Article:

Cherchye, Laurens, Demuynck, Thomas, De Rock, Bram et al. (1 more author) (2020) Revealed Preference Analysis with Normal Goods:Application to Cost-of-Living Indices. American Economic Journal: Microeconomics. pp. 165-188. ISSN 1945-7669

https://doi.org/10.1257/mic.20180133

American Economic Journal: Microeconomics 2020, 12(3): 165–188
https://doi.org/10.1257/mic.20180133

Revealed Preference Analysis with Normal Goods: Application to Cost-of-Living Indices†

By Laurens Cherchye, Thomas Demuynck, Bram De Rock, and Khushboo Surana*

We present a revealed preference methodology for nonparametric demand analysis under the assumption of normal goods. Our methodology is flexible in that it allows for imposing normality on any subset of goods. We show the usefulness of our methodology for empirical welfare analysis through cost-of-living indices. An illustration to US consumption data drawn from the Panel Study of Income Dynamics (PSID) demonstrates that mild normality assumptions can substantially strengthen the empirical analysis. It obtains considerably tighter bounds on cost-of-living indices and a significantly more informative classification of better-off and worse-off individuals after the 2008 financial crisis. (*JEL* D11, D12, E31, G01)

Changing price-income regimes can have a substantive impact on individual demand patterns. The empirical analysis of the associated welfare effects has attracted considerable attention in the applied welfare literature. In the current paper, we propose a structural method for such welfare analysis that is intrinsically nonparametric: it does not impose any parametric/functional structure on the individual utilities but merely exploits the preference information that is directly revealed by the observed consumption behavior. Particularly, we demonstrate that mild normality assumptions on the demand for (a subset of) goods can obtain a significantly informative analysis of individual cost-of-living indices. We show this through an empirical illustration to household demand data taken from the PSID, in which we analyze the welfare effects of the 2008 financial crisis for a sample of singles in the United States.

*Cherchye: Department of Economics, University of Leuven (KU Leuven), E. Sabbelaan 53, B-8500 Kortrijk, Belgium (email: laurens.cherchye@kuleuven.be); Demuynck: ECARES, Université Libre de Bruxelles, Avenue F. D. Roosevelt 50, CP 114, B-1050 Brussels, Belgium (email: thomas.demuynck@ulb.ac.be); De Rock: ECARES, Université Libre de Bruxelles and Department of Economics, University of Leuven (KU Leuven), Avenue F. D. Roosevelt 50, CP 114, B-1050 Brussels, Belgium (email: bderock@ulb.ac.be); Surana: Department of Economics and Related Studies, University of York, York, YO10 5DD, UK (email: khushboo.surana@york.ac.uk). John Asker was coeditor for this article. We thank the editor John Asker and three anonymous referees for many insighful comments, which helped us to substantially improve the paper. Laurens Cherchye gratefully acknowledges the European Research Council (ERC) for his Consolidator Grant 614221. Part of this research is also funded by the Research Fund KU Leuven and the Fund for Scientific Research-Flanders (FWO). Thomas Demuynck acknowledges financial support by the Fonds de la Recherche Scientifique-FNRS under grant nr F.4516.18. Bram De Rock gratefully acknowledges FNRS, FWO, and EOS for their financial support.
†Go to https://doi.org/10.1257/mic.20180133 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

Welfare Evaluation and Counterfactual Demand Analysis.—The structural analysis of welfare effects associated with changing prices and/or incomes requires predicting demand in counterfactual price-income regimes. This issue is standardly addressed by adopting a parametric approach, which assumes a specific functional form for the consumers' utility or expenditure functions.[1] The parameters of this functional form are then estimated from the observed consumption behavior, and these estimations can be used to interpolate or extrapolate demand in unobserved price-income situations. A main problem of this parametric approach is that it crucially relies on some a priori assumed functional form for the individual preferences, which is typically nonverifiable. This implies an intrinsic risk of specification error.

We can avoid this specification risk by adopting the nonparametric revealed preference approach that was initiated by Samuelson (1938) and Houthakker (1950) and further developed by Afriat (1967), Diewert (1973), and Varian (1982). Basically, this nonparametric approach develops testable implications for observed consumption patterns (prices and quantities) that must hold under rational demand behavior associated with *any* well-behaved utility function. These testable implications are then used as a basis for counterfactual demand predictions in the form of set identification (producing bounds on possible demand responses in new price-income regimes). By its very nature, this nonparametric approach avoids the possibility of erroneous conclusions following from a wrongly specified functional form.

Revealed Preference Analysis and Normal Goods.—Although this nonparametric orientation of the revealed preference approach is conceptually appealing, its empirical usefulness is often put into question. Generally, an informative empirical analysis requires a rich dataset with high price variation and low income variation. In many observational settings, however, the opposite holds true (i.e., low price variation combined with high income variation). In such cases, the nonparametric testable implications have little empirical bite, and, correspondingly, the set identification results are not very informative (see, for example, Varian 1982 and Bronars 1987 for detailed discussions). As an implication, the revealed preference methodology is then of limited practical value.

In the current paper, we show that this lack of power can be remediated by assuming normality of the goods that are consumed. Normality is often a natural assumption to make. Basically, a good is normal if its income expansion path is increasing. A convenient feature of our method is that we can impose normality without needing to estimate the expansion path; our nonparametric testable implications apply to *any* expansion path that satisfies normal demand. Moreover, our method applies to settings with any number of goods and can impose normality on any subset of these goods. The only assumption it makes is that normality holds for the observed prices, so avoiding the stronger hypothesis that normality must apply to any (observed or unobserved) price.

[1] Popular functional forms in the literature are the Cobb-Douglas, the translog (Christensen, Jorgenson, and Lau 1975), the almost ideal demand (Deaton and Muellbauer 1980), and quadratic almost ideal demand specification (Banks, Blundell, and Lewbel 1997).

In a recent series of papers, Blundell, Browning, and Crawford (2003, 2007, 2008) and Blundell et al. (2015) also used the assumption of normal demand for observed prices to deal with the power issue associated with empirical revealed preference analysis. However, we see at least two main differences between the method proposed by these authors and our novel method. First, they assume that normality holds for all goods simultaneously, whereas our method is equally applicable to normality for any subset of goods. Second, and more importantly, these authors exploit normality of demand by using (nonparametrically) estimated income expansion paths (assuming a repeated cross-sectional dataset). As indicated above, our method avoids this prior estimation step (and associated statistical issues); it directly applies revealed preference restrictions (for normal demand) to the observed consumption choices. Interestingly, our empirical application shows that our method can yield an informative welfare analysis even with a short time series of (three) consumption observations per individual.

In another closely related paper, Cherchye, Demuynck, and De Rock (2018)—henceforth, CDR—also establish revealed preference conditions for normal demand, with a main focus on the two goods setting. A first crucial difference with the current paper is that CDR consider the stronger assumption that normality holds for all (observed and unobserved) nonnegative prices, whereas we use the substantially weaker assumption that imposes normality (only) for the observed prices. Next, CDR focus on so-called WARP-consistent demand, implying that they do not exploit transitivity of preferences. In the current paper, however, we also explicitly consider the testable implications of transitivity. Rose (1958) showed that transitivity has no empirical bite in the two goods setting. As an implication, our testable implications will be weaker than the ones of CDR if there are only two goods (because of CDR's stronger normality assumption; see above). For more than two goods, transitivity may have empirical bite, and thus, our testable implications may become more restrictive than the ones of CDR. Evidently, whether or not this is the case will crucially depend on the nature of the observed price regimes. Finally, while CDR's conditions are necessary and sufficient for rational demand that satisfies normality when there are two goods, they are only necessary (but not sufficient) for the general setting with more than two goods. By contrast, our testable implications provide a necessary and sufficient characterization of rationality under normal demand that applies to any number of goods.

Empirical Welfare Analysis and Cost-of-living Indices.—We show that our revealed preference method can be used for a meaningful welfare analysis on the basis of cost-of-living indices. We demonstrate this through an empirical application to data drawn from the PSID. We select a balanced panel from the 2007, 2009, and 2011 waves of the PSID to study the welfare effects of the 2008 financial crisis. A large number of studies has analyzed these welfare effects since the onset of the crisis. As the crisis led to a substantial rise in unemployment, the principal focus so far has been on the extensive margin of labor supply (see, for example, Verick 2009, Hurd and Rohwedder 2010, Goodman and Mance 2011, Deaton 2012). By contrast, in our application, we concentrate on individuals who remained employed after the crisis.

More specifically, our structural analysis assumes a model of rational labor supply for singles who spend their potential income on leisure, food, housing, and other goods, hereby imposing normality on all consumption categories except for leisure. To assess the empirical bite of the testable implications associated with normality, we also compute the empirical results for the rational labor supply model without normal demand. Our results show that imposing normality entails a substantially more powerful empirical analysis. In particular, we obtain considerably tighter bounds on cost-of-living indices and a significantly more informative classification of better-off and worse-off individuals after the 2008 crisis.

Outline.—Section I develops the revealed preference characterization of utility maximization under normality assumptions. Section II introduces the cost-of-living index for our empirical welfare analysis. We also define the goodness-of-fit and predictive success measures that we will use to evaluate the empirical performance of our normality assumptions. Section III presents our empirical application to PSID data. Section IV concludes.

I. Rational Demand with Normal Goods

Our main theoretical result defines the testable implications for the observed demand behavior to be consistent with rationality (i.e., utility maximization) and normality of (a subset of) the consumed goods. To this end, we first define the Generalized Axiom of Revealed Preference (GARP) in terms of Hicksian demand bundles that correspond to the observed prices and associated utility levels (for the given quantity bundles). Imposing normality boils down to restricting these Hicksian demand bundles at any observed price regime to be monotone in utility (Fisher 1990). Basically, our testable revealed preference conditions verify whether there exists at least one possible specification of the utility levels and Hicksian demand bundles that satisfy this requirement. If so, we cannot reject the joint hypothesis of normality and rational behavior.

Generalized Axiom of Revealed Preference (*GARP*).—Throughout, we focus on a finite set T of observed prices and corresponding quantities. For each consumption observation $t \in T$, let $q_t \in \mathbb{R}^n_+$ and $p_t \in \mathbb{R}^n_{++}$ denote the (column) vectors of quantities and prices, respectively. This defines the dataset $S = \{(p_t, q_t)\}_{t \in T}$. We say that S is "rationalizable" if there exists a utility function $u(\cdot)$ such that for each observation $t \in T$, q_t maximizes this function $u(\cdot)$ over all affordable bundles for the given prices p_t and outlay $x_t = p_t q_t$. Throughout, we will assume utility functions that are continuous and strictly monotone.

DEFINITION 1: *A dataset $S = \{(p_t, q_t)\}_{t \in T}$ is rationalizable if there exists a continuous and strictly monotone utility function $u : \mathbb{R}^n_+ \to \mathbb{R}$ such that for all $t \in T$ and $x_t = p_t q_t$,*

$$q_t \in \arg\max u(q) \quad \text{subject to} \quad p_t q \leq x_t.$$

Varian (1982) has shown that GARP defines a necessary and sufficient condition for a dataset S to be rationalizable. Thus, checking rationalizability boils down to verifying whether or not the set S satisfies GARP. To formally define this GARP requirement, we will need the following concepts.

DEFINITION 2: *Consider a dataset* $S = \{(p_t, q_t)\}_{t \in T}$. *We say that* q_t, $t \in T$, *is directly revealed preferred to the bundle* q_v, $v \in T$, *if* $p_t q_t \geq p_t q_v$. *We denote this as* $q_t R^D q_v$. *Next, we say that* q_t *is strictly directly revealed preferred to* q_v *if* $p_t q_t > p_t q_v$. *We denote this as* $q_t P^D q_v$. *Finally, we say that* q_t *is revealed preferred to* q_v *if there exists a (possibly empty) sequence* $u, s, \ldots, r \in T$ *such that*

$$q_t R^D q_u, q_u R^D q_s, \ldots, q_r R^D q_v.$$

We denote this as $q_t R q_v$.

Thus, the quantity bundle q_t is directly revealed preferred to the bundle q_v (i.e., $q_t R^D q_v$) if q_v was affordable when bundle q_t was chosen (i.e., $p_t q_t \geq p_t q_v$). If the inequality is strict (i.e., $p_t q_t > p_t q_v$), then q_t is strictly directly revealed preferred to q_v (i.e., $q_t P^D q_v$). Finally, from the direct revealed preference relations, we can define the more general concept of (direct or indirect) revealed preference relations by exploiting transitivity of preferences (i.e., $q_t R q_v$ follows from $q_t R^D q_u, q_u R^D q_s, \ldots, q_r R^D q_v$).

We can now define GARP.

DEFINITION 3: *A dataset* $S = \{(p_t, q_t)\}_{t \in T}$ *satisfies GARP if for all* $t, v \in T$, $q_t R q_v$ *implies not* $q_v P^D q_t$.

In words, a dataset S satisfies GARP if for any two observed bundles q_t and q_v, $q_t R q_v$ implies that q_v is not strictly directly revealed preferred to q_t (i.e., not $q_v P^D q_t$). Intuitively, GARP excludes that bundle q_t is revealed preferred to q_v while, at the same time, q_t was affordable at a strictly lower cost when q_v was purchased.

In what follows, we will focus on a less standard reformulation of the GARP condition in Definition 3. This alternative formulation will be instrumental for our characterization of rationalizable consumer behavior under normal demand. It is contained in the following result.[2]

PROPOSITION 1: *A dataset* $S = \{(p_t, q_t)\}_{t \in T}$ *satisfies GARP if and only if there exist numbers* $(u_t)_{t \in T}$ *such that for all* $s, t \in T$

- *if* $u_t \geq u_s$, *then* $p_s q_s \leq p_s q_t$;
- *if* $u_t > u_s$, *then* $p_s q_s < p_s q_t$.

[2] This equivalent reformulation of GARP has been used in the literature on nonparametric production analysis. We refer to Varian (1984, Theorem 2) for a formal proof of Proposition 1.

The second equivalence shows that a dataset S can be verified by checking the existence of "utility numbers" u_t that satisfy a series of "if-then" conditions. Intuitively, each number u_t represents the consumer's utility level associated with the bundle q_t. If the utility level at observation t is (strictly) above the utility level at observation s (i.e., $u_t \geq (>) \ u_s$), then the bundle q_t must be (strictly) more expensive than the bundle q_s at the prices p_s.

Normality-extended GARP (*N-GARP*).—Let $M \subseteq \{1, \ldots, n\}$ be a subset of the goods that are consumed. We say that a dataset S is *rationalizable by normal demand on the subset M* if there exists a well-behaved utility function that (i) represents each observed bundle q_t as utility maximizing under (ii) the additional requirement that for each good $i \in M$, the income expansion path at the observed prices has a positive slope. Formally, we have the following definition.

DEFINITION 4: *A dataset* $S = \{(p_t, q_t)\}_{t \in T}$ *is rationalizable by normal demand on the subset* M $(M \subseteq \{1, \ldots, n\})$ *if there exists a continuous and strictly monotone utility function* $u : \mathbb{R}^n_+ \to \mathbb{R}$ *and functions* $q_t : \mathbb{R}_+ \to \mathbb{R}^n_+$ *such that for all* $t \in T$ *and* $x_t = p_t q_t$,

- $q_t(x) \in \arg\max u(q)$ *subject to* $p_t q \leq x$,
- $q_t^i(x)$ *is monotone in* x *for all* $i \in M$,
- $q_t = q_t(x_t)$.

In this definition, the function $q_t(\cdot)$ represents the income expansion path at the observed prices p_t, defining the quantities demanded by the consumer at the price-income pair (p_t, x) for any value of x. Definition 4 defines three conditions for the functions $u(\cdot)$ and $q_t(\cdot)$. The first condition states that for all income levels x, $q_t(x)$ maximizes the function $u(\cdot)$ over all affordable bundles at prices p_t and income x. The second condition imposes that $q_t^i(x)$ is increasing in x, meaning that good $i \in M$ is normal at prices p_t. The last condition requires that $q_t(x_t)$ equals the observed demand q_t for the observed income/outlay $x_t \ (= p_t q_t)$ and prices p_t.

In order to better grasp the meaning of our main result (captured by Proposition 2 below), we make use of dual demand theory. If utility functions are continuous and strictly monotone, then every utility maximization problem has a dual expenditure minimization problem where the objective is to minimize expenditures for a given price vector conditional upon a certain level of utility:

$$\underbrace{v(p,x) = \max_q u(q) \quad \text{subject to} \quad pq \leq x}_{\text{primal utility max problem}};$$

$$\underbrace{e(p,u) = \min_q pq \quad \text{subject to} \quad u(q) \geq u}_{\text{dual expenditure min problem}}.$$

The indirect utility function, here denoted by $v(p,x)$, is the inverse of the expenditure function, denoted by $e(p,u)$, in the sense that for all prices p, utility levels u, and income levels x, we have

$$v\big(p,e(p,u)\big) = u \quad \text{and} \quad e\big(p,v(p,x)\big) = x.$$

The expenditure function is increasing in utility u, and the indirect utility function is increasing in income x. In addition, if they are unique, the solution to the utility maximization problem, $q(p,x)$, which is called the Marshallian demand function, and the solution to the expenditure minimization problem, $h(p,u)$, which is called the Hicksian demand function, are related in the following sense:

$$q\big(p,e(p,u)\big) = h(p,u) \quad \text{and} \quad h\big(p,v(p,x)\big) = q(p,x).$$

Let us then consider two income levels x and x', with $x \geq x'$, and a good $i \in M$. If $q^i(p,x)$ satisfies normality, then

$$q^i(p,x) \geq q^i(p,x'),$$

and therefore, by the identity above,

$$h^i\big(p,v(p,x)\big) \geq h^i\big(p,v(p,x')\big).$$

Given that $v(p,x)$ is increasing in income x, this shows that normality of q^i implies that the Hicksian demand function $h^i(p,u)$ is increasing in utility u. Vice versa, if we take two utility levels u and u' with $u \geq u'$, then monotonicity of h^i in u requires

$$h^i(p,u) \geq h^i(p,u') \Leftrightarrow q^i\big(p,e(p,u)\big) \geq q^i\big(p,e(p,u')\big).$$

As $e(p,u)$ is increasing in u, this shows that q^i must be increasing in x, i.e., good i is a normal good. Summarizing, we conclude that monotonicity (normality) of $q^i(p,x)$ in x is equivalent to monotonicity of the Hicksian demand $h^i(p,u)$ in u.[3]

We can use this equivalence to establish the revealed preference characterization of rationalizable behavior as specified in Definition 4. This characterization provides nonparametric testable implications for the observed dataset S to be consistent with utility maximization under the additional assumption of normal demand. In particular, we can show that rationalizability under normal demand holds if and only if the dataset S satisfies the N-GARP.

[3] We refer to Fisher (1990) for a more formal statement of this argument.

DEFINITION 5: *For* $M \subseteq \{1, \ldots, n\}$, *a dataset* $S = \{(p_t, q_t)\}_{t \in T}$ *satisfies N-GARP if there exist numbers* $(u_t)_{t \in T}$ *and vectors* $(h_{t,v})_{t,v \in T}$ $(h_{t,v} \in \mathbb{R}^n_+)$ *such that for all* $r, s, t, v \in T$,

- $h_{t,t} = q_t$;
- *if* $u_t \geq u_v$, *then* $p_r h_{r,v} \leq p_r h_{s,t}$;
- *if* $u_t > u_v$, *then* $p_r h_{r,v} < p_r h_{s,t}$;
- *if* $u_t \geq u_v$, *then* $h^i_{r,v} \leq h^i_{r,t}$ *for all* $i \in M$.

The following proposition contains our main theoretical result.[4]

PROPOSITION 2: *A dataset* $S = \{(p_t, q_t)\}_{t \in T}$ *is rationalizable by normal demand on the subset M* $(M \subseteq \{1, \ldots, n\})$ *if and only if it satisfies N-GARP.*

Similar to Proposition 1, we obtain that rationalizability imposes the existence of utility numbers u_t that satisfy a series of if-then conditions. In our N-GARP definition, each vector $h_{t,v}$ represents the Hicksian demand bundle at prices p_t for the utility level associated with the bundle q_v (captured by the number u_v). In other words, $h_{t,v} = h(p_t, u_v)$.

Rationalizability requires the numbers u_t and vectors $h_{t,v}$ to satisfy the four conditions in Definition 5. The first condition states for each observation $t \in T$ that the Hicksian demand $h_{t,t} = h(p_t, u_t)$ must equal the observed Marshallian demand $q_t = q(p_t, x_t)$. The second and third conditions impose GARP (as formulated in Proposition 1) on the sets $(p_t, h_{t,v})_{t,v \in T}$, which consist of observed prices p_t and Hicksian demand vectors $h_{t,v} = h(p_t, u_v)$. To grasp the intuition behind these conditions, assume that $u_t \geq u_v$. Then, $h_{r,v} = h(p_r, u_v)$ represents the Hicksian demand at prices p_r and utility level u_v, which is situated on the intersection of the indifference curve of u_v and the hyperplane (tangent to this indifference curve) with slope p_r. Now, given that $u_t \geq u_v$, it must be that all bundles that obtain utility level u_t are above this hyperplane (because all bundles below the hyperplane have utility levels below u_v). Formally, for all q with $u(q) = u_t$, we must have $p_r h_{r,v} \leq p_r q$. Then, given that $u(h_{s,t}) = u(h(p_s, u_t)) = u_t$, it follows that $p_r h_{r,v} \leq p_r h_{s,t}$, which gives the second condition. The third condition has a similar interpretation. Finally, the fourth condition requires that the Hicksian quantities for each good $i \in M$ are monotonically increasing in utility, which corresponds to normal demand, i.e., if $u_t \geq u_v$, then $h^i_{r,v} = h^i(p_r, u_v) \leq h^i(p_r, u_t) = h^i_{r,t}$.[5]

Figure 1 presents a graphical illustration of the N-GARP condition for a setting with two normal goods. The figure shows two indifference curves corresponding to utility levels u_1 and u_2, with $u_2 > u_1$. The two budget lines correspond

[4] Online Appendix I contains the proof of Proposition 2.

[5] In principle, we can restrict the normality restriction to be imposed only on certain income regions. For example, suppose that we only want to impose normal demands on the income range $[\underline{y}_r, \bar{y}_r]$ for prices p_r; then it suffices to modify the fourth condition in Definition 5 as follows:

$$\text{if} \quad u_t \geq u_v, \quad p_r h_{r,v} = y_{r,v}, \quad p_r h_{r,t} = y_{r,t}, \quad \text{and} \quad y_{r,v}, y_{r,t} \in [\underline{y}_r, \bar{y}_r],$$

$$\text{then} \quad h^i_{r,v} \leq h^i_{r,t} \quad \text{for all } i \in M.$$

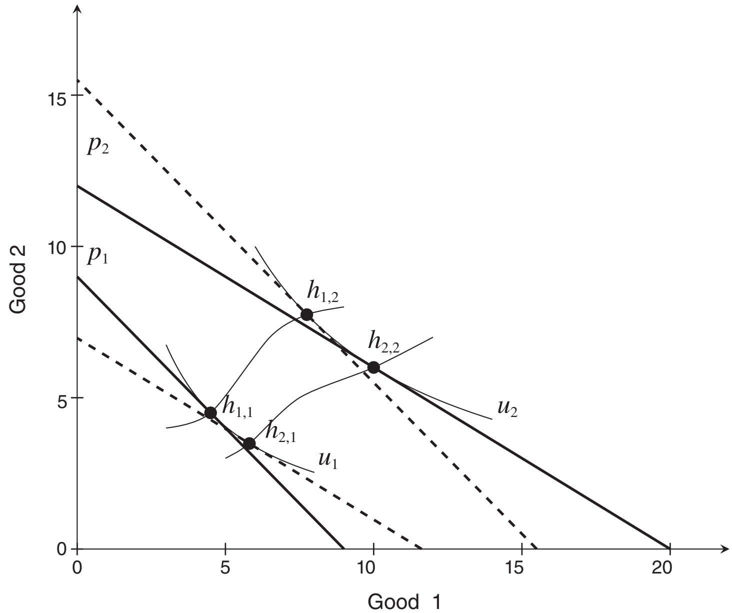

FIGURE 1. ILLUSTRATIVE EXAMPLE

to observation $(p_1,q_1) = (p_1,h_{1,1})$, which obtains utility level u_1, and observation $(p_2,q_2) = (p_2,h_{2,2})$, which obtains utility level u_2. We also depict two auxiliary, dashed budget lines that are parallel to the observed budget lines (i.e., they correspond to the same relative prices). The (unobserved) Hicksian demand $h_{2,1}$ corresponds to the bundle that would give the utility level u_1 at prices p_2. Similarly, $h_{1,2}$ is the bundle that would give utility level u_2 at prices p_1. The N-GARP condition requires that these (observed and unobserved) demands satisfy GARP and that $h_{2,1} \leq h_{2,2}$ and $h_{1,1} \leq h_{1,2}$. In reality, however, we do not observe these indifference curves, and, therefore, the N-GARP condition only imposes that it must be possible to construct hypothetical bundles $h_{1,2}$ and $h_{2,1}$ that satisfy these requirements.

When comparing the conditions in Proposition 1 with those in Definition 5, it is clear that N-GARP generally implies stronger rationalizability requirements than GARP. N-GARP reduces to GARP (only) in the limiting case that does not impose normality for any good. We illustrate the difference between N-GARP and GARP in Example 1, which contains a dataset that satisfies GARP but violates N-GARP. It indicates that imposing normality can yield a more powerful revealed preference analysis. This is an attractive feature as normality assumptions are often little debatable and thus easy to make.

Finally, in online Appendix II, we show that the N-GARP condition in Definition 5 can be reformulated in terms of inequality constraints that are linear in unknowns and characterized by (binary) integer variables. These linear inequality constraints are easily operationalized, which is convenient from an application point of view.[6]

[6]For example, we used the software package IBM ILOG CPLEX Optimization Studio for our empirical application in Section III.

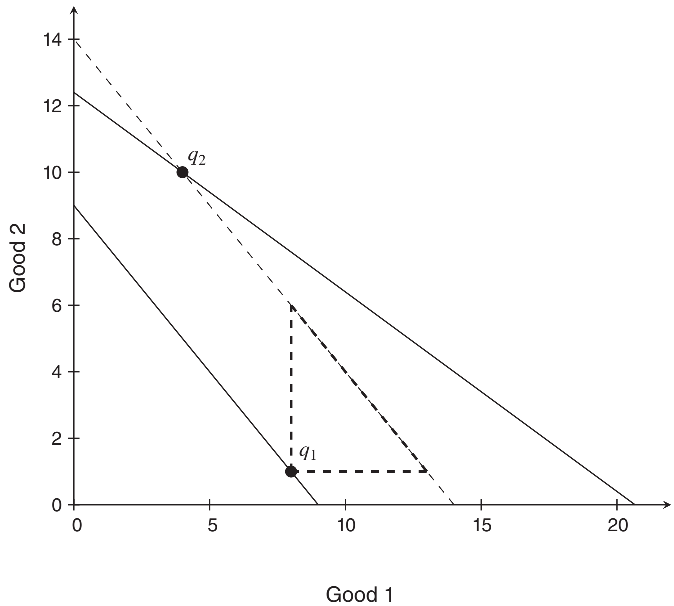

Figure 2. Example Dataset that Violates N-GARP but Not GARP

Example 1: We illustrate the difference between N-GARP and GARP by means of a simple numerical example using a dataset S with two goods $(n = 2)$ and two observations $(T = \{1,2\})$:

$$p_1 = \begin{bmatrix} 4 \\ 4 \end{bmatrix}, \quad p_2 = \begin{bmatrix} 3 \\ 5 \end{bmatrix}, \quad q_1 = \begin{bmatrix} 8 \\ 1 \end{bmatrix}, \quad q_2 = \begin{bmatrix} 4 \\ 10 \end{bmatrix}.$$

Figure 2 depicts the two quantity bundles and associated budget sets. From this figure, it is easy to verify that the set S satisfies GARP. In particular, the budget lines do not cross, which automatically implies consistency with GARP. More formally, referring to Proposition 1, we have $p_1 q_1 = 36$, $p_1 q_2 = 56$, $p_2 q_1 = 29$, and $p_2 q_2 = 62$. Then, using $u_1 = 0.1$ and $u_2 = 0.2$ obtains that all conditions in Proposition 1 are satisfied.

Next, we can show that the same dataset S violates N-GARP for $M = \{1,2\}$, i.e., both goods are assumed to be normal goods. In particular, we prove that there do not exist numbers u_1, u_2 and vectors $h_{1,1}$, $h_{1,2}$, $h_{2,1}$, $h_{2,2}$ that simultaneously meet the four conditions in Definition 5. To see this, we begin by noting that the first N-GARP condition imposes

$$(1) \qquad h_{1,1} = q_1 = \begin{bmatrix} 8 \\ 1 \end{bmatrix}, \quad \text{and} \quad h_{2,2} = q_2 = \begin{bmatrix} 4 \\ 10 \end{bmatrix}.$$

In addition, the second N-GARP condition (using that $u_2 \geq u_2$) imposes

$$(2) \qquad p_1 h_{1,2} \leq p_1 h_{2,2} \quad \text{and} \quad p_2 h_{2,2} \leq p_2 h_{1,2}.$$

Combining (1) and (2) obtains (using superscripts to indicate the quantities of goods 1 and 2)

$$4\,h_{1,2}^{1} + 4\,h_{1,2}^{2} \;\leq\; 56,$$

$$62 \;\leq\; 3\,h_{1,2}^{1} + 5\,h_{1,2}^{2}.$$

These two inequalities together imply

$$(3) \qquad 62 \;\leq\; 3h_{1,2}^{1} + 5h_{1,2}^{2} \;\leq\; 3h_{1,2}^{1} + 5\big(14 - h_{1,2}^{1}\big) \;\Leftrightarrow\; h_{1,2}^{1} \;\leq\; 4.$$

On the other hand, because $p_1 q_{1,1} = 36 < p_1 q_{2,2} = 56$, the third N-GARP condition in Proposition 5 requires

$$u_1 \;<\; u_2.$$

Then, the fourth N-GARP condition imposes (using that goods 1 and 2 are both normal)

$$h_{1,1} \;\leq\; h_{1,2}.$$

Combined with (1), this entails

$$h_{1,2}^{1} \;\geq\; 8,$$

which contradicts (3). Thus, we conclude that N-GARP is violated.

We can also graphically illustrate this N-GARP violation in Figure 2. To see this, we first note that the Hicksian demand $h_{1,2}$ should lie below the dashed line associated with the budget $p_1 q_2$. Also, if both goods are normal at the prices p_1, it must hold that $h_{1,2}$ contains more of both goods 1 and 2 than q_1 (i.e., $h_{1,2} \geq q_1$). Taken together, we conclude that $h_{1,2}$ is situated in the triangular region formed by the thick-dashed lines. Then, the conclusion that N-GARP is violated follows from the observation that no $h_{1,2}$ in this region is consistent with rationalizability of the consumption observation (p_2, q_2). Specifically, any such $q_{1,2}$ is strictly less expensive than the bundle q_2 at prices p_2. As an implication, for the outlay $p_2 q_2$ and prices p_2 associated with the quantity bundle q_2, the consumer could have chosen bundles strictly better than $h_{1,2}$. This implies that $h_{1,2}$ and q_2 cannot yield the same utility value for a strictly monotone utility function.

II. Cost of Living, Goodness-of-fit, and Predictive Success

In this section, we introduce some additional concepts and tools that will be useful for our following application. First, we show how our testable conditions for normal demand can be used to identify cost-of-living indices for comparing individual welfare in alternative price-income regimes. Next, in their original formulation, our revealed preference conditions for rational behavior under normal demand define "exact" tests: data either satisfy the requirements or not. In our empirical application, we will use an Afriat-type Critical Cost Efficiency Index (CCEI) to assess how closely behavior complies with rational behavior. This index will serve as a *goodness-of-fit* measure that has a specific interpretation as capturing the *economic significance* of violations of our testable implications. Finally, we present the *predictive success* measure that we will use in Section III to compare the empirical performance of the alternative normality assumptions under study.

Cost-of-living Indices.—An important application of empirical demand analysis consists of comparing consumers' welfare in alternative price-income regimes. More specifically, for two consumption observations (p_t, q_t) and (p_r, q_r), we not only wish to know which combination is (revealed) "better" by the consumer, but also "how much better." As utility theory is ordinal in nature, there is no unique answer to this last question. A popular method makes use of the money metric utility concept that was introduced by Samuelson (1974). In what follows, we will use this money metric representation of individual utility to compute cost-of-living indices associated with different price-income situations. Technically, we adapt the nonparametric method that was developed by Varian (1982), based on the GARP concept in Definition 3.[7] We will show that our N-GARP characterization in Definition 5 easily allows for computing lower and upper bounds on individuals' cost-of-living indices. This effectively set identifies these indices using the assumption of rationalizability under normal demand.

The money metric utility function gives the minimum expenditure required in observation t (with price-income pair (p_t, x_t)) to attain the same utility level as under some reference price-income regime (p_r, x_r). Formally, it is defined as

$$\mu(p_t; p_r, x_r) \equiv e\big(p_t, v(p_r, x_r)\big),$$

with $e(p, u)$ the expenditure function quantifying the minimum income required to attain utility u at prices p and $v(p, x)$ the indirect utility function giving the maximum utility level at prices p and income x. In our setup, the vector $q_{t,r}$ represents Hicksian demand at price p_t and utility level u_r, which itself equals $v(p_r, x_r)$. Thus, we can simply write

$$\mu(p_t; p_r, x_r) = e(p_t, u_r) = p_t h(p_t, u_r) = p_t h_{t,r}.$$

Then, using our N-GARP characterization of rationalizable consumer behavior under normal demand, we can define upper (or lower) bounds on $\mu(p_t; p_r, x_r)$ by maximizing (or minimizing) $p_t q_{t,r}$ subject to the conditions in Definition 5. This implies optimization problems with a linear objective and linear inequality constraints that are characterized by integer variables (see also online Appendix II). It

[7] Varian (1982) refers to the money metric utility function as income compensation function. He considers welfare comparisons between price-income situations that are possibly unobserved. In the current paper, our focus is on comparing observed price-income situations. Under specific assumptions regarding unobserved prices, it is fairly easy to extend our following reasoning to welfare comparisons that involve unobserved price-income regimes.

defines an interval set of possible values for $\mu(p_t;p_r,x_r)$ under the given normality assumptions.

In a following step, we can compare the welfare of some evaluated individual in consumption observation t and reference observation r by using the cost-of-living index

$$c_{t,r} = \frac{x_t - \mu(p_t;p_r,x_r)}{x_t} = \frac{x_t - p_t h_{t,r}}{x_t}.$$

In this expression, the numerator $x_t - \mu(p_t;p_r,x_r)$ defines the compensating variation associated with the price change from p_r to p_t. It measures the difference between the individual's potential income in the decision situation t (i.e., x_t) and the income needed by the same individual under the prices p_t to be equally well off, as in the reference situation r (i.e., $\mu(p_t;p_r,x_r)$). To obtain the cost-of-living index $c_{t,r}$, we divide this compensating variation by the available income in observation t. This compares the individual's welfare in t relative to r. If $c_{t,r}$ exceeds zero, the individual is better off in t than in r. Conversely, if $c_{t,r}$ is below zero, the individual is worse off in t than in r.

Similar to before, our nonparametric characterization of rationalizable demand behavior allows us to nonparametrically identify upper and lower bounds on $c_{t,r}$ (using set identification of $\mu(p_t;p_r,x_r)$). These nonparametric bounds apply to *any* utility specification that rationalizes the observed consumption behavior in terms of normal demand. In our empirical application, we will conclude that an individual is better off in situation t than in situation r if the lower bound of $c_{t,r}$ is above zero. It means that for every specification of the individual utilities that rationalizes the observed consumption behavior, we obtain a value for $c_{t,r}$ that exceeds zero. Similarly, we can conclude that the individual is worse off in t than in r if the upper bound of $c_{t,r}$ is below zero. Finally, if the lower and upper bounds have opposite signs, we cannot reject the hypothesis that the individual is equally well off in both decision situations: we are unable to robustly (i.e., for any specification of the rationalizing utilities) conclude that the individual is better or worse off in t than in r.

Goodness-of-fit.—The revealed preference characterization in Definition 5 allows us to define sharp tests for rationalizable consumption behavior: either the data satisfy the testable N-GARP conditions or they do not. When the data do not satisfy these exact conditions, it is often interesting to empirically evaluate the degree of violation. For example, it may happen that the data are close to satisfying the exact rationalizability conditions, and we may want to include such *almost rationalizable* data in our further empirical analysis. To this end, we extend Afriat's (1973) notion of CCEI to our specific setting. Intuitively, the CCEI quantifies the goodness-of-fit of the rationalizability conditions in terms of minimal adjustments of the observed expenditure levels that are needed to exclude violations of the nonparametric rationalizability conditions. In other words, it quantifies the error that must be accounted for such that the (corrected) data satisfy the rationality restrictions.[8]

[8] The CCEI was originally introduced by Afriat (1973) and further developed by Varian (1990). Choi et al. (2014) used the CCEI in a large-scale field experiment as a measure of consumers' decision-making quality. Intuitively, they interpret low CCEI values as revealing optimization errors arising from imperfect decision-making

Formally, to apply the CCEI concept to our N-GARP characterization, we introduce a parameter $e \in [0,1]$. Correspondingly, we adjust the last three (if-then) conditions in Definition 5 for which $r = v$ while keeping the other conditions intact. That is, we only change the conditions for which $h_{r,v}$ is equal to the observed bundle q_v. This obtains the following adapted conditions (for all $r,s,t,v \in T$):

- if $u_t \geq u_v$, then $ep_v q_v \leq p_v h_{s,t}$,
- if $u_t > u_v$, then $ep_v q_v < p_v h_{s,t}$,
- if $u_t \geq u_v$, then $eq_v^i \leq h_{v,t}$ for all $i \in M$.

For a given dataset S, the CCEI equals the highest value of e such that the consumption observations satisfy these adjusted rationalizability conditions.[9] Obviously, higher CCEI values signal a better fit of the rationalizability conditions. Next, as argued by Apesteguia and Ballester (2015, section V), the CCEI has two properties that are specifically attractive from a practical point of view. First, it satisfies continuity, which means that it never increases with the number of observations. Second, the CCEI satisfies rationality, which implies that it equals one if and only if the data are (exactly) rationalizable.

Let e^* represent the CCEI of a given dataset S. Then, we can define the adjusted revealed preference test which, by construction, satisfies the modified N-GARP restrictions in Definition 5. For this adjusted test, we can compute cost-of-living indices by using the nonparametric procedure outlined above. In the following section, our main empirical analysis will do so for the individuals with CCEI values $e^* \geq 0.99$, which means that the observed behavior is sufficiently close to rationalizability.

Predictive Success.—One may be inclined to compare the empirical performance of alternative revealed preference conditions by comparing their pass rate, i.e., the proportion of individuals passing the conditions. However, this practice can be very misleading if one rationalizability condition is structurally weaker than the other. For example, any demand behavior that meets N-GARP will by construction also satisfy GARP (but not vice versa). Thus, the pass rate for N-GARP can never exceed the pass rate for GARP.

In order to solve this issue, one should account for the empirical stringency of the revealed preference conditions. A widely used measure for the power of revealed preference conditions is the so-called Bronars index (Bronars 1987). This Bronars power computes the fraction of (simulated) random datasets that violate the rationalizability conditions subject to testing. A random dataset is then constructed by randomly selecting bundles from each of the observed budget hyperplanes. In general, higher power values reveal more stringent revealed preference conditions. Thus, if one condition is weaker than the other, then its power will also be lower. For example, the power of GARP will never exceed the power of N-GARP.

quality. We may use a similar interpretation of the CCEI results in our empirical application in Section III. See also Apesteguia and Ballester (2015) and Halevy, Persitz, and Zrill (2018) for related discussions.

[9] See online Appendix B for more information concerning the computation of the CCEI.

Selten (1991) suggested to combine the pass rate and power of a given test into a single-dimensional measure of "predictive success," which is computed as

$$\text{predictive success} = \text{pass rate} - (1 - \text{power})$$

and always situated between -1 and 1.[10] A well-performing revealed preference condition has a predictive success measure that is close to one, as this reveals both a high pass rate and high power; many observed individuals pass the test while almost no random behavior passes the test. A predictive success measure below zero implies that the pass rate for the randomly generated data exceeds the one for the observed data. This indicates the—obviously undesirable—situation that the revealed preference condition fits random behavior better than actual behavior. In principle, the higher the measure of predictive success, the better the empirical fit of the demand model that is tested. Demuynck (2015) introduced statistical tests for differences in predictive success associated with alternative behavioral models. We will use these statistical tools in our following application.

III. Illustrative Application

To evaluate the welfare effects of the 2008 financial crisis, we make use of a balanced panel drawn from the 2007, 2009, and 2011 waves of the PSID. By considering only three PSID waves, we can show that our methodology enables an informative empirical analysis even for short time series of consumption observations.[11] Moreover, it seems more reasonable to assume stable individual preferences over a shorter time period. In online Appendix IV, we demonstrate the robustness of our main qualitative conclusions for a longer panel containing four consumption observations per individual (adding the 2013 PSID wave to our original dataset). This extra analysis also allows us to study the impact of the crisis over a longer time period.

Data and Setup.—The PSID, which was initiated in 1968, is a widely used survey of a national representative sample of 18,000 individuals living in 5,000 families in the United States. The dataset contains information on income, wealth, health, marriage, childbearing, child development, education, and other sociodemographic variables. Since 1999, the panel also provides additional expenditure information on a detailed set of consumption categories (see Blundell, Pistaferri, and Saporta-Eksten 2016 for more details).

Our empirical analysis specifically focuses on the welfare effects of the 2008 crisis for singles (with and without children). Thus, we exclude couples from our investigation, which also conveniently avoids preference aggregation issues associated with

[10] Selten's measure was popularized for revealed preference tests by Beatty and Crawford (2011).

[11] In principle, it is possible to use our methodology with only two consumption observations per individual. However, it can be shown that in such a case, the N-GARP-based lower bounds on the cost-of-living indices always equal the GARP-based lower bounds by construction. Thus, by using three consumption observations per individual, we can illustrate the usefulness of normality assumptions for obtaining lower bounds that are more informative than the GARP-based bounds.

the welfare analysis of multiperson households.[12] We concentrate on individuals who are situated on the intensive margin of labor supply; that is, our subjects are actively working on the labor market in each period under study. We excluded the self-employed to avoid issues regarding the imputation of wages and the separation of consumption from work-related expenditures. After excluding observations with missing information (e.g., on wages, labor hours, or consumption expenditures), we end up with a sample of 821 individuals.

Table 1 in online Appendix III reports summary statistics for our sample. We assume that individuals spend their full potential income on four consumption categories: food, housing, leisure, and other goods. We compute leisure quantities by assuming that each individual needs eight hours per day for personal care and sleep. Leisure equals the available time that could have been spent on market work but was not (i.e., leisure per week $= (24 - 8) \times 7 -$ market work). We calculate the individuals' weekly expenditures (i.e., nominal dollars per week) on the three remaining consumption categories (food, housing, and other goods) as the reported annual expenditures divided by 52. The price of leisure equals the individual's hourly wage for market work. The prices of food, housing, and other goods are region-specific consumer price indices that have been constructed by the Bureau of Labor Statistics.

For our empirical analysis, we take it that the normality assumption is arguably debatable for leisure. Therefore, our following analysis will focus on two alternative scenarios: a first one in which we assume normality for all four goods (i.e., N-GARP(4)) and a second one in which we only assume normality for the consumption categories food, housing, and other goods (i.e., N-GARP(3)). We effectively do believe it plausible that the nonleisure expenditures are normal, all the more because they pertain to aggregate consumption categories. We will conduct a goodness-of-fit analysis (using the CCEI) as well as a welfare analysis (on the basis of cost-of-living indices) for the N-GARP conditions associated with our normality assumptions. We will compare (in terms of predictive success) our two N-GARP models with the GARP model that makes no use of any normality assumption (recalling that N-GARP reduces to GARP if no good is assumed to be normal).

In our following exercises, we will conduct separate N-GARP-based and GARP-based analyses for all 821 individuals whom we observe. Using our notation of Section II, this defines a dataset S with 3 observations (i.e., $T = \{2007, 2009, 2011\}$) and 4 goods (i.e., $n = 4$) for every single in our sample. By analyzing each individual separately, we fully account for preference heterogeneity across individuals.

Goodness-of-fit.—We begin by using Afriat's CCEI to check data consistency with N-GARP and GARP for the sample of singles under study. Basically, the GARP-based CCEI results reveal how well the assumption of utility maximization fits the observed

[12] Practical welfare analysis of multiperson households often adopts a unitary assumption, which models these households as single decision-makers. However, this unitary assumption has been rejected by a large number of empirical studies (see, for example, Browning and Chiappori 1998 and Dauphin et al. 2011). This suggests the extension of our analysis toward collective household models, with multiperson households consisting of multiple decision-makers, as an interesting avenue of follow-up research. Such an extension can build on Cherchye, De Rock, and Vermeulen (2007, 2011), who developed the revealed preference characterization of rational consumption for collective households.

TABLE 1—CCEI

	N-GARP(3)	N-GARP(4)	GARP
CCEI $=$ 1 (percent)	587 (71.50)	424 (51.64)	782 (95.25)
CCEI $\geq$ 0.99 (percent)	702 (85.51)	595 (72.47)	803 (97.81)
Mean	0.9913	0.9817	0.9987
Standard deviation	0.0296	0.0435	0.0124
Min	0.6774	0.6047	0.7451
25 percent	0.9980	0.9874	1.0000
50 percent	1.0000	1.0000	1.0000
75 percent	1.0000	1.0000	1.0000
Max	1.0000	1.0000	1.0000

behavior, while the N-GARP-based CCEI results indicate the empirical fit of our normality assumptions in addition to utility maximization. As explained in Section II, the CCEI evaluates model fit in terms of necessary adjustments of observed expenditures to obtain data consistency with the (N-GARP and GARP) rationalizability conditions that are subject to evaluation. CCEI values are situated between zero and one, with higher values signaling a better fit.

Table 1 summarizes our CCEI results. The first row shows the number of individuals who satisfy the exact N-GARP and GARP conditions (corresponding to CCEI $=$ 1). The second row reports the number of individuals who are very close to rationalizability (characterized by CCEI $\geq$ 0.99). Generally, the CCEI values for the N-GARP conditions are below the CCEI values for the GARP condition. This should not be surprising because, as explained above, the N-GARP conditions are more stringent than the GARP condition. Importantly, we find that the average CCEI value is very high for both the N-GARP and GARP tests: it equals 0.9913 for N-GARP(3), 0.9817 for N-GARP(4), and 0.9987 for GARP. However, we also observe that the behavior of some individuals turns out to be quite far from exact rationalizability. For example, the minimum CCEI value equals 0.6774 for N-GARP(3), 0.6047 for N-GARP(4), and 0.7451 for GARP.

Next, when comparing our findings for the N-GARP(3) and N-GARP(4) conditions in Table 1, we observe that the N-GARP(3) model provides a better fit. Once more, this is actually not surprising as the models are nested; the N-GARP(4) model imposes stronger normality restrictions than the N-GARP(3) model. From now on, however, we mainly focus on the N-GARP(3) setting where we assume normality only for the consumption categories food, housing, and other goods. The reason for this is that (i) a priori the assumption of leisure being a normal good is often debated, and (ii) by switching from N-GARP(3) to N-GARP(4), we lose around 100 more observations, while the improvement in tightness of the bounds is not significantly improved (see online Appendix IV).

Overall, the results in Table 1 provide rather strong empirical support for N-GARP (as well as GARP) applied to our sample of individuals. In most cases, we need only (very) small expenditure adjustments to obtain consistency with the rationalizability conditions. In our following welfare analysis, we will focus on the subsamples of, respectively, 702 and 595 individuals with N-GARP(3)-based and N-GARP(4)-based CCEI values greater than or equal to 0.99. As explained above, such

Table 2—Predictive Success Measures

	CCEI $=$ 1	CCEI $\geq$ 0.99
N-GARP(4)	0.1784 [0.1442, 0.2126]	0.3207 [0.2901, 0.3513]
N-GARP(3)	0.2550 [0.2241, 0.2859]	0.3131 [0.2890, 0.3372]
GARP	0.1405 [0.1259, 0.1551]	0.1171 [0.1071, 0.1271]
H_1	p-value	p-value
N-GARP(3) $>$ GARP	0.0000	0.0000
N-GARP(4) $>$ GARP	0.0143	0.0000
N-GARP(3) $>$ N-GARP(4)	0.0000	0.7411

high CCEI values signal behavior that is very close to exactly rationalizable, which empirically motivates using the assumption of rationality (with normal demand) for our welfare analysis. Online Appendix IV contains a robustness analysis that only includes the (587) individuals with N-GARP-based CCEI equal to one (i.e., exactly rationalizable behavior). Comfortingly, this additional analysis yields the same main findings.

Predictive Success.—The top part of Table 2 presents the predictive success measures for the various revealed preference conditions that are subject to evaluation. We consider rationalizability tests with CCEI equal to 1 and with CCEI at least 0.99. We also report (between square brackets) 95 percent asymptotic confidence intervals for the predictive success measures (obtained through the method of Demuynck 2015). Reassuringly, we find that all three rationalizability conditions (GARP, N-GARP(3), and N-GARP(4)) have a predictive success that is significantly above zero.

The bottom part of Table 2 provides results on hypotheses tests regarding differences in predictive success for the behavioral models under consideration. We test the null hypothesis of equal predictive success against alternative inequality hypotheses. Our results indicate that both the N-GARP(3) and N-GARP(4) models significantly outperform the GARP model in terms of predictive success. We also check whether the N-GARP(3) model performs better than the N-GARP(4) model. Interestingly, we do find that the hypothesis of equal empirical success is rejected against this alternative hypothesis when considering CCEI equal to one. However, this conclusion no longer holds when focusing on the slightly relaxed setting with CCEI at least 0.99.

Cost-of-living Indices.—We quantify the welfare effects of the 2008 crisis by calculating cost-of-living indices. For each individual in our sample, we estimate the difference in cost of living between 2007 and 2011. More formally, we define this as the difference between the actual income in 2011 and the income that would be required in the same year (at 2011 prices) to be equally well off as in 2007:

$$c_{2011,2007} = \frac{x_{2011} - p_{2011} h_{2011,2007}}{x_{2011}}.$$

TABLE 3—BOUNDS ON $c_{2011,2007}$ FOR THE N-GARP(3) SUBSAMPLE (702 INDIVIDUALS)

	N-GARP(3)-based			GARP-based			
	Lower	Upper	Δ_n	Lower	Upper	Δ_g	$\frac{\Delta_g - \Delta_n}{\Delta_g}$
Mean	−0.037	0.033	0.070	−0.038	0.107	0.144	0.469
Standard deviation	0.288	0.247	0.131	0.288	0.279	0.168	0.391
Min	−3.044	−2.492	0.000	−3.044	−2.489	0.000	0.000
25 percent	−0.120	−0.042	0.008	−0.120	0.000	0.042	0.011
50 percent	−0.005	0.008	0.029	−0.006	0.040	0.094	0.500
75 percent	0.083	0.131	0.076	0.083	0.255	0.193	0.865
Max	0.830	0.897	2.099	0.830	0.899	2.285	1.000

TABLE 4—BOUNDS ON $c_{2011,2007}$ FOR THE N-GARP(4) SUBSAMPLE (595 INDIVIDUALS)

	N-GARP(4)-based			GARP-based			
	Lower	Upper	Δ_n	Lower	Upper	Δ_g	$\frac{\Delta_g - \Delta_{n1}}{\Delta_g}$
Mean	−0.036	0.007	0.043	−0.038	0.114	0.153	0.758
Standard deviation	0.263	0.227	0.117	0.264	0.253	0.176	0.244
Min	−3.044	−1.624	0.000	−3.044	−1.578	0.002	0.000
25 percent	−0.124	−0.084	0.003	−0.124	0.000	0.043	0.636
50 percent	−0.006	0.003	0.012	−0.007	0.046	0.099	0.835
75 percent	0.084	0.113	0.041	0.083	0.260	0.200	0.951
Max	0.831	0.897	2.099	0.830	0.899	2.285	1.000

We use the nonparametric set identification procedure outlined above. Particularly, we compute GARP-based and N-GARP-based lower and upper bound on $c_{2011,2007}$ by using the rationalizability restrictions associated with GARP (in Definition 3) and N-GARP (in Definition 5), respectively. As explained above, we focus on subsamples of "almost rational" individuals with an N-GARP-based CCEI value at least equal to 0.99. These subsamples contain 702 individuals for the N-GARP(3) model and 595 individuals for the N-GARP(4) model.

Tables 3 and 4 give a summary of our results for the sample of individuals under study. Columns 2–7 summarize our N-GARP-based bounds and columns 8–10 our GARP-based bounds. Correspondingly, Δ_n in column 4 and Δ_g in column 7 represent the differences between the respective upper and lower bounds. Finally, column 8 reports on the relative difference between Δ_n and Δ_g. This measures the extent to which the N-GARP-based bounds are tighter than the GARP-based bounds. In a sense, it quantifies the identifying power that specifically follows from our normality assumptions.

We observe that both the N-GARP(3)-based and the N-GARP(4)-based bounds are substantially tighter than the GARP-based bounds. The mean (respectively, median) differences between the N-GARP-based lower and upper bounds are 7 percent and 4.3 percent (respectively, 2.9 percent and 1.2 percent) for the N-GARP(3) and N-GARP(4) subsamples, which is much below the differences of 14.4 percent and 15.3 percent (respectively, 9.4 percent and 9.9 percent) between the GARP-based bounds for the same subsamples. Moreover, the relative difference

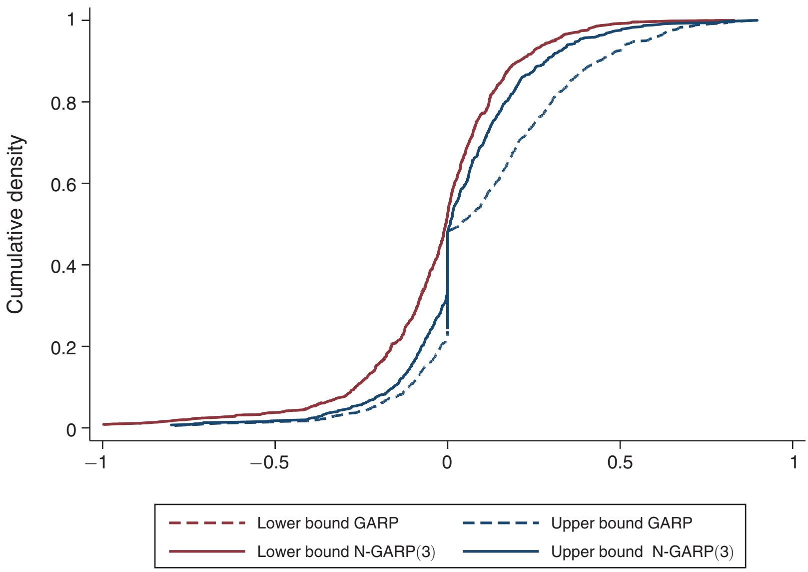

Figure 3. CDF of N-GARP(3)-based Bounds

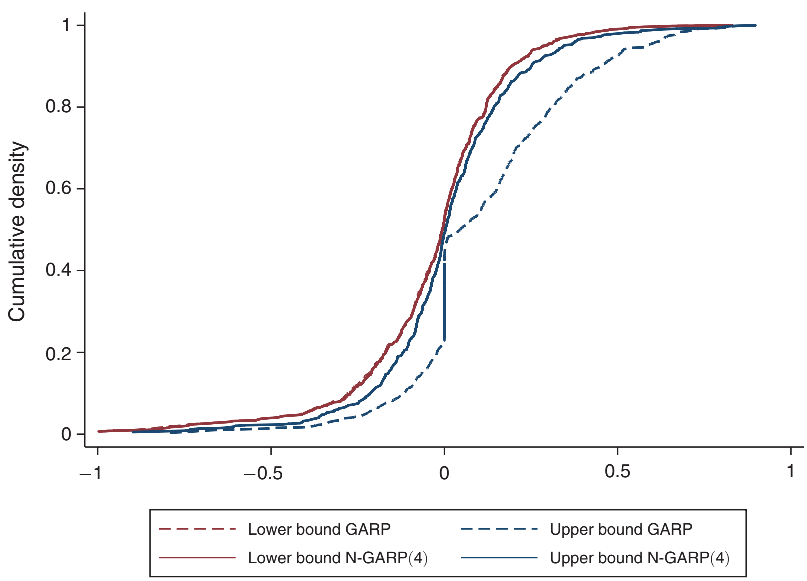

Figure 4. CDF of NGARP(4)-based Bounds

between Δ_n and Δ_g amounts to no less than 50 percent for about half of our sample, again showing a significant increase of identifying power when imposing normality.

As a following exercise, Figures 3 and 4 depict the empirical cumulative distribution functions (CDFs) of our N-GARP-based and GARP-based lower and upper bounds for $c_{2011,2007}$. In line with our results in Tables 3 and 4, the N-GARP-based CDFs are much closer to each other than the GARP-based CDFs. From all this, we

TABLE 5—WORSE-OFF AND BETTER-OFF INDIVIDUALS FOR THE N-GARP(3) SUBSAMPLE

		N-GARP(3)	GARP
UB $<$ 0	Worse-off in 2011	33.05	22.36
LB $>$ 0	Better-off in 2011	47.86	47.58
LB $\leq$ 0 and 0 $\leq$ UB	Cannot say	19.09	30.06

TABLE 6—WORSE-OFF AND BETTER-OFF INDIVIDUALS FOR THE N-GARP(4) SUBSAMPLE

		N-GARP(4)	GARP
UB $<$ 0	Worse-off in 2011	49.08	22.86
LB $>$ 0	Better-off in 2011	47.90	47.56
LB $\leq$ 0 and 0 $\geq$ UB	Cannot say	3.03	29.58

may safely conclude that our (mild) normality assumptions do yield a considerably more informative welfare analysis. Further inspection of Tables 3 and 4 and Figures 3 and 4 reveals that for our specific data, this improvement in identifying power is mostly driven by lower upper bounds (and to a lesser degree by higher lower bounds).

Better-off and Worse-off Individuals.—As explained in Section II, we can state that an individual is better off in 2011 than in 2007 if the lower bound of $c_{2011,2007}$ (LB) exceeds zero, while the individual is worse off in 2011 if the upper bound of $c_{2011,2007}$ (UB) is below zero. These better-off and worse-off classifications are robust in that they hold for any specification of the individual utilities that rationalize the observed consumption behavior. Finally, if the lower and upper bounds have opposite signs (i.e., LB $\leq$ 0 and UB $\geq$ 0), we cannot robustly conclude that the individual is better- or worse-off in 2011.

Rows 2–4 of Tables 5 and 6 give the fractions of individuals who are classified as better off, worse off, and cannot say according to our N-GARP-based (column 3) and GARP-based (column 4) bounds for $c_{2011,2007}$. Using our N-GARP(3)-based and N-GARP(4)-based bounds, we classify respectively 33.05 percent and 49.08 percent of our individuals as worse off and 47.86 percent and 47.90 percent of the individuals as better off, with a residual 19.09 percent and 3.03 percent falling in the cannot say category. By contrast, our GARP-based bounds classify only 22.36 percent (N-GARP(3) subsample) and 22.86 percent (N-GARP(4) subsample) of the individuals as worse off and, respectively, 47.58 percent and 47.56 percent as better off, now leaving about 30 percent of the individuals in the cannot say category. We see that particularly, the fraction of individuals in the worse-off category is substantially higher in the N-GARP-based analyses than in the GARP-based analysis. Correspondingly, the fraction of individuals in the cannot say category is lower in the N-GARP-based classifications than in the GARP-based classification. These findings show that using normality assumptions obtains a significantly more informative classification of individuals after the 2008 crisis. Particularly, the N-GARP restrictions for rational behavior enable a considerably better identification of the individuals who suffered from a welfare loss after the 2008 crisis.

Overall, Tables 5 and 6 provide further support for our earlier conclusion that (mild) normality assumptions can substantially improve the informative value of nonparametric welfare analysis. Moreover, our cost of living estimates reveal considerable some heterogeneity across individuals. In online Appendix IV, we investigate this further by relating these cost of living estimates to observable individual characteristics. A main finding is that individuals with higher potential incomes in 2007 have been hit more severely by the crisis.[13] Next, we also observe that having children correlates significantly with our estimated welfare effects. At this point, it is worth recalling that our empirical analysis considers singles who remained employed after the crisis. This contrasts with existing studies, which mainly focused on the extensive margin of labor supply.

IV. Conclusion

We presented a revealed preference characterization of rational consumer behavior under the assumption of normal demand. The characterization is easily operationalized in practice, and it is flexible in that it can impose normality on any subset of goods. We have also shown the use of our characterization to analyze the welfare effects (in terms of cost-of-living indices) of changing price-income regimes. As normality is often a plausible assumption to make, this provides a useful tool kit to remediate the lack of power that is frequently associated with empirical revealed preference analysis.

We used our novel methodology to evaluate the welfare impact of the 2008 financial crisis for individuals situated on the intensive margin of labor supply. Particularly, we studied the labor supply behavior of a sample of singles drawn from the PSID. Our main focus was on comparing the goodness-of-fit and identifying power of our nonparametric characterization of utility maximization, with and without normality assumptions. We found that the goodness-of-fit results were hardly affected when imposing normality, providing good empirical support for our normality hypotheses. Next, and more importantly, we showed that using mild normality assumptions yields a substantially more powerful empirical welfare analysis: it obtained considerably sharper set identification of individuals' cost-of-living indices and a significantly more informative classification of better-off and worse-off individuals after the 2008 crisis.

REFERENCES

Afriat, Sydney N. 1967. "The construction of utility functions from expendituredata." *International Economic Review* 8 (1): 67–77.

Afriat, S.N. 1973. "On a System of Inequalities in Demand Analysis: An Extension of the Classical Method." *International Economic Review* 14 (2): 460–72.

[13] This is mainly driven by the fact that in our sample, people with higher initial wages suffered from more severe wage drops after the crisis. See our discussion of Table 8 in online Appendix IV.

Apesteguia, Jose, and Miguel A. Ballester. 2015. "A Measure of Rationality and Welfare." *Journal of Political Economy* 123 (6): 1278–1310.

Banks, James, Richard Blundell, and Arthur Lewbel. 1997. "Quadratic Engel Curves and Consumer Demand." *The Review of Economics and Statistics* 79 (4): 527–39.

Beatty, Timothy K.M., and Ian A. Crawford. 2011. "How Demanding Is the Revealed Preference Approach to Demand?" *American Economic Review* 101 (6): 2782–95.

Blundell, Richard, Martin Browning, Laurens Cherchye, Ian Crawford, Bram De Rock, and Frederic Vermeulen. 2015. "Sharp for SARP: Nonparametric Bounds on Counterfactual Demands." *American Economic Journal: Microeconomics* 7 (1): 43–60.

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

Blundell, Richard, Martin Browning, and Ian Crawford. 2007. "Improving Revealed Preference Bounds on Demand Responses." *International Economic Review* 48 (4): 1227–44.

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

Blundell, Richard, Luigi Pistaferri, and Itay Saporta-Eksten. 2016. "Consumption Inequality and Family Labor Supply." *American Economic Review* 106 (2): 387–435.

Bronars, Stephen G. 1987. "The Power of Nonparametric Tests of Preference Maximization." *Econometrica* 55 (3): 693–98.

Browning, M., and P.A. Chiappori. 1998. "Efficient Intra-Household Allocations: A General Characterization and Empirical Tests." *Econometrica* 66 (6): 1241–78.

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

Cherchye, Laurens, Bram De Rock, and Frederic Vermeulen. 2011. "The Revealed Preference Approach to Collective Consumption Behaviour: Testing and Sharing Rule Recovery." *The Review of Economic Studies* 78 (1): 176–98.

Cherchye, Laurens, Thomas Demuynck, and Bram De Rock. 2018. "Normality of Demand in a Two-Goods Setting." *Journal of Economic Theory* 173 : 361–82.

Choi, Syngjoo, Shachar Kariv, Wieland Müller, and Dan Silverman. 2014. "Who Is (More) Rational?" *American Economic Review* 104 (6): 1518–50.

Christensen, Laurits R, Dale W Jorgenson, and Lawrence J Lau. 1975. "Transcendental Logarithmic Utility Functions." *American Economic Review* 65 (3): 367–83.

Dauphin, Anyck, Abdel-Rahmen El Lahga, Bernard Fortin, and Guy Lacroix. 2011. "Are Children Decision-Makers within the Household?" *Economic Journal* 121 (553): 871–903.

Deaton, Angus. 2012. "The Financial Crisis and the Well-Being of Americans: 2011 OEP Hicks Lecture." *Oxford Economic Papers* 64 (1): 1–26.

Deaton, Angus, and John Muellbauer. 1980. "An Almost Ideal Demand System." *The American Economic Review* 70 (3): 312–26.

Demuynck, Thomas. 2015. "Statistical Inference for Measures of Predictive Success." *Theory and Decision* 79 (4): 689–99.

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

Fisher, Franklin M. 1990. "Normal Goods and the Expenditure Function." *Journal of Economic Theory* 51 (2): 431–33.

Goodman, Christopher J., and Steven M. Mance. 2011. "Employment Loss and the 2007–09 Recession: An Overview." *Bureau of Labor Statistics Monthly Labor Review* 134: 3–12.

Halevy, Yoram, Dotan Persitz, and Lanny Zrill. 2018. "Parametric Recoverability of Preferences." *Journal of Political Economy* 126 (4): 1558–93.

Houthakker, Hendrik S. 1950. "Revealed Preference and the Utility Function." *Economica* 17 (66): 159–74.

Hurd, Michael D., and Susann Rohwedder. 2010. "Effects of the Financial Crisis and Great Recession on American Households." NBER Working Paper 16407.

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

Samuelson, Paul A. 1938. "A Note on the Pure Theory of Consumer's Behaviour." *Economica* 5 (17): 61-71.

Samuelson, Paul A. 1974. "Complementarity: An Essay on the 40th Anniversary of the Hicks-Allen Revolution in Demand Theory." *Journal of Economic Literature* 12 (4): 1255–89.

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

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

Varian, Hal R. 1984. "The Nonparametric Approach to Production Analysis." *Econometrica* 52 (3): 579–97.

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

Verick, Sher. 2009. "Who Is Hit Hardest During a Financial Crisis? The Vulnerability of Young Men and Women to Unemployment in an Economic Downturn." IZA Discussion Paper 4359.