# Revealed preference analysis with normal goods: Application to cost of living indices - online appendix

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

## I. Proof of Proposition 2

**Proof.** We begin by showing necessity of our N-GARP conditions in Definition 5, i.e. any observed demand originating from utility maximization under normality must satisfy the conditions in Proposition 2. In a following step, we show sufficiency of the N-GARP conditions by using the auxiliary results stated in Lemmata 1, 2 and 3 below.

Necessity. — Let $S = (p_t, q_t)_{t \in T}$ be rationalizable under normal demand (on the set $M \subseteq \{1, ..., n\}$) by the utility function $u : \mathbb{R}^n_+ \to \mathbb{R}$ and expansion paths $q_t : \mathbb{R}_+ \to \mathbb{R}^n_+$ that are monotone and continuous in $x$ for all goods $i \in M$ and such that $q_t(x_t) = q_t$ for $x_t = p_t q_t$.

For all $t \in T$, define $u_t \equiv u(q_t)$ and, for all $t, v \in T$, define $h_{t,v}$ as the bundle on the intersection of the expansion path $q_t(x)$ and the indifference curve through $q_v$, i.e. $h_{t,v}$ represents the Hicksian demand bundle $h(p_t, u_v)$. Given that the utility function $u(.)$ and the expansion paths $q_t(.)$ are continuous and monotone, this bundle is unique. By definition, we have that the intersection of $q_t(x)$ with the indifference curve through $q_t$ is $q_t$. This gives the first N-GARP condition in Definition 5, i.e. $h_{t,t} = h(p_t, u_t) = q_t$ for all $t \in T$.

We know that $h_{t,v} \equiv h(p_t, u_v)$ solves the corresponding expenditure minimization problem

$$e(p_t, u_v) = \min_h p_t h \text{ s.t. } u(h) \geq u_v.$$

For the second N-GARP condition, let $u_t \geq u_v$ and assume (towards a contradiction) that $p_r h_{r,v} > p_r h_{s,t}$. This means that

$$p_r h_{r,v} = p_r h(p_r, u_v) = e(p_r, u_v) > p_r h_{s,t} = p_r h(p_s, u_t).$$

Given that $h(p_r, u_v)$ is expenditure minimizing at utility level $u_v$ and prices $p_r$, this requires that $u_v > u_t$. Indeed, if this were not the case, then it would have been less expensive to buy $h_{s,t}$ instead of $h_{r,v}$ and still attain at least the same utility level. This is a contradiction, which implies $p_r h_{r,v} \leq p_r h_{s,t}$. We can derive the third N-GARP condition in a directly similar way.

Finally, for the fourth N-GARP condition, we observe that, if $u_t \geq u_v$, then we obtain that $h^i_{r,t} = h^i(p_r, u_t) \geq h^i(p_r, u_v) = h^i_{r,v}$, because the Hicksian demand functions for $i \in M$ are monotone in utility.

SUFFICIENCY. — Suppose the data set $S = \{(p_t, q_t)\}_{t \in T}$ is consistent with the N-GARP conditions in Definition 5 (for the set $M = \{1, ..., n\}$). We want to construct a utility function $u : \mathbb{R}^n_+ \to \mathbb{R}$ and expansion paths $q_t : \mathbb{R}_+ \to \mathbb{R}^n_+$ (which are monotone in $x$ for each good $i \in M$) that generate the observed demand.

Our result is based on an application of Proposition 1, which is taken from Nishimura, Ok and Quah (2017):

PROPOSITION 1 (Nishimura, Ok and Quah): *Let $(q_t(.))_{t \in T}$ be a set of continuous expansion paths (i.e. $q_t : \mathbb{R}_+ \to \mathbb{R}^n_+$ are continuous functions such that, for all $x \in \mathbb{R}_+ : p_t q_t(x) = x$). Then, the following equivalence holds:*
*There exists a continuous and monotone utility function $u : \mathbb{R}^n_+ \to \mathbb{R}$ such that, for all $t \in T$ and $x \in \mathbb{R}_+$,*

$$q_t(x) \in \arg\max_q u(q) \text{ s.t. } p_t q \leq x$$

*if and only if,*
*for all $N \in \mathbb{N}$, all sequences of income values $x_1, \ldots, x_N$ in $\mathbb{R}_+$ and all sequences of observations $t_1, \ldots t_N \in T$, the data sets $(p_n, q_t(x_n))_{n \leq N}$ satisfy GARP.*

Let $(u_t, h_{t,v})_{t,v \in T}$ be the solution of the N-GARP restrictions. The idea is to construct income expansion paths $q_t(x)$ that satisfy the condition of Proposition 1 above. A straightforward idea would be to define $q_t(x)$ by taking a linear interpolation between the various bundles $(h_{t,r})_{r \in T}$. A potential problem with this approach, however, is that the solution to the N-GARP conditions may set $u_s = u_r$ for different observations $r, s \in T$. This means that our expansion path would contain two potentially distinct bundles on the same (counterfactual) indifference curve, which would violate the assumption that $q_t(x)$ is a function.

Given this potential issue, the proof takes three steps. In a first step, we show that feasibility of the N-GARP restrictions is equivalent to feasibility of a similar set of restrictions where all utility values $u_t$ are distinct. In Step 2, we use linear interpolation to define, for each observation $t \in T$, an increasing and continuous income expansion path $q_t(.)$ through the observed bundle $q_t$. Finally, Step 3 shows that these expansion paths satisfy the condition of Proposition 1 above.

**Step 1:** For the ease of interpretation, we separate the indices attached to the utility values from the indices attached to the prices and quantities. To this end, we define $T_u \equiv T$ and $T_p \equiv T$. Let $(u_v, h_{t,v})_{t \in T_p, v \in T_u}$ solve the N-GARP restrictions for the given data set $S = (p_t, q_t)_{t \in T_p}$. Observe that feasibility of N-GARP is equivalent to feasibility of the following problem, which we call $FP(T_u, S, \rho)$ (for $\rho : T_p \to T_u$ defined as $\rho(t) = t$).

PROGRAM 1 ($FP(T_u, S, \rho)$): *There exist numbers* $(u_t)_{t \in T_u}$ *and vectors* $(h_{t,v})_{t \in T_p, v \in T_u}$ ($h_{t,v} \in \mathbb{R}^n_+$) *such that*

1) $\forall t \in T_p$: $h_{t,\rho(t)} = q_t$,
2) $\forall t, v \in T_u, \forall r, s \in T_p$: *if* $u_t \geq u_v$, *then* $p_r h_{r,v} \leq p_r h_{s,t}$,
3) $\forall t, v \in T_u, \forall r, s \in T_p$: *if* $u_t > u_v$, *then* $p_r h_{r,v} < p_r h_{s,t}$,
4) $\forall t, v \in T_u, \forall r \in T_p, \forall i \in M$: *if* $u_t \geq u_v$, *then* $h^i_{r,v} \leq h^i_{r,t}$.

If this problem gives a solution with $u_t = u_v$ for some $t, v \in T_u$ such that $t \neq v$, we can apply Lemma 1 below to show that there exists a solution for the problem $FP(T'_u, S, \rho')$ where $T'_u = T_u - \{v\}$ and

$$\rho'(i) = \begin{cases} \rho(i) & \text{if } i \neq v, \\ t & \text{if } i = v \end{cases}$$

We can repeat this argument $n$ times until $u_t \neq u_v$ for all indices $t, v \in T_u^{(n)}$. In turn, this leads us to define the following feasibility problem.

PROGRAM 2 ($FP(T_u^{(n)}, S, \rho^{(n)})$): *There exist distinct numbers* $(u_t)_{t \in T_u^{(n)}}$ *and vectors* $(h_{t,v})_{t \in T_p, v \in T_u^{(n)}}$ ($h_{t,v} \in \mathbb{R}^n_+$) *such that*

1) $\forall t \in T_p$: $h_{t,\rho^{(n)}(t)} = q_t$,
2) $\forall t, v \in T_u, \forall r, s \in T_p$, *if* $u_t > u_v$, *then* $p_r h_{r,v} < p_r h_{s,t}$,
3) $\forall t, v \in T_u, \forall r \in T_p, \forall i \in M$, *if* $u_t \geq u_v$, *then* $h^i_{r,v} \leq h^i_{r,t}$.

Let $|T_u^{(n)}| = R$ and, for notational convenience, let us re-index the elements of the set $T_u^{(n)}$ to obtain the set $\{1, \ldots, R\}$ such that

$$u_1 < u_2 < \ldots < u_R.$$

Step 2 will start from a solution $(u_v, h_{t,v})_{v \leq R, t \in T_p}$ as obtained from this last problem.

**Step 2:** We construct piecewise linear expansion paths $q_t(x)$ in the following way:

- If $x > p_t h_{t,R}$, then $q_t(x) \equiv \gamma h_{t,R}$ with $\gamma = \frac{x}{p_t h_{t,R}}$.
  We say that $q_t(x)$ is of level $R + 1$. Observe that $p_t q_t(x) = x$.

- If $x \leq p_t h_{t,1}$, then $q_t(x) \equiv \gamma h_{t,1}$ with $\gamma = \frac{x}{p_t h_{t,1}}$.
  We say that $q_t(x)$ is of level 1. Again, observe that $p_t q_t(x) = x$.

- If $p_t q_{t,1} < x \leq p_t h_{t,R}$, then the ordering of the observations and the second condition of $FP(T_u^{(n)}, S, \rho^{(n)})$ above imply that there exists a unique $v \leq R$ such that $p_t h_{t,v-1} < x \leq p_t h_{t,v}$. As such, there exists a unique $\alpha \in (0,1]$ such that
$$x = \alpha(p_t h_{t,v}) + (1-\alpha)(p_t h_{t,v-1}).$$
  Given this $\alpha \in (0,1]$, define
$$q_t(x) \equiv \alpha h_{t,v} + (1-\alpha) h_{t,v-1}.$$
  In this case, we will say that $q_t(x)$ is of level $v$. Also, $p_t q_t(x) = x$.

Observe that, for all goods $i \in M$, the path $q_t^i(x)$ is monotone in $x$. In addition, the expansion path is piecewise linear and, therefore, continuous. Moreover, the expansion path $q_t(x)$ contains all bundles $(h_{t,v})_{v \leq R}$ and, thus, also the observed bundle $q_t$.

**Step 3:** We need to show that, for any $N \in \mathbb{N}$, any sequence of income levels $x_1, x_2, \cdots, x_N$ and any sequence of observations $t_1, \ldots, t_N \in T$, the set $(p_{t_i}, q_{t_i}(x_i))_{i \leq N}$ satisfies GARP. Suppose (towards a contradiction) that the result does not hold. Then, there is a $N \in \mathbb{N}$, a sequence $x_1, x_2, \cdots, x_N$ of income levels, and a sequence $t_1, t_2, \cdots, t_N$ of observations that violate GARP. That is,
$$\begin{array}{rcl} p_{t_1} q_{t_1}(x_1) & \geq & p_{t_1} q_{t_2}(x_2), \\ p_{t_2} q_{t_2}(x_2) & \geq & p_{t_2} q_{t_3}(x_3), \\ & \vdots & \\ p_{t_N} q_{t_N}(x_N) & \geq & p_{t_N} q_{t_1}(x_1), \end{array}$$
with at least one strict inequality. From Lemma 2, we know that the level of the bundles (as defined above) along the cycle cannot increase. Also, it cannot strictly decrease as this would mean that somewhere along the cycle it must strictly increase. This implies that the level of all bundles should be the same, say $r$. We distinguish three cases for $r$:

- If $r = R+1$, then there are $\gamma_1, \dots, \gamma_N$ such that

$$\begin{aligned} &q_{t_1}(x_1) = \gamma_1 h_{t_1,R}, \\ &q_{t_2}(x_2) = \gamma_2 h_{t_2,R}, \\ &\dots \\ &q_{t_N}(x_N) = \gamma_N h_{t_N,R}. \end{aligned}$$

  By Lemma 3, we have $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_n \geq \gamma_N \geq \gamma_1$ with at least one strict inequality, a contradiction.

- If $r = 1$, then there are $\gamma_1, \dots, \gamma_N$ such that

$$\begin{aligned} &q_{t_1}(x_1) = \gamma_1 h_{t_1,1}, \\ &q_{t_2}(x_2) = \gamma_2 h_{t_2,1}, \\ &\dots \\ &q_{t_N}(x_N) = \gamma_N h_{t_N,1}. \end{aligned}$$

  Again, by Lemma 3, we have $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_N \geq \gamma_1$, with at least one strict inequality, a contradiction.

- If $1 < r < R+1$, then there are $\alpha_1, \dots, \alpha_N \in (0,1]$ such that

$$\begin{aligned} &q_{t_1}(x_1) = \alpha_1 h_{t_1,r} + (1-\alpha_1) h_{t_1,r-1}, \\ &q_{t_2}(x_2) = \alpha_2 h_{t_2,r} + (1-\alpha_2) h_{t_2,r-1}, \\ &\dots \\ &q_{t_N}(x_N) = \alpha_N h_{t_N,r} + (1-\alpha_N) h_{t_N,r-1}. \end{aligned}$$

  By Lemma 3, we have $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_N \geq \alpha_1$, with at least one strict inequality, a contradiction.

Thus, we conclude that, for any $N \in \mathbb{N}$, any sequence $x_1, x_2, \cdots, x_N$ of income levels and any sequence $t_1, t_2, \cdots, t_N$ of observations, the set $(p_{t_i}, q_{t_i}(x_i))_{i \leq N}$ satisfies GARP. Then, Proposition 3 implies that there exists a continuous and strictly increasing utility function that rationalizes our constructed expansion paths. ■

LEMMA 1: *Let $T_u$ be a finite index set, let $S = (p_t, q_t)_{t \in T_p}$ be a data set and let $\rho : T_p \to T_u$. Then, the problem $FP(T_u, S, \rho)$ has a solution with $u_k = u_j$ if and only if $FP(T_u - \{j\}, S, \rho')$ has a solution where*

$$\rho'(i) = \begin{cases} \rho(i) & \text{if } \rho(i) \neq j, \\ k & \text{if } \rho(i) = j. \end{cases}$$

**Proof of Lemma 1.** Let $(u_t)_{t \in T_u}, (h_{t,v})_{t \in T_p, v \in T_u}$ be a solution of $FP(T_u, S, \rho)$ with $u_k = u_j$.

Define $(\tilde{u}_t)_{t\in T_u-\{j\}}, (\tilde{h}_{t,v})_{t\in T_p, v\in T_u-\{j\}}$ in the following way:

$$\begin{array}{lll} \tilde{h}_{t,v} & \equiv h_{t,v} & \text{if } \rho(t)\neq j \text{ or } v\neq k, \\ \tilde{h}_{t,v} & \equiv q_t & \text{if } \rho(t)=j \text{ and } v=k, \\ \tilde{u}_v & \equiv u_v & \forall v\in T_u-\{j\}. \end{array}$$

Let us show that this provides a solution for $FP(T_u-\{j\}, S, \rho')$. For the first condition, let $t\in T_p$. If $\rho(t)\neq j$ then $h_{t,\rho'(t)} = h_{t,\rho(t)} = q_t$, as was to be shown. If $\rho(t)=j$ then $h_{t,\rho'(t)} = h_{t,k} = q_t$, as was to be shown.

For the second condition, let $t,v\in T_u-\{j\}$ and assume that $\tilde{u}_t \geq \tilde{u}_v$, i.e. $u_t\geq u_v$. Take $r,s\in T_p$. There are four cases.

- $(\rho(r)\neq j$ or $v\neq k)$ and $(\rho(s)\neq j$ or $t\neq k)$. Then,

$$p_r\tilde{h}_{r,v}\leq p_r\tilde{h}_{s,t}\Leftrightarrow p_r h_{r,v}\leq p_r h_{s,t},$$

as was to be shown.

- $(\rho(r)=j$ and $v=k)$ and $(\rho(s)\neq j$ or $t\neq k)$. Then,

$$p_r\tilde{h}_{r,k}\leq p_r\tilde{h}_{s,t}\Leftrightarrow p_r q_r\leq p_r h_{s,t}\leftrightarrow p_r h_{r,j}\leq p_r h_{s,t}.$$

This holds as $u_t\geq u_v=u_k=u_j$.

- $(\rho(r)\neq j$ or $v\neq k)$ and $(\rho(s)=j$ and $t=k)$. Then,

$$p_r\tilde{h}_{r,v}\leq p_r\tilde{h}_{s,k}\Leftrightarrow p_r h_{r,v}\leq p_r q_s\Leftrightarrow p_r h_{r,v}\leq p_r h_{s,j}.$$

This holds as $u_t=u_k=u_j\geq u_v$.

- $(\rho(r)=j$ and $v=k)$ and $(\rho(s)=j$ and $t=k)$. Then, $u_t=u_k=u_j=u_v$ and

$$p_r\tilde{h}_{r,k}\leq p_r\tilde{h}_{s,k}\Leftrightarrow p_r q_r\leq p_r q_s\Leftrightarrow p_r h_{r,j}\leq p_r h_{s,j}.$$

This holds as $u_j\geq u_j$.

Replacing the weak inequalities by strict inequalities shows that the third condition is satisfied. For the last condition, let $\tilde{u}_t\geq\tilde{u}_v$, i.e. $u_t\geq u_v$. Let $i\in M$ and $r\in T_p$. If $\rho(r)\neq j$ or $(t\neq k$ and $v\neq k)$ then,

$$\tilde{h}^i_{r,v}\leq\tilde{h}^i_{r,t}\Leftrightarrow h^i_{r,v}\leq h^i_{r,v},$$

as was to be shown. If $\rho(r)=j$ and $t=k$ but $v\neq k$, then

$$\tilde{h}^i_{r,v}\leq\tilde{h}^i_{r,k}\Leftrightarrow h^i_{r,v}\leq q^i_r\Leftrightarrow h^i_{r,v}\leq h^i_{r,j}.$$

This holds as $u_t = u_k = u_j \geq u_v$. If $\rho(r) = j$ and $t \neq k$ but $v = k$, then

$$\tilde{h}^i_{r,k} \leq \tilde{h}^i_{r,t} \Leftrightarrow q^i_r \leq h^i_{r,t} \Leftrightarrow h^i_{r,j} \leq h^i_{r,t}.$$

This holds as $u_t \geq u_v = u_k = u_j$. Finally, we have the case that $\rho(r) = j$ and $t = v = k$, but then $\tilde{h}^i_{r,t} = \tilde{h}^i_{r,k} = \tilde{h}^i_{r,v}$ so this case is obviously satisfied. ■

LEMMA 2: *If $p_t q_t(x) \geq p_t q_v(y)$, then the level of $q_v(y)$ is not strictly higher than the level of $q_t(x)$.*

**Proof of Lemma 2.** Let $q_v(y)$ be of level $r$ and $q_t(x)$ be of level $s$. Assume (towards a contradiction) that Lemma 2 does not hold, that is, $r > s$. Then,

- If $r(= R+1) > s(= 1)$, then $p_t h_{t,1} \leq p_t h_{v,R}$, so

$$p_t q_t(x) \leq p_t h_{t,1} \leq p_t h_{v,R} < p_t q_v(y),$$

  a contradiction.

- If $r(= R+1) > s > 1$, then $p_t h_{t,s} \leq p_t h_{v,R}$ and $p_t q_{t,s-1} < p_t q_{v,R}$. As such, if $q_t(x) = \alpha h_{t,s} + (1-\alpha) h_{t,s-1}$ with $\alpha \in (0, 1]$, then

$$p_t q_t(x) = \alpha(p_t h_{t,s}) + (1-\alpha)(p_t h_{t,s-1}) \leq p_t h_{v,R} < p_t q_v(y),$$

  a contradiction.

- If $R + 1 > r > s = 1$, then $p_t h_{t,1} \leq p_t h_{v,r-1}$ and $p_t h_{t,1} < p_t h_{v,r}$. As such, if $q_v(y) = \beta h_{v,r} + (1-\beta) h_{v,r-1}$ with $\beta \in (0, 1]$, then

$$p_t q_t(x) \leq p_t h_{t,1} < \beta p_t h_{v,r} + (1-\beta) p_t h_{v,r-1} = q_v(y).$$

- If $R + 1 > r > s > 1$, then $p_t h_{t,s} \leq p_t h_{v,r-1}$, $p_t h_{t,s} < p_t h_{v,r}$, $p_t h_{t,s-1} < p_t h_{v,r-1}$ and $p_t h_{t,s-1} < p_t h_{v,r}$. This implies that any convex combination of $p_t h_{t,s}$ and $p_t h_{t,s-1}$ must always be strictly smaller than any convex combination of $p_t h_{v,r-1}$ and $p_t h_{v,r}$. As such, if $q_t(x) = \alpha h_{t,s} + (1-\alpha) h_{t,s-1}$ and $q_v(y) = \beta h_{v,r} + (1-\beta) h_{v,r-1}$ with $\alpha, \beta \in (0, 1]$, then

$$\begin{aligned} p_t q_t(x) &= \alpha p_t h_{t,s} + (1-\alpha) p_t h_{t,s-1} \\ &\leq \beta p_t h_{v,r} + (1-\beta) p_t h_{v,r-1} = p_t q_v(y), \end{aligned}$$

  a contradiction.

■

LEMMA 3: *Let $p_t q_t(x) \geq p_t q_v(y)$, with the level of $q_t(x)$ the same as the level of $q_v(y)$. Then:*

- *If both $q_t(x)$ and $q_v(y)$ are of level $R+1$, and $q_t(x) = \gamma h_{t,R}, q_v(y) = \delta h_{v,R}$, we have $\gamma \geq \delta$. In addition, if $p_t q_t(x) > p_t q_v(y)$, then $\gamma > \delta$.*
- *If both $q_t(x)$ and $q_v(y)$ are of level $1$, and $q_t(x) = \gamma h_{t,1}, q_v(y) = \delta h_{v,1}$, we have $\gamma \geq \delta$. In addition, if $p_t q_t(x) > p_t q_v(y)$, then $\gamma > \delta$.*
- *If both $q_t(x)$ and $q_v(y)$ are of level $r$ with $1 < r < R+1$, and $q_t(x) = \alpha h_{t,r} + (1-\alpha) h_{t,r-1}$, $q_v(y) = \beta h_{v,r} + (1-\beta) h_{v,r-1}$ with $\alpha, \beta \in (0,1]$, then we have $\alpha \geq \beta$. In addition, if $p_t q_t(x) > p_t q_v(y)$, then $\alpha > \beta$.*

**Proof of Lemma 3.** We look at the three cases separately:

Suppose that both $q_t(x)$ and $q_v(y)$ are of level $R+1$. From the second N-GARP condition in Definition 5, we know that $p_t h_{t,R} \leq p_t h_{v,R}$. This implies

$$\begin{aligned}
\delta p_t h_{v,R} &= p_t q_v(y) \\
&\leq p_t q_t(x) = \gamma p_t h_{t,R} \\
&\leq \gamma p_t h_{v,R}.
\end{aligned}$$

So, $\delta \leq \gamma$ with a strict inequality if $p_t q_t(x) > p_t q_v(y)$.

Suppose that both $q_t(x)$ and $q_v(y)$ are of level 1. From the second N-GARP condition in Definition 5, we know that $p_t h_{t,1} \leq p_t h_{v,1}$. This implies

$$\begin{aligned}
\delta p_t h_{v,1} &= p_t q_v(y) \\
&\leq p_t q_t(x) = \gamma p_t h_{t,1} \\
&\leq \gamma p_t h_{v,1}.
\end{aligned}$$

So, $\delta \leq \gamma$ with a strict inequality if $p_t q_t(x) > p_t q_v(y)$.

Suppose that both $q_t(x)$ and $q_v(y)$ are of level $r$ with $R+1 > r > 1$. From the second N-GARP condition in Definition 5, we know that $p_t h_{t,r} \leq p_t h_{v,r}$ and $p_t h_{t,r-1} \leq p_t h_{v,r-1}$. As such,

$$\begin{aligned}
\alpha(p_t h_{t,r}) + (1-\alpha)(p_t h_{t,r-1}) &= p_t q_t(x) \\
&\geq p_t q_v(y) \\
&= \beta(p_t h_{v,r}) + (1-\beta)(p_t h_{v,r-1}) \\
&\geq \beta p_t h_{t,r} + (1-\beta) p_t h_{t,r-1}.
\end{aligned}$$

This is equivalent to the condition $(\alpha - \beta)(p_t h_{t,r} - p_t h_{t,r-1}) \geq 0$. The third N-GARP condition in Definition 5 implies that $p_t h_{t,r} > p_t h_{t,r-1}$. As such, it must be that $\alpha \geq \beta$, with a strict inequality if $p_t q_t(x) > p_t q_v(y)$. ■

## II. Practical implementation

Mixed integer programming formulation of N-GARP. — The N-GARP conditions in Definition 5 can be reformulated in terms of linear inequalities that are characterized by (binary) integer variables.

PROPOSITION 2: *A data set $S = \{(p_t, q_t)\}_{t \in T}$ satisfies the N-GARP conditions in Definition 5 if and only if there exist binary numbers $r_{t,v} \in \{0,1\}$ vectors $h_{t,v} \in \mathbb{R}^n_+$, and numbers $u_t \in [0,1]$ such that, for all $r, s, t, v \in T$,*

- $h_{t,t} = q_t$,
- $u_t - u_v < r_{t,v}$,
- $(r_{v,t} - 1) < u_v - u_t$,
- $p_r h_{r,v} - p_r h_{s,t} < r_{v,t} A$,
- $A(r_{t,v} - 1) \leq (p_r h_{s,t} - p_r h_{r,v})$,
- $B(r_{t,v} - 1) \leq h^i_{r,t} - h^i_{r,v}$ *for all* $i \in M$.

*where $A$ is a fixed number greater than any possible value $p_r h_{r,v} (r, v \in T)$ and $B$ is a fixed number greater than any $h^i_{r,v} (i \in M, r, v \in T)$. By default $A$ and $B$ are finite numbers.*

**Proof of Proposition 2.** **Necessity.** Assume that the N-GARP conditions in Definition 5 are satisfied. Let us use the same solution and define $r_{t,v} = 1$ if and only if $u_t \geq u_v$. The the first three conditions above are satisfied by default. By the definition of $A$, the fourth condition is only binding if $r_{v,t} = 0$, which means that $u_t > u_v$. In this case, Definition 5 implies that $p_r h_{r,v} < p_r h_{s,t}$ and the condition holds. Similarly, the fifth condition is binding only if $r_{t,v} = 1$, which implies that $u_t \geq u_v$ and thus that $p_r h_{s,t} \geq p_r h_{r,v}$. Finally, the last condition only binds if $r_{t,v} = 1$, which implies that $u_t \geq u_v$, In this case the last condition of Definition 5 gives $h^i_{r,v} \leq h^i_{r,t}$. We can thus conclude that the conditions of Proposition 2 are feasible whenever Definition 5 is satisfied.
**Sufficiency.** Assume that there exists a solution for the conditions in Proposition 2. Then we can show that the conditions in Definition 5 are also satisfied for the same solution. The first condition in Definition 5 is satisfied by default. For the second condition, if $u_t \geq u_v$ then $r_{t,v} = 1$ by the second condition above and as such the fifth condition implies that $p_r h_{s,t} \geq p_r h_{r,v}$. This shows that the second condition of Definition 5 holds. Next, let $u_t > u_v$. If, towards a contradiction, $p_r h_{r,v} \geq p_r h_{s,t}$, then, by the fourth condition above, $r_{v,t} = 1$. This implies, by the third condition, that $u_v \geq u_t$, a contradiction. This shows that the third condition of Definition 5 holds. For the final condition, let $u_t \geq u_v$. Then, by the second condition above, $r_{t,v} = 1$ and, by the last condition, $h^i_{r,t} \geq h^i_{r,v}$, as was to be shown. ■

COMPUTING THE CCEI. — The CCEI is found by solving the following optimization problem:

$$
\begin{aligned}
&\max \; e \\
\text{s.t. } & 0 \leq e \leq 1 \\
& \forall t \in T : 0 \leq u_t \leq 1 \\
& \forall t \in T : h_{t,t} = q_t \\
& \forall t, v, r, s \in T \text{ such that } r \neq v : u_t \geq u_v \rightarrow p_r h_{r,v} \leq p_r h_{s,t} \\
& \forall t, v, r, s \in T \text{ such that } r \neq v : u_t > u_v \rightarrow p_r h_{r,v} < p_r h_{s,t} \\
& \forall i \in M, \forall t, v, r \in T, \text{ such that } r \neq v : u_t \geq u_v \rightarrow h^i_{r,v} \leq h^i_{r,t} \\
& \forall t, v, r, s \in T \text{ such that } r = v : u_t \geq u_v \rightarrow e p_r q_r \leq p_r h_{s,t} \\
& \forall t, v, r, s \in T \text{ such that } r = v : u_t > u_v \rightarrow e p_r q_r < p_r h_{s,t} \\
& \forall i \in M, \forall t, v, r \in T \text{ such that } r = v : u_t \geq u_v \rightarrow e q^i_r \leq h^i_{r,t}.
\end{aligned}
$$

The if–then conditions can be reformulated in terms of linear restrictions with binary variables, following our reasoning leading up to Proposition 2. As a result, the above optimization problem can be reformulated as a mixed integer linear programming problem.

## III. Data

Table 1 provides a summary of the data set that we use in our empirical application. As explained in the main text, we assume that the individuals spend their full potential incomes on four different consumption categories: leisure, food, housing and other goods. Table 1 reports information on prices, quantities, incomes and some demographics for our sample of 821 singles.

We compute leisure quantities by assuming that each individual needs 8 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)*7 - market work). Food expenditures include food at home, delivered and eaten away from home. Housing expenditures include mortgage and loan payments, rent, property tax, insurance, utilities, cable tv, telephone, internet charges, home repairs and home furnishing. Others expenditures include health, transportation, education and childcare. 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.

.

Table 1—Summary statistics

| | mean | std. dev. | min | max |
|---|---|---|---|---|
| age in 2007 | 37.95 | 13.38 | 18.00 | 81.00 |
| is male | 0.34 | 0.47 | 0.00 | 1.00 |
| has home in 2007 | 0.36 | 0.48 | 0.00 | 1.00 |
| has children | 0.31 | 0.46 | 0.00 | 1.00 |
| number of children in 2007 | 0.54 | 0.96 | 0.00 | 6.00 |
| years of education in 2007 | 13.53 | 2.10 | 6.00 | 17.00 |
| quantity food in 2011 | 0.43 | 0.27 | 0.00 | 1.99 |
| quantity food in 2009 | 0.41 | 0.26 | 0.00 | 2.13 |
| quantity food in 2007 | 0.44 | 0.30 | 0.00 | 2.25 |
| quantity housing in 2011 | 1.20 | 2.06 | 0.00 | 56.28 |
| quantity housing in 2009 | 1.08 | 0.69 | 0.00 | 7.06 |
| quantity housing in 2007 | 1.17 | 1.38 | 0.00 | 22.60 |
| quantity other in 2011 | 0.72 | 0.66 | 0.00 | 6.94 |
| quantity other in 2009 | 0.82 | 1.24 | 0.00 | 22.86 |
| quantity other in 2007 | 0.82 | 0.75 | 0.00 | 6.03 |
| quantity leisure in 2011 | 71.35 | 11.00 | 16.00 | 111.00 |
| quantity leisure in 2009 | 72.98 | 10.12 | 22.00 | 111.00 |
| quantity leisure in 2007 | 70.31 | 12.15 | 12.00 | 105.00 |
| price food in 2011 | 226.53 | 4.00 | 220.43 | 233.20 |
| price food in 2009 | 217.00 | 4.35 | 211.09 | 224.35 |
| price food in 2007 | 201.09 | 4.44 | 195.48 | 207.76 |
| price housing in 2011 | 213.27 | 17.03 | 199.98 | 248.68 |
| price housing in 2009 | 211.90 | 16.48 | 197.21 | 243.76 |
| price housing in 2007 | 204.13 | 15.99 | 193.38 | 236.24 |
| price other in 2011 | 238.61 | 2.58 | 235.89 | 241.36 |
| price other in 2009 | 209.32 | 3.98 | 205.15 | 214.13 |
| price other in 2007 | 205.29 | 2.57 | 202.62 | 208.21 |
| price leisure in 2011 | 20.55 | 17.58 | 0.50 | 180.85 |
| price leisure in 2009 | 19.66 | 15.32 | 2.05 | 165.52 |
| price leisure in 2007 | 16.46 | 11.95 | 2.15 | 149.29 |
| expenditures in 2007 | 1649.61 | 1070.51 | 289.31 | 13231.01 |
| expenditures in 2009 | 1919.20 | 1294.08 | 245.85 | 13179.54 |
| expenditures in 2011 | 1973.16 | 1442.99 | 181.25 | 13235.89 |
| full potential income 2007 | 1842.97 | 1338.09 | 240.80 | 16720.48 |
| full potential income 2009 | 2202.09 | 1715.59 | 229.60 | 18538.24 |
| full potential income 2011 | 2301.66 | 1968.76 | 56.00 | 20255.20 |
| nonlabor income 2007 | -193.36 | 513.20 | -3489.47 | 4213.92 |
| nonlabor income 2009 | -282.88 | 617.01 | -5358.70 | 4887.96 |
| nonlabor income 2011 | -328.50 | 802.93 | -7999.98 | 10699.09 |

## IV. Additional empirical results

In this appendix, we first provide several robustness checks of our empirical results discussed in Section 4 of the main text. These checks largely confirm our principal conclusions. In a following step, we conduct a regression analysis that relates our estimated cost of living indices to observable individual characteristics. This provides an (exploratory) investigation of who has been affected by the 2008 crisis. To avoid an overload of empirical results, we only present the results for N-GARP(3).

Cost of living indices. — As a first robustness check, Table 2 summarizes our N-GARP(3)-based and GARP-based estimated bounds on $c_{2011,2007}$ for the 587 individuals whose behavior is exactly rationalizable under normal demand (i.e., N-GARP(3)-based CCEI equals 1). We observe that the results are closely similar to the ones contained in Table 3 in the main text.

Table 2—Bounds on $c_{2011,2007}$ for individuals with N-GARP(3)-based CCEI = 1

| | N-GARP(3)-based | | | GARP-based | | | |
|---|---|---|---|---|---|---|---|
| | min | max | $\Delta_n$ | min | max | $\Delta_g$ | $\frac{\Delta_g-\Delta_n}{\Delta_g}$ |
| mean | -0.048 | 0.029 | 0.077 | -0.049 | 0.099 | 0.148 | 0.425 |
| std. dev. | 0.308 | 0.262 | 0.138 | 0.308 | 0.292 | 0.176 | 0.379 |
| min | -3.044 | -2.492 | 0.000 | -3.044 | -2.489 | 0.001 | 0.000 |
| 25% | -0.128 | -0.050 | 0.012 | -0.128 | 0.000 | 0.042 | 0.000 |
| 50% | -0.012 | 0.000 | 0.035 | -0.012 | 0.000 | 0.095 | 0.402 |
| 75% | 0.082 | 0.137 | 0.085 | 0.082 | 0.249 | 0.196 | 0.821 |
| max | 0.830 | 0.897 | 2.099 | 0.830 | 0.899 | 2.285 | 0.998 |

Better-off and worse-off individuals. — As a following robustness check of our results in Section 4, we consider the classification of worse-off, better-off and cannot-say individuals for two alternative scenarios: the first scenario uses the N-GARP(3)-based and GARP-based classifications for the 587 individuals of which the N-GARP(3)-based CCEI equals 1 (also included in Table 2); the second scenario uses the GARP-based classification for the 782 individuals whose behavior is exactly rationalizable when not imposing normality on any good (i.e., GARP-based CCEI equals 1).

The results for the two scenarios are summarized in Table 3. Comfortingly, we find that the results in Table 3 are generally close to the ones in Table 5 that we discuss in the main text. Again, it suggests that our main qualitative conclusions are robust.

TABLE 3—WORSE-OFF AND BETTER-OFF INDIVIDUALS FOR INDIVIDUALS WITH N-GARP-BASED CCEI=1 AND GARP-BASED CCEI=1

| | | N-GARP-CCEI=1 (587 individuals) | | GARP-CCEI=1 (782 individuals) |
|---|---|---|---|---|
| | | N-GARP | GARP | GARP |
| UB $< 0$ | Worse off in 2011 | 33.39 | 22.66 | 22.38 |
| LB $> 0$ | Better off in 2011 | 45.32 | 44.97 | 48.59 |
| LB $\leq 0$ and $0 \leq$ UB | Cannot say | 21.29 | 32.37 | 29.03 |

FOUR PSID WAVES: 2007, 2009, 2011 AND 2013. — Next, we check robustness of our main findings for a longer panel containing four consumption observations per individual (adding the 2013 PSID wave to our original data set). The following Tables 4, 5 and 6 have a directly analogous interpretation as the Tables 1, 3 and 5 that we discussed in the main text.

Generally, we can conclude that the results in Tables 4, 5 and 6 are fairly close to those in Tables 1, 3 and 5. For our application, adding a consumption observation (i.e., PSID wave) per individual only moderately affects our goodness-of-fit and cost of living results.

TABLE 4—CRITICAL COST EFFICIENCY INDEX (CCEI); 4 WAVES

| | N-GARP(3) | GARP |
|---|---|---|
| CCEI=1 | 376 (54.26%) | 632 (91.20%) |
| CCEI $\geq$ 0.99 | 495 (71.43%) | 666 (96.10%) |
| mean | 0.9789 | 0.9975 |
| std. dev. | 0.0492 | 0.0160 |
| min | 0.6235 | 0.7456 |
| 25% | 0.9849 | 1.0000 |
| 50% | 1.0000 | 1.0000 |
| 75% | 1.0000 | 1.0000 |
| max | 1.0000 | 1.0000 |

WHO IS AFFECTED BY THE CRISIS ?. — Generally, our cost of living estimates reveal quite some heterogeneity across individuals. In what follows, we investigate this further by relating the N-GARP(3)-based cost of living estimates to observable individual characteristics. This can provide additional insight into which types of individuals (on the intensive margin of labor supply) were particularly hit by

Table 5—Bounds on $c_{2011,2007}$; 4 waves

| | N-GARP(3) | | | GARP | | | |
|---|---|---|---|---|---|---|---|
| | min | max | $\Delta_n$ | min | max | $\Delta_g$ | $\frac{\Delta_g - \Delta_n}{\Delta_g}$ |
| mean | -0.071 | 0.016 | 0.087 | -0.073 | 0.084 | 0.156 | 0.458 |
| std. dev. | 0.519 | 0.279 | 0.404 | 0.519 | 0.302 | 0.441 | 0.379 |
| min | -9.450 | -2.514 | 0.000 | -9.450 | -2.489 | 0.000 | 0.000 |
| 25% | -0.126 | -0.065 | 0.008 | -0.126 | -0.008 | 0.036 | 0.054 |
| 50% | -0.010 | 0.001 | 0.029 | -0.012 | 0.005 | 0.084 | 0.458 |
| 75% | 0.081 | 0.122 | 0.076 | 0.078 | 0.226 | 0.196 | 0.832 |
| max | 0.831 | 0.838 | 8.735 | 0.830 | 0.900 | 9.371 | 1.000 |

Table 6—Worse-off and better-off individuals; 4 waves

| classification by bounds of: | | N-GARP(3) | GARP |
|---|---|---|---|
| UB $\leq 0$ | Worse off in 2011 | 38.59 | 27.47 |
| LB $\geq 0$ | Better off in 2011 | 46.67 | 45.66 |
| LB $\leq 0$ and $0 \leq$ UB | Cannot-say | 14.75 | 26.87 |

the crisis. We conduct three regression exercises: our first exercise uses interval regression and explicitly takes the (difference between) lower and upper bounds into account, our second exercise is a simple OLS regression that uses the average of the lower and upper bounds as the dependent variable, and our last exercise is a logit regression that explains the probability of being better-off (versus worse-off) after the 2008 crisis (using our N-GARP(3)-based classification as worse-off or better-off to define the dependent variable). Further, to distinguish between short-run and longer-run effects of the crisis, we ran our regressions for two cost of living indices: $c_{2009,2007}$ (capturing the short run effect) and $c_{2011,2007}$ (capturing the longer run effect). We use the N-GARP(3)-based bound estimates for the 702 individuals with a CCEI-value at least equal to 0.99 (with $c_{2011,2007}$-values summarized in Table 5).

Table 7 summarizes our findings. We see that individuals with higher labor incomes (i.e., wages) and nonlabor incomes are generally associated with lower cost of living indices, and are less likely to be better off in both the short run and the longer run when compared to their pre-crisis utility level. Next, while we find no significant short run effect related to region of residence (captured by the dummy variables North Central, South and West, using North East as the reference category) or industry (captured by the dummy variables construction and services), we do see that individuals residing in the West region are generally worse off in the longer run, while the opposite holds true for individuals working in the service sector.

Next, we observe that many individual characteristics that are statistically significant in the short run become insignificant in the longer run. For example, homeowners and single parents are better off than non-home owners and childless singles in the short run. However, these effects fade out in the longer run. Similarly, being a single male parent corresponds to a significantly negative crisis effect in the short run, but this effect disappears in the longer run.

Table 7—Welfare effects and individual characteristics

| | interval | | simple OLS | | logit | |
|---|---|---|---|---|---|---|
| | $c_{2009,2007}$ | $c_{2011,07}$ | $c_{2009,2007}$ | $c_{2011,2007}$ | $c_{2009,2007}$ | $c_{2011,2007}$ |
| full potential income | -0.0001*** | -0.0001*** | -0.0001*** | -0.0001*** | -0.0011*** | -0.0013*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| nonlabor income | -0.0004*** | -0.0004*** | -0.0004*** | -0.0004*** | -0.0039*** | -0.0045*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.001) | (0.001) |
| years of education | 0.0030 | 0.0024 | 0.0027 | 0.0029 | 0.0524 | 0.0182 |
| | (0.004) | (0.004) | (0.004) | (0.005) | (0.058) | (0.064) |
| North Central | -0.0247 | -0.0341 | -0.0215 | -0.0334 | -0.5653 | -0.8655* |
| | (0.027) | (0.022) | (0.028) | (0.023) | (0.409) | (0.419) |
| South | -0.0091 | -0.0080 | -0.0095 | -0.0086 | -0.2440 | -0.5950 |
| | (0.025) | (0.021) | (0.026) | (0.022) | (0.379) | (0.398) |
| West | -0.0322 | -0.0663* | -0.0301 | -0.0616* | -0.7972 | -1.0248* |
| | (0.029) | (0.027) | (0.030) | (0.027) | (0.426) | (0.445) |
| home owner | 0.0312* | 0.0153 | 0.0324* | 0.0193 | 0.3120 | 0.2324 |
| | (0.015) | (0.016) | (0.016) | (0.017) | (0.243) | (0.245) |
| male | 0.0153 | 0.0003 | 0.0155 | 0.0006 | 0.0422 | -0.2751 |
| | (0.017) | (0.017) | (0.018) | (0.018) | (0.261) | (0.263) |
| with children | 0.0647*** | 0.0367 | 0.0665*** | 0.0327 | 0.6361* | 0.2555 |
| | (0.018) | (0.020) | (0.018) | (0.021) | (0.279) | (0.279) |
| male × with children | -0.1448** | -0.0004 | -0.1454** | 0.0075 | -1.5709 | 0.0422 |
| | (0.045) | (0.061) | (0.046) | (0.062) | (0.867) | (0.885) |
| age | 0.0008 | -0.0006 | 0.0009 | -0.0006 | 0.0107 | -0.0137 |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.009) | (0.008) |
| worked in construction industry | -0.0039 | -0.0130 | -0.0003 | -0.0049 | -0.1187 | -0.1568 |
| | (0.023) | (0.029) | (0.024) | (0.031) | (0.392) | (0.378) |
| worked in service industry | 0.0227 | 0.0294 | 0.0251 | 0.0355* | 0.1404 | 0.1542 |
| | (0.015) | (0.015) | (0.016) | (0.018) | (0.228) | (0.233) |
| Constant | 0.0416 | 0.1149 | 0.0412 | 0.1041 | 0.7332 | 2.7100** |
| | (0.065) | (0.069) | (0.067) | (0.074) | (0.963) | (0.957) |
| Observations | 628 | 628 | 628 | 628 | 510 | 501 |
| $R^2$ | | | 0.415 | 0.437 | | |

*Note:* All variables are observations in year 2007. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 8 shows pairwise correlation coefficients between wages in 2007 and relative changes in wages, leisure hours, expenditures on leisure, expenditures on food, housing expenditures and other expenditures (measuring the relative change in variable $y$ as $\frac{y_{11}-y_{07}}{y_{07}}$). We see that people with higher initial wages (in 2007) generally experienced larger wage drops (and thus income drops) than people

TABLE 8—PAIRWISE CORRELATION COEFFICIENTS

| | | a | b | c | d | e | f | g |
|---|---|---|---|---|---|---|---|---|
| wage in 2007 | a | **1** | | | | | | |
| relative increase in wage | b | **-0.2296** | **1** | | | | | |
| | | 0 | | | | | | |
| relative increase in leisure hours | c | -0.0298 | **0.0782** | **1** | | | | |
| | | 0.4303 | 0.0383 | | | | | |
| relative increase in leisure expenditures | d | **-0.1761** | **0.9002** | **0.3932** | **1** | | | |
| | | 0 | 0 | 0 | | | | |
| relative increase in food expenditures | e | -0.0556 | 0.0487 | 0.0159 | 0.0376 | **1** | | |
| | | 0.144 | 0.2008 | 0.6765 | 0.3232 | | | |
| relative increase in house expenditures | f | -0.0429 | 0.0205 | 0.0357 | 0.0182 | 0.0195 | **1** | |
| | | 0.2572 | 0.5883 | 0.3457 | 0.6301 | 0.609 | | |
| relative increase in other expenditures | g | **-0.0669** | 0.0146 | 0.0236 | 0.0147 | 0.0553 | 0.0129 | **1** |
| | | 0.0789 | 0.7016 | 0.5352 | 0.7003 | 0.149 | 0.7356 | |

*Note:* Bold represents significant correlation (at 10%)

with lower initial wages. This explains the negative regression coefficient for the initial full income in Table 7: if a higher initial potential income corresponds to a greater loss in total income, it is also associated with a more pronounced utility loss.

## REFERENCES

**Nishimura, Hiroki, Efe A Ok, and John K-H Quah.** 2017. "A Comprehensive Approach to Revealed Preference Theory." *The American Economic Review*, 107: 1239–1263.