Winkler, Fischer, Basmann, and Krasner (1973) ran a field experiment in the female ward of a psychiatric hospital. Patients (who were diagnosed psychotic) could exchange tokens for different consumption goods. By varying the value of the tokens, the authors induce a variety of different budget sets, and record the patients' purchases. Battalio *et al.* found that many patients satisfied GARP. In particular, if one allows for small measurement errors in quantities, then almost all the psychiatric patients' behavior is consistent with GARP. The authors also look at the dynamic behavior of the few patients who violate GARP, and argue that the patients reacted to price changes in the correct direction, but that they failed to fully adjust to the new prices. It is possible that if patients had had sufficient time to get used to the price change, then no violations would have been observed.

Sippel (1997) presented individuals with a menu of choices, determined by standard budget sets. Subjects were according to the random decision selection mechanism. Individuals were paid in consumption goods, and were required to consume the goods at the experiment. Overall, subjects were found to be inconsistent with classical demand theory. By setting an AEI of 95%, most subjects were found to be consistent with the predictions of preference maximization, but the power of the test also decreased substantially. The study then chose perturbed demand close to actual demand, finding that the number of violations changes little after perturbation, suggesting that inconsistency cannot be due to error alone.

Andreoni and Miller (2002) are motivated by the standard experimental finding that agents are too generous when asked to split a given amount of money. In "dictator game" experiments, subjects are asked to split x dollars between themselves and another subject, who is completely anonymous. Andreoni and Miller run such dictator-game experiments, and test if the observed altruistic behavior is utility-maximizing, by a utility that depends on the monetary rewards received by both agents. They find that 98% of subjects make choices that are consistent with utility maximization. Andreoni and Miller go further than most revealed-preference exercises in estimating a parametric function of a utility function accounting for subject's choices (about half the subjects can be classified as using a linear, CES (constant elasticity of substitution), or Leontief utility).

Harbaugh, Krause, and Berry (2001) perform an experiment using seven- and eleven-year-old children, as well as college students. They test each of these populations for compliance with GARP, after making them choose from several budget sets. They find a relatively small number of violations of GARP. Seven-year-olds violate GARP much more than older children (but still exhibit a relatively small number of violations). Eleven-year-old children behave close to rationally, and there are few differences between college-age adults and eleven-year-old children. It is interesting that the authors find that violations of GARP are largely uncorrelated with the results of a test that measures mathematical ability in children.

Choi, Kariv, Müller, and Silverman (2014) run an experiment on choices from budgets, and correlate the degree of consistency with GARP (as measured by the Afriat efficiency index) with the demographic and socioeconomic characteristics of the subjects. They work with a large sample of over 2,000 Dutch households. In their experiments, subjects are presented with 25 different budgets, each one chosen randomly. They find an average AEI of .88, and that almost half the population has an AEI above .95; so the population is quite close to satisfying GARP. The meat of their study lies in correlating AEI with the socioeconomic characteristics of the subjects. They find that men as well as highly educated and rich individuals score higher AEI than others. They also find that high AEI scores predict high levels of wealth: a standard deviation increase in AEI predicts 15–19% higher household wealth.

5.4 CHAPTER REFERENCES

A basic reference to matching in statistics is Rubin (1973). Famulari (1995) is an example of the use of matching in revealed preference analysis. Famulari (2006) uses a similar methodology to study labor supply from the viewpoint of revealed preference. She uses nonlinear budget sets to capture the effects of taxes on labor supply. A different methodological approach is taken by

Hoderlein and Stoye (2014), who focus on bounding the joint distribution of the population of consumer characteristics, and by Kitamura and Stoye (2013), who translate the problem into the framework analyzed in Chapter 7. Hoderlein (2011) should also be mentioned: he focuses on testing the implications of rationality for the Slutsky matrix of demand.

One issue we have not dealt with here is that consumption data usually reflect the decisions of households, not of single agents. We shall consider models of collective decision making in Chapter 9, 10, and 11, but the focus on those chapters will not be on consumption data. In the consumption setting, the issue is addressed by Cherchye, De Rock, and Vermeulen (2007, 2009), based on the model of Chiappori (1988). See also Browning and Chiappori (1998).

Afriat (1967) proposed Afriat's efficiency index. Varian (1990) extended the index as explained above, and gives it an interpretation in terms of consistency with rationality and errors in measurement (see also Varian, 1985 and Epstein and Yatchew, 1985). Afriat's index is also called the critical cost efficiency index, or CCEI in the literature. The critical cost efficiency index (CCEI) terminology was introduced by Varian (1991), based on Afriat (1972) (a paper on optimality in cost and production, see Chapter 6). It is used heavily in applied work: see Choi, Fisman, Gale, and Kariv (2007) and Choi, Kariv, Müller, and Silverman (2014).

The Money Pump Index is proposed by Echenique, Lee, and Shum (2011), who show that the MPI has an interpretation as a statistical test. Smeulders, Cherchye, De Rock, and Spieksma (2013) propose a variant of MPI that is computationally easy. An alternative measure of a violation of GARP results from computing the smallest set of observations one would need to delete from the data in order for it to satisfy GARP: see Houtman and Maks (1985), and more recently Dean and Martin (2013). An early related work in abstract choice theory is Basu (1984).

We have omitted a discussion of Varian's procedure for estimating demand responses by using revealed preference inequalities: see Varian (1982), Knoblauch (1992) and Blundell, Browning, and Crawford (2008). The discussion in Section 5.2.2 on the Engel-curve corrections approach to the problem of power is related to this procedure, and is taken from Blundell, Browning, and Crawford (2003).

Bronars' index was proposed in Bronars (1987). The model of random choice as a model of irrational behavior was proposed by Becker (1962). It is curious, however, that the point of Becker's paper is that random behavior is close to being rational. This point calls into question the idea of using random choices as a measure of the power of GARP. The "bootstrap" approach to power described in the text is due to Andreoni and Miller (2002).

There are alternative power indices formulated by Famulari (1995) and by Andreoni, Gillen, and Harbaugh (2013). Andreoni, Gillen, and Harbaugh's test, in particular, rests on a clever "reversion" of AEI to measure how far an observation that satisfies GARP is from not satisfying it. Beatty and Crawford

(2011) propose a measure of power that is related to Bronars', but based on the ideas in Selten (1991).

In addition to the evaluation of power based on a collection of budgets, it is sensible to use the data on actual choices to get an idea for the power of GARP. Andreoni, Gillen, and Harbaugh (2013) propose econometric methodologies for carrying out such evaluations of power.

The selection of empirical papers in Section 5.3 is obviously arbitrary. It is worth mentioning some papers that study the correlation between subjects' pass rates for GARP, and the presence of factors that may impede subjects' cognitive abilities. In particular, the paper of Burghart, Glimcher, and Lazzaro (2013) finds that subjects in experiments who have consumed substantial amounts of alcohol still pass GARP. The paper by Castillo, Dickinson, and Petrie (2014) in contrast compares subjects who are sleepy with those who are fully alert: again pass rates for GARP are the same across treated and non-treated subjects. It is important to note that both papers do find an effect of the treatment on agents' risk-taking behavior.